ARTICLES
https://doi.org/10.1038/s41559-021-01485-1

nature ecology & evolution

Check for updates

The global distribution and environmental drivers of aboveground versus belowground plant biomass

Haozhi Ma[1,5], Lidong Mo[1,5], Thomas W. Crowther[1], Daniel S. Maynard[1], Johan van den Hoogen[1], Benjamin D. Stocker[1,2], César Terrer[3,4] and Constantin M. Zohner[1]✉

A poor understanding of the fraction of global plant biomass occurring belowground as roots limits our understanding of present and future ecosystem function and carbon pools. Here we create a database of root-mass fractions (RMFs), an index of plant below- versus aboveground biomass distributions, and generate quantitative, spatially explicit global maps of RMFs in trees, shrubs and grasses. Our analyses reveal large gradients in RMFs both across and within vegetation types that can be attributed to resource availability. High RMFs occur in cold and dry ecosystems, while low RMFs dominate in warm and wet regions. Across all vegetation types, the directional effect of temperature on RMFs depends on water availability, suggesting feedbacks between heat, water and nutrient supply. By integrating our RMF maps with existing aboveground plant biomass information, we estimate that in forests, shrublands and grasslands, respectively, 22%, 47% and 67% of plant biomass exists belowground, with a total global belowground fraction of 24% (20–28%), that is, 113 (90–135) Gt carbon. By documenting the environmental correlates of root biomass allocation, our results can inform model projections of global vegetation dynamics under current and future climate scenarios.

The extent and distribution of carbon storage in plants is central to our understanding of the structure and function of the terrestrial biosphere[1,2]. At present, our understanding of global plant biomass stems primarily from satellite and inventory data that characterize the aboveground components of ecosystems[1,3–6]. However, a major component of plant biomass cannot be detected by satellites or direct aboveground observations because it exists belowground as roots, which are central to the functioning of terrestrial ecosystems[7] and greatly affect ecosystem nutrient, water and carbon cycling[8–11]. Plant investment in root biomass has generally been explored using local-scale observations[12,13], such that variation at regional and global scales remains unknown. Without empirical, spatially explicit models of the relative proportion of plant biomass distributed to roots (root-mass fraction, RMF), aboveground biomass cannot be adequately scaled to total plant biomass[6,14,15].

At the global scale, patterns of biomass fractions in roots, wood and leaves are expected to reflect a functional balance between the distribution of resources that exist above (light, CO_2) and below (water, nutrient) the soil surface[16,17]. Indeed, plants tend to show proportionally higher biomass allocation to roots when belowground resources are scarce[16–20]. Because belowground limitations should be most pronounced in cold regions (where nutrient supply is low) and in dry environments[13,21–23], RMFs are expected to increase with latitude[24] and in arid regions with low water availability[13,22]. Given such relationships, changing climate and associated shifts in above- versus belowground resource availability may alter relative plant investments in root biomass. Due to large variations in the longevity of different plant tissues, such shifts drive feedbacks between climate and the terrestrial carbon cycle[25]. Yet, despite the central role of biomass allocation in shaping vegetation responses to altered environmental conditions, we lack a quantitative, spatially explicit understanding of the factors driving RMF[13,19,26], which is essential for improving confidence in future carbon cycle projections[25]. Characterizing the environmental drivers of the variation in RMF across the globe is thus critical for projecting plant biomass accumulation[23], productivity[27] and nutrient cycling[11] under current or future climate change scenarios[6].

Here we use a large dataset of plant RMFs to generate a spatially explicit understanding of the proportional plant biomass that exists in roots and upscale these predictions to the globe to quantify total belowground plant biomass. First, we constructed a database of 17,814 plot-level RMF measurements, aggregated to 6,803 spatially distinct samples from forest ($n = 5{,}170$), grassland ($n = 1{,}293$), and shrubland ($n = 340$) vegetation (Fig. 1). Using 63 global layers of climate, topography, soil and vegetation features (Extended Data Fig. 1), we then trained random forest models to characterize the environmental factors that drive geographic variations in RMFs within each vegetation type and interpolate these relationships across the globe. To comprehend the interactive mechanisms that have ultimately shaped variations in biomass partitioning within and across regions, we additionally use recursive partitioning and regression methods accounting for spatial autocorrelation. Finally, we compare our empirical RMF data with vegetation model simulations to pinpoint important deviations in model predictions versus observed biomass fractions.

Global mapping of RMFs and model performance

There was considerable variation in RMF within each vegetation type (Fig. 2a–c), which can rival the amount of variation observed across types. Within forests, the highest RMFs are found in boreal regions and seasonally dry tropical and subtropical regions, such as the Mediterranean region, east Brazil, California, and Southern Africa (Fig. 3a). Within grasslands, RMFs in tropical areas are, on average, ~50% lower than in temperate environments, and the

[1]Institute of Integrative Biology, ETH Zurich (Swiss Federal Institute of Technology), Zurich, Switzerland. [2]Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Birmensdorf, Switzerland. [3]Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, CA, USA. [4]Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA. [5]These authors contributed equally: Haozhi Ma, Lidong Mo. ✉e-mail: constantin.zohner@gmail.com

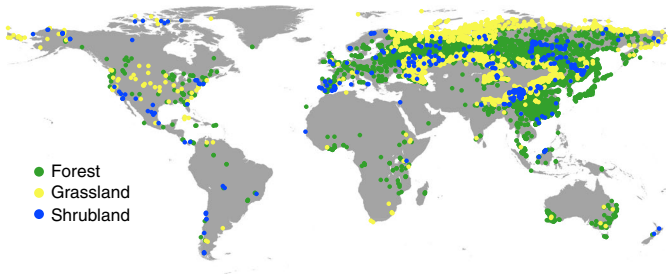


Fig. 1 | RMF sample locations in forests, shrublands and grasslands. A total of 17,814 RMF records (13,620 forest, 2,468 grassland and 1,726 shrubland) were collected and aggregated into 6,893 spatially unique data points used for geospatial modelling (5,170 forest, 1,293 grassland and 340 shrubland records; Methods).

highest RMFs are found in montane and temperate steppe regions with low temperature and soil moisture, such as the Qinghai-Tibetan and Mongolian plateaus (Fig. 3c). In shrublands, RMFs are highest in the boreal and arid biomes (Fig. 3e), while low RMFs occur in flooded savannas and tropical and subtropical ecosystems.

Based on a spatial validation procedure (spatially buffered leave-one-out cross-validation, LOO-CV), our random forest models explained 16–39% of the spatial variation in RMFs within vegetation types (LOO-CV $R^2 = 0.16$, 0.34 and 0.39 for forests, shrublands and grasslands, respectively; Fig. 2d–f; Extended Data Fig. 2a–c shows random tenfold cross-validation results). We used LOO-CV to account for the potential effects of spatial autocorrelation on model validation statistics[28], with buffer radii of 250 km in forests and shrublands and 500 km in grasslands. These buffer ranges were based on the extent of autocorrelation in the model residuals (Extended Data Fig. 2g–i shows residual semivariograms and Fig. 2g–i shows LOO-CV R^2s for buffer radii from 1 to 500 km).

We then used these predictive relationships to spatially upscale observed RMFs over existing forest, shrubland and grassland areas across the globe to reveal the global distribution of RMFs for each vegetation type and combined across all existing vegetation (excluding cropland) (Fig. 3). Given the considerable variability in these pixel-level RMF estimates, the predictive uncertainty is relatively high at the single-pixel level. As such, although our map is projected at 30 arcsec (~1 km^2 at the Equator) resolution, high-precision predictions require distances larger than 10 km (Extended Data Fig. 2d–f). Yet, it is worth noting that, even at these distances, systemic bias could potentially cause these predictions to deviate from the true mean. We attempted to minimize this possibility by collecting data from a wide number of sources that would be unlikely to represent the same systematic bias, but the risk remains, particularly in areas with low sampling density. For example, due to the low sampling density (Fig. 1), our models had high prediction uncertainty in African grasslands (Fig. 3d).

The environmental controls of RMFs

To generate a mechanistic understanding of the environmental factors driving the global variation in RMFs, we explored the directional effects of eight environmental features that have previously been shown to influence RMFs, including annual mean temperature and water availability (annual mean soil moisture and aridity index)[21–23] (Fig. 4 and Extended Data Figs. 3–6). Our analyses show that, across all vegetation types, temperature and water availability are the dominant drivers of RMFs at the global scale (Fig. 4a–c), supporting the idea that belowground biomass increases when belowground resources are scarce[18,19,29,30]. However, we also identify important interactions among these main drivers that can obscure the simple directional effects at a global scale (Figs. 4d–l and 5).

In forests, previous meta-analyses[21] and intraspecific comparisons[13,31] showed an effect of soil water availability on RMFs. However, a recent global study on forest biomass distribution did not detect an effect of water availability and instead indicated mean annual temperature as the main driver of RMF variation[23]. Our global analyses can now reconcile this discrepancy, revealing strong interactions between temperature and soil moisture (Figs. 4d,g,j and 5 and Extended Data Fig. 4a,b) that probably result from geographic variation in the interplay between water, temperature and nutrient limitation[19,29,32]. In dry areas, temperature exhibited a positive effect on RMFs (Figs. 4d,j and 5a and Extended Data Fig. 4a), probably because the extent of water limitation there is exacerbated by hot temperatures, leading to an increased investment into building roots. In contrast, in wet areas, nutrient limitation should be the driving force of above- versus belowground biomass investment, explaining why we see increasing RMFs toward cold areas where nutrients are at low supply and are less mobile. The normalized difference vegetation index (NDVI), an indicator of photosynthetic activity and canopy cover[33], was negatively correlated with RMFs (Extended Data Fig. 3), suggesting reduced allocation to roots in dense forests where aboveground competition for light is high.

Shrublands follow similar trends to forests. The geographic gradients of RMFs are systematically affected by the interplay between temperature and water availability (Fig. 4e,h and Extended Data Fig. 5). As in forests, we find that the negative effect of temperature on RMFs strongly increases toward moist regions (Fig. 4e and Extended Data Fig. 5a). In grasslands, soil moisture and temperature also represent dominant controls of RMFs. In contrast to forests and shrublands, the negative effect of soil moisture on RMFs is most pronounced at intermediate temperature levels and the negative effect of temperature is most pronounced at dry locations (Fig. 4f,i and Extended Data Fig. 6), probably reflecting functional differences between woody and herbaceous species (grasses), as grasses might be more tolerant to low temperature and drought.

Despite the general impacts of water availability and temperature on RMFs, soil characteristics, such as sand content, depth to bedrock and soil organic carbon-to-nitrogen (C:N) ratio, also play an important role in the geographic RMF gradient across all vegetation types (Extended Data Figs. 3–6). For instance, RMFs in forests are higher in regions with sandy and deep soil (Extended Data Figs. 3 and 4d). High RMFs in sandy soils can probably be explained by reduced soil moisture due to the low water-holding capacity and an increased need for root stability, while lack of space belowground probably explains why shallow soils correlate with low RMFs. The C:N ratio of soil organic matter, a proxy for plant N availability[34,35], was positively correlated with RMFs in shrublands, but showed no correlation in forests and a negative correlation with RMFs in grasslands (Extended Data Fig. 3). Despite usually being nutrient poor, arid regions typically have very low soil C:N ratios as a result of a small organic C pool[36], explaining the negative correlation between soil C:N ratio and RMFs in grasslands and emphasizing the need for a direct characterization of nitrogen availability to represent global variations in plant-available nutrients.

Comparison with existing biomass models

To examine the possible utility of these spatial models, we explored the extent to which the observed variation in forest root-mass fractions is reflected in existing biomass models. We assessed the ability of four dynamic global vegetation models that contribute to the dynamic global vegetation model project TRENDY (version 8)[37] as well as the latest static global biomass prediction[6] to replicate the observed RMFs in forest systems. The selected dynamic vegetation models do not currently simulate the spatial variation in RMFs that is reflected in our data (Extended Data Fig. 7; all R^2 values < 0). The most recent static model of global above- and belowground plant carbon[6] also failed to explain the observed spatial variation in RMFs

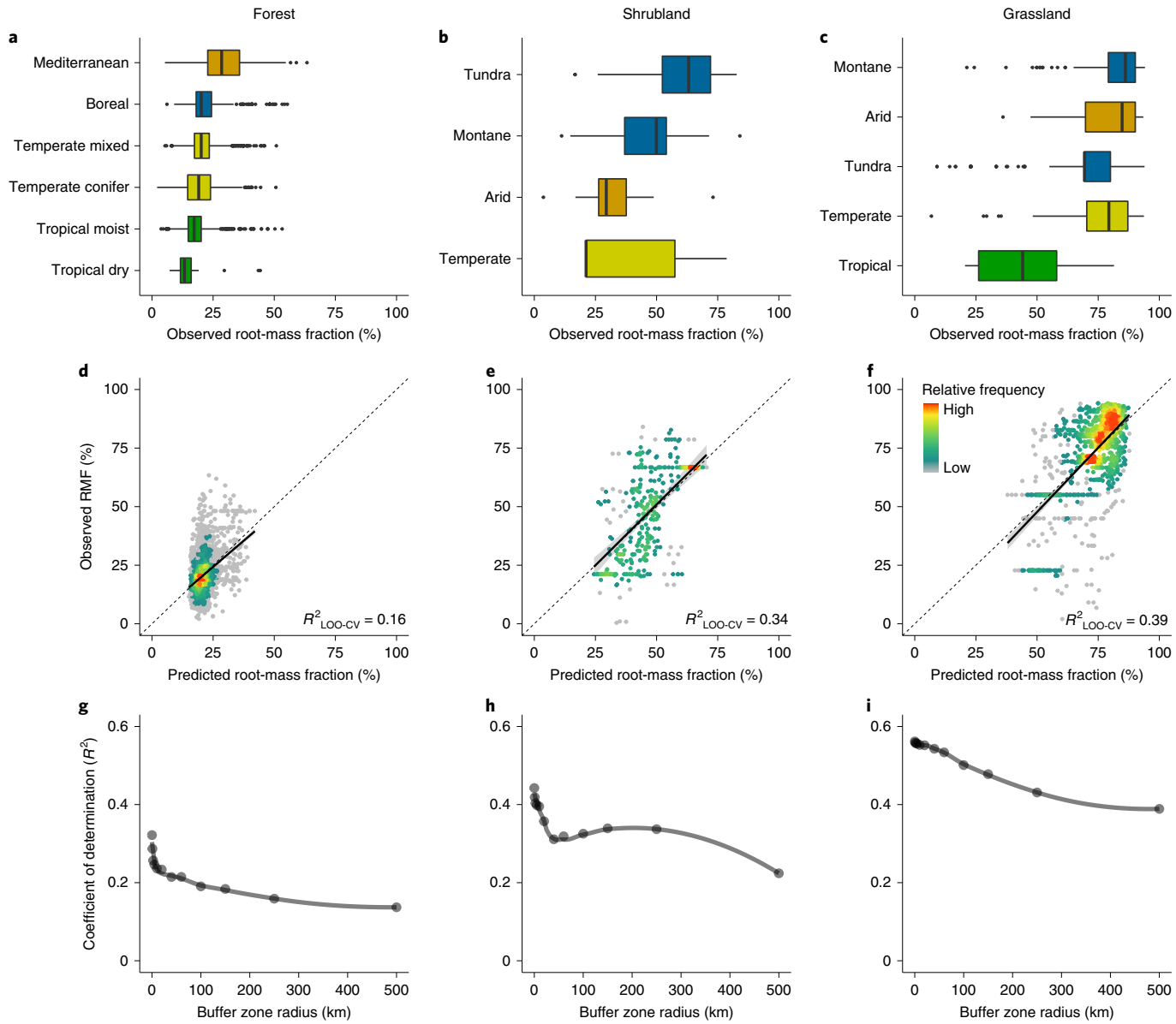


Fig. 2 | RMF variation and model validation. a–c, Boxplots of observed RMFs per vegetation type and biome ($n = 5{,}170$ forest (**a**), 340 shrubland (**b**) and 1,293 grassland (**c**) observations). **d–f**, Heat plots showing the relationships between predicted and observed RMFs in forests (**d**), shrublands (**e**), and grasslands (**f**) based on spatially buffered LOO-CV. Solid lines indicate fitted relationships based on ordinary least squares regression (coefficient-of-determination values relative to the 1:1 line [equation (2)] shown in the bottom right corner); dashed diagonal lines indicate a 1:1 relationship between observed and predicted points. **g–i**, LOO-CV coefficient-of-determination values (equation (2)) for buffer radii for data exclusion from 1 to 500 km in forests (**g**), shrublands (**h**) and grasslands (**i**). For each data point, all data within the respective buffer radius were excluded and all sample points out of the buffer zone were used to train the model.

(Extended Data Fig. 7b, $R^2 < 0$) and overpredicted RMFs by, on average, ~9.8%. As such, consideration of these new, emerging patterns in belowground biomass may be useful for enhancing the performance of global dynamic vegetation models. By capturing some of this unexplained variation in RMF across the globe, we hope that our predictive models can be used to benchmark and parameterize process-based models to represent the meaningful variation in plant belowground biomass investment across time and space.

Discussion

By reflecting the morphological adaptations of plants in different ecosystem types to varying resource availability, our global models of RMF are directly relevant for building a thorough understanding of the plant carbon sink, and associated plant trait variation over time and space. For instance, roots often have a lower litter quality (high C:N ratio) relative to leaves and aboveground wood[38,39]. Various studies additionally point towards a trade-off between plant growth rate and RMF, with high RMFs being associated with lower growth rates and higher stress tolerance[38,40]. Given the significance of RMFs for ecosystem functioning on multiple levels[7,27], our study provides a baseline for future research allowing for upscaling these feedbacks to the global scale. Moreover, this study represents a critical step toward assessing the impacts of climate change on terrestrial carbon distributions. The strong interactive effects of temperature and water availability on RMFs observed here indicate that warmer and drier climates in the future might

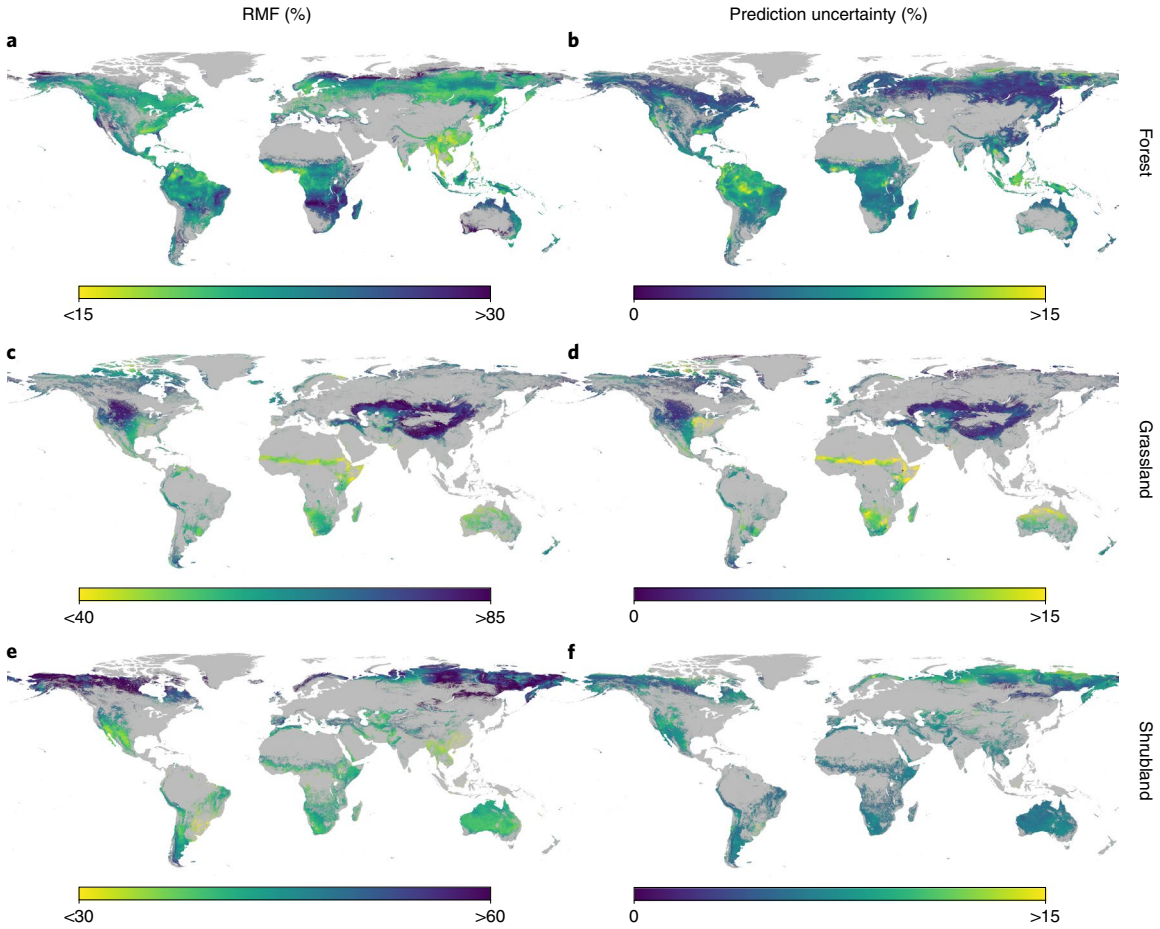


Fig. 3 | The global distribution of RMFs in forests, grasslands and shrublands. a,**c**,**e**, Predicted RMFs from random forest models based on empirical data from 5,170 forest (**a**), 1,293 grassland (**c**) and 340 shrubland (**e**) records. **b**,**d**,**f**, Bootstrapped variation coefficients (standard deviation divided by mean) in forests (**b**), grasslands (**d**) and shrublands (**f**). In forests, high uncertainty occurs in Southeastern Asia and central Amazonia. In grasslands, high uncertainty is found in African savannah and steppes, where we lack sampling data. All maps are projected at 30 arcsec (~1 km^2) resolution.

reshape the relative biomass allocation to root systems to maintain the functional balance of aboveground versus belowground resources and sinks.

The systematic variation in global plant RMFs that we detect now offers the opportunity to derive quantitative, spatially explicit estimates of total belowground biomass from existing aboveground biomass products. To exemplify this, we combined existing maps of aboveground plant biomass with land-cover maps[41] and our vegetation-type-specific RMF estimates to approximate the variation in root biomass across the globe[15] (Fig. 6 and Extended Data Fig. 8). We estimate a total belowground biomass of 93.9 (minimum to maximum = 77.5–113.8), 11.5 (9.1–14.4) and 5.2 (3.7–7.0) GtC for forests, shrublands and grasslands, respectively. As such, the contribution of belowground biomass to the total terrestrial plant biomass ranges around 24% (20–28%), highlighting the importance of belowground plant carbon pools for global biogeochemical cycling. This estimation of the total proportion of belowground biomass largely agrees with previous estimates that ranged from 20% to 30% at the global scale[6,15,26,29,42,43].

While our model predictions well reflect the observed variation in RMFs across large spatial gradients, there are still limits regarding our ability to accurately predict at finer spatial scales. First, although our observations support much of the mechanistic and experimental research in plant biomass investment, we must highlight that the correlations observed here are not necessarily indicative of causal relationships. In addition, a considerable proportion of the variation in root allocation remains unexplained, underscoring the need for further research and data collection to improve predictions of plant biomass partitioning. All RMF uncertainty estimates in this study reflect the bootstrap prediction uncertainty (Fig. 3b,d,f) or the intermodel consistency (Extended Data Fig. 9) and thus cannot represent other potentially important sources of uncertainty, such as measurement error and sampling bias. Identifying ways to incorporate these other sources of uncertainty is a key outstanding challenge and will be critical to improving future modelling efforts. Some of the unexplained variation in our models is probably due to missing meta-information about the sites where the RMFs were observed. Because plot-level metadata is neither available nor could it be used for global extrapolation, our models had to rely on interpolated global layers of environmental conditions. These layers might accurately describe plot-level variations in aboveground parameters, such as mean annual temperature, but are probably introducing considerable uncertainties with respect to soil information, in particular nutrient content. Yet, despite these covariate uncertainties, the out-of-sample model fit of 16–39% (obtained from spatial LOO-CV; Fig. 2d–i) demonstrates the important role that climate and soil characteristics play in shaping plant biomass distributions above- and belowground.

In conclusion, our analysis reveals the spatial distribution of plant root allocation at a global scale that can inform our understanding of vegetation carbon storage and dynamics.

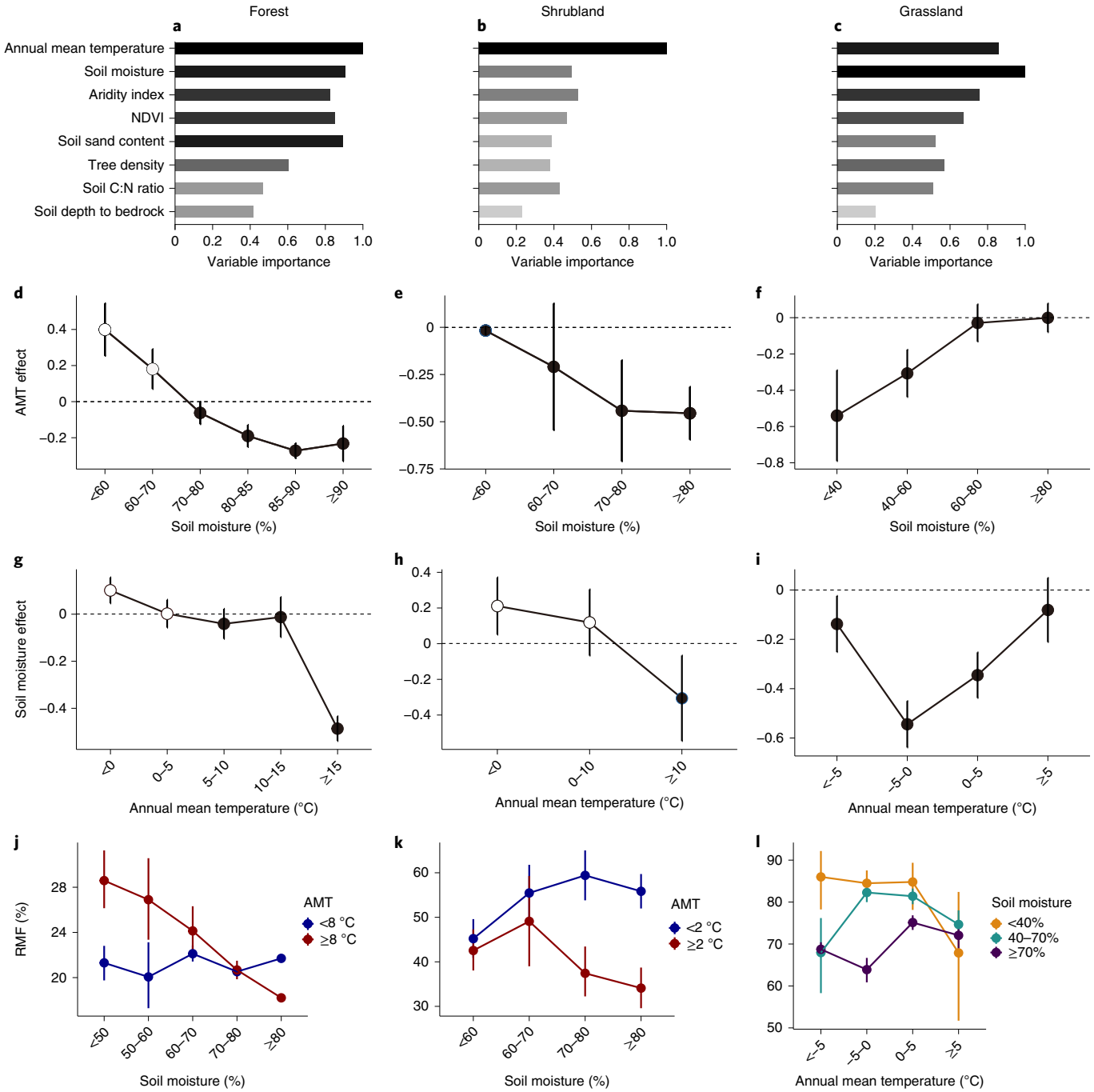


Fig. 4 | Relationships between environmental variables and RMFs. a–c, Variable importance metrics from random forest models for the effects of eight environmental variables on RMFs in forests (**a**), shrublands (**b**) and grasslands (**c**). **d–i**, The effects of annual mean temperature (AMT) (**d–f**) and soil moisture (**g–i**) on RMFs for different soil-moisture or temperature classes, respectively. Values represent mean standardized coefficients (±2 standard errors) bounded between −1 and 1, where negative values indicate a negative effect (closed circles) and positive values a positive effect on RMFs (open circles). **j–l**, Illustration of changes in mean RMFs (±95% confidence intervals) along temperature and soil-moisture gradients. For forests (**j**) and shrublands (**k**), RMFs are shown in response to soil moisture for two temperature classes (red = high temperature, blue = low temperature); for grasslands (**l**), RMFs are shown in response to annual mean temperature for three soil-moisture classes (purple = high soil moisture, teal = intermediate and orange = low).

The geographic gradients in above- versus belowground biomass distribution contribute to our understanding of limiting factors to plant growth and offer a new fine-scale framework to estimate the total belowground carbon stocks of plants. We estimate that approximately one quarter of the total plant carbon is stored belowground, stressing the importance of root systems for the terrestrial carbon sink. These estimates of root biomass are directly determined by existing estimates of aboveground biomass, which are characterized by considerable uncertainty. As we continue to expand our understanding of global aboveground biomass, these predictive models of proportional belowground biomass contributions will be valuable for improving our predictions of total plant biomass and ecosystem functioning under current and future climate scenarios.

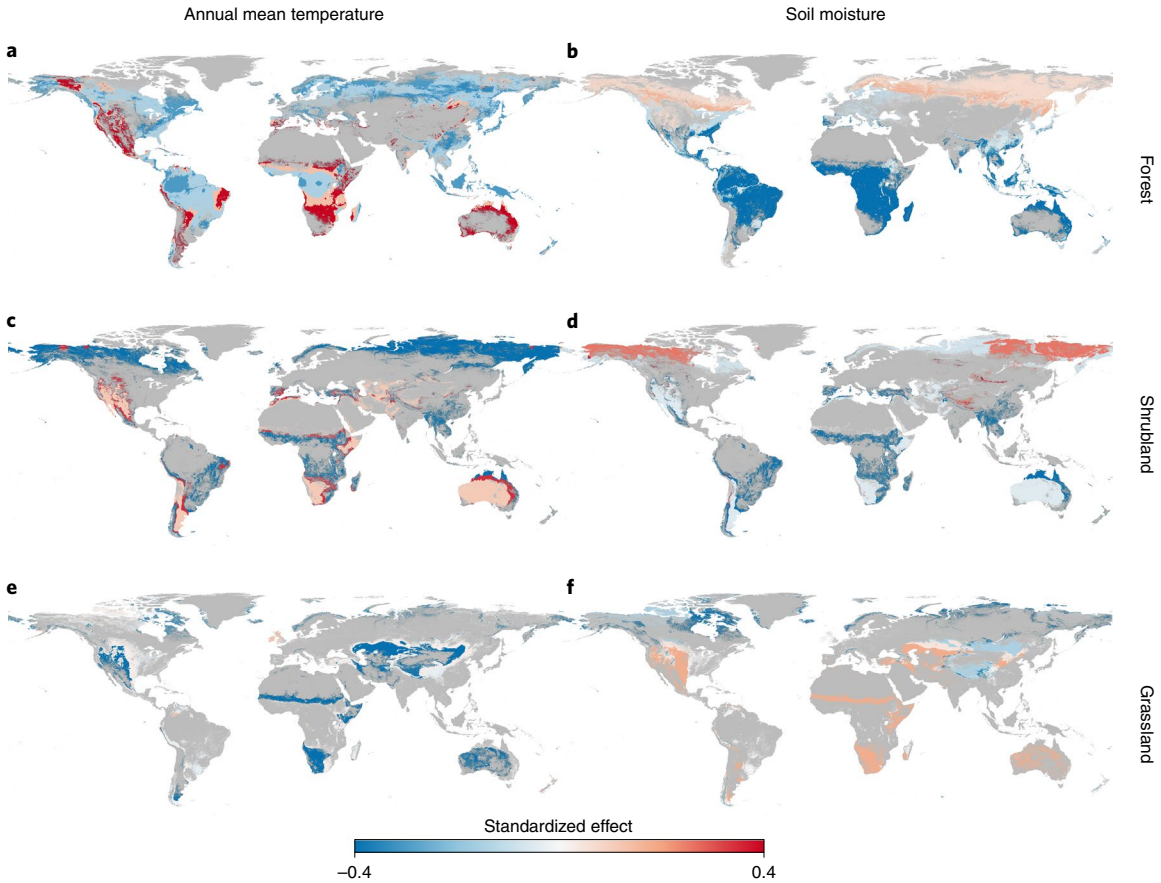


Fig. 5 | Spatial variations in the environmental correlates of RMFs. a–f, Maps show the relationships between RMFs and annual mean temperature (left panels) or soil moisture (right panels) in forests (**a**,**b**), shrublands (**c**,**d**) and grasslands (**e**,**f**). Effect sizes (standardized coefficients) for each region are based on recursive partitioning trees (Extended Data Figs. 4–6). Red colours represent a positive effect of annual mean temperature or soil moisture on RMFs and blue colours represent a negative effect.

Methods

Data collection. *Collection of plot-level RMFs.* To collect data on the biomass ratio between roots and shoots (root–shoot ratio, RSR) from the literature, we set ubiquitous criteria for all three vegetation types (forests, shrublands and grasslands) and specific criteria for some of the vegetation types. Data was only used if all of the following four criteria were met: (1) RSR values were available at the plot level (and not only for selected individuals); (2) data came from field measurements (allometry-derived RSRs were excluded). If the study consisted of multiple treatments, only data from the control treatment was selected; (3) coordinate information was given; and (4) measurements were taken after 1960. For datasets that included information on more than one vegetation type, the abovementioned criteria were evaluated separately for the subset of trees, grasses and shrubs within the dataset. Furthermore, to minimize the influence of ontogenetic drift on RSRs in forests (higher root-to-shoot biomass ratios in seedlings than in adults[44]), we (1) excluded studies that focused on tree seedlings and (2) excluded seedlings or tree individuals within studies with a diameter at breast height (DBH) smaller than 10 cm. This selection process resulted in a total of 17,814 plot-level RSR records, with 13,620 forest records from refs. [12,21,45–95], 2,468 grassland records from refs. [26,63,81,82,96–113] and 1,726 shrubland records from refs. [58,63,81,82,91,96,99,105,114–184]. For each vegetation type, we aggregated records from the same location by calculating the mean and removed obvious outliers by visual inspection (0.7%, 2.0% and 4.9% of the forest, grassland and shrubland data, respectively), resulting in spatially distinct data for a total of 5,170 forest, 1,293 grassland and 340 shrubland records (Fig. 1). These aggregated measurements thus reflect the ratio of living plant biomass existing below- versus aboveground in a given location. We then converted all root–shoot ratio records to root-mass fractions (RMFs), restricting the data to values from 0 to 100%:

$$\mathrm{RMF} = \frac{\mathrm{RSR}}{\mathrm{RSR}+1} \times 100\%. \quad (1)$$

Environmental covariates. In total, 63 environmental layers, including climate, soil, anthropogenic, vegetation and topographic variables, were used as covariates in our analyses (Extended Data Fig. 1 and ref. [185] provide covariate layer details). All layers were standardized to 30 arcsec resolution (~1 km² at the Equator).

Geospatial modelling of root-mass fractions. *Random forest models.* To train global models of RMFs for each vegetation type, we ran a series of random forest machine learning models using the R package h2o[186]. We used a grid-search procedure to iteratively explore the results of a suite of random forest models trained on the 63 covariates. A list of hyperparameters was selected, varying in (1) the number of trees (50, 100 or 150), (2) the number of variables sampled at each split (2 to 10) and (3) the minimum observations per leaf (2 to 5), resulting in a total of 108 models for each vegetation type. During the grid-search procedure, we assessed the performance of each model using random tenfold cross validation. Coefficient-of-determination values (R^2) for each fold were used to compute mean and standard deviation values for the cross-validated model. We selected the best ten models for each vegetation type based on the highest mean R^2 values.

All R^2 values reported throughout the manuscript represent the coefficient of determination relative to the 1:1 line of observed versus predicted values. This is equivalent to a standardized mean squared error:

$$R^2 = 1 - \frac{\mathrm{RSS}}{\mathrm{TSS}} = 1 - \frac{n \times \mathrm{MSE}}{\mathrm{TSS}} = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(O_i - \bar{O})^2} \quad (2)$$

where RSS refers to the residual sum of squares, TSS the total sum of squares, MSE the mean squared error, O_i the observations, P_i the predictions and n the number of observations. This metric reflects systematic deviation between model predictions and observed values. Negative values indicate that the model performs worse than using the mean across observations.

To create the final maps, we used an ensemble approach, whereby, for each vegetation type, we averaged the global predictions from the best ten random forest models. The final R^2 values (based on the ensemble of the cross-validated predictions) for RMFs in forests, shrublands and grasslands were 0.33, 0.47 and 0.58, respectively (Extended Data Fig. 2a–c). By taking the average prediction across multiple models, ensemble methods minimize the influence of any single

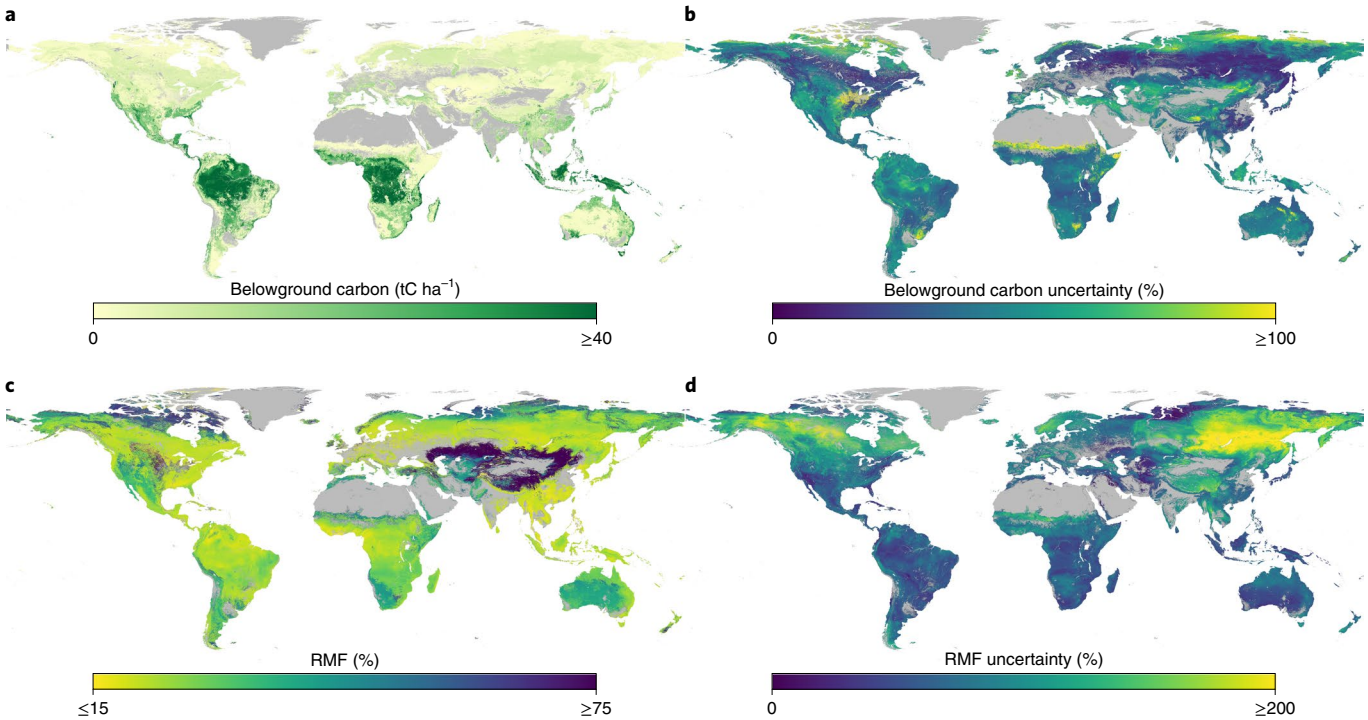


Fig. 6 | The global distribution of belowground plant biomass. a, Average belowground plant biomass per area (tonnes of carbon per hectare) scaled by the relative biomass contribution of each vegetation type (Methods). **b**, Relative prediction uncertainty calculated as the difference between the highest and lowest belowground carbon estimates per pixel (based on the minimum and maximum estimates for each vegetation type) divided by the mean carbon estimate in the respective pixel. **c**,**d**, Average pixel-level RMF distribution (**c**) and associated uncertainty (**d**) integrating forest, grassland and shrubland RMFs. To obtain pixel-level averages, RMFs of the three functional types were weighted by the relative contribution of each type to the aboveground biomass in the respective pixel (Methods). With this map, pixel-level estimates of aboveground biomass can be directly scaled to belowground biomass. The uncertainty map shows the difference between maximum and minimum RMFs divided by averaged RMFs. All maps are projected at 30 arcsec (~1 km^2) resolution.

prediction, thereby stabilizing variation and minimizing bias that can otherwise arise from extrapolation or in-fit overfitting when using a single machine learning model[187]. Moreover, by quantifying the variation across these ensemble predictions (intermodel consistency), we can identify areas that have low consensus across multiple models (Extended Data Fig. 9).

To implement this ensemble approach, the mean predicted value across the ten best-fitting models was used as the final model prediction for each vegetation type for each pixel, and the coefficient of variation across these ten models was used to characterize intermodel consistency (see "Evaluating model uncertainties"). All predictions were masked by the global distribution of forests, shrublands and grasslands (≥10% coverage), based on a 30 arcsec (1 km^2) consensus land-cover map[41], representing the percentage of area covered by each vegetation type per pixel.

We also generated a unified RMF map, reflecting a weighted RMF average across vegetation types (Fig. 6c). To do so, we accounted for the relative aboveground biomass contribution of each vegetation type to the pixel-level total biomass (see "Belowground biomass estimation").

To test for the potential effect of spatial autocorrelation in the model residuals, which can cause overestimation of model predictive power[28], we ran spatially buffered LOO-CV for a series of buffer radii from 1 to 500 km. In this procedure, each data point is predicted by a model that uses all data outside the buffer radius of the respective data point as training data, resulting in 5,170, 1,293 and 340 independent model runs for forests, grasslands and shrublands, respectively. For each vegetation type, we used the hyperparameter settings of the best-performing random forest model (based on random tenfold cross validation, see above). As in the other models, model performance was evaluated by the coefficient of determination relative to the 1:1 line (equation (2)) of observed versus predicted values (Fig. 2d–i).

To assess the range of autocorrelation, we created semivariograms of model residuals (Extended Data Fig. 2g–i). The residual semivariograms of our random cross-validation and leave-one-out cross-validation models show autocorrelation ranges on the order of ~150 to 500 km. Based on these results, we chose buffer radii of 250 km for forests and shrublands and 500 km for grasslands, which represent the range of autocorrelation of RMF residuals.

All models were run in R version 3.5.1 (ref. [188]). Geospatial mapping was performed in Google Earth Engine[189].

Evaluating model uncertainties. To represent the confidence of our estimates for each pixel based on sampling uncertainty, we performed a stratified bootstrapping procedure using all data within each vegetation type. We set the number of bootstrap iterations to 100, that is, all pixels of the raster layer data (93,693,855 in forests, 60,977,011 in shrublands and 49,464,610 in grasslands) were classified 100 times. Each of the three datasets (forest, shrubland and grassland) were sampled with replacement, using biome as stratification category to proportionally represent the major bioclimatic zones in each of the 100 bootstrap samples. The size of each bootstrap sample was equal to the size of the original datasets. Each of the 100 bootstrap samples were then used to classify the composite image. Finally, these 100 images were used to calculate per-pixel coefficient-of-variation values (standard deviation divided by the mean predicted value) as a measure of sampling uncertainty (hereafter referred to as bootstrap prediction uncertainty; Fig. 3b,d,f).

To identify pixel-level uncertainties in the unified RMF map, combining all three vegetation types, we first calculated the difference between the upper (97.5% quantile) and lower (2.5% quantile) RMF prediction for each pixel and vegetation type, based on the bootstrapping procedure. We then obtained the lower and upper RMFs within each pixel, based on the lower and upper RMFs in each vegetation type, accounting for the relative aboveground biomass contribution of each vegetation type to the pixel-level total biomass (Fig. 6d and see "Belowground biomass estimation").

We additionally created a spatially explicit visualization of the variation in predictions across the ten best random forest models, hereafter referred to as intermodel consistency, by calculating the per-pixel coefficient of variation (standard deviation divided by the average) across the ten best random forest models (Extended Data Fig. 9).

To assess the extent of extrapolation in our models, that is, how well our training data represents the full multivariate environmental covariate space, we performed a principal-component-analysis–based approach following van den Hoogen et al.[185]. We projected the covariates composite into the same space using the centring values, scaling values and eigenvectors from the principal component analysis of the training data. Then, we created convex hulls for each of the bivariate combinations from the top principal components (which collectively covered more than 90% of the sample space variation), and classified whether each pixel falls within or outside each of these convex hulls. We used 17 principal components

(PCs) with 136 combinations for all covariates for the forest dataset, 15 PCs with 105 combinations for all covariates for the shrubland dataset, and 16 PCs with 120 combination for all covariates for the grassland dataset. This analysis revealed that 94%, 70% and 78% of land pixels falling in the extent of the respective vegetation type are covering at least 90% of the environmental conditions present in our training data locations in forest, shrubland and grassland, respectively (Extended Data Fig. 10).

The maps are available online at: https://code.earthengine.google.com/2b17d6844f6077ea25021988cbc9b4c7.

Environmental drivers of geographic variation in root-mass fractions. *Variable selection.* To explore the effect of magnitude and direction of the main environmental drivers of RMFs within forests, shrublands and grasslands, we selected 8 uncorrelated environmental factors among our 63 covariates, representing climate, vegetation and soil features that are of high ecological importance and that have shown significant relationships with RMF in previous studies[21–23,82]: annual mean temperature, aridity index, NDVI, tree density, soil organic C:N ratio, soil depth to bedrock, soil-moisture profile (the available water for the plant divided by the total water-holding capacity of the soil) and soil sand content. To test for multicollinearity among covariates, we calculated variance inflation factors using the R package HH[190]. For all vegetation types, variance inflation factors were less than three, indicating sufficient independence among predictor variables.

Variable importance testing. We then ran a random forest model, implemented in the h2o R package[186], and quantified the relative importance of each of the eight variables, as determined by calculating whether that variable was selected to split on during the tree-building process, and how much the squared error (over all trees) improved as a result. To test whether our subset of 8 variables is representative of the main drivers of RMFs across our full random forest model including all 63 covariates, we compared the R^2 values (from random tenfold cross validation) of the reduced 8-covariate model with the full 63-covariate model. The results show that the reduced model captures 78%, 79% and 83% of the variation explained by our full model for forests, grasslands and shrublands, respectively, indicating that our subset of 8 independent variables represents the main drivers of our full model.

To further explore how the effects of environmental variables on RMFs vary depending on the magnitude of other environmental drivers, we ran recursive partitioning analyses using the packages partykit[191] and ggparty[192]. Using a machine learning (that is, decision tree) algorithm, we explored the context dependency of the slope and intercept of a univariate linear model for the effect of any of the top four predictor variables (Fig. 4a–c) on RMFs, while the three remaining variables were evaluated as potential split points (Extended Data Figs. 4–6 and Fig. 5). The minimum node size (minimum number of observations contained in each terminal node) was set to ~10% of the total data, that is, 500 for forests, 120 for grasslands, and 30 for shrublands. These analyses demonstrate that the effects of one variable are highly dependent on other variables; the slope of the regression even changed from negative to positive depending on other explanatory variables (for example, Extended Data Fig. 4a). To exemplify the strong interactive effects of temperature and water availability, we additionally characterized the effects of mean annual temperature on RMFs within different soil-moisture classes and vice versa (Fig. 4d–l).

Partial-regression analysis. We also tested for consistent global trends of the eight variables on RMFs by univariate partial-linear-regression models. For forests, we ran bootstrapped partial-linear-regression models to remove the spatial autocorrelation present in our full dataset (see below) by randomly subsampling 1,700 points (33%) with replacement out of a total of 5,170 points 1,000 times. For each subsample, we fit a partial-regression model for each covariate of interest and calculated the corresponding slope. The results were then aggregated by calculating the mean coefficient (slope) across all model runs. For grasslands and shrublands, we used the whole dataset for our univariate partial-linear models. These partial-regression models show the overall effect of one variable when controlling for the global variation in all other variables. We tested for spatial autocorrelation in the residuals that could potentially bias significance tests and parameter estimates[193] by performing Moran's I tests. We detected significant spatial autocorrelation in the full forest data ($I=0.27$, $P<0.001$), and our bootstrapping procedure successfully removed this spatial autocorrelation (Moran's I P value >0.05) for 98% of the subsampled datasets. There was no significant spatial autocorrelation in grasslands ($I=0.04$, $P=0.36$) and shrublands ($I=0.15$, $P=0.13$).

As illustrated by the strong interaction between soil moisture and annual mean temperature (Fig. 4), a non-significant effect in these global linear models (Extended Data Fig. 3) does not mean that the respective variable does not affect RMFs; it rather indicates that there is no globally coherent trend detectable for this variable. For example, in shrublands, soil moisture had a positive effect on RMFs in cold regions and a negative effect on RMFs in warm regions (Fig. 4h). This dependency on other variables therefore explains why our global partial-regression models show no consistent general trend for this variable (Extended Data Fig. 3).

Belowground biomass estimation. Belowground biomass was estimated by integrating global estimates of aboveground biomass (Intergovernmental Panel on Climate Change (IPCC) tier 1)[15] with our high-resolution maps of RMFs. To account for the occurrence of different vegetation types within pixels, we divided the aboveground biomass of each pixel into the relative proportion comprising different vegetation forms, derived from a high-resolution land-cover map[41]. Generally, total aboveground biomass of each pixel (AGB_i) can be partitioned by the following formula:

$$AGB_i = AGB_f + AGB_g + AGB_s \quad (3)$$

where AGB_i, AGB_g and AGB_s refer to the aboveground biomass of forests, grasses and shrubs for each pixel. For each vegetation type, the aboveground biomass can be rewritten as the product of aboveground biomass density (ABD; biomass per unit area), focal vegetation coverage (C), and pixel area (A_i). As such, aboveground biomass for each vegetation type can be expressed as:

$$AGB_f = ABD_f \times C_f \times A_i \quad (4)$$

$$AGB_g = ABD_g \times C_g \times A_i \quad (5)$$

$$AGB_s = ABD_s \times C_s \times A_i \quad (6)$$

and we can therefore rewrite equation (3) to:

$$AGB_i = (ABD_f \times C_f + ABD_g \times C_g + ABD_s \times C_s) \times A_i. \quad (7)$$

Because the IPCC tier 1 biomass layer does not account for different vegetation types within pixels[15], we estimated the biomass that each vegetation type contributes within each pixel by using fixed biomass ratios per unit area for each ecofloristic zone[15]:

$$\alpha = \frac{ABD_f}{ABD_g}, \quad (8)$$

$$\beta = \frac{ABD_s}{ABD_g} \quad (9)$$

where α is the biomass ratio between forests and grasslands and β is the biomass ratio between shrublands and grasslands. To obtain the biomass ratios α and β, we used aboveground biomass values for each vegetation type and ecofloristic zone published in the IPCC report[15]. By combining these ratios with the spatial coverage of each vegetation type within pixels[41], we were able to allocate the total aboveground biomass of each pixel known from equation (7) to the three vegetation forms as:

$$AGB_i = (C_f \times \alpha \times ABD_g + ABD_g \times C_g + C_s \times \beta \times ABD_g) \times A_i \quad (10)$$

where A_i is the area of the pixel and C_g, C_f and C_s refer to the relative grassland, forest and shrubland coverage.

Because all parameters in equation (10) are known besides ABD_g, we can solve this equation for each pixel to estimate ABD_g. We then used this value to obtain the corresponding aboveground densities for forest and shrubland using equations (8) and (9). Finally, we used these density values to obtain the total aboveground biomass for each vegetation type for each pixel, using equations (3–6)[13,15].

Using these pixel-level estimates of aboveground biomass for each vegetation type, we then calculated belowground biomass (BGB) per pixel by scaling these values by our RMF estimates using equation (1) (Extended Data Fig. 8a,c,e).

$$BGB_f = AGB_f \times RSR_f \quad (11)$$

$$BGB_g = AGB_g \times RSR_g \quad (12)$$

$$BGB_s = AGB_s \times RSR_s \quad (13)$$

BGB uncertainty within vegetation types was based on the upper and lower biomass estimates in each pixel and vegetation type, which were calculated from the upper and lower RMFs based on the bootstrapping procedure (Extended Data Fig. 8b,d,f).

To compute the total BGB in each pixel (Fig. 6a), BGB values of each vegetation type were aggregated via:

$$BGB_i = BGB_f + BGB_g + BGB_s. \quad (14)$$

Total BGB uncertainty in each pixel was then derived from the sum of the upper or lower biomass estimates for each vegetation type (Fig. 6b).

To compute total belowground biomass at the global scale, we summed pixel-level BGBs:

$$\text{Total BGB} = \sum \text{BGB}_i. \quad (15)$$

Similarly, belowground biomass of each vegetation type was summed by:

$$\text{Total BGB}_\text{f} = \sum \text{BGB}_\text{f}, \quad (16)$$

$$\text{Total BGB}_\text{g} = \sum \text{BGB}_\text{g}, \quad (17)$$

$$\text{Total BGB}_\text{s} = \sum \text{BGB}_\text{s}. \quad (18)$$

To obtain a pixel-level RMF map that accounts for all three vegetation types (Fig. 6c), the belowground biomass of each pixel was divided by the total (below- and aboveground) biomass of each pixel:

$$\text{RMF}_i = \frac{\text{BGB}_i}{\text{AGB}_i + \text{BGB}_i}. \quad (19)$$

Comparison of observed RMFs with dynamic global vegetation model outputs. We derived root-mass fractions as root biomass divided by total vegetation biomass and averaged (mean) over the years 1986–2015 from outputs of TRENDY (version 8) global vegetation model simulations following the S2 setup (that is, forced with observed climate and CO_2, while land use and land-use change were held constant at year 1700 state). The resulting maps represent 'ecosystem-level' root-mass fractions, aggregated across all plant functional types (PFT) by grid cell. To represent forest RMFs, we used only simulated RMFs of grid cells with a forest-cover fraction (determined by model outputs of total fractional plant coverage of all tree PFTs) of >50% (Extended Data Fig. 7). We tested the robustness of model–data comparison using an alternative threshold of >90% forest cover. Both thresholds yielded the same results, with R^2s of observed versus predicted RMFs consistently <0. Models were selected for our analysis if outputs were provided for all required variables (root biomass, total vegetation biomass and PFT cover fraction).

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The root–shoot ratio data underlying this study are available at https://github.com/haozhima95/Global_mapping_root_shoot_ratio/tree/master/RSR_data. Citations for the root–shoot ratio data are provided in the methods.

Code availability

The code used for this study is available at https://github.com/haozhima95/Global_mapping_root_shoot_ratio.git.

Received: 30 October 2020; Accepted: 6 May 2021;
Published online: 24 June 2021

References

1. Erb, K. H. et al. Unexpectedly large impact of forest management and grazing on global vegetation biomass. *Nature* **553**, 73–76 (2018).
2. Luyssaert, S. et al. Old-growth forests as global carbon sinks. *Nature* **455**, 213–215 (2008).
3. Drake, J. B. et al. Above-ground biomass estimation in closed canopy Neotropical forests using lidar remote sensing: factors affecting the generality of relationships. *Glob. Ecol. Biogeogr.* **12**, 147–159 (2003).
4. Lefsky, M. A. et al. Lidar remote sensing of above-ground biomass in three biomes. *Glob. Ecol. Biogeogr.* **11**, 393–399 (2002).
5. Duncanson, L. et al. The importance of consistent global forest aboveground biomass product validation. *Surv. Geophys.* **40**, 979–999 (2019).
6. Spawn, S. A., Sullivan, C. C., Lark, T. J. & Gibbs, H. K. Harmonized global maps of above and belowground biomass carbon density in the year 2010. *Sci. Data* **7**, 112 (2020).
7. Ottaviani, G. et al. The neglected belowground dimension of plant dominance. *Trends Ecol. Evol.* **35**, 763–766 (2020).
8. Jackson, L. E., Burger, M. & Cavagnaro, T. R. Roots, nitrogen transformations, and ecosystem services. *Annu. Rev. Plant Biol.* **59**, 341–363 (2008).
9. Gill, R. A. & Jackson, R. B. Global patterns of root turnover for terrestrial ecosystems. *New Phytol.* **147**, 13–31 (2000).
10. Robinson, D. Implications of a large global root biomass for carbon sink estimates and for soil carbon dynamics. *Proc. R. Soc. Lond. B* **274**, 2753–2759 (2007).
11. Bardgett, R. D., Mommer, L. & De Vries, F. T. Going underground: root traits as drivers of ecosystem processes. *Trends Ecol. Evol.* **29**, 692–699 (2014).
12. Ribeiro, S. C. et al. Above- and belowground biomass in a Brazilian Cerrado. *For. Ecol. Manage.* **262**, 491–499 (2011).
13. Mokany, K., Raison, R. J. & Prokushkin, A. S. Critical analysis of root:shoot ratios in terrestrial biomes. *Glob. Chang. Biol.* **12**, 84–96 (2006).
14. Saatchi, S. S. et al. Benchmark map of forest carbon stocks in tropical regions across three continents. *Proc. Natl Acad. Sci. USA* **108**, 9899–9904 (2011).
15. Ruesch, A. S. & Gibbs, H. H. K. *New IPCC Tier-1 Global Biomass Carbon Map for the Year 2000* (Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, 2008).
16. Chen, J. L. & Reynolds, J. F. A coordination model of whole-plant carbon allocation in relation to water stress. *Ann. Bot.* **80**, 45–55 (1997).
17. Franklin, O. et al. Modeling carbon allocation in trees: a search for principles. *Tree Physiol.* **32**, 648–666 (2012).
18. Bloom, A. J., Chapin, F. S. & Mooney, H. A. Resource limitation in plants—an economic analogy. *Annu. Rev. Ecol. Syst.* **16**, 363–392 (1985).
19. Poorter, H. et al. Biomass allocation to leaves, stems and roots: meta-analyses of interspecific variation and environmental control. *New Phytol.* **193**, 30–50 (2012).
20. Reich, P. in *Plant Roots: The Hidden Half* (eds. Waisel, Y. et al.) 205–220 (Marcel Dekker, 2006).
21. Ledo, A. et al. Tree size and climatic water deficit control root to shoot ratio in individual trees globally. *New Phytol.* **217**, 8–11 (2018).
22. Qi, Y., Wei, W., Chen, C. & Chen, L. Plant root-shoot biomass allocation over diverse biomes: a global synthesis. *Glob. Ecol. Conserv.* **18**, e00606 (2019).
23. Reich, P. B. et al. Temperature drives global patterns in forest biomass distribution in leaves, stems, and roots. *Proc. Natl Acad. Sci. USA* **111**, 13721–13726 (2014).
24. De Frenne, P. et al. Latitudinal gradients as natural laboratories to infer species' responses to temperature. *J. Ecol.* **101**, 784–795 (2013).
25. Luo, Y. Terrestrial carbon-cycle feedback to climate warming. *Annu. Rev. Ecol. Evol. Syst.* **38**, 683–712 (2007).
26. Jackson, R. B. et al. A global analysis of root distributions for terrestrial biomes. *Oecologia* **108**, 389–411 (1996).
27. Malhi, Y., Doughty, C. & Galbraith, D. The allocation of ecosystem net primary productivity in tropical forests. *Philos. Trans. R. Soc. Lond. B* **366**, 3225–3245 (2011).
28. Roberts, D. R. et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* **40**, 913–929 (2017).
29. Cairns, M. A., Brown, S., Helmer, E. H. & Baumgardner, G. A. Root biomass allocation in the world's upland forests. *Oecologia* **111**, 1–11 (1997).
30. McCarthy, M. C. & Enquist, B. J. Consistency between an allometric approach and optimal partitioning theory in global patterns of plant biomass allocation. *Funct. Ecol.* **21**, 713–720 (2007).
31. Barton, C. V. M. & Montagu, K. D. Effect of spacing and water availability on root:shoot ratio in *Eucalyptus camaldulensis*. *For. Ecol. Manage.* **221**, 52–62 (2006).
32. Enquist, B. J. & Niklas, K. J. Global allocation rules for patterns of biomass partitioning in seed plants. *Science* **295**, 1517–1520 (2002).
33. Goward, S. N., Tucker, C. J. & Dye, D. G. North American vegetation patterns observed with the NOAA-7 advanced very high resolution radiometer. *Vegetatio* **64**, 3–14 (1985).
34. Manzoni, S., Jackson, R. B., Trofymow, J. A. & Porporato, A. The global stoichiometry of litter nitrogen mineralization. *Science* **321**, 684–686 (2008).
35. Kaiser, C., Franklin, O., Dieckmann, U. & Richter, A. Microbial community dynamics alleviate stoichiometric constraints during litter decay. *Ecol. Lett.* **17**, 680–690 (2014).
36. Jiao, F., Shi, X. R., Han, F. P. & Yuan, Z. Y. Increasing aridity, temperature and soil pH induce soil C-N-P imbalance in grasslands. *Sci. Rep.* **6**, 19601 (2016).
37. Sitch, S. et al. Recent trends and drivers of regional sources and sinks of carbon dioxide. *Biogeosciences* **12**, 653–679 (2015).
38. De Deyn, G. B., Cornelissen, J. H. C. & Bardgett, R. D. Plant functional traits and soil carbon sequestration in contrasting biomes. *Ecol. Lett.* **11**, 516–531 (2008).
39. Tjoelker, M. G., Craine, J. M., Wedin, D., Reich, P. B. & Tilman, D. Linking leaf and root trait syndromes among 39 grassland and savannah species. *New Phytol.* **167**, 493–508 (2005).
40. Personeni, E. & Loiseau, P. How does the nature of living and dead roots affect the residence time of carbon in the root litter continuum? *Plant Soil* **267**, 129–141 (2004).
41. Tuanmu, M. N. & Jetz, W. A global 1-km consensus land-cover product for biodiversity and ecosystem modelling. *Glob. Ecol. Biogeogr.* **23**, 1031–1045 (2014).

42. Pan, Y., Birdsey, R. A., Phillips, O. L. & Jackson, R. B. The structure, distribution, and biomass of the world's forests. *Annu. Rev. Ecol. Evol. Syst.* **44**, 593–622 (2013).
43. Jackson, R. B., Mooney, H. A. & Schulze, E. D. A global budget for fine root biomass, surface area, and nutrient contents. *Proc. Natl Acad. Sci. USA* **94**, 7362–7366 (1997).
44. Genet, H., Bréda, N. & Dufrêne, E. Age-related variation in carbon allocation at tree and stand scales in beech (*Fagus sylvatica* L.) and sessile oak (*Quercus petraea* (Matt.) Liebl.) using a chronosequence approach. *Tree Physiol.* **30**, 177–192 (2009).
45. De Castro, E. A. & Kauffman, J. B. Ecosystem structure in the Brazilian Cerrado: a vegetation gradient of aboveground biomass, root mass and consumption by fire. *J. Trop. Ecol.* **14**, 263–283 (1998).
46. Ding, B. & Sun, J. Study on biomass of Korean pine plantation in east mountain areas of northeast China. *Bull. Bot. Res.* **9**, 149–157 (1989).
47. Ding, B., Liu, S. & Cai, T. Studies on biological productivity of artificial forests of Dahurian larches. *Chin. J. Plant Ecol.* **14**, 226–236 (1990).
48. Ding, B. & Sun, J. Accumulation and distribution of productivity and nutrient element in natural Manchurian ash. *J. Northeast For. Univ.* **4**, 1–9 (1989).
49. Dossa, E. L., Fernandes, E. C. M., Reid, W. S. & Ezui, K. Above- and belowground biomass, nutrient and carbon stocks contrasting an open-grown and a shaded coffee plantation. *Agrofor. Syst.* **72**, 103–115 (2008).
50. Epron, D. et al. Do changes in carbon allocation account for the growth response to potassium and sodium applications in tropical *Eucalyptus* plantations? *Tree Physiol.* **32**, 667–679 (2012).
51. Fonseca, W., Rey Benayas, J. M. & Alice, F. E. Carbon accumulation in the biomass and soil of different aged secondary forests in the humid tropics of Costa Rica. *For. Ecol. Manage.* **262**, 1400–1408 (2011).
52. Goodman, R. C. et al. Amazon palm biomass and allometry. *For. Ecol. Manage.* **310**, 994–1004 (2013).
53. Greenland, D. J. & Kowal, J. M. L. Nutrient content of the moist tropical forest of Ghana. *Plant Soil* **12**, 154–173 (1960).
54. He, Y. et al. Carbon storage capacity of monoculture and mixed-species plantations in subtropical China. *For. Ecol. Manage.* **295**, 193–198 (2013).
55. Aiba, M. & Nakashizuka, T. Variation in juvenile survival and related physiological traits among dipterocarp species co-existing in a Bornean forest. *J. Veg. Sci.* **18**, 379–388 (2007).
56. Jha, K. K. Carbon storage and sequestration rate assessment and allometric model development in young teak plantations of tropical moist deciduous forest, India. *J. For. Res.* **26**, 589–604 (2015).
57. Kalita, R. M., Das, A. K. & Nath, A. J. Allometric equations for estimating above- and belowground biomass in Tea (*Camellia sinensis* (L.) O. Kuntze) agroforestry system of Barak Valley, Assam, northeast India. *Biomass Bioenergy* **83**, 42–49 (2015).
58. Kenzo, T. et al. Development of allometric relationships for accurate estimation of above- and below-ground biomass in tropical secondary forests in Sarawak, Malaysia. *J. Trop. Ecol.* **25**, 371–386 (2009).
59. Kenzo, T. et al. Allometric equations for accurate estimation of above-ground biomass in logged-over tropical rainforests in Sarawak, Malaysia. *J. For. Res.* **14**, 365–372 (2009).
60. Kraenzel, M., Castillo, A., Moore, T. & Potvin, C. Carbon storage of harvest-age teak (*Tectona grandis*) plantations, Panama. *For. Ecol. Manage.* **173**, 213–225 (2003).
61. Kuyah, S., Dietz, J., Muthuri, C., van Noordwijk, M. & Neufeldt, H. Allometry and partitioning of above- and below-ground biomass in farmed eucalyptus species dominant in Western Kenyan agricultural landscapes. *Biomass Bioenergy* **55**, 276–284 (2013).
62. Liu, S., Cai, Y. & Cai, T. in *Long-term Research on Forest Ecosystems* (ed. Zhou, X.) 419–427 (Northeast Forestry Univ. Press, 1991).
63. Luo, T. et al. Root biomass along subtropical to alpine gradients: global implication from Tibetan transect studies. *For. Ecol. Manage.* **206**, 349–363 (2005).
64. Markesteijn, L. & Poorter, L. Seedling root morphology and biomass allocation of 62 tropical tree species in relation to drought- and shade-tolerance. *J. Ecol.* **97**, 311–325 (2009).
65. McNicol, I. M. et al. Development of allometric models for above and belowground biomass in swidden cultivation fallows of northern Laos. *For. Ecol. Manage.* **357**, 104–116 (2015).
66. Aiba, M. & Nakashizuka, T. Sapling structure and regeneration strategy in 18 *Shorea* species co-occurring in a tropical rainforest. *Ann. Bot.* **96**, 313–321 (2005).
67. Menaut, J. C. & Cesar, J. Structure and primary productivity of Lamto savannas, Ivory Coast. *Ecology* **60**, 1197–1210 (1979).
68. Morais, V. A. et al. Estoques de carbono e biomassa de um fragmento de cerradão em Minas Gerais, Brasil. *Cerne* **19**, 237–245 (2013).
69. Mugasha, W. A. et al. Allometric models for prediction of above- and belowground biomass of trees in the miombo woodlands of Tanzania. *For. Ecol. Manage.* **310**, 87–101 (2013).
70. Návar, J. Plasticity of biomass component allocation patterns in semiarid Tamaulipan thornscrub and dry temperate pine species of northeastern Mexico. *Polibotánica* **31**, 121–141 (2011).
71. Njana, M. A., Eid, T., Zahabu, E. & Malimbwi, R. Procedures for quantification of belowground biomass of three mangrove tree species. *Wetl. Ecol. Manage.* **23**, 749–764 (2015).
72. Nogueira Junior, L. R., Engel, V. L., Parrotta, J. A., de Melo, A. C. G. & Ré, D. S. Equações alométricas para estimativa da biomassa arbórea em plantios mistos com espécies nativas na restauração da Mata Atlântica. *Biota Neotrop.* **14**, 1–9 (2014).
73. Peichl, M. & Arain, M. A. Above- and belowground ecosystem biomass and carbon pools in an age-sequence of temperate pine plantation forests. *Agric. For. Meteorol.* **140**, e20130084 (2006).
74. Battles, J. J. et al. Vegetation composition, structure, and biomass of two unpolluted watersheds in the Cordillera de Piuchué, Chiloé Island, Chile. *Plant Ecol.* **158**, 5–19 (2002).
75. Ryan, C. M., Williams, M. & Grace, J. Above- and belowground carbon stocks in a miombo woodland landscape of Mozambique. *Biotropica* **43**, 423–432 (2011).
76. Saint-André, L. et al. Age-related equations for above- and below-ground biomass of a *Eucalyptus* hybrid in Congo. *For. Ecol. Manage.* **205**, 199–214 (2005).
77. Aryal, D. R., De Jong, B. H. J., Ochoa-Gaona, S., Esparza-Olguin, L. & Mendoza-Vega, J. Carbon stocks and changes in tropical secondary forests of southern Mexico. *Agric. Ecosyst. Environ.* **195**, 220–230 (2014).
78. Schepaschenko, D. et al. A dataset of forest biomass structure for Eurasia. *Sci. Data* **4**, 170070 (2017).
79. Schroth, G., D'Angelo, S. A., Teixeira, W. G., Haag, D. & Lieberei, R. Conversion of secondary forest into agroforestry and monoculture plantations in Amazonia: consequences for biomass, litter and soil carbon stocks after 7 years. *For. Ecol. Manage.* **163**, 131–150 (2002).
80. Schulze, E. D. et al. Rooting depth, water availability, and vegetation cover along an aridity gradient in Patagonia. *Oecologia* **108**, 503–511 (1996).
81. Stolbovoi, V. & McCallum, I. *Land resources of Russia* [CD] (International Institute for Applied Systems Analysis and the Russian Academy of Science, 2002); http://www.iiasa.ac.at/Research/FOR/russia_cd/guide.htm
82. Wang, L. et al. Biomass allocation patterns across China's terrestrial biomes. *PLoS ONE* **9**, e93566 (2014).
83. Wauters, J. B., Coudert, S., Grallien, E., Jonard, M. & Ponette, Q. Carbon stock in rubber tree plantations in Western Ghana and Mato Grosso (Brazil). *For. Ecol. Manage.* **255**, 2347–2361 (2008).
84. Williams-Linera, G. Biomass and nutrient content in two successional stages of tropical wet forest in Uxpanapa, Mexico. *Biotropica* **15**, 275–284 (1983).
85. Xu, Y. et al. Improving allometry models to estimate the above- and belowground biomass of subtropical forest, China. *Ecosphere* **6**, 289 (2015).
86. Youkhana, A. H. & Idol, T. W. Allometric models for predicting above- and belowground biomass of *Leucaena*-KX2 in a shaded coffee agroecosystem in Hawaii. *Agrofor. Syst.* **83**, 331–345 (2011).
87. Zhang, H. et al. Biogeographical patterns of biomass allocation in leaves, stems, and roots in China's forests. *Sci. Rep.* **5**, 15997 (2015).
88. Castellanos, J., Maass, M. & Kummerow, J. Root biomass of a dry deciduous tropical forest in Mexico. *Plant Soil* **131**, 225–228 (1991).
89. Zheng, Z., Feng, Z., Cao, M., Li, Z. & Zhang, J. Forest structure and biomass of a tropical seasonal rain forest in Xishuangbanna, southwest China. *Biotropica* **38**, 318–327 (2006).
90. Návar, J. Root stock biomass and productivity assessments of reforested pine stands in northern Mexico. *For. Ecol. Manage.* **338**, 139–147 (2015).
91. Wang, X., Fang, J. & Zhu, B. Forest biomass and root–shoot allocation in northeast China. *For. Ecol. Manage.* **255**, 4007–4020 (2008).
92. Chen, D. K., Zhou, X. F., Zhao, H. X., Wang, Y. H. & Jing, Y. Y. Study on the structure, function and succession of the four types in natural secondary forest. *J. Northeast For. Univ.* **2**, 1–20 (1982).
93. Chidumayo, E. N. Estimating tree biomass and changes in root biomass following clear-cutting of *Brachystegia-Julbernardia* (miombo) woodland in central Zambia. *Environ. Conserv.* **41**, 54–63 (2014).
94. Coll, L., Potvin, C., Messier, C. & Delagrange, S. Root architecture and allocation patterns of eight native tropical species with different successional status used in open-grown mixed plantations in Panama. *Trees* **22**, 585–596 (2008).
95. Das, D. K. & Chaturvedi, O. P. Structure and function of *Populus deltoides* agroforestry systems in eastern India: 1. dry matter dynamics. *Agrofor. Syst.* **65**, 215–221 (2005).
96. Ni, J. Estimating net primary productivity of grasslands from field biomass measurements in temperate northern China. *Plant Ecol.* **174**, 217–234 (2011).
97. Olson, R. et al. *NPP Multi-Biome: Summary Data from Intensive Studies at 125 Sites, 1936–2006* (ORNL DAAC, accessed 19 June 2019); https://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=1352

98. Perez, C. A. & Frangi, J. L. Grassland biomass dynamics along an altitudinal gradient in the pampa. *J. Range Manage.* **53**, 518–528 (2007).
99. Perez-Quezada, J. F. F., Delpiano, C. A. A., Snyder, K. A. A., Johnson, D. A. A. & Franck, N. Carbon pools in an arid shrubland in Chile under natural and afforested conditions. *J. Arid Environ.* **75**, 29–37 (2011).
100. Pornon, A., Boutin, M. & Lamaze, T. Contribution of plant species to the high N retention capacity of a subalpine meadow undergoing elevated N deposition and warming. *Environ. Pollut.* **245**, 235–242 (2019).
101. Ramakrishnan, P. S. & Ram, S. C. Vegetation, biomass and productivity of seral grasslands of Cherrapunji in north-east India. *Vegetatio* **74**, 47–53 (1988).
102. Shaver, G. R., Laundre, J. A., Giblin, A. E. & Nadelhoffer, K. J. Changes in live plant biomass, primary production, and species composition along a riverside toposequence in Arctic Alaska, USA. *Arct. Alp. Res.* **28**, 363–379 (2006).
103. Smith, J. M. B. & Klinger, L. F. Aboveground:belowground phytomass ratios in Venezuelan paramo vegetation and their significance. *Arct. Alp. Res.* **17**, 189–198 (2006).
104. Sun, J. et al. Effects of grazing regimes on plant traits and soil nutrients in an alpine steppe, northern Tibetan Plateau. *PLoS ONE* **9**, e108821 (2014).
105. Wang, P. et al. Belowground plant biomass allocation in tundra ecosystems and its relationship with temperature. *Environ. Res. Lett.* **11**, 055003 (2016).
106. Yang, Y., Fang, J., Ji, C. & Han, W. Above- and belowground biomass allocation in Tibetan grasslands. *J. Veg. Sci.* **20**, 177–184 (2009).
107. Yang, Y., Fang, J., Ma, W., Guo, D. & Mohammat, A. Large-scale pattern of biomass partitioning across China's grasslands. *Glob. Ecol. Biogeogr.* **19**, 268–277 (2010).
108. Geng, H. L., Wang, Y. H., Wang, F. Y. & Jia, B. R. The dynamics of root-shoot ratio and its environmental effective factors of recovering *Leymus chinensis* steppe vegetation in Inner Mongolia, China. *Acta Ecol. Sin.* **28**, 4629–4634 (2008).
109. Hui, D. & Jackson, R. B. Geographical and interannual variability in biomass partitioning in grassland ecosystems: a synthesis of field data. *New Phytol.* **169**, 85–93 (2006).
110. Jouquet, P., Tavernier, V., Abbadie, L. & Lepage, M. Nests of subterranean fungus-growing termites (Isoptera, Macrotermitinae) as nutrient patches for grasses in savannah ecosystems. *Afr. J. Ecol.* **43**, 191–196 (2005).
111. Leonid, U. et al. Impact of climate and grazing on biomass components of eastern Russia typical steppe. *J. Integr. Agric.* **13**, 1183–1192 (2014).
112. Lucash, M. S., Farnsworth, B. & Winner, W. E. Response of sagebrush steppe species to elevated CO_2 and soil temperature. *West. N. Am. Nat.* **65**, 80–86 (2005).
113. Luo, W. et al. Patterns of plant biomass allocation in temperate grasslands across a 2500-km transect in northern China. *PLoS ONE* **8**, e71749 (2013).
114. Barbour, M. G. Desert dogma reexamined: root/shoot productivity and plant spacing. *Am. Midl. Nat.* **89**, 41–57 (1973).
115. Becker, P., Sharbini, N. & Yahya, R. Root architecture and root:shoot allocation of shrubs and saplings in two lowland tropical forests: implications for life-form composition. *Biotropica* **31**, 93–101 (1999).
116. Becker, P. & Castillo, A. Root architecture of shrubs and saplings in the understory of a tropical moist forest in lowland Panama. *Biotropica* **22**, 242–249 (1990).
117. Beier, C. et al. Carbon and nitrogen balances for six shrublands across Europe. *Glob. Biogeochem. Cycles* **23**, GB4008 (2009).
118. Bhatt, Y. D., Rawat, Y. S. & Singh, S. P. Changes in ecosystem functioning after replacement of forest by *Lantana* shrubland in Kumaun Himalaya. *J. Veg. Sci.* **5**, 67–70 (1994).
119. Caldwell, M. M., White, R. S., Moore, R. T. & Camp, L. B. Carbon balance, productivity, and water use of cold-winter desert shrub communities dominated by C_3 and C_4 species. *Oecologia* **29**, 275–300 (1977).
120. De Viñas, I. C. R. et al. Biomass of root and shoot systems of *Quercus coccifera* shrublands in eastern Spain. *Ann. For. Sci.* **57**, 803–810 (2000).
121. Caravaca, F., Figueroa, D., Alguacil, M. M. & Roldán, A. Application of composted urban residue enhanced the performance of afforested shrub species in a degraded semiarid land. *Bioresour. Technol.* **90**, 65–70 (2003).
122. Caravaca, F., Figueroa, D., Azcón-Aguilar, C., Barea, J. M. & Roldán, A. Medium-term effects of mycorrhizal inoculation and composted municipal waste addition on the establishment of two Mediterranean shrub species under semiarid field conditions. *Agric. Ecosyst. Environ.* **97**, 95–105 (2003).
123. Carrasco, L., Azcón, R., Kohler, J., Roldán, A. & Caravaca, F. Comparative effects of native filamentous and arbuscular mycorrhizal fungi in the establishment of an autochthonous, leguminous shrub growing in a metal-contaminated soil. *Sci. Total Environ.* **409**, 1205–1209 (2011).
124. Carrillo-Garcia, Á., Bashan, Y. & Bethlenfalvay, G. J. Resource-island soils and the survival of the giant cactus, cardon, of Baja California Sur. *Plant Soil* **218**, 207–214 (2000).
125. Carrión-Prieto, P. et al. Mediterranean shrublands as carbon sinks for climate change mitigation: new root-to-shoot ratios. *Carbon Manage.* **8**, 67–77 (2017).
126. Deng, L., Han, Q. S., Zhang, C., Tang, Z. S. & Shangguan, Z. P. Above-ground and below-ground ecosystem biomass accumulation and carbon sequestration with *Caragana korshinskii* Kom plantation development. *Land Degrad. Dev.* **28**, 906–917 (2017).
127. Perkins, S. R. & Owens, M. K. Growth and biomass allocation of shrub and grass seedlings in response to predicted changes in precipitation seasonality. *Plant Ecol.* **168**, 107–120 (2003).
128. Gargaglione, V., Peri, P. L. & Rubio, G. Allometric relations for biomass partitioning of *Nothofagus antarctica* trees of different crown classes over a site quality gradient. *For. Ecol. Manage.* **259**, 1118–1126 (2010).
129. Hao, H. M. et al. Effects of shrub patch size succession on plant diversity and soil water content in the water-wind erosion crisscross region on the Loess Plateau. *Catena* **144**, 177–183 (2016).
130. Herwitz, S. R. & Olsvig-Whittaker, L. Preferential upslope growth of *Zygophyllum dumosum* Boiss. (Zygophyllaceae) roots into bedrock fissures in the northern Negev desert. *J. Biogeogr.* **16**, 457–460 (1989).
131. Hoffmann, A. & Kummerow, J. Root studies in the Chilean matorral. *Oecologia* **32**, 57–69 (1978).
132. Holl, K. D. Effects of above- and below-ground competition of shrubs and grass on *Calophyllum brasiliense* (Camb.) seedling growth in abandoned tropical pasture. *For. Ecol. Manage.* **109**, 187–195 (1998).
133. Hollister, R. D. & Flaherty, K. J. Above- and below-ground plant biomass response to experimental warming in northern Alaska. *Appl. Veg. Sci.* **13**, 378–387 (2010).
134. Kizito, F. et al. Seasonal soil water variation and root patterns between two semi-arid shrubs co-existing with pearl millet in Senegal, West Africa. *J. Arid Environ.* **67**, 436–455 (2006).
135. Kummerow, J., Krause, D. & Jow, W. Root systems of chaparral shrubs. *Oecologia* **29**, 163–177 (1977).
136. León, M. F., Squeo, F. A., Gutiérrez, J. R. & Holmgren, M. Rapid root extension during water pulses enhances establishment of shrub seedlings in the Atacama Desert. *J. Veg. Sci.* **22**, 120–129 (2011).
137. Li, C. P. & Xiao, C. W. Above- and belowground biomass of *Artemisia ordosica* communities in three contrasting habitats of the Mu Us Desert, northern China. *J. Arid Environ.* **70**, 195–207 (2007).
138. Liang, Y. M., Hazlett, D. L. & Lauenroth, W. K. Biomass dynamics and water use efficiencies of five plant communities in the shortgrass steppe. *Oecologia* **80**, 148–153 (1989).
139. Zan, Q., Wang, Y., Liao, B. & Zheng, D. Biomass and net productivity of *Sonneratia apetala*, *S. caseolaris* mangrove man-made forest. *Wuhan Bot. Res.* **19**, 391–396 (2001).
140. Liao, B., Zheng, D. & Zheng, S. Studies on the biomass of *Sonneratia caseolaris* stand. *For. Res.* **3**, 47–54 (1990).
141. Lufafa, A. et al. Allometric relationships and peak-season community biomass stocks of native shrubs in Senegal's Peanut Basin. *J. Arid Environ.* **73**, 260–266 (2009).
142. Lusk, C. H. Leaf area and growth of juvenile temperate evergreens in low light: species of contrasting shade tolerance change rank during ontogeny. *Funct. Ecol.* **18**, 820–828 (2004).
143. Marsh, A. S., Arnone, J. A., Bormann, B. T. & Gordon, J. C. The role of *Equisetum* in nutrient cycling in an Alaskan shrub wetland. *J. Ecol.* **88**, 999–1011 (2000).
144. Martínez, F. et al. Belowground structure and production in a Mediterranean sand dune shrub community. *Plant Soil* **201**, 209–216 (1998).
145. Marziliano, P. A. et al. Estimating belowground biomass and root/shoot ratio of *Phillyrea latifolia* L. in the Mediterranean forest landscapes. *Ann. For. Sci.* **72**, 585–593 (2015).
146. Mauchamp, A., Montaña, C., Lepart, J., Rambal, S. & Montana, C. Ecotone dependent recruitment of a desert shrub, *Flourensia cernua*, in vegetation stripes. *Oikos* **68**, 107–116 (1993).
147. Mendoza-Ponce, A. & Galicia, L. Aboveground and belowground biomass and carbon pools in highland temperate forest landscape in central Mexico. *Forestry* **83**, 497–506 (2010).
148. Miller, P. C. & Ng, E. Root:shoot biomass ratios in shrubs in southern California and central Chile. *Madrono* **24**, 215–223 (1977).
149. Mooney, H. A. & Rundel, P. W. Nutrient relations of the evergreen shrub, *Adenostoma fasciculatum*, in the California chaparral. *Bot. Gaz.* **140**, 109–113 (1979).
150. Moro, M. J., Pugnaire, F. I., Haase, P. & Puigdefábregas, J. Effect of the canopy of *Retama sphaerocarpa* on its understorey in a semiarid environment. *Funct. Ecol.* **11**, 425–431 (1997).
151. Negreiros, D., Fernandes, G. W., Silveira, F. A. O. & Chalub, C. Seedling growth and biomass allocation of endemic and threatened shrubs of rupestrian fields. *Acta Oecol.* **35**, 301–310 (2009).
152. Nie, X., Yang, Y., Yang, L. & Zhou, G. Above- and belowground biomass allocation in shrub biomes across the northeast Tibetan Plateau. *PLoS ONE* **11**, e0154251 (2016).
153. Nobel, P. S., Quero, E. & Linares, H. Root versus shoot biomass: responses to water, nitrogen, and phosphorus applications for *Agave lechuguilla*. *Bot. Gaz.* **150**, 411–416 (1989).

154. Pacaldo, R. S., Volk, T. A. & Briggs, R. D. Greenhouse gas potentials of shrub willow biomass crops based on below- and aboveground biomass inventory along a 19-year chronosequence. *Bioenergy Res.* **6**, 252–262 (2013).
155. Padilla, F. M., Miranda, J. D., Jorquera, M. J. & Pugnaire, F. I. Variability in amount and frequency of water supply affects roots but not growth of arid shrubs. *Plant Ecol.* **204**, 261–270 (2009).
156. Portsmuth, A., Niinemets, Ü., Truus, L. & Pensa, M. Biomass allocation and growth rates in *Pinus sylvestris* are interactively modified by nitrogen and phosphorus availabilities and by tree size and age. *Can. J. For. Res.* **35**, 2346–2359 (2005).
157. Roth, G. A., Whitford, W. G. & Steinberger, Y. Jackrabbit (*Lepus californicus*) herbivory changes dominance in desertified Chihuahuan Desert ecosystems. *J. Arid Environ.* **70**, 418–426 (2007).
158. Ruiz-Peinado, R., Moreno, G., Juarez, E., Montero, G. & Roig, S. The contribution of two common shrub species to aboveground and belowground carbon stock in Iberian dehesas. *J. Arid Environ.* **91**, 22–30 (2013).
159. Rundel, P. W. Biomass, productivity, and nutrient allocation in subalpine shrublands and meadows of the Emerald Lake Basin, Sequoia National Park, California. *Arct. Antarct. Alp. Res.* **47**, 115–123 (2015).
160. Millikin, C. S. & Bledsoe, C. S. Biomass and distribution of fine and coarse roots from blue oak (*Quercus douglasii*) trees in the northern Sierra Nevada foothills of California. *Plant Soil* **214**, 27–38 (1999).
161. Saura-Mas, S. & Lloret, F. Adult root structure of Mediterranean shrubs: relationship with post-fire regenerative syndrome. *Plant Biol.* **16**, 147–154 (2014).
162. Schenk, H. J. & Mahall, B. E. Positive and negative plant interactions contribute to a north-south-patterned association between two desert shrub species. *Oecologia* **132**, 402–410 (2002).
163. Silva, J. S., Rego, F. C. & Martins-Loução, M. A. Belowground traits of Mediterranean woody plants in a Portuguese shrubland. *Ecol. Mediterr.* **28**, 5–13 (2002).
164. Simões, M. P., Madeira, M. & Gazarini, L. Biomass and nutrient dynamics in Mediterranean seasonal dimorphic shrubs: strategies to face environmental constraints. *Plant Biosyst.* **146**, 500–510 (2012).
165. Tao, Y., Zhang, Y. M. & Downing, A. Similarity and difference in vegetation structure of three desert shrub communities under the same temperate climate but with different microhabitats. *Bot. Stud.* **54**, 59 (2013).
166. Toscano, S., Scuderi, D., Giuffrida, F. & Romano, D. Responses of Mediterranean ornamental shrubs to drought stress and recovery. *Sci. Hortic.* **178**, 145–153 (2014).
167. Trubat, R., Cortina, J. & Vilagrosa, A. Nutrient deprivation improves field performance of woody seedlings in a degraded semi-arid shrubland. *Ecol. Eng.* **37**, 1164–1173 (2011).
168. Van Wijk, M. T., Williams, M., Gough, L., Hobbie, S. E. & Shaver, G. R. Luxury consumption of soil nutrients: a possible competitive strategy in above-ground and below-ground biomass allocation and root morphology for slow-growing arctic vegetation? *J. Ecol.* **91**, 664–676 (2003).
169. Walker, L. R., Clarkson, B. D., Silvester, W. B. & Clarkson, B. R. Colonization dynamics and facilitative impacts of a nitrogen-fixing shrub in primary succession. *J. Veg. Sci.* **14**, 277–290 (2003).
170. Wang, B. & Yang, X. S. Comparison of biomass and species diversity of four typical zonal vegetations. *J. Fujian Coll. For.* **29**, 345–350 (2009).
171. Wang, M. & Li, H. Quantitative study on the soil water dynamics of various forest plantations in the Loess Plateau region in northwestern Shanxi. *Acta Ecol. Sin.* **2**, 178–184 (1995).
172. Wang, P. et al. Seasonal changes and vertical distribution of root standing biomass of graminoids and shrubs at a Siberian tundra site. *Plant Soil* **407**, 55–65 (2016).
173. Whittaker, R. H. & Woodwell, G. M. Dimension and production relations of trees and shrubs in the Brookhaven Forest, New York. *J. Ecol.* **56**, 1–25 (1968).
174. Xu, H., Li, Y., Xu, G. & Zou, T. Ecophysiological response and morphological adjustment of two Central Asian desert shrubs towards variation in summer precipitation. *Plant Cell Environ.* **30**, 399–409 (2007).
175. Yan, Z. Biomass and its allocation in a 28-year-old *Castanopsis kawakamii* plantation. *J. Fujian Coll. For.* **2**, 114–118 (1996).
176. Gong, Y. et al. Carbon storage and vertical distribution in three shrubland communities in Gurbantünggüt Desert, Uygur Autonomous Region of Xinjiang, northwest China. *Chin. Geogr. Sci.* **22**, 541–549 (2012).
177. Yu, Y., Shi, D., Qiuyi, J., He, L. & Cheng, G. On the biomass of secondary *Schima superba* forest in Hangzhou. *J. Zhejiang For. Coll.* **2**, 157–161 (1993).
178. Kato, T. et al. Carbon dioxide exchange between the atmosphere and an alpine meadow ecosystem on the Qinghai-Tibetan Plateau, China. *Agric. Meteorol.* **124**, 121–134 (2004).
179. Li, Z., Zhu, Q. & Li, J. A comparison of photosynthetic carbon sequestration of four shrubs in Ningxia. *Pratacultural Sci.* **29**, 352–357 (2012).
180. Zhu, X., Shi, Q. & Li, Y. A preliminary study on the Qinghai's treasure house of the forest biomass and shrubs. *Sci. Technol. Qinghai Agric. For.* **1**, 15–20 (1993).
181. Liao, B. & Zheng, D. Study on the forest biomass and productivity of olive wood. *For. Res.* **4**, 22–29 (1991).
182. Liu, B., Liu, Z., Lü, X., Maestre, F. T. & Wang, L. Sand burial compensates for the negative effects of erosion on the dune-building shrub *Artemisia wudanica*. *Plant Soil* **374**, 263–273 (2014).
183. Alguacil, M. M., Hernández, J. A., Caravaca, F., Portillo, B. & Roldán, A. Antioxidant enzyme activities in shoots from three mycorrhizal shrub species afforested in a degraded semi-arid soil. *Physiol. Plant.* **118**, 562–570 (2003).
184. Axe, M. S., Grange, I. D. & Conway, J. S. Carbon storage in hedge biomass—a case study of actively managed hedges in England. *Agric. Ecosyst. Environ.* **250**, 81–88 (2017).
185. van den Hoogen, J. et al. Soil nematode abundance and functional group composition at a global scale. *Nature* **572**, 194–198 (2019).
186. Erin, L. et al. h2o: R Interface for the 'H2O' Scalable Machine Learning Platform. R package v.3.32.0.2 (2020); https://github.com/h2oai/h2o-3
187. Sagi, O. & Rokach, L. Ensemble learning: a survey. *WIREs Data Min. Knowl. Discov.* **8**, e1249 (2018).
188. R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2019).
189. Gorelick, N. et al. Google Earth Engine: planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **202**, 18–27 (2017).
190. Heiberger, R. M. HH: Statistical Analysis and Data Display: Heiberger and Holland (2020).
191. Hothorn, T. & Zeileis, A. partykit: A modular toolkit for recursive partytioning in R. *J. Mach. Learn. Res.* **16**, 3905–3909 (2015).
192. Borkovec, M. & Madin, N. ggparty: 'ggplot' visualizations for the 'partykit' package (2019).
193. Dormann, C. F. Effects of incorporating spatial autocorrelation into the analysis of species distribution data. *Glob. Ecol. Biogeogr.* **16**, 129–138 (2007).
194. Hutchinson, M., Xu, T., Houlder, D., Nix, H. & McMahon, J. *ANUCLIM 6.0 User's Guide* (Australian National Univ., 2009).
195. Fick, S. E. & Hijmans, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* **37**, 4302–4315 (2017).
196. *Global Aridity and PET database* (CGIAR-CSI, accessed 15 May 2018); http://www.cgiarcsi.community/data/global-aridity-and-pet-database
197. CIESIN *Gridded Population of the World, version 4 (GPWv4): Population Density Adjusted to Match 2015 Revision UN WPP Country Totals* (NASA SEDAC, 2018); https://doi.org/10.7927/H4HX19NJ
198. Venter, O. et al. Global terrestrial human footprint maps for 1993 and 2009. *Sci. Data* **3**, 160067 (2016).
199. *SoilGrids* (ISRIC, accessed 15 May 2018); https://www.soilgrids.org
200. Entekhabi, D. et al. The soil moisture active passive (SMAP) mission. *Proc. IEEE* **98**, 704–716 (2010).
201. Fan, Y., Li, H. & Miguez-Macho, G. Global patterns of groundwater table depth. *Science* **339**, 940–943 (2013).
202. Batjes, N. H. Harmonized soil property values for broad-scale modelling (WISE30sec) with estimates of global soil carbon stocks. *Geoderma* **269**, 61–68 (2016).
203. Schaaf, C. & Wang, Z. *MCD43A1 MODIS/Terra+Aqua BRDF/Albedo Model Parameters Daily L3 Global - 500m V006* (NASA LP DAAC, 2015); https://doi.org/10.5067/MODIS/MCD43A1C.006
204. Didan, K. *MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V006* (NASA LP DAAC, 2015).
205. Crowther, T. W. et al. Mapping tree density at a global scale. *Nature* **525**, 201–205 (2015).

Acknowledgements

We thank J.-F. Bastin, P. B. Reich, R. B. Jackson and Y. Zeng for their constructive comments on this study. This work was supported by grants to C.M.Z. from the ETH Zurich Postdoctoral Fellowship programme, L.M. from the China Scholarship Council and T.W.C. from DOB Ecology. B.D.S. was funded by the Swiss National Science Foundation grant no. PCEFP2_181115. C.T. was supported by a Lawrence Fellow award through the Lawrence Livermore National Laboratory, the US Department of Energy under contract DE-AC52-07NA27344 and the Lawrence Livermore National Laboratory LDRD (Laboratory Directed Research & Development) Program under project no. 20-ERD-055.

Author contributions

H.M., L.M., T.W.C. and C.M.Z. conceived and developed the study and wrote the manuscript. H.M. and L.M. collected the data. H.M. and L.M. performed the analyses. D.S.M., J.v.d.H., B.S. and C.T. gave input on the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41559-021-01485-1.

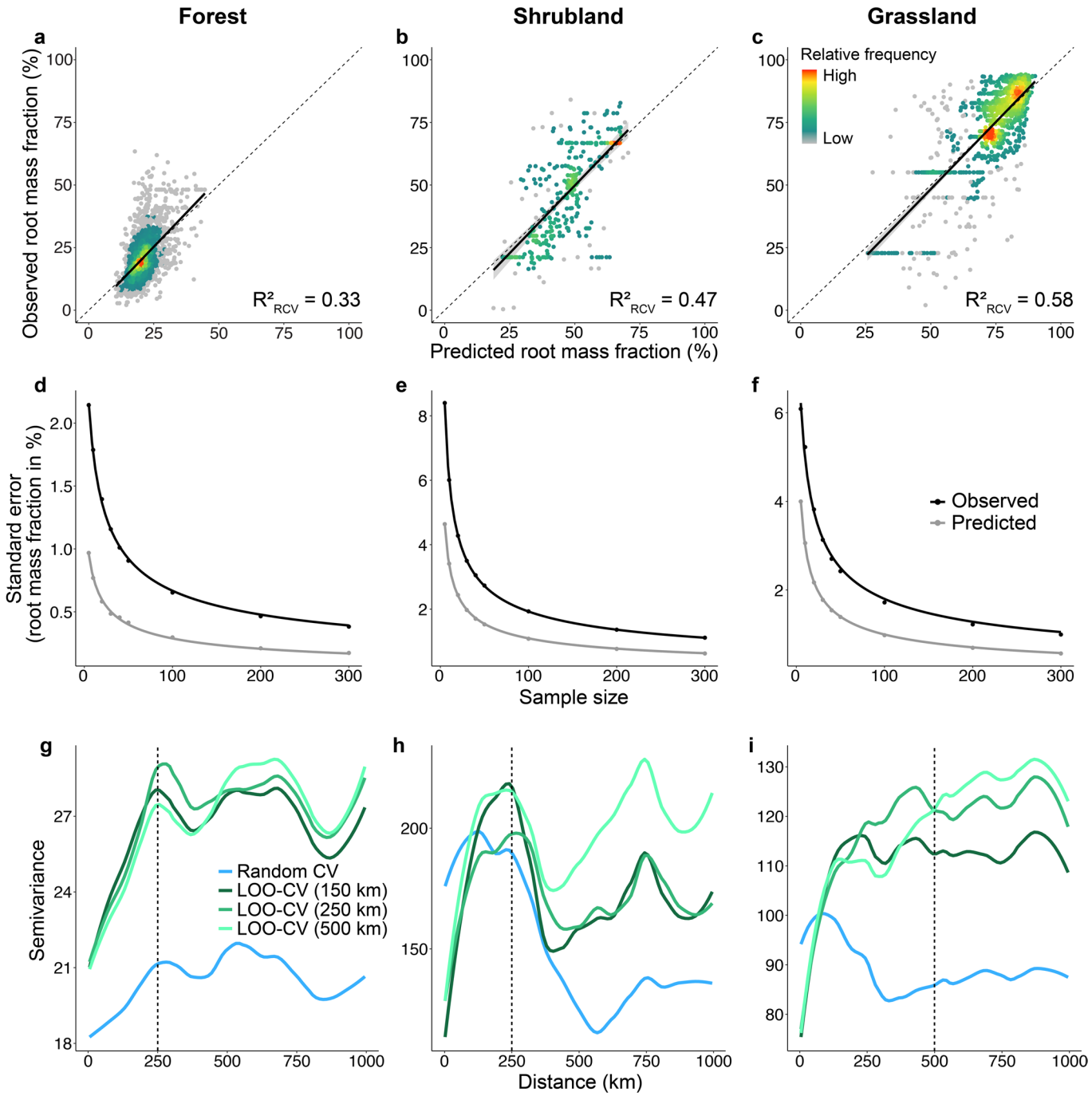
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41559-021-01485-1.

Correspondence and requests for materials should be addressed to C.M.Z.

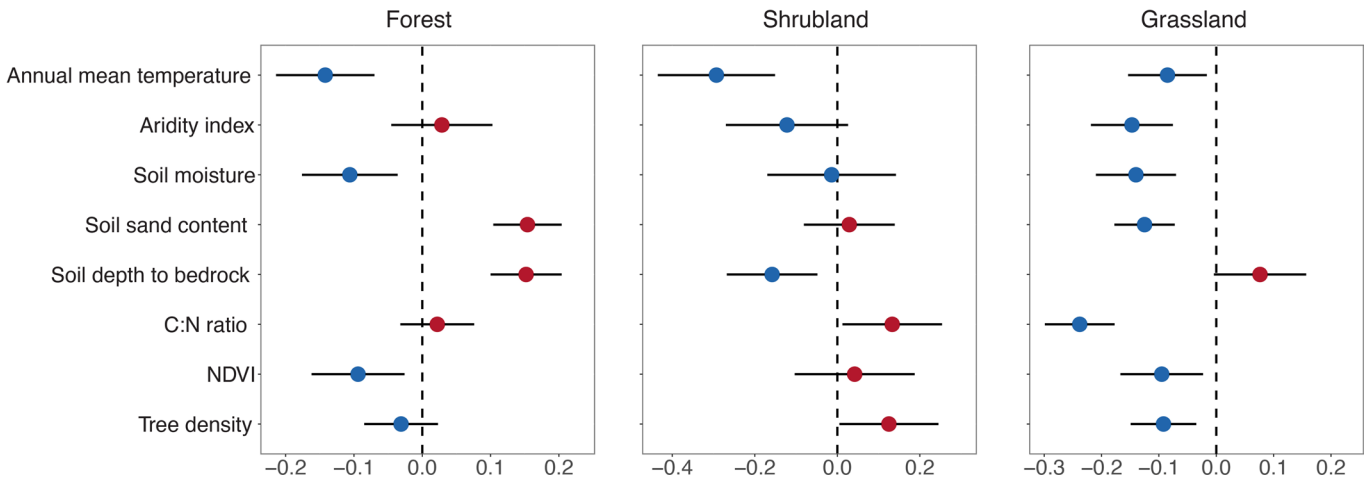
Peer review information *Nature Ecology & Evolution* thanks the anonymous reviewers for their contribution to the peer review of this work.

Variable	Variable group	Original spatial resolution	Units	Source
Annual mean radiation	Atmospheric	≈1km	W/m2	ref[194]
Highest weekly radiation	Atmospheric	≈1km	W/m2	
Lowest weekly radiation	Atmospheric	≈1km	W/m2	
Radiation of coldest quarter	Atmospheric	≈1km	W/m2	
Radiation of driest quarter	Atmospheric	≈1km	W/m2	
Radiation of warmest quarter	Atmospheric	≈1km	W/m2	
Radiation of wettest quarter	Atmospheric	≈1km	W/m2	
Radiation seasonality	Atmospheric	≈1km	W/m2	
Annual mean temperature	Climatic	≈1km	℃	ref[195]
Annual precipitation	Climatic	≈1km	mm	
Isothermality	Climatic	≈1km	%	
Maximum temperature of warmest month	Climatic	≈1km	℃	
Mean diurnal range	Climatic	≈1km	℃	
Mean temperature of coldest quarter	Climatic	≈1km	℃	
Mean temperature of driest quarter	Climatic	≈1km	℃	
Mean temperature of warmest quarter	Climatic	≈1km	℃	
Mean temperature of wettest quarter	Climatic	≈1km	℃	
Minimum temperature of coldest month	Climatic	≈1km	℃	
Precipitation of coldest quarter	Climatic	≈1km	mm	
Precipitation of driest month	Climatic	≈1km	mm	
Precipitation of driest quarter	Climatic	≈1km	mm	
Precipitation of warmest quarter	Climatic	≈1km	mm	
Precipitation of wettest month	Climatic	≈1km	mm	
Precipitation of wettest quarter	Climatic	≈1km	mm	
Precipitation seasonality	Climatic	≈1km	mm	
Temperature annual range	Climatic	≈1km	℃	
Temperature Seasonality	Climatic	≈1km	℃	
Global aridity index	Climatic	≈1km	Unitless	ref[196]
Global potential evapotranspiration (PET)	Climatic	≈1km	PET value	
Global population density	Anthropogenic	≈1km	persons/km2	ref[197]
Human development percentage	Anthropogenic	≈1km	%	ref[41]
Human footprint	Anthropogenic	≈1km	Unitless	ref[198]
Bulk density at depth 0~15cm	Soil	≈250m	kg/m3	ref[199]
Cation exchange capacity of soil at depth 0~15cm	Soil	≈250m	cmolc/kg	
Clay content at depth 0~15cm	Soil	≈250m	%	
Coarse fragments volumetric at depth 0~15cm	Soil	≈250m	%	
Depth to bedrock	Soil	≈250m	cm	
Probability of occurrence in the R horizon	Soil	≈250m	%	
Sand content at depth 0~15cm	Soil	≈250m	Mass fraction in %	
Silt content at depth 0~15cm	Soil	≈250m	Mass fraction in %	
Soil organic carbon content at depth 0~15cm	Soil	≈250m	g/kg	
Soil organic carbon stock at depth 0~15cm	Soil	≈250m	tones/hectare	
Soil pH*10 in H2O at depth 0~15cm	Soil	≈250m	pH*10	
Soil moisture profile	Soil	≈1km	%	ref[200]
Soil water table depth	Soil	≈1km	m	ref[201]
Soil Organic C:N ratio	Soil	≈1km		ref[202]
Soil total Nitrogen	Soil	≈1km	g/kg	
Eastness	Topographic	≈1km	-	Global Multi-resolution Terrain Elevation Data 2010
Elevation	Topographic	≈1km	Meters	
Hillshade	Topographic	≈1km	Grayscale value	
Northness	Topographic	≈1km	-	
Slope	Topographic	≈1km	Degree	
BRDF Albedo mandatory quality band 1	Vegetative	≈500m	reflectance band value	ref[203]
BRDF Albedo mandatory quality band 2	Vegetative	≈500m	reflectance band value	
BRDF Albedo mandatory quality band 3	Vegetative	≈500m	Reflectance band value	
BRDF Albedo mandatory quality band 4	Vegetative	≈500m	reflectance band value	
BRDF Albedo mandatory quality band 5	Vegetative	≈500m	reflectance band value	
BRDF Albedo mandatory quality band 6	Vegetative	≈500m	reflectance band value	
BRDF Albedo mandatory quality band 7	Vegetative	≈500m	reflectance band value	
Global 16-day enhanced vegetation indices (EVI)	Vegetative	≈250m	EVI value	ref[204]
Global 16-day vegetation indices (NDVI)	Vegetative	≈250m	NDVI value	
Tree density	Vegetative	≈1km	Number of trees/ha	ref[205]

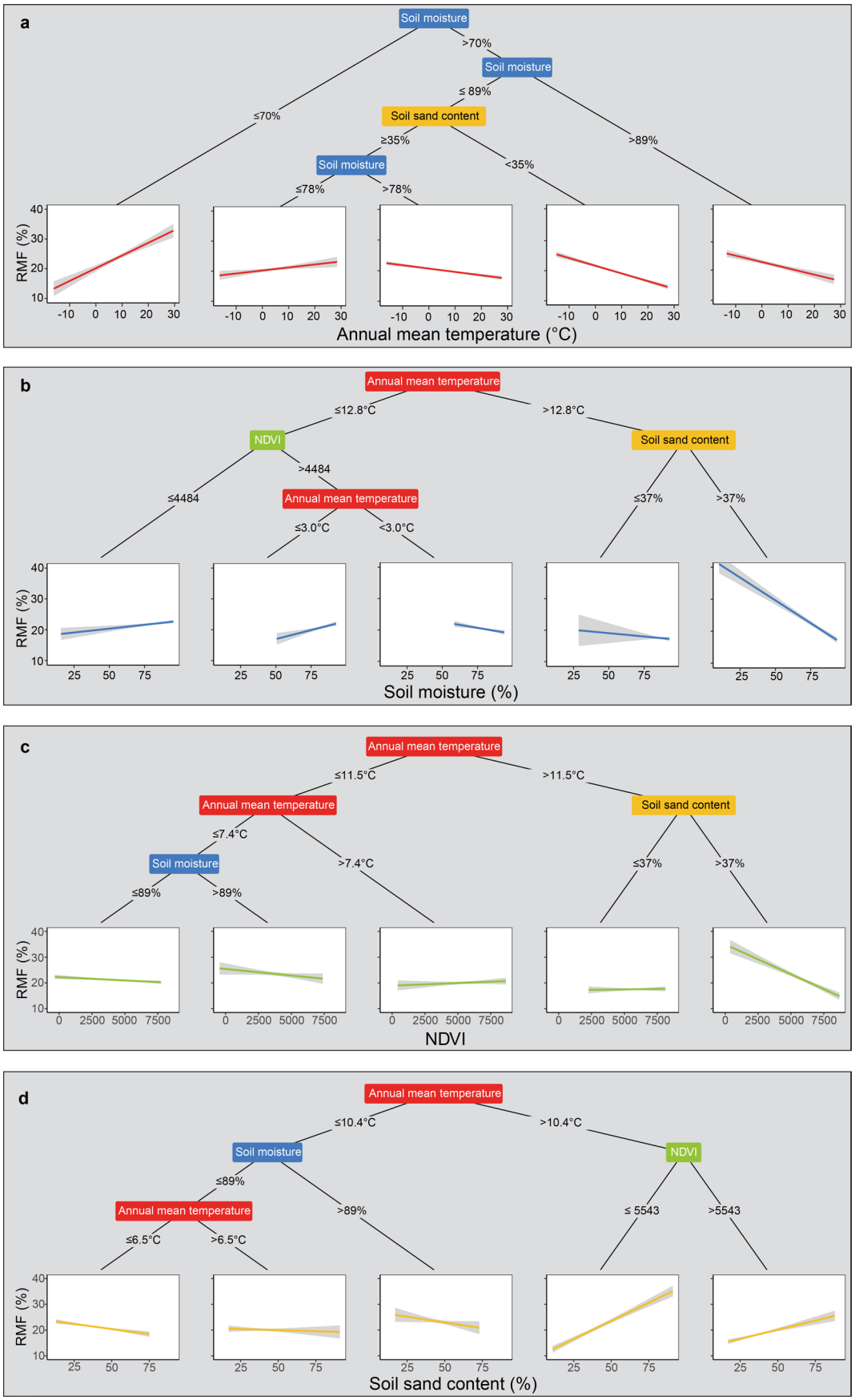
Extended Data Fig. 1 | Information on the 63 selected covariate layers used to model root mass fractions[194–205].



Extended Data Fig. 2 | Random 10-fold cross-validation (RCV) of the spatial root-mass fraction models and spatial autocorrelation of model residuals. **a–c**, Heat plots showing the relationships between predicted and observed RMFs in forests (**a**), shrublands (**b**), and grasslands (**c**) based on RCV. Solid lines indicate fitted relationships based on ordinary least squares regression [coefficient of determination values relative to the 1:1 line (equation 2) shown in the bottom right corner], dashed diagonal lines indicate a 1:1 relationship between observed and predicted points. **d–f**, The standard errors of the observed (black) and predicted (grey) mean values of root mass fractions decrease with increasing sample size. The operation was repeated with 1,000 random seeds for the observed and predicted mean values, and the calculated standard errors of the mean are shown. Note, 'sample size' in D–F refers to the number of pixels, and thus denotes square kilometres. **g–i**, Semivariograms illustrating spatial autocorrelation of model residuals in forests (**g**), shrublands (**h**) and grasslands (**i**). Semivariances of residuals were computed based on random 10-fold cross validation (blue) and spatial leave-one-out cross-validation (LOO-CV) with buffer radii of 150km (dark green), 250km (green) and 500km (light green). Dashed vertical lines indicate the buffer radii of the final validation model reported throughout the text.

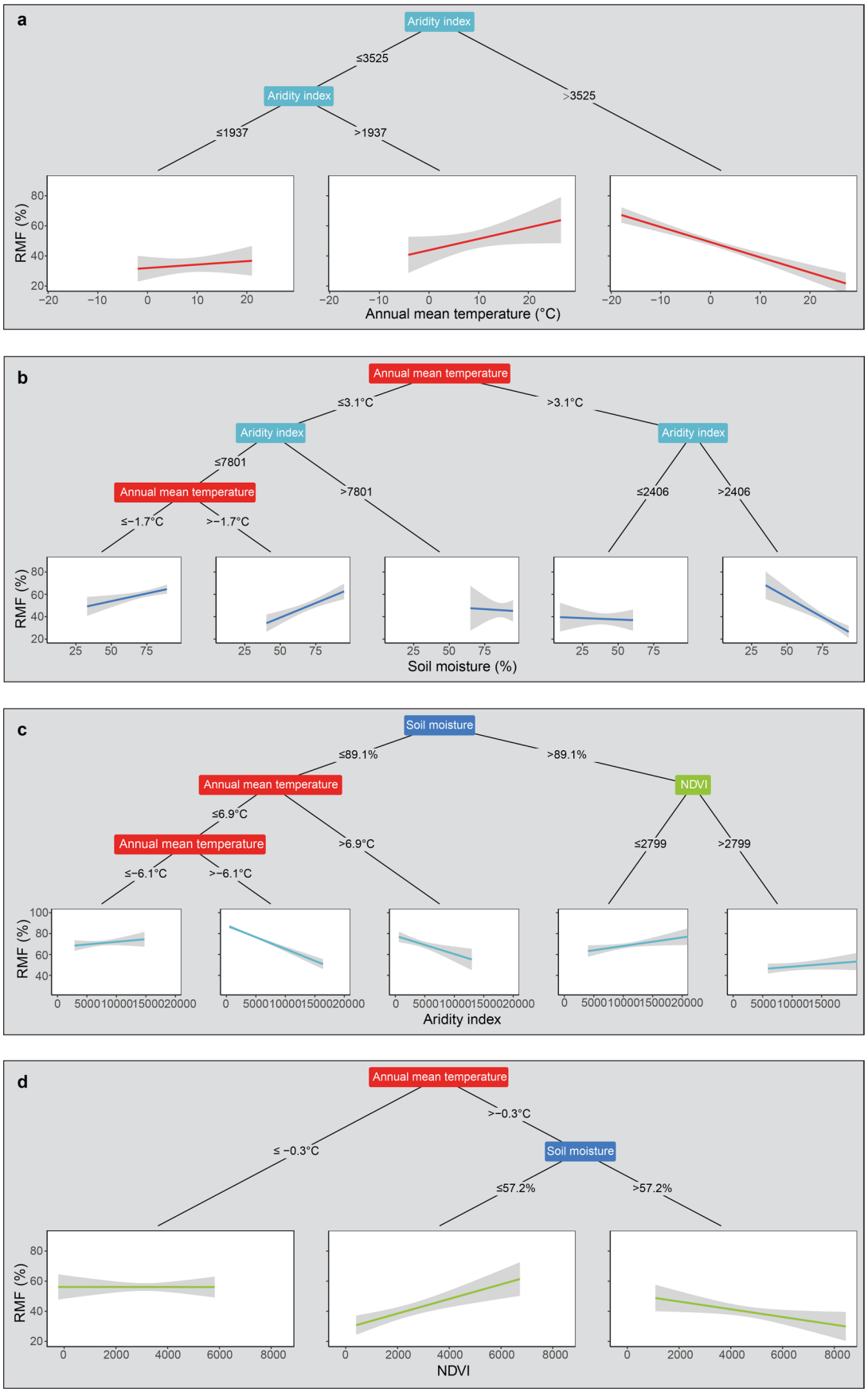


Extended Data Fig. 3 | Partial regression coefficients for the effects of 8 environmental covariates from linear multiple regression models. To reduce the influence of spatial autocorrelation, a bootstrapping procedure was applied for the forest data (see Methods). Red dots indicate positive effects on RMFs, blue dots indicate negative effects. Error bars reflect two standard errors either side of the mean partial regression coefficient.



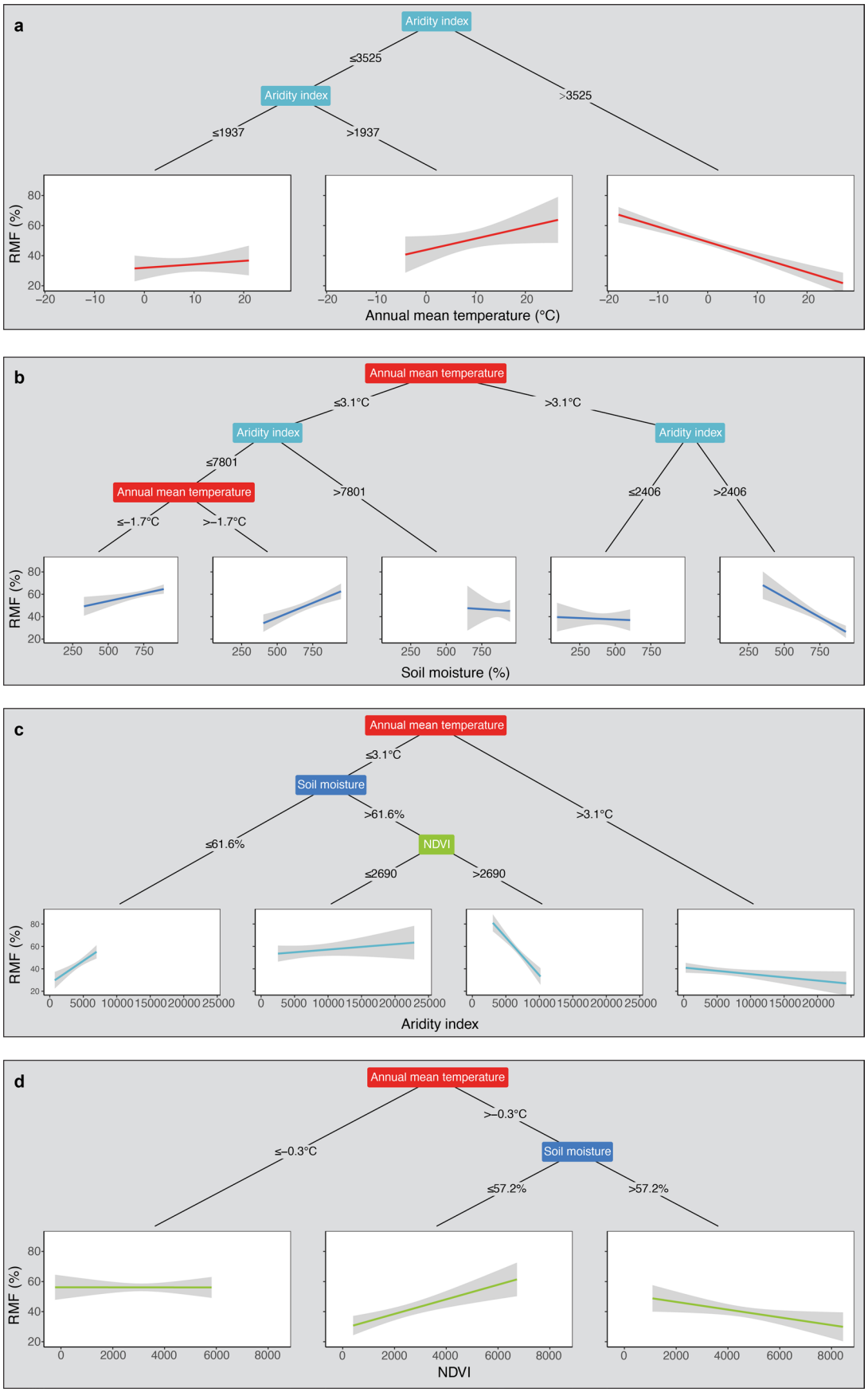
Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Recursive partitioning trees for the univariate effects of annual mean temperature (a), soil moisture (b), NDVI (c), and sand content (d) on RMFs in forests. These four variables were chosen on basis of the random forest variable importance metric (Fig. 3a) and, for each model, the remaining three variables were evaluated as potential split points. The number of independent observations contained in each terminal node was constrained to ≥10% of the total data (500 observations). Regression plots show slopes and 95% confidence intervals.



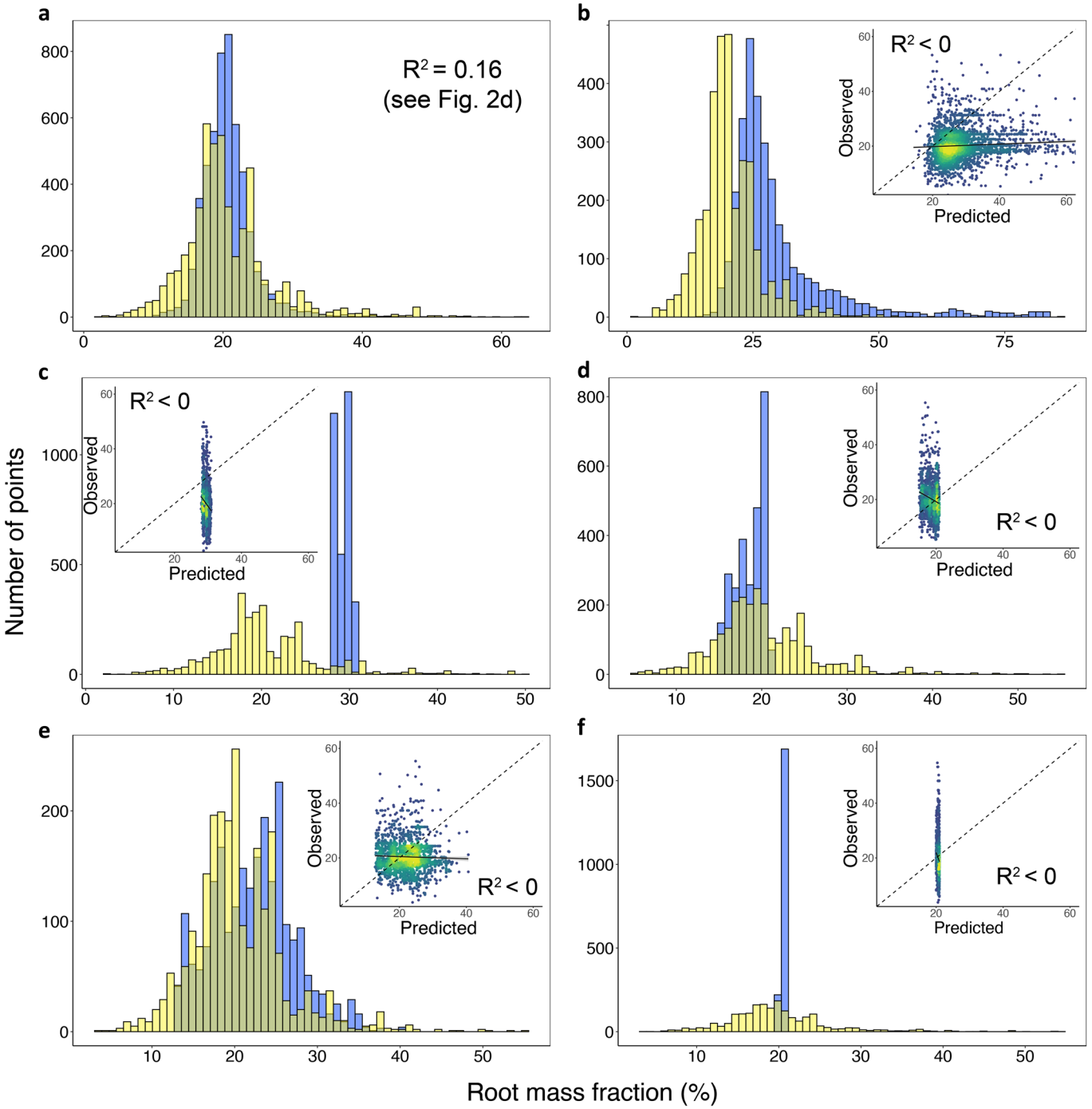
Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Recursive partitioning trees for the univariate effects of annual mean temperature (a), soil moisture (b), aridity index (c), and NDVI (d) on RMFs in shrublands. These four variables were chosen on basis of the random forest variable importance metric (Fig. 3b) and, for each model, the remaining three variables were evaluated as potential split points. The number of independent observations contained in each terminal node was constrained to ≥10% of the total data (30 observations). Regression plots show slopes and 95% confidence intervals.

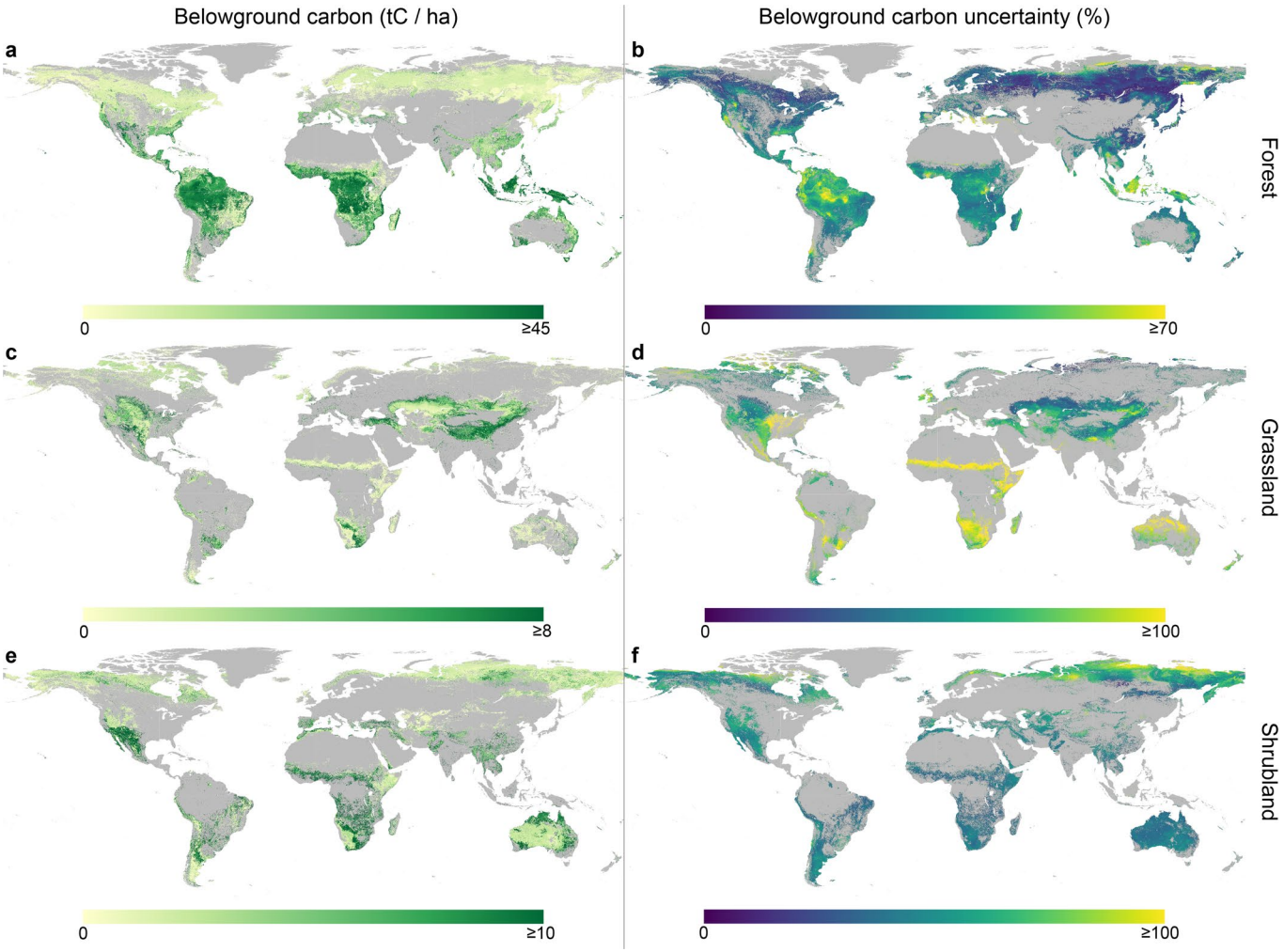


Extended Data Fig. 6 | See next page for caption.

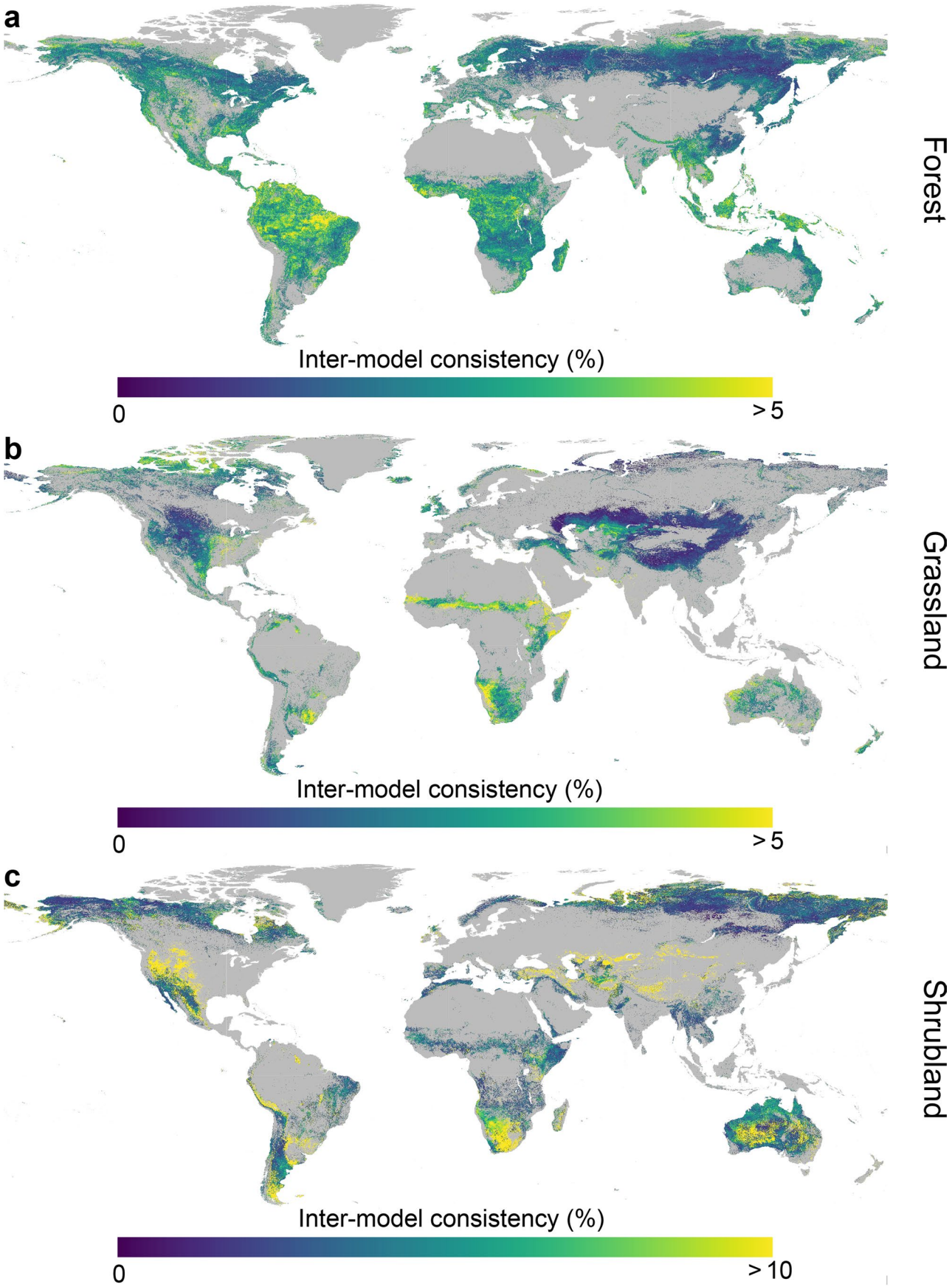
Extended Data Fig. 6 | Recursive partitioning trees for the univariate effects of annual mean temperature (a), soil moisture (b), aridity index (c), and NDVI (d) on RMFs in grasslands. These four variables were chosen on basis of the random forest variable importance metric (Fig. 3c) and, for each model, the remaining three variables were evaluated as potential split points. The number of independent observations contained in each terminal node was constrained to ≥10% of the total data (120 observations). Regression plots show slopes and 95% confidence intervals.



Extended Data Fig. 7 | Comparison of observed forest RMFs with predicted RMFs from dynamic global vegetation models and a current-generation biomass map. The blue bars represent histograms of predicted RMF values based on our LOO-CV procedure (**a**), current-generation biomass estimates[6] (**b**), and the vegetation models CABLE-POP (**c**), CLASS-CTEM (**d**), ISAM (**e**) and ORCHIDEE (**f**). Yellow bars represent observed values. Insets show scatter plots of predicted versus observed RMFs with solid lines indicating fitted relationships, dashed diagonal lines indicating a 1:1 relationship between observed and predicted points. For the vegetation models, forest was defined as pixels with a tree cover fraction higher than 50%.

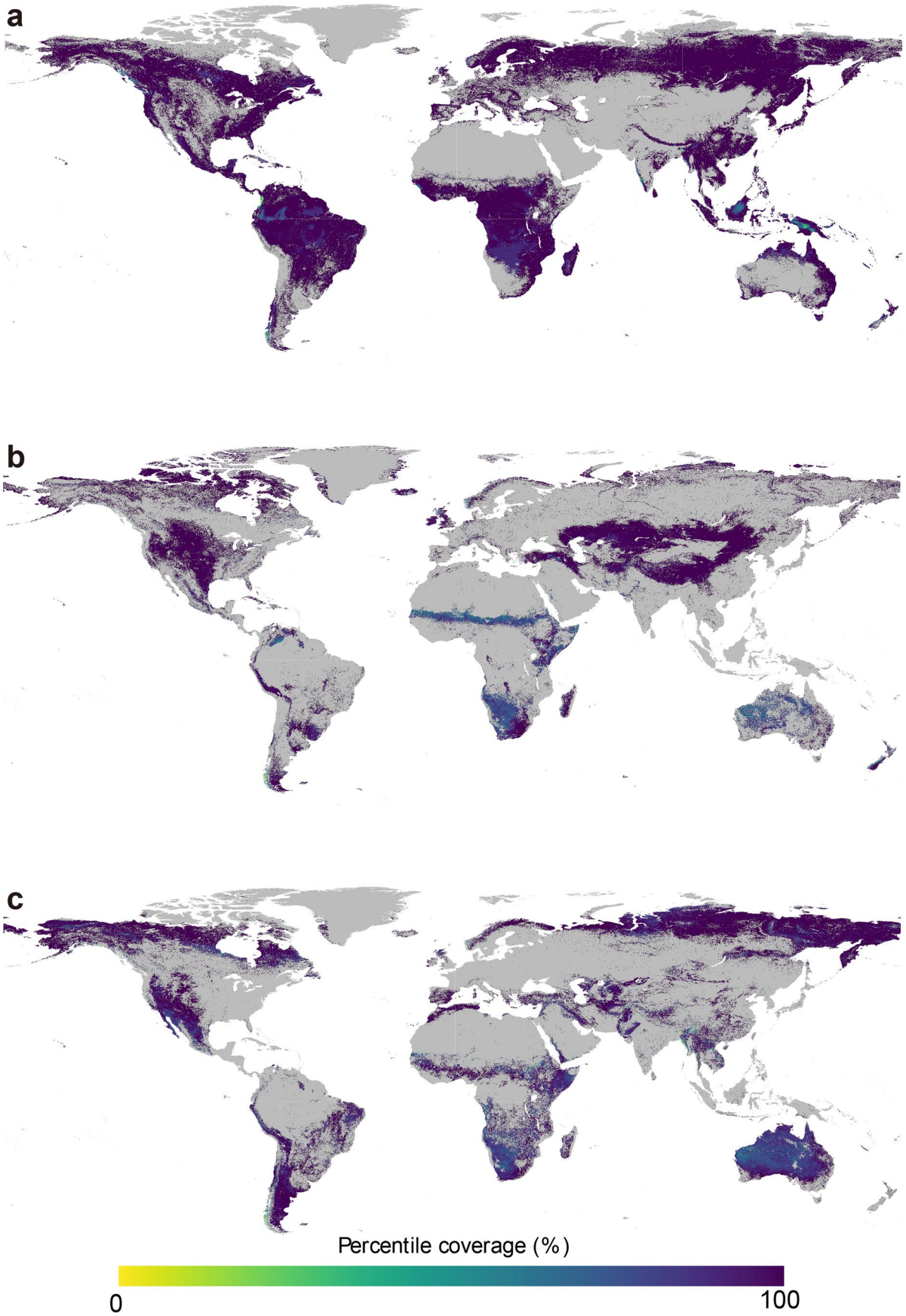


Extended Data Fig. 8 | The global distribution of belowground plant biomass and associated uncertainties in forests (a, b), grasslands (c, d), and shrublands (e, f). a, c, e, Belowground plant biomass (in tons carbon per hectare). **b, d, f** Associated uncertainties in belowground carbon, calculated as the predicted biomass range (based on 2.5% and 97.5% RMF quantiles derived from the bootstrapped RMF models) divided by the mean predicted biomass in each pixel. Maps are projected at 30 arc-seconds (~1 km^2) resolution.



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | Root mass fraction inter-model consistency in forests (a), grasslands (b) and shrublands (c). Inter-model consistency was calculated as the coefficient of variation (standard deviation divided by mean, in %) of the predictions of the 10 best models. Maps are projected at 30 arc-seconds (~1 km^2) resolution.



Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | The extent of interpolation and extrapolation across all terrestrial pixels in which the respective vegetation type, forest (a), grassland (b) and shrubland (c) occurs. Values represent the percentage of interpolation based on principal component analysis, that is, the percentage of bands that fall into the convex hull space.

nature research

Corresponding author(s): Constantin M. Zohner

Last updated by author(s): Apr 15, 2021

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Data was collected manually by searching the literature. Citations for the root-shoot ratio data are provided in the methods.
Data analysis	The analyses were run in R [R Core Team (2021) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-proje ct.org. Accessed 17 Jan 2021] and Google Earth Engine (https://earthengine.google.com). The code used for this study is available at https://github.com/haozhima95/Global_mapping_root_shoot_ratio.git.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The root-shoot ratio data underlying this study are available at https://github.com/haozhima95/Global_mapping_root_shoot_ratio/tree/master/RSR_data. Citations for the root-shoot ratio data are provided in the methods.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	We use a global dataset of plant root mass fractions to generate a spatially explicit understanding of the proportional plant biomass that exists in roots and upscale these predictions to the globe to quantify total belowground plant biomass. Our database consists of 17,814 plot-level root mass fraction measurements, aggregated to 6,803 spatially distinct samples from forest (N = 5170, from 52 studies), grassland (N = 1293, from 22 studies), and shrubland (N = 340, from 82 studies) biomes. Using 63 global layers of climate, topography, soil, and vegetation features, we trained random forest models to characterize the environmental factors that drive geographic variations in RMFs within each vegetation type and interpolate these relationships across the globe. We additionally use recursive partitioning and regression methods accounting for spatial autocorrelation and compare our empirical root mass fraction data against vegetation model simulations.
Research sample	Plot-level root-shoot biomass ratios in trees, grasses and shrubs.
Sampling strategy	No statistical methods were used to predetermine sample size.
Data collection	Data was collected through literature search. Citations for the root-shoot ratio data are provided in the methods.
Timing and spatial scale	Root-shoot ratio data from all continents except Antarctica, collected after the year 1960 were used in this study.
Data exclusions	Data was only used if all of the following four criteria were met: 1) RSR values were available at the plot level (and not only for selected individuals); 2) data came from field measurements (allometry-derived RSRs were excluded). If the study consisted of multiple treatments, only data from the control treatment was selected; 3) coordinate information was given; and 4) measurements were taken after 1960. For datasets that included information on more than one vegetation type, the abovementioned criteria were evaluated separately for the subset of trees, grasses and shrubs within the dataset. Furthermore, to minimize the influence of ontogenetic drift on RSRs in forests (higher root-to-shoot biomass ratios in seedlings than in adults45), we 1) excluded studies that only focused on seedlings of trees and 2) excluded seedlings or tree individuals with a DBH smaller than 10cm within studies.
Reproducibility	n/a
Randomization	n/a
Blinding	n/a

Did the study involve field work? ☐ Yes ☒ No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging