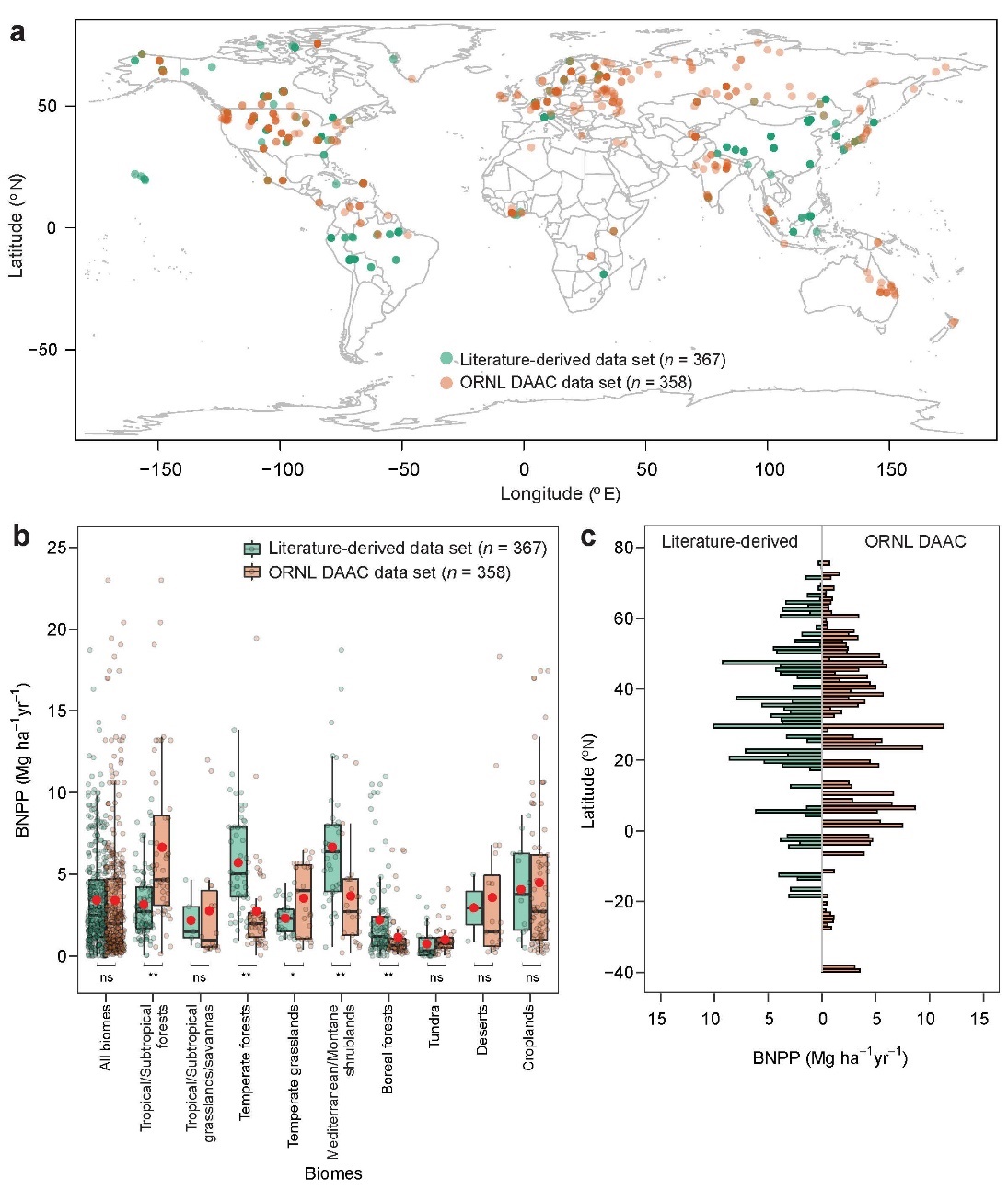
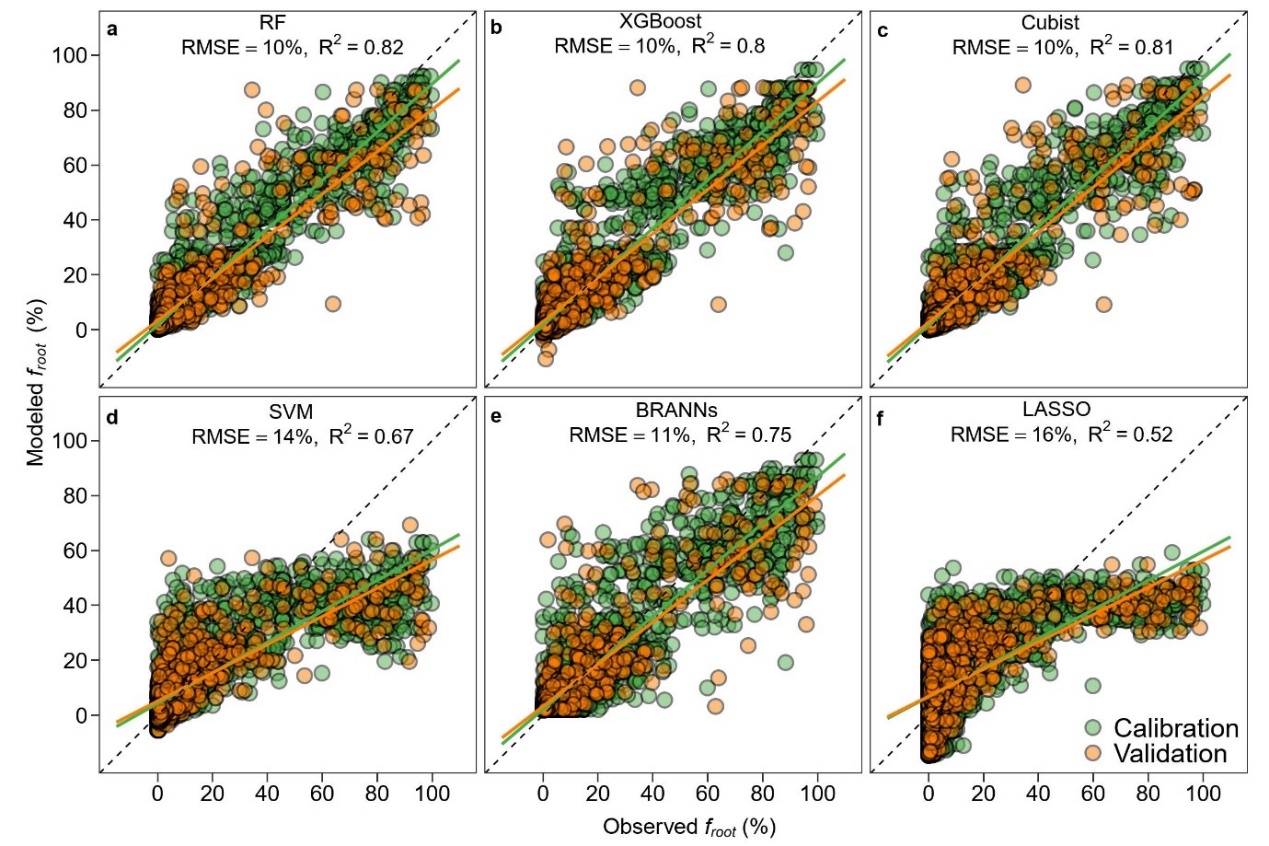
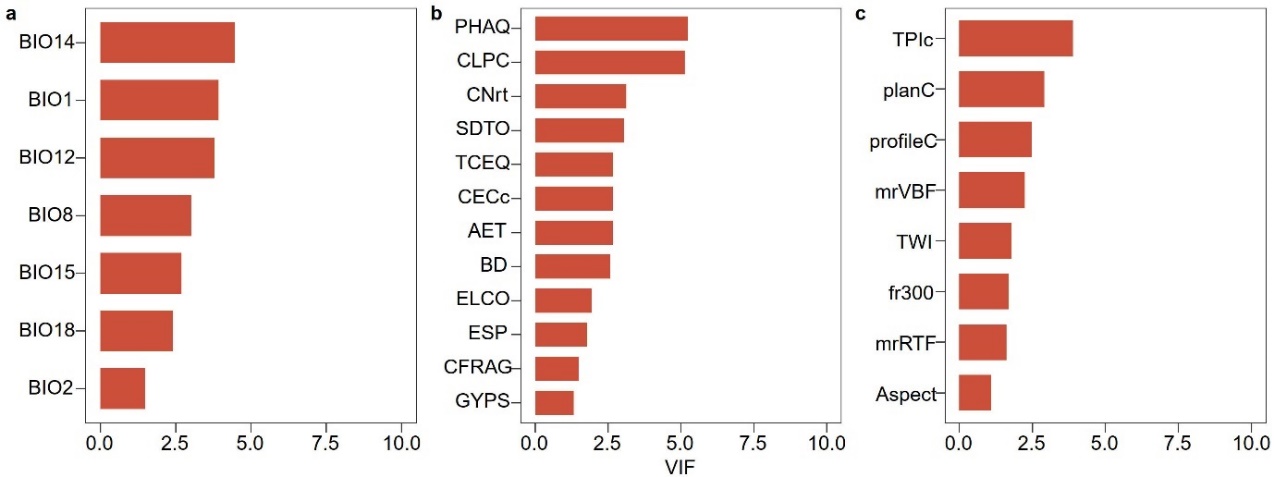
**Supplementary information**

****

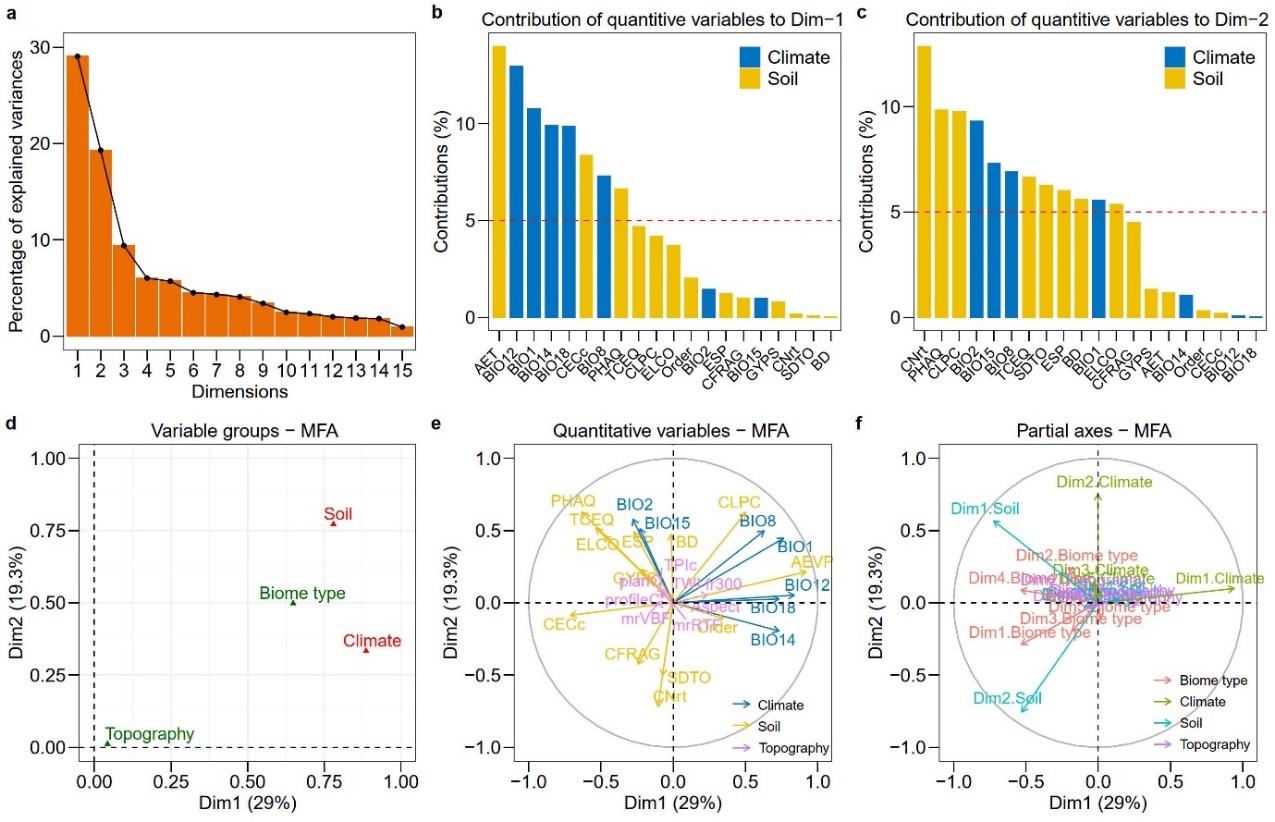
**Fig. S1. Comparison of belowground net primary production (BNPP) from two data sources. a**, spatial location of literature-derived and ORNL DAAC data; **b**, the distribution of BNPP across biomes; **c**, latitudinal distribution of average BNPP aggregated in each 1 degree of latitude. Boxplots show the median and interquartile range with whiskers extending to 1.5 times of the interquartile range, and red dots show averages. Literature-derived data set is collected by conducting a comprehensive literature review; ORNL DAAC data set is an online data source publically available at <https://daac.ornl.gov/cgi-bin/dataset_lister.pl?p=13>. *ns* below the boxes in **b** indicates insignificant difference between the two data sources; while \* and \*\* indicate significant difference at *P < 0.05* and *P < 0.01*, respectively.



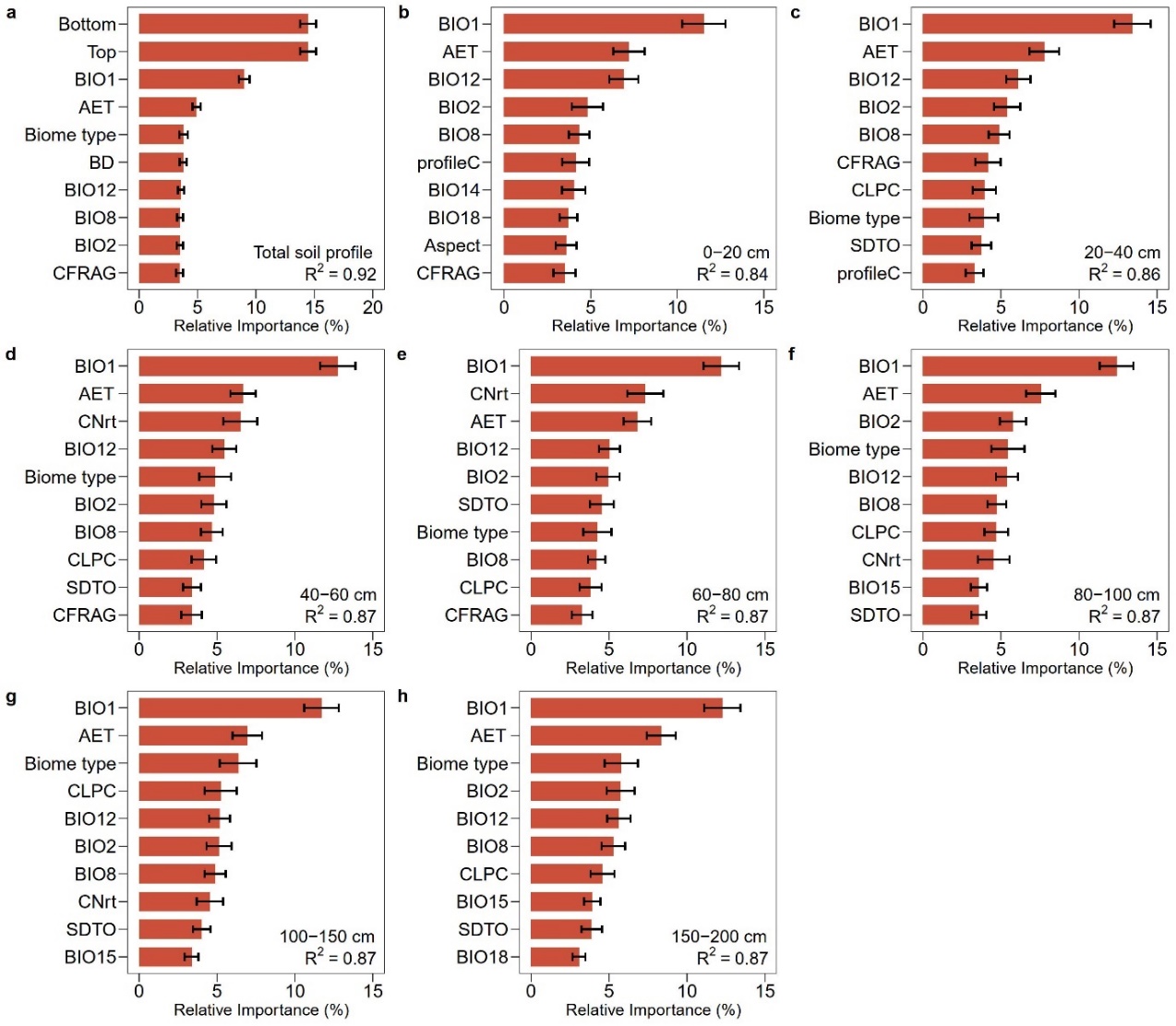
**Fig. S2. Performance of machine learning models to predict root biomass distribution. a**, random forest (RF); **b**, extreme gradient boosting (XGBoost); **c**, Cubist; **d**, support vector machines (SVM); **e**, bayesian regularized artificial neural networks (BRANNs); **f**, lasso regression (LASSO). For each model, observations from randomly selected 80% of the soil profiles (i.e., 559 soil profiles from the Rootobs data set) are used for model calibration, with the data from the remaining 20% soil profiles for validation. RMSE and R2 represent the rooted mean squared error and determination coefficient of validation, respectively.



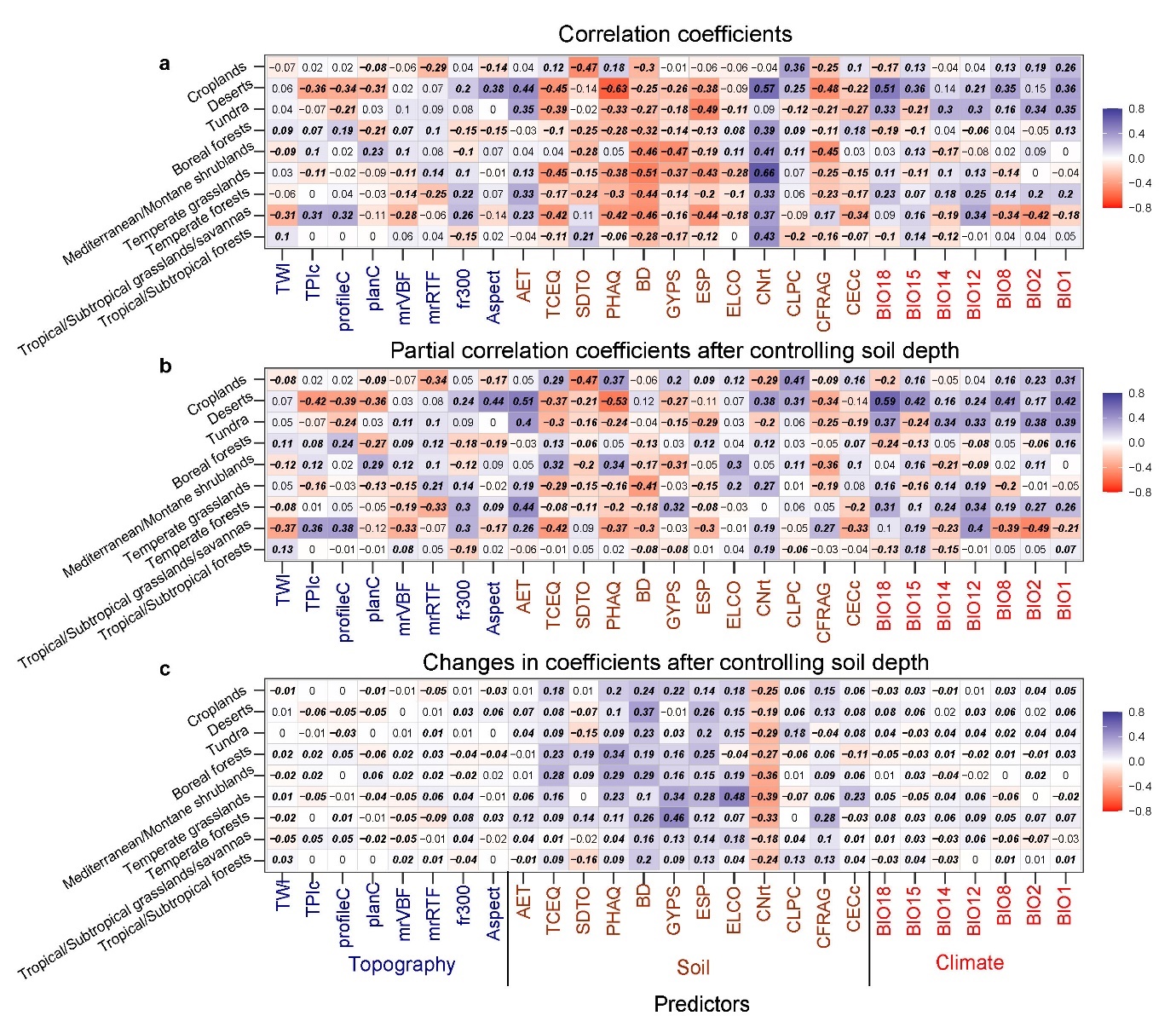
**Fig. S3. Selected variables by variance inflation factor (VIF) for predicting belowground net primary production (BNPP).** **a**,selected climatic variables; **b**, selected edaphic variables; **c**, selected topographic variables. Detailed descriptions for the variables are presented in Table S1. Variables with a VIF value larger than 10 indicate that the variables have high intercorrelations with the selected variables and are excluded from the modelling.



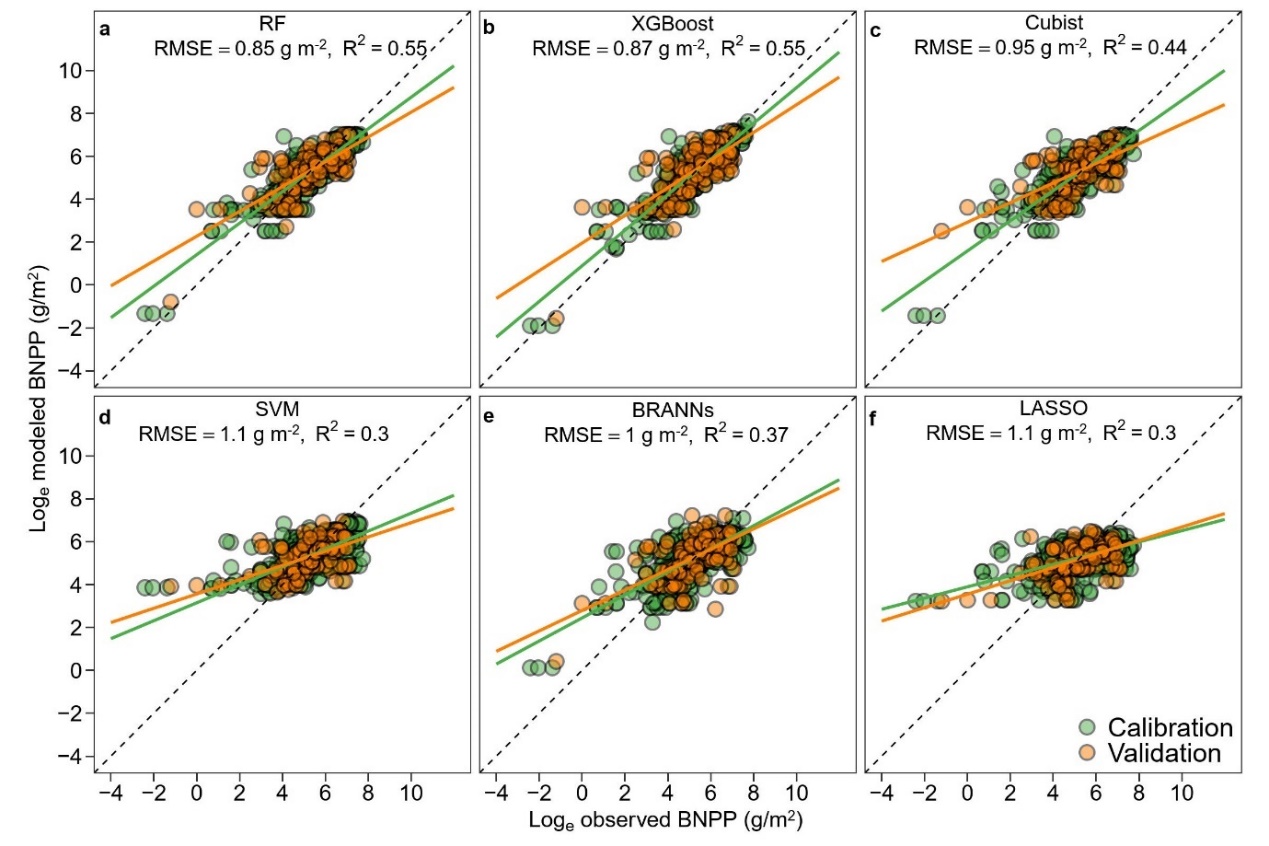
**Fig. S4. Results of multiple factor analysis (MFA) of climatic, edaphic, and topographic variables to reduce variable dimension.** Detailed descriptions for the variablesare presented in Table S1. **a** The percentage of explained variances by different dimensions; **b** and **c**, the contribution of variables to dimensions 1 and 2 (i.e., Dim1 and Dim2); **d**, the contribution of each group of variables (i.e., climatic, edaphic and topographic variables, biome type) to Dim1 and Dim2; **e**, correlation between individual variables and dimensions, demonstrating the relative importance of individual climatic, edaphic and topographic variables; **f**, partial axes analysis between the principal axes (Dim 1 and 2) and the ones obtained from analyzing each group of variables using principal component analysis.



**Fig. S5.** **The relative importance of the top 10 most important variables for predicting depth distribution of BNPP.** **a**, the whole soil profile; **b-h**, the 0-20, 20-40, 40-60, 60-80, 80-100, 100-150 and 150-200 cm layers, respectively. The error bars show one standard deviation.

****

**Fig. S6. The linear relationship between the depth distribution of BNPP (LogeBNPP) and the selected most important variables.** The proportional depth distribution of BNPP is estimated by Rootobs. Detailed descriptions for the variablesare presented in Table S1.



**Fig. S7. Performances of fitted machine learning models to predict belowground net primary production (BNPP). a**, random forest (RF); **b**, extreme gradient boosting (XGBoost); **c**, Cubist; **d**, support vector machines (SVM); **e**, bayesian regularized artificial neural networks (BRANNs); **f**, lasso regression (LASSO). For each model, observations from randomly selected 80% of the soil profiles (i.e., 725 soil profiles from the NPPobs data set) are used for model calibration, with the data from the remaining 20% soil profiles for validation. RMSE and R2 represent the rooted mean squared error and determination coefficient of validation. Data are log-transformed.

**Table S1. Datasets used in this study.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Datasets | Product name | Variable | Spatial resolution | Source |
| NPP dataset | NPPobs | NPP, ANPP, BNPP | field data | 1. Peer-reviewed papers (Supplementary references)  2. ORNL DAAC NPP : <https://daac.ornl.gov/cgi-bin/dataset_lister.pl?p=13> |
| Root biomass dataset (Rootobs) | Rootobs | root biomass and its depth distribution | field data | Schenk and Jackson (50) |
| Climate database | WorldClim 2.1 | Edaphic variables (BIO1: BIO19) | 0.5 arc-min  (~1km) | <http://www.worldclim.com/version2> |
| Soil properties database | ISRIC-WISE: WISE30sec, v1.0 | Edaphic variables | 0.5 arc-min  (~1km) | <https://www.isric.org/explore/wise-databases> |
| TerraClimate | Edaphic variables | 2.5 arc-min  (~4km) | <http://www.climatologylab.org/terraclimate.html> |
| Topography database | SRTM-DEM v4.1 | Topographical  variables | 5 arc-sec  (~90m) | <http://srtm.csi.cgiar.org> |
| Biome types | MODIS Land Cover -Product MCD12Q2 &  Terrestrial Ecoregions of the World | Biomes | 0.5 arc-min  (~1km) | <https://modis.gsfc.nasa.gov/data/dataprod/mod12.php>  <https://www.worldwildlife.org/biome-categories/terrestrial-ecoregions> |
| Global soil order | soil order | soil order | 2 arc-min  (~4km) | <https://go.nature.com/3hgdsgb> |

**Table S2. Environmental covariates used in this study.**

|  |  |  |  |
| --- | --- | --- | --- |
| Covariates | Code | Description | Unit |
| Edaphic variables | CFRAG | Coarse fragments (>2mm) | % |
| BD | Bulk density | g cm-3 |
| ORGC | Organic carbon | g kg-1 |
| SDTO | Sand content | % |
| CLPC | Clay content | % |
| STPC | Silt content | % |
| TAWC | Available water capacity | cm m-1 |
| TOTN | Total nitrogen | g kg-1 |
| CNrt | C:N ratio | - |
| PHAQ | pH measured in H2O | - |
| CECS | Cation exchange capacity | cmol kg-1 |
| ECEC | Effective cation exchange capacity | cmol kg-1 |
| CECc | Cation exchange capacity of clay size fraction | cmol kg-1 |
| TEB | Total exchangeable bases | cmol kg-1 |
| BSAT | Base saturation | % |
| ESP | Exchangeable sodium percentage | % |
| ALSA | Aluminum saturation | % |
| TCEQ | Total carbonate equivalent | % |
| GYPS | Gypsum content | % |
| ELCO | Electrical conductivity | dS m-1 |
| AE | Actual daily soil evaporation | mm |
| Climatic variables | BIO1 | Mean annual temperature | °C |
| BIO2 | Mean diurnal range | °C |
| BIO3 | Isothermality (BIO2/ BIO7×100) | % |
| BIO4 | Temperature seasonality (standard deviation×100) | °C |
| BIO5 | Max temperature of warmest month | °C |
| BIO6 | Min temperature of coldest month | °C |
| BIO7 | Temperature annual range (BIO5–BIO6) | °C |
| BIO8 | Mean temperature of wettest quarter | °C |
| BIO9 | Mean temperature of direst quarter | °C |
| BIO10 | Mean temperature of warmest quarter | °C |
| BIO11 | Mean temperature of coldest quarter | °C |
| BIO12 | Annual precipitation | mm |
| BIO13 | Precipitation of wettest month | mm |
| BIO14 | Precipitation of driest month | mm |
| BIO15 | Precipitation seasonality (coefficient of variation) | % |
| BIO16 | Precipitation of wettest quarter | mm |
| BIO17 | Precipitation of driest quarter | mm |
| BIO18 | Precipitation of warmest quarter | mm |
| BIO19 | Precipitation of coldest quarter | mm |
| Biome  Soil order | - | Biome type | - |
| - | US Department of Agriculture soil taxonomy | - |
| Topographical  variables | Aspect | The direction in which a land surface slope faces expressed in degrees from north | - |
| fm | 300 m focal median of percent slope | - |
| fr1000 | 1000 m elevation range | m |
| fr300 | 300 m elevation range | m |
| mrRTF | Multi-resolution ridge top flatness | - |
| mrVBF | Multi-resolution valley bottom flatness index | - |
| PI | The Prescott Index | - |
| planC | Plan curvature | ° |
| profileC | Profile curvature | ° |
| SD | Mean monthly solar radiation | MJ m−2 day−1 |
| SP | Percent slope | - |
| TPI | Topographic position index | - |
| TWI | Topographic wetness index calculated as ln(specific catchment area/slope) and estimates the relative wetness within a catchment | - |

**Table S3. Coefficients and summary statistics for the model using soil depth to predict the depth allocation of absolute belowground net primary productivity (BNPP). The model is written as: BNPP=a×exp(-Depth/b).**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Biome type | a (mean ± s.e.) | b (mean ±s.e.) | *P* | R2 |
| Tropical/Subtropical forests | 5.5±0.4 | 17.5±1.2 | <0.01 | 0.43 |
| Tropical/Subtropical grasslands/savannas | 1.9±0.4 | 26.4±5.7 | <0.01 | 0.27 |
| Temperate forests | 6.2±0.5 | 15.7±1.1 | <0.01 | 0.52 |
| Temperate grasslands | 3.1±0.3 | 19.2±1.6 | <0.01 | 0.56 |
| Mediterranean/Montane shrublands | 5.8±0.5 | 21.2±1.8 | <0.01 | 0.48 |
| Boreal forests | 6.1±1.2 | 11.5±1.5 | <0.01 | 0.23 |
| Tundra | 3.2±0.9 | 9.27±1.5 | <0.01 | 0.41 |
| Deserts | 2.3±0.4 | 30.8±6.8 | <0.01 | 0.25 |
| Croplands | 5.5±0.6 | 17.5±1.8 | <0.01 | 0.36 |
| All biomes | 4.9±0.2 | 16.4±0.6 | <0.01 | 0.34 |

**Table S4.** The overall relative importance (%, mean ± one standard deviation) of three groups of environmental variables (i.e., climate, soil and topograph) for predicting BNPP in the whole soil profile as well as in different soil layer depths using random forest models.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Soil layer | Depth | Climate | Soil | Topography |
| Total soil profile | 29.0±0.9 | 25.7±0.8 | 27.7±0.7 | 13.0±0.4 |
| 0-20 cm | - | 38.6±1.9 | 33.2±2.0 | 24.5±2.1 |
| 20-40 cm | - | 38.5±1.6 | 35.8±1.9 | 20.5±1.7 |
| 40-60 cm | - | 36.3±1.7 | 38.4±2.0 | 19.3±1.5 |
| 60-80 cm | - | 35.3±1.7 | 40.3±2.0 | 19.0±1.6 |
| 80-100 cm | - | 37.7±1.7 | 37.3±1.8 | 18.4±1.5 |
| 100-150 cm | - | 36.3±1.7 | 38.0±1.8 | 18.3±1.4 |
| 150-200 cm | - | 39.0±1.8 | 36.1±1.6 | 18.1±1.4 |