Supplementary Materials

Mapping global distributions, environmental controls, and uncertainties of apparent top- and subsoil organic carbon turnover times

Lei Zhang *et al.*

Contents of this file:

Text S1 to S6

Figures S1 to S27

Tables S1 and S6

References

Supplementary Text

Text S1. Approach for filling the missing values in bulk density measurements

Although we filtered soil sample data from multiple databases for ensuring all contain the soil organic content observation values at multiple soil layers. The measurements of bulk density (BD) were not always available for all soil layers of the profiles considered in our study. Following the approach provided in FAO - CMCC (2017), some pedotransfer functions were adopted to estimate BD if they are missing. In particular, for layers with a SOC content lower than 12%, according to Hiederer & Köchy (2011), an exponential function was used to calculate BD for global data either in the topsoil or subsoil. For soil layers with a SOC content higher than 12%, it was used the log-transformed function. The abovementioned pedotransfer functions used depending on the SOC content and the considered soil layer to estimate BD are summarized in the following equation.

$$\text{BD} = \begin{cases} e^{(a \cdot SOC + b)} & \textit{if SOC content} \le 12\% \\ -0.285 \cdot \ln(SOC) + 1.456 & \textit{if SOC content} > 12\% \textit{ and soil depth} \le 30cm \\ -0.291 \cdot \ln(SOC) + 1.389 & \textit{if SOC content} > 12\% \textit{ and soil depth} > 30cm \end{cases} \quad \text{(S1)}$$

where $a = -0.034$, $b = 0.100$.

Text S2. Non-extracted aboveground NPP in croplands

When calculating the aboveground NPP (ANPP) in cropland areas, a decent fraction of ANPP is removed as the harvest products, and it should be additionally considered. The way for estimating the non-extracted ANPP in croplands is as follows. We first confirmed that, in the aboveground and belowground biomass carbon maps (Spawn et al., 2020), the crop's yield and the harvest index (crop-specific) were used to estimate the maximum standing biomass of crops prior to harvest. Following the work of Spawn et al., we adopted the crop-specific harvest index provided by Wolf et al. (2015) and Spawn et al. (2020) to estimate the unharvested residue in croplands. First, we used the EARTHSTAT dataset (Monfreda et al., 2008) to extract the spatial coverage of different croplands with different crop types. Then, we considered the aboveground residue retention rate in areas with different developing levels. According to the published literature, the residue retention rates can be assumed as 30% in developing regions such as Asia and Africa (e.g., Erenstein, 2011; Jiang et al., 2012; Baudron et al., 2014), and 60% in other regions (e.g., Scarlat et al., 2010; Lokupitiya et al., 2012; Baudron et al., 2015). This way was also adopted in a previous study (Wang et al., 2016). In addition, the manure input is an extra carbon input in croplands. We used the global crop-specific nitrogen (N) fertilization dataset

from Adalibieke et al. (2023), to determine the consumption of N fertilizer input from manure application. The C:N ratio of farmyard manure provided by FAO (Roman et al., 2015) was adopted for converting the amount of N to the carbon.

For the grassland and forest areas, we acknowledge that considering the human-induced decrease (e.g. grazing, deforestation) or increase (e.g. dung excretions during grazing, afforestation) of carbon input co-exist, and the available accurate global datasets that quantify these impacts are still not well developed, so these impacts were not included in our study for the limited observations and datasets.

Text S3. Fraction of aboveground NPP transfers into topsoil

When calculating carbon input flux for topsoil, apart from belowground NPP (BNPP), the carbon input also contains a portion of carbon from surface litterfall, which should be additionally considered. Although there are no spatially-explicit values of this proportion provided directly, we found currently available datasets and literatures that help us generate estimations as follows.

It should be noted here that the organic litter on the surface was not incorporated in the soil (White, 2006), therefore, we need to know the fraction of ANPP converted into litterfall. For having a biome-specific estimation on it, we collected datasets provided by Holland et al. (2015) and Jia et al. (2016), to get a total number of 813 geo-referenced observations with the values of litterfall (kg m^{-2} yr^{-1}). Combining these observations with the ANPP values we processed before and the global biome map provided by (Dinerstein et al., 2017), we compiled a biome-specific estimated result of the fraction of litterfall in ANPP, which is shown in Table S5 as below. Considering the difference between tree species (needle-leaved, broadleaved, evergreen and deciduous), we further used MODIS land cover data product to divide the litterfall sample data by a more detailed classification. Table S6 shows the differences of this value in forests with different leaf types.

Then, we need to know the fraction of litterfall allocated to topsoil. Such data are very limited so that we cannot be supported to get a spatially detailed estimation. Refer to Community Land Model (CLM) (Oleson et al., 2013), also some evidences from Millennial model (Abramoff et al., 2022) and MEMS 2.0 model (Zhang et al., 2021), we adopted a value of 0.5 as the respiration fraction (*rf*) for litter pool as suggested from CLM that based on a model structure used in Century model (Oleson et al., 2013). Therefore, it can be considered that approximately half of litterfall released as CO_2 back to the atmosphere, thereby the left half

of it allocated into the topsoil. Finally, the fraction of ANPP that transferred as carbon input into the topsoil (denoted as fr_a), can be determined as the biome-specific value of the fraction of litterfall in ANPP multiplied by $(1 - rf)$. For subsoil, fr_a was set to be zero.

Text S4. Calculation of variable importance that contributes τ pattern

The relative importance (*RI*) of each covariate (the contribution of each predictor) was calculated by using permutation-based feature importance. The permutation feature importance is defined to be the decrease in the model accuracy when a single feature (covariate) value is randomly shuffled. The feature importance can be computed on a held-out validation set. This approach can avoid the problem that the feature might be given high importance on the training set but not on the held-out validation set caused by the model overfitting. The *RI* of a certain covariate can be calculated as follows:

$$RI_i = s - \frac{1}{K}\sum_{k=1}^{K} s_{k,i} \tag{S2}$$

Where RI_i is the relative importance of the i-th covariate, s is the score of model performance fitted on the original data without permutation, $s_{k,i}$ is the score of model fitted on the data with shuffled values of the i-th covariate in the k-th reputation, and K is the number of times to permute a covariate. The score of the model was set as the metric of R^2 on the withheld validation set.

The *RI* of each environmental covariate to the global variations of τ in top- and subsoil layers was quantified. To further reflect the difference in influences of the four variable categories on τ, we integrated the importance of variables in each category. The variable with the highest *RI* in each category was used to represent the importance of that category. Compared to taking the sum or mean of *RI* for each category, this way avoids to some extent over or underestimate of importance caused by the difference of variable importance within each category and the different number of variables included among categories. Then, the importance values of all categories were scaled to a sum of 1, and the percentage number represents the importance of each category. The same procedure of calculating variable importance was also conducted for data in different biomes (Fig. S23 and S24).

Text S5. Detecting interactive effects among climate and edaphic factors on τ

Given the potentially non-significant effects of a variable (e.g. the global impact of precipitation) on τ in a global-level linear model, it does not necessarily imply that the respective variable has no influence on τ. Instead, this lack of significance may arise from its divergent effects when interacting with other factors. This led to an interest in exploring the reasons behind such nonlinear and uncoherent driving mechanisms. Considering the extrinsic climate effects on soil carbon turnover can be interacted by regionally intrinsic soil characteristics (Doetterl et al., 2015), we used the following approach to exemplify the interactive effects of MAT and MAP on top- and subsoil τ in response to the changes of other factors (namely interactive factors).

The variables of MAP/MAT, CLAY+SILT, C:N, CEC, and soil pH were selected (according to the analytical results of variable importance) as the important interactive factors for detect nonlinear MAT/MAP effect on τ. Each selected interactive factor was divided into ten intervals by using Jenks natural breaks classification method (Jenks, 1967), and all sample data were divided into ten subsets according the intervals that their interactive factors fall into. Then, the ten partial correlation coefficients between τ and two climatic variables within all intervals were connected to generate a line chart, which reveals how effects of temperature and water availability on τ strongly interact with edaphic factors (Fig. 3).

Text S6. Local-level analysis of dominating factors

In addition to the overall importance of covariates analyzed at the global level, the local-level importance for four categories of variables was also assessed. We quantified *RI* of covariates based on the RF model trained by sample data at local scales stratified by the global terrestrial ecoregions (Dinerstein et al., 2017). The same procedure of calculating importance of variables and categories mentioned in above was conducted in each ecoregion. This approach enables the globally geographical pattern of the regional importance of covariates yield. The global maps of the dominant factors (the category of covariate) influencing τ at top- and subsoil layers were generated by interpolating results in all windows. Additionally, since the sample sizes in different windows are not same (Fig. S26), the relative density of samples in each local window was considered. We determined the relative density to be the ratio of sample points fall within each window to a number of 200 which is considered as a sufficient size. All ratio that larger than 1 were controlled to be 1, so that the relative density value is between 0 and 1. The colors represent the local importance of the four categories of covariates in the maps was adjusted to display lighter proportionally to the lower the sample density.

Supplementary Figures

Fig. S1
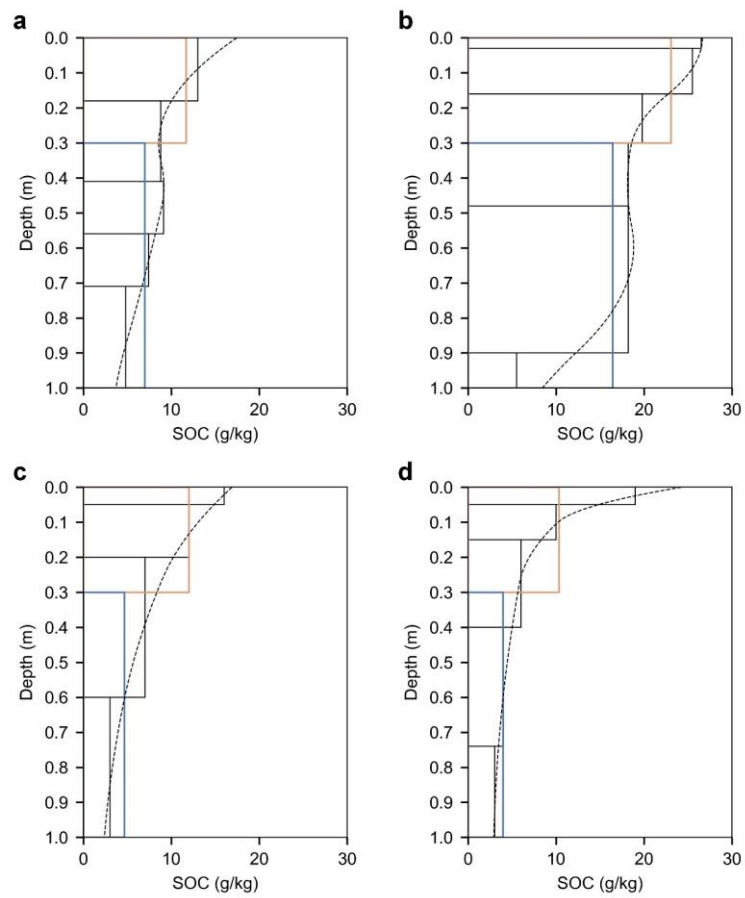
Figure S1. Examples of the fitted variation of soil organic carbon content (SOC, g kg^{-1}) with depth (m). The figure shows observed SOC values (black bars) and fitted splines (dashed lines) by using the equal-area smoothing spline algorithm referred to Bishop et al. (1999) and Malone et al. (2009) at four randomly selected sample sites. Orange and blue lines represent the estimated averages at top- (0-0.3 m) and subsoil (0.3-1 m) layers.

Fig. S2
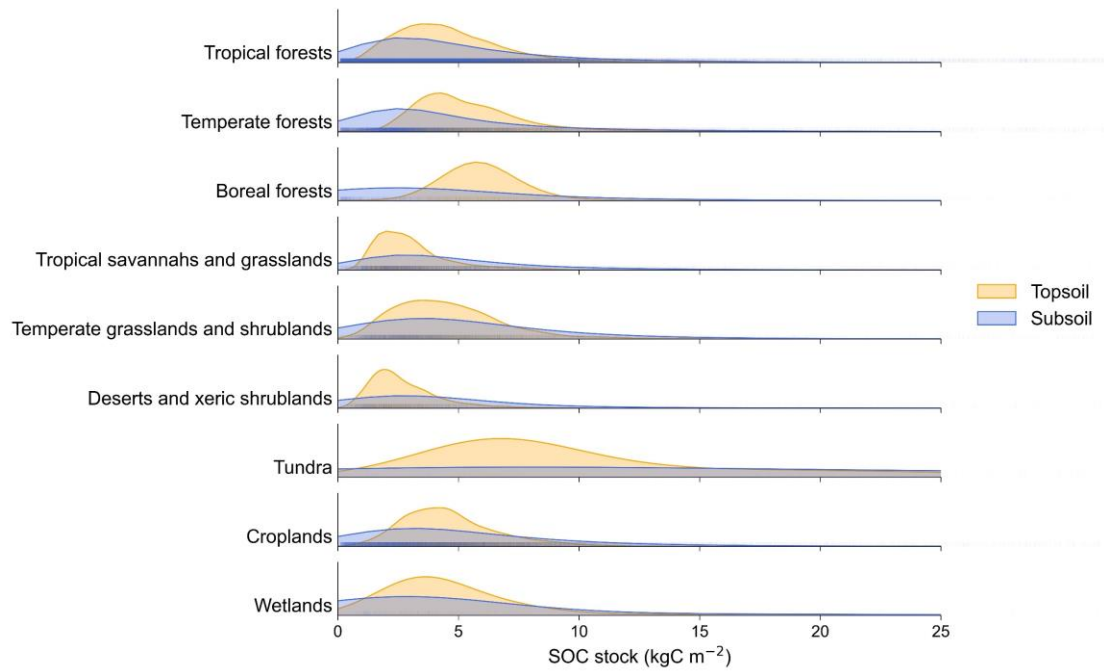
Figure S2. Kernel density estimates of soil organic carbon stock (SOCS, kgC m⁻²) in top- and subsoil by biome, based on data aggregated by soil profile observations.

Fig. S3
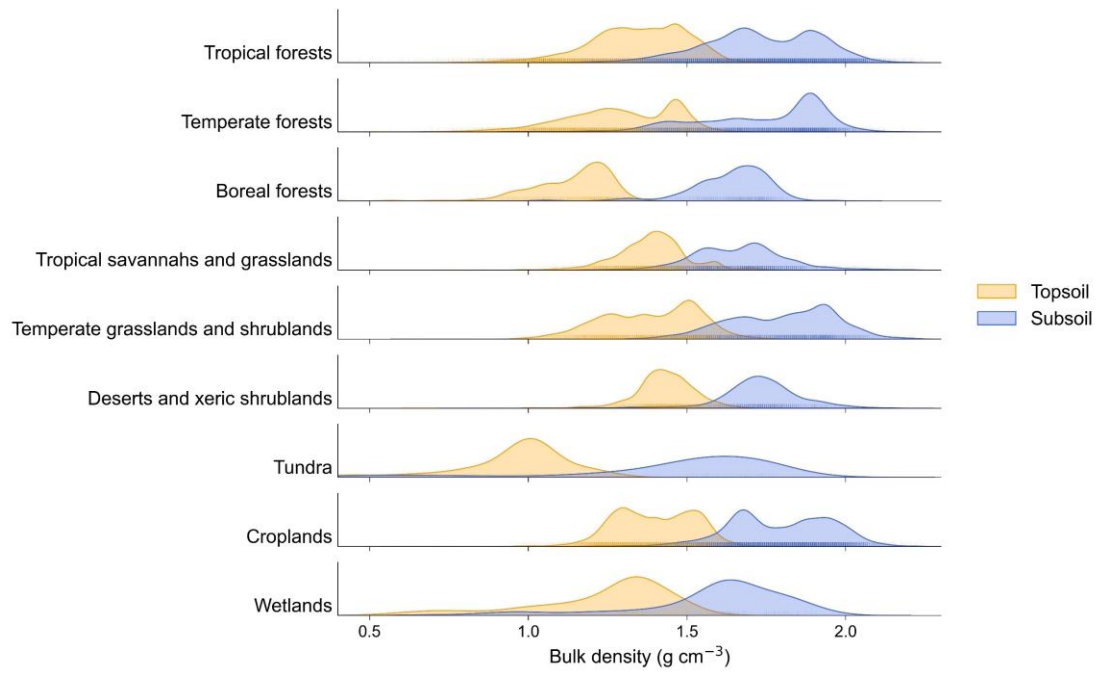
Figure S3. Kernel density estimates of bulk density (g cm^{-3}) in top- and subsoil by biome, based on data aggregated by soil profile observations.

Fig. S4
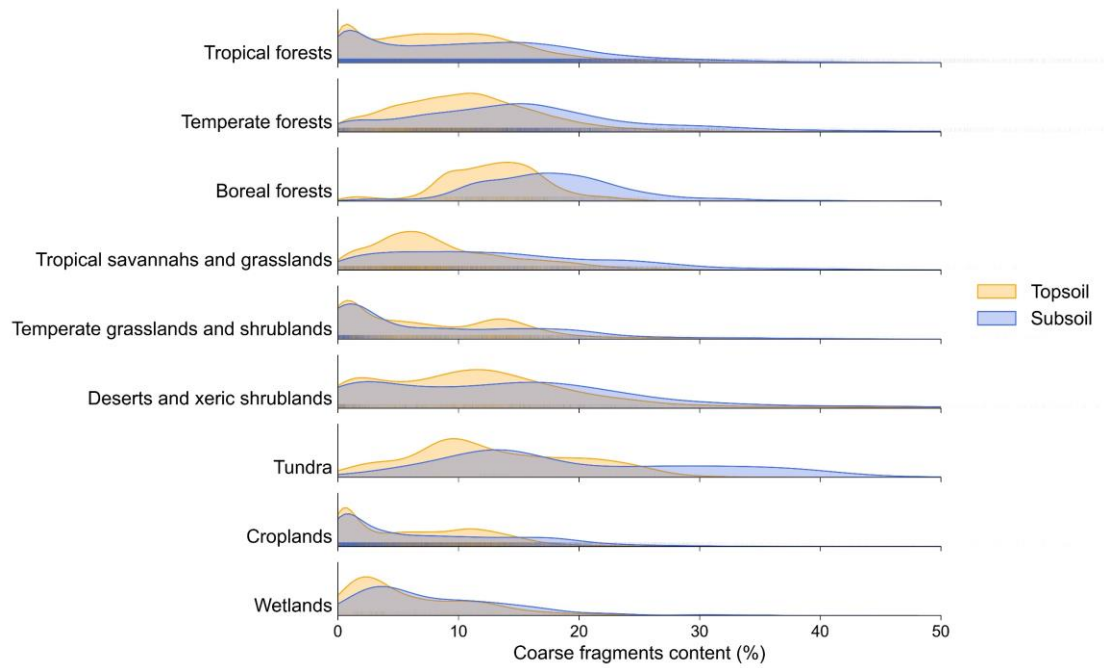
Figure S4. Kernel density estimates of coarse fragments (> 2 mm) content (%) in top- and subsoil by biome, based on data aggregated by soil profile observations.

Fig. S5
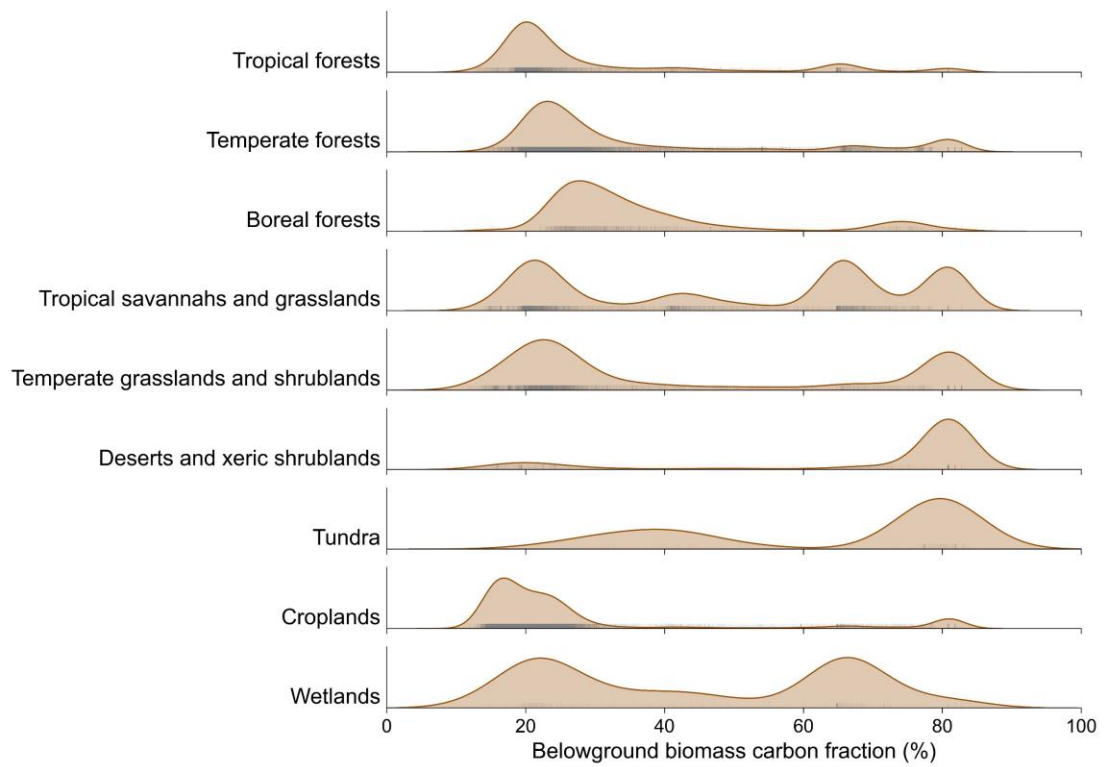
Figure S5. Kernel density estimates of belowground biomass carbon fraction (%) by biome, based on data aggregated by locations of soil profile observations. The root-to-shoot ratio at each sample location was collected from the harmonized global maps of above and belowground biomass carbon density provided by Spawn et al. (2020).

Fig. S6
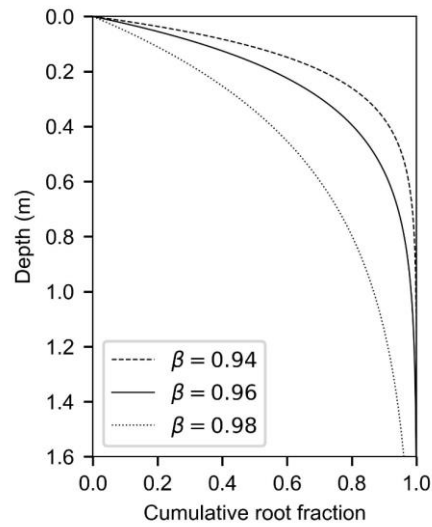
Figure S6. An illustration of the root distribution estimation through the information of belowground extents from the Root Systems of Individual Plants (RSIP) and global the global root traits (GRooT) databases. The cumulative root mass fraction along the depth was estimated by using a asymptotic nonlinear model (Eq. 6) considering the maximum rooting depth (Gale and Grigal, 1987; Zeng, 2001).

Fig. S7
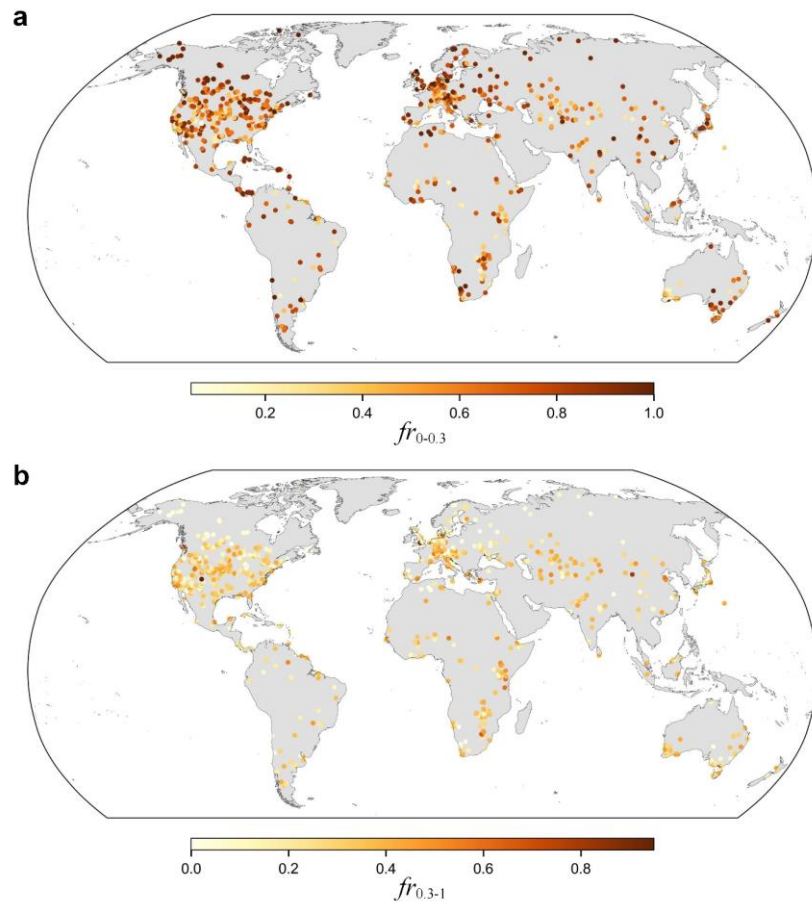
Figure S7. Global distribution of estimated fraction of root from 0 to 0.3 m (a) and from 0.3 to 1 m (b) at available root profiles and root size observations. The observations with a total number of 1,732 geo-referenced measurements were collected from the root profiles with the root distribution information provided by Schenk & Jackson, (2002), the extended Root Systems of Individual Plants (RSIP) database provided by Tumber-Dávila et al. (2022), and the global root traits (GRooT) database provided by Guerrero-Ramírez et al. (2021).

Fig. S8
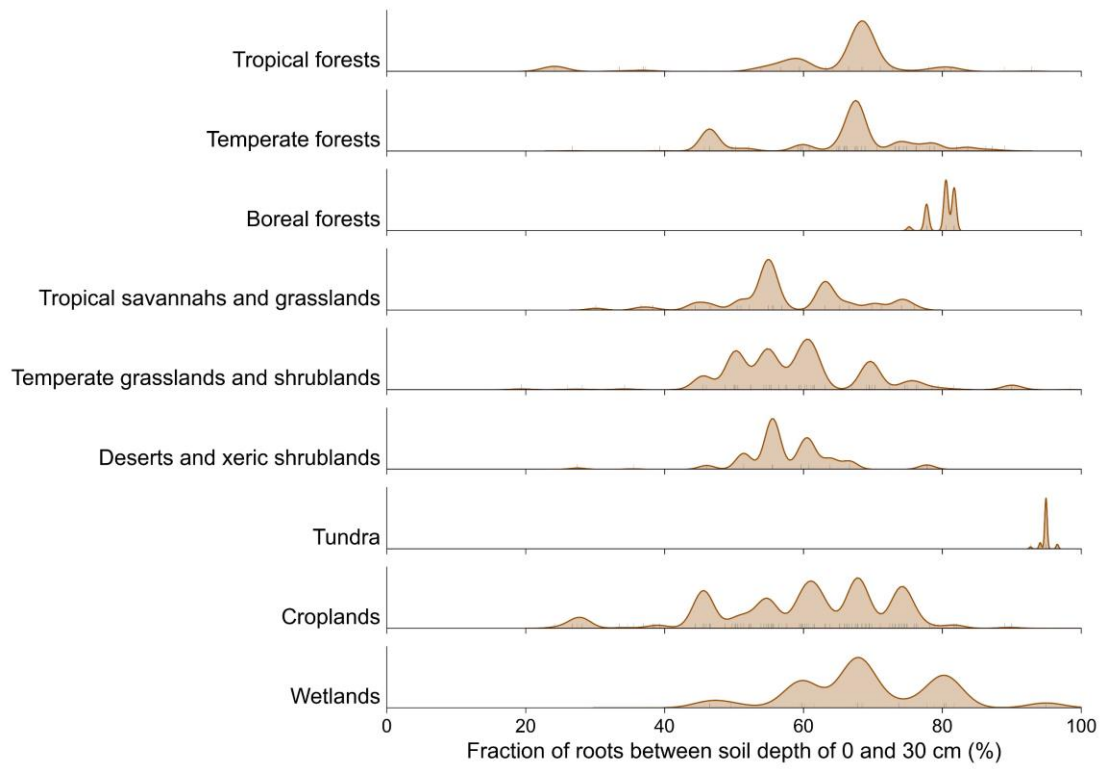
Figure S8. Kernel density estimates of root fraction in topsoil layer (fr_{0-30}) (%) by biome, based on data aggregated by locations of all observations.

Fig. S9
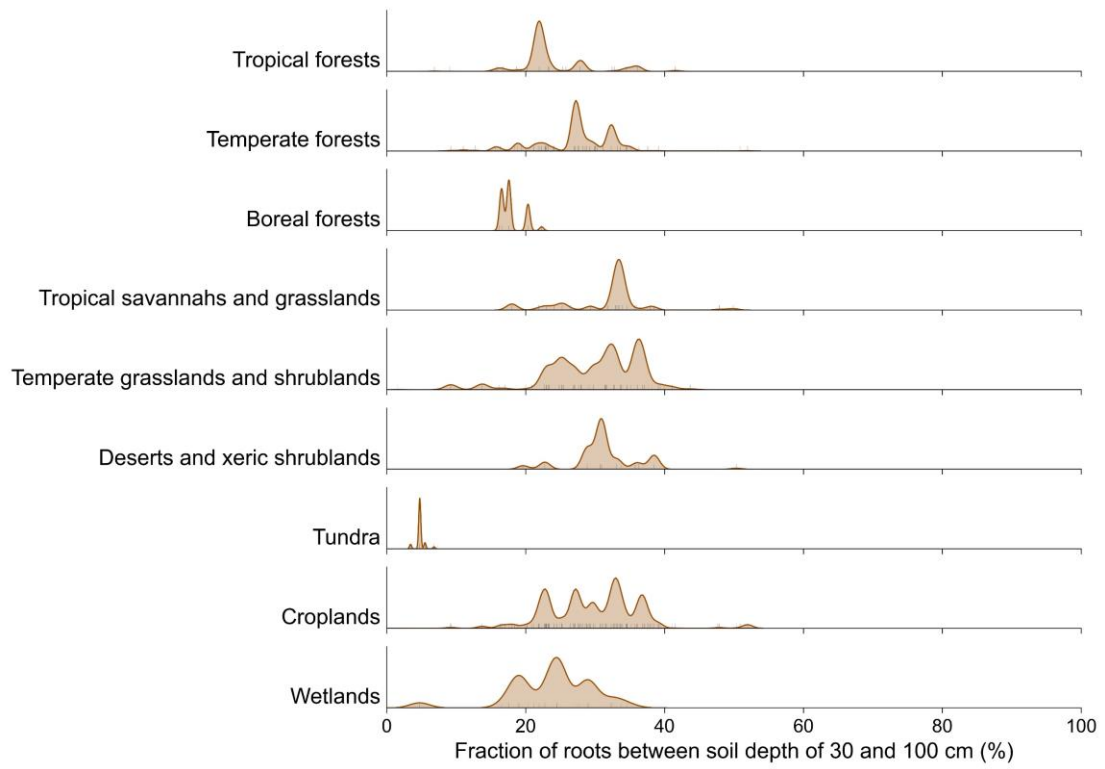
Figure S9. Kernel density estimates of root fraction in subsoil layer ($fr_{30\text{-}100}$) (%) by biome, based on data aggregated by locations of all observations.

Fig. S10
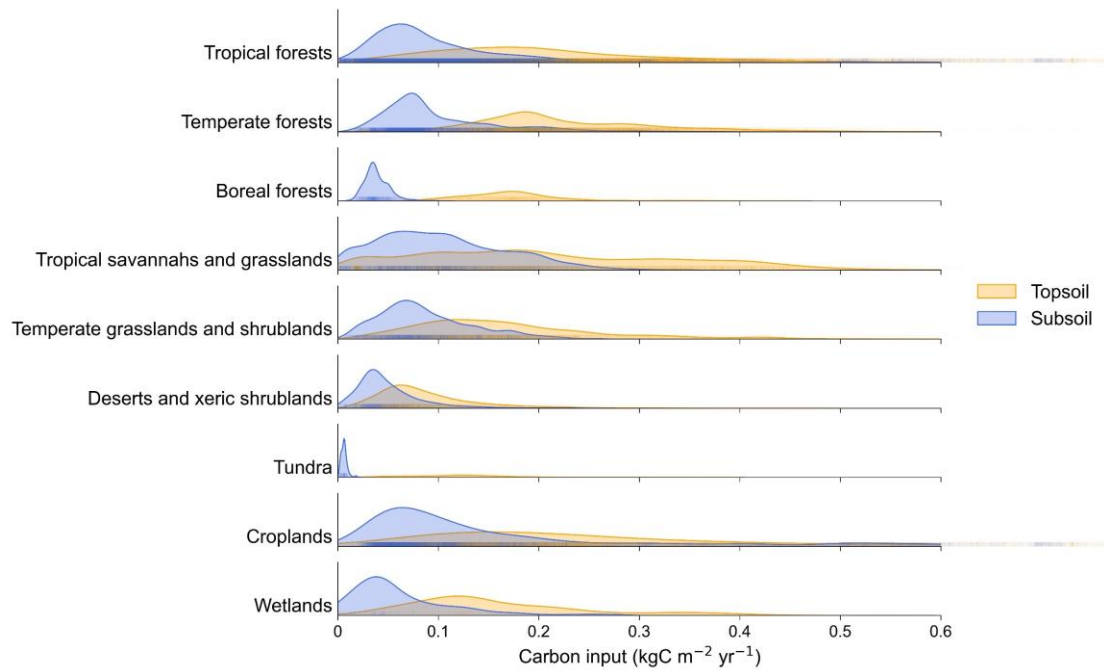
Figure S10. Kernel density estimates of carbon input at top- and subsoil calculated by our method in different biomes, based on data aggregated by soil profile observations combined with multiple remote sensing derived data products.

Fig. S11
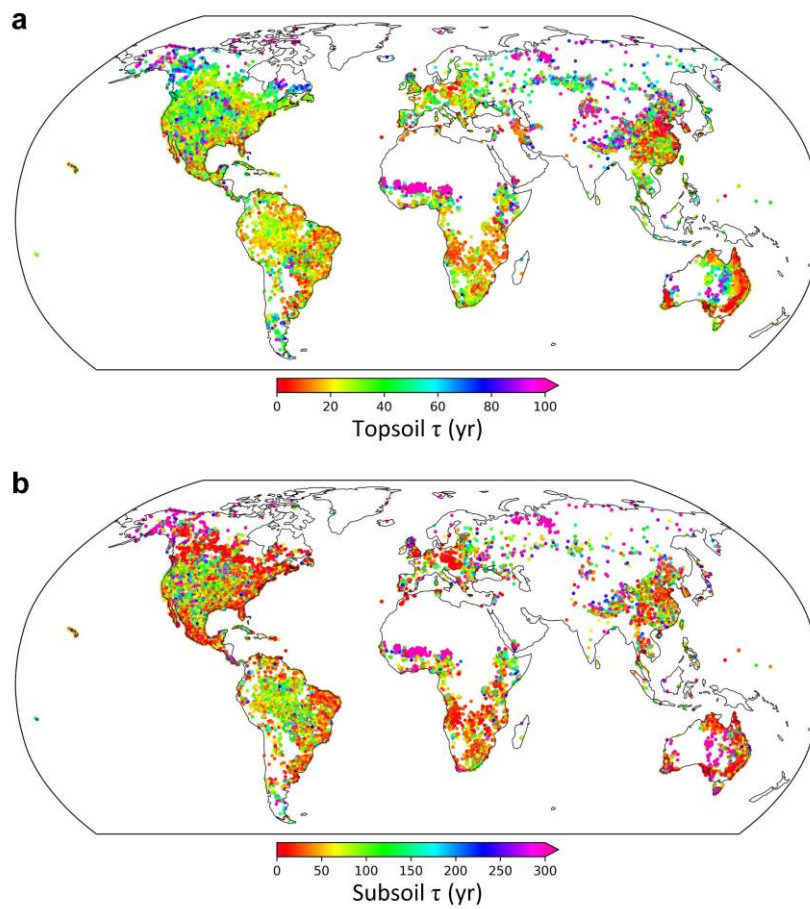
Figure S11. Soil organic carbon (SOC) turnover times (τ) in top- (a) and subsoil (b) layers at sample locations. Values of τ at all sample locations were calculated as the ratio of the total size of SOC stock over the input flux of carbon at that soil layer (see "Materials and Methods" for details of calculations).

Fig. S12
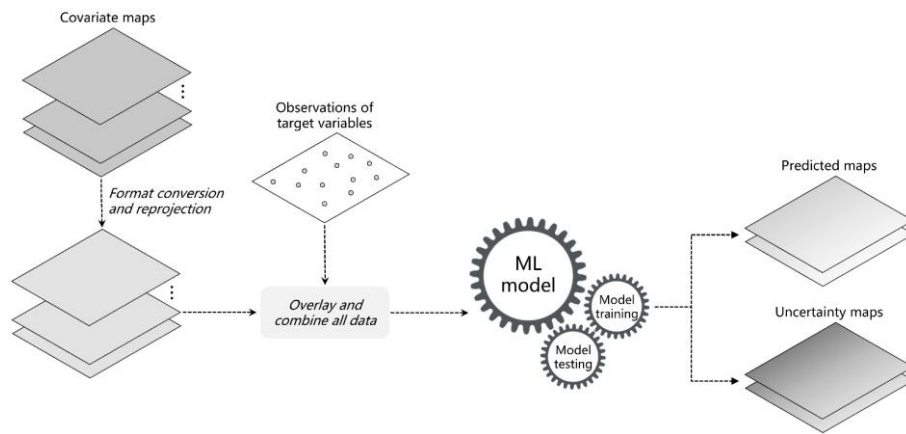
Figure S12. Schematic illustration of the geospatial predictive mapping process. With georeferenced observations of target variables at sample locations and environmental covariate maps as the input, the machine learning models are trained and tested for generating the relationship between the target variable and covariates. The resulting maps produced include the spatially explicit prediction of the target variable and the uncertainty associated with the prediction. Adapted from Hengl et al. (2018).

Fig. S13
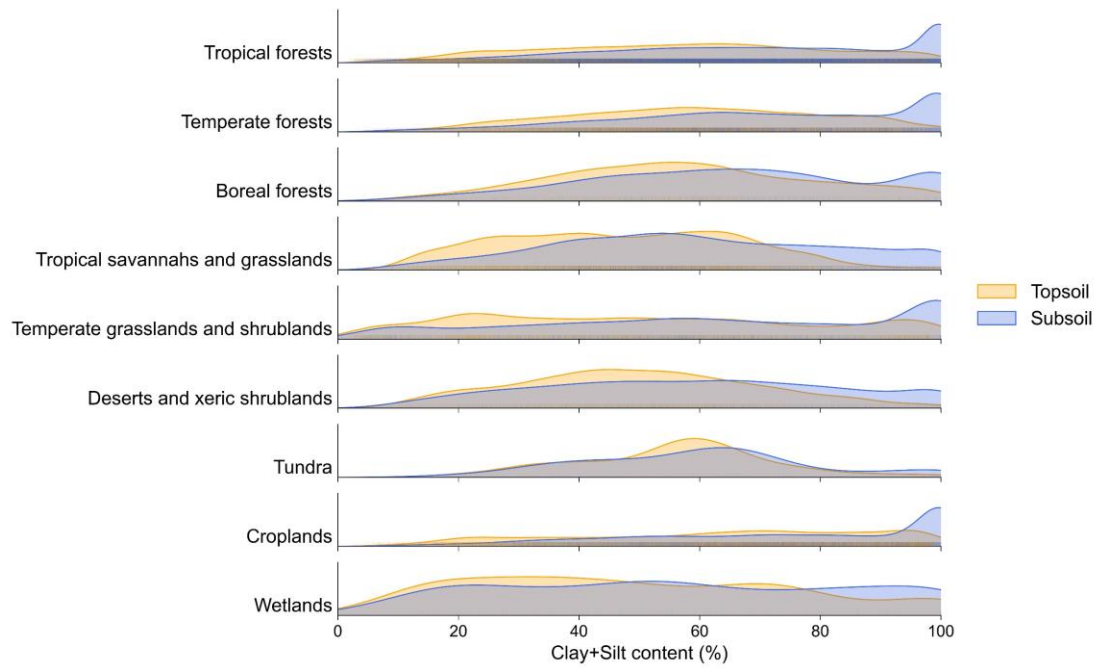
Figure S13. Kernel density estimates of fine particle-size fraction (CLAY+SILT) content (%) in top- and subsoil by biome, based on data aggregated by soil profile observations.

Fig. S14
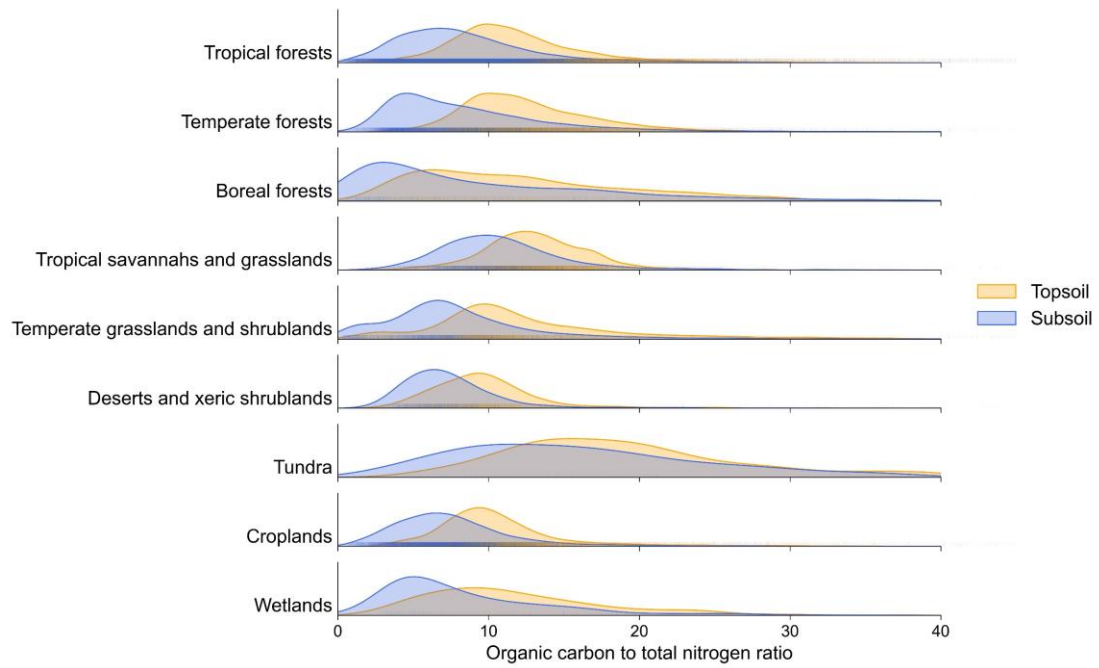
Figure S14. Kernel density estimates of C:N (organic carbon to total nitrogen ratio) in top- and subsoil by biome, based on data aggregated by soil profile observations.

Fig. S15
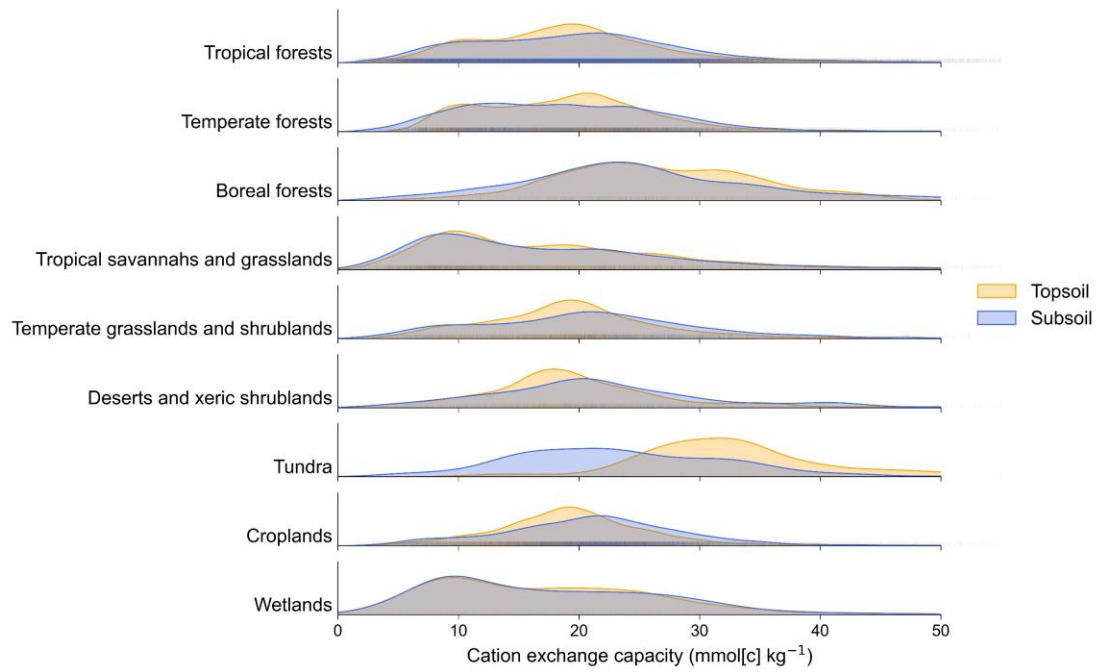
Figure S15. Kernel density estimates of CEC (cation exchange capacity of the soil) in top-and subsoil by biome, based on data aggregated by soil profile observations.

Fig. S16
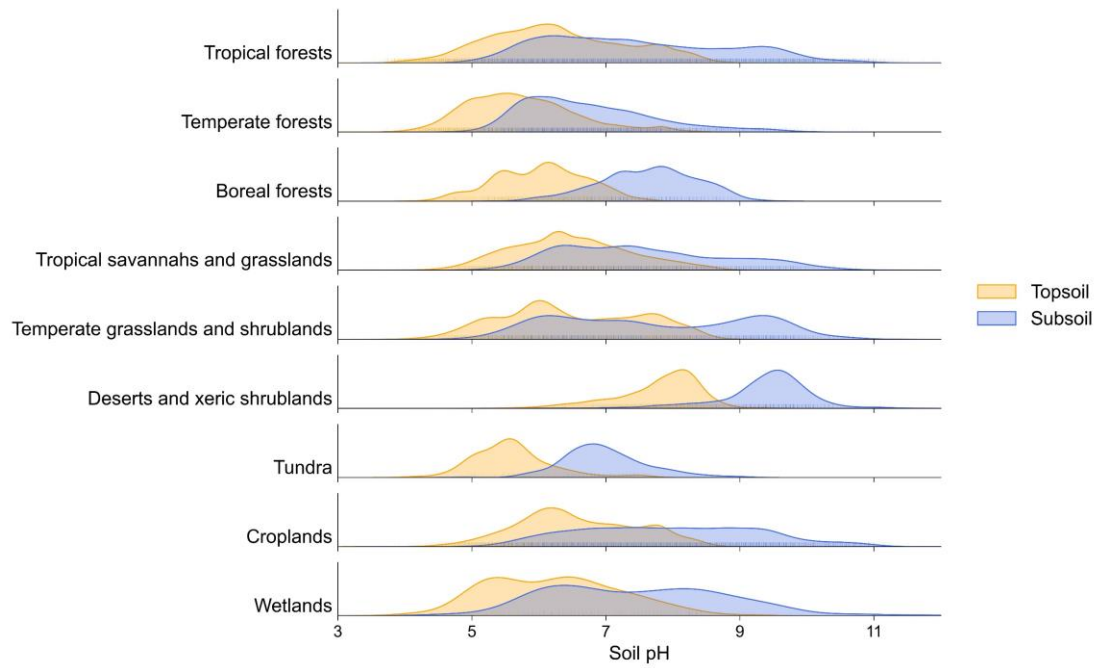
Figure S16. Kernel density estimates of soil pH in top- and subsoil by biome, based on data aggregated by soil profile observations.

Fig. S17
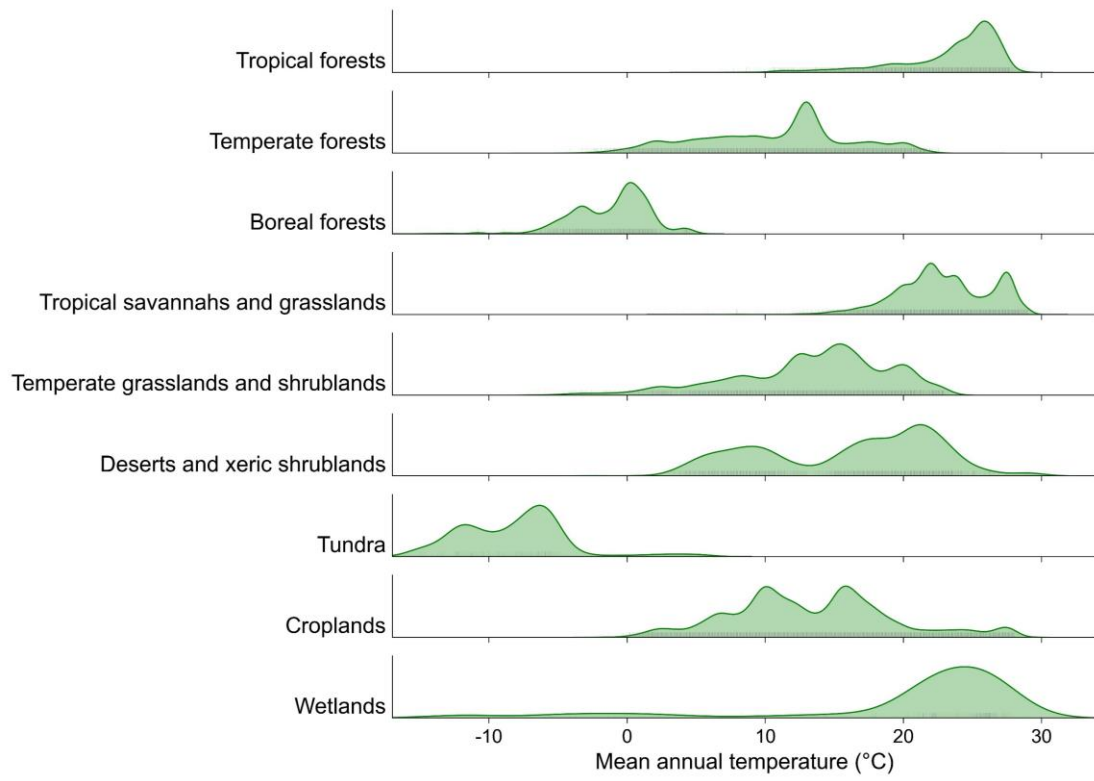
Figure S17. Kernel density estimates of mean annual temperature (°C) by biome, based on data aggregated by locations of soil profile observations.

Fig. S18
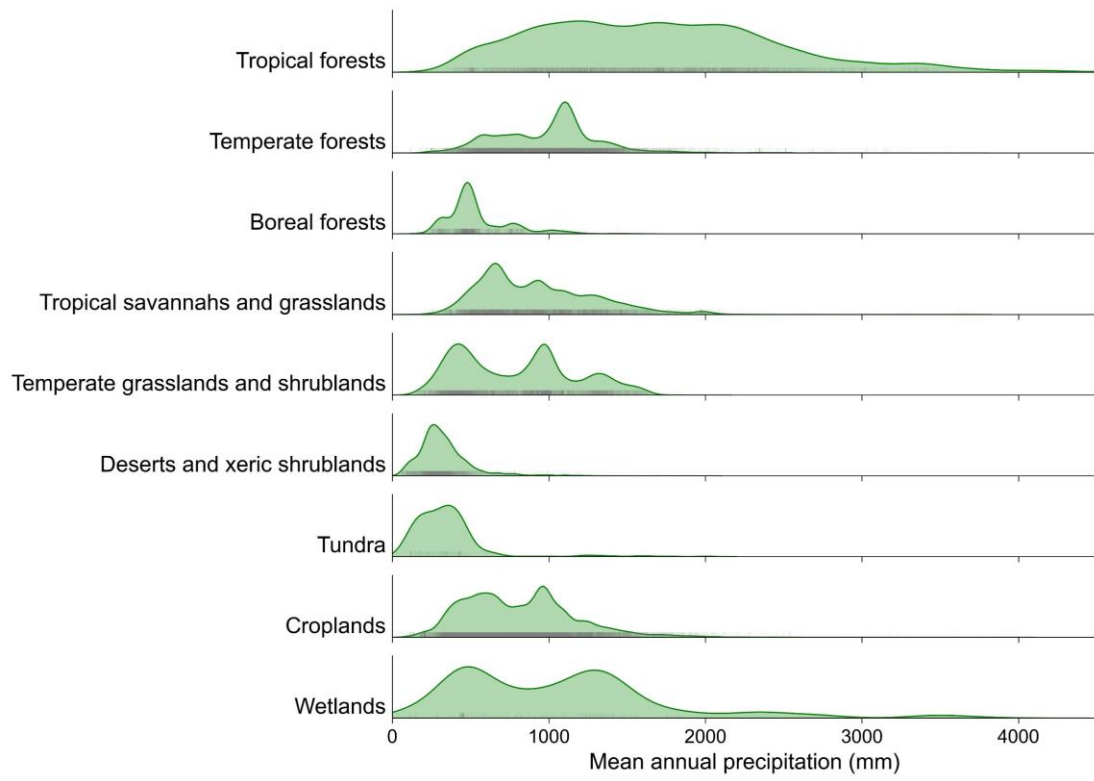
Figure S18. Kernel density estimates of mean annual precipitation (mm) by biome, based on data aggregated by locations of soil profile observations.

Fig. S19
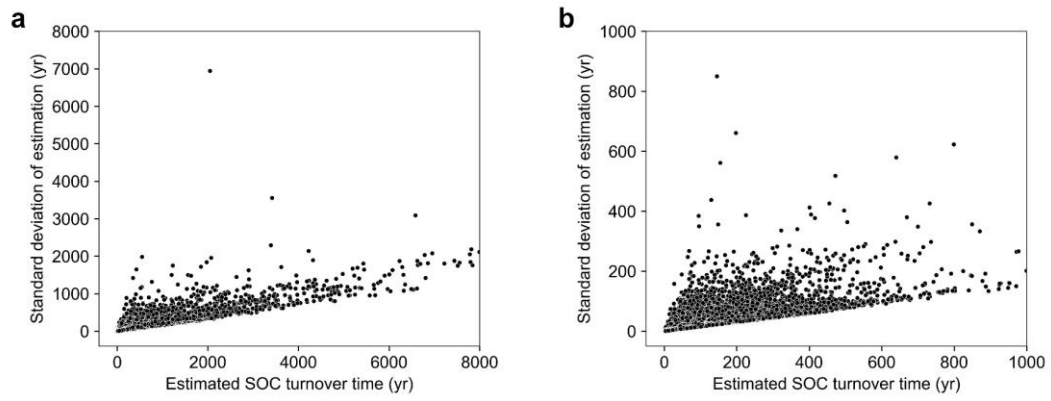
Figure S19. The relationship between the estimated values of SOC turnover time and their estimation error (standard deviation) at sample locations. The error was determined by the cumulative uncertainty of the inputs that participate in calculation of SOC turnover time according to the error propagation theory.

Fig. S20
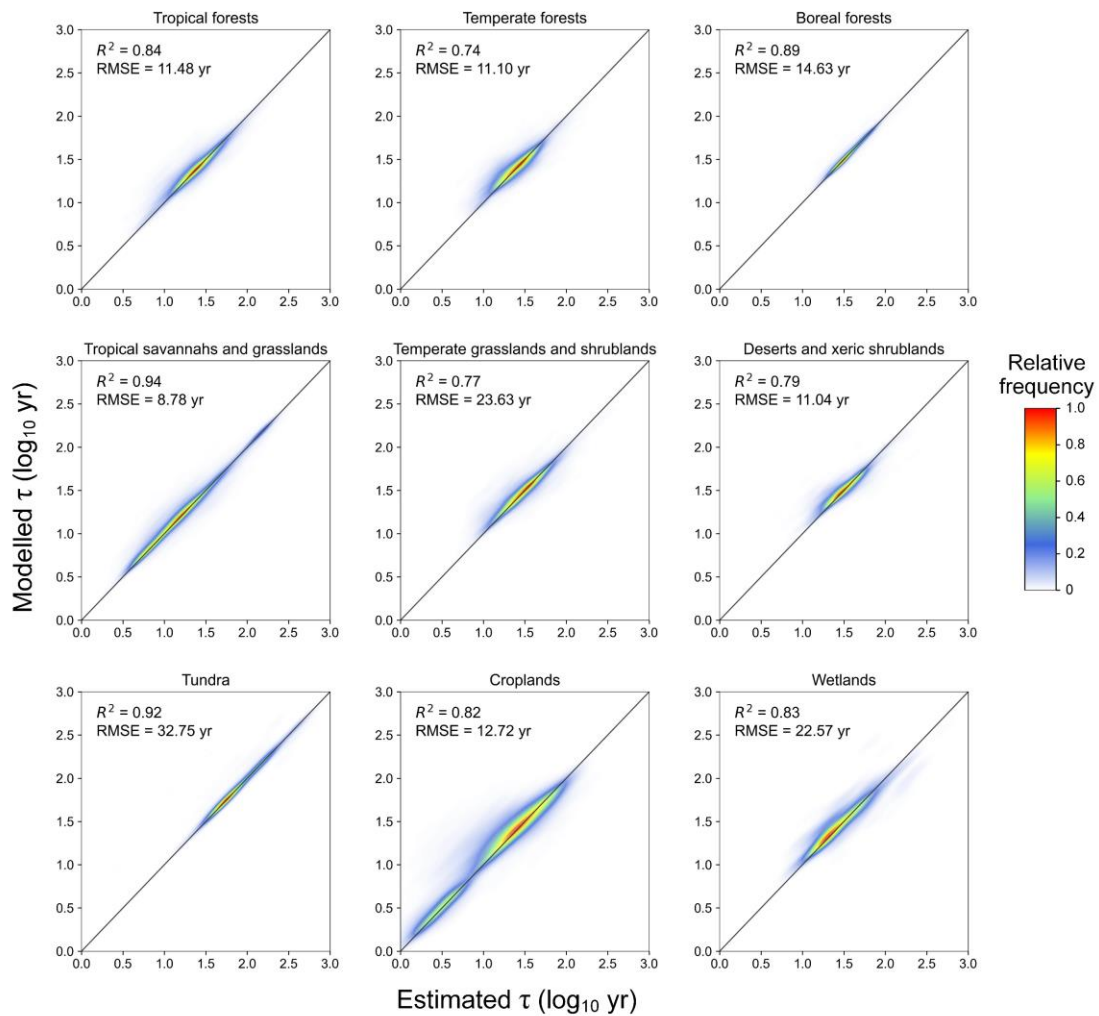
Figure S20. Plots of measured versus predicted values of SOC turnover times (τ, yr) at topsoil (0–0.3 m) layer in different biomes. R^2 values represent the coefficient of determination relative to the 1:1 line (solid diagonal lines) of the estimated τ values based on observed data and the predicted values for all validation samples using 10-fold cross-validation. Colors indicate the proportion (relative frequency) of data points at each pixel. Validation results show that biomes with a greater number of sample data generally have larger variability.
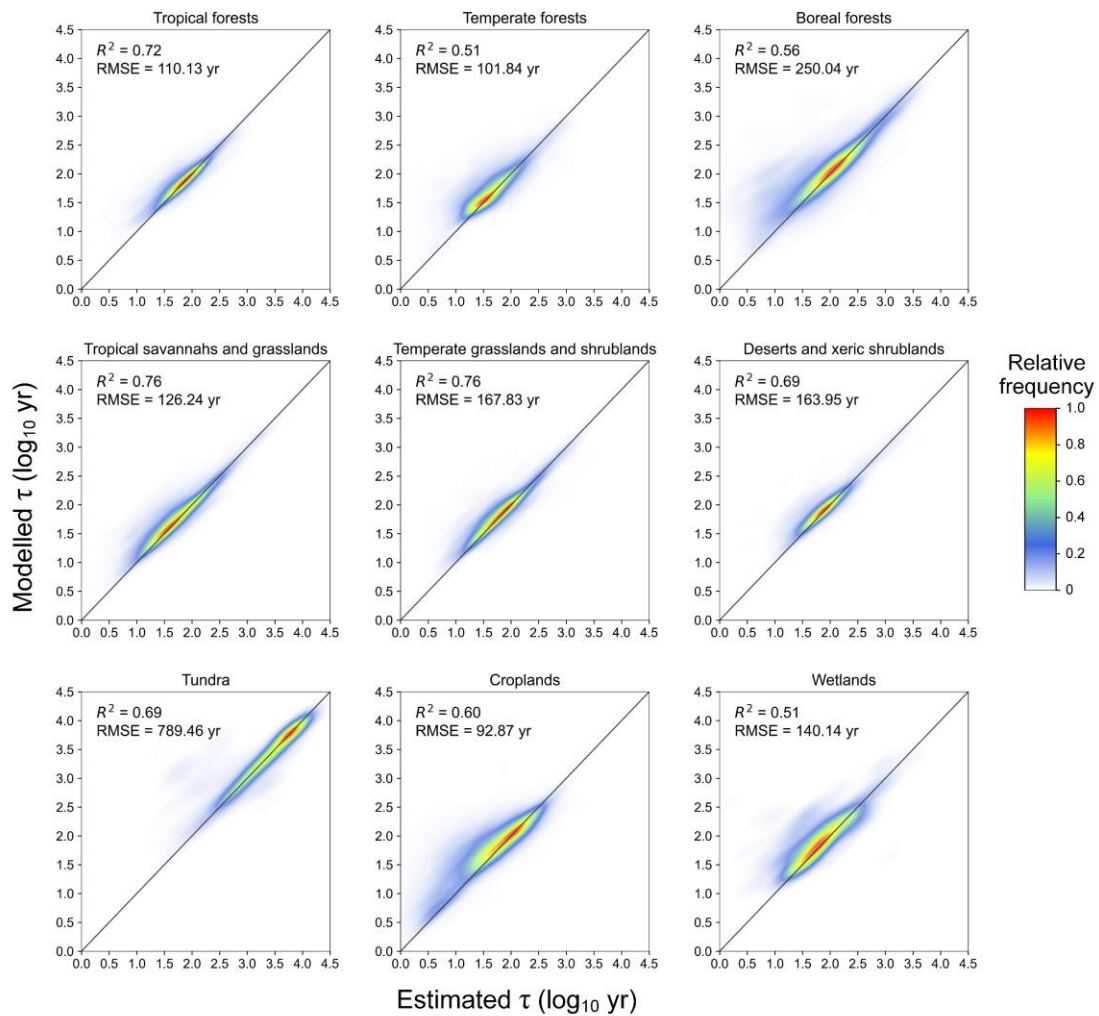
Figure S21. Plots of measured versus predicted values of SOC turnover times (τ, yr) at subsoil (0.3–1 m) layer in different biomes. R^2 values represent the coefficient of determination relative to the 1:1 line (solid diagonal lines) of the estimated τ values based on observed data and the predicted values for all validation samples using 10-fold cross-validation. Colors indicate the proportion (relative frequency) of data points at each pixel. Validation results show that biomes with a greater number of sample data generally have larger variability.
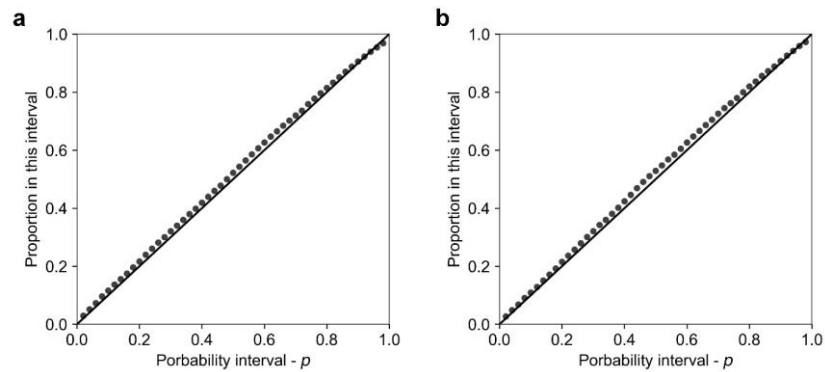
Figure S22. Plots of the proportion of observed τ values at top- (a) and subsoil (b) layers in validation set falling within probability intervals (accuracy plot (Goovaerts, 2001)**).** The values on *x*-axis represent a series of symmetric *p*-probability intervals (PI) bounded by the $(1 - p)/2$ and $(1 + p)/2$ quantiles. The values on *y*-axis represent the probability that the actual target values (the observed τ values in validation set) fall into each *p*-PI. A correct modelling of the prediction uncertainty would entail that the proportion of validation points covered in each prediction interval approximately equals the value of that PI, for all values of *p*. The plotted points substantially higher than the 1:1 line suggests that the uncertainty was overestimated, and the substantially lower values indicate that it was underestimated.

Fig. S23
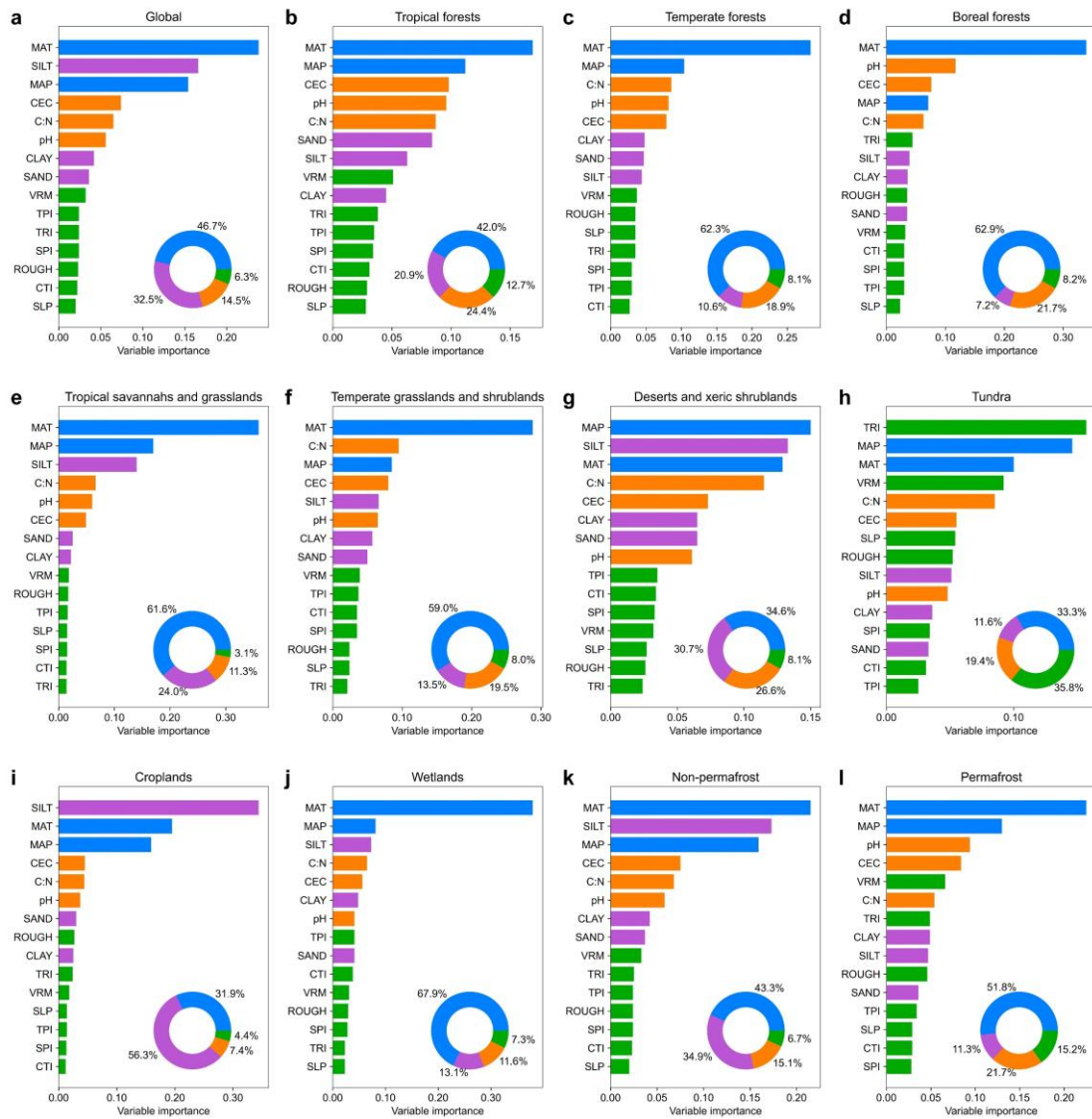
Figure S23. The influences of climate, soil physical and chemical properties, and topography factors on the topsoil organic carbon turnover time (τ). Variable importance metrics at the global scale and main biomes are calculated by using the random forest model for evaluating the effects of these environmental variables on τ at topsoil layer.

Fig. S24
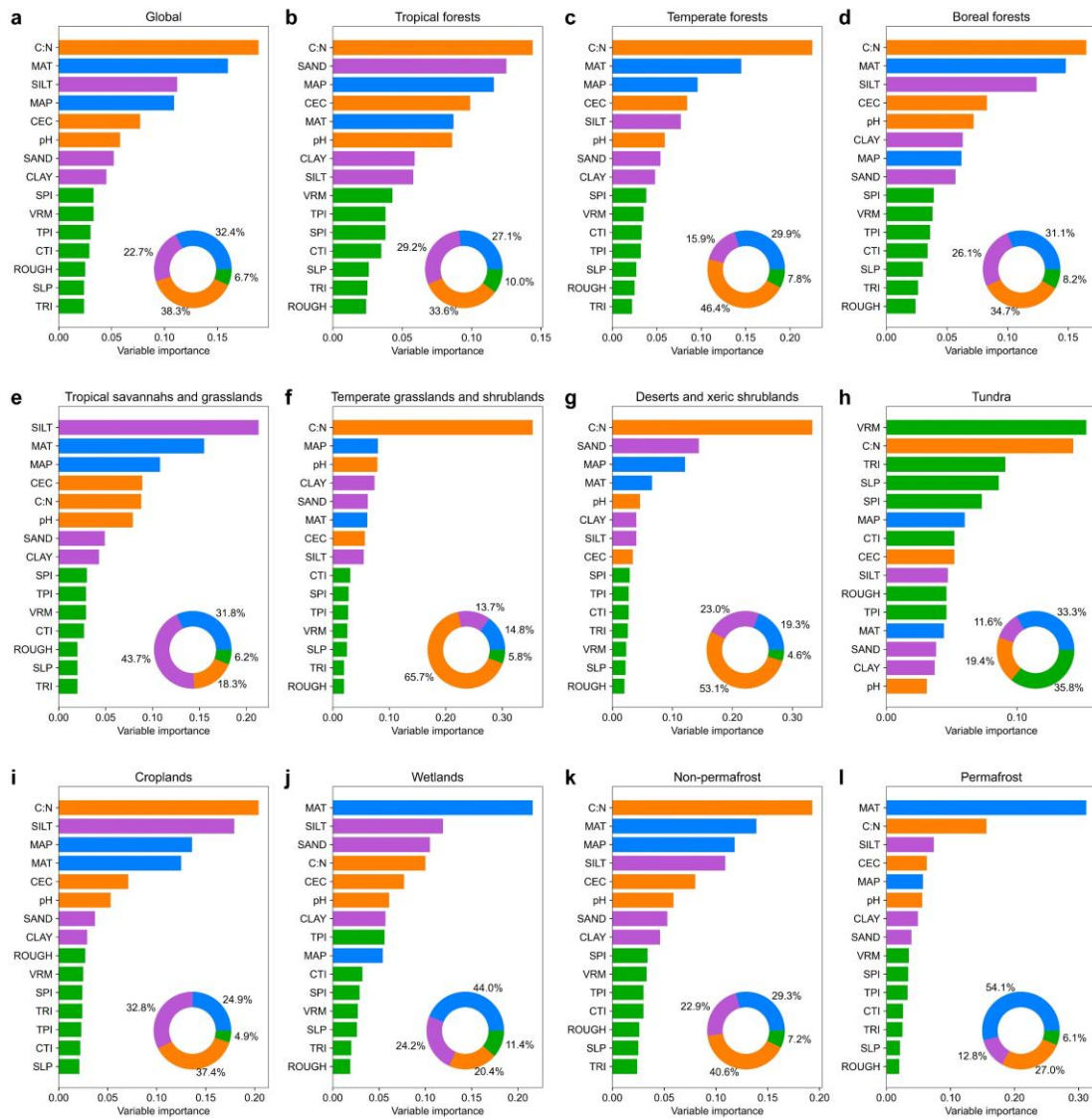
Figure S24. The influences of climate, soil physical and chemical properties, and topography factors on the subsoil organic carbon turnover time (τ). Variable importance metrics at the global scale and main biomes are calculated by using the random forest model for evaluating the effects of these environmental variables on τ at subsoil layer.

Fig. S25
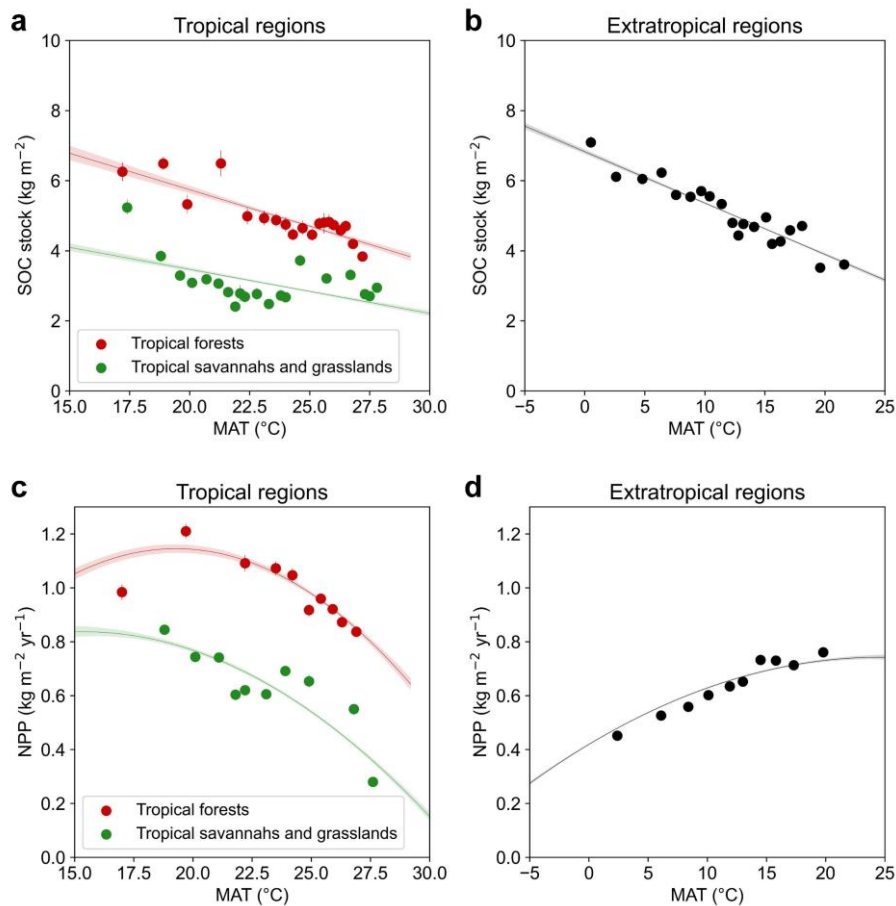
Figure S25. Relationships of SOC stock and NPP with temperature in tropical and extratropical regions. Points are the averaged SOC stock or NPP values at soil sample locations within each bin of mean annual temperature (MAT) values (MAT were divided into 20 discrete bins). Solid lines or curves (using 2nd order polynomial) represent the results of their regression fits. This reflects a uniquely tropical phenomenon contrasting to extratropical regions. The tropical plants are often near their thermal optimum for photosynthesis (related to the role of vapor pressure deficit [VPD]-induced stomatal closure), and the rising temperature can lead to higher respiration rates, thereby potentially reducing the net carbon gain in tropical ecosystems. Whereas in extratropical regions, it usually shows that temperature becomes a more critical limiting factor, with NPP generally increasing with temperature up to a certain point.

Fig. S26
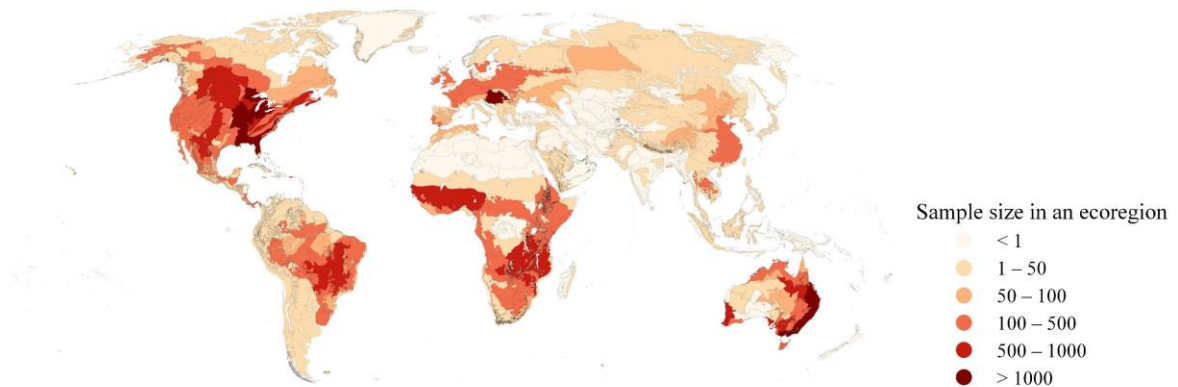
Figure S26. Sample size in each ecoregion across the globe. The global terrestrial ecoregions were classified according to the map provided by Dinerstein et al. (2017). The darkness of colors represents the number of samples fall in a certain ecoregion.
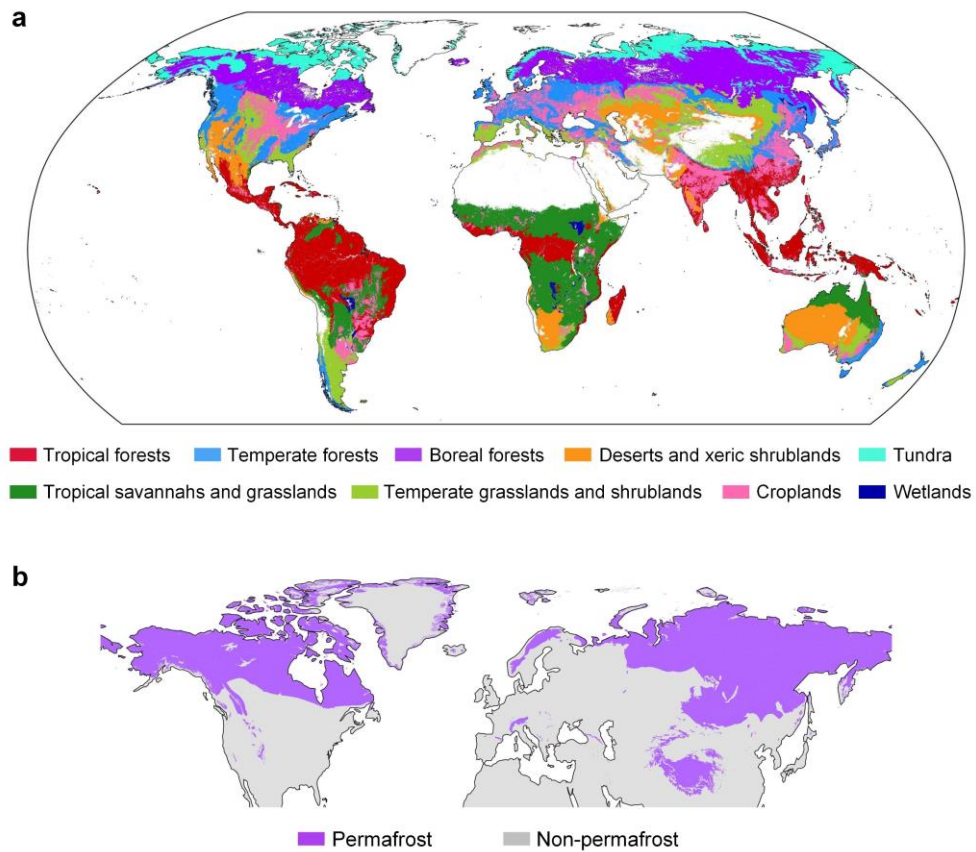
Figure S27. The global biome map used in this study. The nine biome types (a) were reclassified from the terrestrial biome map provided by Dinerstein et al. (2017). All forests, savannahs, grasslands and shrublands were reclassified according to latitude into tropical (-23°S ~ 23°N), temperate (23° ~ 50°) or boreal (> 50°) categories. Cropland areas were extracted from MODIS Land Cover MCD12Q1 product (Friedl and Sulla-Menashe, 2019). Areas with low NPP (below 10 gC m^{-2} yr^{-1}) were removed in this map. The global permafrost and non-permafrost areas were classified by using the Circum-Arctic permafrost and ground ice map provided by Brown et al. (2002).

Supplementary Tables

Table S1. Overview of datasets used for the calculation of SOC turnover time at sample locations.

Variable name	Description	Unit	Source
SOC	Soil organic carbon content	g kg⁻¹	WoSIS (Batjes et al., 2020), NCSCD (Hugelius et al., 2013), China soil survey (Zhang et al., 2013; Liu et al., 2022), and SoilGrids (Hengl et al., 2017; Poggio et al., 2021)
BD	Bulk density	g cm⁻³	
CF	Coarse fragments content	%	
NPP	Net primary production	kg m⁻² yr⁻¹	The MOD17A3HGF Version 6.1 product (Running and Zhao, 2019)
RSR	Root-shoot ratio	%	Harmonized global above and belowground biomass carbon density maps (Spawn et al., 2020)
$fr_{D_u - D_l}$	Fraction of root in a soil layer between upper (D_u) and lower depth (D_l)	%	Global root profile database (Schenk and Jackson, 2002), RSIP database (Tumber-Dávila et al., 2022), and GRooT database (Guerrero-Ramírez et al., 2021)

Note: WoSIS, standardized soil profile data from World Soil Information Service (WoSIS) registered in the ISRIC (World Soil Information) data repository; NCSCD, the Northern Circumpolar Soil Carbon Database. SoilGrids, the global gridded soil information that uses state-of-the-art machine learning methods to map the spatial distribution of soil properties across the globe.

Table S2. Details on four categories of environmental variables for predictive mapping of SOC turnover time (τ) and analyzing the driving mechanism of τ.

Variable category	Variable name	Code	Description	Unit	Source
Climate	Temperature	MAT	Mean annual temperature	°C	WorldClim (Fick and Hijmans, 2017)
	Precipitation	MAP	Mean annual precipitation	mm	
Soil physical property	Clay	CLAY	Proportion of clay particles (< 0.002 mm) in the fine earth fraction	g kg^{-1}	WoSIS (Batjes et al., 2020), NCSCD (Hugelius et al., 2013), SoilGrid (Hengl et al., 2017; Poggio et al., 2021)
	Sand	SAND	Proportion of sand particles (> 0.05 mm) in the fine earth fraction	g kg^{-1}	
	Silt	SILT	Proportion of silt particles (> 0.002 mm and < 0.05 mm) in the fine earth fraction	g kg^{-1}	
Soil chemical property	Cation exchange capacity	CEC	Cation Exchange Capacity of the soil	cmol$_c$ kg^{-1}	
	C N ratio	C:N	Organic carbon to total nitrogen ratio	-	
	pH	pH	Soil pH	-	
Topography	Elevation	ELEV	Elevation	m	GTOPO30 (Gesch et al., 1999)
	Slope	SLP	Slope	°	Geomorpho90m (Amatulli et al., 2020)
	Compound topographic index	CTI	Compound topographic index	-	
	Terrain ruggedness index	TRI	Terrain ruggedness index	-	
	Vector ruggedness measure	VRM	Vector ruggedness measure	-	
	Roughness	ROUGH	Roughness	-	
	Topographic position index	TPI	Topographic position index	-	
	Stream power index	SPI	Stream power index	-	

Note: WorldClim, a database of high spatial resolution global weather and climate data. WoSIS, the standardised soil profile data from World Soil Information Service (WoSIS) registered in the ISRIC (World Soil Information) data repository; NCSCD, the Northern Circumpolar Soil Carbon Database. SoilGrid, the global gridded soil information that uses state-of-the-art machine learning methods to map the spatial distribution of soil properties across the globe; GTOPO30, a global digital elevation model (DEM) developed by the U.S. Geological Survey's (USGS) Center for Earth Resources Observation and Science (EROS). Geomorpho90m, a global dataset comprising of different geomorphometric features derived from the high-resolution DEM.

Table S3. Differences of our data-derived SOC turnover time (τ) and τ from selected Earth system models from CMIP6. The values of τ difference report ratio of ESM-derived τ to our observation-derived τ in four statistical indices by aggregating the global estimates.

Model	τ difference				Proportion of area on global land with the τ difference less than 1
	P 2.5	**Median**	**P 97.5**	**SD**	
ACCESS-ESM1-5	0.09	0.42	1.80	0.47	92%
CESM2-WACCM	0.09	0.47	2.44	0.62	89%
IPSL-CM6A-LR	0.02	0.25	1.15	0.37	97%
NorESM2-LM	0.08	0.42	2.70	0.63	89%
Average	0.07	0.39	2.02	0.52	92%

Note: P 2.5, P 97.5 and SD represent the range of τ difference values between the 2.5th and 97.5th percentiles and standard deviation from the aggregation of all grid cells globally.

Table S4. Differences of our data-derived SOC turnover time (τ) and τ from selected Earth system models from CMIP6 in different biomes. The values of τ difference report ratio of ESM-derived τ to our observation-derived τ in four statistical indices by aggregating the global estimates.

Biome	τ difference			
	P 2.5	Median	P 97.5	SD
Tropical forests	0.12	0.29	0.70	0.20
Temperate forests	0.22	0.53	0.90	0.23
Boreal forests	0.15	0.54	1.71	0.61
Tropical savannahs and grasslands	0.07	0.43	0.99	0.53
Temperate grasslands and shrublands	0.20	0.50	1.32	0.43
Deserts and xeric shrublands	0.06	0.53	3.15	0.96
Tundra	0.07	0.41	2.39	0.80
Croplands	0.16	0.46	1.03	0.30
Wetlands	0.15	0.47	2.34	0.78

Note: P 2.5, P 97.5 and SD represent the range of τ difference values between the 2.5[th] and 97.5[th] percentiles and standard deviation from the aggregation of all grid cells globally.

Table S5. The estimated mean of the fraction of litterfall in ANPP for each biome.

Biome	Fraction of litterfall in ANPP
Boreal Forests/Taiga	0.59
Deserts & Xeric Shrublands	0.90
Flooded Grasslands & Savannas	0.50
Mangroves	0.50
Mediterranean Forests, Woodlands & Scrub	0.41
Montane Grasslands & Shrublands	0.71
Temperate Broadleaf & Mixed Forests	0.73
Temperate Conifer Forests	0.70
Temperate Grasslands, Savannas & Shrublands	0.65
Tropical & Subtropical Coniferous Forests	0.70
Tropical & Subtropical Dry Broadleaf Forests	0.68
Tropical & Subtropical Grasslands, Savannas & Shrublands	0.59
Tropical & Subtropical Moist Broadleaf Forests	0.65
Tundra	0.56

Table S6. The estimated mean of the fraction of litterfall in ANPP for forest areas with different leaf types.

Biome	Leaf type	Fraction of litterfall in ANPP
Temperate forests	Deciduous Broadleaf	0.70
	Evergreen Broadleaf	0.55
	Evergreen Needleleaf	0.71
Tropical forests	Deciduous Broadleaf	0.68
	Evergreen Broadleaf	0.69
	Evergreen Needleleaf	0.83

References

Abramoff, R. Z., Guenet, B., Zhang, H., Georgiou, K., Xu, X., Viscarra Rossel, R. A., Yuan, W., and Ciais, P.: Improved global-scale predictions of soil carbon stocks with Millennial Version 2, Soil Biology and Biochemistry, 164, 108466, https://doi.org/10.1016/j.soilbio.2021.108466, 2022.

Adalibieke, W., Cui, X., Cai, H., You, L., and Zhou, F.: Global crop-specific nitrogen fertilization dataset in 1961–2020, Sci Data, 10, 617, https://doi.org/10.1038/s41597-023-02526-z, 2023.

Amatulli, G., McInerney, D., Sethi, T., Strobl, P., and Domisch, S.: Geomorpho90m, empirical evaluation and accuracy assessment of global high-resolution geomorphometric layers, Sci Data, 7, 162, https://doi.org/10.1038/s41597-020-0479-6, 2020.

Batjes, N. H., Ribeiro, E., and van Oostrum, A.: Standardised soil profile data to support global mapping and modelling (WoSIS snapshot 2019), Earth System Science Data, 12, 299–320, https://doi.org/10.5194/essd-12-299-2020, 2020.

Baudron, F., Jaleta, M., Okitoi, O., and Tegegn, A.: Conservation agriculture in African mixed crop-livestock systems: Expanding the niche, Agriculture, Ecosystems & Environment, 187, 171–182, https://doi.org/10.1016/j.agee.2013.08.020, 2014.

Baudron, F., Delmotte, S., Corbeels, M., Herrera, J. M., and Tittonell, P.: Multi-scale trade-off analysis of cereal residue use for livestock feeding vs. soil mulching in the Mid-Zambezi Valley, Zimbabwe, Agricultural Systems, 134, 97–106, https://doi.org/10.1016/j.agsy.2014.03.002, 2015.

Bishop, T. F. A., McBratney, A. B., and Laslett, G. M.: Modelling soil attribute depth functions with equal-area quadratic smoothing splines, Geoderma, 91, 27–45, https://doi.org/10.1016/S0016-7061(99)00003-8, 1999.

Brown, J., Ferrians, O., Heginbottom, J. A., and Melnikov, E.: Circum-Arctic Map of Permafrost and Ground-Ice Conditions, Version 2, https://doi.org/10.7265/SKBG-KF16, 2002.

Dinerstein, E., Olson, D., Joshi, A., Vynne, C., Burgess, N. D., Wikramanayake, E., Hahn, N., Palminteri, S., Hedao, P., Noss, R., Hansen, M., Locke, H., Ellis, E. C., Jones, B., Barber, C. V., Hayes, R., Kormos, C., Martin, V., Crist, E., Sechrest, W., Price, L., Baillie, J. E. M., Weeden, D., Suckling, K., Davis, C., Sizer, N., Moore, R., Thau, D., Birch, T., Potapov, P., Turubanova, S., Tyukavina, A., de Souza, N., Pintea, L., Brito, J. C., Llewellyn, O. A., Miller, A. G., Patzelt, A., Ghazanfar, S. A., Timberlake, J., Klöser, H., Shennan-Farpón, Y., Kindt, R., Lillesø, J.-P. B., van Breugel, P., Graudal, L., Voge, M., Al-Shammari, K. F., and Saleem, M.: An Ecoregion-Based Approach to Protecting Half the Terrestrial Realm, BioScience, 67, 534–545, https://doi.org/10.1093/biosci/bix014, 2017.

Doetterl, S., Stevens, A., Six, J., Merckx, R., Van Oost, K., Casanova Pinto, M., Casanova-Katny, A., Muñoz, C., Boudin, M., Zagal Venegas, E., and Boeckx, P.: Soil carbon storage controlled by interactions between geochemistry and climate, Nature Geosci, 8, 780–783, https://doi.org/10.1038/ngeo2516, 2015.

Erenstein, O.: Cropping systems and crop residue management in the Trans-Gangetic Plains: Issues and challenges for conservation agriculture from village surveys, Agricultural Systems, 104, 54–62, https://doi.org/10.1016/j.agsy.2010.09.005, 2011.

FAO - CMCC: Global Soil Organic Carbon Database, FAO, Rome, Italy, 47 pp., 2017.

Fick, S. E. and Hijmans, R. J.: WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas, International Journal of Climatology, 37, 4302–4315, https://doi.org/10.1002/joc.5086, 2017.

Friedl, M. A. and Sulla-Menashe, D.: MCD12Q1 MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 500m SIN Grid V006, https://doi.org/10.5067/MODIS/MCD12Q1.006, 2019.

Gale, M. R. and Grigal, D. F.: Vertical root distributions of northern tree species in relation to successional status,

Can. J. For. Res., 17, 829–834, https://doi.org/10.1139/x87-131, 1987.

Gesch, D. B., Verdin, K. L., and Greenlee, S. K.: New land surface digital elevation model covers the Earth, Eos, Transactions American Geophysical Union, 80, 69–70, https://doi.org/10.1029/99EO00050, 1999.

Goovaerts, P.: Geostatistical modelling of uncertainty in soil science, Geoderma, 103, 3–26, https://doi.org/10.1016/S0016-7061(01)00067-2, 2001.

Guerrero-Ramírez, N. R., Mommer, L., Freschet, G. T., Iversen, C. M., McCormack, M. L., Kattge, J., Poorter, H., van der Plas, F., Bergmann, J., Kuyper, T. W., York, L. M., Bruelheide, H., Laughlin, D. C., Meier, I. C., Roumet, C., Semchenko, M., Sweeney, C. J., van Ruijven, J., Valverde-Barrantes, O. J., Aubin, I., Catford, J. A., Manning, P., Martin, A., Milla, R., Minden, V., Pausas, J. G., Smith, S. W., Soudzilovskaia, N. A., Ammer, C., Butterfield, B., Craine, J., Cornelissen, J. H. C., de Vries, F. T., Isaac, M. E., Kramer, K., König, C., Lamb, E. G., Onipchenko, V. G., Peñuelas, J., Reich, P. B., Rillig, M. C., Sack, L., Shipley, B., Tedersoo, L., Valladares, F., van Bodegom, P., Weigelt, P., Wright, J. P., and Weigelt, A.: Global root traits (GRooT) database, Global Ecology and Biogeography, 30, 25–37, https://doi.org/10.1111/geb.13179, 2021.

Hengl, T., Jesus, J. M. de, Heuvelink, G. B. M., Gonzalez, M. R., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A., Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G. B., Ribeiro, E., Wheeler, I., Mantel, S., and Kempen, B.: SoilGrids250m: Global gridded soil information based on machine learning, PLOS ONE, 12, e0169748, https://doi.org/10.1371/journal.pone.0169748, 2017.

Hengl, T., Nussbaum, M., Wright, M. N., Heuvelink, G. B. M., and Gräler, B.: Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables, PeerJ, 6, e5518, https://doi.org/10.7717/peerj.5518, 2018.

Hiederer, R. and Köchy, M.: Global soil organic carbon estimates and the harmonized world soil database, Publications Office, LU, 2011.

Holland, E. A., Post, W. M., Matthews, E. G., Sulzman, J. M., Staufer, R., and Krankina, O. N.: A Global Database of Litterfall Mass and Litter Pool Carbon and Nutrients, 2.150694 MB, https://doi.org/10.3334/ORNLDAAC/1244, 2015.

Hugelius, G., Tarnocai, C., Broll, G., Canadell, J. G., Kuhry, P., and Swanson, D. K.: The Northern Circumpolar Soil Carbon Database: spatially distributed datasets of soil coverage and soil carbon storage in the northern permafrost regions, Earth System Science Data, 5, 3–13, https://doi.org/10.5194/essd-5-3-2013, 2013.

Jenks, G. F.: The Data Model Concept in Statistical Mapping, International Yearbook of Cartography, 7, 186–190, 1967.

Jia, B., Zhou, G., and Xu, Z.: Forest litterfall and its composition: a new data set of observational data from China, Ecology, 97, 1365–1365, https://doi.org/10.1890/15-1604.1, 2016.

Jiang, D., Zhuang, D., Fu, J., Huang, Y., and Wen, K.: Bioenergy potential from crop residues in China: Availability and distribution, Renewable and Sustainable Energy Reviews, 16, 1377–1382, https://doi.org/10.1016/j.rser.2011.12.012, 2012.

Liu, F., Wu, H., Zhao, Y., Li, D., Yang, J.-L., Song, X., Shi, Z., Zhu, A.-X., and Zhang, G.-L.: Mapping high resolution National Soil Information Grids of China, Science Bulletin, 67, 328–340, https://doi.org/10.1016/j.scib.2021.10.013, 2022.

Lokupitiya, E., Paustian, K., Easter, M., Williams, S., Andrén, O., and Kätterer, T.: Carbon balances in US croplands during the last two decades of the twentieth century, Biogeochemistry, 107, 207–225, https://doi.org/10.1007/s10533-010-9546-y, 2012.

Malone, B. P., McBratney, A. B., Minasny, B., and Laslett, G. M.: Mapping continuous depth functions of soil

carbon storage and available water capacity, Geoderma, 154, 138–152, https://doi.org/10.1016/j.geoderma.2009.10.007, 2009.

Monfreda, C., Ramankutty, N., and Foley, J. A.: Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000, Global Biogeochemical Cycles, 22, https://doi.org/10.1029/2007GB002947, 2008.

Oleson, K., Lawrence, D., Bonan, G., Drewniak, B., Huang, M., Koven, C., Levis, S., Li, F., Riley, W., Subin, Z., Swenson, S., Thornton, P., Bozbiyik, A., Fisher, R., Heald, C., Kluzek, E., Lamarque, J.-F., Lawrence, P., Leung, L., Lipscomb, W., Muszala, S., Ricciuto, D., Sacks, W., Sun, Y., Tang, J., and Yang, Z.-L.: Technical description of version 4.5 of the Community Land Model (CLM), [object Object], https://doi.org/10.5065/D6RR1W7M, 2013.

Poggio, L., de Sousa, L. M., Batjes, N. H., Heuvelink, G. B. M., Kempen, B., Ribeiro, E., and Rossiter, D.: SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty, SOIL, 7, 217–240, https://doi.org/10.5194/soil-7-217-2021, 2021.

Roman, P., Martinez, M. M., and Pantoja, A.: Farmer´s Compost Handbook: Experiences in Latin America, FAO, Rome, Italy, 112 pp., 2015.

Running, S. and Zhao, M.: MOD17A3HGF MODIS/Terra Net Primary Production Gap-Filled Yearly L4 Global 500 m SIN Grid V006, https://doi.org/10.5067/MODIS/MOD17A3HGF.006, 2019.

Scarlat, N., Martinov, M., and Dallemand, J.-F.: Assessment of the availability of agricultural crop residues in the European Union: Potential and limitations for bioenergy use, Waste Management, 30, 1889–1897, https://doi.org/10.1016/j.wasman.2010.04.016, 2010.

Schenk, H. J. and Jackson, R. B.: The Global Biogeography of Roots, Ecological Monographs, 72, 311–328, https://doi.org/10.1890/0012-9615(2002)072[0311:TGBOR]2.0.CO;2, 2002.

Spawn, S. A., Sullivan, C. C., Lark, T. J., and Gibbs, H. K.: Harmonized global maps of above and belowground biomass carbon density in the year 2010, Sci Data, 7, 112, https://doi.org/10.1038/s41597-020-0444-4, 2020.

Tumber-Dávila, S. J., Schenk, H. J., Du, E., and Jackson, R. B.: Plant sizes and shapes above and belowground and their interactions with climate, New Phytologist, nph.18031, https://doi.org/10.1111/nph.18031, 2022.

Wang, G., Luo, Z., Han, P., Chen, H., and Xu, J.: Critical carbon input to maintain current soil organic carbon stocks in global wheat systems, Sci Rep, 6, 19327, https://doi.org/10.1038/srep19327, 2016.

Wolf, J., West, T. O., Le Page, Y., Kyle, G. P., Zhang, X., Collatz, G. J., and Imhoff, M. L.: Biogenic carbon fluxes from global agricultural production and consumption, Global Biogeochemical Cycles, 29, 1617–1639, https://doi.org/10.1002/2015GB005119, 2015.

Zeng, X.: Global Vegetation Root Distribution for Land Modeling, Journal of Hydrometeorology, 2, 525–530, https://doi.org/10.1175/1525-7541(2001)002<0525:GVRDFL>2.0.CO;2, 2001.

Zhang, G., Wan, Q., Zhang, F., Wu, K., Cai, C., Zhang, M., Li, D., Zhao, Y., and Yang, J.: Criteria for establishment of soil family and soil series in Chinese Soil Taxonomy, 土壤学报英文版, 50, 826–834, https://doi.org/10.11766/trxb201303180124, 2013.

Zhang, Y., Lavallee, J. M., Robertson, A. D., Even, R., Ogle, S. M., Paustian, K., and Cotrufo, M. F.: Simulating measurable ecosystem carbon and nitrogen dynamics with the mechanistically defined MEMS 2.0 model, Biogeosciences, 18, 3147–3171, https://doi.org/10.5194/bg-18-3147-2021, 2021.