



# Global soil-climate-biome diagram: linking surface soil properties to climate and biota

Xia Zhao<sup>1</sup>, Yuanhe Yang<sup>1</sup>, Haihua Shen<sup>1</sup>, Xiaoqing Geng<sup>1</sup>, Jingyun Fang<sup>1,2\*</sup>

<sup>1</sup>State Key Laboratory of Vegetation and Environmental Change, Institute of Botany, Chinese Academy of Sciences, Beijing, China 100093

<sup>2</sup>College of Urban and Environmental Sciences, and Key Laboratory for Earth Surface Processes of the Ministry of Education, Peking University, Beijing, China 100871

*Correspondence to:* Jingyun Fang ([jyfang@urban.pku.edu.cn](mailto:jyfang@urban.pku.edu.cn))

**Abstract.** Surface soils interact strongly with both climate and biota and provide fundamental ecosystem services that maintain food, climate, and human security. However, the quantitative linkages between soil properties, climate, and biota at the global scale remain unclear. By compiling a comprehensive global soil database, we mapped eight major soil properties (bulk density; clay, silt, and sand fractions; soil pH; soil organic carbon [SOC] density; soil total nitrogen [STN] density; and soil C:N mass ratios) in the surface (0-30 cm) soil layer based on machine learning algorithms, and demonstrated the quantitative linkages between surface soil properties, climate, and biota at the global scale (i.e., global soil-climate-biome diagram). On the diagram, bulk density increased significantly with higher mean annual temperature (MAT) and lower mean annual precipitation (MAP); soil clay fraction increased significantly with higher MAT and MAP; Soil pH decreased with higher MAP and lower MAT, and the critical MAP for the transition from alkaline to acidic soil decreased with decreasing MAT; SOC density and STN density both were jointly affected by MAT and MAP, showing an increase at lower MAT and a saturation tendency towards higher MAP. Surface soil physical and chemical properties also showed remarkable variations across biomes. The soil-climate-biome diagram suggests the co-evolution of the soil, climate, and biota under global environmental change.

## 1. Introduction

As a critical component of the Earth system, soils influence all aspects of ecosystem processes and provide fundamental ecosystem services that maintain food, climate, and human security (Amundson et al., 2015; Milne et al., 2015; Adhikari and Hartemink, 2016). Soil physical properties, such as bulk density and soil texture, are important for green water retention and the preservation of carbon (C) and nutrients (Hassink, 1997; Sposito et al., 1999; Castellano and Kaye, 2009; Stockmann et al., 2013), whereas soil chemical properties, such as soil acidity (pH), organic C, and nutrient contents, are essential regulators of biogeochemical cycles and climate feedbacks (Davidson and Janssens, 2006; Chapin et al., 2009; Milne et al., 2015). As the most biogeochemically active soil layer, surface soil dominates the soil function and interacts strongly with climate and vegetation (Jenny, 1941; Weil and Brady, 2016). Therefore, assessing the physical and chemical properties in



surface soil could provide insights of global soil functions and support soil stewardship for societal well-being in the context of unprecedented pressure on soils (Batjes, 2009; Sanchez et al., 2009; Koch et al., 2013).

In the context of rapid environmental change, there is an increasing need for high-quality, high-resolution, and timely updated global mapping of soil properties (Grunwald et al., 2011). Based on global database of soil properties (e.g., the Harmonized World Soil Database [HWSD]), multiple linear regression models have been widely used for soil mapping (Batjes 2009; Hengl et al. 2014). Although recent progress has been made by compiling larger numbers of soil profiles and performing accuracy assessments, the corresponding maps of global soil properties are subject to weak relationships between soil properties and explanatory variables (Hengl et al., 2014). Moreover, some attempts have been made to predict global soil properties based on Earth system models, but these predictions frequently showed large variation among different models and agreed poorly with observational data (Todd-Brown et al., 2013; Tian et al., 2015). Recently, machine learning algorithms, such as random forest (RF) analyses have been successfully applied to develop spatially explicit estimates of soil organic C (SOC) (Grimm et al., 2008; Wiesmeier et al., 2011; Ding et al., 2016; Hengl et al., 2017). Compared with multiple linear regression models, RF analysis has several advantages, such as the ability to model non-linear relationships, handle both categorical and continuous predictors, and resist overfitting and noise features (Breiman, 2001). It is thus necessary to re-evaluate global patterns of soil properties using machine learning algorithms.

The underlying stability of soil systems is controlled by their inherent balance between mass inputs and losses of C and nutrients, which strongly feeds back on climate and biota (Weil and Brady, 2016; Amundson et al., 2015). By overlapping the global spatial distribution of climate types, biome types, and soil orders, Rohli et al. (2015) first quantified the percentage of global land surface that is covered by the combinations of climate types, biomes, and soil orders. However, appropriate linkages of soil properties, climate, and biota remain to be developed in a common diagram. In the context of significant progress in digital soil mapping techniques and the rapidly growing quantity of recorded soil information (Sanchez et al., 2009; Grunwald et al., 2011; Arrouays et al., 2014; Hengl et al., 2014; Shangguan et al., 2014), an opportunity is present to assess the quantitative linkages between soil properties, climate, and biota at the global scale.

In this study, we first compiled a global soil database (GSD, see Materials and Methods) that contains more than 28000 soil profiles for seven soil physical and chemical properties in surface soil layer (0-30 cm), including bulk density ( $\text{g cm}^{-3}$ ); sand, silt and clay fractions (%); soil pH; SOC density ( $\text{kg m}^{-2}$ ), and soil total nitrogen (STN) density ( $\text{kg m}^{-2}$ ). Using RF machine learning algorithms, we then established global soil maps for eight soil properties (the above mentioned seven soil properties plus C:N ratios, being estimated based on SOC density and STN density) at 1-km resolution and evaluated the corresponding uncertainties. The essential role of climate in shaping the spatial pattern of global biomes has been well demonstrated by Whittaker biome diagram (Whittaker, 1962). As climate and vegetation are two key soil-forming factors (Jenny, 1941), we further developed a global soil-climate-biome diagram by plotted each soil property on a climate basis within a modified



Whittaker biome diagram (Whittaker, 1962). Specifically, our objectives were to (i) map the physical and chemical properties of global surface soils, and (ii) determine the linkages between surface soil properties, climate and biota at the global scale.

## 2. Materials and Methods

### 5 2.1 Data set

Soil property data were compiled to establish a comprehensive database of worldwide soil profile information (Global soil database, GSD). Our GSD includes existing sources of soil profile data from the International Soil Reference and Information Centre-World Inventory of Soil Emission (ISRIC-WISE) Potential database (version 3.2; Batjes, 2009), soil reference profiles of Canada (Pan et al., 2011), the Land Resources of Russia/International Institute for Applied Systems  
10 Analysis (IIASA) (<http://nsidc.org/data/ggd601.html>), the International Soil Carbon Network (ISCN 2012, <http://www.fluxdata.org/nscn/Data/AccessData/SitePages/Carbonto1M.aspx>), the Soil Profile Analytical Database of Europe (SPADE), the Northern Circumpolar Soil Carbon Database (NCSCD, Tarnocai et al., 2009), the Second State Soil Survey of China (National Soil Survey Office, 1998), literature-retrieved soil data on the forests of China (Yang et al., 2014), field  
15 campaign data on the grasslands of northern China (from our research team; Yang et al., 2008, 2010), and field survey data of Australia (Wynn et al., 2006) (see Table S1 for more detailed information on these data sources).

The GSD compiled ground-truth soil data in the world and includes more than 28000 soil profiles (Fig. 1; Table S1). Although the total sample number and spatial distribution of the profile data are similar to those of the WISE30sec (Batjes, 2016), the GSD includes more specific soil data from China. Nonetheless, both databases include limited profiles for some  
20 regions of the world, notably Australia, Sahara and the northern territories of both Canada and Russia (Fig. 1). The GSD includes observed data on four soil physical properties (bulk density [ $\text{g cm}^{-3}$ ]; sand, silt and clay fractions [%]) and three chemical properties (soil pH; SOC density [ $\text{kg C m}^{-2}$ ]; and STN density [ $\text{kg N m}^{-2}$ ] in surface soil layer (see Table S2 and Fig. S1 for more detailed information). Moreover, the GSD contains pedologic information on the sampled soil profiles, mean annual temperature (MAT), mean annual precipitation (MAP), seasonality of air temperature (TS, calculated as  
25  $100 \times \text{SD}_{\text{monthly}} / \text{Mean}_{\text{monthly}}$ ) (Xu & Hutchinson, 2011), seasonality of precipitation (PS), mean annual normalized difference vegetation index (NDVI), elevation (global digital elevation map [DEM]), slope, and land use type for each recorded site (see Table S3 for more details). For sites with missing reports on climate or topographical data, profile coordinates were used to derive data at each site using a selection of GIS layers, from the WorldClim database for MAT and MAP and GTOP30 DEM-derived surfaces.

30



## 2.2 Region-specific random forest model

The random forest (RF) model is a data mining algorithm to make predictions based on an ensemble of randomized classification and regression trees (Breiman, 2001). We mapped soil properties based on a region-specific RF approach that yields spatially explicit estimates of each pixel (see Fig. S2 for more details on the workflow of this approach). To overcome spatial biases of the database (for example, heavy sampling in the USA), we divided the global land into 11 regions: Africa, Australia, Canada and Alaska, East Asia, Europe, Mexico, Russia, South America, tropical Asia, the USA, and West Asia (Table S2). In each region, we first constructed a RF model using the regional datasets of GSD and then used the model to estimate the spatial distribution of each soil property at a resolution of 1 km. Predictions were based on eight environmental variables, including MAT, MAP, TS, PS, vegetation cover conditions (NDVI), elevation, slope, and land use type (see Table S3 for more details on the data sources of each variable). Soil property data below 5% quantile and above 95% quantile were excluded as outliers and were not used for modeling. In addition, soil C:N ratio for each pixel was calculated based on predicted SOC and STN densities.

Because a large number of regression trees are constructed, one major advantage of RF model is that the risk of overfitting can be reduced. Another advantage is that the prediction depends on only three user-defined parameters: the number of trees (ntree), the minimum number of data points at each terminal node (nodesize), and the number of features sampled for splitting at each node (mtry). We used ntree = 1000 (default ntree = 500) in order to achieve more stable results (Grimm et al., 2008). For nodesize and mtry, we used the default set for RF regression. Also called a “black box” approach, one major disadvantage of RF model is that the relationships between the response and predictor variables cannot be interpreted individually for every RF tree. Nevertheless, the importance of each environmental variable was estimated by the mean change in prediction accuracy before and after permuting each variable.

## 2.3 Uncertainty analysis

In each region, we used 10-fold cross-validation to estimate the average mapping accuracy for each target soil property. The modelling accuracy for each bootstrap sample was evaluated by the amount of variation explained by the models ( $R^2$ ) and by the root mean square error (RMSE) calculated based on the observational and predicted soil property in the independent validation dataset (Table S4). Model uncertainties were assessed based on the bootstrap method. A robust estimate was derived by averaging the 10-fold cross-validation samples, and the uncertainty of the estimates was calculated as the standard deviation (SD) of the 10-fold cross-validation (Fig. S3).

## 2.4 Statistical analysis

Based on the results of the ensemble models, the spatial distribution of each soil property (bulk density, sand, silt, clay, pH, SOC density, STN density and C:N ratio) and their uncertainty were mapped at a resolution of 1 km. We also plotted each



soil property on a climate space illustrated by a modified Whittaker biome diagram. Moreover, we explored quantitative linkages between each soil property and climate parameters, including MAP, MAT, temperature seasonality and precipitation seasonality. We further compared these soil properties across main biomes, including tropical forest, temperate forest, boreal forests, tropical savannahs and grasslands, temperate grasslands and shrublands, tundra, permanent wetlands, deserts and croplands. All statistical analyses were performed using Matlab 2015a (The MathWorks Inc., Natick, MA, USA). Values are presented as mean  $\pm$  standard deviations, if not specially noted.

### 3. Results

#### 3.1 Global mapping of soil properties

Our results indicate that model predictions (Fig. S3) agreed well with the observed data across most regions, and that the ensemble models generally explained 30~60% of the variation in soil properties (Table S4). The eight soil properties showed great spatial heterogeneity across the globe in the upper 30-cm (Fig. 2). For instance, bulk density showed low values in the northern latitudes in the Eurasian continent, whereas high values occurred in the USA, North Africa, West Asia, and India (Fig. 2a). The clay fraction exhibited lower values at higher latitudes, whereas higher levels of sand fraction occurred at lower latitudes (Figs. 2b, 2c). The pH value of the surface soil was high (generally  $> 7.0$ ) in arid regions and it was relatively low (generally  $< 6.0$ ) in most forested regions (Fig. 2e).

The spatial patterns of SOC density and STN density were generally similar; both showed greater values at higher latitudes in the northern hemisphere and no consistent change with latitude in the southern hemisphere (Figs. 2f, 2g). Specifically, SOCD and STND both showed highest values in the northern high latitudes, while low values occurred in semiarid and desert regions. Soil C:N ratio showed the highest values at high latitudes in northern hemisphere, while it was the lowest values in arid regions in Northern Africa, West Asia and Southern Europe (Fig. 2h). On average, the global means of SOC density and STN density were  $6.94$  (SD=  $4.42$ )  $\text{kg C m}^{-2}$  and  $0.53$  (SD=  $0.23$ )  $\text{kg N m}^{-2}$  in surface soils, respectively (Table 1). In the surface soil layer, global stocks of SOC and STN were estimated to be  $797 \pm 4.1$  Pg C ( $10^{15}$  g, or billion tons) and  $64 \pm 0.4$  Pg N, respectively (Table 1).

#### 3.2 Global soil-climate-biome diagram

By putting data of surface soil properties on the Whittaker climate-biome diagram (Whittaker, 1962), we then documented the linkages between soil properties and climate across global biomes. We call this as the global soil-climate-biome diagram (Fig. 3). As showed in Fig. 3, bulk density generally decreased with lower MAT and higher MAP (Fig. 3a, also see Fig. 4a, 4b); sand fraction was inversely related to MAP and MAT (Fig. 3b; also Fig. 4c, 4d), whereas the clay fraction showed an opposite pattern (Fig. 3c, Fig. 4e, 4f); and soil pH increased with higher MAT in arid climate ( $\text{MAT} \leq 400$  mm) (Figs. 3e,



- 5a), while it decreased significantly with higher MAP both in cold ( $\text{MAT} \leq 10\text{ }^\circ\text{C}$ ) and warm ( $\text{MAT} > 10\text{ }^\circ\text{C}$ ) climates (Figs. 3e, 5b). The critical MAP for the transition from alkaline to acidic soil ( $\text{pH} = 7.0$ ) showed a non-linear increase with MAT and reached to a maximum of 400-500 mm when MAT exceeded  $10\text{ }^\circ\text{C}$  (Fig. 5c).
- 5 Generally, SOC density in the upper 30-cm soil layer decreased significantly with MAT at arid ( $\text{MAT} \leq 400\text{ mm}$ ) or humid climates ( $\text{MAT} > 400\text{ mm}$ ) (Figs. 3f), whereas it increased with MAP in accordance with a saturation curve, showing a higher saturation threshold ( $14.5$  vs.  $8.0\text{ kg C m}^{-2}$ ) in cold climates ( $\text{MAT} \leq 10\text{ }^\circ\text{C}$ ) compared with warm climates ( $\text{MAT} > 10\text{ }^\circ\text{C}$ ) (Figs. 6b). Similarly, STN density decreased significantly with MAT (Figs. 6c) and increased with MAP in accordance with a saturation curve, showing a higher saturation threshold ( $0.80$  vs.  $0.65\text{ kg N m}^{-2}$ ) in cold than warm climates (Figs. 6d). Combining the trends of SOC density and STN density, the C:N ratio of the upper 30-cm layer increased with MAT at a climate of  $\text{MAT} < 0\text{ }^\circ\text{C}$  and then decreased (Fig. 6e). In contrast, the C:N ratio increased with MAP in accordance with a saturation curve, showing a higher saturation threshold ( $18:1$  vs.  $12:1$ ) in cold climates compared with warm climates (Fig. 6f).
- 15 Soil properties showed different values across and within biomes throughout the world (Table 1; Fig. 3). Mean bulk density was lowest in the tundra and boreal forest but it was highest in the desert and tropical thorn scrub and woodland (Tables 1). Mean sand fraction was highest in boreal forest, whereas mean clay fraction was highest in the tropical rainforests (Tables 1). Soil pH was generally lower than 5.5 in tropical forest, boreal forest and tundra, but mean pH values could approach and even exceed 7.0 in dry biomes, such as the desert, grassland and savanna (Table 1). Moreover, means of SOC and STN densities both showed high values in boreal forest and tundra, but they were extremely low in the desert and tropical thorn scrub and woodland (Table 1). Overall, mean soil C:N ratio showed the highest values in tundra and boreal forest ( $> 15$ ), while it was lowest in desert, temperate shrubs and grasslands ( $\leq 10$ ) (Table 1; Figs. 3h).

## 4. Discussion

### 4.1 Global soil-climate-biome diagram for soil physical properties

- 25 The soil-climate-biome diagram demonstrates strong quantitative linkages between surface soil physical properties and climate variables at the global scale. Compared with variables associated with topography (e.g., elevation and slope), vegetation activity (i.e., NDVI) and land cover (i.e., land use type), climate variables (such as MAT, MAP, TS and PS) are stronger predictors of bulk density and soil texture (Fig. 7a-c). This is likely due to the essential effect of temperature and precipitation on physical, chemical and biological processes during soil formation (Weil and Brady, 2016). For instance, clay fraction increased significantly with MAT and MAP (Figs. 4) due to an enhancement of soil erosion, weathering and leaching processes (Jenny, 1941).
- 30



Other factors, such as historical tectonics, glaciations and soil ages, could also affect soil physical properties (Jenny, 1941; Weil and Brady, 2016), but they are often spatially correlated with current climate variables and biome distribution, making it difficult to separate their role from the latter. For instance, the effect of glaciations is stronger, the soil age is younger and air temperature is lower towards higher latitudes. Likewise, the role of tectonics in rejuvenating younger soils might also be mixed by corresponding climate conditions across altitudinal gradients. By exploring variations in soil physical properties with elevation in tropical regions, we found a significant decrease in bulk density and clay fraction (Fig. S4). These altitudinal gradients were consistent with the results of field studies (Dieleman et al., 2013) and also mirrored a similar trend across latitudes.

#### 4.2 Key role of climate in determining global patterns of surface soil pH

Our results indicated that MAP was the most important surrogate for soil pH prediction (Fig. 7d). Such a pattern may be due to the increased leaching of exchangeable base cations across large-scale precipitation gradients (Jenny et al., 1941). Interestingly, our further analysis showed that the critical levels of MAP for the transition from alkaline to acidic soil decreased non-linearly with lower MAT owing to changing water balance (Fig. 5). Specifically, the critical MAP ranged from 400-500 mm when the MAT exceeded 10 °C and could decrease to 50-100 mm when MAT was close to 0 °C, highlighting significant interactions between MAP and MAT. Such a pattern was supported by a recent study, which revealed that the transition from alkaline to acidic soil occurred when the MAP began to exceed the mean annual potential evapotranspiration (Slessarev et al., 2016). It should be noted that, other factors besides climate variables, such as acid deposition may also contribute to regional-scale patterns of soil pH, especially in Europe, eastern North America and southern China, where have received high-level acid deposition (Bouwman et al., 2002; Vet et al., 2014; Du et al., 2015).

#### 4.3 Climate as drivers of SOC and STN in global surface soils

Our results indicate an increase in SOC density with higher MAP and lower MAT (Fig. 6), which is in agreement with the pattern of previous estimates (Post et al., 1982; Gray et al., 2009). Such a pattern reflects the fact that soil C stock depends on the balance between plant inputs (e.g., litterfall and other plant debris) and microbially mediated metabolic losses of CO<sub>2</sub> to atmosphere (Stockmann et al., 2013), which were strongly controlled by climate (Davidson and Janssens, 2006; Bond-Lamberty and Thomson, 2010). Our analysis indicates that climate variables (e.g., MAT, MAP) are the strongest predictors of SOC density (Fig. 7e). Besides climatic variables, substrate quality (e.g., C:N ratio) is another important factor that affects SOC accumulation (Fierer et al., 2005). Our analysis indicates that higher SOC density is associated with higher soil C:N ratios (Figs. 2, 3). Across global terrestrial biomes, SOC density showed high values in boreal forests and tundra due to the slower decomposition compared with biomass inputs (Hobbie et al., 2000; Hashimoto et al., 2015; Bloom et al., 2016), but these values were extremely low in drylands due to low plant cover and productivity (Delgado-Baquerizo et al., 2013). This pattern can be partially attributed to the fact that soil substrate quality, as indicated by C:N ratio (Jensen et al., 2005; Craine et al., 2007), is more favourable for C storage in tundra and boreal forest (C:N ratio > 16:1) compared with desert and



temperate shrubs and grasslands (C:N ratio  $\leq 10:1$ ) (Table 1). As an overall effect of climate shifts and substrate quality, SOC accumulation increases poleward with longer C turnover time (Carvalhais et al., 2014). Similarly, we also found an increase in SOC density with elevation (Fig. S4f), likely due to a shift in climate regime and vegetation type.

5 Due to a prevailing N limitation in terrestrial ecosystems, the C and N cycles are tightly linked (LeBauer and Treseder, 2008; Gruber and Galloway, 2008; Chapin et al., 2009). Accordingly, our results indicate a strong correlation between STN density as SOC density (Fig. 8) and demonstrate a similar pattern of STN density as SOC density (Fig. 3). Further analysis also evidences an interaction between MAT and MAP in shaping the patterns of SOC density and STN density. For instance, SOC density and STN density both showed a tendency of saturation with higher MAP, while the saturation thresholds were  
10 higher under  $\text{MAT} \leq 10^\circ\text{C}$  compared with  $\text{MAT} > 10^\circ\text{C}$  (Fig. 6). Specifically, the saturation threshold for SOC density under  $\text{MAT} \leq 10^\circ\text{C}$  ( $14.5 \text{ kg C m}^{-2}$ ) were nearly twice of that under  $\text{MAT} > 10^\circ\text{C}$  ( $8.0 \text{ kg C m}^{-2}$ ) (Fig. 6b), while saturation threshold for STN density under  $\text{MAT} \leq 10^\circ\text{C}$  ( $0.80 \text{ kg N m}^{-2}$ ) were slightly (23%) higher than that under  $\text{MAT} > 10^\circ\text{C}$  ( $0.65 \text{ kg N m}^{-2}$ ) (Fig. 6d). These critical levels of SOCD and STND imply a saturation threshold of SOC and STN stocks  
15 indicate that the SOC pool has an upper limit with respect to C input levels because of a threshold of SOM stabilization efficiency (Stewart et al., 2008; Kimetu et al., 2009). These thresholds of soil C saturation can help to estimate soil C sequestration potential and provide important guidelines for regional soil steward and ecosystem management.

#### 4.4 Uncertainties in mapping soil properties at the global scale

In this study, we used machine learning algorithms to map global soil properties at a 1-km resolution. Although this  
20 approach could overcome uncertainties derived from large variations in mapping unit, several limitations still existed in our analysis. First, the limited sample size in certain area may lead to estimation uncertainties. Particularly, the accuracy of the region-specific RF model partially depends on the number of sampling sites and the evenness of the spatial pattern. The limited number and uneven distribution of the soil profile may thus constrain the accuracy of region-specified RF models, especially in regions such as Russia and South America (Fig. S3).

25 Second, the various approaches used to measure soil properties among regions and nations may also generate uncertainties in global-scale mapping. Specifically, soil properties were measured using various approaches and compiled for several decades, and no straightforward solutions exist for harmonizing the data at the global level (Maire et al., 2015; Batjes, 2016). Similar to other studies, the errors due to varied sampling and measurement methods across time may lead to uncertainties in  
30 our analysis and also hinder reliable hindcast and forecast estimates at the global scale (Grunwald et al., 2011). Therefore, future international actions are needed to survey soil properties using uniform sampling design and measurement methods to improve the current mapping of global soil properties.





Third, the uncertainties may also arise from the limited independent variables used in this study. Although essential surrogate variables of climate, topography, vegetation activity, and land cover (see method section) were incorporated in our analysis, we still could not account for the role of soil ages and parental material characteristics due to the lack of global-scale dataset. For instance, soil mineralogy, being a function of parent material, climate and soil age (Jenny, 1941), has been demonstrated to be important in determining the quantity of SOC storage and its turnover time during long-term soil development (Torn et al., 1997). Soil age may also play an important role in forming soil property (Jenny, 1941), but it is hard to evaluate its individual role in regulating spatial patterns of soil properties due to its strong interactions with climate variables. Therefore, future studies should make more efforts to consider these variables when predicting spatial patterns of soil physical and chemical properties at the global scale.

## 10 5. Conclusion

By compiling a comprehensive global soil database, we mapped eight soil properties based on machine learning algorithms and assessed the quantitative linkages between soil properties, climate, and biota at the global scale. Our region-specific random forest model generated reliable predictions, and can thus provide useful inputs to regional and global biogeochemical models. Our results also produced a global soil-climate-biome diagram, which improved the understanding of the strong correspondence between soil, climate, and biota. Given that significant changes in major soil properties may have occurred and will continue due to global environmental change (Trumbore and Czimczik, 2008; Chapin et al., 2009; Todd-Brown et al., 2013; Luo et al., 2016, 2017), more efforts should be made to understand the co-evolution of soil, climate and biota in view of the global soil-climate-biome diagram.

## Acknowledgement

20 This work is supported by National Key Research and Development Program (2017YFC0503901), the National Natural Science Foundation of China (31321061 and 31330012) and Key Research Program of Frontier Sciences, CAS (QYZDY-SSW-SMC011). All data will be available online as a supplement when this paper is published.

## References

- Adhikari, K., and Hartemink, A. E.: Linking soils to ecosystem services—A global review. *Geoderma*, 262, 101–111, 2016.
- 25 Amundson, R.: The carbon budget in soils. *Annu. Rev. Earth Pl. Sc.*, 29(1), 535–562, 2001.
- Amundson, R., Berhe, A. A., Hopmans, J. W., Olson, C., Sztein, A. E., and Sparks, D. L.: Soil and human security in the 21<sup>st</sup> century. *Science*, 348(6235), 1261071, 2015.



- Arrouays, D., McKenzie, N., Hempel, J., de Forges, A. R., and McBratney, A. B. (Eds.): GlobalSoilMap: basis of the global spatial soil information system. CRC Press, 2014.
- Batjes, N. H.: Harmonized soil profile data for applications at global and continental scales: updates to the WISE database. *Soil Use and Manage.*, 25(2), 124–127, 2009.
- 5 Batjes, N. H.: Harmonized soil property values for broad-scale modelling (WISE30sec) with estimates of global soil carbon stocks. *Geoderma*, 269, 61–68, 2016.
- Bloom, A. A., Exbrayat, J. F., van der Velde, I. R., Feng, L., and Williams, M.: The decadal state of the terrestrial carbon cycle: Global retrievals of terrestrial carbon allocation, pools, and residence times. *P. Natl. Acad. Sci. U.S.A.*, 113(5), 1285–1290, 2016.
- 10 Bond-Lamberty, B., and Thomson, A.: Temperature-associated increases in the global soil respiration record. *Nature*, 464(7288), 579–582, 2010.
- Bouwman, A. F., Van Vuuren, D. P., Derwent, R. G., and Posch, M.: A global analysis of acidification and eutrophication of terrestrial ecosystems. *Water Air Soil Poll.*, 141(1), 349–382, 2002.
- Breiman, L.: Random forests. *Mach. Learn.*, 45(1), 5–32, 2001.
- 15 Carvalhais, N., et al.: Global covariation of carbon turnover times with climate in terrestrial ecosystems. *Nature*, 514, 213–217, 2014.
- Castellano, M. J., and Kaye, J. P.: Global within-site variance in soil solution nitrogen and hydraulic conductivity are correlated with clay content. *Ecosystems*, 12(8), 1343–1351, 2009.
- Chapin III, S., McFarland, J., McGuire, D. A., Euskirchen, E. S., Ruess, R. W., and Kielland, K.: The changing global carbon cycle: linking plant–soil carbon dynamics to global consequences. *J. Ecol.*, 97(5), 840–850, 2009.
- 20 Craine, J. M., Morrow, C., and Fierer, N.: Microbial nitrogen limitation increases decomposition. *Ecology*, 88(8), 2105–2113, 2007.
- Davidson, E. A., and Janssens, I. A.: Temperature sensitivity of soil carbon decomposition and feedbacks to climate change. *Nature*, 440(7081), 165, 2006.
- 25 Delgado-Baquerizo, M., Maestre, F. T., Gallardo, A., Bowker, M. A., Wallenstein, M. D., Quero, J. L., and García-Palacios, P.: Decoupling of soil nutrient cycles as a function of aridity in global drylands. *Nature*, 502(7473), 672–676, 2013.
- Dieleman, W. I., Venter, M., Ramachandra, A., Krockenberger, A. K., and Bird, M. I.: Soil carbon stocks vary predictably with altitude in tropical forests: implications for soil carbon storage. *Geoderma*, 204, 59–67, 2013.
- Ding, J., Li, F., Yang, G., Chen, L., Zhang, B., Liu, L., and Ji, C.: The permafrost carbon inventory on the Tibetan Plateau: a new evaluation using deep sediment cores. *Global Change Biol.*, 22(8), 2688–2701, 2016.
- 30 Du, E., De Vries, W., Liu, X., Fang, J., Galloway, J. N., and Jiang, Y.: Spatial boundary of urban ‘acid islands’ in southern China. *Scientific reports*, 5, 12625, 2015.
- Fierer, N., Craine, J. M., McLauchlan, K., and Schimel, J. P.: Litter quality and the temperature sensitivity of decomposition. *Ecology*, 86(2), 320–326, 2005.



- Gray, J. M., Humphreys, G. S., and Deckers, J. A.: Relationships in soil distribution as revealed by a global soil database. *Geoderma*, 150(3), 309–323, 2009.
- Grimm, R., Behrens, T., Märker, M., and Elsenbeer, H.: Soil organic carbon concentrations and stocks on Barro Colorado Island—Digital soil mapping using Random Forests analysis. *Geoderma*, 146(1), 102–113, 2008.
- 5 Gruber, N., and Galloway, J. N.: An Earth-system perspective of the global nitrogen cycle. *Nature*, 451(7176), 293–296, 2008.
- Grunwald, S., Thompson, J. A., and Boettinger, J. L.: Digital soil mapping and modeling at continental scales: Finding solutions for global issues. *Soil Sci. Soc. Am. J.*, 75(4), 1201–1213, 2011.
- Hashimoto, S., Carvalhais, N., Ito, A., Migliavacca, M., Nishina, K., and Reichstein, M.: Global spatiotemporal distribution of soil respiration modeled using a global database. *Biogeosciences*, 12, 4121–4132, 2015.
- 10 Hassink, J.: The capacity of soils to preserve organic C and N by their association with clay and silt particles. *Plant Soil*, 191(1), 77–87, 1997.
- Hengl, T., de Jesus, J. M., MacMillan, R. A., Batjes, N. H., Heuvelink, G. B., Ribeiro, E., et al.: SoilGrids1km—global soil information based on automated mapping. *PLoS One*, 9(8), e105992, 2014.
- 15 Hengl, T., Heuvelink, G. B., Kempen, B., Leenaars, J. G., Walsh, M. G., Shepherd, K. D., and Tondoh, J. E.: Mapping soil properties of Africa at 250 m resolution: random forests significantly improve current predictions. *PloS One*, 10(6), e0125814, 2015.
- Hengl, T., de Jesus, J. M., Heuvelink, G. B., Gonzalez, M. R., Kilibarda, M., Blagotić, A., and Guevara, M. A.: SoilGrids250m: Global gridded soil information based on machine learning. *PloS One*, 12(2), e0169748, 2017.
- 20 Hobbie, S. E., Schimel, J. P., Trumbore, S. E., and Randerson, J. R.: Controls over carbon storage and turnover in high - latitude soils. *Global Change Biol.*, 6(S1), 196–210, 2000.
- Jenny, H.: *Factors of soil formation: a system of quantitative pedology*. McGraw-Hill Book Company Inc, New York, 1941.
- Jensen, L. S., Salo, T., Palmason, F., Breland, T. A., Henriksen, T. M., Stenberg, B., and Esala, M.: Influence of biochemical quality on C and N mineralisation from a broad variety of plant materials in soil. *Plant Soil*, 273(1-2), 307–326, 2005.
- 25 Kimetu, J. M., Lehmann, J., Kinyangi, J. M., Cheng, C. H., Thies, J., Mugendi, D. N., and Pell, A.: Soil organic C stabilization and thresholds in C saturation. *Soil Biol. Biochem.*, 41(10), 2100–2104, 2009.
- Koch, A., McBratney, A., Adams, M., Field, D., Hill, R., Crawford, J., and Angers, D.: Soil security: solving the global soil crisis. *Global Policy*, 4(4), 434–441, 2013.
- Kopittke, P. M., Dalal, R. C., Finn, D., and Menzies, N. W.: Global changes in soil stocks of carbon, nitrogen, phosphorus, and sulphur as influenced by long - term agricultural production. *Global Change Biol.*, 23(6), 2509–2519, 2017.
- 30 LeBauer, D. S., and Treseder, K. K.: Nitrogen limitation of net primary productivity in terrestrial ecosystems is globally distributed. *Ecology*, 89(2), 371–379.–8040, 2008.
- Luo, Y., Ahlström, A., Allison, S. D., Batjes, N. H., Brovkin, V., Carvalhais, N., et al.: Toward more realistic projections of soil carbon dynamics by Earth system models. *Global Biogeochem. Cy.*, 30(1), 40–56, 2016.



- Luo, Z., Wang, E., and Sun, O.J.: Uncertain future soil carbon dynamics under global change predicted by models constrained by total carbon measurements. *Ecol. Appl.*, 27(3), 1001–1009, 2017.
- Maire, V., Wright, I. J., Prentice, I. C., Batjes, N. H., Bhaskar, R., Bodegom, P. M., and Reich, P. B.: Global effects of soil and climate on leaf photosynthetic traits and rates. *Global Ecol. Biogeogr.*, 24(6), 706–717, 2015.
- 5 Milne, E., Banwart, S. A., Noellemeyer, E., Abson, D. J., Ballabio, C., Bampa, F., and Black, H.: Soil carbon, multiple benefits. *Environ. Dev.*, 13, 33–38, 2015.
- Pan, Y., Birdsey, R. A., Fang, J., Houghton, R., Kauppi, P. E., Kurz, W. A. and Ciais, P.: A large and persistent carbon sink in the world's forests. *Science*, 333(6045), 988–993, 2011.
- Post, W. M., Emanuel, W. R., Zinke, P. J. and Stangenberger, A. G.: Soil carbon pools and world life zones. *Nature*,  
10 298(5870), 156–159, 1982.
- Post, W. M., Pastor, J., Zinke, P. J., and Stangenberger, A.G.: Global patterns of soil nitrogen storage. *Nature*, 317(6038), 613–616, 1985.
- Weil, R. R., and Brady, N. C.: *The nature and properties of soils* (15<sup>th</sup> Edition). Pearson Education, 2016.
- Rohli, R. V., Joyner, T. A., Reynolds, S. J., and Ballinger, T. J.: Overlap of global Köppen–Geiger climates, biomes, and soil  
15 orders. *Phys. Geogr.*, 36(2), 158–175, 2015.
- Sanchez, P. A., Ahamed, S., Carré F., Hartemink, A. E., Hempel, J., Huising, J. and Minasny, B.: Digital soil map of the world. *Science*, 325(5941), 680–681, 2009.
- Shangguan, W., Dai, Y., Duan, Q., Liu, B., and Yuan, H.: A global soil data set for earth system modeling. *J. Adv. Model Earth Sy.*, 6(1), 249–263, 2014.
- 20 Slessarev, E. W., Lin, Y., Bingham, N. L., Johnson, J. E., Dai, Y., Schimel, J. P., and Chadwick, O. A.: Water balance creates a threshold in soil pH at the global scale. *Nature*, 540(7634), 567–569, 2016.
- Sposito, G., Skipper, N. T., Sutton, R., Park, S. H., Soper, A. K., and Greathouse, J. A.: Surface geochemistry of the clay minerals. *P. Natl. Acad. Sci. U.S.A.*, 96(7), 3358–3364, 1999.
- Stewart, C. E., Paustian, K., Conant, R. T., Plante, A. F., and Six, J.: Soil carbon saturation: concept, evidence and evaluation.  
25 *Biogeochemistry*, 86(1), 19–31, 2007.
- Stewart, C. E., Paustian, K., Conant, R. T., Plante, A. F., Six, J.: Soil carbon saturation: evaluation and corroboration by long-term incubations. *Soil Biol. Biochem.*, 40(7), 1741–1750, 2008.
- Stockmann, U., Adams, M. A., Crawford, J. W., Field, D. J., Henakaarchchi, N., Jenkins, M., and Wheeler, I. (2013). The knowns, known unknowns and unknowns of sequestration of soil organic carbon. *Agr., Ecosyst. Environ.*, 164, 80–99, 2013.
- 30 Tarnocai, C., Canadell, J. G., Schuur, E. A. G., Kuhry, P., Mazhitova, G., and Zimov, S.: Soil organic carbon pools in the northern circumpolar permafrost region. *Global Biogeochem. Cy.*, 23, GB2023, 2009.
- Tian, H., Lu, C., Yang, J., Banger, K., Huntzinger, D. N., Schwalm, C. R., et al.: Global patterns and controls of soil organic carbon dynamics as simulated by multiple terrestrial biosphere models: Current status and future directions. *Global Biogeochem. Cy.*, 29(6), 775–792, 2015.



- Todd-Brown, K. E., Randerson, J. T., Post, W. M., Hoffman, F. M., Tarnocai, C., Schuur, E. A., and Allison, S. D.: Causes of variation in soil carbon simulations from CMIP5 Earth system models and comparison with observations. *Biogeosciences*, 10(3), 1717-1736, 2013.
- Torn, M. S., Trumbore, S. E., Chadwick, O. A., Vitousek, P. M., and Hendricks, D. M.: Mineral control of soil organic carbon storage and turnover. *Nature*, 389(6647), 170–173, 1997.
- Trumbore, S. E., and Czimczik, C. I.: An uncertain future for soil carbon. *Science*, 321, 1455-1456, 2008.
- Vet, R., Artz, R. S., Carou, S., Shaw, M., Ro, C. U., Aas, W., and Hou, A.: A global assessment of precipitation chemistry and deposition of sulfur, nitrogen, sea salt, base cations, organic acids, acidity and pH, and phosphorus. *Atmos. Environ.*, 93, 3–100, 2014.
- Weil, R. R., and Brady, N. C.: *The nature and properties of soils*. Pearson Education, USA, 2016.
- Whittaker, R. H.: Classification of natural communities. *Bot. Rev.*, 28(1), 1–239, 1962.
- Wiesmeier, M., Barthold, F., Blank, B., and Kögel-Knabner, I.: Digital mapping of soil organic matter stocks using Random Forest modeling in a semi-arid steppe ecosystem. *Plant Soil*, 340(1-2), 7–24, 2011.
- Wynn, J. G., Bird, M. I., Vellen, L., Grand - Clement, E., Carter, J., and Berry, S. L.: Continental - scale measurement of the soil organic carbon pool with climatic, edaphic, and biotic controls. *Global Biogeochem. Cy.*, 20, GB1007, doi:10.1029/2005GB002576, 2006.
- Xu, T., and Hutchinson, M.: ANUCLIM version 6.1 user guide. The Australian National University, Fenner School of Environment and Society, Canberra, Australia, 2001.
- Yang, Y., Fang, J., Tang, Y., Ji, C., Zheng, C., He, J., and Zhu, B.: Storage, patterns and controls of soil organic carbon in the Tibetan grasslands. *Global Change Biol.*, 14(7), 1592–1599, 2008.
- Yang, Y., Fang, J., Ma, W., Smith, P., Mohammat, A., Wang, S., and Wang, W. E. I.: Soil carbon stock and its changes in northern China's grasslands from 1980s to 2000s. *Global Change Biol.*, 16(11), 3036–3047, 2010.
- Yang, Y., Li, P., Ding, J., Zhao, X., Ma, W., Ji, C., and Fang, J.: Increased topsoil carbon stock across China's forests. *Global Change Biol.*, 20(8), 2687–2696, 2014.

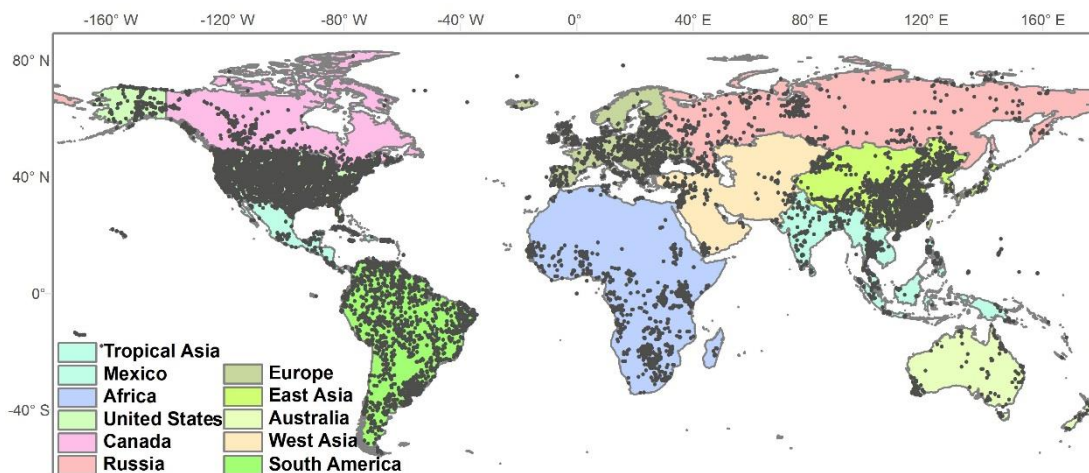


Figure 1. Global distribution of 28222 soil profiles included in the global soil database (GSD).

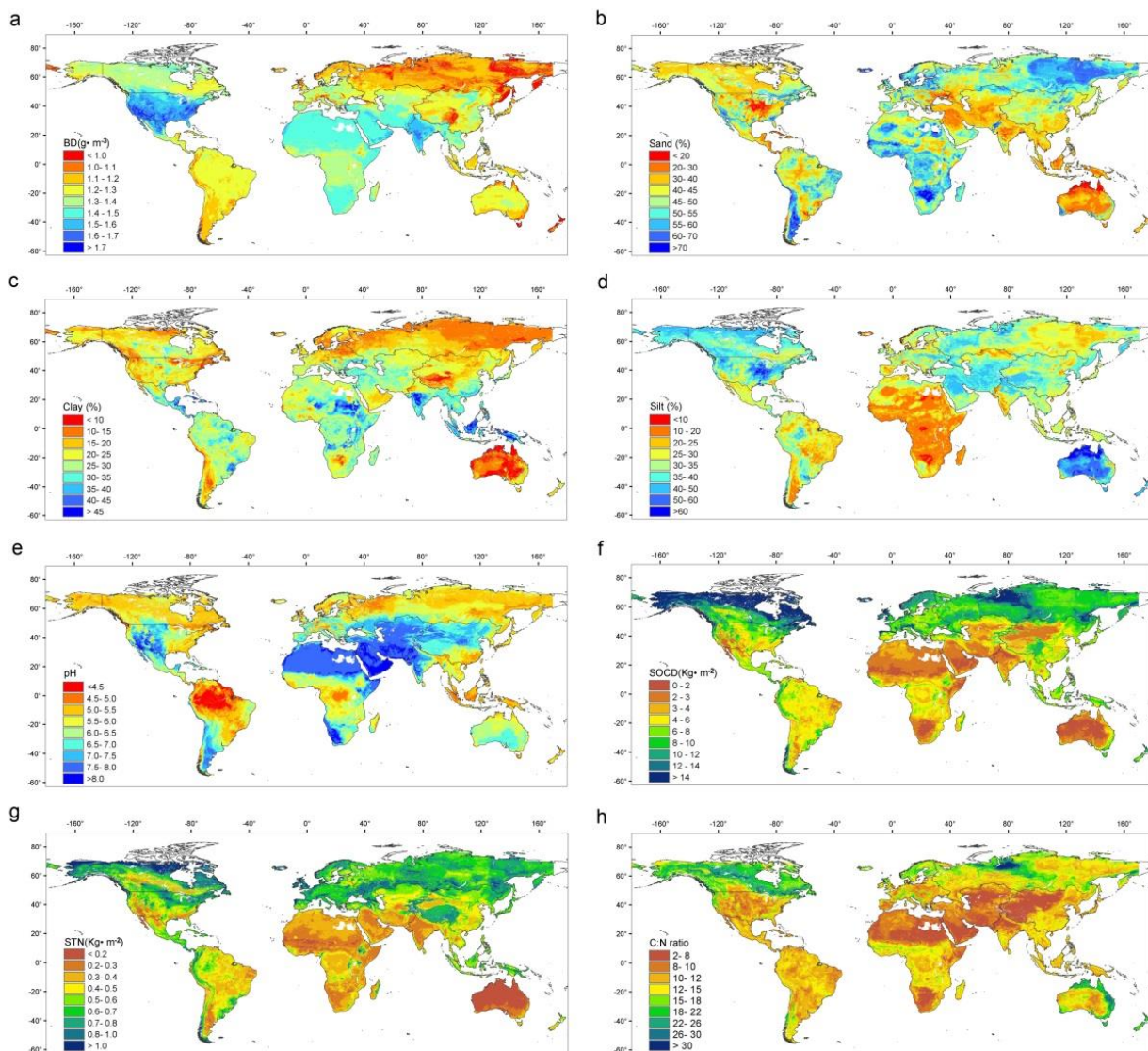


Figure 2. Map of worldwide soil properties in the upper 30-cm soil layer. a, BD (bulk density,  $\text{g m}^{-3}$ ); b, Sand fraction (%); c, Silt fraction (%); d, Clay fraction (%); e, pH; f, SOCD (soil organic carbon density,  $\text{kg m}^{-2}$ ); g, STND (soil total nitrogen density,  $\text{kg m}^{-2}$ ); and h, C:N ratio.

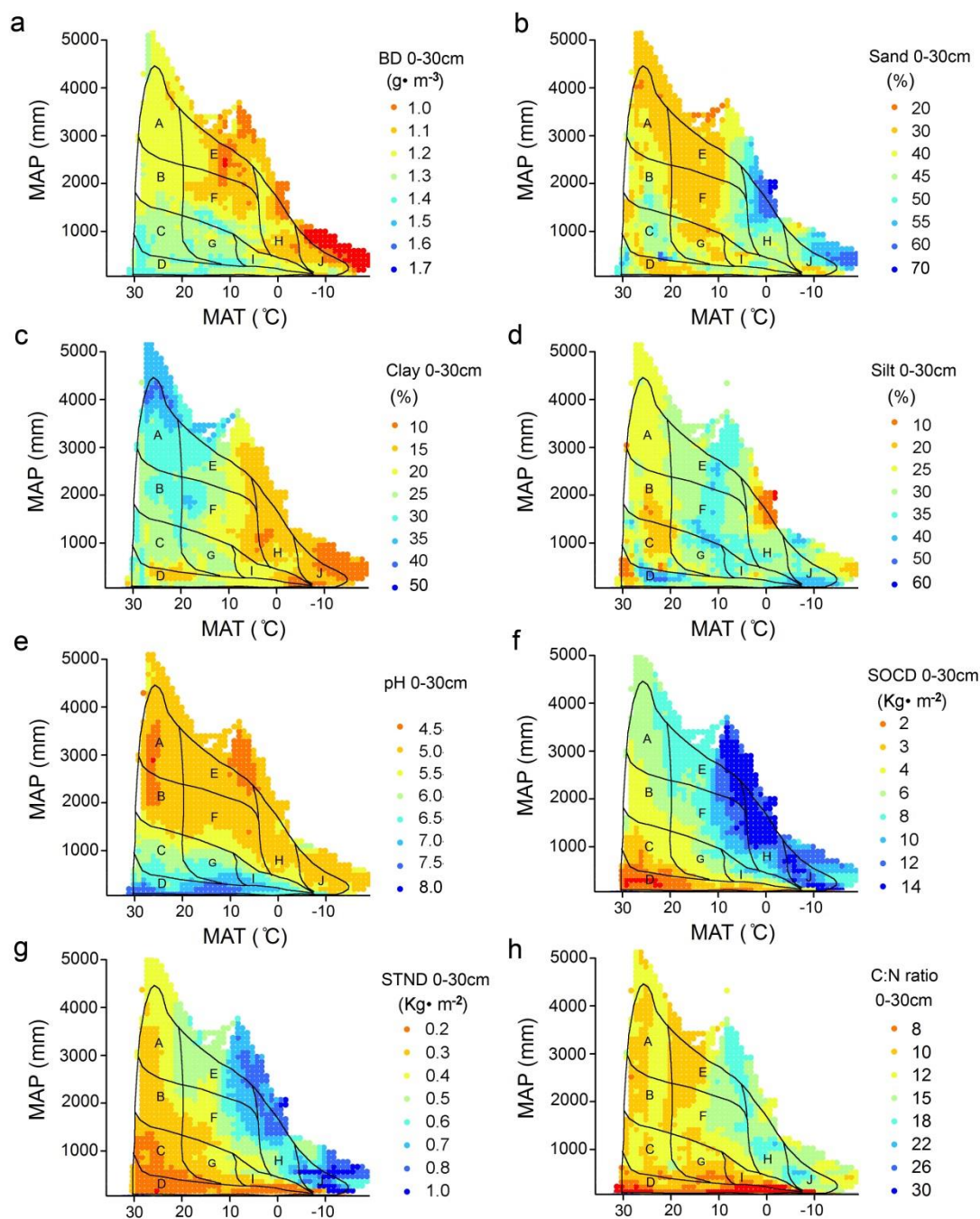


Figure 3. Changes in soil properties of the upper 30-cm layer on the Whittaker biome diagram. Each square shows the average C density within each 1 °C of MAT and 100 mm of MAP. Each biome type in the modified Whittaker biome diagram is indicated by a capital letter. A, Tropical rainforest; B, Tropical seasonal forest; C, Tropical thorn scrub and woodland; D, Desert; E, Temperate rainforest; F, Temperate forest; G, Savanna; H, Boreal forest; I, Grassland; and J, Tundra.



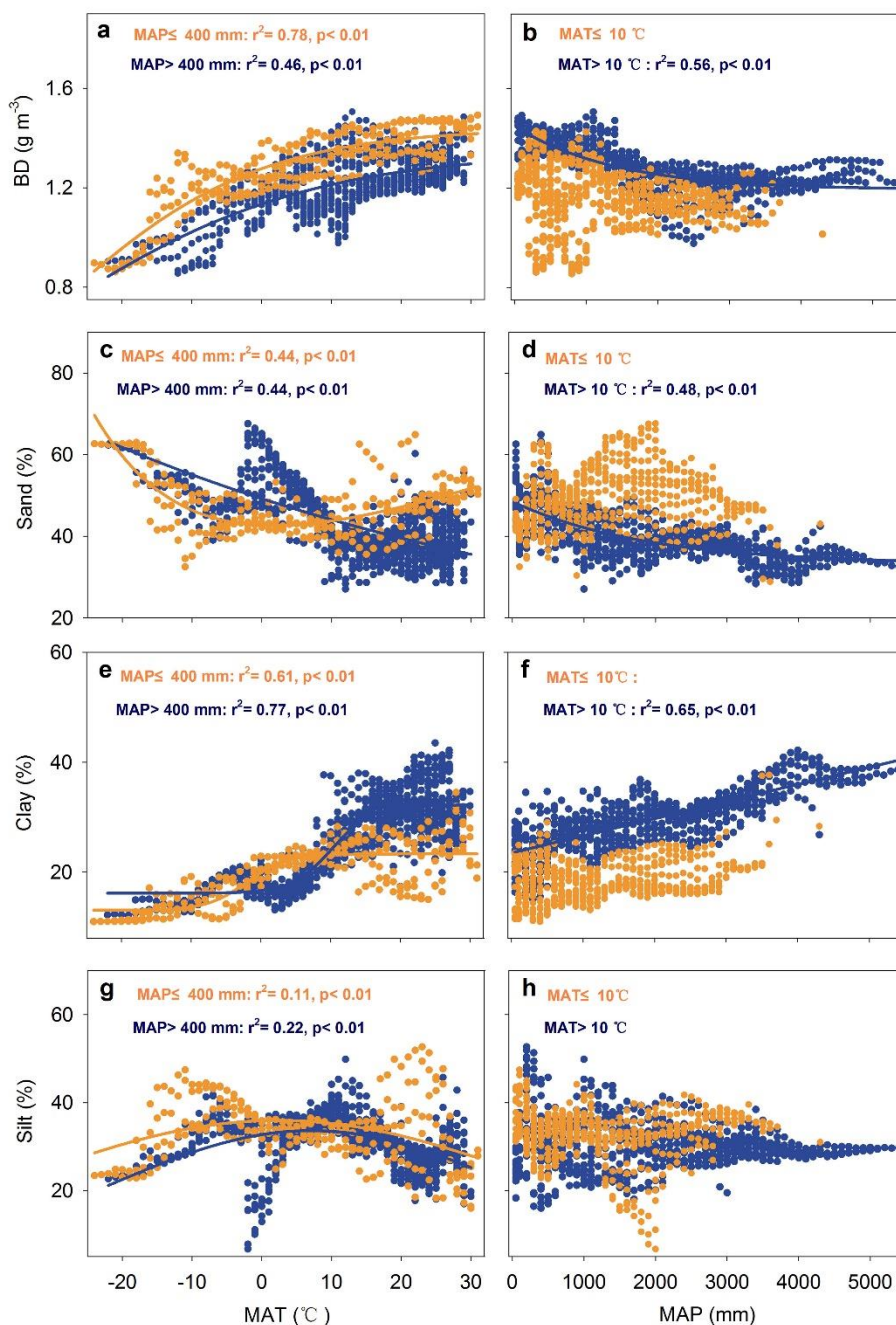


Figure 4. Changes in upper 30-cm soil bulk density (BD,  $\text{g m}^{-3}$ ), sand fraction (%), clay fraction (%) and silt fraction (%) with mean annual precipitation (MAP) and mean annual temperature (MAT). We used 400 mm of MAP as a threshold of transition from arid to humid climate, and 10  $^{\circ}\text{C}$  of MAT as a threshold of transition from cool to warm climate.

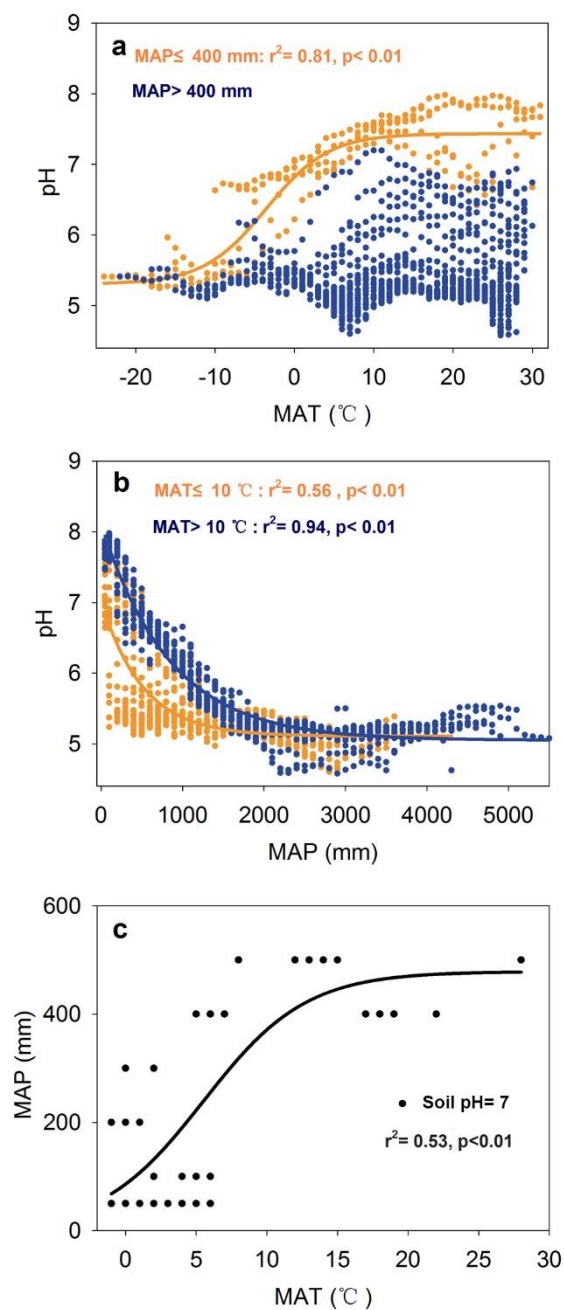


Figure 5. Changes in soil pH with climate. a, mean annual temperature (MAT); b, mean annual precipitation (MAP); and c, changes in critical levels of MAP at soil pH=7.0 with MAT. We used MAP of 400 mm as a threshold of transition from arid to humid climate, and MAT of 10 °C as a threshold of transition from cool to warm climate.

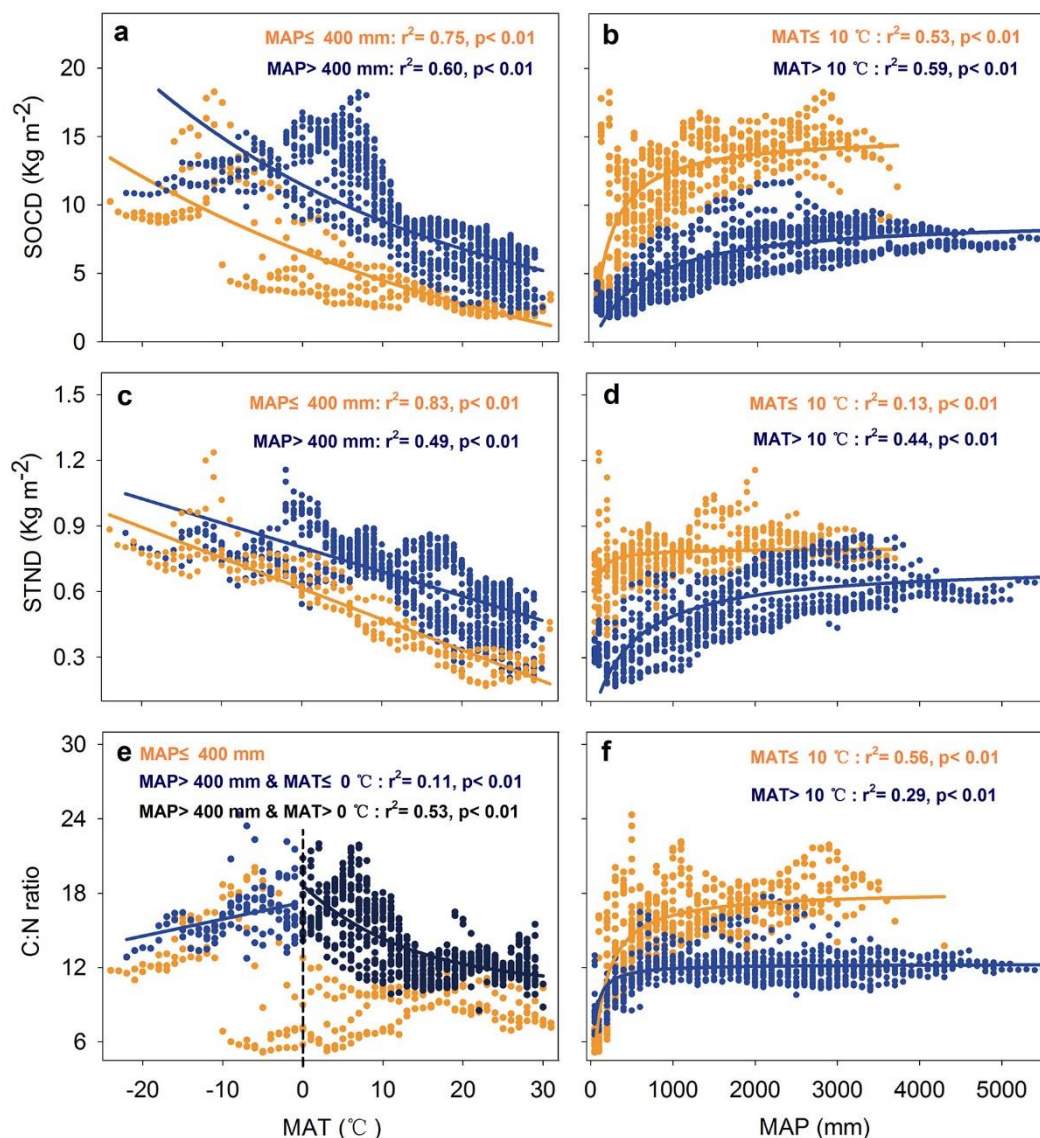


Figure 6. Changes in upper 30-cm soil organic carbon density (SOCD), soil total nitrogen density (STND) and C:N ratios with mean annual precipitation (MAP) and mean annual temperature (MAT). We used MAP of 400 mm as a threshold of transition from arid to humid climate, and MAT of 10  $^{\circ}\text{C}$  as a threshold of transition from cool to warm climate. (a) and (b),  
 5 SOCD; (c) and (d), STND; and (e) and (f), C:N ratios. Each dot shows the average value within each 1  $^{\circ}\text{C}$  MAT and 100 mm MAP.

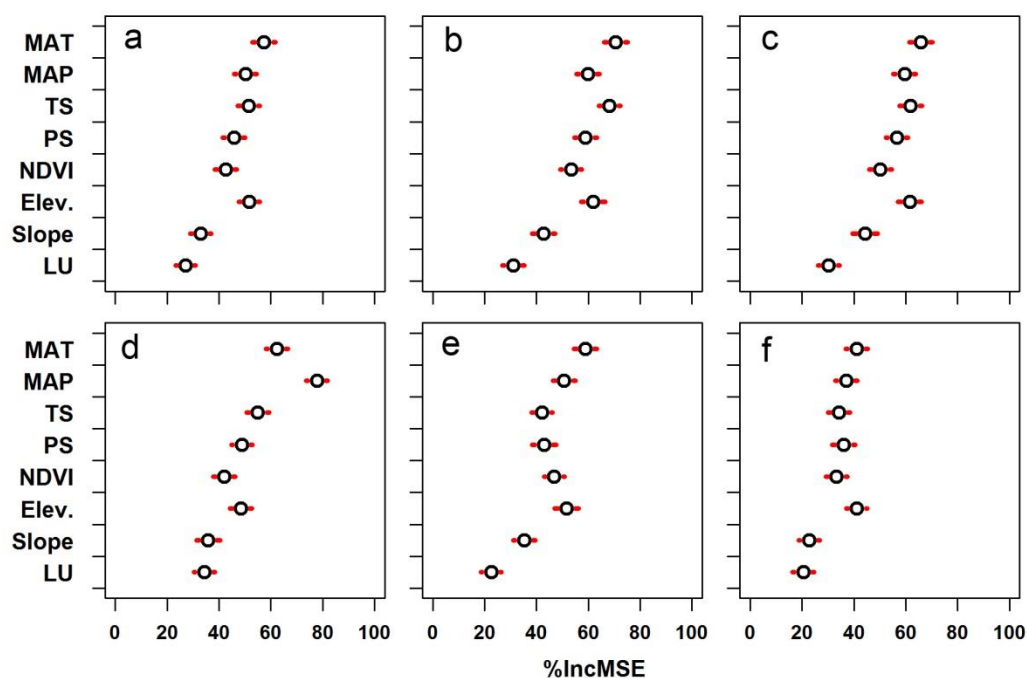


Figure 7. Importance of variables, denoted by the percent increase in mean-squared error (%IncMSE), for each soil property estimation RF model constructed from the training dataset in the top 30-cm layer. a, BD (bulk density,  $\text{g m}^{-3}$ ); b, Sand fraction (%); c, Clay fraction (%); d, pH; e, SOCD (soil organic carbon density,  $\text{kg m}^{-2}$ ); f, STND (soil total nitrogen density,  $\text{kg m}^{-2}$ ). MAT, MAP, TS, PS, Elev. and LU indicate mean annual temperature, mean annual precipitation, annual temperature seasonality, annual precipitation seasonality, elevation and land use type, respectively.

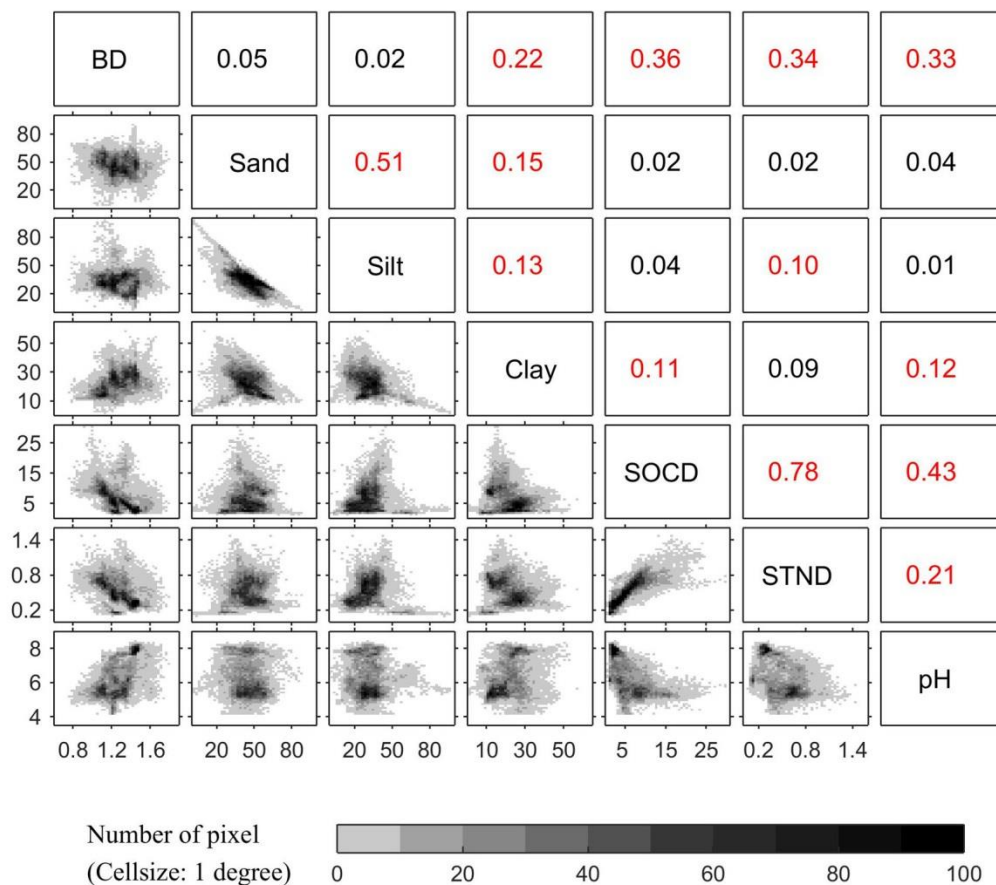


Figure 8. Correlation between surface soil properties.  $R^2$  between each two soil properties is shown in the upper plots with red color indicating  $R^2 > 0.1$ . BD, SOCD, and STND indicate bulk density, soil organic carbon density, and soil total nitrogen density, respectively.

**Table 1.** Mean values of surface (0-30 cm) soil properties by biome of the world.

Biome	Area (10 <sup>6</sup> ha)	Bulk density (g·m <sup>-3</sup> )	Sand (%)	Silt (%)	Clay (%)	pH	SOC (kg·m <sup>-2</sup> )	STND (kg·m <sup>-2</sup> )	C:N ratio	SOC stock (Pg)	STN stock (Pg)
TroF	1877	1.27±0.10	42.42±9.37	27.05±7.25	30.53±5.97	5.30±0.74	5.67±1.62	0.48±0.13	11.79±1.53	107±0.56	9±0.05
TemF	992	1.28±0.21	45.43±8.94	34.97±6.89	19.59±5.65	5.80±0.80	8.82±3.46	0.64±0.18	13.99±4.07	87±0.47	6±0.04
BF	1435	1.16±0.12	50.01±8.77	32.77±6.03	17.22±4.52	5.36±0.31	11.11±3.56	0.66±0.15	17.07±4.98	159±1.94	10±0.21
TSG	1915	1.34±0.10	47.92±13.06	25.64±16.68	26.44±9.05	6.25±0.80	3.82±1.65	0.32±0.13	12.28±4.05	73±0.62	6±0.04
TGS	1148	1.30±0.16	45.14±11.13	34.71±10.68	20.15±6.73	6.97±0.58	5.36±2.45	0.55±0.20	10.19±3.27	62±0.41	6±0.05
Deserts	2674	1.40±0.12	43.28±10.66	33.67±12.89	23.05±6.80	7.45±0.63	2.78±1.09	0.34±0.12	8.64±2.47	74±1.19	9±0.10
Tundra	644	1.16±0.16	47.57±8.92	37.02±7.35	15.41±3.12	5.44±0.34	13.78±4.51	0.81±0.19	17.18±5.06	89±2.07	5±0.10
Croplands	1984	1.34±0.15	40.70±11.08	33.13±9.87	26.16±7.04	6.40±0.74	6.54±2.80	0.58±0.22	11.36±2.19	130±0.50	11±0.09
PW	159	1.23±0.12	41.24±8.74	33.45±7.59	25.31±6.16	5.83±0.53	9.77±4.16	0.72±0.23	13.62±4.11	16±0.16	1±0.04
Total	12829	1.29±0.16	45.20±10.84	32.17±10.97	22.63±7.89	6.18±1.01	6.94±4.42	0.53±0.23	12.63±4.68	797±4.10	64±0.41

Notes: We include croplands and permanent wetlands in this table, although they are not single biomes. Abbriations: TroF, Tropical forests; TemF, Temperate forests; BF, Boreal forests; TSG, Tropical savannahs and grasslands; TGS, Temperate grasslands and shrublands; PW, Permanent wetlands. Spatial variability of soil properties within each biome was estimated as standard deviations. Uncertainties of total SOC and STN stocks were estimated as standard deviations based on 10-fold cross-validation.