Global patterns and drivers of soil dissolved organic carbon concentrations

Tianjing Ren[1,2] and Andong Cai[1]

[1]Institute of Environment and Sustainable Development in Agriculture, Chinese Academy of Agricultural Sciences, Beijing, 100081, China

[2]Institute of Soil Science and Plant Cultivation-State Research Institute, Department Soil Science and Environmental Analyses, Czartoryskich St. 8, 24–100 Puławy, Poland

***Corresponding author:** caiandong@caas.cn

Tel: +86-10-82106022, Fax: +86-10-8210622

Abstract

Dissolved organic carbon (DOC) constitutes the most active carbon pool in soils and plays critical roles in soil carbon cycling, plant productivity, and global climate change. Accurately assessing soil DOC quantity is essential to elucidate ecosystem functions and services. However, global driving factors and the spatial distribution of soil DOC remain poorly quantified, largely due to limited large-scale data. Here, we compile a comprehensive global database of soil DOC concentrations, encompassing 12,807 observations extracted from 975 scientific publications published between 1984 and 2020. We also record detailed geographic locations, climatic variables, and soil properties as predictors. Machine learning techniques were employed, including 10-fold cross-validation and evaluating model performance by R-squared and root-mean-square error, to predict the relative importance of various predictors and the global distribution of soil DOC concentrations. Worldwide soil DOC concentrations ranged from 0.04 to 7859 mg kg^{-1}, averaging 222.78 mg kg^{-1}. The ~~12~~ 14 selected predictors, including elevation, soil properties, ~~month,~~ and climate~~, and ecosystem~~, explained ~~65~~ 63 percent of the variance in soil DOC concentrations. Elevation played the most important predictor for soil DOC prediction, followed by soil organic carbon, seasonal variability of temperature, and soil clay. Soil DOC decreases initially but increases when soil clay exceeds 20% and seasonal variability of temperature exceeds 0.7. ~~Among these predictors, elevation, soil clay, and soil organic carbon were the most influential.~~ Using these findings, a global map of predicted soil DOC concentrations was produced at a 0.05° by 0.05° resolution. Global soil DOC concentrations generally increased from the equator to the poles, and the topsoil layer (0-30 cm) holed ~~12.17~~13.47 Pg of soil DOC, with substantial variations across continents. These results informed soil management practices strategies, ecosystem services evaluations, and climate change mitigation efforts. Furthermore, we envisioned integrating our database with other carbon pools to advance understanding of total soil carbon turnover and to refine Earth system models. The dataset is publicly available at https://doi.org/10.6084/m9.figshare.28574183 (Ren and Cai, 2024).

1. **Introduction**

With global changes over the last few decades, terrestrial ecosystems, which serve as the fundamental safeguard for biodiversity and function as a carbon sink, have become increasingly vital in mitigating global climate warming (Lee et al., 2023). Soils anchor the largest dynamic carbon reservoir~~Cumulatively, soil carbon pools constitute the largest carbon reservoirs~~ in terrestrial ecosystems, with the 0-1 meter storing 1,500-2,400 Pg of carbon, which is triple the atmospheric carbon stock (880 Pg) and quadruple the biotic carbon pools (450-650 Pg)~~containing three to four times more carbon than the ambient atmospheric carbon pool~~ (Lal, 2004; Zhou et al., 2024a). Sub-decadal perturbations as small as ±1% in soil carbon stocks could release 15-24 Pg C, which is equivalent to 1.5-2.4 years of anthropogenic emissions and could trigger nonlinear climate feedbacks (Schlesinger and Bernhardt, 2020). ~~Even minor fluctuations in soil carbon can significantly affect on biogeochemical cycles and the global C balance.~~ Dissolved organic carbon (DOC), a molecular continuum spanning labile metabolites (e.g., glucose, citrate) to mineral-stabilized colloids~~composed of simple organic acids and complex macromolecular substances~~, is recognized as the most active carbon pool in soil (Ren et al., 2024b). Currently, the portion of organic carbon that is water-soluble and able to pass through a 0.45 μm microporous filter membrane is referred to as DOC (Gmach et al., 2020; Guo et al., 2020a). Despite constituting 0.1-2% of total soil organic carbon, DOC mediates three disproportionately critical processes: fuelling 65-80% of heterotrophic respiration via rapid turnover, controlling mineral-organic complexation that stabilizes 40-60% of persistent carbon, and exporting 0.25-0.75 Pg C yr^{-1} to aquatic systems—a flux comparable to landuse change emissions~~Although soil DOC typically accounts for less than 2% of the soil carbon pool, it provides a substantial source of carbon and energy for soil microorganisms, while playing a key role in soil carbon sequestration, transport, and stabilization mechanisms~~ (Drake et al., 2018; Nakhavali et al., 2021; Ren et al., 2024b). Lateral DOC fluxes create a terrestrial - aquatic carbon conveyor belt equivalent to 50% of the Amazonian carbon sink, while also modifying water chemistry through pH buffering and metal complexation~~The lateral transport of DOC is crucial for linking terrestrial and aquatic ecosystems and for evaluating terrestrial carbon budgets~~ (Fichot et al., 2023). Thus, an accurate assessment of soil DOC concentrations is vital, given its unique properties, roles, and broad variability, which can span up to three orders of magnitude (Nakhavali et al., 2020; Ren et al., 2024b). Despite significant variations in soil DOC concentrations, their global distribution has not yet been systematically quantified. Bridging this knowledge gap is essential for more accurate representations of the carbon cycle in Earth system models.

Soil DOC concentration is regulated by a kinetic equilibrium between production processes~~depends on the dynamic balance between sources~~ (~~e.g., leachates from decomposing~~ plant litter leaching, rhizodeposition~~root secretions~~, and microbial necromass~~decomposition~~ release~~products~~) and removal pathways~~losses~~ (microbial mineralization, mineral adsorption, and hydrological leaching~~migration and microbial decomposition~~). ~~Any factor that alters this balance also influences soil DOC concentrations.~~ Disruption of this equilibrium, whether caused by altered substrate inputs or shifted microbial metabolic demands, reshapes DOC pool dynamics (Sokol et al., 2022). ~~Extensive research has shown that~~ Hierarchical controls shaped DOC dynamics: climatic drivers set thermal-hydrological boundaries, vegetation types modulate organic matter stoichiometry, and soil properties dictate mineral-mediated stabilization ~~soil DOC concentration is affected by climate, vegetation type, and soil properties~~ (Fichot et al., 2023; Ren et al., 2024b; Smreczak and Ukalska-Jaruga, 2021)~~, each playing a distinct role in shaping DOC dynamics~~. ~~For example, c~~Climate, often characterized by annual mean temperature and precipitation, is recognized as a primary driver of soil DOC concentrations (Lønborg et al., 2020). Temperature and precipitation directly influence soil DOC through effects on microbial activity, organic matter decomposition rates, solubility, and mobility, and indirectly shape DOC dynamics by influencing vegetation growth and soil structure (Ren et al., 2023). Vegetation type affects soil DOC primarily by altering the quantity and quality of organic matter inputs (Zhao et al., 2022). Together, climate and vegetation type profoundly affect soil biological, chemical, and physical properties, all closely with the formation and decomposition of soil DOC (Cotrufo and Lavallee, 2022). Some studies have reported large temporal variations in soil DOC concentrations at certain field sites (Ding et al., 2022; Zhao et al., 2022), with significantly higher DOC concentrations in summer and autumn than in winter and spring. Seasonal effects on soil DOC concentrations are closely associated with factors such as precipitation, soil moisture, and substrate availability (Ren et al., 2023). In warmer seasons, soil DOC production can increase due to active organic matter decomposition, driven by higher microbial activity, as well as greater DOC contributions from root exudation during periods of more active plant photosynthesis. Although relationships between soil DOC concentrations and environmental factors have been observed at local and regional scales, the relative importance of these factors at the global scale remains unclear. This lack of understanding hinders the development of effective strategies for soil carbon management and climate change mitigation.

Accurate mapping of soil DOC provides critical baseline data for addressing global challenges spanning climate-carbon feedbacks, agricultural sustainability, and aquatic ecosystem management~~is essential for addressing pressing~~

~~global challenges, including climate warming, food security, and eutrophication in aquatic systems~~ (Guo et al., 2020b; Langeveld et al., 2020). Current global soil DOC inventories remain limited in both spatial resolution and mechanistic representation. Existing maps derived from conventional geostatistical approaches, such as those by by Guo et al. (2020b) and Langeveld et al. (2020), exhibit three fundamental limitations that constrain their utility for process-based modeling. ~~To the best of our knowledge, few global maps of the spatial distribution of soil DOC exist (Guo, Z. et al., 2020; Langeveld et al., 2020). However, these maps are subject to considerable uncertainties due to limited data, restrictive factor selections, and low interpretation rates.~~ First, the global soil DOC maps produced by Guo et al. (2020b) and Langeveld et al. (2020) rely on relatively few observational data points (2890 and 762 pairs, respectively), with over 80% of training data clustered in North America and Western Europe, while tropical regions and continental interiors remain under sampled. ~~There is a lack of valid observational data for Africa, South America, Eastern Europe, and Central Asia.~~ Africa, South America, Eastern Europe, and Central Asia collectively contribute less than 5% of the global calibration datasets in these studies. Second, ~~when assessing the global distribution of soil DOC concentrations, Guo, Z. et al. (2020) and Langeveld et al. (2020) have not considered the impact of seasonal changes, even though soil DOC concentrations can vary substantially with shift of season.~~ they employ static representations of DOC dynamics, neglecting well-documented seasonal fluctuations driven by plant phenology and hydrologic pulses. Field observations demonstrate that temperate forest soils can exhibit 2-3 fold increases in DOC concentrations during autumn litterfall periods compared to spring thaw events. Third, current models oversimplify vertical DOC gradients by treating topsoil (0-30 cm) as homogeneous layers, despite empirical evidence showing exponential decreases in DOC with depth~~topsoil DOC concentrations were treated as constant value by Guo, Z. et al. (2020) and Langeveld et al. (2020), overlooking the dynamic nature of soil DOC, which decrease with increasing depth~~. In reality, soil DOC concentrations are higher in surface soils (0-10 cm) and decline with depth, exhibiting a clear vertical gradient. Finally, traditional linear regression methods used in these studies capture~~Guo, Z. et al. (2020) and Langeveld et al. (2020) have explained~~ only 30-40%~~about one-third~~ of observed~~the variation in~~ soil DOC variability, as they fail to account for threshold responses to environmental drivers such as soil pH transitions below 5.2 that trigger dissolved organic matter flocculation~~by using multivariate linear equations~~. Recent advancements in machine learning has enabled researchers to apply such techniques because of their capacities to automate feature extraction, handle large datasets, and identify complex patterns, ultimately offering significant advantages in predictive accuracy and adaptive learning.

To advance our knowledge of global soil DOC patterns and drivers, we developed a global database of soil DOC concentrations, comprising 12,807 samples from 975 published studies. Using Random Forest algorithms, we quantified the relative importance of environmental factors and predicted soil DOC concentrations on a global scale. The specific aims of this study were: (1) to determine global patterns of soil DOC concentrations, (2) to identify the primary factors controlling soil DOC concentrations on a global scale and to estimate total global soil DOC storage.

2. Material and method

2.1 Data sources and processing

We searched for publications up to December 2022 using Google Scholar (https://scholar.google.com), the Web of Science (http://apps.webofknowledge.com), and the China Knowledge Resource Integrated Database (http://www.cnki.net/) using the following search terms: (dissolved organic carbon OR dissolved organic matter OR "DOC" OR "DOM") and soil, up to December 2022. The data flow through the selection phases is shown in Fig. S1. To ensure a standardized and minimally biased dataset, we applied the following inclusion criteria: First, we included only data from terrestrial ecosystems (excluding oceans and rivers) to maintain consistency in environmental factors and ecological interactions. Second, we used only topsoil data (0-30 cm) to ensure data representativeness and quantity. Third, we recorded duplicate results from different articles only once to avoid overrepresentation of certain research groups or locations. Finally, we included agricultural soils affected by human activities such as tilling and fertilization but excluded industrial and urban soils to avoid complexity introduced by industrial and urban settings. We extracted data presented solely in figures using the digitizer function of Origin 2019. Before extracting the target data, we employed the Isolation Forest method for anomaly detection. The algorithm constructs random binary trees, where anomalies are typically isolated more rapidly, while normal points require more splitting steps.

Based on these criteria, we compiled a total of 12,807 DOC observations based on 1610 sites from 975 publications (Fig. 1a). We also collected data on experimental sites (longitude, latitude, and altitude), climate (mean annual temperature [MAT] and mean annual precipitation [MAP]), biomes (wetland, forest, shrubland, tundra, grassland, and cropland) and soil properties (soil organic carbon, texture, and pH) (Table 1). These environmental factors are used as predictors. When environmental factors were not reported in original publication, the missing data were extracted from grid datasets according to the geographic coordinates of each observed site (Table S1). We extracted elevation, MAT, ~~and~~ MAP, monthly evaporation (ETM), seasonal variability of precipitation (SVP), and seasonal variability of temperature (SVT) data from WorldClim Version 2 (https://www.worldclim.com/) with resolution of 1 km × 1 km, ~~biome~~ecosystem data from NASA's Socioeconomic Data and Applications Center (https://sedac.ciesin.columbia.edu) with resolution of 1 km × 1 km, soil properties from OpenLandMap version 2.0.0 (https://openlandmap.org) with resolution of 0.25 km × 0.25 km, and microbial biomass carbon data from the open database of figshare (https://doi.org/10.6084/m9.figshare.19556419) with resolution of 1 km × 1 km. Despite bias, ~~There~~ there is a significant linear relationship between the measured values and the corresponding extracted values (Fig. S2). Noteworthy, this bias could introduce some uncertainty to the results. Overall, our study sites spanned a wide range of latitudes (−64.81° to 78.85°) and longitudes (−159.66° to 175.95°) (Table 1), encompassing a large climate gradient with MAT from −11.16 to 28.00℃ and MAP from 30 to 4200 mm.

2.2 Data standardization

For our database, the DOC concentrations were quantified using a mix of physical and chemical techniques. Physical methods included soil solution collection using lysimeters or ceramic suction. Chemical methods employed various solvents like distilled water, potassium chloride (KCl), or potassium sulfate (K_2SO_4) as described by Li et al. (2018)**.** Over 74.32% of the DOC was determined using chemical techniques, which highlighted their reliability. For consistency, the DOC values derived from physical approaches was converted to chemical method values using the following equation:

$$DOC_{soil} = (DOC_{solution} \times V \times 1000)/W \times [1/(V \times (1\text{-}W) \times BD \times 1000000)] \quad (1)$$

where, DOC_{soil} represents soil DOC concentration determined by chemical methods (mg g^{-1}); $DOC_{solution}$ is the concentration measured by physical methods (mg L^{-1}); W denotes the volumetric soil moisture ($m^3 m^{-3}$); V is the volume of the soil column for solution extraction (m^3); and BD is the soil bulk density (g cm^{-3}). The factor 1000 converts m^3 to L, and 1,000,000 converts m^3 to cm^3 following established by Guo (Guo et al., 2020b). This standardization allowed for a consistent comparison and analysis of the DOC data across various studies.

2.3 Predictive modeling

The driving factors of soil DOC concentrations were divided into ~~four~~ four categories: elevation, climate, ecosystem, and soil properties~~, and temporal variation~~. Soil properties included physical attributes (clay, sand, bulk density, and depth), chemical attributes (SOC, pH), and a biological attributes (microbial biomass carbon) attributes. ~~Temporal~~

~~variation was represented by month.~~ Climate comprised MAT, MAP, ~~and elevation~~ETM, SVP, and SVT. Ecosystems encompassed wetland, forest, shrubland, tundra, grassland, and cropland. In our predictive models, correlated predictors could substitute for each other, causing their importance to be shared and thus potentially underestimated. Consequently, we excluded soil ~~total nitrogen,~~ silt~~, and the aridity index~~ because they were correlated with soil ~~soil organic carbon,~~ sand~~, and MAP, respectively~~ (Fig. S3). Further, we did not include some variables (e.g., soil moisture, soil porosity, ferroaluminum oxide, microbial structures, microbial diversity, and carbon cycling enzymes) because they were rarely report in the target papers.

To develop and optimize a predictive model for soil DOC, we employed an array of regression methods, which encompassed three linear and four nonlinear approaches (Table S2). The linear methods included a least absolute shrinkage and selection operator (LEAPS), elastic net (ENET), and standard linear modeling (LM) to identify the most important predictor variables, while minimizing overfitting. The nonlinear methods included the random forest (RF) algorithm, boosted tree (BOOSTED), bagged tree (Bagged), and cubist (CUBIST) models. Each model had intrinsic feature selection processes, and we fine-tuned them to improve accuracy and control complexity. During optimization phase, various actions were implemented. LEAPS models were educated to accommodate the largest number of variables. We applied penalties for feature condensation (diminishing the role of less impactful variables in the resultant linear formula) between 0 and 0.1, incremented by 0.01, to discipline the models. RF growth was restricted at a maximum of 1,000 trees and limited the number of predictors to one-third of the maximum possible, ensuring a balance between complexity and manageability. BOOSTED models underwent training with 10 to 100 trees, each having between 1 to 7 nodes. We incorporated shrinkage rates of 0.01 or 0.1, with a maximum tree size of 5. For CUBIST model, we explored neighboring values from 1 to 9 in increments of 2 and varied community sizes from 1 to 100, refining predictive accuracy. In every instance, the models were evaluated using Monte Carlo cross-validation with 100 iterations, employing a 70/15/15 split between training, validation, and testing sets (Fig. 2b and Fig. ~~S6~~S7 and 8). The root mean square error and R^2 values were calculated to evaluate model accuracy and residual variance, which served as criteria for ranking model performance (Table S2). A 10-fold cross-validation method was used to evaluate model performance. A flowchart for model selection process was shown in Fig. S5. Finally, the RF model was used to predict soil DOC concentrations. The factor of ecosystems was excluded based on the IncNodePurity of RF model (Fig. S6).

To evaluate the effects of independent variables on soil DOC, a variable importance analysis was conducted using permutation variable importance measurements. This analysis was performed with the variable importance tool integrated into the R packages for the RF model that exhibited the highest predictive quality. In essence, this method assessed prediction errors within the model by calculating mean square errors for each regression tree. The models' variable importance scores assessed the influence of predictor variables on the outcomes. For enhanced comparability of all model inputs, the independent environmental variables were scaled to a 0–100% range to facilitate comparisons of their proportional contribution to the model's predictions. For evaluate the sensitivity analysis of model predictions, the Sobol index, a variance of based global sensitivity analysis method, was used to assesses how model input parameters impact output results (Fig. S9). It breaks down the system's total variance into contributions from individual inputs and their combinations.

Partial dependence analyses were employed to examine the relationships between predicted soil DOC and independent variables across their entire value ranges in the RF model. These analyses allowed us to isolate the effects of specific independent variables by removing the influence of the others. Partial dependence plots offered insights into the average marginal effects of one or more independent variables on model predictions. For instance, these plots could reveal whether relationships were linear, monotonic, or more complex. By examining curvature and inflection points, we could identify where variable exerted strong, immediate effects or where their influences were more subtle and possibly mediated by other variables. We reported the x-axis as a standardized value, ensuring a clear progression from low to high values. When we generated partial dependence with RF, several uncertainties arose. The high model complexity sometimes slowed predictions, especially with many trees. The limited interpretability of the RF models could complicate understanding partial dependence. Sensitivity to noise potentially led to overfitting and reduced accuracy. Variable importance measurements could also be biased by varying feature scales or categories, potentially skewing interpretations of feature-outcome relationships. For explore the interaction effects between key drivers of derived soil DOC concentration, SHapley Additive exPlanations (SHAP) is used to interpret machine learning model predictions by calculating the contribution of features to the model's predictions (Fig. 4). SHAP values can be further decomposed into main effects and interaction effects, where interaction effects reveal the interactions between features. SHAP interaction values are obtained by first defining an explainer using the TreeExplainer function (by passing the model to it), and then deriving the interaction values from this explainer. These values can be interpreted similarly to standard SHAP values, explicitly quantifying how individual features and their pairwise interactions contribute to specific predictions.

2.4 Global soil DOC mapping

The global distribution of soil DOC and the relative uncertainties of our predictions were generated by combining our DOC dataset with the RF model, which incorporated global climate~~, vegetation,~~ and soil-rasterized datasets (Figs. ~~4~~5, ~~S5~~ S11 and Table S1). We first produced factor maps from the key input variables, focusing on the ~~12~~ 14 distinct variables associated with each raster cell. Subsequently, the factor maps were employed to derive a spatially detailed global map of soil DOC. To achieve global-scale mapping, we processed the driving factors at a 0.05° resolution to calculate soil DOC values. Areas that did not meet the following criteria were excluded from our prediction: (1) absence of data for any essential predictors, (2) soil order and biomes not aligning with the previously discussed aggregated land use systems, or (3) locations in climate zones outside the scope of our model's focus. Duo to the different spatial resolution of input variables data, resampling techniques enables the conversion of raster data between spatial resolutions to facilitate spatial analysis and modeling. The core principle of resampling involves estimating pixel values at new resolutions through interpolation or other mathematical methods. Specifically, down-sampling (high-to-low resolution conversion) requires aggregating values from multiple high-resolution pixels into a single low-resolution pixel. Up-sampling (low-to-high resolution conversion) necessitates generating new pixel values through interpolation algorithms. To evaluate uncertainty due to data resampling and unexplained variability not accounted for by the independent variables, we analyzed finer-resolution (5 km²) grids where driving factors were available at this detailed. This analysis clarified the overall uncertainty inherent in our global soil DOC estimation. The corresponding map of relative uncertainty of prediction was built by displaying the standard deviation divided by the mean prediction, based on our final random forest RF model. The standard deviation reflected the range of possible predictions derived from the iterative build-up of decision trees after 500 model runs.

Soil DOC concentration varied significantly with ~~temporal changes~~ecosystems (Table 2) and soil depth (Fig. ~~2~~3). Ecosystems were divided into wetland, forest, shruland, tundra, grassland, and cropland (Fig. S10).~~Sampling time (month) was used to represent seasonal variations in soil DOC concentration.~~ Soil DOC concentration decreased with soil depth and reached a turning point at approximately 10 cm (Fig. ~~2~~3). Therefore, when extrapolating the RF model to the entire globe, we used a month range from 1 to 12 and depths of 5 (0–10 cm) and 20 (10–30 cm). From this, we generated a total of ~~24~~ 12 maps of global soil DOC concentration. We combined these ~~24~~ 12 maps into a single map representing the global distribution of soil DOC concentration based on soil depth. Finally, we calculated the global soil DOC stock using the following equation applied to the combined map of global soil DOC concentration:

$$SOC_s = \sum SOC_i \times BD_i \times (1 - f) \times T \times M_i \qquad (2)$$

where SOC_s is SOC stock and SOC_i is SOC concentration. The subscript i is the number of global grid. BD, f, and T are soil bulk density, the volumetric percentage of coarse fraction (>2 mm), and the depth of soil layer, respectively. M is the effective area of each grid.

3. Results

3.1 Soil DOC concentrations in different ecosystems globally

A total of 12,807 soil DOC observations were compiled from 975 publications that spanned six continents, as well as major biomes and terrestrial ecosystems (Fig. 1). We found that the natural logarithm of soil DOC concentrations conformed to a normal distribution (Fig. 1b). Global soil DOC concentrations ranged from 0.04 to 7859 mg kg^{-1}. The global average, median, and standard deviation were 222.78, 101.01, and 445.78 mg kg^{-1}, respectively (Table 2). We observed that soil DOC concentrations varied across ecosystems. Tundra had the highest average and median soil DOC concentrations at 470.78 and 241.90 mg kg^{-1}, respectively. Grassland averaged 327.77 mg kg^{-1} with a median of 126.48 mg kg^{-1}, while forest averaged 256.18 mg kg^{-1} with a median of 115.51 mg kg^{-1}. Wetland averaged 218.53 mg kg^{-1} with a median of 107.11 mg kg^{-1}, cropland averaged 165.98 mg kg^{-1} with a median of 83.00 mg kg^{-1}, and shrubland averaged 160.24 mg kg^{-1} with a median of 127.84 mg kg^{-1} (Table 2).

3.2 Model performance and drivers of soil DOC concentrations

We estimated RMSE and R^2 for all tuned models and used these statistics to analyze residual variance and accuracy, as well as to rank model performance (Table S2). To facilitate interpretation of uncertainty, we also calculated relative RMSE by dividing the absolute error by the global mean soil DOC concentration. RF model resulted in the best performance within one standard error of the minimal RMSE and were thus used for further analyses of variable importance. The residual plot of train, validation, and test data for RF model were randomly distributed near zero (Fig. ~~S7~~S8). Overall, nonlinear models (R^2 = 0.~~42~~41–0.63~~5~~; RMSE = ~~250~~248~~–332~~327) outperformed linear models (R^2= 0.10~~1~~–0.11~~08~~; RMSE = ~~410~~401~~–427~~411) (Table S2). The RF model yielded the lowest RMSE within one standard deviation range and was therefore selected for subsequent analyses of variable importance (~~Fig~~Table~~.~~ S2~~3~~). The relative importance of soil DOC drivers and the global map of soil DOC distribution were derived from the RF model outputs (Fig. 4 and Fig. S11).

~~We based the relative importance of soil DOC drivers and the global map of soil DOC on averaged RF model results.~~

The RF model explained ~~65~~63% of the variability in soil DOC concentrations across all sites and achieved the lowest RMSE compared with other models (Fig. 2 and~~,~~ Table S2). Elevation played the most important predictor for soil DOC prediction among the selected 14 variables, followed by SOC, SVT, and soil clay. The relative importance of MAP, SVP, MBC, soil pH, soil sand, and soil C:N was gradually diminishing. ~~Soil properties were the most important predictor categories, with elevation and soil clay content exerting the greatest influence.~~ Meantime, elevation~~ and~~, ~~-~~SOC, SVT, soil~~soil organic~~ sand and soil clay were the ~~most~~ more sensitivity factors of RF model than the other predictor (Fig. ~~S8~~S9). ~~Although less influential, other predictors still contributed; among them, soil organic carbon and soil pH were notable. Although mean annual precipitation, and mean annual temperature, microbial biomass carbon, bulk density, sand, depth, month, and ecosystem also affected soil DOC, their contributions were lower than those of the top four predictors. Elevation has strong interaction with soil pH, bulk density, and microbial biomass carbon (Fig. S9).~~ Partial dependence analysis produced results (Fig. 3) similar to Pearson correlation analyses (Fig. S4). We found a positive correlation between soil DOC and both elevation and soil organic carbon, although there were fewer data points corresponding to higher elevations and greater soil organic carbon values (Fig. 3f). Soil DOC showed a trend of decreasing first and then increasing with the increase of MAT (0-30 °C), SVT (0-1.5), and soil clay (0-50%) (Fig. 3a, d and h). Soil DOC showed a trend of decreasing first and then stabilizing with the increase of soil depth and soil pH (4-8.5). The inflection point of soil depth and soil pH was 10 cm and 5.8, respectively (Fig. 3i and k). ~~, and a negative correlation between soil DOC and both mean annual temperature and soil pH (Fig. 3h).~~

Elevation, SOC, SVT, and soil clay had strong negative interactions with MAT (Fig. 4). This means as the MAT variable increases, the influence of the other variables is weakened. Elevation had a positive interaction with bulk density, suggesting they work together to affect soil DOC.

3.3 Global soil DOC patterns

The RF model has the ability to predict soil DOC in wetland (R^2=0.87), forest (R^2=0.~~84~~85), shruland (R^2=0.~~84~~85), tundra (R^2=0.~~70~~77), grassland (R^2=0.~~95~~96), and cropland (R^2=0.90)~~ ecosystems~~ (Fig. S10). We observed significant spatial heterogeneity in predicted global soil DOC concentrations (Fig. ~~4a~~5a). Soil DOC concentrations increased from the equator toward the poles (Fig. ~~4b~~5b). High soil DOC concentrations were found in high-altitude plateaus and mountain ranges at low latitudes, including the Andes, African Highlands, and West Indies (Fig. ~~4a~~5a). The global average soil DOC concentration was ~~237.56~~224.72 mg kg^{-1} (Table 3), and the topsoil (0-30 cm) DOC stock was ~~12.17~~13.74 Pg. Asia had the highest soil DOC concentration (~~274.43~~259.03 mg kg^{-1}), followed by North America (~~263.63~~250.66 mg kg^{-1}), South America (219.83 mg kg^{-1}), Europe (~~227.34~~208.28 mg kg^{-1}) and Oceania (206.36 mg kg^{-1})~~, and South America (215.81 mg kg^{-1})~~. ~~Oceania and~~ Africa had the lowest soil DOC concentrations (~~198.13 and 186.35~~166.73 mg kg^{-1}~~, respectively~~). For predicted soil DOC stocks, Asia and North America remained ranked first and second at ~~4.8~~ 4.93 and ~~2.45~~2.93 Pg, respectively. Despite its relatively low predicted soil DOC concentrations, Africa ranked third in total DOC stock (~~2.07~~ 2.37 Pg) because of its large land area. South America followed at ~~1.37~~1.76 Pg, while Europe and Oceania had the lowest stocks at ~~0.88~~0.98 and ~~0.59~~0.76 Pg, respectively.

4 Discussions

~~4.1 Variations in soil DOC between ecosystems~~

~~Given the substantial number of measurements included in our study (12,807 observations), the range of topsoil (0–30 cm) DOC concentrations (0.04–7859 mg kg^{-1}) was broader than previously reported for a database of 2,890 observations (Guo, Z. et al., 2020). Our global median soil DOC concentration was 101.01 mg kg^{-1} (Table 2), in contast to a previously reported average of only 77.39 mg kg^{-1}. For different ecosystems, the median soil DOC concentrations of wetlands, tundra, and shrublands in our study aligned with previously reported values (Guo, Z. et al., 2020), primarily due to the relatively small number of observations for these ecosystems, with tundra comprising only 1% of our database. However, significant differences emerged in forests, grasslands, and croplands compared with previous data. Tundra had the highest soil DOC concentration (Table 2). This can be attributed to low soil temperatures and limited microbial activity, which slow the decomposition of organic material and lead to higher soil DOC concentrations (Propster et al., 2023). In addition, prolonged soil freezing in tundra areas reduces evaporation and oxygen supply, further slowing organic decomposition. Soil DOC concentrations were also relatively high in grassland, forest, and shrub ecosystems because leaves, dead branches, and plant root exudates provide abundant organic C inputs (Cai et al., 2021). However, our results indicated that DOC concentrations in forest soils were consistently lower than in grasslands (Table 2). Grassland ecosystems often have higher plant diversity, including legumes and weeds, whose residue decomposition contributes to increased DOC concentrations (Perrot et al., 2023). In contrast, the cooler conditions in forest soils limit microbial activity and slow organic matter decomposition, reducing DOC consumption. Additionally, grassland soils tend to have better water conditions, promoting higher microbial activity and organic matter breakdown, thus increasing DOC concentrations (Deng et al., 2023). Differences in land use and management, forests being less disturbed while grasslands may be more~~

~~frequently disturbed by grazing, can also influence soil organic matter decomposition and DOC levels. These combined factors of vegetation type, microbial activity, water conditions, and land use practices result in varying soil DOC concentrations between these two ecosystems. The lowest median soil DOC concentration appeared in cropland ecosystems, likely due to decreased soil organic matter inputs resulting from frequent tillage and harvesting, as well as accelerated DOC decomposition caused by tillage (Ren et al., 2024). Meanwhile, our median soil DOC concentration for croplands was 83.00 mg kg⁻¹, whereas a previous value was 60.58 mg kg⁻¹ (Guo, Z. et al., 2020). This discrepancy may be due to the previous database having only 13% cropland observations, while our cropland observations were approximately ten times larger. In summary, our study builds on earlier work by incorporating a more extensive dataset that better represents heterogeneous global conditions.~~

4.~~2~~ 1 Effects of elevation and ~~climate and controlled~~ soil properties on soil DOC concentrations

The ~~two~~ most critical predictors of soil DOC concentrations among the selected 14 variables ~~were~~ were elevation ~~and soil clay content~~ (Fig. 2)~~.~~ , with soil DOC concentrations exhibiting a significant positive correlation with elevation after controlling for confounding variables (Fig. 3f). This finding contrasted with several previous studies that prioritized precipitation regimes (Guo et al., 2020b) or soil texture (Angst et al., 2021) as primary soil DOC drivers, suggesting that elevation effects may have been obscured in large-scale analyses lacking environmental stratification. Three interconnected mechanisms may explain this pattern of elevation effects. First, ~~In high-altitude regions,~~ decreasing~~lower~~ temperatures at high-altitude regions (0.6 ℃/100m adiabatic lapse rate) limit the metabolic activity of microorganisms (Davidson and Janssens, 2006), slowing the decomposition of soil DOC and favoring soil DOC accumulation through reduced mineralization. Additionally, these regions typically receive more precipitation, which increases soil moisture and helps protect soil DOC from rapid breakdown. ~~Typically, high-altitude regions host vegetation types characterized by longer growth cycles and greater litterfall (Pesántez et al., 2018; Wei et al., 2024).~~High-altitude regions often experience distinct precipitation patterns and soil moisture conditions compared with lower elevations (Li et al., 2023). Higher precipitation and lower evaporation rates may promote greater dissolution and leaching of organic matter, thereby increasing soil DOC concentrations (He et al., 2021; Lu et al., 2019). Second, the altitudinal shift in vegetation communities, particularly the transition to coniferous species and ericaceous shrubs at higher elevations, enhances labile carbon inputs through distinct litter chemistry (higher phenolic compounds and lower C:N ratios), which created a positive feedback loop for DOC production (Pesántez et al., 2018; Wei et al., 2024). ~~Decomposing plant residues contribute to SOC, a portion of which is converted to DOC. Consequently, differences in vegetation type and productivity also influence soil DOC concentrations (Camino‐Serrano et al., 2014; Rahbek et al., 2019). We also found that forest and grassland sites above 2,000 m, which accounted for 73% of the high DOC observations.~~Third, the orographic precipitation effect and persistent cloud immersion at higher elevations maintain soil moisture conditions that simultaneously stimulate DOC release from organic matter while limiting its lateral export through reduced drainage flux (Michalzik et al., 2001). Moreover, high-altitude areas are generally less disturbed by humans activities, which may help preserve soil DOC. Our results also indicated that soils in low-latitude plateaus and mountain ranges (e.g., Tibetan Plateau, Andes, African Highlands, and West Indies) exhibited higher DOC concentrations (Fig. ~~4a~~5a). These results fundamentally recalibrated our understanding of topographic controls on soil carbon cycling, which provided a mechanistic basis for predicting climate feedbacks in vertically stratified landscapes.~~(Sanders et al., 2021)(Awedat et al., 2021)(Awedat et al., 2021)~~

~~Soils with high clay content have a strong adsorption capacity that more effectively retains DOC and reduces its loss. Clay also provides a suitable habitat for microorganisms, affecting microbial communities' structure and activity and thus regulating the rate of soil DOC turnover.~~ The effects of soil clay content on DOC concentrations are complex, involving adsorption, water retention, microbial activities, and organic matter protection mechanisms (Kaiser and Zech, 2000; Singh et al., 2017). Generally, high clay content fosters DOC accumulation through the adsorption and stabilization of organic matter (Gmach et al., 2019; Kalbitz et al., 2000). Our findings revealed a nonlinear threshold control of soil clay content on soil DOC with minimum DOC concentrations occurring at 20% clay (Fig. 3h), which was a pedogenic tipping point where the dominant regulatory mechanisms shift from physicochemical stabilization to biogeochemical accumulation. In soils with clay content below this threshold, increasing clay promotes organo-mineral association through Fe/Al-oxide bridging and exponential growth of specific surface area (Sanders et al., 2021), which effectively sequester labile organic carbon into micro-aggregates while suppressing soil DOC release. Beyond 20% clay, however, the emergence of impermeable microstructures reduces oxygen diffusion, establishing anaerobic microsites that inhibit phenol oxidase activity and accumulate phenolic metabolites (Awedat et al., 2021). This shift coincides with clay-organic co-precipitation dynamics: high-clay soils (>25%) exhibit stronger preferential dissolution of Fe-OM complexes during redox oscillations (Awedat et al., 2021). Furthermore, ~~the effects of SOC and soil pH on DOC should not be overlooked (Fig. 2a).~~ SOC serves as the main source of DOC, so higher SOC results in more DOC release through microbial metabolism (Kalbitz et al., 2000; Neff and Asner, 2001). ~~Variations in soil pH can alter the charge of soil colloids, influencing adsorption-desorption mechanisms and thus affecting DOC solubility (Andersson and Nilsson, 2001; Cheng et al., 2020; Kaiser et al., 2005). Overall, soil DOC concentration arises from interactions among soil and climate factors, as well as biological, chemical, physical, and human influences at various spatial and temporal scales. Each factor plays a distinct role in shaping DOC dynamics.~~

4.2 Effects of climate on soil DOC concentrations

Seasonal temperature variability (SVT) was the predominant climatic driver of soil DOC, exhibiting a nonlinear threshold response where soil DOC concentrations initially decline but shift to an increasing trend beyond an SVT threshold of 0.7 after accounting for confounding factors (Fig. 3d). This contrasts sharplied with previous studies that primarily attributed soil DOC fluctuations to mean annual temperature or precipitation (Guo et al., 2020b) or emphasized moisture variability over thermal regimes (Li et al., 2018). This makes our work the first study to identify SVT-driven biphasic DOC behavior in global terrestrial ecosystem. Three interconnected mechanisms could explain this pattern. First, moderate SVT levels (<0.7) likely enhance microbial carbon use efficiency by promoting enzymatic acclimation to predictable thermal fluctuations, which reduce soil DOC accumulation through efficient mineralization (Ren et al., 2024a). Second, surpassing the 0.7 SVT threshold destabilizes microbial communities through repeated thermal shocks, which increase cell lysis and releasing labile organic compounds into the soil matrix (Zhou et al., 2024b). Third, extreme temperature variability alters soil physical structure by disrupting aggregate stability and exposes previously protected organic matter to solubilization during thermal contraction-expansion cycles (Six et al., 2004). The observed DOC rebound at high SVT aligns with plant root exudation strategies under thermal stress, which suggested that vegetation may compensate for microbial carbon loss by releasing soluble metabolites to maintain rhizosphere functionality (Kruthika et al., 2024). Overall, the identified SVT threshold (0.7) serves as an early warning indicator for ecosystems approaching critical thermal instability, particularly in climate transition zones where seasonal temperature swings are intensifying. Practically, this threshold could guide land management strategies. For instance, prioritizing organic amendments or shade crops in regions with SVT >0.7 may mitigate soil DOC leaching risks.

4.3 Global patterns of soil DOC

Using our soil DOC concentration dataset, we quantified the soil DOC concentrations (0-30 cm) in terrestrial ecosystems, identified their key driving factors, and produced global predictions. Global DOC stocks in the topsoil are estimated at 13.74~~12.17~~ Pg C, accounting for 0.~~775~~87% of global soil organic carbon, which is significantly higher than previous estimates (Guo et al., 2020b). Our predictions indicated that soil DOC concentrations decreased markedly toward lower latitudes, particularly in the Northern Hemisphere. Previous global maps of soil DOC concentrations failed to capture this latitudinal trend, likely due to limited spatial coverage (Guo et al., 2020b; Langeveld et al., 2020). Our predicted map shows that soil DOC concentrations increase with latitude. In high-latitude regions, low temperatures limit microbial activity, which slows the decomposition of organic matter and leads to more organic carbon being retained in dissolved form (Patoine et al., 2022) , thereby increasing soil DOC concentrations. In addition, soils in high-latitude areas are often moist or frozen due to low temperatures, limiting oxygen supply and further inhibiting microbial decomposition (Zhou et al., 2024c). These moist or frozen conditions also help protect organic matter, reducing its decomposition and contributing to DOC accumulation. Thus, low temperatures and specific moisture conditions in high-latitude regions jointly result in relatively high soil DOC concentrations. However, substantial heterogeneity exists at regional and local scales. For instance, despite their similar latitudes, soil DOC concentrations in Northern Europe were significantly lower than in Siberia, primarily due to differences in climatic conditions. Northern Europe's maritime climate, with mild temperatures and evenly distributed precipitation, promotes higher microbial activity and accelerates organic matter decomposition. In contrast, Siberia's cold subarctic climate results in lower soil temperatures that limit microbial activity and slow organic matter decomposition, leading to greater DOC retention (Jin and Ma, 2021). Furthermore, soils in Siberia are often frozen, restricting oxygen supply and further inhibiting decomposition, thereby contributing to DOC accumulation (Raudina et al., 2022). Climatic conditions thus play a key role in explaining the significant differences in soil DOC concentrations between these regions. Regional variations may also be related to topographic conditions. Higher soil DOC concentrations on the Tibetan Plateau compared with Eastern China may result from high elevation and low MAT in the plateau (Fig. ~~4a~~5a). In contrast, other studies reported lower DOC levels in Arctic regions, which may have been due to omitting DOC concentration measurements in dry or frozen soils (Langeveld et al., 2020). Our predictive model offered higher accuracy in estimating global soil DOC storage because our comprehensive dataset included DOC concentrations in both dry soil and soil solutions, providing a robust data foundation. In addition, we used the optimal model by comparing various linear and nonlinear models to predict global soil DOC.

~~**4.4 Implications for carbon cycling models**~~

~~Carbon cycling models are key tools for predicting how soil organic carbon responds to future global changes. Considerable uncertainty exists in simulating and predicting soil organic carbon cycles in many current Earth system models, largely due to model structure, model parameters, and initial conditions (Luo et al., 2015). Regarding model structure, the soil carbon pools in models cannot be directly separated through experiments, which hamper the quantification of many parameters (Bailey et al., 2018). By integrating global soil DOC concentration data and coupling it with particulate organic carbon, mineral-associated organic carbon, and microbial biomass carbon pools, future models can establish a quantifiable structure based on measurable pools. Our study reveals key factors affecting soil DOC concentrations, such as elevation, soil clay content, and soil organic carbon, can be incorporated into carbon cycle models to improve their predictive capabilities. Moreover, this research provides a detailed global distribution map of soil DOC, which is essential for model parameterization and validation, particularly in regions where data are scarce.~~

4.~~5~~ 4 Limitations and predictive uncertainties

Although we compiled a comprehensive global soil DOC concentration dataset, identified key drivers, and made a global prediction, our study had certain limitations. First, certain ecosystems remained underrepresented; for instance, tundra accounted for only 1% of our database, while shrublands, grasslands, and wetlands collectively constituted only 21%. This underrepresentation may reduce the accuracy of predictions for different ecosystems. Second, although we considered the subsoil at the beginning of dataset, we did not explore this further due to the limited availability of data and considerations of predictive accuracy. We intend to continue expanding the subsoil DOC database in future work. Third, there was a deficiency in some predictive variables; although we had extracted missing data through gridded datasets, this inevitably introduced uncertainty in predictions, particularly for soil variables. Fourth, although data standardization enables consistent comparison and analysis of soil DOC across different measurement methods, there were potential issues such as the possible loss of original data characteristics, dependence on accurate parameters, overgeneralization, increasing the complexity of data interpretation, and introducing bias. Finally, despite employing advanced machine learning methods with multiple predictors to predict the global soil DOC, 35% of soil DOC concentration variability remains unexplained. However, these limitations also highlighted areas for future soil DOC research. Future research should enhance the collection of deep soil samples to address the current data scarcity and more accurately quantify the DOC reserves across the entire soil profile. There is a particular need to increase sample collection in key regions such as Siberia and Africa.

5 Data availability

The global soil DOC in this study and raw dataset of driving factors can be downloaded at https://doi.org/10.6084/m9.figshare.28574183 (Ren and Cai, 2024).

6 Conclusions

Through the development of a comprehensive soil DOC dataset, we quantified soil DOC concentrations in terrestrial ecosystems, identified their driving factors, and made global predictions. After~~ to~~ comparing multiple predictive models, we selected the Random Forest model as the best performer for mapping soil DOC concentrations. The results indicated that tundra exhibited the highest DOC concentrations, while shrubland and cropland soils had relatively lower concentrations. Elevation played the most important predictor for soil DOC prediction, followed by SOC, SVT, and soil clay. There was a nonlinear threshold response of soil DOC to soil clay and SVT, which initially decline but shift to an increasing trend beyond an soil clay threshold of 20% and SVT threshold of 0.7 after accounting for confounding factors. ~~Climate factors (elevation) and soil properties (clay content, SOC, pH) jointly regulated DOC variations.~~ We predicted that the soil DOC concentration increased significantly from the equator to the poles, and estimated that the DOC stocks in the topsoil of terrestrial ecosystems were ~~12.17~~13.74 Pg. The global soil DOC database we created served as a critical resource for future research and enhanced our understanding of the roles of soil in the global carbon cycle. This database provided valuable data support for climate change research, ecosystem management, agricultural sustainability, environmental policymaking, and the improvement of biogeochemical models. It aided in addressing soil degradation, improving food security, and tackling global environmental challenges.

Author contributions

Andong Cai designed this study. Tianjing Ren collected the data. Tianjing Ren and Andong Cai discussed analyzing methods. Andong Cai conducted the analysis. Tianjing Ren drafted the manuscript. All authors discussed the results and contributed to the manuscript.

Competing interests

The contact author has declared that neither they have any competing interests.

Acknowledgements

We would like to thank Frank Boehm at NanoApps Consulting2341York Ave. Vancouver, BC, Canada for his assistance with English language and grammatical editing.

Financial support

This work was financially supported by the National Key Research and Development Program of China (2022YFD2300500).

References

Andersson, S., & Nilsson, S. I. (2001). Influence of pH and temperature on microbial activity, substrate availability of soil-solution bacteria and leaching of dissolved organic carbon in a mor humus. Soil Biology and Biochemistry, 33, 1181-1191. https://doi.org/10.1016/S0038-0717(01)00022-0.

Angst, G., Pokorn, J., Mueller, C. W., Prater, I., Preusser, S., Kandeler, E., Meador, T., Straková, P., Hájek, T., & Buiten, G. V. (2021). Soil texture affects the coupling of litter decomposition and soil organic matter formation. Soil Biology and Biochemistry, 159, 108302. https://doi.org/10.1016/j.soilbio.2021.108302.

Awedat, A. M., Zhu, Y., Bennett, J. M., & Raine, S. R. (2021). The impact of clay dispersion and migration on soil hydraulic conductivity and pore networks. Geoderma, 404, 115297. https://doi.org/10.1016/j.geoderma.2021.115297.

Bailey, V. L., Bond-Lamberty, B., DeAngelis, K., Grandy, A. S., Hawkes, C. V., Heckman, K., Lajtha, K., Phillips, R. P., Sulman, B. N., Todd-Brown, K. E. O., & Wallenstein, M. D. (2018). Soil carbon cycling proxies: Understanding their critical role in predicting climate change feedbacks. Global Change Biology, 24, 895-905. https://doi.org/10.1111/gcb.13926.

Cai, A., Liang, G., Yang, W., Zhu, J., Han, T., Zhang, W., & Xu, M. (2021). Patterns and driving factors of litter decomposition across Chinese terrestrial ecosystems. Journal of Cleaner Production, 278, 123964. https://doi.org/10.1016/j.jclepro.2020.123964.

Camino-Serrano, M., Gielen, B., Luyssaert, S., Ciais, P., Vicca, S., Guenet, B., Vos, B. D., Cools, N., Ahrens, B., Altaf Arain, M., Borken, W., Clarke, N., Clarkson, B., Cummins, T., Don, A., Pannatier, E. G., Laudon, H., Moore, T., Nieminen, T. M., Janssens, I. (2014). Linking variability in soil solution dissolved organic carbon to climate, soil type, and vegetation type. Global Biogeochemical Cycles, 28, 497-509. https://doi.org/10.1002/2013gb004726.

Cheng, X., Hou, H., Li, R., Zheng, C., & Liu, H. (2020). Adsorption behavior of tetracycline on the soil and molecular insight into the effect of dissolved organic matter on the adsorption. Journal of Soils and Sediments, 20, 1846-1857. https://doi.org/10.1007/s11368-019-02553-7.

Cotrufo, M. F., & Lavallee, J. M. (2022). Soil organic matter formation, persistence, and functioning: A synthesis of current understanding to inform its conservation and regeneration. Advances In Agronomy, 172, 1-66. https://doi.org/10.1016/bs.agron.2021.11.002.

Davidson, E. A., & Janssens, I. A. (2006). Temperature sensitivity of soil carbon decomposition and feedbacks to climate change. Nature, 440, 165-173. https://doi.org/10.1038/nature04514.

Deng, M., Li, P., Liu, W., Chang, P., Yang, L., Wang, Z., Wang, J., & Liu, L. (2023). Deepened snow cover increases grassland soil carbon stocks by incorporating carbon inputs into deep soil layers. Global Change Biology, 29, 4686-4696. https://doi.org/10.1111/gcb.16798.

Ding, H., Hu, Q., Cai, M., Cao, C., & Jiang, Y. (2022). Effect of dissolved organic matter (DOM) on greenhouse gas emissions in rice varieties. Agriculture, Ecosystems & Environment, 330, 107870. https://doi.org/10.1016/j.agee.2022.107870.

Drake, T. W., Raymond, P. A., & Spencer, R. G. (2018). Terrestrial carbon inputs to inland waters: A current synthesis of estimates and uncertainty. Limnology and Oceanography Letters, 3, 132-142. https://doi.org/10.1002/lol2.10055.

Fichot, C. G., Tzortziou, M., & Mannino, A. (2023). Remote sensing of dissolved organic carbon (DOC) stocks, fluxes and transformations along the land-ocean aquatic continuum: Advances, challenges, and opportunities. Earth-science Reviews, 242, 104446. https://doi.org/10.1016/j.earscirev.2023.104446.

Gmach, M. R., Cherubin, M. R., Kaiser, K., & Cerri, C. E. P. (2020). Processes that influence dissolved organic matter in the soil: a review. Scientia Agricola, 77, e20180164. https://doi.org/10.1590/1678-992x-2018-0164.

Guo, B., Zheng, X., Yu, J., Ding, H., Pan, B., Luo, S., & Zhang, Y. (2020). Dissolved organic carbon enhances both soil N2O production and uptake. Global Ecology and Conservation, 24, e01264. https://doi.org/10.1016/j.gecco.2020.e01264.

Guo, Z., Wang, Y., Wan, Z., Zuo, Y., He, L., Li, D., Yuan, F., Wang, N., Liu, J., Song, Y., Song, C., Xu, X., & Hickler, T. (2020). Soil dissolved organic carbon in terrestrial ecosystems: Global budget, spatial distribution and controls. Global Ecology and Biogeography, 29, 2159-2175. https://doi.org/10.1111/geb.13186.

He, X., Augusto, L., Goll, D. S., Ringeval, B., Wang, Y., Helfenstein, J., Huang, Y., Yu, K., Wang, Z., Yang, Y., & Hou, E. (2021). Global patterns and drivers of soil total phosphorus concentration. Earth System Science Data, 13, 5831-5846. https://doi.org/10.5194/essd-13-5831-2021.

Jin, H., & Ma, Q. (2021). Impacts of permafrost degradation on carbon stocks and emissions under a warming climate: a review. Atmosphere, 12, 1425. https://doi.org/10.3390/atmos12111425.

Kaiser, K., Guggenberger, G., Haumaier, L., & Zech, W. (2005). Dissolved organic matter sorption on sub soils and minerals studied by 13C-NMR and DRIFT spectroscopy. European Journal of Soil Science, 48, 301-310. https://doi.org/10.1111/j.1365-2389.1997.tb00550.x.

Kaiser, K., & Zech, W. (2000). Dissolved organic matter sorption by mineral constituents of subsoil clay fractions. Journal of Plant Nutrition and Soil Science, 163, 531-535. https://doi.org/10.1002/1522-2624.

Kalbitz, K., Solinger, S., Park, J.-H., Michalzik, B., & Matzner, E. (2000). Controls on the dynamics of dissolved organic matter in soils: a review. Soil Science, 165, 277-304. https://doi.org/10.1097/00010694-200004000-00001.

Kruthika, S., Ashu, A., Anand, A., Sammi Reddy, K., Vara Prasad, P. V., & Gurumurthy, S. (2024). Unveiling the Role of Root Exudates in Plant Adaptation to Drought and Heat Stress. Journal of Crop Health, 76, 941-955. https://doi.org/10.1007/s10343-024-01013-8.

Lal, R. (2004). Soil carbon sequestration impacts on global climate change and food security. science, 304, 1623-1627. https://doi.org/10.1126/science.1097396.

Langeveld, J., Bouwman, A. F., van Hoek, W. J., Vilmin, L., Beusen, A. H. W., Mogollón, J. M., & Middelburg, J. J. (2020). Estimating dissolved carbon concentrations in global soils: a global database and model. Sn Applied Sciences, 2, 1626.https://doi.org/10.1007/s42452-020-03290-0.

Lee, H., Calvin, K., Dasgupta, D., Krinner, G., Mukherji, A., Thorne, P., Trisos, C., Romero, J., Aldunce, P., & Barret, K. (2023). IPCC, 2023: Climate Change 2023: Synthesis Report, Summary for Policymakers. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC, Geneva, Switzerland. 1-34. https://cir.nii.ac.jp/crid/1360019997669261824.

Li, J., Wu, B., Zhang, D., & Cheng, X. (2023). Elevational variation in soil phosphorus pools and controlling factors in alpine areas of Southwest China. Geoderma, 431, 116361. https://doi.org/10.1016/j.geoderma.2023.116361.

Li, S., Zheng, X., Liu, C., Yao, Z., Zhang, W., & Han, S. (2018). Influences of observation method, season, soil depth, land use and management practice on soil dissolvable organic carbon concentrations: A meta-analysis. Science of the Total Environment, 631-632, 105-114. https://doi.org/10.1016/j.scitotenv.2018.02.238.

Lønborg, C., Carreira, C., Jickells, T., & Álvarez-Salgado, X. A. (2020). Impacts of global change on ocean dissolved organic carbon (DOC) cycling. Frontiers in Marine Science, 7, 466. https://doi.org/10.3389/fmars.2020.00466.

Lu, S., Xu, Y., Fu, X., Xiao, H., Ding, W., & Zhang, Y. (2019). Patterns and drivers of soil respiration and vegetation at different altitudes in Southern China. Applied Ecology & Environmental Research, 17. https://doi.org/10.15666/aeer/1702_30973106.

Luo, Y., Keenan, T. F., & Smith, M. (2015). Predictability of the terrestrial carbon cycle. Global change biology, 21, 1737-1751. https://doi.org/10.1111/gcb.12766.

Michalzik, B., Kalbitz, K., Park, J. H., Solinger, S., & Matzner, E. (2001). Fluxes and concentrations of dissolved organic carbon and nitrogen – a synthesis for temperate forests. Biogeochemistry, 52, 173-205. https://doi.org/10.1023/A:1006441620810.

Nakhavali, M., Lauerwald, R., Regnier, P., Guenet, B., Chadburn, S., & Friedlingstein, P. (2021). Leaching of dissolved organic carbon from mineral soils plays a significant role in the terrestrial carbon balance. Global Change Biology, 27, 1083-1096. https://doi.org/10.1111/gcb.15460.

Neff, J. C., & Asner, G. P. (2001). Dissolved organic carbon in terrestrial ecosystems: synthesis and a model. Ecosystems, 4, 29-48. https://doi.org/10.1007/s100210000058.

Patoine, G., Eisenhauer, N., Cesarz, S., Phillips, H. R. P., Xu, X., Zhang, L., & Guerra, C. A. (2022). Drivers and trends of global soil microbial carbon over two decades. Nature Communications, 13, 4195. https://doi.org/10.1038/s41467-022-31833-z.

Perrot, T., Rusch, A., Gaba, S., & Bretagnolle, V. (2023). Both long-term grasslands and crop diversity are needed to limit pest and weed infestations in agricultural landscapes. Proceedings of the National Academy of Sciences, 120, e2300861120. https://doi.org/10.1073/pnas.2300861120.

Pesántez, J., Mosquera, G. M., Crespo, P., Breuer, L., & Windhorst, D. (2018). Effect of land cover and hydro-meteorological controls on soil water DOC concentrations in a high-elevation tropical environment. Hydrological Processes, 32, 2624-2635. https://doi.org/10.1002/hyp.13224.

Propster, J. R., Schwartz, E., Hayer, M., Miller, S., Monsaint-Queeney, V., Koch, B. J., Morrissey, E. M., Mack, M. C., & Hungate, B. A. (2023). Distinct growth responses of tundra soil bacteria to short-term and long-term warming. Applied and Environmental Microbiology, 89, e01543-01522. https://doi.org/10.1128/aem.01543-22.

Rahbek, C., Borregaard, M. K., Colwell, R. K., Dalsgaard, B., Holt, B. G., Morueta-Holme, N., Nogues-Bravo, D., Whittaker, R. J., & Fjeldså, J. (2019). Humboldt's enigma: What causes global patterns of mountain biodiversity? Science, 365, 1108-1113. https://doi.org/10.1126/science.aax0149.

Raudina, T. V., Smirnov, S. V., Lushchaeva, I. V., Istigechev, G. I., Kulizhskiy, S. P., Golovatskaya, E. A., Shirokova, L. S., & Pokrovsky, O. S. (2022). Seasonal and spatial variations of dissolved organic matter biodegradation along the aquatic continuum in the Southern Taiga bog complex, Western Siberia. Water, 14, 3969. https://doi.org/10.3390/w14233969.

Ren, C., Zhou, Z., Delgado-Baquerizo, M., Bastida, F., Zhao, F., Yang, Y., Zhang, S., Wang, J., Zhang, C., Han, X., Wang, J., Yang, G., & Wei, G. (2024). Thermal sensitivity of soil microbial carbon use efficiency across forest biomes. Nature Communications, 15, 6269. https://doi.org/10.1038/s41467-024-50593-6.

Ren, T., & Cai, A. (2024). Global patterns and drivers of soil dissolved organic carbon concentrations. Earth System Science Data Discussions., 2024, 1-25. https://doi.org/10.5194/essd-2024-343.

Ren, T., Tang, S., Han, T., Wang, B., Zhou, Z., Liang, G., Li, Y. e., & Cai, A. (2023). Positive rhizospheric effects on soil carbon are primarily controlled by abiotic rather than biotic factors across global agroecosystems. Geoderma, 430, 116337. https://doi.org/10.1016/j.geoderma.2023.116337.

Ren, T., Ukalska-Jaruga, A., Smreczak, B., & Cai, A. (2024). Dissolved organic carbon in cropland soils: A global meta-analysis of management effects. Agriculture, Ecosystems & Environment, 371, 109080. https://doi.org/10.1016/j.agee.2024.109080.

Sanders, E. D., Pereira, A., & Paulino, G. H. (2021). Optimal and continuous multilattice embedding. Science Advances, 7, eabf4838. https://doi.org/doi:10.1126/sciadv.abf4838

Schlesinger, W. H., & Bernhardt, E. S. (2020). The atmosphere. Biogeochemistry, 12, 51-97. https://doi.org/10.1016/B978-0-12-814608-8.00003-7.

Singh, M., Sarkar, B., Hussain, S., Ok, Y. S., Bolan, N. S., & Churchman, G. J. (2017). Influence of physico-chemical properties of soil clay fractions on the retention of dissolved organic carbon. Environmental Geochemistry and Health, 39, 1335-1350. https://doi.org/10.1007/s10653-017-9939-0.

Six, J., Bossuyt, H., Degryze, S., & Denef, K. (2004). A history of research on the link between (micro)aggregates, soil biota, and soil organic matter dynamics. Soil and Tillage Research, 79, 7-31. https://doi.org/https://doi.org/10.1016/j.still.2004.03.008.

Smreczak, B., & Ukalska-Jaruga, A. (2021). Dissolved organic matter in agricultural soils. Soil Science Annual. 72,1. https://doi.org/10.37501/soilsa/132234.

Sokol, N. W., Slessarev, E., Marschmann, G. L., Nicolas, A., Blazewicz, S. J., Brodie, E. L., Firestone, M. K., Foley, M. M., Hestrin, R., & Hungate, B. A. (2022). Life and death in the soil microbiome: how ecological processes influence biogeochemistry. Nature Reviews Microbiology, 20, 415-430. https://doi.org/10.1038/s41579-022-00695-z.

Wei, D., Tao, J., Wang, Z., Zhao, H., Zhao, W., & Wang, X. (2024). Elevation-dependent pattern of net CO(2) uptake across China. Nature Communications, 15, 2489. https://doi.org/10.1038/s41467-024-46930-4.

Zhao, X., Tian, P., Sun, Z., Liu, S., Wang, Q., & Zeng, Z. (2022). Rhizosphere effects on soil organic carbon processes in terrestrial ecosystems: A meta-analysis. Geoderma, 412, 115739. https://doi.org/10.1016/j.geoderma.2022.115739.

Zhou, Z., Ren, C., Wang, C., Delgado-Baquerizo, M., Luo, Y., Luo, Z., Du, Z., Zhu, B., Yang, Y., Jiao, S., Zhao, F., Cai, A., Yang, G., & Wei, G. (2024). Global turnover of soil mineral-associated and particulate organic carbon. Nature Communications, 15, 5329. https://doi.org/10.1038/s41467-024-49743-7.

Zhou, Z., Wang, C., Cha, X., Zhou, T., Pang, X., Zhao, F., Han, X., Yang, G., Wei, G., & Ren, C. (2024a). The biogeography of soil microbiome potential growth rates. Nature Communications, 15, 9472. https://doi.org/10.1038/s41467-024-53753-w.

Figure 1 Global distribution of soil dissolved organic carbon (DOC) concentration according to our site-level dataset. The dataset contains 12807 sets of data (**a, b**), which covers major wetland (1106), forest (4867), shrubland (385), tundra (130), grassland (1192), cropland (5125) terrestrial biomes (**c**). The dashed red line within the subplot (**b**) signifies the average soil DOC concentration, which is 223 mg kg^{-1}.

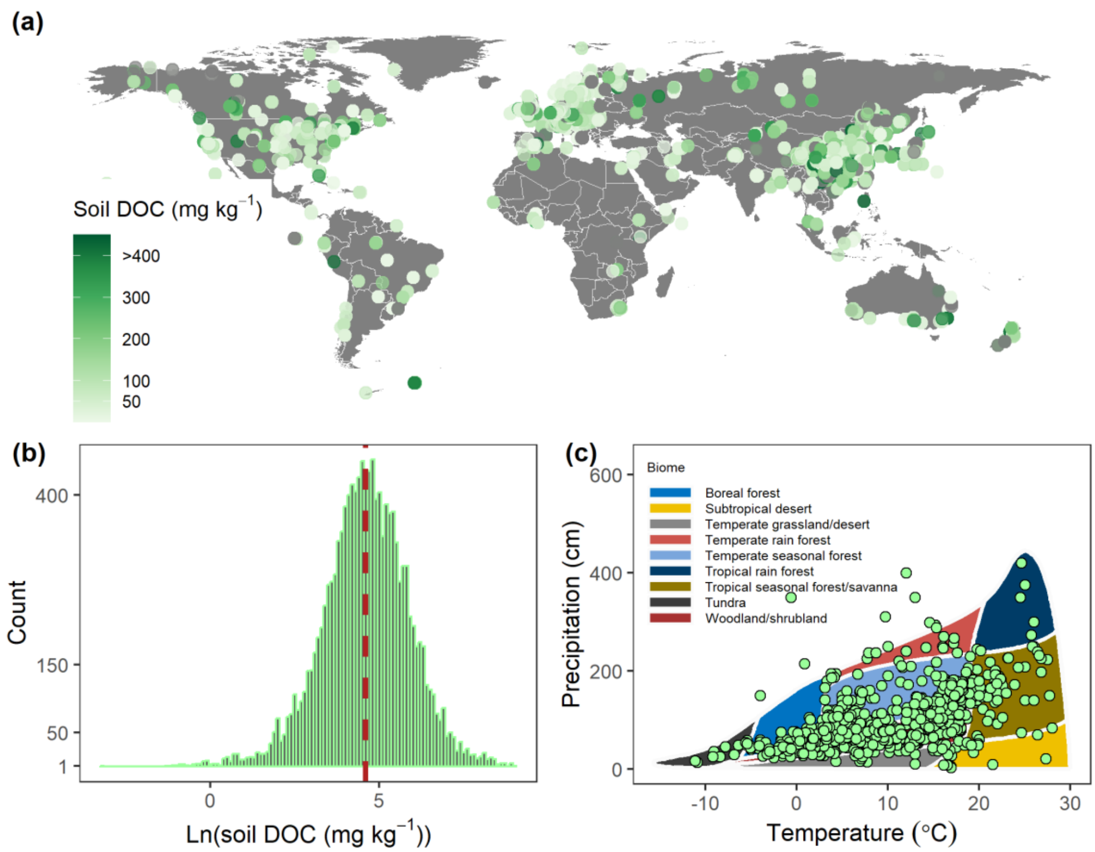

Figure 2 Result of the random forest model predicting soil dissolved organic carbon (DOC) concentration. (**a**) The relative importance of predictors in the random forest model. (**b**) Predicted vs. observed soil DOC concentration. The dashed line indicates the 1:1 line and the blue line indicates the regression line between predicted and observed values. MAT, mean annual temperature; MAP, mean annual precipitation; SVP, seasonal variability of precipitation; SVT, seasonal variability of temperature; ETM, monthly evaporation; SOC, soil organic carbon; BD, bulk density; MBC, microbial biomass carbon content; and C:N, ratio of carbon to nitrogen.

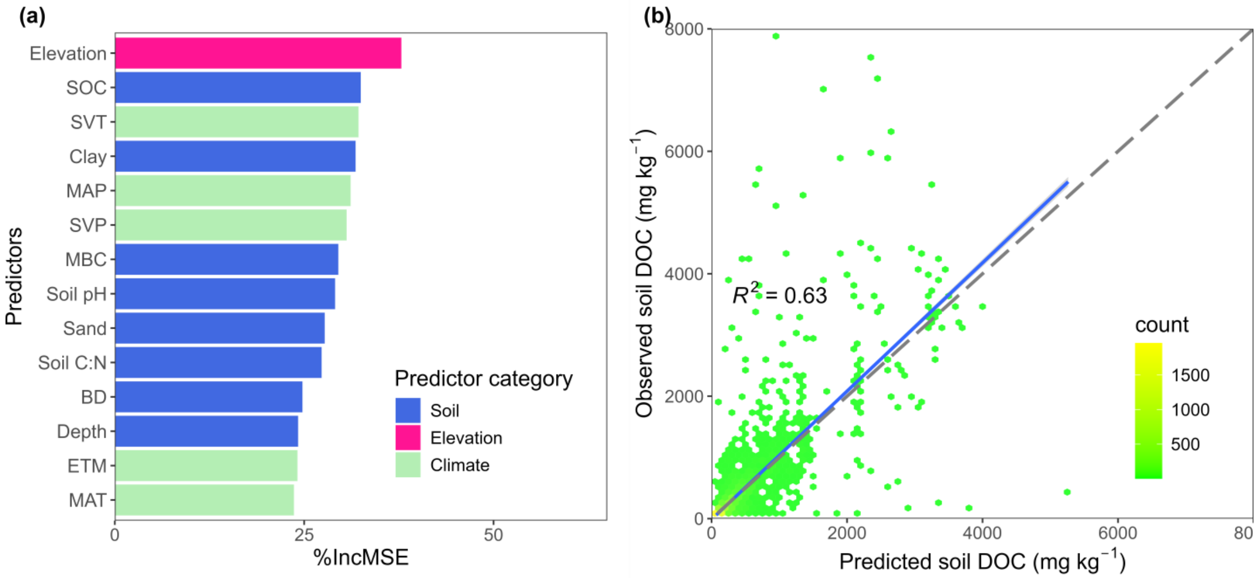

Figure 3 Partial dependence of predictors from random forest algorithm. Soil dissolved organic carbon (DOC) concentration in relation to mean annual temperature (MAT), mean annual precipitation (MAP), elevation, seasonal variability of precipitation (SVP), seasonal variability of temperature (SVT), monthly evaporation (ETM), elevation, soil sand content, soil clay content, soil depth, soil organic carbon (SOC) content, soil pH, bulk density, microbial biomass carbon content (MBC), and ration of soil carbon to nitrogen (C:N) (**a, b, c, d, e, f, g, h, i, j, k, l, m, and n** respectively). The histogram in each plot represents the data distribution of the X-axis indicator.

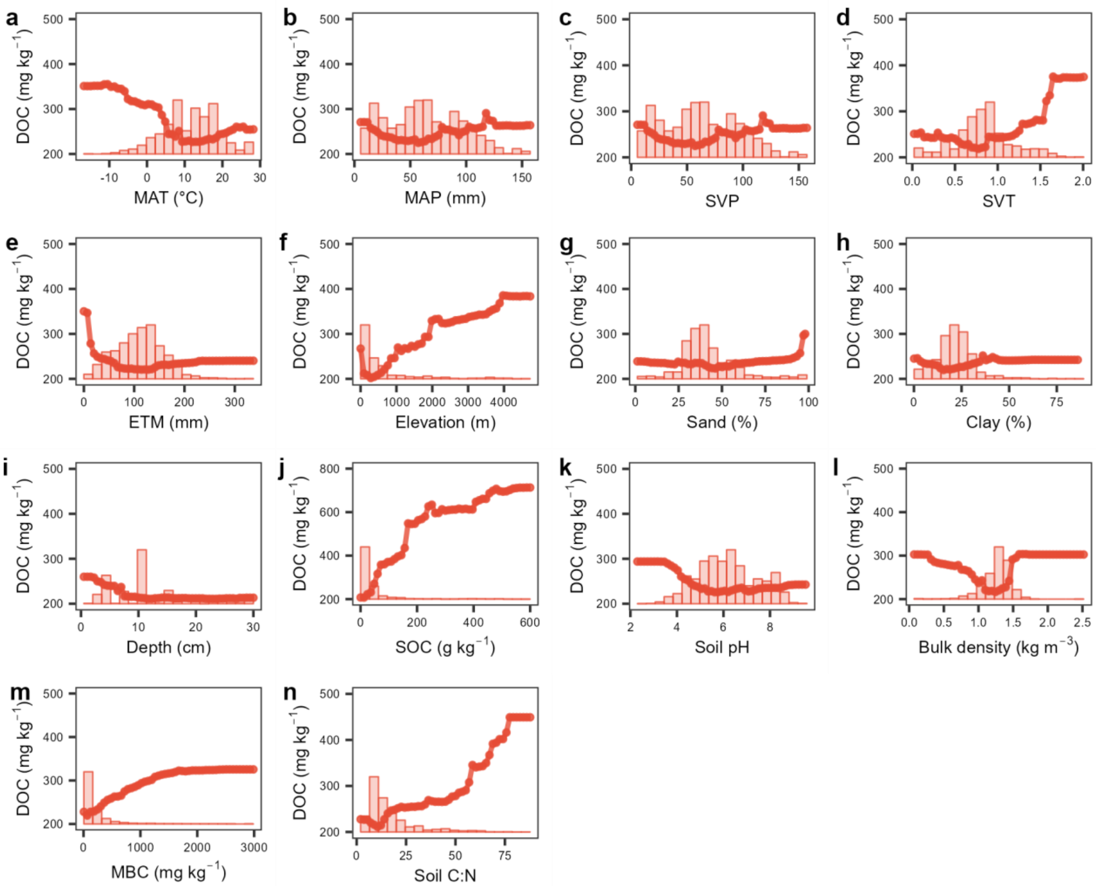

Figure 4 Interaction effects between key drivers of derived soil dissolved organic carbon concentration. key drivers included mean annual temperature (MAT), mean annual precipitation (MAP), elevation, seasonal variability of precipitation (SVP), seasonal variability of temperature (SVT), monthly evaporation (ETM), elevation, soil sand content, soil clay content, soil depth, soil organic carbon (SOC) content, soil pH, bulk density, microbial biomass carbon content (MBC), and ration of soil carbon to nitrogen (C:N).

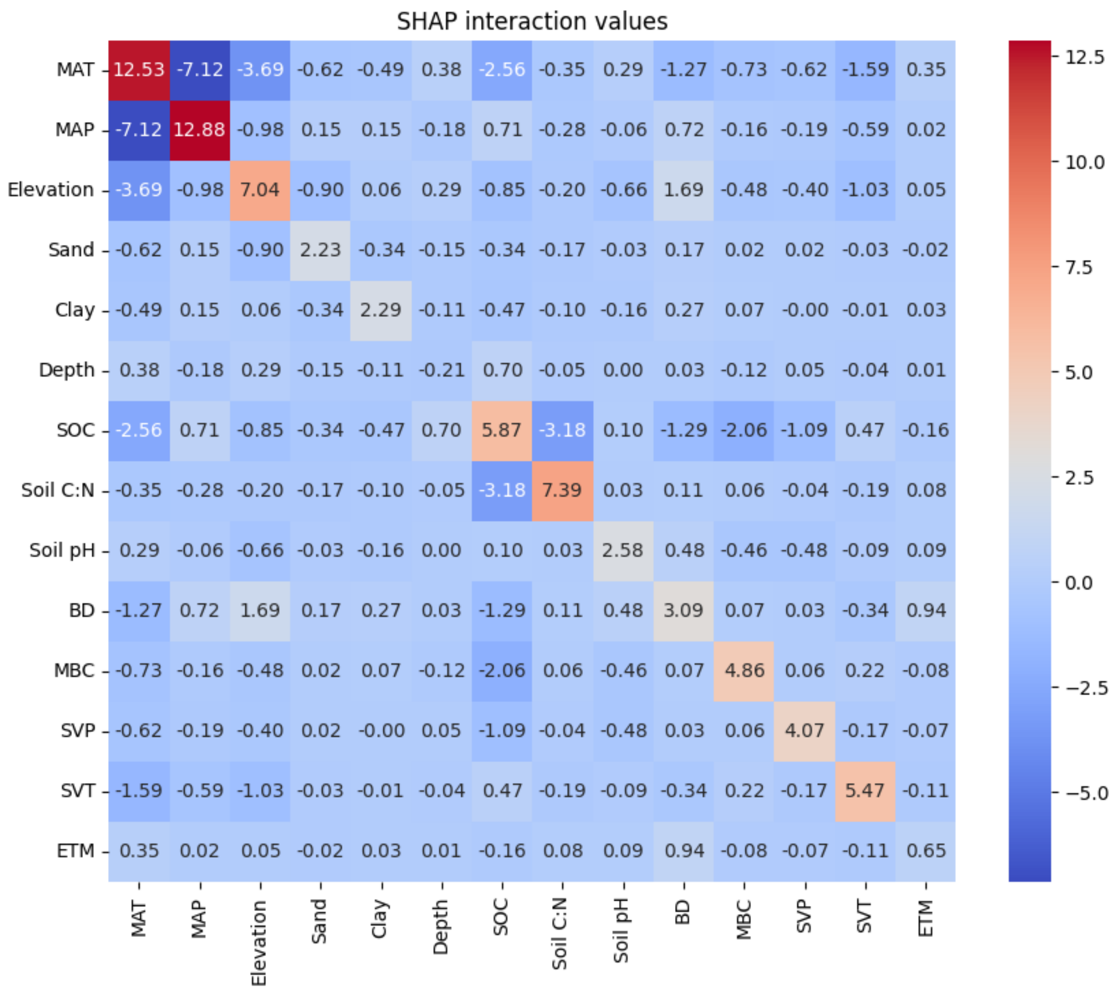

Figure 5 Prediction of soil dissolved organic carbon (DOC) concentration in global ecosystems. (**a**) Global map of predicted soil DOC concentration. (**b**) Latitudinal patterns of soil DOC concentration. Blue line indicates the locally weighted regressions between latitude and soil DOC concentration in the predicted global map. Values in the predicted map reflect soil DOC concentration within a grid cell resolution of 0.05° × 0.05°. A value in the grid is the averaged from the result of random forest model.

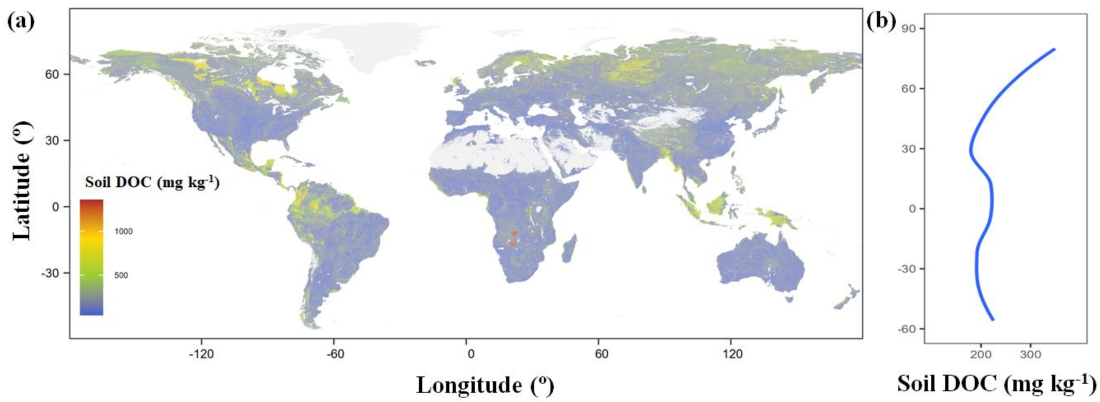

Table 1. Variables information of soil dissolved organic carbon dataset in global terrestrial ecosystems. n/a refers to values that are not applicable.

Variables	Description	Unit	Number	Range	Mean
No.	Unique identification number of each record	n/a	12807	1 to 12807	6404
Latitude	Latitude of study site	°	12807	-64.81 to 78.85	34.89
Longitude	Latitude of study site	°	12807	-159.66 to 175.95	107.05
MAT	Mean annual temperature	℃	9948	-11.16 to 28.00	11.84
MAP	Mean annual precipitation	mm	10325	30 to 4200	1071
Elevation	Altitude of study site	m	5578	4 to 4730	881
Ecosystems	Community by the dominant plant species		7	n/a	n/a
Soil sand	Soil sand content	%	4062	1 to 98	45
Soil silt	Soil silt content	%	4025	1 to 95	33
Soil clay	Soil clay content	%	4316	0 to 89	22
Soil depth	Mean depth of soil sample	cm	12807	0.53 to 30.00	11.36
SOC	Soil organic carbon	g kg^{-1}	9136	0.23 to 598.50	38.74
TN	Soil total nitrogen	g kg^{-1}	7089	0.00 to 33.30	2.57
Soil pH	Measure by 1:2.5 H_2O,	n/a	8266	2.30 to 9.59	6.16
BD	Soil bulk density	kg m^{-3}	4380	0.07 to 2.52	1.29
MBC	Soil microbial biomass carbon	mg kg^{-1}	4218	5.93 to 2986	413
Date	Observation month of DOC	month	12807	1 to 12	6.50
DOC_{phy}	Measure by physical method	mg kg^{-1}	3289	0.28 to 3181	155.99
DOC_{che}	Measure by chemical process	mg kg^{-1}	9518	0.04 to 7859	245.83
DOC	Soil dissolved organic carbon	mg kg^{-1}	12807	0.04 to 7859	222.78

Table 2. Global soil dissolved organic carbon concentration (mg kg-1) for major ecosystems. 25% and 75% represent the 25th and 75th percentiles of one group, respectively. SD, Standard deviation; SE, Standard error.

Ecosystems	Mean	SD	SE	Skewness	Kurtosis	25%	Median	75%
Wetland	218.53	340.35	10.23	5.15	39.41	46.40	107.11	266.51
Forest	256.18	531.72	7.62	7.09	69.72	47.60	115.51	246.55
Shrubland	160.24	131.51	6.70	3.40	22.58	76.53	127.84	205.50
Tundra	470.78	721.70	63.30	4.67	29.59	86.91	241.09	577.00
Grassland	327.77	674.43	19.53	4.16	18.03	54.62	126.48	303.63
Cropland	165.98	272.51	3.81	6.53	73.25	40.51	83.00	178.81
Global	222.78	445.78	3.93	7.16	73.67	45.86	101.01	226.47

Table 3. Analysis of the predicted global map of soil dissolved organic carbon. The area-weighted average soil dissolved organic carbon concentration was calculated based on our predicted map. Converting soil dissolved organic carbon concentration to soil dissolved organic carbon content and stock used the soil bulk density and land area.

Continent	Soil DOC concentration ($mg\ kg^{-1}$)	Soil DOC content ($g\ m^{-2}$)	Soil DOC stock (Pg)
Asia	259.03	103.26	4.93
North America	250.66	111.29	2.93
Europe	208.28	89.97	0.98
South America	219.83	92.33	1.76
Oceania	206.36	91.62	0.76
Africa	166.73	72.77	2.37
Global	224.72	97.75	13.74