The age distribution of global soil carbon inferred from radiocarbon measurements

Zheng Shi[1,2*], Steven D. Allison[1,2], Yujie He[2], Paul A. Levine[2], Alison M. Hoyt[3], Jeff Beem-Miller[3], Qing Zhu[4], William R. Wieder[5], Susan Trumbore[3], James T. Randerson[2]

[1] Department of Ecology and Evolutionary Biology, University of California Irvine, Irvine, CA, United States

[2] Department of Earth System Science, University of California Irvine, Irvine, CA, United States

[3] Department of Biogeochemical Processes, Max-Planck-Institute for Biogeochemistry, Jena, Germany

[4]Climate and Ecosystem Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

[5]Climate and Global Dynamics Laboratory, National Center for Atmospheric Research, Boulder, CO, USA

*Correspondence to: zshi7@uci.edu

Soils contain more carbon as organic material than the atmosphere and vegetation combined, so increased flow of carbon from the atmosphere into soil pools could help mitigate anthropogenic CO_2 emissions and climate change. Yet we do not know how quickly soils might respond because the age distribution of soil carbon is uncertain. Here we used 789 radiocarbon ($\Delta^{14}C$) profiles, along with other geospatial information, to create a globally-gridded dataset of mineral soil $\Delta^{14}C$ and mean age. We find that soil depth is a primary driver of $\Delta^{14}C$, whereas climate (e.g. mean annual temperature) is a major control on the spatial pattern of $\Delta^{14}C$ in surface soil. Integrated to a depth of 1-meter, global soil carbon has a mean age of 4830±1730 years, with older carbon in deeper layers and permafrost regions. In contrast, vertically-resolved land models simulate $\Delta^{14}C$ values that imply younger carbon ages and more rapid carbon turnover. Our data-derived estimates of older mean soil carbon age suggest that soils will accumulate less carbon than predicted by current Earth system models over the 21st century. Reconciling these models with the global distribution of soil radiocarbon will require better representation of the mechanisms controlling carbon persistence in soils.

Soils offer promise for carbon sequestration. Elevated atmospheric CO_2 concentration, nitrogen deposition, and improved land management can increase vegetation production[1,2], leading to increased soil carbon storage. Initiatives such as “4 per mille”—0.4% annual growth of soil organic carbon with improved agricultural practice—depend on this carbon storage potential to mitigate climate warming[3]. Land surface models that include CO_2 fertilization often predict soil carbon accumulation even under the highest radiative forcing scenario[4]. On the other hand, experimental and chronosequence studies have shown limited soil carbon sequestration despite increased carbon input[5-7], and soils may lose carbon due to warming and land use change[8,9]. Therefore, whether increased plant productivity will increase soil carbon storage in a warming climate remains uncertain.

Accurately estimating the age of carbon in soils is critical for evaluating sequestration potential. To be useful for CO_2 emissions mitigation, soil carbon pools must react to increased carbon inputs on decadal to centennial timescales. Assuming first-order loss rates remain constant, increases in carbon inputs eventually lead to a proportional increase in carbon stock. To a first approximation, older carbon pools, with mean ages of thousands to tens of thousands of years, have substrate inputs and outputs that are small compared to the total amount of carbon stored in the pool[5]. With these pools, it can take thousands of years for carbon to accumulate. In contrast, young carbon pools with mean ages of decades to a few centuries can accumulate new carbon more quickly. While these pools could sequester carbon on timescales relevant for climate mitigation, their smaller sizes and higher rates of carbon turnover may limit carbon storage capacity.

Radiocarbon measurements can be used to estimate rates of soil carbon cycling on decadal to millennial timescales[10]. Fast-cycling soil carbon pools derived from the atmosphere during the last few decades show a fingerprint of "bomb" carbon from atmospheric weapons testing[11]. By contrast, natural radiocarbon decay provides information about soil carbon cycling on timescales from centuries to millennia.

Leveraging these principles, we analyzed 789 vertical soil profiles from the International Soil Radiocarbon Database (ISRaD)[12]. This approach builds on an analysis by He et al.[13] in which Earth system models constrained by soil radiocarbon predicted less carbon uptake in response to rising atmospheric CO_2. Their analysis raised questions about the environmental drivers of soil radiocarbon and how those drivers are represented in earth system models. To address these questions, we leveraged the new ISRaD database to generate the first global, spatially- and depth-resolved data product for soil radiocarbon. We used the data product to calculate the age distribution of global soil carbon, analyze the environmental drivers of biome-level variability in soil radiocarbon, and test predictions from state-of-the-art earth system models.

We express soil radiocarbon as $\Delta^{14}C$, the difference in $^{14}C/^{12}C$ ratio between the sample and an absolute standard expressed in parts per thousand[14]. Positive $\Delta^{14}C$ indicates the presence of bomb carbon, whereas negative $\Delta^{14}C$ indicates that radioactive decay of ^{14}C overwhelms any incorporation of bomb carbon into the sample. Radiocarbon measurements covered all major land biomes (Supplementary Fig. 1a) with a wide range of mean annual temperature and precipitation (Supplementary Fig. 1b). Most of the soil profiles reported in ISRaD were sampled in the first 100 cm during 1990-2010, and 75% of the profiles included more than one vertical horizon (Supplementary Fig. 2).

Relative importance of the environmental drivers

To produce globally-gridded maps of $\Delta^{14}C$ and age, we used a machine learning approach that linked measurements of soil $\Delta^{14}C$ with variation in environmental factors (see Methods). Because soil sampling date affects $\Delta^{14}C$, we used a one-pool model to normalize all the $\Delta^{14}C$

measurements to the year 2000, around which most of the data were collected (Supplementary Fig. 2a), before conducting the statistical analysis (see Methods). A random forest model showed that depth was the primary control on soil $\Delta^{14}C$, followed by mean annual temperature and precipitation (Supplementary Fig. 3a). Soil $\Delta^{14}C$ decreased with greater soil depth and increased with greater mean annual temperature and precipitation (Supplementary Fig. 4). Mechanisms driving the decline in $\Delta^{14}C$ with depth could be changes in microbial activity, smaller carbon substrate inputs from plants, and increased carbon stabilization by mineral sorption[15,16]. Soil depth and clay content may be important proxies for physical protection as suggested in previous studies[17]. However, the minor role of clay content in our analysis suggests that other depth-dependent variables such as the type of clay, cation exchange capacity[18], and mineral chemistry[19,20] may be more important determinants of soil $\Delta^{14}C$. Further investigation into these mechanisms would advance our predictive understanding of soil carbon dynamics.

For surface soils (0 – 30 cm), mean annual temperature was a dominant control on the spatial variation of $\Delta^{14}C$ (Supplementary Fig. 3b). Mechanistically, warmer temperatures may allow for a longer growing season, higher levels of net primary production, greater soil carbon inflows, and more rapid decomposition of labile carbon pools that are not closely bound to mineral surfaces. The importance of this variable is consistent with previous work showing that climate regulates the global spatial pattern of turnover times for ecosystem carbon[21] and soil carbon[22]. For deeper soils (with a depth greater than 30 cm), $\Delta^{14}C$ was mainly controlled by depth, but also by temperature, precipitation, and clay content (Supplementary Fig. 3c). Depth may have emerged from the model as a more important factor than temperature in this layer because of a greater vertical range that includes more variation in soil mineral content, vertical transport processes, and carbon inputs from root turnover.

Global soil radiocarbon $\Delta^{14}C$

Based on the relationships with environmental drivers in our random forest model, we scaled up $\Delta^{14}C$ measurements from individual soil profiles to create global maps (Methods). Soils had less negative (or more positive) values of carbon-weighted $\Delta^{14}C$ in tropical regions than in temperate and boreal regions (Fig. 1, a to c; Supplementary Fig. 5). Carbon in subsurface soils consistently had more negative $\Delta^{14}C$ than carbon in surface soils (Fig. 1, b and c). Most surface soils in the tropics had a $\Delta^{14}C$ greater than 0‰ (Fig. 1b), whereas all subsurface soils had negative $\Delta^{14}C$ values (Fig. 1c). The carbon-weighted $\Delta^{14}C$ was -244±48‰ globally, with values of -97±24‰ in surface soil and -391±56‰ in subsurface soil (Table 1).

Mean annual temperature structured the spatial variation of $\Delta^{14}C$ in our global maps, with a sharp increase near -4°C and then further, more gradual increases between 0 and 25°C (Supplementary Fig. 6a). Among different biomes, tundra had the most negative $\Delta^{14}C$, with median values of -249‰ and -624‰, respectively, for surface and subsurface soils. Tropical forests had the greatest $\Delta^{14}C$ in surface soils with a median value of 7‰, and intermediate values in subsurface soils, with a median of -250‰. Permafrost soils had considerably more negative $\Delta^{14}C$ than non-permafrost soils (Table 1). In addition to temperature, mean annual precipitation also influenced $\Delta^{14}C$ at a regional scale. For example, drier grasslands and shrublands, and wetter boreal and temperate forests had more negative $\Delta^{14}C$ (Supplementary Fig. 6b).

The depth profiles of soil $\Delta^{14}C$ also differed among biomes (Supplementary Fig. 7). Tundra and boreal forest ecosystems had much stronger depletion of radiocarbon in deeper soil layers where sub-zero temperatures and permafrost processes regulate carbon cycling[23]. In deep tropical forest soils, the random forest model was not able to fully capture low observed $\Delta^{14}C$ values—which occur despite warm temperatures—suggesting that more detailed information about vertical transport and mineral stabilization mechanisms is needed in future modeling efforts.

Mean age of global soil carbon

To convert $\Delta^{14}C$ into mean soil carbon age, we fit a one-pool carbon model to the $\Delta^{14}C$ estimate in each grid cell and depth interval using the time series of atmospheric $\Delta^{14}C$ over the past 50 ky[24] (Methods). Globally, the carbon-weighted mean age of mineral soil carbon was 4830±1730 (mean ± standard deviation) years between 0 and 100 cm depth (Table 1). Surface soils (0 – 30 cm) had a younger mean age (1390±310 years) than subsurface soils (8280±2820 years; 30 – 100 cm). Use of a two-pool model to estimate mean age yielded similar but slightly older estimates (Supplementary Fig. 8). Mean age varied as a function of latitude (Fig.1, d to f and Supplementary Fig. 9), mean annual temperature (Supplementary Fig. 10a) and among biomes (Table 1, Supplementary Fig. 10b). Our estimated age distribution for soil carbon in tropical forests was comparable to another independent estimate derived from ^{13}C[25]. In permafrost regions, soil carbon ages were considerably older than in other regions, ranging from about 2800 years for the surface layer to over 15,000 years for the subsurface layer (Table 1, Fig. 1, e and f, and Fig. 2). To a depth of 100 cm, about 24% (450 Pg out of 1848 Pg) of global soil carbon was younger than 1000 years (Fig. 2), with nearly all of this carbon confined to the 0 – 30 cm surface layer (Figs. 1e and 2). In contrast, nearly all subsurface soil carbon (1005 Pg out of 1008 Pg) was older than 1000 years (Figs. 1F and 2), meaning that it is probably unresponsive to changes in carbon inputs from 21st century global environmental change.

Comparisons between models and data

The global three-dimensional structure of soil $\Delta^{14}C$ provides a new way to constrain land surface models that resolve soil carbon vertically. We compared our gridded $\Delta^{14}C$ dataset with two state-of-the-art global land models, version 5 of the Community Land Model (CLM5)[26] and version 1.0 of the land model within the Energy Exascale Earth System Model (ELM v1.0)[27]. Compared to the gridded dataset, the land surface models overestimated $\Delta^{14}C$ in both surface and subsurface soil layers (Fig. 3, and Supplementary Figs. 11-13), and in each biome (Supplementary Tables 1-2). In surface soils, over 60% of carbon in each of the models had positive $\Delta^{14}C$ values compared to only about 14% of carbon in the globally gridded dataset (Fig. 3, a, c and e). The two models also predicted that about 50% of subsurface soil carbon had $\Delta^{14}C$ more positive than -200‰ (Fig. 3, d and f), whereas this amount was less than 10% in the data-derived product (Fig. 3b). The over-estimation in the two models occurred in all biomes, with larger positive biases in tropical biomes and smaller positive biases in boreal forest and tundra biomes.

The over-estimation of $\Delta^{14}C$ in the two models is likely a consequence of positive biases in fresh carbon inflows at depth, vertical substrate diffusion[28], and carbon turnover in slow and passive carbon pools[13]. The two models employ a similar decomposition cascade whereby plant litter passes through pools with successively longer turnover times. Moreover, aboveground litterfall is distributed throughout the soil column following rooting depth profiles for each plant functional type[29], and this parameterization may provide a larger than expected input of modern soil carbon to deeper soil horizons.

Differences in other model parameters result in distinct spatial distributions of soil carbon stocks and ages. Specifically, ELM uses a smaller value for z_τ, the e-folding depth that reduces the intrinsic decomposition rate for soil carbon in deeper soil horizons[29], whereas CLM5 has higher z_τ, but applies stronger soil moisture limitations on decomposition[26,30]. Globally, the ELM parameterization provides more negative $\Delta^{14}C$ values, especially in deeper soils (Supplementary Figs. 11, 13; and Tables 1-2), albeit not for mechanistically satisfying reasons. To match the ^{14}C observations, our analysis suggests the models should retain a smaller fraction of fresh litter inputs in soil carbon pools with long turnover times. Also, the turnover times of these 'slow' or 'passive' pools that comprise the majority of soil carbon should be much greater. In developing improved models, however, a mechanistic representation of carbon cycling is needed that recognizes the potential vulnerability of key reservoirs, including carbon stored in permafrost soils[8,23].

Although soil carbon is heterogeneous, consisting of multiple fast- and slow-cycling pools, our $\Delta^{14}C$ data provide a key constraint on the slow pools that make up the bulk of soil organic carbon. Previous estimates of turnover based on the ratio of carbon stocks to inputs[22,30] imply faster cycling and younger ages of soil carbon compared to our results. The discrepancy arises because most net primary production cycles through relatively small soil carbon pools on timescales of years to decades. Such a "leaky" response to increased carbon input is also supported by empirical studies[5-7]. The bulk of soil carbon, in contrast, is supplied by a very slow trickle of inputs that are stabilized on millennial timescales. However, CLM5 and ELM both assume that a larger fraction of recent photosynthate is retained in the soil system as indicated by their positive biases in $\Delta^{14}C$ (Fig. 3). Due to these biases, the global models may be too responsive to new carbon inputs and may over-estimate the effect of CO_2 fertilization on productivity and potential soil carbon sequestration[13]. The millennium-scale mean age of global soil carbon, coupled with limited retention of bomb carbon over the past 70 years, implies that soil carbon is unlikely to increase as much as predicted in land surface models with CO_2 fertilization over the next few decades. Nevertheless, the depth-resolved models are better at predicting soil carbon age compared to models that omit soil depth[31], and a clear path now exists for improving these models using observations from ISRaD[12].

Despite its old age, soil carbon in many ecosystems may still be vulnerable to climate and land use change. For example, permafrost thaw in tundra and boreal forest may allow for the rapid decomposition and release of previously protected deep soil carbon[8]. Similarly, disturbance associated with the expansion of global agriculture accelerates decomposition through the physical destruction of soil aggregates and by exposing deep soil carbon to microbial decay[9,35]. More frequent and severe fire disturbance can also contribute to losses of soil carbon[36].

For more than 25 years, soil science has upheld a paradigm that mineral soil carbon mainly consists of pools with decadal and centennial turnover times. Despite a growing awareness of old soil carbon stabilized in deep soils, expert assessments and influential models such as Century have considered carbon with millennium turnover times to be a relatively small fraction of bulk soil organic matter[32,33]. Yet we show that in deeper soils, which represent more than half of the global soil carbon stock, pools with multi-millennium ages are dominant, yielding a global mean deep-soil age over 8000 years. Even in surface soils from 0-30 cm, our mean age estimate of over 1300 years suggests that millennial-scale carbon pools may equal or exceed centennial pools. Future work could further constrain the distribution of turnover times by combining data (such as respiration[34]) that constrain faster C pools with bulk soil isotopic measurements[25].

Our study shows that old soil carbon pools identified in site-level studies extend to the global scale and that soil carbon is older than predicted by state-of-the-art earth system models. Radiocarbon age can serve as a critical, independent benchmark that will improve model predictions of soil carbon turnover and storage as climate changes. Such improvements will require that models represent mechanisms consistent with radiocarbon measurements, particularly the stabilization of deep, old soil carbon. In addition, the spatial patterns revealed in our analyses should catalyze new research to uncover fundamental mechanisms of soil carbon preservation and loss around the globe.

References

1 Zhu, Z. *et al.* Greening of the Earth and its drivers. *Nature Climate Change* **6**, 791-795 (2016).

2 Schimel, D., Stephens, B. B. & Fisher, J. B. Effect of increasing CO_2 on the terrestrial carbon cycle. *Proceedings of the National Academy of Sciences* **112**, 436-441 (2015).

3 Minasny, B. *et al.* Soil carbon 4 per mille. *Geoderma* **292**, 59-86 (2017).

4 Todd-Brown, K. E. O. *et al.* Changes in soil organic carbon storage predicted by Earth system models during the 21st century. *Biogeosciences* **11**, 2341-2356 (2014).

5 Schlesinger, W. H. Evidence from chronosequence studies for a low carbon-storage potential of soils. *Nature* **348**, 232-234 (1990).

6 van Groenigen, K. J. *et al.* Faster turnover of new soil carbon inputs under increased atmospheric CO_2. *Global Change Biology* **23**, 4420-4429 (2017).

7 Richter, D. D., Markewitz, D., Trumbore, S. E. & Wells, C. G. Rapid accumulation and turnover of soil carbon in a re-establishing forest. *Nature* **400**, 56-58 (1999).

8 Plaza, C. *et al.* Direct observation of permafrost degradation and rapid soil carbon loss in tundra. *Nat Geosci* (2019).

9 Sanderman, J., Hengl, T. & Fiske, G. J. Soil carbon debt of 12,000 years of human land use. *Proceedings of the National Academy of Sciences* **114**, 9575-9580 (2017).

10 Trumbore, S. Radiocarbon and soil carbon dynamics. *Annual Review of Earth and Planetary Sciences* **37**, 47-66 (2009).

11 Torn, M. S., Trumbore, S. E., Chadwick, O. A., Vitousek, P. M. & Hendricks, D. M. Mineral control of soil organic carbon storage and turnover. *Nature* **389**, 170-173 (1997).

12 Lawrence, C. R. *et al.* An open-source database for the synthesis of soil radiocarbon data: International Soil Radiocarbon Database (ISRaD) version 1.0. *Earth Syst. Sci. Data* **12**, 61-76 (2020).

13 He, Y. *et al.* Radiocarbon constraints imply reduced carbon uptake by soils during the 21st century. *Science* **353**, 1419-1424 (2016).

14 Stuiver, M. & Polach, H. A. Discussion reporting of ^{14}C data. *Radiocarbon* **19**, 355-363 (1977).

15 Hemingway, J. D. *et al.* Mineral protection regulates long-term global preservation of natural organic carbon. *Nature* **570**, 228-231 (2019).

16 Kramer, M. G. & Chadwick, O. A. Climate-driven thresholds in reactive mineral retention of soil carbon at the global scale. *Nature Climate Change* **8**, 1104-1108 (2018).

17 Xu, X. *et al.* Soil properties control decomposition of soil organic carbon: results from data-assimilation analysis. *Geoderma* **262**, 235-242 (2016).

18 Rasmussen, C. *et al.* Beyond clay: towards an improved set of variables for predicting soil organic matter content. *Biogeochemistry* **137**, 297-306 (2018).

19 Lalonde, K., Mucci, A., Ouellet, A. & Gélinas, Y. Preservation of organic matter in sediments promoted by iron. *Nature* **483**, 198-200 (2012).

20 Gentsch, N. *et al.* Temperature response of permafrost soil carbon is attenuated by mineral protection. *Global Change Biology* **24**, 3401-3415 (2018).

21 Carvalhais, N. *et al.* Global covariation of carbon turnover times with climate in terrestrial ecosystems. *Nature* **514**, 213-217 (2014).

22 Fan, N. *et al.* Apparent ecosystem carbon turnover time: uncertainties and robust features. *Earth Syst. Sci. Data Discuss.* **2020**, 1-25 (2020).

23 Schuur, E. A. G. *et al.* Climate change and the permafrost carbon feedback. *Nature* **520**, 171-179 (2015).

24 Reimer, P. J. *et al.* IntCal13 and Marine13 radiocarbon age calibration curves 0–50,000 years cal BP. *Radiocarbon* **55**, 1869-1887 (2013).

25 Balesdent, J. *et al.* Atmosphere–soil carbon transfer as a function of soil depth. *Nature* **559**, 599-602 (2018).

26 Lawrence, D. M. *et al.* The Community Land Model Version 5: description of new features, benchmarking, and impact of forcing uncertainty. *Journal of Advances in Modeling Earth Systems* **11**, 4245-4287 (2019).

27 Zhu, Q. *et al.* Representing nitrogen, phosphorus, and carbon interactions in the E3SM Land Model: development and global benchmarking. *Journal of Advances in Modeling Earth Systems* **11**, 2238-2258 (2019).

28 Chen, J. *et al.* Comparison with global soil radiocarbon observations indicates needed carbon cycle improvements in the E3SM Land Model. *Journal of Geophysical Research: Biogeosciences* **124**, 1098-1114 (2019).

29 Koven, C. D. *et al.* The effect of vertically resolved soil biogeochemistry and alternate soil C and N models on C dynamics of CLM4. *Biogeosciences* **10**, 7109-7131 (2013).

30 Koven, C. D., Hugelius, G., Lawrence, D. M. & Wieder, W. R. Higher climatological temperature sensitivity of soil carbon in cold than warm climates. *Nature Climate Change* **7**, 817-822 (2017).

31 Sierra, C. A., Hoyt, A. M., He, Y. & Trumbore, S. E. Soil organic matter persistence as a stochastic process: age and transit time distributions of carbon in soils. *Global Biogeochemical Cycles* **32**, 1574-1588 (2018).

32 Davidson, E. A. & Janssens, I. A. Temperature sensitivity of soil carbon decomposition and feedbacks to climate change. *Nature* **440**, 165-173 (2006).

33 Parton, W. J., Stewart, J. W. B. & Cole, C. V. Dynamics of C, N, P and S in grassland soils: a model. *Biogeochemistry* **5**, 109-131 (1988).

34 Bond-Lamberty, B., Bailey, V. L., Chen, M., Gough, C. M. & Vargas, R. Globally rising soil heterotrophic respiration over recent decades. *Nature* **560**, 80-83 (2018).

35 Hobley, E., Baldock, J., Hua, Q. & Wilson, B. Land-use contrasts reveal instability of subsoil organic carbon. *Global Change Biology* **23**, 955-965 (2017).

36 Pellegrini, A. F. A. *et al.* Fire frequency drives decadal changes in soil carbon and nitrogen and ecosystem productivity. *Nature* **553**, 194-198 (2018).

Correspondence author: Zheng Shi (email: zshi7@uci.edu)

Acknowledgements

This work was supported by the European Research Council (Horizon 2020 Research and Innovation Programme, grant agreement 695101, to SET and JTR), by the US DOE Office of Science Biological and Environmental Research RUBISCO Science Focus Area (to JTR and QZ) and award DE-SC0014374 (to SDA and JTR), and by a NASA Earth and Space Science Fellowship (to PAL).

Author contributions

ZS, YH, SDA, ST, and JTR designed the study; ZS and YH analyzed the data using machine learning and other approaches; PAL, WRW, and QZ provided analysis of the land surface models; JB-M, AMH, PAL, and SET contributed to the development of the version of the ISRaD dataset used here; ZS, SDA, and JTR wrote the paper with significant contributions from all of the authors.

Competing interests

Authors declare no competing interests.

Figure captions

Fig. 1. Global distribution of soil $\Delta^{14}C$ and mean carbon age. Carbon-weighted average $\Delta^{14}C$ and mean age in the top 1 meter (**a** and **d**), surface soil (0 – 30 cm; **b** and **e**) and subsurface soil (30 – 100 cm; **c** and **f**) are shown at a 0.5° × 0.5° spatial resolution, derived from a random forest model trained with 789 soil radiocarbon profiles.

Fig. 2. Age distribution of global soil carbon. The histogram shows the distribution of mean carbon ages derived from the globally gridded $\Delta^{14}C$ dataset for surface (0-30 cm, blue) and subsurface (30-100 cm, green) layers. Soil carbon content was estimated from the mean of two global databases, the Harmonized World Soil Database and SoilGrids.

Fig. 3. Comparison of land surface model predictions of soil $\Delta^{14}C$ with the data-derived product developed here for different depths and biomes. Histograms show the distribution of soil carbon proportion in each biome as a function of $\Delta^{14}C$ for the data-derived product (panels **a** and **b**) and for the two global land models (ELM and CLM; **c** – **f**). for the two depth intervals. Comparisons for surface soils (0-30 cm) are shown for panels in the left column and comparisons for subsurface soils (30-100cm) are shown in the right column.

Table 1. Summary statistics of soil carbon, $\Delta^{14}C$ in year 2000, and mean carbon age in each biome. The values of $\Delta^{14}C$ and mean age for each biome (and for permafrost and non-permafrost regions) are the median and 5% to 95% range (in parentheses). Global mean and standard deviation (mean ± sd) of $\Delta^{14}C$ and mean age is weighted by soil carbon content in each biome and soil layer. Mean and standard deviation of soil carbon content for each biome were derived

from two global carbon datasets (Harmonized World Soil Database and SoilGrids) as described in the methods.

Methods

1. Data source and processing

We analyzed soil $\Delta^{14}C$ measurements from the International Soil Radiocarbon Database (ISRaD). ISRaD is an open community repository for soil radiocarbon data[12]. The $\Delta^{14}C$ we used is from soil organic carbon, and not total carbon, which would include carbonates.

We retrieved $\Delta^{14}C$ measurements from ISRaD v1.0.0 on September 24, 2019 (doi: https://doi.org/10.5281/zenodo.2613911; ISRaD_extra data product, v1-2019-09-24). The dataset consisted of 789 mineral soil profiles (organic horizons were not included) from around the world for the major land cover types we used in our analysis (Supplementary Fig. 1). Each profile had on average 4 individual samples representing different depths, yielding a total of 3335 unique $\Delta^{14}C$ measurements. Metadata were also collected along with each profile, including climate (mean annual temperature and precipitation), land cover type, soil properties (soil depth, soil order, and clay content at different depth), sampling year, and location (longitude, latitude). We note that peatland and desert soil profiles are under-represented and were excluded from the dataset.

We processed the radiocarbon data in the following steps.

i. We standardized the radiocarbon reporting nomenclature. In some studies, ^{14}C activity was reported as fraction modern (F_m). In such cases, we converted F_m to $\Delta^{14}C$ (equation 1) and used $\Delta^{14}C$ as the common unit.

$$\Delta^{14}C = [F_m \times e^{\lambda(1950-Yc)} - 1] \times 1000 \qquad (1)$$

Where λ is 1/ (true mean-life) of radiocarbon = 1/8267 = 0.00012097. Y_c is the year of collection.

When uncalibrated radiocarbon ages were reported, they were converted to fraction modern values using

$$F_m = e^{(-age/8033)} \qquad (2)$$

and F_m was converted to $\Delta^{14}C$ using equation 1. Data reported as calibrated dates were not included. These calculations were performed within the ISRaD_extra data product.

ii. When the sampling year was not reported, we assumed it was the publication year minus 3 based on the mean interval from articles reporting both sampling and publication year.

iii. When the mean annual temperature and precipitation were not reported, we extracted ten-year average temperature and precipitation data (1990 – 2000) from a global-gridded database (Climatic Research Unit, Harris et al. 2014) using the geographic coordinates of each site location.

iv. We assigned one of 8 land cover types using the site description when available. Land cover types were tundra, boreal forest, temperate forest, tropical forest, grassland, shrubland, savanna and cropland (Supplementary Fig. 14). See section 2 for details on categorizing the land cover types.

v. When soil clay content was not reported, we extracted it from the SoilGrids database[37] using the geographic coordinates of each site location and depth. Note that the SoilGrids database has been updated (December 24th, 2018) and data are available at https://landgis.opengeohub.org.

vi. For soil order, we used the USDA soil taxonomy system[38]. Missing soil order data were extracted from Global Soil Regions Map database with a resolution of 2 minutes (FAO-UNESCO, https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/use/?cid=nrcs142p2_054013).

vii. Soil depth was calculated as the midpoint between the top and bottom of the reported depth interval. For example, if the soil sample was from the depth interval 10-20 cm, the soil depth was calculated as $(10+20)/2 = 15$ cm.

viii. Each $\Delta^{14}C$ measurement was normalized to the year 2000 using a steady state one-pool model and the observed time series of atmospheric $\Delta^{14}C$. Past atmospheric $\Delta^{14}C$ records were obtained from the Intcal13 calibration curve (50kyr – 0 BP)[24]. Modern data from 1950 were obtained from Vermunt and Schauinsland stations[39] extended through 2012[40,41]. To normalize $\Delta^{14}C$ to year 2000, we first constructed the relationship between turnover time and $\Delta^{14}C$ (shown in Supplementary Fig. 15) to derive turnover time for each $\Delta^{14}C$ value. Then we normalized the original $\Delta^{14}C$ by running the one-pool model with the respective turnover time to year 2000. Supplementary figure 16 shows the comparison between the original and normalized $\Delta^{14}C$.

2. Statistical modeling, prediction, and sources of uncertainty

Statistical modeling to identify key factors that influence vertical and spatial variability in soil $\Delta^{14}C$ was accomplished using machine learning techniques implemented in the Python environment for statistical computing (*Scikit-Learn*). We used a suite of algorithms including three generalized linear models, support vector regression, and two bagging and boosting ensemble methods. For model fitting, we used all soil profiles with predictors including mean annual temperature and precipitation, land cover type, soil depth, soil order, and soil clay content. Land cover type and soil order are categorical variables and were converted to binary variables for each class. A 5-fold cross validation based on soil profiles showed that random forest performed the best, accounting for about 69% of the variation in the profile dataset (Supplementary Table 3). Therefore, we used the random forest algorithm for our main analysis.

The random forest algorithm used 300 decision trees, with the maximum depth of 18. The learned hyperparameter values were derived using the grid search cross validation method from the *sklearn* library. With the random forest algorithm, importance scores for each predictor were calculated using the feature_importances function from *Scikit-Learn*. These scores reflect how important each predictor is in determining the fitted values of $\Delta^{14}C$.

Finally, we used the predictive model to extrapolate $\Delta^{14}C$ across the land surface at each 1 cm vertical increment to a soil depth of 1 meter. First, we trained the random forest machine learning algorithm with the observational data. The model features in the dataset included mean annual temperature and precipitation, land cover type, soil depth, soil order and clay content. Then, we applied the trained model to global databases of mean annual temperature, mean annual precipitation, land cover type, soil clay content, soil order and soil depth to generate a global dataset of soil $\Delta^{14}C$ (Supplementary Table 4). The gridded driver variables used for global extrapolation were all regridded to a spatial resolution of 0.5º × 0.5º. Specifically, we calculated 10-year average annual temperature and precipitation during 1990-2000 from the Climate

Research Unit (CRU) v. 3.23[42] as the climate driving data. The land cover map was obtained from MODIS Land cover MCD12Q1 product [43]. Note that 16 land cover types from MODIS were combined into 10 types for consistency with reported observations (Supplementary Fig. 14). Soil order data were extracted from the Global Soil Regions Map database[38]. Soil clay content was obtained from the SoilGrids database[37]. There are four depth intervals in the first meter (0-10cm, 10-30cm, 30-60cm and 60-100cm) for soil clay content in SoilGrids. The trained model was then used to predict mineral soil $\Delta^{14}C$ at each 1 cm increment to a depth of 1 m.

Note that the data-derived global gridded $\Delta^{14}C$ is subject to uncertainties from the machine learning algorithm, errors in the predictors of climate, soil properties, and land cover type, as well as uncertainty in the soil carbon content for the weighted $\Delta^{14}C$ estimates. We quantified these main uncertainty sources at both grid scale (Supplementary Fig. 17) and biome levels (Supplementary Table 5). To estimate the uncertainty from the algorithm, we calculated the absolute differences in global-gridded $\Delta^{14}C$ in each regression tree and our gridded product (baseline); to estimate the uncertainty by each of the key drivers, we first computed global gridded $\Delta^{14}C$ by holding out the given driver, and then calculated the absolute difference between the $\Delta^{14}C$ predictions and the baseline estimate. We found that uncertainties caused by excluding temperature were always greater than those caused by excluding precipitation, followed by those caused by excluding soil clay content (Supplementary Fig. 17). These results are consistent with our analysis of relative importance of different variables (Supplementary Figure 3).

In addition to uncertainties at the level of individual grid points, we have further quantified the uncertainties of $\Delta^{14}C$ at the biome level and for our global estimates (Supplementary Table 5). Weighting $\Delta^{14}C$ by different soil carbon datasets created the largest uncertainty in our global estimates of $\Delta^{14}C$, and including or excluding temperature and precipitation generated the largest uncertainty at a biome level. In addition, it is important to note that uneven sampling of soils in the ISRaD database, including relatively few sites in tundra and boreal forests, represents an important source of uncertainty and influences some of the breakpoints that emerge near -4°C in projections of $\Delta^{14}C$ and age shown in Supplementary Figs. 6 and 10. To reduce uncertainties in the age distribution of global soil carbon in future work, more $\Delta^{14}C$ profile measurements are needed in high latitude ecosystems as well as along moisture and temperature gradients in Africa and other sparsely sampled areas of the tropics (Supplementary Fig. 1). More accurate gridded maps of soil carbon content and other soil properties are essential for developing more accurate statistical and mechanistic models of soil carbon cycle and mean age.

3. Mean age calculation

Interpretation of carbon dynamics from radiocarbon data requires the use of models. The most effective way to use $\Delta^{14}C$ as a constraint on carbon cycling is to directly simulate this tracer within a land surface model within each ecosystem pool and soil layer and compare these predicted values to radiocarbon measurements. This is the approach we take to evaluate carbon cycling within CLM5 and ELM1.0. However, we also used the $\Delta^{14}C$ dataset directly to estimate global three-dimensional structure of the mean age of soil carbon. This approach, while requiring simplifying assumptions, can help with building an intuitive understanding of the processes regulating soil carbon dynamics at a global scale.

We estimated mean age as the turnover time in a one-pool, homogeneous, steady state model that was fit to the $\Delta^{14}C$ value in each 1-cm soil layer. Specifically, we assumed a steady state of soil carbon and radiocarbon at the beginning of the model run (i.e., 50 ky BP) and ran the model until the year 2000 with the atmospheric history of $\Delta^{14}C$. Then we determined the relationship between turnover time and $\Delta^{14}C$ in year 2000 (Supplementary Fig. 15). This relationship was used to derive the mean age for each layer. Note that when $\Delta^{14}C$ is greater than about 85‰, the calculation generates two mean turnover times (Supplementary Fig. 15A). We selected the longer one in our analysis, as measurements of bulk radiocarbon emphasize the carbon in mineral-associated organic matter that dominate total soil C content. In studies that applied multi-pool modeling to soil that had been divided into fractions according to density, the mineral-associated organic matter was associated with the longer turnover times[44-46].

This approximation of mean age is justified because the $\Delta^{14}C$ of the bulk soil carbon is primarily determined by pools of the most slowly cycling carbon. It is well known that soil carbon is not homogeneous, so our assumption of a single pool is simplistic but still informative. In theory, the mean age of material within a reservoir with multiple carbon residence times can be computed by using an impulse response approach[47]. The temporal integral of the product of the fractional mass remaining in the system with the time since entry of the impulse provides a direct measure of mean age. In practice this approach requires perfect knowledge of the different components that are cycling through the reservoir and their individual turnover times. Nonetheless, we tested a two-pool model constrained by bulk $\Delta^{14}C$ and estimated its mean carbon age. The mean carbon age (MCA) in the two-pool model is calculated using $MCA = \frac{1}{K_1} + \frac{1}{K_2} - \frac{1}{\alpha \times K_1 + K_2}$[48], where K_1

and K_2 are the turnover rates of the two carbon pools and α is carbon transfer coefficient from the first pool to the second pool. We found that the mean carbon age estimated in the one-pool model was within the uncertainty of mean age in the two-pool model, especially for young soil carbon (Supplementary Fig. 8). For comparison with carbon cycle models, we recommend directly simulating the three-dimensional structure of the gridded $\Delta^{14}C$ data set, following our approach described in section 5.

4. Carbon-weighted $\Delta^{14}C$ and mean age along depth and across land cover types

For each grid cell, we calculated carbon-weighted $\Delta^{14}C$ and mean age in the three depth intervals (0 – 100 cm, 0 – 30 cm, and 30 – 100 cm) using soil carbon datasets from SoilGrids[37] and the Harmonized World Soil Database (HWSD)[49]. Note that soil carbon content in SoilGrids has been updated (December 24th, 2018) and is available at https://landgis.opengeohub.org. Both datasets were re-gridded to 0.5° to match the resolution of our $\Delta^{14}C$ maps. There are four soil layers (0 – 10 cm, 10 – 30 cm, 30 – 60 cm and 60 – 100 cm) in the SoilGrids database and two soil layers (0 – 30 cm and 30 – 100 cm) in HWSD. To calculate the vertical, carbon-weighted $\Delta^{14}C$ and mean age for 0 – 100 cm at each grid cell we used SoilGrids with equation 3 and HWSD with equation 4:

$$X_{w,\,0\text{–}100} = C_{0\text{–}10} / C_{0\text{–}100} \times X_{uw,\,0\text{–}10} + C_{10\text{–}30} / C_{0\text{–}100} \times X_{uw,\,10\text{–}30} + C_{30\text{–}60} / C_{0\text{–}100} \times X_{uw,\,30\text{–}60} + C_{60\text{–}100} / C_{0\text{–}100} \times X_{uw,\,60\text{–}100} \quad (3)$$

$$X_{w,\,0\text{–}100} = C_{0\text{–}30} / C_{0\text{–}100} \times X_{uw,\,0\text{–}30} + C_{30\text{–}100} / C_{0\text{–}100} \times X_{uw,\,30\text{–}100} \quad (4)$$

Where *w* stands for weighted, *uw* is unweighted, X is $\Delta^{14}C$ or mean age, and C is soil carbon content. Due to lack of depth resolution in HWSD, we only used soil carbon from SoilGrids for the weighting within the depth intervals of 0 – 30 cm and 30 – 100 cm (equations 5 and 6).

$$X_{w,\,0\text{–}30} = C_{0\text{–}10} / C_{0\text{–}30} \times X_{uw,\,0\text{–}10} + C_{10\text{–}30} / C_{0\text{–}30} \times X_{uw,\,10\text{–}30} \quad (5)$$

$$X_{w,\,30\text{–}100} = C_{30\text{–}60} / C_{30\text{–}100} \times X_{uw,\,30\text{–}60} + C_{60\text{–}100} / C_{30\text{–}100} \times X_{uw,\,60\text{–}100} \quad (6)$$

To calculate the global mean of $\Delta^{14}C$ and mean carbon age in the three depth intervals we describe in the main text (0 – 30 cm, 30 – 100 cm, and 0 – 100 cm), we weighted $\Delta^{14}C$ or mean age from each biome based on the carbon content of that biome according to equation 7:

$$X_{global} = \sum_{i=1}^{8} \left(\frac{C_{lc,depth\ interval}}{C_{total,depth\ interval}} \times X_{lc}\right) \quad (7)$$

where X_{global} is the globally weighted soil $\Delta^{14}C$ or mean age for each of the three depth intervals; C_{lc} is total carbon content in each of the 8 land cover types; and X_{lc} is $\Delta^{14}C$ or mean age bootstrapped randomly 1000 times from its distribution in each land cover type. We then computed the mean and standard deviation of the global weighted $\Delta^{14}C$ and mean age. Note that we created an average of X_{global} by weighting spatially across different biomes by both HWSD and SoilGrids.

We also provided the median and 5% to 95% range for the $\Delta^{14}C$ and mean age within each land cover type and permafrost versus non-permafrost regions. The permafrost map was generated by the National Snow and Ice Data Center[50] and is accessible at https://neo.sci.gsfc.nasa.gov/view.php?datasetId=PermafrostNSIDC&date=2002-02-01.

5. Global land surface models

Soil radiocarbon content, $\Delta^{14}C$ in year 2000, simulated in global land models were compared with our gridded dataset at 0 – 30 and 30 – 100 cm depth intervals. Two depth-resolved global land models were used, the land model from the Energy Exascale Earth System Model version 1.0 with the Equilibrium Chemistry Approximation (ELMv1-ECA)[27] and the Community Land

Model version 5.0 (CLM5)[26]. Both simulate global terrestrial carbon and radiocarbon cycles with explicit representation of soil depth and both models were based on similar initial structure and parameterization[29]. These two models are among a handful of published global models with explicit depth and radiocarbon modules for soil carbon cycling. In addition, both models have been assessed using the International Land Model Benchmarking (ILAMB) system[51].

For the ELMv1-ECA simulation, we initialized the model with a 500-year spin-up simulation, with the first 300 years using the accelerated decomposition procedure, followed by a transient simulation from 1901 to 2010 with Global Soil Wetness Project Phase 3 climate forcing and observed atmospheric CO_2, nitrogen deposition, and ^{14}C, without land use change. The spin-up used 1850 (pre-industrial) conditions for land cover and atmospheric chemistry (CO_2, aerosols, and nitrogen deposition), and a constant atmospheric ^{14}C of zero per mil. The model simulated vertical profiles of SOC ^{14}C globally on 1.9° × 2.5° grids with ten soil layers from 0-3.5 m depth[29].

For CLM5, the initial conditions were also generated by spinning up the model to steady state for 1850 conditions. As with ELM, atmospheric chemistry and land cover were for the year 1850 but climate forcing was for 1901-1920. The transient simulation spanned the period 1850-2014 with Global Soil Wetness Project Phase 3 climate forcing at about a 1° resolution. Land use and land-cover change, atmospheric CO_2 and ^{14}C concentration, and nitrogen deposition were specified from transient datasets[52], which are consistent with the second generation land-use harmonization (LUH2) and CMIP6 protocols[53]. CLM5 simulates vertical profiles of soil ^{14}C with variable soil depth (0-8.5 m) and up to 20 soil layers[54]. Relative to the parameterization used in

ELM and previous versions of CLM, CLM5 applies a lower e-folding depth for soil C decay in deeper soil horizons and applies a stronger soil moisture constraint on decomposition rates[30].

For comparison with our data product, we integrated $\Delta^{14}C$ in the two models for the two depth intervals (0 – 30 cm and 30 – 100 cm) weighted by soil carbon. Because this method assumes uniform density throughout each model layer, it may underestimate the contribution of the lowest layer (82 – 138 cm), but we believe it is a fairly small difference. We did not regrid the spatial resolutions in the two models to the same resolution as the data. Because both models use similar land cover types as our data product, we overlaid the same MODIS-derived map on the two model grids to obtain the biome-level estimates from the models.

Data availability

The gridded maps of soil $\Delta^{14}C$ and mean carbon age are available on the ISRaD website (https://soilradiocarbon.org) and archived at Zenodo (www.zenodo.org).

Code availability

All code relating to this study is available from the corresponding author upon request.

References

37 Hengl, T. *et al.* SoilGrids250m: Global gridded soil information based on machine learning. *Plos One* **12**, e0169748 (2017).

38 Soil Survey Staff. Soil taxonomy: a basic system of soil classification for making and interpreting soil surveys. 2nd edition. *Natural Resources Conservation Service. U.S. Department of Agriculture Handbook 436.* (1999).

39 Levin, I. & Kromer, B. Twenty years of atmospheric ^{14}CO2 observations at Schauinsland station, Germany. *Radiocarbon* **39**, 205-218 (1997).

40 Levin, I., Kromer, B. & Hammer, S. Atmospheric Δ^{14}CO2 trend in Western European background air from 2000 to 2012. *Tellus B: Chemical and Physical Meteorology* **65**, 20092 (2013).

41 Hua, Q., Barbetti, M. & Rakowski, A. Z. Atmospheric radiocarbon for the period 1950–2010. *Radiocarbon* **55**, 2059-2072 (2013).

42 Harris, I., Jones, P. D., Osborn, T. J. & Lister, D. H. Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 Dataset. *International journal of climatology* **34**, 623-642 (2014).

43 Friedl, M. A. *et al.* MODIS Collection 5 global land cover: algorithm refinements and characterization of new datasets. *Remote Sens. Environ.* **114**, 168-182 (2010).

44 Sierra, C. A. *et al.* Predicting decadal trends and transient responses of radiocarbon storage and fluxes in a temperate forest soil. *Biogeosciences* **9**, 3013-3028 (2012).

45 Schrumpf, M. *et al.* Storage and stability of organic carbon in soils as related to depth, occlusion within aggregates, and attachment to minerals. *Biogeosciences* **10**, 1675-1691 (2013).

46 Gaudinski, J. B., Trumbore, S. E., Davidson, E. A. & Zheng, S. H. Soil carbon cycling in a temperate forest: radiocarbon-based estimates of residence times, sequestration rates and partitioning of fluxes. *Biogeochemistry* **51**, 33-69 (2000).

47 Manzoni, S. & Porporato, A. Soil carbon and nitrogen mineralization: theory and models across scales. *Soil Biology and Biochemistry* **41**, 1355-1379 (2009).

48 Sierra, C. A., Müller, M., Metzler, H., Manzoni, S. & Trumbore, S. E. The muddle of ages, turnover, transit, and residence times in the carbon cycle. *Global Change Biology* **23**, 1763-1773 (2017).

49 FAO, IIASA, ISRIC, ISSCAS & JRC. Harmonized World Soil Database (version 1.2), edited by: FAO, Rome, Italy and IIASA, Laxenburg, Austria. (2012).

50 Brown, J., Ferrians Jr, O., Heginbottom, J. & Melnikov, E. *Circum-Arctic map of permafrost and ground-ice conditions*. (US Geological Survey Reston, VA, 1997).

51 Collier, N. *et al.* The International Land Model Benchmarking (ILAMB) system: design, theory, and implementation. *Journal of Advances in Modeling Earth Systems* **10**, 2731-2754 (2018).

52 Bonan, G. B. *et al.* Model structure and climate data uncertainty in historical simulations of the terrestrial carbon cycle (1850–2014). *Global Biogeochemical Cycles* **33**, 1310-1326 (2019).

53 Lawrence, D. M. *et al.* The Land Use Model Intercomparison Project (LUMIP) contribution to CMIP6: rationale and experimental design. *Geosci. Model Dev.* **9**, 2973-2998 (2016).

54 Brunke, M. A. *et al.* Implementing and evaluating variable soil thickness in the Community Land Model, Version 4.5 (CLM4.5). *Journal of Climate* **29**, 3441-3461 (2016).

Table 1. Summary statistics of soil carbon, Δ^{14}C in year 2000, and mean carbon age in each biome. The values of Δ^{14}C and mean age for each biome (and for permafrost and non-permafrost regions) are the median and 5% to 95% range (in parentheses). Global mean and standard deviation (mean ± sd) of Δ^{14}C and mean age is weighted by soil carbon content in each biome and soil layer. Mean and standard deviation of soil carbon content for each biome were derived from two global carbon datasets (Harmonized World Soil Database and SoilGrids) described in the methods.

	Surface soil (0 – 30 cm)			**Subsurface soil (30 – 100 cm)**		
Biome	Soil carbon (Pg C)	Δ^{14}C (‰)	Age (years)	Soil carbon (Pg C)	Δ^{14}C (‰)	Age (years)
Boreal Forest	192±99	-86 (-228, -36)	1020 (650, 2750)	251±166	-385 (-652, -291)	5920 (3740, 22250)
Temperate Forest	46±11	-9 (-72, 46)	440 (200, 920)	42±12	-229 (-334, -157)	2710 (1680, 4670)
Tropical Forest	93±15	7 (-48, 35)	390 (260, 770)	102±37	-250 (-325, -166)	2970 (1790, 4310)
Grassland	75±14	-102 (-218, -16)	1200 (500, 2640)	75±27	-361 (-585, -253)	5380 (3050, 14690)
Cropland	114±19	-58 (-171, 7)	770 (380, 1850)	124±31	-287 (-383, -167)	3690 (1820, 5940)
Shrubland	29±4	-49 (-108, -23)	680 (490, 1240)	26±7	-258 (-384, -147)	3180 (1550, 6080)
Savanna	103±13	-20 (-144, 24)	510 (270, 1620)	107±25	-241 (-439, -119)	2860 (1240, 7960)
Tundra	188±112	-249 (-295, -142)	3490 (1660, 4310)	282±215	-624 (-706, -424)	16890 (6820, 28470)
Permafrost	322±176	-217 (-287, -75)	2770 (940, 4200)	443±320	-603 (-698, -358)	15440 (5150, 28270)
Non-permafrost	517±104	-42 (-150, 25)	660 (290, 1660)	565±184	-274 (-391, -149)	3420 (1590, 6190)
Global mean*	840±280	-97±24	1390±310	1008±505	-391±56	8280±2820

* Global weighted Δ^{14}C was -244±48‰ and mean age was 4830±1730 years for mineral soil carbon down to 1 m depth.

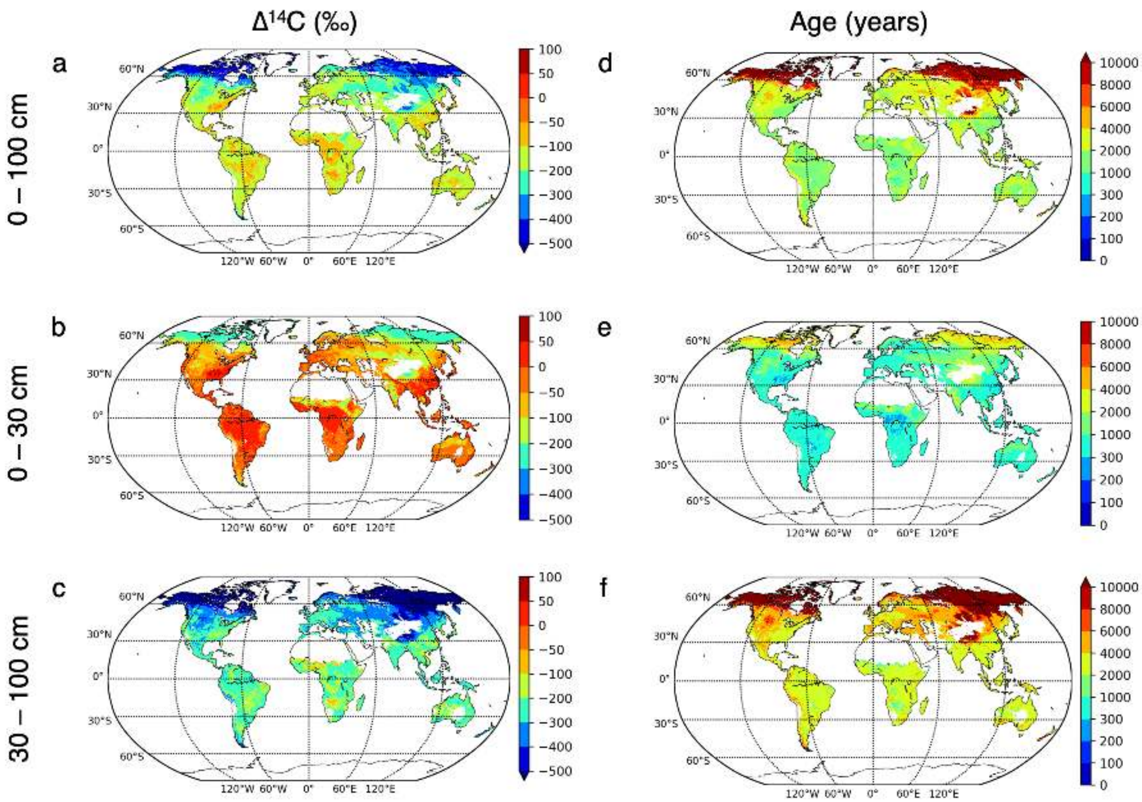


Fig. 1. Global distribution of soil Δ^{14}C and mean carbon age. Carbon-weighted average Δ^{14}C and mean age in the top 1 meter (**a** and **d**), surface soil (0 – 30 cm; **b** and **e**) and subsurface soil (30 – 100 cm; **c** and **f**) are shown at a 0.5° × 0.5° spatial resolution, derived from a random forest model trained with 789 soil radiocarbon profiles.

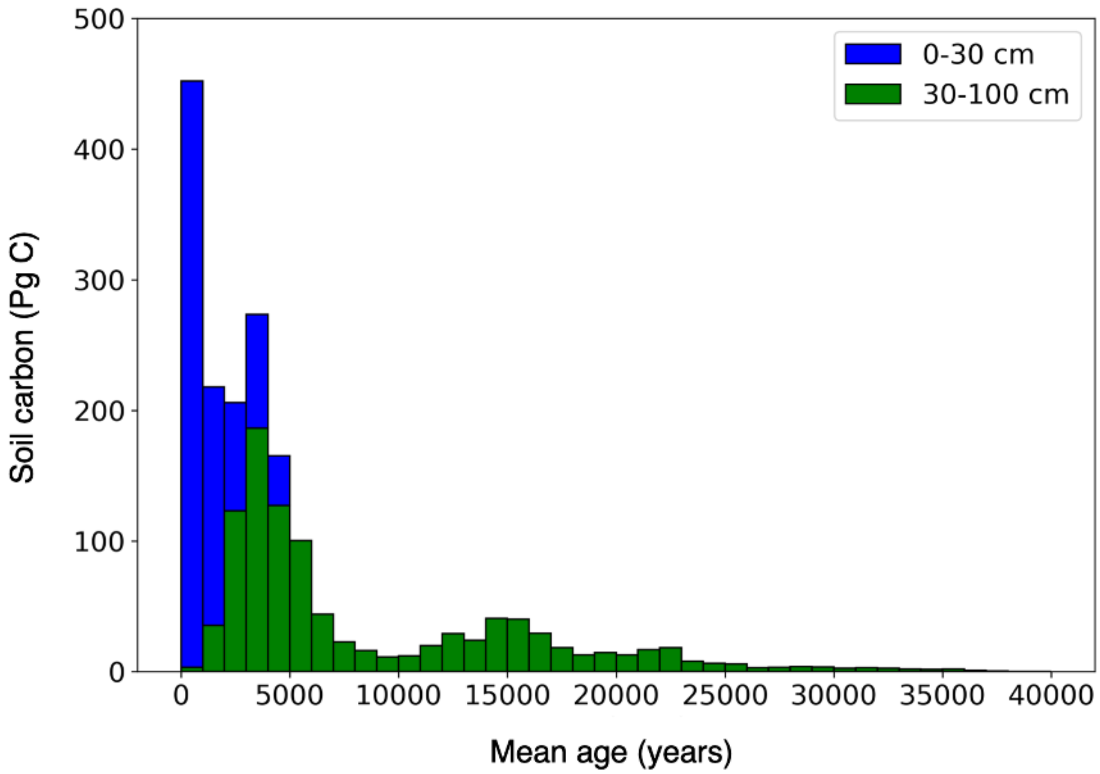


Fig. 2. Age distribution of global soil carbon. Histogram shows mean carbon age derived from the globally gridded $\Delta^{14}C$ dataset for surface (0-30 cm, blue) and subsurface (30-100 cm, green) layers. Soil carbon is the mean of two global databases, Harmonized World Soil Database and SoilGrids.

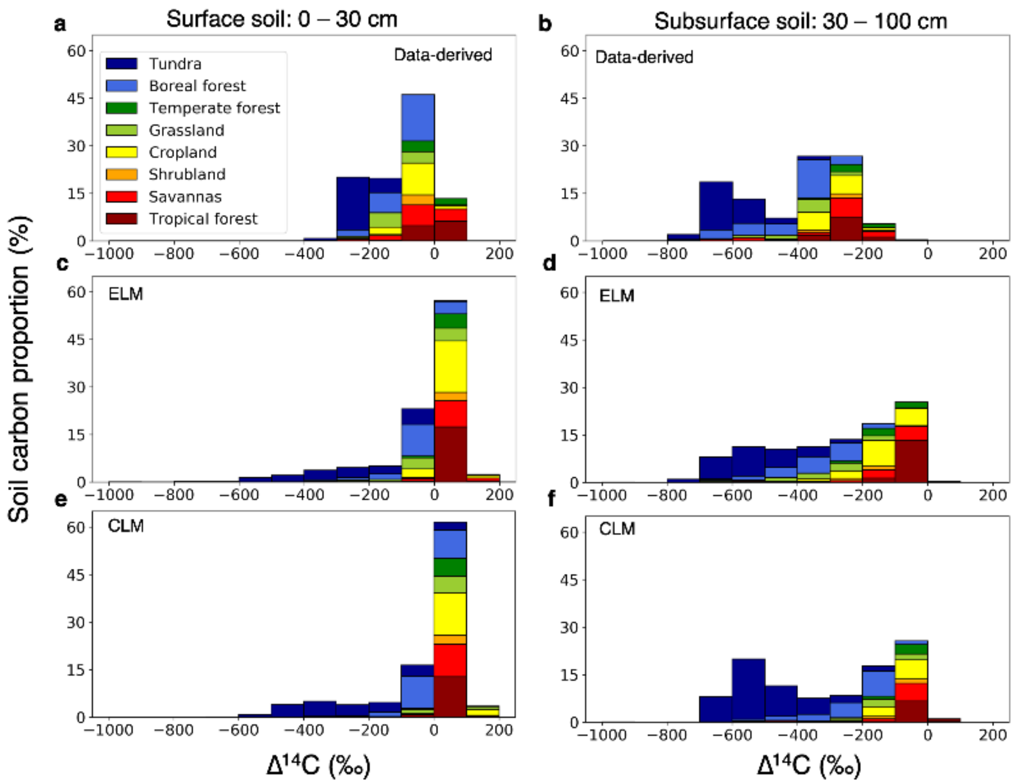


Fig. 3. Comparison of land surface model predictions of soil $\Delta^{14}C$ with the data product developed here for different depths and biomes. Histograms show the distribution of soil carbon in each biome as a function of $\Delta^{14}C$ for the globally gridded data (data-derived; a and b) and the two global land models (ELM v1.0 and CLM5; c – f), for the two depth intervals.

Supplementary Materials for

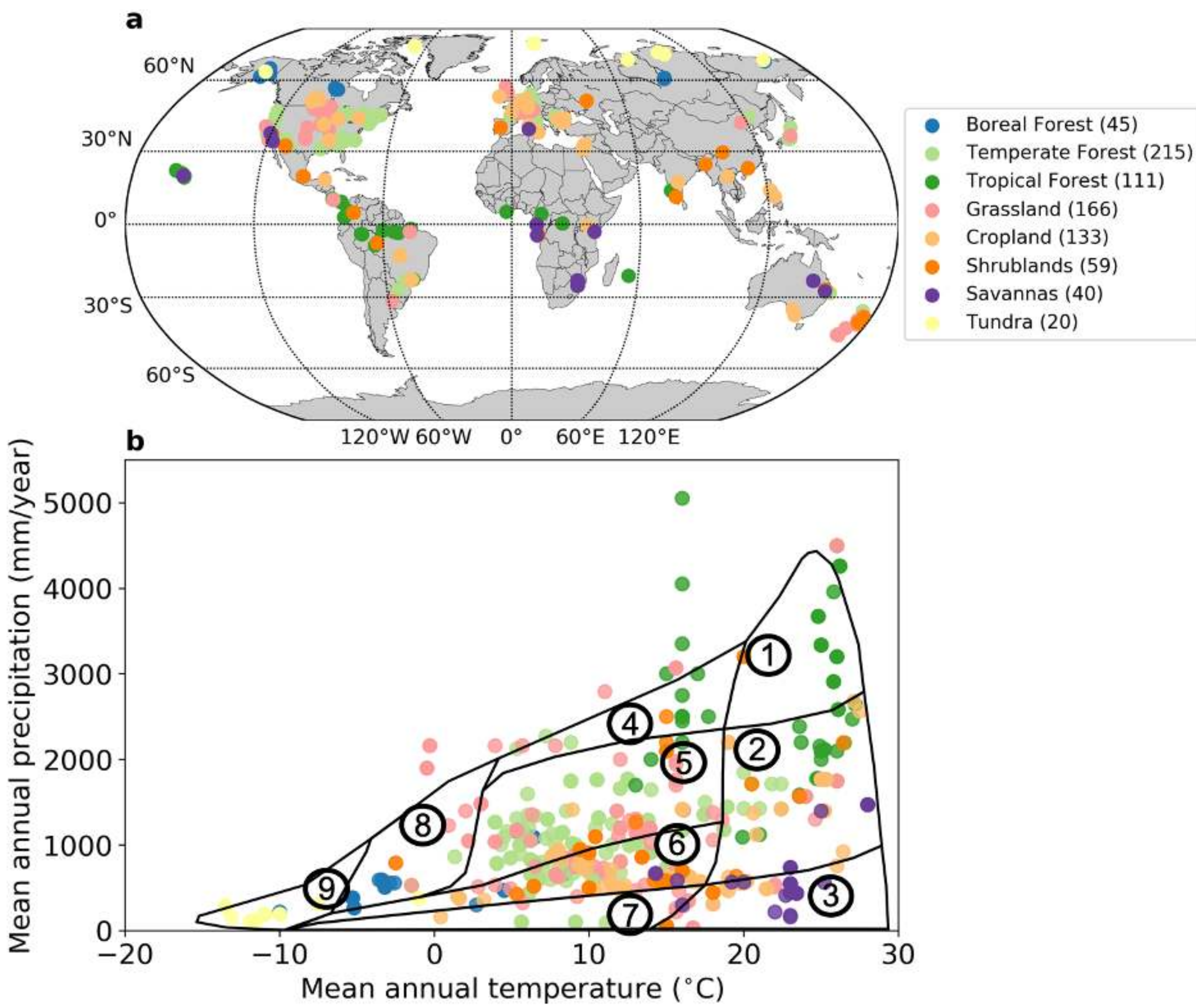
Millennial timescales of soil carbon cycling implied by global radiocarbon measurements

Zheng Shi*, Steven D. Allison, Yujie He, Paul A. Levine, Alison M. Hoyt, Jeff Beem-Miller,
Qing Zhu, William R. Wieder, Susan Trumbore, James T. Randerson
*Correspondence to: zshi7@uci.edu

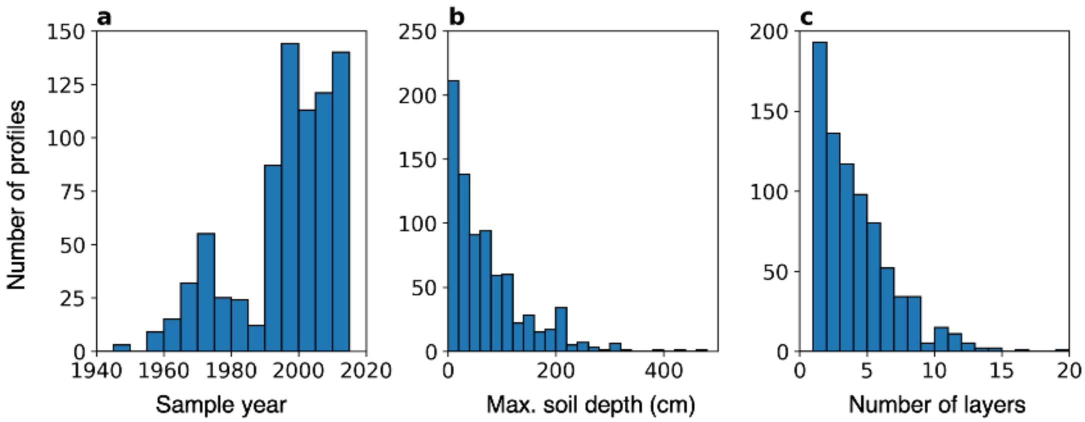
This file includes:

Supplementary Figures 1 to 17
Supplementary Tables 1 to 5

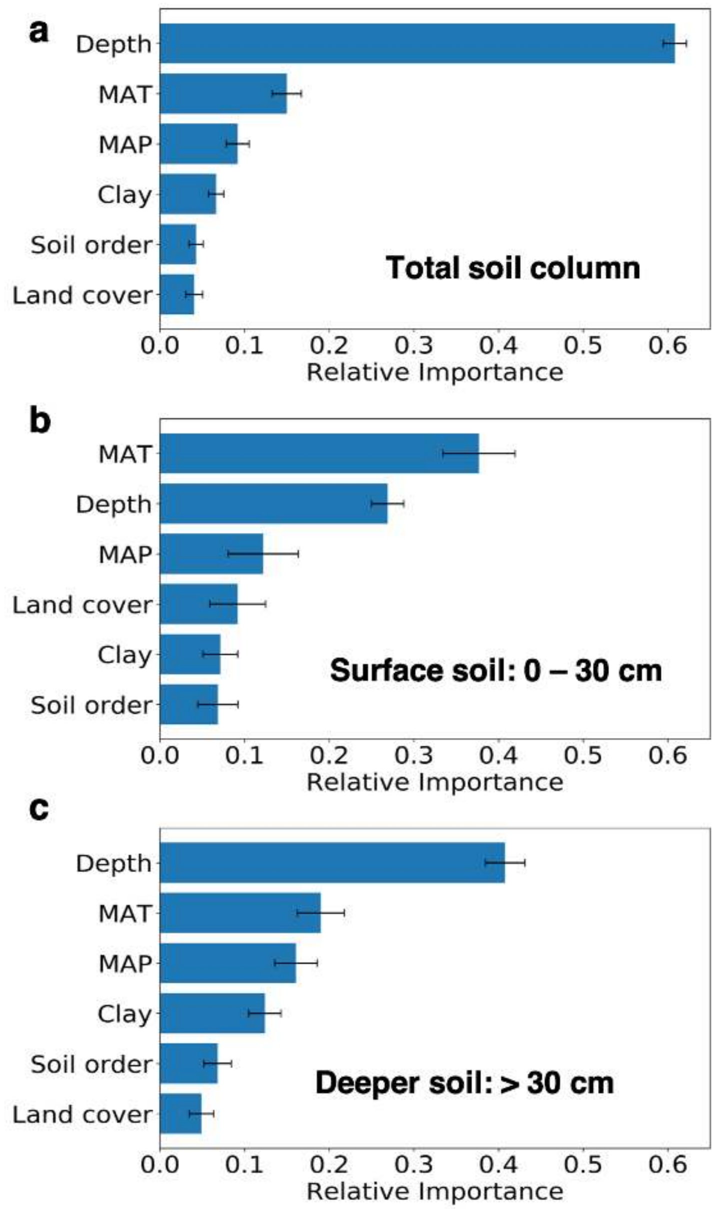
Supplementary figures:



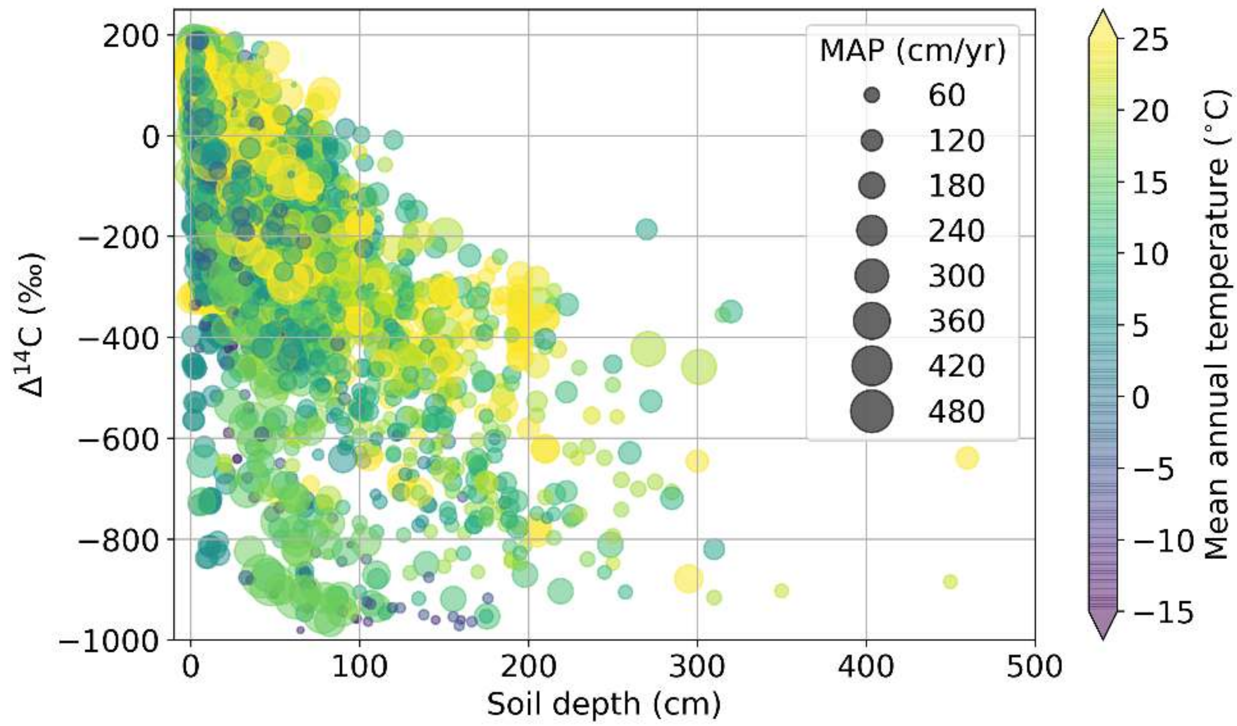
Supplementary Fig. 1. Location and climate of soil radiocarbon measurements. (**a**) A total of 789 soil profiles span all major land cover types and climate zones. (**b**) Climate of the soil profiles varies widely in mean annual temperature and mean annual precipitation. Black lines delineate Whittaker's biomes[37] according to mean annual temperature and precipitation. The biomes are: 1, tropical rainforest; 2, tropical seasonal rainforest/savanna; 3, subtropical desert; 4, temperate rainforest; 5, temperate seasonal forest; 6, woodland/shrubland; 7, temperate grassland/desert; 8, boreal forest; and 9, tundra.



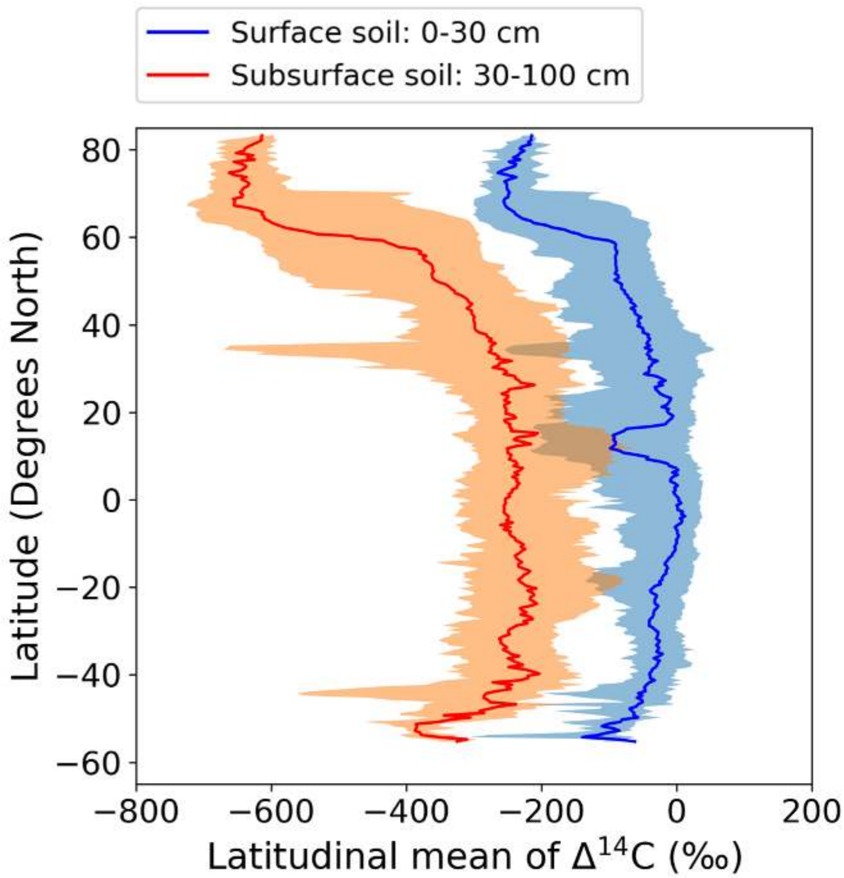
Supplementary Fig. 2. Frequency distribution of soil profiles. (**a**) The distribution of sample years. One archived soil profile sampled in 1900 is not shown here. (**b**) The distribution of maximum mineral soil depth (relative to the top of mineral soil). One soil profile with maximum depth of 600 cm is not shown here. (**c**) The distribution of number of layers in each soil profile. Not shown are 10 soil profiles that have more than 20 layers.



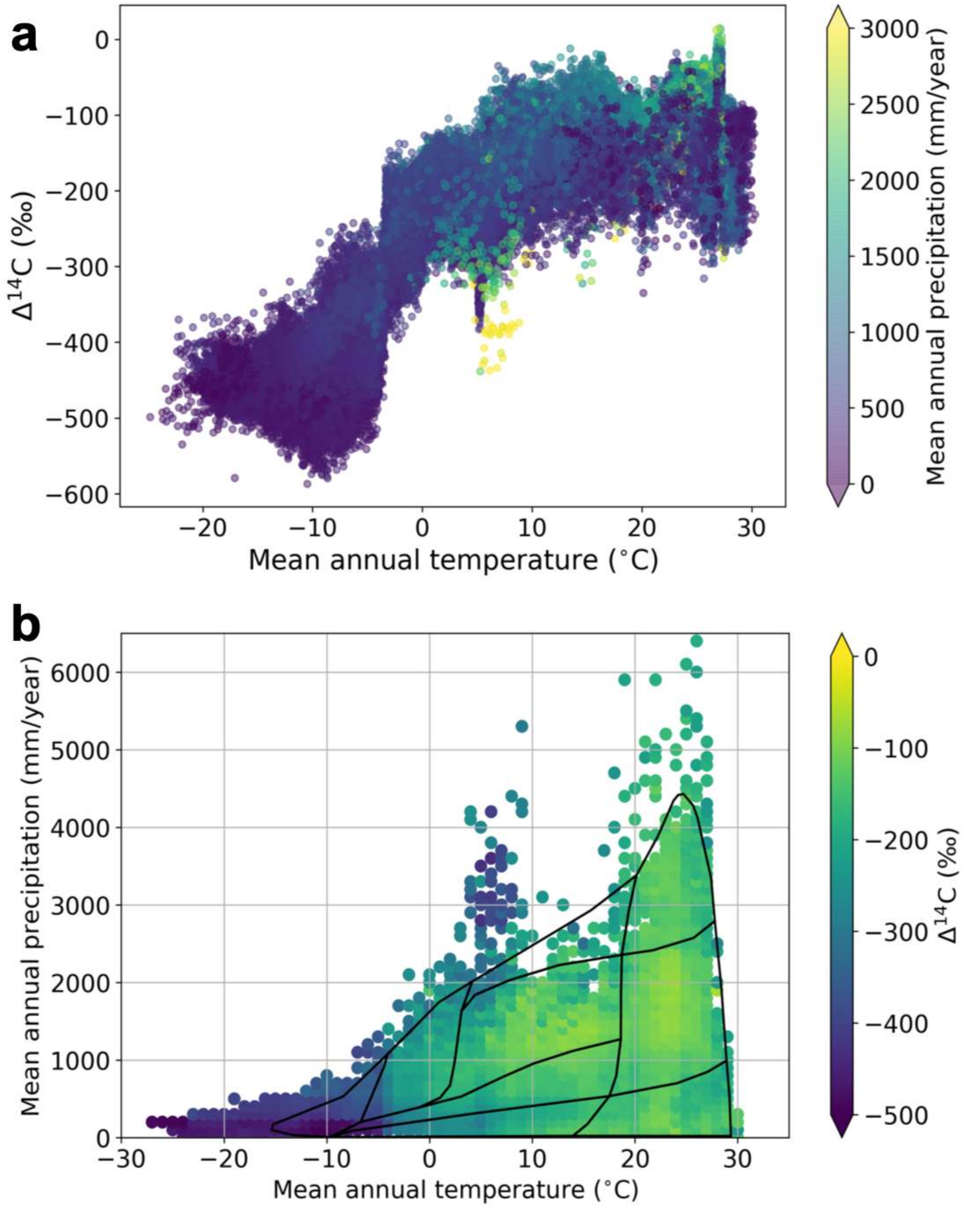
Supplementary Fig. 3. Relative variable importance based on the Random Forest algorithm for $\Delta^{14}C$. Three soil depth intervals include total soil column (**a**), surface soil (0 – 30 cm; **b**), and deeper soil (> 30 cm; **c**).



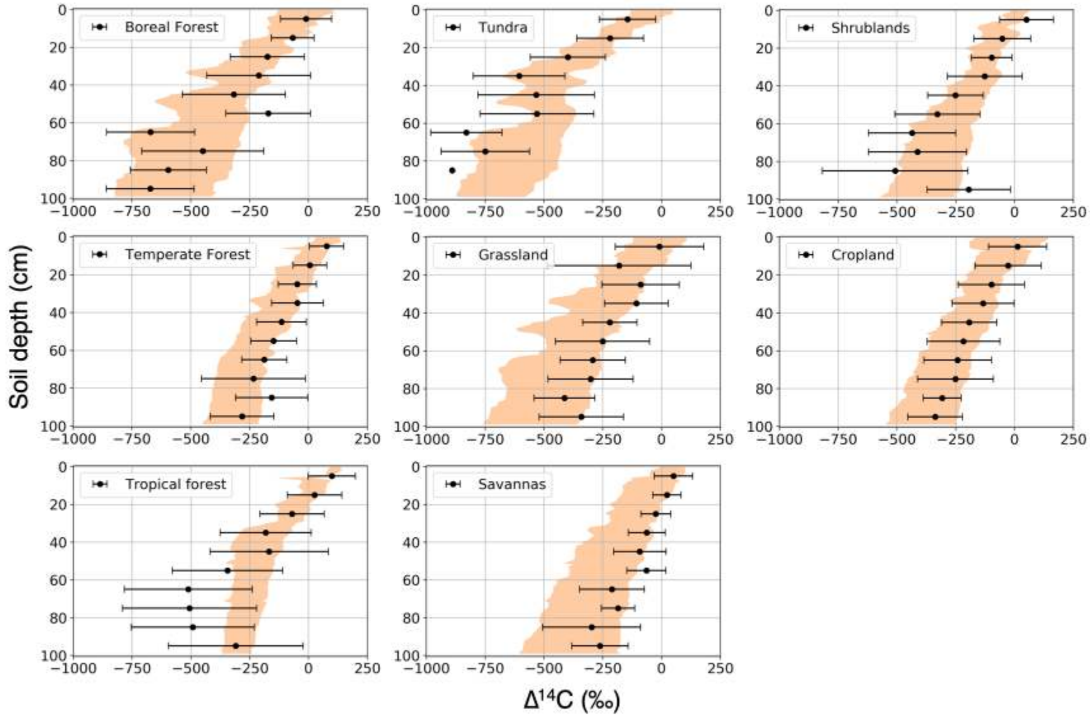
Supplementary Fig. 4. Relationships between measured $\Delta^{14}C$ and soil depth, mean annual temperature, and mean annual precipitation (MAP). $\Delta^{14}C$ decreases with soil depth, but increases with temperature and precipitation. Note that $\Delta^{14}C$ is the value normalized to year 2000.



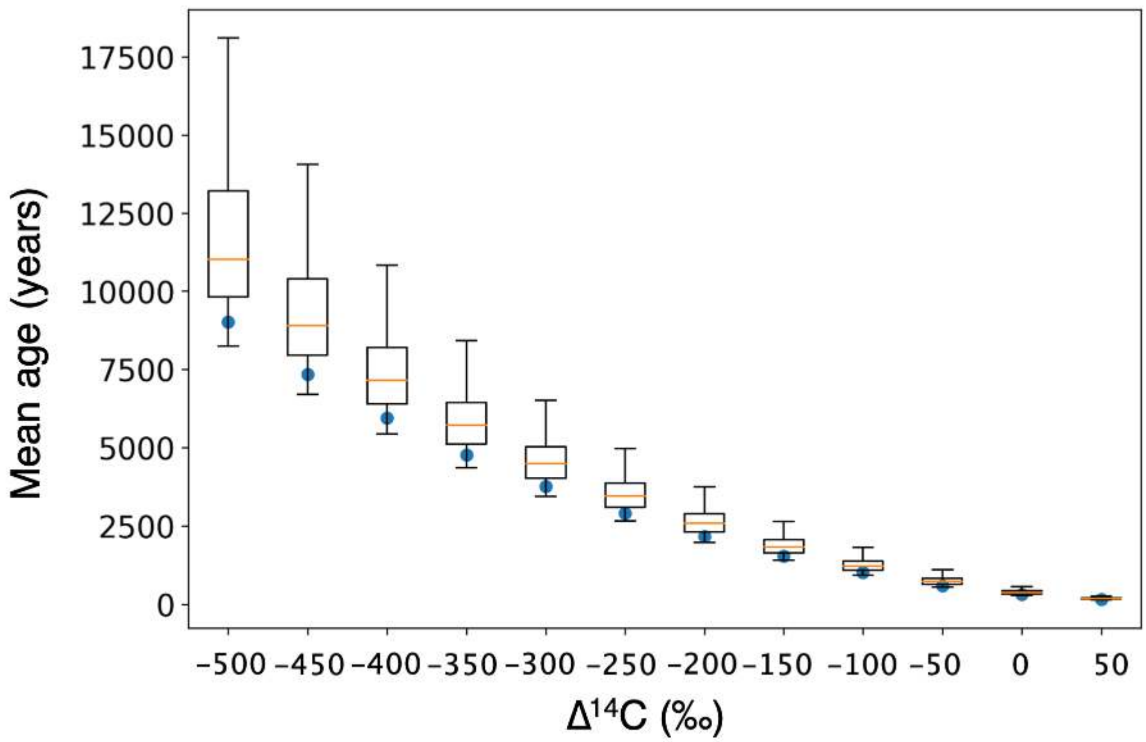
Supplementary Fig. 5. Latitudinal distribution of globally gridded $\Delta^{14}C$ in surface (0–30 cm) and subsurface (30–100 cm) soils. Lines and shaded area are median and the 5th–95th percentiles, respectively. The peaks in low- and mid-latitudes were mainly caused by the dry regions close to the Sahara and Taklamakan deserts.



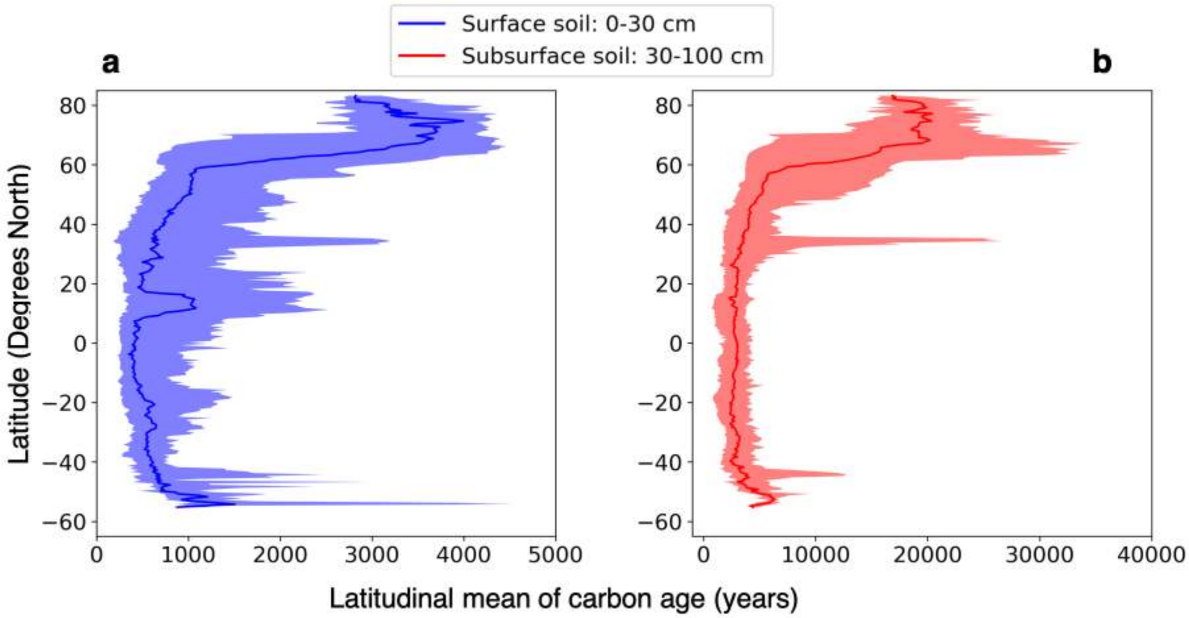
Supplementary Fig. 6. Distribution of soil radiocarbon $\Delta^{14}C$. $\Delta^{14}C$ (0 – 100 cm) varies with climatic space (**a**) and land cover type (**b**). Black lines delineate Whittaker's biomes according to mean annual temperature and precipitation. See Supplementary Fig. 1b for biome types.



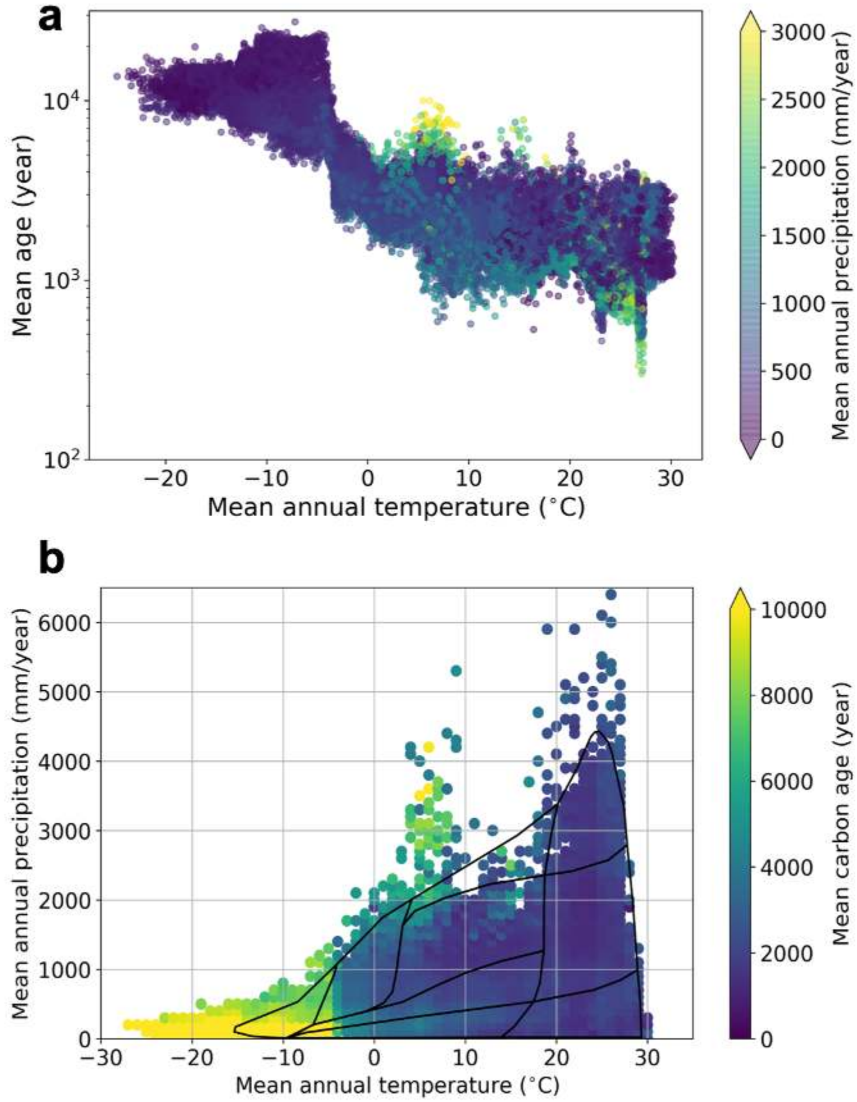
Supplementary Fig. 7. Depth distribution of $\Delta^{14}C$ (‰) in different land cover types. Black circles and error bars are observations with mean and standard deviation binned over 10 cm depth intervals. Note that there were no observations within 90 – 100 cm in the tundra biome. Shaded areas are the 5th–95th percentiles in each biome at 1-cm depth intervals from the global gridded $\Delta^{14}C$.



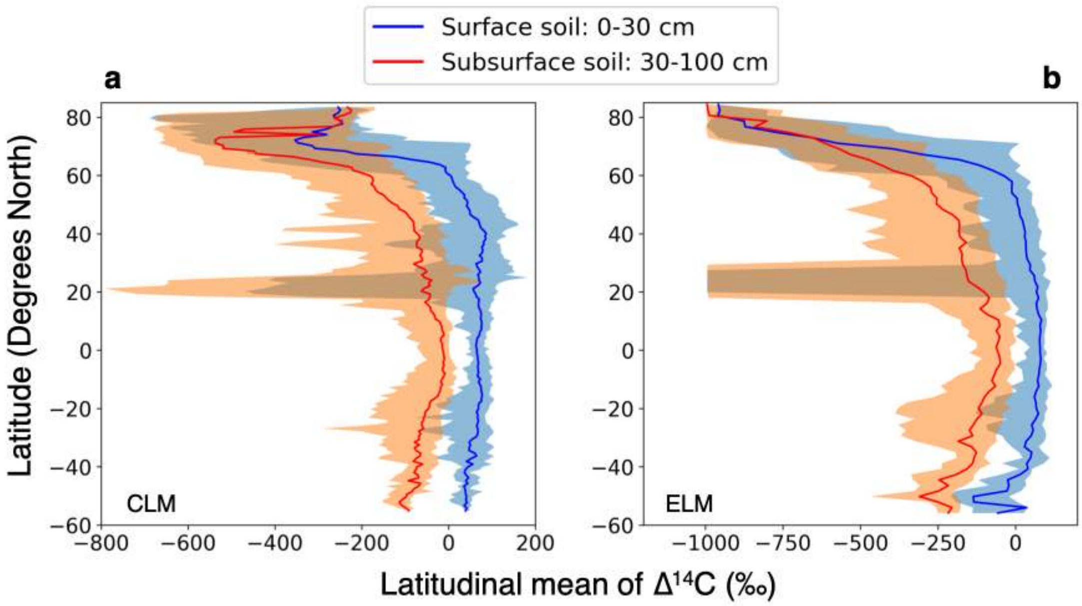
Supplementary Fig. 8. Comparison in mean carbon age between one-pool and two-pool models constrained by Δ^{14}C. The blue dots are mean age estimated in the one-pool model, and the box plots (whiskers are 5%-95% confidence interval) are mean age estimated in the two-pool model. The uncertainty stems from variation in turnover time of the two pools ($1/K_1$ and $1/K_2$) and the transfer coefficient (α) between the two pools. The mean carbon age (MCA) is calculated using $MCA = \frac{1}{K_1} + \frac{1}{K_2} - \frac{1}{\alpha \times K_1 + K_2}$.



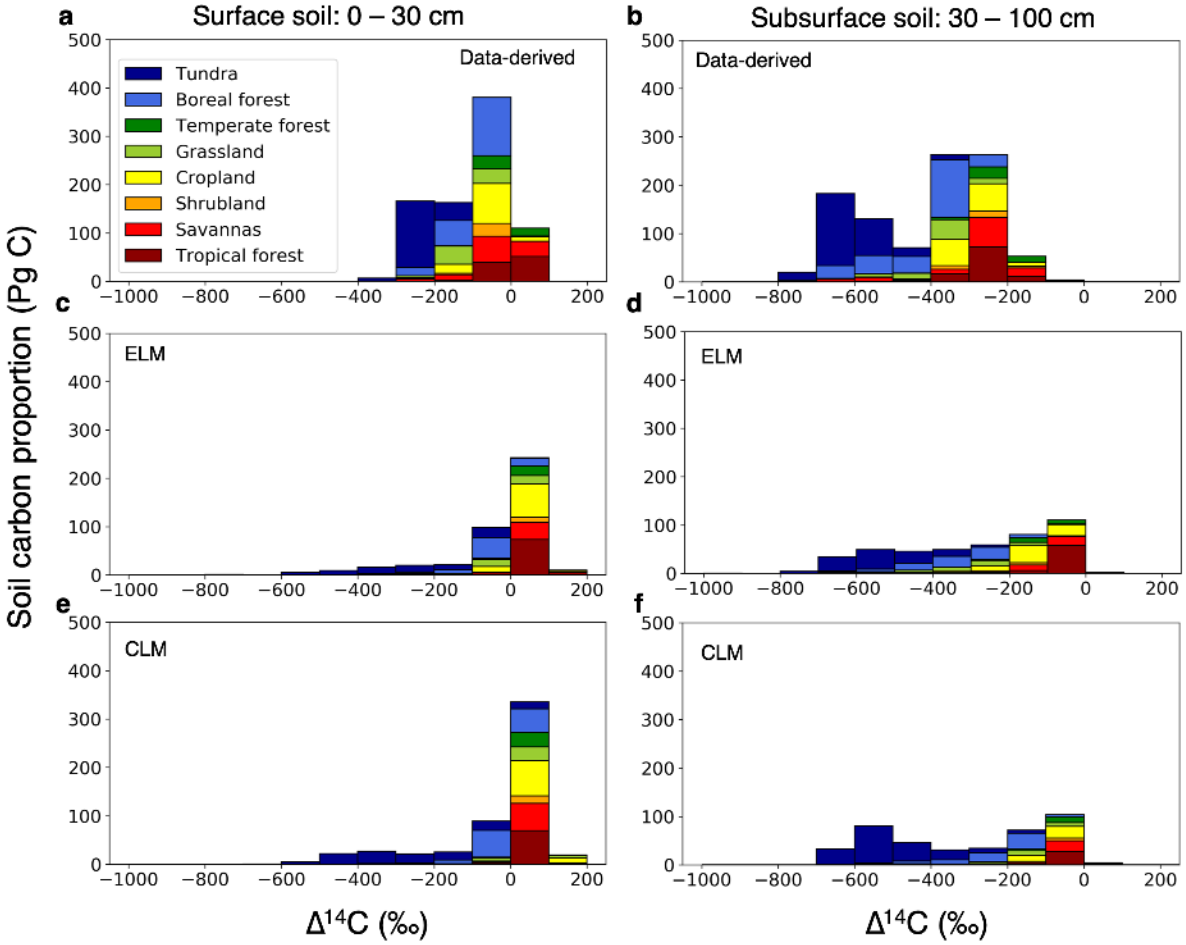
Supplementary Fig. 9. Latitudinal distribution of globally gridded soil carbon mean age. Panel **a** shows mean age in surface (0–30 cm) soil and panel **b** shows mean age in subsurface (30–100 cm) soil. Lines and shaded area are median and the 5th–95th percentiles, respectively.



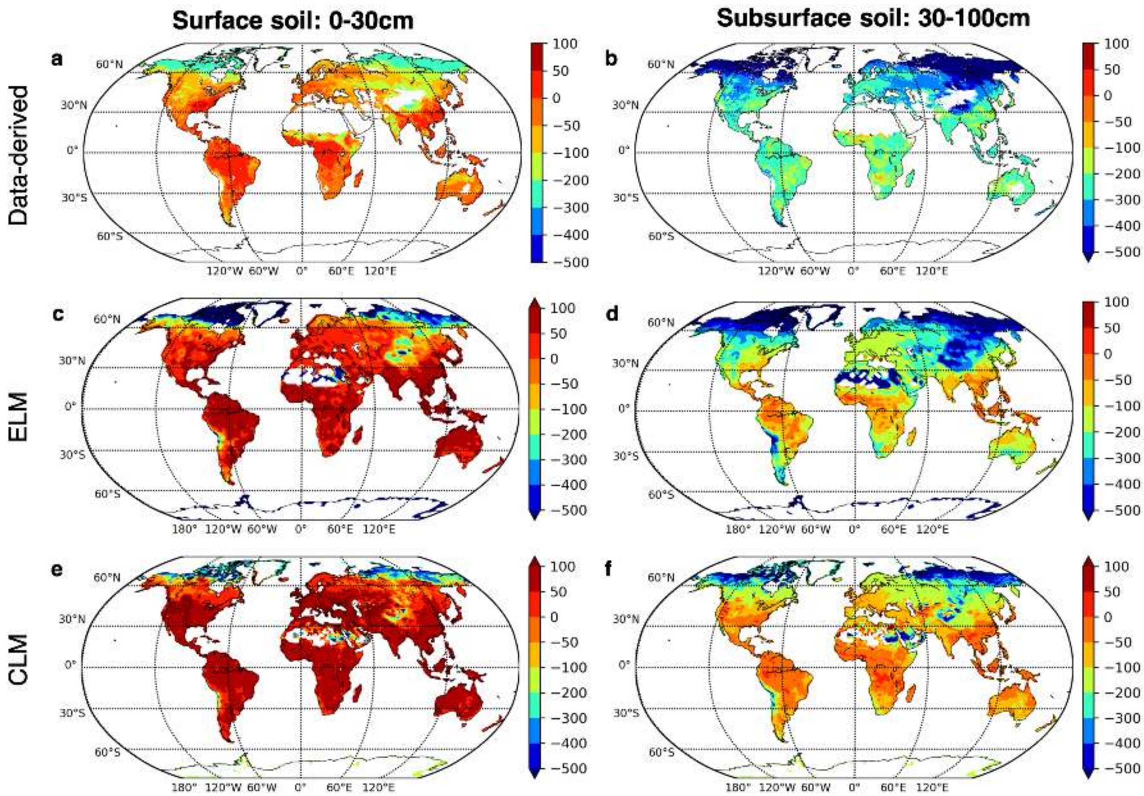
Supplementary Fig. 10. Distribution of mean carbon age. Mean carbon age (0 – 100 cm) varies with climatic space (**a**) and land cover type (**b**). Black lines delineate Whittaker's biomes according to mean annual temperature and precipitation. See Supplementary Fig. 1b for biome types.



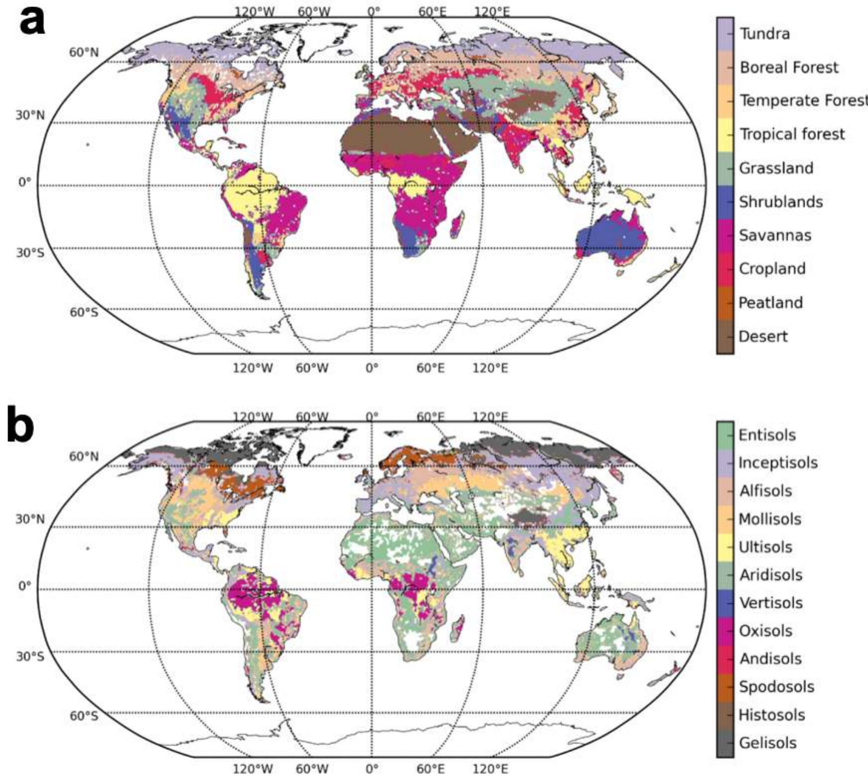
Supplementary Fig. 11. Latitudinal distribution of Δ^{14}C. a, Δ^{14}C in CLM in surface (0 – 30 cm) and subsurface (30 – 100 cm) soils. **b**, Δ^{14}C in ELM surface and subsurface soils. Lines and shaded area are median and the 5th–95th percentiles, respectively. In both models, Δ^{14}C becomes very negative in the Sahara Desert (near 20°N) because low soil moisture levels reduce the rate constant for decomposition and because of challenges in spinning up the models in regions with low carbon inputs. This does not appreciably modify carbon cycling in the model because levels of NPP and carbon storage are also very low in this region.



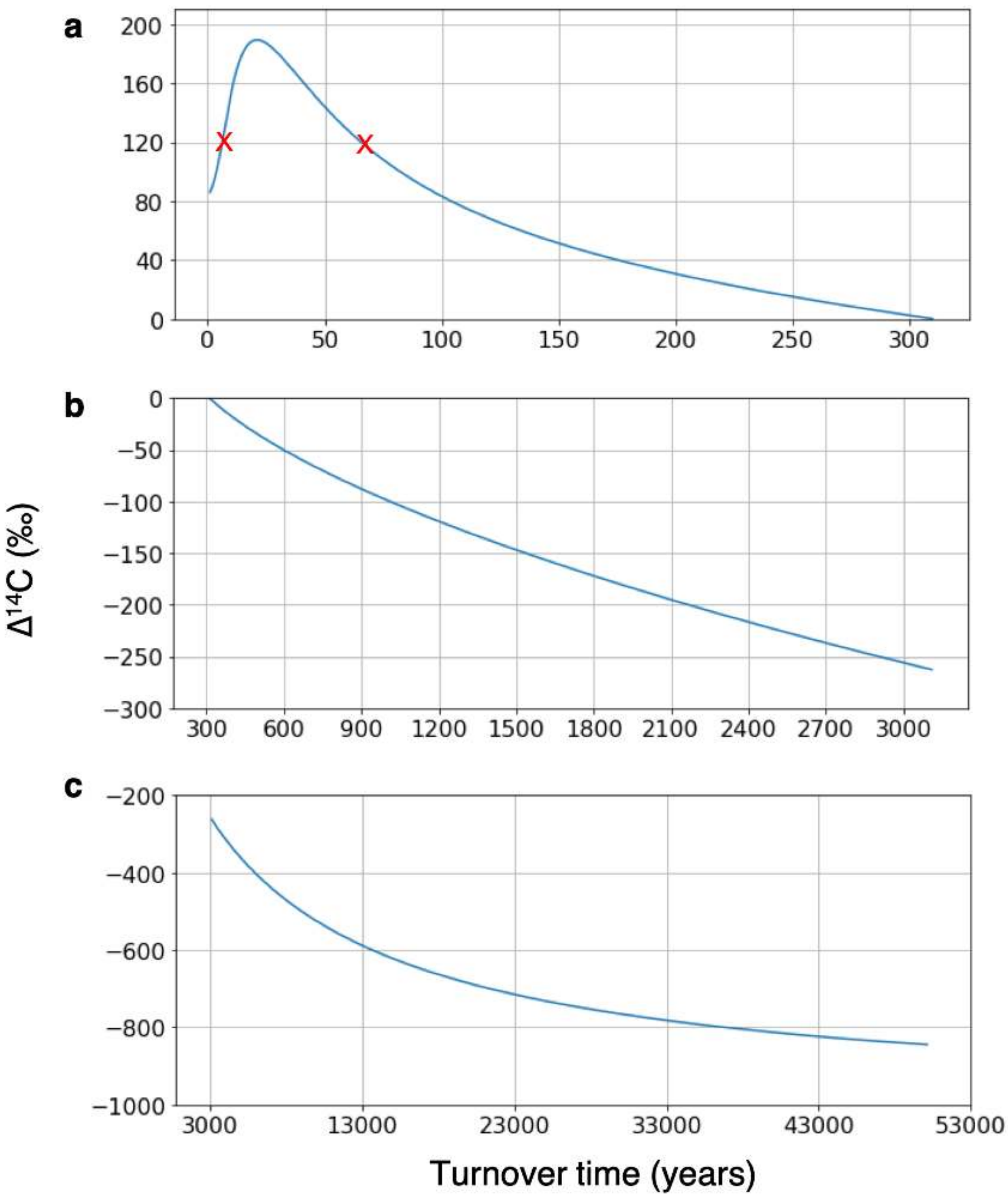
Supplementary Fig. 12. $\Delta^{14}C$ distribution of global soil carbon. Histograms show the distribution of carbon mass binned by $\Delta^{14}C$ for our data products (**a, b**) and the two global land models (ELM: **c, d** and CLM: **e, f**) at the biome level in the two depth intervals.



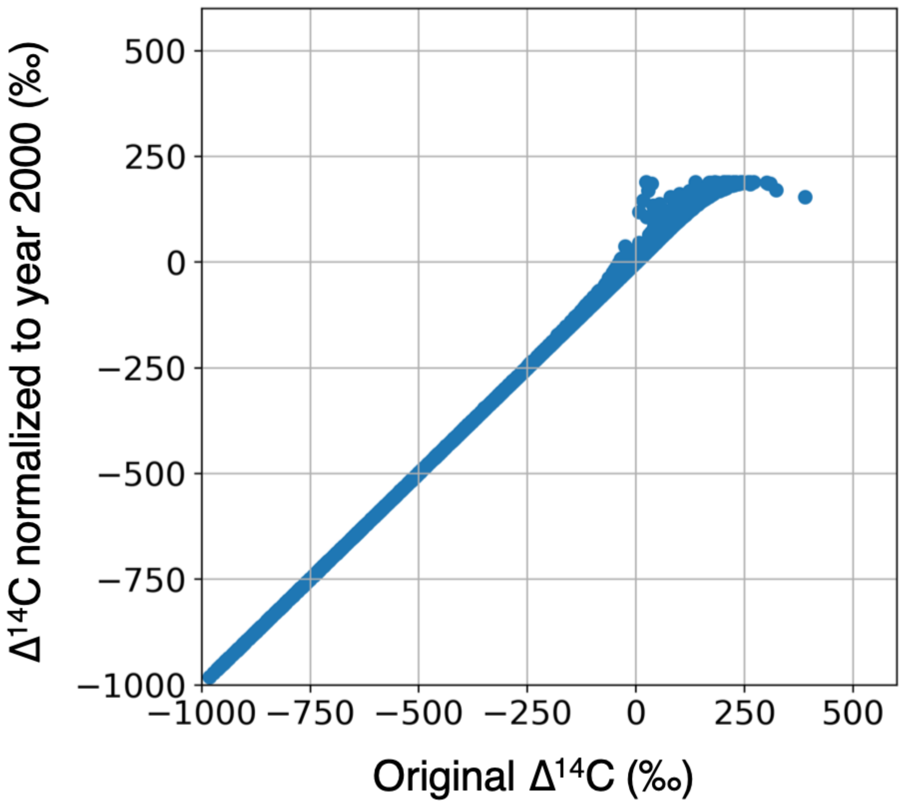
Supplementary Fig. 13. Global distribution of $\Delta^{14}C$. Comparisons between our global gridded products (**a, b**) and the two depth-resolved global land models ELM v1.0 (**c, d**) and CLM5 (**e, f**) in surface (0 – 30 cm) and subsurface soils (30 – 100 cm).



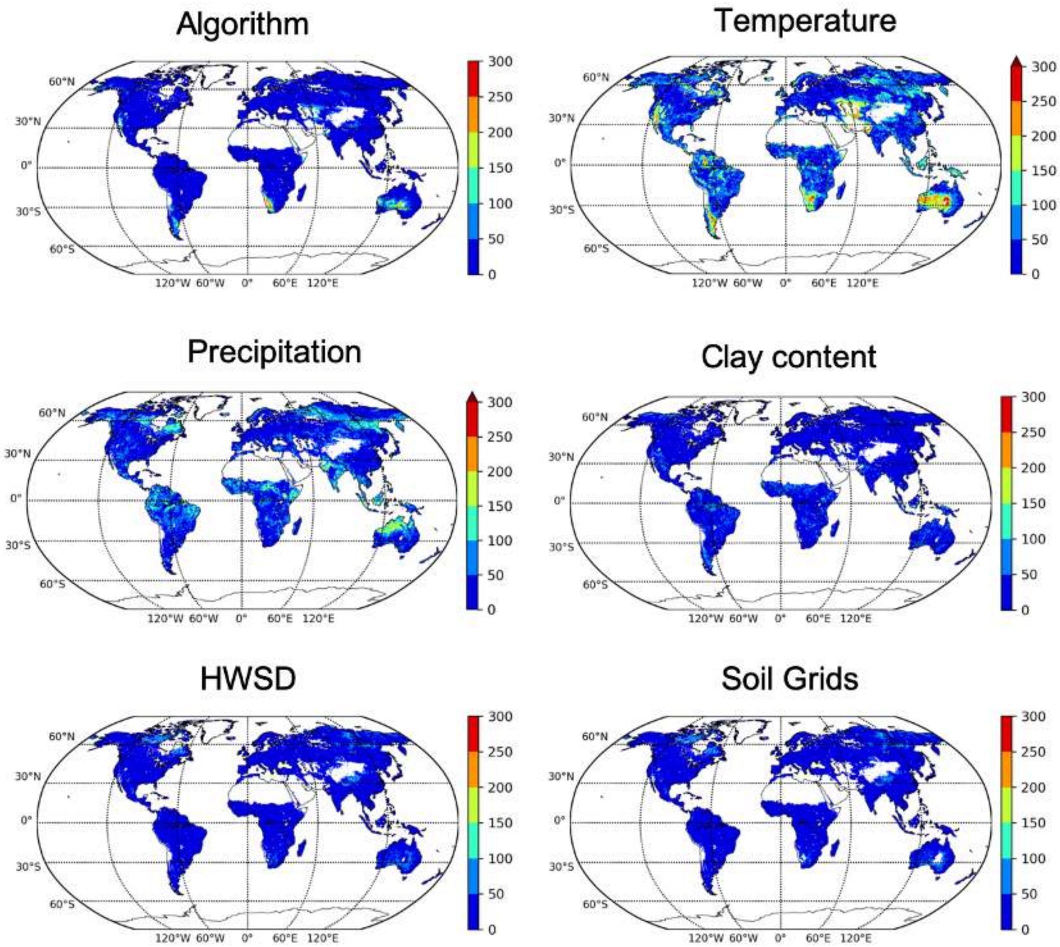
Supplementary Fig. 14. Land cover (a) and soil order (b) data used in this study. The land cover map was modified from MODIS Land cover MCD12Q1 product [38]. The 16 land cover types from MODIS were combined into 10 types for consistency with reported observations. All forests and woody savannas were re-categorized based on latitude as boreal (>50°N), temperate (> 23° and < 50°N and S) or tropical forests (< 23° N and S); open and closed shrublands were combined as shrubland (<50° N and S) or tundra (>50° N). The rest were unchanged. Note that desert and peatland were not included in the analysis due to small sample sizes.



Supplementary Fig. 15. Relationship between turnover time and $\Delta^{14}C$ in year 2000 generated by a one-pool steady state model. The relationships are for $\Delta^{14}C$ and turnover time up to 300 years (**a**), 3000 years (**b**), and 50000 years (**c**). Panel A shows the two possible solutions (red X's) for $\Delta^{14}C$ values greater than about 85‰. Turnover time and mean age are equivalent for a 1 pool model.



Supplementary Fig. 16. Comparison between the original $\Delta^{14}C$ and the $\Delta^{14}C$ after normalization to year 2000. Negative $\Delta^{14}C$ values are linearly related to normalized $\Delta^{14}C$, whereas there was a strong nonlinear relationship for positive $\Delta^{14}C$ values.



Supplementary Fig. 17. Uncertainty of $\Delta^{14}C$. We quantified uncertainty as the absolute difference with the data-derived global $\Delta^{14}C$ (0 – 100 cm). The absolute differences were calculated for each regression tree (algorithm). differences introduced by each driver was calculated while holding out the temperature, precipitation, or soil clay content, respectively. Absolute differences were also calculated for global-gridded $\Delta^{14}C$ weighted by the two different soil carbon datasets (HWSD and SoilGrids).

Supplementary Table 1 Summary statistics of soil carbon and $\Delta^{14}C$ in year 2000 in global land model CLM5. The estimates of $\Delta^{14}C$ for each biome are the median and 5% to 95% range (in brackets). Global mean and standard deviation (mean ± sd) of $\Delta^{14}C$ and mean age is weighted by soil carbon content in each biome.

	Surface soil (0 – 30 cm)		**Subsurface soil (30 – 100 cm)**	
Biome	Soil carbon (Pg C)	$\Delta^{14}C$ (‰)	Soil carbon (Pg C)	$\Delta^{14}C$ (‰)
Boreal Forest	113	5 (-108, 58)	72	-174 (-371, -81)
Temperate Forest	33	64 (8, 108)	17	-54 (-143, -5)
Tropical Forest	70	64 (44, 77)	30	-12 (-66, 2)
Grassland	41	66 (-41, 134)	22	-104 (-274, -21)
Cropland	85	80 (33, 121)	38	-68 (-172, -8)
Shrubland	18	69 (-1, 108)	9	-67 (-169, -10)
Savanna	63	77 (4, 103)	32	-31 (-167, 12)
Tundra	122	-202 (-520, 40)	184	-416 (-632, -149)
Global mean*	545	-10±46	404	-231±81

*: Global weighted $\Delta^{14}C$ was -104±65‰ for the soil down to 1 m depth.

Supplementary Table 2 Summary statistics of soil carbon and $\Delta^{14}C$ in year 2000 in global land model ELM1.0. The estimates of $\Delta^{14}C$ for each biome are the median and 5% to 95% range (in brackets). Global mean and standard deviation (mean ± sd) of $\Delta^{14}C$ and mean age is weighted by soil carbon content in each biome.

	Surface soil (0 – 30 cm)		**Subsurface soil (30 – 100 cm)**	
Biome	Soil carbon (Pg C)	$\Delta^{14}C$ (‰)	Soil carbon (Pg C)	$\Delta^{14}C$ (‰)
Boreal Forest	71	-29 (-241, 35)	75	-291 (-499, -160)
Temperate Forest	22	50 (-32, 85)	21	-123 (-272, -48)
Tropical Forest	76	80 (31, 94)	64	-49 (-115, -13)
Grassland	38	9 (-173, 97)	34	-255 (-532, -102)
Cropland	84	51 (-28, 99)	73	-125 (-298, -30)
Shrubland	13	48 (-48, 95)	11	-172 (-350, -84)
Savanna	44	65 (-24, 105)	37	-88 (-238, -12)
Tundra	76	-289 (-787, -22)	119	-523 (-750, -301)
Global mean*	424	-55±61	434	-285±60

*: Global weighted $\Delta^{14}C$ -169±65‰ for the soil down to 1 m depth.

Supplementary Table 3 Statistical model performance with five-fold cross validation. Using the assembled $\Delta^{14}C$ measurements, we applied generalized linear models (ordinary least square, ridge regression and lasso regression), support vector machines (e.g. support vector regression) and ensemble methods (e.g. random forests and gradient boosted regression tree). R^2 and mean absolute error were calculated from 5-fold cross-validation to assess model performance.

Models	**R^2**	**Mean absolute error (‰)**
Random forest	0.69±0.08	140.6±19.5
Gradient boosted regression tree	0.67±0.06	146.4±11.0
Support vector regression	0.58±0.12	163.1±22.1
Ordinary least square	0.56±0.11	168.1±22.7
Ridge regression	0.55±0.11	170.5±22.4
Lasso regression	0.54±0.10	172.0±22.2

Supplementary Table 4 Variables and data sources used in the random forest model

Variable	Product name	Original resolution	Reference
Mean annual temperature	Climatic Research Unit TS v. 3.23	0.5°	Harris *et al.* 2014 [39]
Mean annual precipitation	Climatic Research Unit TS v. 3.23	0.5°	Harris *et al.* 2014 [39]
Land cover	MODIS Land cover MCD12Q1	500 m	Friedl *et al.* 2010 [38]
Soil order	Global Soil Regions map	2'	FAO-UNESCO [40]
Soil clay content*	Global Soil Grids	250m	Hengl *et al.* 2017 [41]

*: Note that the SoilGrids database has been updated (December 24th, 2018) and data are available at https://landgis.opengeohub.org

Supplementary Table 5 Comparisons of $\Delta^{14}C$ at the biome level in different scenarios. Biome-level median $\Delta^{14}C$ (0 – 100 cm) was computed for each scenario. Baseline is our data-derived global $\Delta^{14}C$. The scenarios of temperature, precipitation, and clay content are estimated by holding out temperature, precipitation, and clay content in the random forest algorithm. The scenario of algorithm is the mean of ensemble trees (i.e., 300). HWSD and Soil Grids are the estimates weighted by HWSD and SoilGrids. (Unit of $\Delta^{14}C$: ‰).

	Boreal Forest	Temperate Forest	Tropical Forest	Grassland	Cropland	Shrubland	Savanna	Tundra	Total
Baseline	-237	-106	-116	-226	-171	-140	-122	-437	-244
Temperature	-216	-131	-174	-254	-186	-266	-152	-443	-269
Precipitation	-250	-109	-65	-205	-152	-124	-86	-468	-256
Clay content	-242	-98	-95	-234	-177	-136	-110	-445	-251
Algorithm	-223	-105	-118	-233	-174	-190	-123	-428	-250
HWSD	-230	-111	-114	-222	-172	-163	-124	-411	-205
Soil Grids	-245	-103	-118	-230	-170	-115	-120	-462	-280