Article Type: Primary Research Articles

Observation-based global soil heterotrophic respiration indicates underestimated turnover and sequestration of soil carbon by terrestrial ecosystem models

Yue He[1], Jinzhi Ding[2*], Tsechoe Dorji[2], Tao Wang[2], Juan Li[3], Pete Smith[4]

[1] *Sino-French Institute for Earth System Science, College of Urban and Environmental Sciences, Peking University, Beijing 100871, China.*

[2] *Key Laboratory of Alpine Ecology, Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing 100085, China.*

[3] *School of Life Sciences, Lanzhou University, Lanzhou 730000, China.*

[4] *Institute of Biological and Environmental Sciences, School of Biological Sciences, University of Aberdeen, 23 St Machar Drive, Room G45 Aberdeen, AB24 3UU, Scotland, UK.*

Keywords: soil heterotrophic respiration, carbon turnover time, Random Forest, carbon cycling, terrestrial carbon sink, terrestrial ecosystem models

Manuscript for *Global Change Biology*

* Correspondence: Dr. Jinzhi Ding, Email: jzding@itpcas.ac.cn

Abstract

Soil heterotrophic respiration (R_h) refers to the flux of CO_2 released from soil to atmosphere as a result of organic matter decomposition by soil microbes and fauna. As one of the major fluxes in the global carbon cycle, large uncertainties still exist in the estimation of global R_h, which further limits our current understanding of carbon accumulation in soils. Here, we applied a Random Forest algorithm to create a global dataset of soil R_h, by linking 761 field observations with both abiotic and biotic predictors. We estimated that global R_h was 48.8 ± 0.9 Pg C yr^{-1} for 1982-2018, which was 16% less than the ensemble mean (58.6 ± 9.9 Pg C yr^{-1}) of 16 terrestrial ecosystem models. By integrating our observational R_h with independent soil carbon stock datasets, we obtained a global mean soil carbon turnover time of 38.3 ± 11 yr. Using observation-based turnover times as a constraint, we found that terrestrial ecosystem models simulated faster carbon turnovers, leading to a 30% (74 Pg C) underestimation of terrestrial ecosystem carbon accumulation for the past century, which was especially pronounced at high latitudes. This underestimation is equivalent to 45% of the total carbon emissions (164 Pg C) caused by global land use change at the same time. Our analyses highlight the need to constrain ecosystem models using observation-based and locally adapted R_h values to obtain reliable projections of the carbon sink capacity of terrestrial ecosystems.

Introduction

Soil carbon sequestration, the process by which carbon dioxide (CO_2) acquired by plants from the atmosphere is transferred and stored into the soil carbon reservoir, can partially offset anthropogenic carbon emissions (Bond-Lamberty and Thomson, 2010; Stockmann et al., 2013). The capacity of soil carbon sequestration is mediated partly by carbon input through plant photosynthesis, and partly by soil carbon turnover through which the accumulated carbon is depleted by heterotrophic respiration (R_h) (Chen et al., 2015; Todd-Brown et al., 2013). Many previous studies have reported that the uncertainty in carbon turnover is greater than the uncertainty in the carbon input, and this has been proved to be the main reason for the large inter-model spreads of predicted carbon stock dynamics and their response to future climate change (Anav et al., 2013; Carvalhais et al., 2014; Todd-Brown et al., 2013; Wu et al., 2018). Thus, an improved quantification of soil R_h would be a key step towards the reliable prediction of real-world soil carbon dynamics.

R_h refers to the release of CO_2 from the soil to the atmosphere through the decomposition of organic matter by soil microbes and fauna (Bond-Lamberty et al., 2018; Chapin et al., 2006; Subke et al., 2006). As one of the major fluxes in the terrestrial carbon cycle, there is large uncertainty associated with current estimates of soil R_h. On the one hand, some previous studies estimated global R_h indirectly using an empirical relationship with soil respiration (R_s), where large uncertainties remain in the universality of the global constant R_s-R_h relationship (Baggs, 2006; Bond-Lamberty et

al., 2004; Hashimoto et al., 2015; Subke et al., 2006; Warner et al., 2019). For example, a widely used R_s-R_h equation provided by Bond-Lamberty et al. (2004) was established using only 54 forest sites, calling into question the applicability to non-forested ecosystems around the world. On the other hand, although many studies have upscaled global soil R_h directly based on putative environmental drivers, to our knowledge, the key role of soil biotic decomposers (e.g., soil microbes and fauna) have not been considered and investigated (Feng et al., 2022; Moinet et al., 2016; Tang et al. 2020a; Yan et al., 2018; Ye et al., 2019; Zhang and Zhang, 2016).The emergence of global gridded soil biotic property datasets has thus provided an opportunity to more accurately predict soil R_h (He et al., 2020; van den Hoogen et al., 2019).

In addition, terrestrial ecosystem model development has strongly focused on photosynthesis processes, whereas the soil carbon dynamics have shown a relatively weak convergence in modeling results (Carvalhais et al., 2014; Todd-Brown et al., 2013, 2014). For example, global R_h simulated by the MsTMIP models (Tian et al., 2015) varies from 35 to 69 Pg C yr^{-1}, values from the TRENDY models (Sitch et al., 2015) varies from 48 to 72 Pg C yr^{-1}, and those from the CMIP5 models (Hashimoto et al., 2015) varies from 42 to 73 Pg C yr^{-1}. Such uncertainty in soil R_h propagated through the model predictions could cause substantial variation in estimates of soil carbon sequestration, and further affects global carbon-climate feedbacks (Anav et al., 2013; Friend et al., 2014; Todd-Brown et al., 2013).

The primary goal of this study was to constrain the magnitude of simulated carbon sequestration capacity in terrestrial ecosystem models combined with R_h observations. Specifically, we first present a new dataset of soil R_h using a Random Forest (RF) algorithm with comprehensive in situ R_h observations and specific consideration of soil biotic factors. On this basis, we are then able to generate a new data-driven map of soil carbon turnover time (τ_{soil}) by adopting previously published soil carbon-stock (cSoil) datasets and the newly derived R_h. Taking these observation-based R_h and τ_{soil} as benchmarks, we the evaluate the performance of the terrestrial ecosystem models and constrained the simulated changes in terrestrial carbon storage during the past century. The schematic overview of the aforementioned procedures is shown in Figure S1.

Materials and Methods

Site-level R_h measurements

This work was mainly based on a global soil respiration database (SRDB) compiled by Bond-Lamberty and Thomson (2010a, b), which is currently on version 5 (https://github.com/bpbond/srdb; Jian et al., 2021). Due to the lack of data records included in SRDB for China because of language barriers, we additionally collected R_h observations from the China Knowledge Resource Integrated Database (CNKI; https://www.cnki.net/, last access: March 2020). We evaluated and filtered all the observations, retaining only those data records that matched the following criteria: (1) measurements were made for at least one year; (2) multi-year measurements within the same site were considered independent; (3) explicit geographic coordinates and the

measurement year were provided; (4) measurements with no experimental manipulation; (5) measurements based on alkali absorption were excluded because of the verified low accuracy of the method (Bekku et al., 1997; Bond-Lamberty et al., 2018; Haynes and Gower, 1995; Pumpanen et al., 2004). (6) four extremely high R_h measurements (lying outside 6 standard deviations of all the data records) were excluded as outliers from our analysis. In total, we retained a set of 761 site-level R_h observations from all around the world (Figure S2), including 56 data points from the tropics, 327 from temperate zones and 376 from boreal zones (Figure S3).

Predictor variables

In total, 19 predictors, including climate, vegetation, and soil abiotic and biotic variables, were used for the global up-scaling of soil R_h in our analysis (Table S1). Specifically, mean annual temperature (MAT) and mean annual precipitation (MAP) were acquired from the Climatic Research Unit (CRU TS 4.04; Harris et al., 2020), and dry month length (DryMonth) was defined using the CRU TS v4.04 dataset, as the number of months per year with monthly potential evapotranspiration larger than precipitation. Mean annual leaf area index (LAI) and growing season length (GSL) were derived from the Global Inventory Modeling and Mapping Studies version 4 (GIMMS v4; Zhu et al., 2013). Litter production (Litter) was derived from He et al. (2021) by upscaling field observations to the global scale. Note that the litter production data used in our study did not cover cropland; instead, we substituted the litter production with NPP in cropland areas, which may overestimate the carbon input to the

soil system since a large proportion of NPP is harvested for human use (Haberl et al., 2007). Annual nitrogen deposition (Ndep) was obtained from the North American Carbon Program Multiscale Synthesis and Terrestrial Model Intercomparison Project (MsTMIP; Huntzinger et al., 2013). Soil abiotic variables such as soil water content (SWC) were obtained from The Global Land Evaporation Amsterdam Model (GLEAM v3; Martens et al., 2016). Sand fractions (Sand), silt fractions (Silt), clay fractions (Clay), cation-exchange capacity (CEC), soil water PH (PH), and soil organic carbon content (SOC) in the 0-1m layer were all calculated from the SoilGrids250m project (Hengl et al., 2017). Total nitrogen density (TotN) and C: N ratio (C:N) were obtained from the ISRIC-WISE soil profile database (Batjes, 2016). We also applied two global gridded soil biotic property datasets which included the Fungal: Bacterial ratio (F:B) and nematode density (Nematode) from previously published upscaling studies (He et al., 2020; van den Hoogen et al., 2019). Also, we collected all these predictors for each given site from corresponding studies in the literature; if unavailable, they were extracted from global-scale gridded productions (Table S1) based on the specific site geographic coordinates and measurement years. All the predictor variables were then resampled to a common spatial resolution of 0.5° using the nearest neighbor method.

The vegetation distribution map used in this study is based on the 17 International Geosphere - Biosphere Programme (IGBP) vegetation classes derived from the Moderate Resolution Imaging Spectroradiometer (MODIS) land cover product (MCD12C1; Friedl et al., 2010) in 2001. These classes consist of 11 natural vegetation

types, 3 unnatural vegetation types and 3 non-vegetated types. Here, the 11 natural vegetation types were reclassified into 9 categories: Evergreen Needleleaf Forest (ENF); Evergreen Broadleaf Forest (EBF); Deciduous Needleleaf forest (DNF); Deciduous Broadleaf forest (DBF); Mixed forest (MF); Shrublands, including Open shrubland and Closed shrubland (Shrub); Savannas, including Woody Savanna and Savanna (Savanna); Grassland; Cropland, including Cropland and Cropland/Natural vegetation Mosaic (Crop). Grasslands, savannas and shrublands north of 55^{o} N were categorized as Tundra (Figure S2). In addition, bare ground and sparse vegetated areas, defined as pixels with 1982-2018 annual mean normalized difference vegetation index (NDVI) lower than 0.1, were masked out.

Geospatial modelling of R_h

Machine learning algorithms can directly "learn" information from data without assuming the functional relationship between dependent and independent variables in advance, and have thus been widely applied in geosciences in recent years, especially in up-scaling modelling of carbon and water fluxes (Jung et al., 2017; Yao et al., 2018; Zeng et al., 2012). To predict global R_h, we trained 761 R_h observations using all the 19 selected predictor variables via the RF model (Breiman, 2001). We fine-tuned the RF model for the following two parameters: ntree (the number of decision trees grown in RF; ntree = 100, 150 or 200) and mtry (the minimum number of variables randomly sampled per tree leaf; mtry = 2, 3, 4 or 5). To improve the RF model accuracy, we developed a stratified 10-fold cross validation, by grouping our dataset into three

175 climate zone classes (i.e., tropical, temperate and boreal zones) and ensuring each class
176 was approximately equally represented across each fold. The model performance was
177 then evaluated based on the goodness of fit (R^2), Nash-Sutcliffe efficiency (NSE) and
178 the root mean squared error (RMSE). Finally, we ranked all the prediction factors based
179 on the increased node purity and calculated the relative importance of each factor from
180 the percentage contribution of all the predictors. The average predicted map across the
181 RF model ensemble was used as the final product, and the standard deviation across all
182 the RF models was then considered to be an indication of the prediction uncertainty
183 (Figure S4). All the processes were analyzed using the "RandomForest" package in
184 RStudio.
185
186 To further detect the combined relative importance of the four predictor groups
187 (i.e., climate, vegetation, soil abiotic and soil biotic variables), we first conducted a
188 principal component analysis (PCA) on each group, ensuring the same number of
189 predictors in each group and removing the correlation between predictors. Note that
190 vegetation type was excluded from the PCA analysis since it is an unordered categorical
191 variable. All the predictors were normalized before the PCA analysis. Since there were
192 only two soil biotic properties used in this study, we used the first two principal
193 components (i.e., PC1 and PC2) from each group and the corresponding 761 soil R_h
194 records to reconstruct a new RF model (R^2 = 0.58). The relative importance for both
195 PC1 and PC2 from all the four groups were quantified. Finally, the sums of the relative
196 importance of each group was regarded as their combined importance (Figure S5).

Independent observational R_h and soil carbon stock datasets

Four previously derived data-driven global R_h datasets were used for comparison with the global R_h data produced in this study (Table S2). Specially, Hashimoto et al. (2015) firstly established a R_s model driven by monthly temperature and precipitation, and then indirectly derived global R_h based on a globally constant R_s-R_h empirical relationship. Similarly, Warner et al. (2019) produced two global R_h datasets based on two separate R_s-R_h empirical relationships. The average soil R_h derived from Warner et al. (2019) was used here. Tang et al. (2020a) directly upscaled global R_h using the RF algorithm, based on the SRDBv4 database and nine environmental abiotic variables. As an alternative, R_h was also estimated from a mass balance approach, using net ecosystem production (NEP) from atmospheric inversions, gross primary production (GPP) derived from solar-induced fluorescence, and carbon use efficiency (CUE) from a model-data fusion system (Konings et al., 2019); however, this method has large uncertainties stemming from these input datasets.

Recent advances in observational global soil carbon stock datasets provided us with a unique opportunity to create a new data-oriented map for τ_{soil} (Batjes, 2016; Bloom et al., 2016; Hengl et al., 2017; Nachtergaele et al., 2010; Sanderman et al., 2017). We used five different state-of-the-art global cSoil datasets for the 0-1m soil layer to estimate global τ_{soil}, assuming that all the terrestrial ecosystem models contained soil organic carbon within the top 1m soil layer (Todd-Brown et al., 2013).

Detailed information for each cSoil dataset is given in Table S3. Since global estimates of cSoil have large differences which may arise from the different upscaling methods and in situ observations used to derive global maps (Figure S6), we used the average estimate of all five cSoil datasets to ensure robustness.

In addition, the observation-based cSoil does not include the carbon stock of litter (cLitter), because the sample of soils were sieved and the remaining recognizable litter detritus were mainly removed before the determination of soil organic carbon. We thus combined the observation-based cSoil with an independent cLitter dataset from a Data Assimilation Linked Ecosystem Carbon Model (DALEC2; Bloom et al., 2016) to better represent the soil carbon dynamics and to be comparable with DGVMs.

Terrestrial ecosystem models

Global net primary productivity (NPP), R_h, vegetation carbon stock (cVeg) and cSoil outputs from 16 dynamic global vegetation models (DGVMs) were used in this study. The model outputs were obtained from the S2 experiments of the TRENDY (version 9) multi-model inter-comparison project, in which the models were forced by time-varying climate and CO_2 concentrations, while land use and land-use change were held constant at pre-industrial levels. For the DGVMs that have simulated cLitter pools, the cSoil value used here was estimated as the sum of cSoil and cLitter. All the DGVM outputs were harmonized to a common 0.5° grid using the nearest neighbor method. Detailed information for all 16 DGVMs is given in Table S4.

241

242 **Calculation of the vegetation and soil carbon turnover time**

243 At realistic non-steady state, τ is commonly defined as the ratio between the
244 storage of a carbon pool and its corresponding outflux based upon first-order kinetics
245 (Koven et al., 2015; Schwartz, 1979). Without considering natural and anthropogenic
246 induced disturbances, the outflux of the vegetation carbon pool is approximately equal
247 to litter production, and the outflux of the soil carbon pool is approximately equal to
248 soil R_h. Thus, the turnover time of vegetation and soil can be calculated as follows:

249 $$\tau_{veg} = \frac{cVeg}{I_{veg} - \triangle cVeg} = \frac{cVeg}{litter} \quad (1)$$

250 $$\tau_{soil} = \frac{cSoil}{I_{soil} - \triangle cSoil} = \frac{cSoil}{R_h} \quad (2)$$

251 where cVeg and cSoil are the size of vegetation and soil (including litter) carbon
252 pools, respectively; I_{veg} and I_{soil} refer to the influx of vegetation and soil carbon
253 pools, respectively; litter represents the litter production, and R_h represents the soil
254 heterotrophic respiration. When estimating the observation-based τ_{soil}, the uncertainty
255 range denotes that stemming from the inter-model standard deviation of five
256 observation-based soil carbon stock datasets.

257

258 **Constraining the changes in global terrestrial carbon stocks during 1901-2018**

259 Here, we developed a reduced-complexity model, the 2-box model (2BoxModel),
260 to approximate the DGVM-based vegetation and soil carbon dynamics during the past
261 century for each of the 16 terrestrial ecosystem models. The 2BoxModel assumes that
262 all the output carbon fluxes of the vegetation carbon pool enter the soil carbon pool,

and all the output carbon fluxes of the soil carbon pool are released to the atmosphere in the form of soil R_h. Therefore, we can simplify the carbon-cycle processes of the terrestrial ecosystem into the following vegetation-soil carbon turnover theoretical formulations:

$$\begin{cases} \frac{\mathrm{dcVeg}}{\mathrm{d}t} = \mathrm{NPP}(t) - \frac{\mathrm{cVeg}(t)}{\tau_{\mathrm{veg}}(t)} \\ \frac{\mathrm{dcSoil}}{\mathrm{d}t} = \frac{\mathrm{cVeg}(t)}{\tau_{\mathrm{veg}}(t)} - \frac{\mathrm{cSoil}(t)}{\tau_{\mathrm{soil}}(t)} \end{cases} \quad (3)$$

where,

cVeg(t) denotes the vegetation carbon stock in year t;

cSoil(t) denotes the soil carbon stock (including the litter carbon stock) in year t;

NPP(t) denotes the net primary productivity in year t;

$\tau_{veg}(t)$ denotes the vegetation carbon turnover time in year t;

$\tau_{soil}(t)$ denotes the soil carbon turnover time in year t;

(t represents the year within the period 1902-2018, with t=2, 3, …119)

Second, we assumed that both the DGVM-simulated τ_{veg} and τ_{soil} in 1901 were the same as the values in 1902, and both the DGVM-simulated cVeg and cSoil were in a steady state in 1901. Thus, cVeg(1) and cSoil(1) can be estimated as NPP(1) × $\tau_{veg}(1)$, NPP(1) × $\tau_{soil}(1)$, respectively. We then reconstructed the global terrestrial ecosystem carbon stocks (i.e., cTotal = cVeg + cSoil) for 1902-2018 for each DGVM using Eq.(3). As a result, we found that the 2BoxModel could generally emulate the terrestrial carbon dynamics of the original DGVM outputs (Figures S7a and S8a).

Third, we replaced the DGVM-simulated τ_{veg} and τ_{soil} during 1901-2018 with the observation-based τ_{veg} (10.3 ± 1.4 yr; He et al., 2021) and τ_{soil} (38.3 ± 11 yr), both of which were then used as constraints, while leaving the modelled interannual variation and trend of carbon turnover times unchanged in the 2BoxModel. Specifically, since the DGVM-simulated τ_{veg} and τ_{soil} were generally biased compared to observations, we calculated a correction factor, defined as the difference between observation-based turnover times and DGVM-simulated turnover times, to adjust the DGVM-simulated results (Eq.(4) and (5)). The global correction factors for each model were then used to recalculate τ_{veg} and τ_{soil} during 1901-2018. Using these bias-corrected carbon turnover times, we could thus constrain both the vegetation carbon stock changes (ΔcVeg) and soil carbon stock changes (ΔcSoil) relative to 1901 from the original DGVM model outputs via the 2BoxModel (Figures S7b and S8b). The sum of ΔcVeg and ΔcSoil represents the total terrestrial carbon accumulation (ΔcTotal) relative to 1901 after constraint.

$$\tilde{\tau}_{veg_{(t)}}^{sim} = (\tau_{veg_{(t0)}}^{obs} - \tau_{veg_{(t0)}}^{sim}) + \tau_{veg_{(t)}}^{sim} \quad (4)$$

$$\tilde{\tau}_{soil_{(t)}}^{sim} = (\tau_{soil_{(t0)}}^{obs} - \tau_{soil_{(t0)}}^{sim}) + \tau_{soil_{(t)}}^{sim} \quad (5)$$

where,

$\tau_{veg_{(t0)}}^{obs}$ and $\tau_{soil_{(t0)}}^{obs}$ denote the multi-year global mean observation-based vegetation and soil carbon turnover time during 1982-2018, respectively.

$\tau_{veg_{(t0)}}^{sim}$ and $\tau_{soil_{(t0)}}^{sim}$ denote the multi-year global mean DGVM-simulated vegetation and soil carbon turnover time during 1982-2018, respectively.

$\tau_{veg_{(t)}}^{sim}$ and $\tau_{soil_{(t)}}^{sim}$ denote the yearly global mean DGVM-simulated vegetation and

306 soil carbon turnover time during 1901-2018, respectively.

307

308 To assess the robustness of the modelled terrestrial carbon stock changes during
309 the past century implied by the uncertainty of observation-based carbon turnover-time
310 constraints, we conducted a Monte Carlo sensitivity experiment. Specifically,
311 observation-based τ_{veg} values were randomly drawn from a normal distribution with a
312 mean equal to 10.3 and standard deviation equal to 1.4 by 100 Monte Carlo bootstrap
313 samples. Similarly, observation-based τ_{soil} values were randomly drawn from a normal
314 distribution with a mean equal to 38.3 and standard deviation equal to 11 by 100 Monte
315 Carlo bootstrap samples. We thus obtained a 10000-strong ensemble of global terrestrial
316 ecosystem carbon stocks for each model during 1901-2018 from crossing the
317 combinations of 100 τ_{veg} samples and 100 τ_{soil} samples used as constraints in the
318 2BoxModel. For each model, the average constrained terrestrial ecosystem carbon
319 stocks from the 10000-strong 2BoxModel ensemble was used as the final prediction,
320 and the standard deviation was regarded as an indication of the model uncertainty.

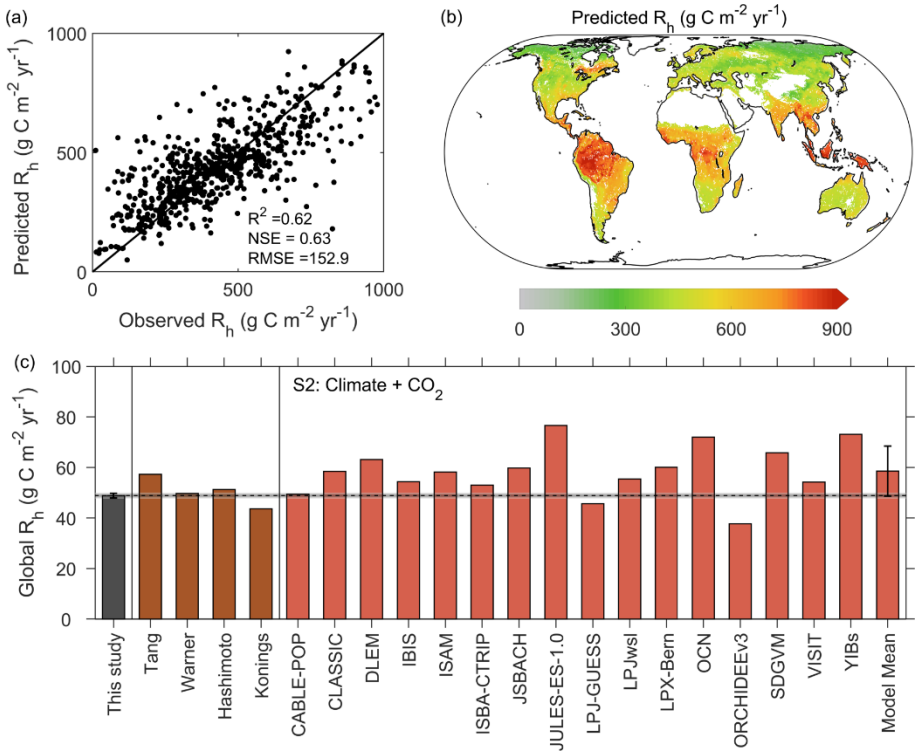
321

322 In addition, we repeated the above steps for each vegetation type, rather than the
323 entire global vegetated area, by replacing the DGVM-simulated τ_{veg} and τ_{soil} with the
324 corresponding observation-based values and tested the contribution of each vegetation
325 type to the changes of global total cVeg and cSoil after constraint (Figure S9). Cropland
326 was excluded from this vegetation-specific analysis, as this non-natural vegetation type
327 is strongly affected by human activities.

Results

A new global estimate of soil R_h

A new observation-based, global value of R_h for 1982-2018 was estimated at 48.8 ± 0.9 Pg C yr^{-1} (mean ± standard deviation), which is close to the ensemble mean (50.5 ± 4.9 Pg C yr^{-1}) obtained by previous studies (Figure 1; Table S2). The global distribution of R_h has a latitudinal gradient with higher values occurring in the tropics (Figure S10a). The highest values occur in South and Central America, Africa, and south and east Asia, while the lowest values are mainly at high latitudes or in arid regions with low water availability. Specifically, we used a RF model by linking 19 putative predictors with 761 field observations of R_h distributed around the world (Figure S2): a much larger data set than those used in previous studies (Table S2; see Materials and Methods). The final stratified 10-fold cross validation revealed good performance of our RF model in predicting soil R_h, with the goodness of fit (R^2), Nash-Sutcliffe efficiency (NSE) and the root mean squared error (RMSE) of 0.62, 0.63 and 152.9 g C m^{-2} yr^{-1}, respectively (Figure 1a).



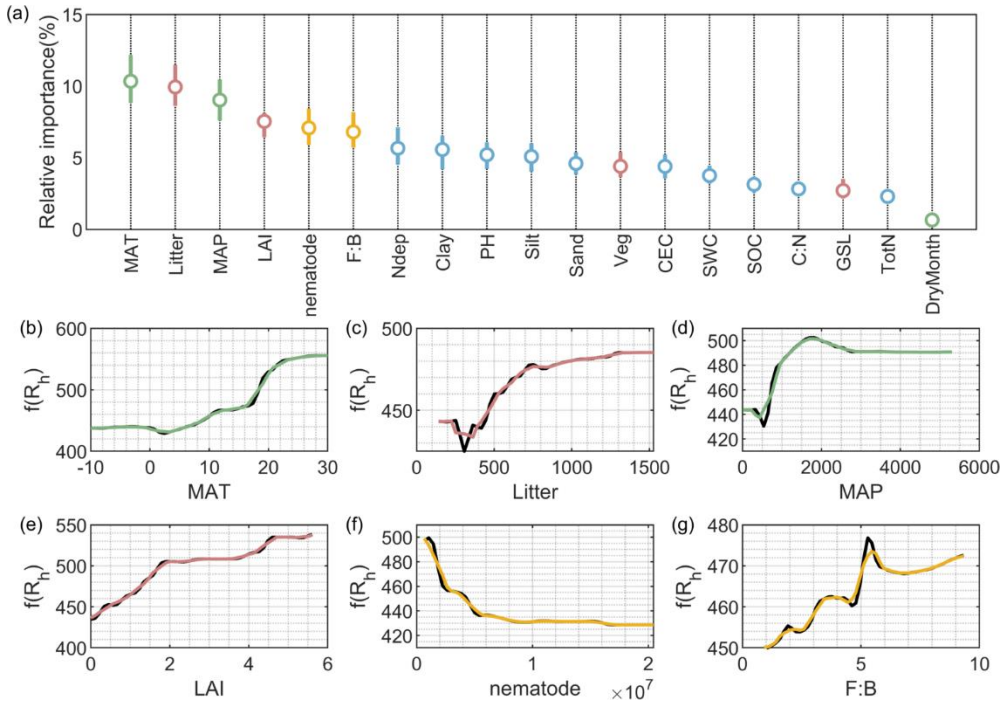
345 **Figure 1.** Global distribution of soil heterotrophic respiration (R_h) derived from
346 observation-based estimates and terrestrial ecosystem models. **a.** Comparison between
347 predicted and observed annual mean R_h using Random Forest (RF) algorithm. **b.** Global
348 distribution of annual mean R_h predicted by RF during 1982-2018. **c.** Comparison of
349 global mean R_h from this study with the other four previous empirical estimates and 16
350 terrestrial ecosystem models. The horizontal dashed line and the shaded area indicate
351 our newly estimated global R_h and its uncertainty range over 1982-2018, respectively.
352 Four previous empirical estimates of global R_h are all recalculated from corresponding
353 publications to match the temporal extent of this study as closely as possible (details in
354 Table S2). The error bar on the multi-model mean R_h denotes the inter-model standard
355 deviation.
356

Using this observation-based R_h as a benchmark, we evaluated the performance of 16 process-based DGVMs and found that most of them (14 out of 16) overestimated the global R_h. The multi-model average of global soil R_h (58.6 ± 9.9 Pg C yr^{-1}) was 20% higher than our observation-based estimate (Figure 1c). Additionally, when the spatial similarities between our observation-based and the DGVM-simulated R_h were examined using a Taylor diagram (Taylor, 2001) (Figure S11), we found that the correlation coefficient (R) ranged from 0.41 (LPJwsl) to 0.72 (LPJ-Bern), the standard deviation (SD) of the DGVM-simulated values ranged from 154 g C m^{-2} yr^{-1} (LPJ-GUESS) to 441 g C m^{-2} yr^{-1} (YIBs), whilst the RMSE of the DGVM-simulated global R_h ranged from 137 g C m^{-2} yr^{-1} (LPJ-GUESS) to 385 g C m^{-2} yr^{-1} (YIBs). In contrast, closer spatial agreement between our newly-derived R_h and the other four previous empirical estimates were observed, with the spatial correlation coefficient ranging from 0.62 (Konings et al., 2019) to 0.87 (Tang et al., 2020a). The SD of the four previous R_h estimates ranged from 111 g C m^{-2} yr^{-1} (Hashimoto et al., 2015) to 312 g C m^{-2} yr^{-1} (Konings et al., 2019), while the RMSE of them ranged from 70 g C m^{-2} yr^{-1} (Tang et al., 2020a) to 253 g C m^{-2} yr^{-1} (Konings et al., 2019). None of the DGVMs reproduced the global pattern of the observation-based global R_h well, suggesting poor performance of the DGVMs in reproducing soil R_h.

Dominant factors controlling R_h changes

The relative importance of the main factors for predicting soil R_h are displayed in Figure 2. Here we focused on the six most important factors: Mean annual temperature

(MAT), litter production (Litter), mean annual precipitation (MAP), leaf area index (LAI), soil Fungal: Bacterial ratio (F:B) and nematode density (Nematode) (Figure 2a). Generally, climate condition is the most important factor shaping R_h variations (29.4%; Figure S5). The partial dependence analyses showed that MAT had a strong positive influence on R_h, with this influence being strongest in the middle range of temperatures (around 0~20°C; Figure 2b). MAP also exerted a strong positive control on R_h near the lower end of its range of values, but this gradually decreased towards the higher end of the range (Figure 2d). The overall effect of vegetation properties played a secondary role (25.6%; Figure S5), where Litter and LAI are the two most important vegetation-related predictors. Litter had a strong positive control on R_h at relatively low values of litter and this gradually vanished at higher values (Figure 2c, d). For the case of LAI, a proxy for vegetation productivity, we found that soil R_h linearly increased with increasing LAI (Figure 2e).



393 **Figure 2.** The performance of Random Forest (RF) trained with heterotrophic
394 respiration (R_h) observations and predictors. **a**. Relative importance of all the predictors.
395 Values shown are the mean relative importance across all the RF models, with error
396 bars denoting 1-sigma standard deviation (see Materials and Methods). Climate,
397 vegetation, soil abiotic and biotic properties are shown in green, red, blue and orange,
398 respectively. **b-g**. Partial-dependence plots for the six most important predictors of soil
399 R_h in the RF model. The y-axis represents the marginal effect of each predictor while
400 holding all other predictors constant, on the predicted R_h (i.e., $f(R_h)$). Black lines and
401 blue lines denote raw and smoothed partial dependence, respectively. Abbreviations:
402 CEC, Cation-exchange capacity; C:N, C: N ratio; Clay, Clay fractions (Clay);
403 DryMonth, Dry month length; F:B, Fungal: Bacterial ratio; GSL, Growing season
404 length; Litter, Litter production; LAI, Mean annual leaf area index; MAP, Mean annual

precipitation; MAT, Mean annual temperature; Ndep, Annual nitrogen deposition; Nematode, Nematode Density; PH, Soil water PH; Veg, Vegetation type; Sand, Sand fractions; Silt, Silt fractions; SWC, Soil water content; SOC, Soil organic carbon content; TotN, Total nitrogen density.

The role of soil carbon decomposers associated with the soil carbon cycle was identified in our study (Figure 2a, f, g). Compared to soil abiotic variables, the soil biotic variables (Nematode and F:B) had stronger effects on the spatial variations of soil R_h (Figure 2f, g). F:B showed generally positive correlations with R_h, although there were fluctuations detected right across the range of values (Figure 2g). The density of nematodes showed a negative exponential relationship to the frequencies of R_h with a sharp decrease in its lower range, indicating little influence of nematodes on soil R_h when the density of nematodes was greater than 8×10^6 m^{-2} (Figure 2f). It's worth noting that although the rank of the relative individual influence of soil abiotic properties is small, their total relative contribution (22.5%) is nearly the same to that of soil biotic variables (22.6%) (Figure S5).

A global data-oriented map of soil carbon turnover time

As shown in Figure 3a, global mean τ_{soil} inferred from the observation-based R_h was 38.3 ± 11 yr (mean ± standard deviation) for the 0-1m soil layer during 1982-2018. The τ_{soil} value in boreal zones was generally higher than that in tropical zones, varying as a monotonically-increasing function of latitude (Figure 3 and S10b). Our new

estimate of τ_{soil} was comparable to the other three independent observation-based estimates (see Materials and Methods) which ranged from 33 yr (Tang et al., 2020a) to 38 yr (Warner et al., 2019). However, the τ_{soil} value of 43 yr based on Konings et al. (2019) was substantially higher than our estimate (Figure 3).

An evaluation of the DGVM performance found that all the simulated τ_{soil} were smaller than our observation-based τ_{soil}. The multi-model mean global τ_{soil} was 23 ± 13 yr, with values from individual models ranging from 13 yr (JSBACH) to 36 yr (ISBA-CTRIP) (Figure 3c). Moreover, the simulated τ_{soil} showed faster carbon turnovers in tropical forest regions and at high latitudes. Interestingly, the largest inter-model variability (Figure S10b) and the largest data-model differences were both most pronounced over the high latitudes of Northern Hemisphere (Figure 3b).

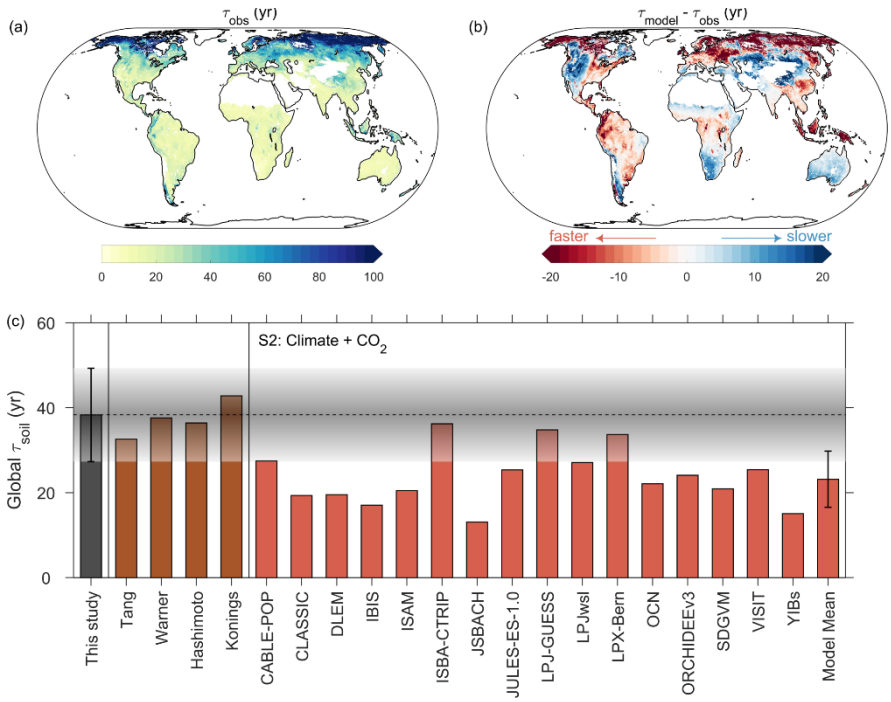


Figure 3. Global distribution of soil carbon turnover time (τ_{soil}) derived from observation-based estimates and terrestrial ecosystem models. **a**. Global distribution of annual mean observation-based τ_{soil} during 1982-2018. **b**. The difference between the multi-model mean (τ_{model}) and observation-based soil carbon turnover time (τ_{obs}). **c**. Comparison of global annual mean τ_{soil} from this study with the other four previous empirical estimates and 16 terrestrial ecosystem models. The horizontal dashed line and the shaded area indicate our newly estimated global τ_{soil} and its uncertainty range over 1982-2018, respectively. The error bar on the multi-model mean τ_{soil} denotes the inter-model standard deviation.

Constrained terrestrial carbon sequestration by terrestrial ecosystem models

Generally, we concluded that DGVMs exhibit a poor performance in reproducing soil carbon dynamics (Figures 1 and 3). Driven by soil R_h of organic carbon substrates, τ_{soil} is largely underestimated by DGVMs, suggesting that soil organic carbon might decompose too rapidly in model simulations. The vegetation carbon turnover time (τ_{veg}) might also be underestimated. A previous study has verified that there is a general tendency for DGVMs to have an unrealistically fast turnover of global vegetation carbon, especially at high latitudes (He et al., 2021). However, the degree to which the high rates of vegetation and soil turnover may affect the accumulation of terrestrial carbon in the models remains an unanswered question. It is, therefore, essential to use observation-based τ_{veg} and τ_{soil} to constrain the terrestrial ecosystem models and ultimately gain a better prediction of the terrestrial carbon sequestration capacity. We

thus applied a reduced complexity 2-box model (2BoxModel) that integrated 16 terrestrial ecosystem models with observation-based τ_{veg} (He at al., 2021) and τ_{soil} for the 0-1m soil layer to constrain global terrestrial ecosystem carbon dynamics during 1901-2018. The results showed that the 2BoxModel could generally emulate the terrestrial carbon dynamics obtained from the original DGVM outputs (Figures 4, S7 and S8), indicating its reliability (see Materials and Methods).

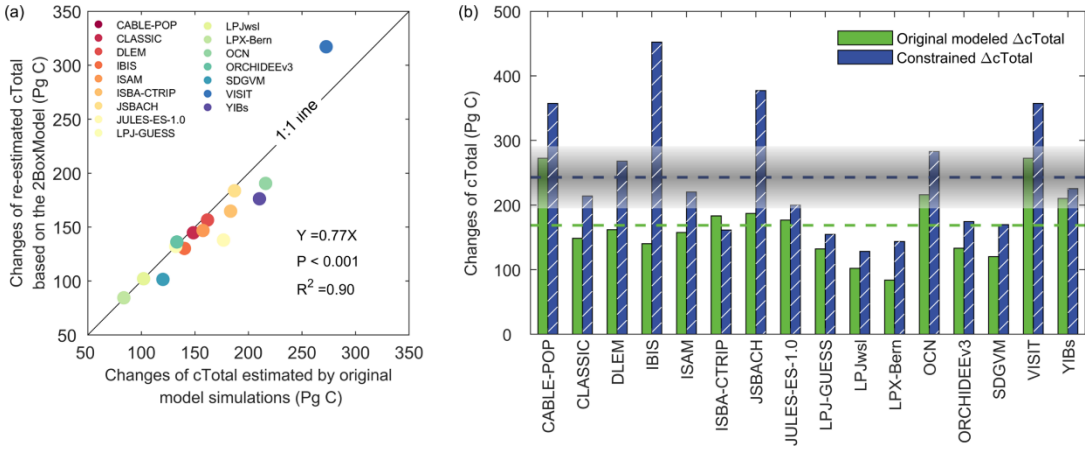


Figure 4. Comparison of the changes in ecosystem total carbon stock (ΔcTotal) from the dynamic global vegetation models (DGVMs) and the corresponding constrained values based on the 2BoxModel during 1901-2018. **a**. Reproduction of the modelled ΔcTotal using DGVM-simulated vegetation and soil carbon turnover times based on the 2BoxModel. **b**. Constraining the DGVM-simulated ΔcTotal by correcting the mean biases of vegetation and soil carbon turnover times based on the 2BoxModel. The horizontal green and blue dashed lines denote the multi-model mean ΔcTotal before and after constraint, respectively. The shaded area indicates the uncertainty range (± 1 standard deviation) of the constrained ΔcTotal.

By correcting the biases of τ_{veg} and τ_{soil} for each model, we found that the underestimation of carbon turnover times resulted in a 30% (74 Pg C) underestimation of the accumulation of terrestrial ecosystem carbon storage changes relative to 1901 (ΔcTotal; from 170 Pg C of changes in the original DGVM-simulated ensemble mean to 244 Pg C of changes after constraint). The underestimation of ΔcTotal is equivalent to 45% of the total cumulative carbon emissions (164 Pg C, Friedlingstein et al., 2020) caused by global land use change during 1901-2018. Specifically, the unconstrained and observational τ_{veg}-constrained outputs (99 vs 102 Pg C) showed only minor changes in the accumulation of vegetation carbon stock (ΔcVeg; Figure S7b). In contrast, the unconstrained and observational τ_{soil}-constrained outputs (71 vs 142 Pg C) showed a significant difference in the accumulation of soil carbon stock (ΔcSoil; Figure S8b), indicating that underestimated τ_{soil} was the main reason for the underestimation of the DGVM-simulated ecosystem carbon sink capacity.

The underestimation of ΔcTotal was particularly pronounced in permafrost regions with an underestimation of 51% in Tundra and 50% in Evergreen Needleleaf Forest, respectively (Figure S9). These results suggest that, to accurately simulate future atmospheric carbon dioxide dynamics and the carbon-concentration feedback, the representation of carbon decomposition processes and turnover time in DGVMs requires improvement, particularly in high-latitude regions (Wider et al., 2013; Wider et al., 2014). Of particular note is that, in the 2BoxModel, the turnover of both soil and

vegetation carbon pools are represented as first-order decay processes, with only one pool each for the vegetation and soil systems. This approach may be too simplified, and its applicability to multi-pool systems remains to be evaluated.

Discussion

Attribution of R_h variations to soil abiotic and biotic factors

Soil respiration involves many complex biogeophysical and biogeochemical processes that are either directly or indirectly regulated by a variety of factors (Davidson and Janssens, 2006). Similar to many previous studies, we found that both climate (e.g., MAT and MAP) and vegetation (e.g., LAI and litter production) significantly affected soil R_h variations (Hursh et al., 2017; Tang et al., 2020a, b; Wang et al., 2010), but we also identified the critical role of soil biotic properties on soil carbon decomposition, which has, in the past, generally been ignored when estimating soil R_h.

Specifically, we found that the variation of soil R_h was highly related to the structure and dynamics of soil microbe and fauna communities, with these soil biotic factors having a larger influence than any single soil abiotic factors (Figure 2). F:B is often used as an important indicator of the changes in microbial community structure and functionality (Bardgett and McAlister, 1999; Gordon et al., 2008; He et al., 2020; Rousk et al., 2010), characterizing the relative dominance of the two most important microorganism groups, i.e., fungi and bacteria. Since the carbon utilization strategies

522 of fungi and bacteria are different, changes in F:B would further affect the
523 decomposition rate of soil carbon (Strickland and Rousk, 2010). First, there is no doubt
524 that soil systems under warm and moist climate hold larger soil R_h, since favoruable
525 hydrothermal environments and abundant plant litter input support microbial growth
526 and metabolism. In addition, fungi are more abundant in the soils with abundant plant
527 residues because they are mainly responsible for the initial decomposition of these
528 residues and are also competent to break and decompose recalcitrant organic
529 components, whereas bacteria are considered to favour easily decomposed organic soils
530 with organic matter of low C:N (De Boer et al., 2005; Gao et al., 2018; Högberg et al.,
531 2007; Lauber et al., 2008; Meidute et al., 2008; Schneider et al., 2012). Therefore, as
532 our results showed, greater soil R_h is accompanied by a higher F:B. This positive
533 relationship could also be found among vegetation types, as forests characterized by
534 larger R_h are generally associated with higher F:B, while grasslands characterized by
535 lower R_h are typically dominated by bacteria (Deng et al., 2018).
536
537 In addition, soil fauna can also regulate soil respiration but *via* complex pathways.
538 On the one hand, soil R_h could be increased by accelerated substrates decomposition
539 through fragmentation (Ohashi et al., 2017), while, on the other hand, R_h could be
540 indirectly, negatively affected by grazing on microbial biomass (Cragg and Bardgett,
541 2001; Eisenhauer et al., 2011). Our results show that soil R_h decreased in response to
542 the increase in nematode density (Figure 2f), that is to say, the decrease of soil R_h
543 induced by the suppressed decomposition through overgrazing was greater than the

increase caused by the enhanced fragmentation capability of nematodes. Overall, we found compelling evidence that the soil fauna and the microbial community influence the broad pattern of decomposition and organic matter turnover in soil, which emphasizes the need to incorporate and improve soil biotic diagnostics into terrestrial ecosystem model evaluation in future studies.

Although none of a single soil abiotic factor rank the top six important predictors of soil R_h, their combined effect is non-negligible (Figure S5). For example, soil texture types (proportions of sand, silt and clay content) exert a strong influence in regulating soil decomposition rate. Soils with a larger percentage of clay content commonly have larger specific surface area, and thus have a greater ability to hold and retain water and nutrients, as well as contact with microbes and enzymes (Balogh et al., 2011; Xu and Shang, 2016). Soil R_h is also affected by N deposition, which favours the rate of mineralization in N-restricted ecosystems, and thus promoting soil R_h (Allen and Schlesinger, 2004), while excessive N deposition may lead to soil nutrient imbalance and soil acidification (Tian and Niu, 2015). In addition, although other substrate quality, such as litter nitrogen concentration, is also an important determinant of soil R_h, its spatial gridded information remains unavailable, highlighting that the development of global databases of such properties is important.

Poor performance of soil carbon dynamics by terrestrial ecosystem models

Based on 761 observations of soil R_h and 19 predictors, we mapped annual R_h

566 across the world at 0.5° resolution using a RF algorithm from 1982-2018. Subsequently,
567 we provided a spatially explicit estimate of τ_{soil} using the newly derived R_h. These
568 observation-based estimates indicated varying degrees of incapability of the terrestrial
569 ecosystem models to reproduce the spatial characteristics of both observational R_h and
570 τ_{soil}. In general, the DGVMs underestimated τ_{soil} due to the overestimation of R_h, as
571 indicated by the significant correlation between the deviations of the modelled R_h to
572 the observed R_h and those of τ_{soil} (Figure S12). However, the deviations in modelled
573 τ_{soil} could be attributed to biases in either the soil carbon pool or R_h or both. Therefore,
574 for each grid cell over the globe, we estimated the percentage bias of modelled τ_{soil}
575 dependent on the bias of R_h only, as well as the percentage bias dependent on the bias
576 of the soil carbon pool only. The results showed that R_h was the dominant factor
577 affecting the bias of modelled τ_{soil} over 62% of the global vegetated area (Figure S13).
578
579 This large discrepancy between modelled and observed R_h (as well as the resultant
580 τ_{soil}) is partly a consequence of the incomplete representation of key biogeochemical
581 processes, e.g. soil biotic activities, in DGVMs. Our results demonstrate that the
582 variation of soil R_h was strongly related to modifications of soil microbe and fauna
583 communities. Most DGVMs, however, did not explicitly represent the influence of the
584 soil microbial and faunal communities on soil carbon decomposition, indicating the
585 significant challenge facing the modeling of soil organic carbon dynamics (Schmidt et
586 al., 2011; Wieder et al., 2013, 2014). Instead, they simulate the release of CO_2 from soil
587 to atmosphere based solely on abiotic functions, by linking soil carbon content to the

588 controlling drivers such as temperature, soil moisture, as well as substrate content (Shao
589 et al., 2013). The main reason for the exclusion of microbial and faunal physiology in
590 models is the lack of mechanistic understanding of their complex feedback responses.
591 It is notable that a growing number of ecologically meaningful and functionally relevant
592 microbial models have been proposed, but with great differences in expression forms.
593 For example, Wieder et al. (2015) constrained the decomposition rates of two litter and
594 three soil organic matter pools based on an introduction of microbial functional groups
595 (i.e., copiotrophic vs. oligotrophic growth strategies), while Wang et al. (2013) directly
596 simulated the dynamics of physically measurable soil carbon pools, such as particulate,
597 dissolved, mineral-associated organic matter and microbial biomass. In addition, soil
598 fauna also plays a key role in soil carbon dynamics through directly or indirectly
599 influencing the decomposition of substrates and the structures and activities of
600 microbial community (Grandy et al., 2016). So far, the feasibility of the microbial-
601 explicit models, let alone the inclusion of soil fauna with more complex food webs, is
602 largely unknown and warrants further investigation (Bradford et al., 2016). A more
603 comprehensive understanding of the soil carbon dynamics, especially the faunal-
604 microbial interactions, is thus required, potentially through the continued cooperation
605 between the experimentalists and modellers, to further provide robust mechanism and
606 parameterization support for modelling studies.
607
608 Moreover, our results showed that larger uncertainties of both R_h and τ_{soil} occurred
609 in high latitude regions. This may be related to the poor performance for current

terrestrial ecosystem models in simulating permafrost dynamics, e.g., soil free-thaw processes which could regulate soil decomposition rates (Koven et al., 2011; Schuur et al., 2015; Todd-Brown et al., 2013; Yan et al., 2018). In addition, the vertical distribution of soil organic carbon and associated depth-dependent environmental controls (e.g., active layer thickness in permafrost areas) have been largely overlooked in current DGVMs (Tian et al., 2015). Previous studies generally assumed that the modelled soil carbon storage was contained within 0-1m in order to simplify the comparison with observation-based cSoil datasets. However, there is a huge amount of "old carbon" held in deeper soil layers, especially in permafrost regions (Mishra et al., 2021). The Northern Circumpolar Soil Carbon Database (NCSCD) indicated that permafrost region alone contained 472 Pg C in 0-1m depth, but 1035 Pg C in 0-3m depth (Hugelius et al., 2014). Given the large temperature increases, along with dramatically increased decomposition rates of thawed soil carbon at high-latitudes, further improvements in the representation of permafrost carbon dynamics is a crucial step toward improving the reliability of predictions of the dynamics of future terrestrial carbon cycling.

In summary, we found that DGVMs have generally overestimated soil R_h, and consequently underestimated τ_{soil}, especially in high latitude permafrost regions, which would ultimately translate into an underestimation of 30% in the carbon sink strength of global terrestrial ecosystems for the past century. Our analysis suggests that

terrestrial ecosystems may have accumulated more carbon than predicted by current terrestrial ecosystem models over the twenty-first century, implying that there is a greater possibility of achieving climate mitigation targets through soil carbon sequestration in the future than previously thought.

Conflict of interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study is supported by National Natural Science Foundation of China (grant number: 41988101), National Key R&D Program of China (2019YFA0607304), National Natural Science Foundation of China (Grant number: 42022004 and 41901085) and the Second Tibetan Plateau Scientific Expedition and Research Program (2019QZKK0606).

Data Availability Statement

Soil heterotrophic respiration data from 1982 to 2018 is openly available at https://doi.org/10.5061/dryad.b2rbnzsj9. The TRENDYv9 data are available from Stephen Sitch (s.a.sitch@exeter.ac.uk) or Pierre Friedlingstein (p.friedlingstein@exeter.ac.uk) upon reasonable request. Data source for all the environmental predictors used in the Random Forest can be found in Materials and Methods section.

654

655 **Author contributions**

656 J.Z.D conceived the research; Y.H. performed statistical analyses; Y.H. and J.Z.D
657 wrote the first draft of the paper; all authors contributed to the text and revisions, and
658 approved the final manuscript.

659

660 **References**

661 Allen, A. S. & Schlesinger, W. H. (2004). Nutrient limitations to soil microbial biomass
662 and activity in loblolly pine forests. *Soil Biology and Biochemistry*, 36, 581–589.

663 Anav, A., Friedlingstein, P., Kidston, M., Bopp, L., Ciais, P., Cox, P., Jones, C., Jung,
664 M., Myneni, R., & Zhu, Z. (2013). Evaluating the land and ocean components of
665 the global carbon cycle in the CMIP5 earth system models. *Journal of Climate*,
666 26(18), 6801–6843.

667 Baggs, E. M. (2006). Partitioning the components of soil respiration: A research
668 challenge. *Plant and Soil*, 284, 1–5.

669 Balogh, J., Pintér, K., Fóti, S., Cserhalmi, D., Papp, M., & Nagy, Z. (2011). Dependence
670 of soil respiration on soil moisture, clay content, soil organic matter, and CO_2
671 uptake in dry grasslands. *Soil Biology and Biochemistry*, 43(5), 1006–1013.

672 Bardgett, R. D., & McAlister, E. (1999). The measurement of soil fungal: bacterial
673 biomass ratios as an indicator of ecosystem self-regulation in temperate meadow
674 grasslands. *Biology and Fertility of Soils*, 29(3), 282–290.

675 Batjes, N. H. (2016). Harmonized soil property values for broad-scale modelling

(WISE30sec) with estimates of global soil carbon stocks. *Geoderma*, 269, 61–68.

Bekku, Y., Koizumi, H., Oikawa, T., & Iwaki, H. (1997). Examination of four methods for measuring soil respiration. *Applied Soil Ecology*, 5(3), 247–254.

Bloom, A. A., Exbrayat, J. F., Van Der Velde, I. R., Feng, L., & Williams, M. (2016). The decadal state of the terrestrial carbon cycle: Global retrievals of terrestrial carbon allocation, pools, and residence times. *Proceedings of the National Academy of Sciences of the United States of America*, 113(5), 1285–1290.

Bond-Lamberty, B., & Thomson, A. (2010a). A global database of soil respiration data. *Biogeosciences*, 7(6), 1915–1926.

Bond-Lamberty, B., & Thomson, A. (2010b). Temperature-associated increases in the global soil respiration record. *Nature*, 464(7288), 579–582.

Bond-Lamberty, B, Wang, C., & Gower, S. T. (2004). A global relationship between the heterotrophic and autotrophic components of soil respiration? *Global Change Biology,* 10(10), 1756–1766.

Bond-Lamberty, B. (2018). New Techniques and Data for Understanding the Global Soil Respiration Flux. *Earth's Future*, 6(9), 1176–1180.

Bradford, M. A., Wieder, W. R., Bonan, G. B., Fierer, N., Raymond, P. A., & Crowther, T. W. (2016). Managing uncertainty in soil carbon feedbacks to climate change. *Nature Climate Change*, 6(8), 751–758.

Carvalhais, N., Forkel, M., Khomik, M., Bellarby, J., Jung, M., Migliavacca, M., Mu, M., Saatchi, S., Santoro, M., Thurner, M., Weber, U., Ahrens, B., Beer, C., Cescatti, A., Randerson, J. T., & Reichstein, M. (2014). Global covariation of carbon

698 turnover times with climate in terrestrial ecosystems. *Nature*, 514, 213-217.
699 Chapin, F. S., Woodwell, G. M., Randerson, J. T., Rastetter, E. B., Lovett, G. M.,
700 Baldocchi, D. D., Clark, D. A., Harmon, M. E., Schimel, D. S., Valentini, R., Wirth,
701 C., Aber, J. D., Cole, J. J., Goulden, M. L., Harden, J. W., Heimann, M., Howarth,
702 R. W., Matson, P. A., McGuire, A. D., … Schulze, E. D. (2006). Reconciling
703 carbon-cycle concepts, terminology, and methods. *Ecosystems*, 9(7), 1041–1050.
704 Chen, Y., Xia, J., Sun, Z., Li, J., Luo, Y., Gang, C. & Wang, Z. (2015). The role of
705 residence time in diagnostic models of global carbon storage capacity: model
706 decomposition based on a traceable scheme. *Scientific Reports*, 5(1), 16155.
707 Ciais, P., Yao, Y., Gasser, T., Baccini, A., Wang, Y., Lauerwald, R., Peng, S., Bastos, A.,
708 Li, W., Raymond, P. A., Canadell, J. G., Peters, G. P., Andres, R. J., Chang, J., Yue,
709 C., Dolman, A. J., Haverd, V., Hartmann, J., Laruelle, G., … Zhu, D. (2021).
710 Empirical estimates of regional carbon budgets imply reduced global soil
711 heterotrophic respiration. *National Science Review*, 8(2), nwaa145.
712 Cragg, R. G., & Bardgett, R. D. (2001). How changes in soil faunal diversity and
713 composition within a trophic group influence decomposition processes. *Soil*
714 *Biology and Biochemistry,* 33(15), 2073–2081.
715 Davidson, E. A., & Janssens, I. A. (2006). Temperature sensitivity of soil carbon
716 decomposition and feedbacks to climate change. *Nature*, 440(7081), 165–173.
717 De Boer, W., Folman, L. B., Summerbell, R. C., & Boddy, L. (2005). Living in a fungal
718 world: Impact of fungi on soil bacterial niche development. *FEMS Microbiology*
719 *Reviews*, 29(4), 795–811.

720 Deng, M., Liu, L., Jiang, L., Liu, W., Wang, X., Li, S., Yang, S., & Wang, B. (2018).
721 Ecosystem scale trade-off in nitrogen acquisition pathways. *Nature Ecology &*
722 *Evolution*, 2(11), 1724–1734.
723 Eisenhauer, N., Schlaghamerský, J., Reich, P. B., & Frelich, L. E. (2011). The wave
724 towards a new steady state: effects of earthworm invasion on soil microbial
725 functions. *Biological Invasions*, 13(10), 2191–2196.
726 Exbrayat, J.-F., Pitman, A. J., Zhang, Q., Abramowitz, G., & Wang, Y.-P. (2013).
727 Examining soil carbon uncertainty in a global model: response of microbial
728 decomposition to temperature, moisture and nutrient limitation. *Biogeosciences*,
729 10(11), 7095–7108.
730 Feng, Y., Zhang, J., Berdugo, M., Guirado, E., Guerra, C. A., Egidi, E., Wang, J., Singh,
731 B. K., & Delgado-Baquerizo, M. (2022). Temperature thresholds drive the global
732 distribution of soil fungal decomposers. *Global Change Biology*, 28(8), 2779-2789.
733 Friedl, M. A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., &
734 Huang, X. (2010). MODIS Collection 5 global land cover: Algorithm refinements
735 and characterization of new datasets. *Remote Sensing of Environment*, 114(1),
736 168–182.
737 Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Hauck, J., Olsen, A.,
738 Peters, G. P., Peters, W., Pongratz, J., Sitch, S., Le Quéré, C., Canadell, J. G., Ciais,
739 P., Jackson, R. B., Alin, S., Aragão, L. E. O. C., Arneth, A., Arora, V., Bates, N.
740 R., … Zaehle, S. (2020). Global Carbon Budget 2020. *Earth System Science Data*,
741 12(4), 3269–3340.

742 Friend, A. D., Lucht, W., Rademacher, T. T., Keribin, R., Betts, R., Cadule, P., Ciais, P.,
743 Clark, D. B., Dankers, R., Falloon, P. D., Ito, A., Kahana, R., Kleidon, A., Lomas,
744 M. R., Nishina, K., Ostberg, S., Pavlick, R., Peylin, P., Schaphoff, S., …
745 Woodward, F. I. (2014). Carbon residence time dominates uncertainty in terrestrial
746 vegetation responses to future climate and atmospheric CO_2. *Proceedings of the*
747 *National Academy of Sciences of the United States of America*, 111(9), 3280–3285.
748 Gao, D., Peng, B., Fan, Z., Pei, G., & Bai, E. (2018). Different winter soil respiration
749 between two mid-temperate plantation forests. *Forest Ecology and Management*,
750 409(72), 390–398.
751 Gordon, H., Haygarth, P. M., & Bardgett, R. D. (2008). Drying and rewetting effects
752 on soil microbial community composition and nutrient leaching. *Soil Biology and*
753 *Biochemistry*, 40(2), 302–311.
754 Grandy, A. S., Wieder, W. R., Wickings, K., & Kyker-Snowman, E. (2016). Beyond
755 microbes: Are fauna the next frontier in soil biogeochemical models? Soil Biology
756 and Biochemistry, 102, 40–44.
757 Haberl, H., Erb, K. H., Krausmann, F., Gaube, V., Bondeau, A., Plutzar, C., Gingrich,
758 S., Lucht, W., & Fischer-Kowalski, M. (2007). Quantifying and mapping the
759 human appropriation of net primary production in earth's terrestrial ecosystems.
760 *Proceedings of the National Academy of Sciences of the United States of America*,
761 104(31), 12942–12947.
762 Harris, I., Osborn, T. J., Jones, P., & Lister, D. (2020). Version 4 of the CRU TS monthly
763 high-resolution gridded multivariate climate dataset. *Scientific Data*, 7(1), 1–18.

764 Hashimoto, S., Carvalhais, N., Ito, A., Migliavacca, M., Nishina, K., & Reichstein, M.
765 (2015). Global spatiotemporal distribution of soil respiration modelled using a
766 global database. *Biogeosciences*, 12(13), 4121–4132.
767 Haynes, B. E., & Gower, S. T. (1995). Belowground carbon allocation in unfertilized
768 and fertilized red pine plantations in northern Wisconsin. *Tree Physiology*, 15(5),
769 317–325.
770 He, L., Mazza Rodrigues, J. L., Soudzilovskaia, N. A., Barceló, M., Olsson, P. A., Song,
771 C., Tedersoo, L., Yuan, F., Yuan, F., Lipson, D. A., & Xu, X. (2020). Global
772 biogeography of fungal and bacterial biomass carbon in topsoil. *Soil Biology and*
773 *Biochemistry*, 151, 108024.
774 He, Y., Wang, X., Wang, K., Tang, S., Xu, H., Chen, A., Ciais, P., Li, X., Peñuelas, J.,
775 & Piao, S. (2021). Data-driven estimates of global litter production imply slower
776 vegetation carbon turnover. *Global Change Biology*, 27, 1678–1688.
777 Hengl, T., De Jesus, J. M., Heuvelink, G. B. M., Gonzalez, M. R., Kilibarda, M.,
778 Blagotić, A., Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B.,
779 Guevara, M. A., Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G. B.,
780 Ribeiro, E., Wheeler, I., Mantel, S., & Kempen, B. (2017). SoilGrids250m: Global
781 gridded soil information based on machine learning. PLoS ONE. 12, e0169748.
782 Högberg, M. N., Högberg, P., & Myrold, D. D. (2007). Is microbial community
783 composition in boreal forest soils determined by pH, C-to-N ratio, the trees, or all
784 three? *Oecologia*, 150(4), 590–601.
785 Hugelius, G., Strauss, J., Zubrzycki, S., Harden, J. W., Schuur, E. A. G., Ping, C. L.,

786 Schirrmeister, L., Grosse, G., Michaelson, G. J., Koven, C. D., O'Donnell, J. A.,
787 Elberling, B., Mishra, U., Camill, P., Yu, Z., Palmtag, J., & Kuhry, P. (2014).
788 Estimated stocks of circumpolar permafrost carbon with quantified uncertainty
789 ranges and identified data gaps. *Biogeosciences*, 11(23), 6573–6593.
790 Huntzinger, D. N., Schwalm, C., Michalak, A. M., Schaefer, K., King, A. W., Wei, Y.,
791 Jacobson, A., Liu, S., Cook, R. B., Post, W. M., Berthier, G., Hayes, D., Huang,
792 M., Ito, A., Lei, H., Lu, C., Mao, J., Peng, C. H., Peng, S., … Zhu, Q. (2013). The
793 North American carbon program multi-scale synthesis and terrestrial model
794 intercomparison project - Part 1: Overview and experimental design. *Geoscientific*
795 *Model Development*, 6(6), 2121–2133.
796 Hursh, A., Ballantyne, A., Cooper, L., Maneta, M., Kimball, J., & Watts, J. (2017). The
797 sensitivity of soil respiration to soil temperature, moisture, and carbon supply at
798 the global scale. *Global Change Biology*, 23(5), 2090–2103.
799 Jian, J., Vargas, R., Anderson-Teixeira, K., Stell, E., Herrmann, V., Horn, M., Kholod,
800 N., Manzon, J., Marchesi, R., Paredes, D., & Bond-Lamberty, B. (2021). A
801 restructured and updated global soil respiration database (SRDB-V5). *Earth*
802 *System Science Data*, 13(2), 255–267.
803 Jung, M., Reichstein, M., Schwalm, C. R., Huntingford, C., Sitch, S., Ahlström, A.,
804 Arneth, A., Camps-Valls, G., Ciais, P., Friedlingstein, P., Gans, F., Ichii, K., Jain,
805 A. K., Kato, E., Papale, D., Poulter, B., Raduly, B., Rödenbeck, C., Tramontana,
806 G., … Zeng, N. (2017). Compensatory water effects link yearly global land CO_2
807 sink changes to temperature. *Nature*, 541(7638), 516–520.

Konings, A. G., Anthony Bloom, A., Liu, J., Parazoo, N. C., Schimel, D. S., & Bowman, K. W. (2019). Global satellite-driven estimates of heterotrophic respiration. *Biogeosciences*, 16(11), 2269–2284.

Koven, C. D., Chambers, J. Q., Georgiou, K., Knox, R., Negron-Juarez, R., Riley, W. J., Arora, V. K., Brovkin, V., Friedlingstein, P., & Jones, C. D. (2015). Controls on terrestrial carbon feedbacks by productivity versus turnover in the CMIP5 Earth System Models. *Biogeosciences*, 12(17), 5211–5228.

Lauber, C. L., Strickland, M. S., Bradford, M. A., & Fierer, N. (2008). The influence of soil properties on the structure of bacterial and fungal communities across land-use types. *Soil Biology and Biochemistry*, 40(9), 2407–2415.

Martens, B., Miralles, D., Lievens, H., van der Schalie, R., de Jeu, R., Férnandez-Prieto, D., Beck, H., Dorigo, W., & Verhoest, N. (2016). GLEAM v3: satellite-based land evaporation and root-zone soil moisture. *Geoscientific Model Development Discussions*, 8, 1903–1925.

Meidute, S., Demoling, F., & Bååth, E. (2008). Antagonistic and synergistic effects of fungal and bacterial growth in soil after adding different carbon and nitrogen sources. *Soil Biology and Biochemistry*, 40(9), 2334–2343.

Mishra, U., Hugelius, G., Shelef, E., Yang, Y., Strauss, J., Lupachev, A., Harden, J. W., Jastrow, J. D., Ping, C. L., Riley, W. J., Schuur, E. A. G., Matamala, R., Siewert, M., Nave, L. E., Koven, C. D., Fuchs, M., Palmtag, J., Kuhry, P., Treat, C. C., … Orr, A. (2021). Spatial heterogeneity and environmental predictors of permafrost region soil organic carbon stocks. *Science Advances*, 7(9), eaaz5236.

Moinet, G. Y. K., Cieraad, E., Hunt, J. E., Fraser, A., Turnbull, M. H., & Whitehead, D. (2016). Soil heterotrophic respiration is insensitive to changes in soil water content but related to microbial access to organic matter. *Geoderma*, 274, 68–78.

Nachtergaele, F. A., van Velthuizen, H. B., Batjes, N. C., Dijkshoorn, K. C., van Engelen, V. C., Fischer, G. B., Jones, A. D., Montanarella, L. D., Petri, M. A., Prieler, S. B., Shi, X. E., Teixeira, E. D., & Wiberg A, D. D. (2010). The harmonized world soil database Food and Agriculture Organization of the United Nations. *Science, Soil Solutions for a Changing World*, 8, 34–37.

Ohashi, M., Maekawa, Y., Hashimoto, Y., Takematsu, Y., Hasin, S., & Yamane, S. (2017). CO_2 emission from subterranean nests of ants and termites in a tropical rain forest in Sarawak, Malaysia. *Applied Soil Ecology*, 117, 147–155.

Pumpanen, J., Kolari, P., Ilvesniemi, H., Minkkinen, K., Vesala, T., Niinistö, S., Lohila, A., Larmola, T., Morero, M., & Pihlatie, M. (2004). Comparison of different chamber techniques for measuring soil CO_2 efflux. *Agricultural and Forest Meteorology*, 123, 159–176.

Rousk, J., Brookes, P. C., & Bååth, E. (2010). The microbial PLFA composition as affected by pH in an arable soil. Soil Biology and Biochemistry, 42(3), 516–520.

Sanderman, J., Hengl, T., & Fiske, G. J. (2017). Soil carbon debt of 12,000 years of human land use. *Proceedings of the National Academy of Sciences*, 114(36), 9575–9580.

Schmidt, M. W. I., Torn, M. S., Abiven, S., Dittmar, T., Guggenberger, G., Janssens, I. A., Kleber, M., Kögel-Knabner, I., Lehmann, J., Manning, D. A. C., Nannipieri, P.,

852 Rasse, D. P., Weiner, S., & Trumbore, S. E. (2011). Persistence of soil organic
853 matter as an ecosystem property. *Nature*, 478(7367), 49–56.
854 Schneider, T., Keiblinger, K. M., Schmid, E., Sterflinger-Gleixner, K., Ellersdorfer, G.,
855 Roschitzki, B., Richter, A., Eberl, L., Zechmeister-Boltenstern, S., & Riedel, K.
856 (2012). Who is who in litter decomposition? Metaproteomics reveals major
857 microbial players and their biogeochemical functions. *The ISME Journal*, 6(9),
858 1749–1762.
859 Schuur, E. A. G., McGuire, A. D., Schädel, C., Grosse, G., Harden, J. W., Hayes, D. J.,
860 Hugelius, G., Koven, C. D., Kuhry, P., Lawrence, D. M., Natali, S. M., Olefeldt,
861 D., Romanovsky, V. E., Schaefer, K., Turetsky, M. R., Treat, C. C., & Vonk, J. E.
862 (2015). Climate change and the permafrost carbon feedback. *Nature*, 520, 171–
863 179.
864 Schwartz, S. E. (1979). Residence times in reservoirs under non-steady-state conditions:
865 Application to atmospheric SO_2 and aerosol sulfate1. *Tellus*, 31(6), 530–547.
866 Shao, P., Zeng, X., Sakaguchi, K., Monson, R. K., & Zeng, X. (2013). Terrestrial carbon
867 cycle: Climate relations in eight CMIP5 earth system models. *Journal of Climate*,
868 26(22), 8744–8764.
869 Sitch, S., Friedlingstein, P., Gruber, N., Jones, S. D., Murray-Tortarolo, G., Ahlström,
870 A., Doney, S. C., Graven, H., Heinze, C., Huntingford, C., Levis, S., Levy, P. E.,
871 Lomas, M., Poulter, B., Viovy, N., Zaehle, S., Zeng, N., Arneth, A., Bonan, G., …
872 Myneni, R. (2015). Recent trends and drivers of regional sources and sinks of
873 carbon dioxide. *Biogeosciences*, 12(3), 653–679.

874 Stockmann, U., Adams, M. A., Crawford, J. W., Field, D. J., Henakaarchchi, N., Jenkins,
875 M., Minasny, B., McBratney, A. B., de Remy de Courcelles, V., Singh, K., Wheeler,
876 I., Abbott, L., Angers, D. A., Baldock, J., Bird, M., Brookes, P. C., Chenu, C.,
877 Jastrow, J. D., Lal, R., … Zimmermann, M. (2013). The knowns, known unknowns
878 and unknowns of sequestration of soil organic carbon. *Ecosystems and*
879 *Environment*, 164, 80–99.
880 Strickland, M. S., & Rousk, J. (2010). Considering fungal: Bacterial dominance in soils
881 - Methods, controls, and ecosystem implications. *Soil Biology and Biochemistry*,
882 42(9), 1385–1395.
883 Subke, J. A., Inglima, I., & Cotrufo, M. F. (2006). Trends and methodological impacts
884 in soil CO_2 efflux partitioning: A meta analytical review. *Global Change Biology*,
885 12(6), 921–943.
886 Tang, J., & Riley, W. J. (2015). Weaker soil carbon–climate feedbacks resulting from
887 microbial and abiotic interactions. *Nature Climate Change*, 5(1), 56–60.
888 Tang, X., Du, J., Shi, Y., Lei, N., Chen, G., Cao, L., & Pei, X. (2020b). Global patterns
889 of soil heterotrophic respiration – A meta-analysis of available dataset. *Catena,*
890 191(4), 104574.
891 Tang, X., Fan, S., Du, M., Zhang, W., Gao, S., Liu, S., Chen, G., Yu, Z., & Yang, W.
892 (2020a). Spatial and temporal patterns of global soil heterotrophic respiration in
893 terrestrial ecosystems. *Earth System Science Data*, 12(2), 1037–1051.
894 Tian, H., Lu, C., Yang, J., Banger, K., Huntzinger, D. N., Schwalm, C. R., Michalak, A.
895 M., Cook, R., Ciais, P., Hayes, D., Huang, M., Ito, A., Jain, A. K., Lei, H., Mao, J.,

Pan, S., Post, W. M., Peng, S., Poulter, B., … Zeng, N. (2015). Global patterns and controls of soil organic carbon dynamics as simulated by multiple terrestrial biosphere models: Current status and future directions. *Global Biogeochemical Cycles*, 29(6), 775–792.

Tian, D. & Niu, S. (2015) A global analysis of soil acidification caused by nitrogen addition. E*nvironmental Research Letters*, 024019.

Todd-Brown, K. E. O., Randerson, J. T., Hopkins, F., Arora, V., Hajima, T., Jones, C., Shevliakova, E., Tjiputra, J., Volodin, E., & Wu, T. (2014). Changes in soil organic carbon storage predicted by Earth system models during the 21st century. *Biogeosciences*, 11(8), 2341–2356.

Todd-Brown, K. E. O., Randerson, J. T., Post, W. M., Hoffman, F. M., Tarnocai, C., Schuur, E. A. G., & Allison, S. D. (2013). Causes of variation in soil carbon simulations from CMIP5 Earth system models and comparison with observations. *Biogeosciences*, 10(3), 1717–1736.

van den Hoogen, J., Geisen, S., Routh, D., Ferris, H., Traunspurger, W., Wardle, D. A., de Goede, R. G. M., Adams, B. J., Ahmad, W., Andriuzzi, W. S., Bardgett, R. D., Bonkowski, M., Campos-Herrera, R., Cares, J. E., Caruso, T., de Brito Caixeta, L., Chen, X., Costa, S. R., Creamer, R., … Crowther, T. W. (2019). Soil nematode abundance and functional group composition at a global scale. *Nature*, 572, 194–198.

Wang, G., Post, W. M., & Mayes, M. A. (2013). Development of microbial-enzyme-mediated decomposition model parameters through steady-state and dynamic

analyses. *Ecological Applications,* 23(1), 255–272.

Wang, W., Chen, W., & Wang, S. (2010). Forest soil respiration and its heterotrophic and autotrophic components: Global patterns and responses to temperature and precipitation. *Soil Biology & Biochemistry.* 42, 1236–1244.

Warner, D. L., Bond-Lamberty, B., Jian, J., Stell, E., & Vargas, R. (2019). Spatial Predictions and Associated Uncertainty of Annual Soil Respiration at the Global Scale. *Global Biogeochemical Cycles*, 33(12), 1733–1745.

Wieder, W. (2014). Soil carbon: Microbes, roots and global carbon. *Nature Climate Change*, 4(12), 1052–1053.

Wieder, W. R., Bonan, G. B., & Allison, S. D. (2013). Global soil carbon projections are improved by modelling microbial processes. *Nature Climate Change*, 3(10), 909–912.

Wieder, W. R., Grandy, A. S., Kallenbach, C. M., Taylor, P. G., & Bonan, G. B. (2015). Representing life in the Earth system with soil microbial functional traits in the MIMICS model. *Geoscientific Model Development*, 8(6), 1789–1808.

Wu, D., Piao, S., Liu, Y., Ciais, P., & Yao, Y. (2018). Evaluation of CMIP5 earth system models for the spatial patterns of biomass and soil carbon turnover times and Their linkage with climate. *Journal of Climate*, 31(15), 5947–5960.

Xu, M., & Shang, H. (2016). Contribution of soil respiration to the global carbon equation. *Journal of Plant Physiology*, 203, 16–28.

Yan, Z., Bond-Lamberty, B., Todd-Brown, K. E., Bailey, V. L., Li, S., Liu, C., & Liu, C. (2018). A moisture function of soil heterotrophic respiration that incorporates

microscale processes. *Nature Communications*, 9(1), 2562.

Yao, Y., Wang, X., Li, Y., Wang, T., Shen, M., Du, M., He, H., Li, Y., Luo, W., Ma, M., Ma, Y., Tang, Y., Wang, H., Zhang, X., Zhang, Y., Zhao, L., Zhou, G., & Piao, S. (2018). Spatiotemporal pattern of gross primary productivity and its covariation with climate in China over the last thirty years. *Global Change Biology*, 24, 184–196.

Ye, J., Bradford, M. A., Dacal, M., Maestre, F. T., & García-Palacios, P. (2019). Increasing microbial carbon use efficiency with warming predicts soil heterotrophic respiration globally. *Global Change Biology*, 25(10), 3354–3364.

Zeng, Z., Piao, S., Lin, X., & Yin, G. (2012). Global evapotranspiration over the past three decades: estimation based on the water balance equation combined with empirical models. *Environmental Research Letters*, 7, 014026.

Zhang, F.G., & Zhang, Q.G. (2016). Microbial diversity limits soil heterotrophic respiration and mitigates the respiration response to moisture increase. *Soil Biology and Biochemistry*, 98, 180–185.

Zhu, Z., Bi, J., Pan, Y., Ganguly, S., Anav, A., Xu, L., Samanta, A., Piao, S., Nemani, R.R. & Myneni, R. B. (2013). Global data sets of vegetation leaf area index (LAI) 3g and fraction of photosynthetically active radiation (FPAR) 3g derived from global inventory modeling and mapping studies (GIMMS) normalized difference vegetation index (NDVI3g) for the period 1981 to 2011. *Remote Sensing*, 5, 927–948.