Spatial biases reduce the ability of earth system models to simulate soil heterotrophic respiration fluxes

Bertrand Guenet[1,*], Jérémie Orliac[1], Lauric Cécillon[1,2], Olivier Torres[3], Laura Sereni[4], Philip A. Martin[5], Pierre Barré[1], Laurent Bopp[3]

[1] Laboratoire de Géologie, Ecole normale supérieure, CNRS, IPSL, Université PSL, 24 Rue Lhomond, 75005 Paris, France

[2] Laboratoire ECODIV, Univ. Normandie, UNIROUEN, INRAE, Rouen, France.

[3] LMD-IPSL, Ecole Normale Supérieure, Université PSL, CNRS, Ecole Polytechnique, Sorbonne Université, Paris, France.

[4] INRAE, AgroParisTech, Université Paris-Saclay, UMR1402 ECOSYS, Ecotoxicology Team, 78026, Versailles, France.

[5] Basque Centre for Climate Change (BC3), Edificio sede no 1, planta 1, Parque científico UPV/EHU, Barrio Sarriena s/n, 48940, Leioa, Bizkaia, Spain.

Correspondence to: Bertrand Guenet (guenet@geologie.ens.fr)

Abstract. Heterotrophic respiration (Rh) is, at a global scale, one of the largest CO_2 fluxes between the earth's surface and atmosphere and may increase in the future. The previous generation of Earth System Models (ESMs) were able to reproduce global fluxes relatively well, but at that time no gridded products were available to perform an in-depth evaluation. The capacity of the new generation of ESMs used within the Coupled Model Intercomparison Project Phase 6 (CMIP6) to reproduce this flux has not been evaluated, meaning that the realism of resulting CO_2 flux estimates is unclear. In this study, we combine recently released observational data on Rh and ESM simulations to evaluate the ability of 13 ESMs from CMIP6 to reproduce Rh. Only four of the 13 tested ESMs were able to reproduce the total Rh flux but spatial analysis underlined important bias compensation for most of the ESMs which generally showed an overestimation in tropical regions and an underestimation in arid regions. To identify the main drivers of the bias, we performed an analysis of the residuals and found that mean annual precipitation was the most important driver explaining the difference between ESM simulations and observation-derived

product of Rh with higher bias between ESM simulations and Rh products where precipitation was high. Based on our results, next-generation ESMs should focus on improving the response of Rh to soil moisture.

1 Introduction

Soil organic carbon stocks represent around three times the amount of carbon in the atmosphere (Scharlemann et al., 2014). This soil carbon is used as a substrate by soil microorganisms to obtain their energy and feed their metabolism, which account for the majority of heterotrophic soil organism biomass. Annual fluxes that result from the respiration of these heterotrophic organisms (hereafter referred to as heterotrophic respiration) are estimated (Warner et al., 2019; Hashimoto et al., 2015; Konings et al., 2019; Ciais et al., 2021) to be five times higher than annual anthropogenic emissions (Friedlingstein et al., 2022) and roughly similar to annual terrestrial net primary production (Zhao et al., 2005). Thus, due to the size of fluxes relating to heterotrophic respiration, even minor changes in soil organic carbon dynamics can lead to significant impacts on carbon feedbacks and, ultimately, on climatic changes. As a result, modification of soil organic carbon stocks due to human activities is considered to be an important driver of future climate trajectories (Chabbi et al., 2017).

Although heterotrophic respiration fluxes are important, the way this flux is represented in Earth System Models (ESMs), which aim to simulate the most important drivers of the earth's climate system, is currently challenged. This is because important drivers are missing and new approaches have been proposed (Huang et al., 2021; Wieder et al., 2015) they lack sufficient evaluation on long term time series (Le Noë et al., 2023). Thus, the identifying the accuracy of predictions of heterotrophic respiration fluxes by ESMs is a key topic in order to help constrain carbon-climate feedbacks in ESMs. The ability of ESMs to reproduce this flux was previously studied by Shao et al., (2013) but this was done using the previous generation of ESMs using the simulations done during the Coupled Model Intercomparison Project Phase 5 (CMIP5). Moreover, at that time no gridded products were available. Since then, the models have been greatly improved and assessing how accurately current ESMs reproduce the fluxes associated with heterotrophic respiration is therefore of major importance. Until now, it was not possible to undertake a robust spatial assessment because of the lack of observation-derived gridded products of Rh. In recent years, new gridded products derived either from (i) upscaling of local observation or (ii) calculations using atmospheric inversions and satellite observations. These products provide the opportunity to evaluate the simulations of ESMs used within the Coupled Model Intercomparison Project Phase 6 (CMIP6) against observation-derived products for heterotrophic respiration. CMIP is a key initiative which aims to compare current ESMs and is a central element of national and international assessments of climate change (Masson-Delmotte et al., 2021).

In this study we have two major aims:

1. Compare predictions of the total flux of heterotrophic respiration from 13 earth system models with three recent gridded products of heterotrophic respiration derived from observations and identify the spatial biases of heterotrophic respiration in the models.

2

2. Identify the major drivers of the heterotrophic respiration bias in earth system models to propose new developments for the next generation of earth system models using a model residue approach to disentangle the main effect.

2 Materials and Methods

2.1 Earth System Models simulations.

In this study, we used the model outputs from the 6th Coupled-Model Intercomparison Project (CMIP6) (Eyring et al., 2016) which coordinates global climate model simulations of the past, current, and future climate. CMIP6 proposes historical simulation spanning from 1850 to 2014. Historical simulations are driven from an initial point chosen in control integration (*piControl*). We chose to use the latest CMIP6 results for basic initial state (r1i1p1f1). We chose outputs from thirteen ESMs that provide heterotrophic respiration fluxes (BCC-CSM2-MR, BCC-ESM1, CanESM5, CESM2, CNRM-ESM2-1, E3SM-1-1-ECA, IPSL-CM6A-LR, MIROC-ES2L, MPI-ESM1-2-LR, NorCPM1, NorESM2-LM, SAM0-UNICON and UKESM1-0-LL). The variable used is "rh" corresponding to the total heterotrophic respiration on land. We computed annual average over the 1990-2010 period which corresponds to the period in which most of the observations in the global Soil Respiration Database (Bond-Lamberty and Thomson, 2010) v3.0 were made. Two observation products we used were obtained using those data.

2.2 Observation-derived products.

In this study we used three observation derived products (Warner et al., 2019; Konings et al., 2019; Hashimoto et al., 2015). In Warner et al. (2019), the authors predicted annual soil respiration and associated uncertainty across terrestrial areas at a resolution of 1 km using a quantile regression forest algorithm trained with observations from the global Soil Respiration Database (Bond-Lamberty and Thomson, 2010) v3.0 (commit number 651770 in GitHub, https://github.com/bpbond/srdb) spanning from 1961 to 2011 but mostly after 1990. Then they deduced Rh from the soil respiration using two different methods (Bond-Lamberty et al., 2004; Subke et al., 2006). They therefore proposed two Rh maps derived from a unique mean map of Rs from quantile regression forest model. Here, we decided to use the mean of two approaches as a reference for Warner et al. (2019) Rh results. The second product we used Hashimoto et al. (2015) is also based on the Soil Respiration Database (Bond-Lamberty and Thomson, 2010) v3.0 but in this case they derived the Rh flux using a climate-driven model of soil respiration derived from the Raich's model (Raich et al., 2002). They provided a 0.5° resolution product at a monthly step time between 1965 and 2012. In our case, we used the yearly average over the period. The third product used, Konings et al. (2019), estimated Rh as a residual remote-sensing data exploiting recent advances in carbon-flux estimations. In contrast with the two other products which are bottom-up, the Konings et al., (2019) product proposes a top-down approach combining net ecosystem productivity estimates from atmospheric inversions with an optimally scaled gross primary productivity dataset derived from satellite observations. Rh is then derived using the CARbon DAta MOdel fraMework, (CARDAMOM). Their result is a monthly evaluation of Rh, between January 2010 and December 2012, at a resolution 4°×5°.

3

2.3 Data treatment and regridding.

Since the ESMs outputs and products were not at the same resolution, we chose a reference for map-grid resolution. The coarsest resolution was from Konings et al.'s (2019) product with a 4°×5° resolution grid. Reduced the resolution of every Rh map to match this would cause a substantial loss of information. Thus, we increased the resolution of those datasets and decreased the very fine scale maps to an arbitrary reference corresponding to the CNRM-ESM2-1 model which runs at 0.7° resolution. We chose to set the reference at the maximum resolution available among CMIP6's ESMs predicting Rh. We used the common regridding routine Climate Data Operators (CDO) remapdis (nco module) that performs regridding by distance-weighted average remapping and conserve latitudinal and longitudinal means. The CDO software is a collection of multiple operators for standard processing of climate and forecast model data. The operators include simple functions (statistical and arithmetic) to be used for data selection, subsampling, and spatial interpolation. To avoid coastal pixels encroaching into oceans, we weighted each pixel by the proportion of its are covered by land. The sum of Rh over the lands was compared before and after regridding to ensure that it was conservative. When comparing the original and the regridded version of the Konings et al. (2019)'s product we observed very similar pattern (Fig. 1).

2.4 Comparison between models' outputs and heterotrophic respiration products.

To estimate the ability of the CMIP6's model to reproduce soil heterotrophic respiration, we first compared the global flux summed over all the grid cells and averaged over 1990-2010 period in Pg C yr^{-1}. We also compared the Rh maps after regridding, averaged over the 1990-2010 period. We also performed latitudinal and longitudinal means calculus including oceanic zero-values. Secondly, we wanted to assess spatial bias distribution. Therefore, we i) compared CMIP6 model average with observation products and ii) compared each CMIP6 models with observation products. Thus, we first represented the model average (over the period 1990-2010) and all the observation derived products on a same figure with their associated latitudinal and longitudinal means. We also calculated the 25th and 75th quantiles of latitudinal and longitudinal CMIP6 model means. Then, we computed the difference for each individual CMIP6 model with the median of the three observation products. To compare the ESMs with the observation products, we calculated the root mean square error (RMSE) and the R^2 using the median of the observation products. Finally, we also calculated the median absolute deviation (MAD) for each grid cells and we calculated the number of pixels for each model that fit within the median ± MAD.

2.5 ESM's model residual analysis.

We defined the ESM's model residuals as the median of the difference between each individual CMIP6's model output and the observation-based products median calculated for each grid cell. The ESM's model residuals were calculated in three steps:

1. We first calculated the median for each cell (i) that we called *Rh_obs$_i$* using the three observation-derived products using eq. 1 with *Rh_Hashimoto et al. (2015)$_i$*, *Rh_Warner et al. (2019)$_i$* and *Rh_Konings et al. (2019)$_i$* being the heterotrophic respiration given for the grid cell I by Hashimoto et al., (2015), Warner et al. (2019) and Konings et al. (2019, respectively). We consider this median as our best-estimate.

4

$$Rh\_obs_i = Med(Rh\_Hashimoto\ et\ al.\ (2015)_i, Rh\_Warner\ et\ al.\ (2019)_i, Rh\_Konings\ et\ al.\ (2019)_i) \quad (1)$$

2. Next, we calculated the residual between each CMIP6's model output and our best-estimate that we called Res_X_i (X being the model's name) for each grid cell i using eq. 2.

$$Res\_X_i = Rh\_X_i - Rh\_obs_i \qquad (2)$$

3. Finally, we calculated the ESM's model residuals (Res_i) as the median of this model specific residuals (Res_X_i) using eq. 3.

$$Rh\_X_i - Rh\_obs_i \qquad (3)$$

Using the ESM's model residuals, we performed a statistical analysis to identify the main drivers of disagreement between predictions and observations. Following this we undertook a two-step methodology. First, we compared several linear generalized least square models with different spatial structures (gaussian, exponential, spherical, linear or rational (gls package, (Venables and Ripley, 2002))) and without spatial structures to estimate the effect of spatial correlation. Based on Akaike information criterion (AIC) values we selected the rational quadratic spatial correlation structure that had the smallest AIC values for the second step of the analysis. Then, we used generalized additive mixed model with ESM's model residuals as variable to explain and mean annual temperature (MAT), mean annual precipitation (MAP), observation derived SOC, ESM's model residuals on NPP and lithology as predictors variables. MAT and MAP are derived from the Global Soil Wetness Project Phase 3 (GSWP3) reanalysis (http://hydro.iis.u-tokyo.ac.jp/GSWP3/ last access: April 5 2022). SOC was taken from the Soilgrid250m product (Hengl et al., 2017). ESM's model residuals on NPP are calculated as the median of the difference between ESM's NPP and NPP from the global inventory monitoring and modelling studies group (GIMMS). Lithology maps from the global lithological map (GLiM) (Hartmann and Moosdorf, 2012) was used but since lithology was not significant (p>0.05) and the model has a lower AIC without this variable, it was not included in the final generalized additive mixed model presented here. All statistical analyses were carried out using R v3.5 (R Core Team, 2018).

3 Results

3.1 Global heterotrophic respiration flux and spatial biases

Global heterotrophic respiration flux simulated by the 13 ESMs ranges from 29 to 78 Pg C yr^{-1} (Fig. 2), whereas the equivalent estimates for observationally derived products estimate range from 43 to 51 Pg C yr^{-1}. The multi-model mean of the ESMs

5

(49 Pg C yr^{-1}) falls within the range of the observation-derived products. However, only four out of 13 ESMs (BCC-CSM2-MR, CNRM-ESM2-1, IPSL-CM6A-LR, and SAM0-UNICOM) simulate an overall heterotrophic respiration flux that is within the range of the observation-derived products (Fig 2). When comparing the model observation products' with the median ± MAD from the observation products (46 ± 7 Pg C yr^{-1}), seven out of the 13 ESMs predicted an heterotrophic respiration within this range (Fig. 2). The R^2 between the model outputs and the median of the observation products range between 0.57 for E3SM-1-1-ECA and 0.82 for MIROC-ES2L (Table 1). When using RMSE to compare the model outputs and the median of the observation products was 170.9 gC m^{-2} yr^{-1} for IPSL-CM6A-LR and 345.1 gC m^{-2} yr^{-1} for CanESM5 (Table 1). Finally, we also estimated the number of pixels that fell within the median ± MAD and using this metric BCC-ESM1 was the best performing model followed by BCC-CSM2-MR and CNRM-ESM2-1.

Despite similar global-scale values, regional-scale differences between the observation-derived products are much larger (Fig. 3). The Konings et al. (2019) product estimates large heterotrophic fluxes in the tropics and lower fluxes in other regions such as the west coast of Northern America or central Asia, as compared to the Warner et al. (2019) and the Hashimoto et al. (2015) products that share similar spatial patterns (Fig. 4). The mean of the 13 ESMs simulations also gives a much larger heterotrophic respiration fluxes over the tropics in particular over South-East Asia compared to any of the three observation-derived products. In general, the heterotrophic respiration fluxes from the 13 ESMs mean is closer to Konings et al. (2019) product over the tropics but closer to the Warner et al. (2019) and the Hashimoto et al., (2015) products over temperate regions. For boreal regions, the three observations-derived products and the 13 ESMs means are very close.

To generate our best-estimate of heterotrophic respiration fluxes from the three observation-derived products we calculated the median for each cell. Thus, we obtained the spatially distributed best-estimate. At each grid cell, we then compared each ESM with the observation-derived products median (Fig. 5). This evaluation indicates that, compared to observation-based products, ESMs (apart from the ESM NorCPM1) tend to overestimate heterotrophic respiration flux in tropical regions (approx. 1,000 gC m^{-2} yr^{-1} for MPI-ESM1-2-LR over the Amazon or 1,500 gC m^{-2} yr^{-1} for UKESM1-0-LL over South-East Asia, for instance). Models perform relatively well in temperate regions with bias close to 0 gC m-2 yr-1 for BCC-ESM-1 over North America and Europe. Important discrepancies were observed for boreal regions with some models underestimating the heterotrophic respiration fluxes to a large degree (e.g. NorCPM1 or SAM0-UNICON) and one overestimating the fluxes (MPI-ESM1-2-LR). The BCC models (BCC-CSM2-MR and BCC-ESM1) performed quite well over this region. Importantly, the four models that predict a global heterotrophic respiration flux within the range given by the observation-derived products (BCC-CSM2-MR, CNRM-ESM2-1, IPSL-CM6A-LR and SAM0-UNICOM), do not perform well at finer scales - with over-estimation of the flux in some regions and under estimation in others. Therefore, this good global-scale performance masks spatial bias compensation.

3.2 Identification of the major drivers of the heterotrophic respiration bias in earth system models.

In order to improve predictions of heterotrophic respiration fluxes in future ESMs we need to understand the spatial biases we observed and determine their causes. To explore these biases, we performed a statistical analysis based on a generalized additive mixed model of the ESMs residuals defined as the median of the difference between each CMIP6 model's output and the median of the observation-based products calculated in each grid cell. ESMs share a very common approach based on first order kinetics with soil organic decomposition driven by soil moisture and temperature (Varney et al., 2022; Todd-Brown et al., 2014). This approach is derived from the very first attempts to describe soil organic decomposition with mathematical equations (Henin and Dupuis, 1945) and is still the most used to describe this process (Manzoni and Porporato, 2009; Wutzler et al., 2008). Since SOM decomposition schemes in ESMs are very similar, comparing each model individually can be redundant and not very informative and less generalizable. To allow broader conclusions and suggestions to improve ESMs performances, we decided to perform the residual analysis on the ESMs median rather on each individual model.

The main drivers of heterotrophic respiration are soil carbon availability, soil moisture and temperature, carbon inputs and mineralogy (Doetterl et al., 2015). To explain our model residues we used soil organic carbon, net primary production residuals calculated using similar methods to heterotrophic respiration flux residuals, mean annual precipitation, mean annual temperature and lithology. Our method identified the main drivers of ESMs residuals as soil organic carbon, net primary production residuals, mean annual precipitation, and mean annual temperature (Fig. 6). The effect of lithology was not statistically significant (p>0.05) and the model has a lower AIC without this variable and so we did not include lithology in the final model presented here. The residuals due to soil organic carbon stock are close to zero for soil with a low carbon stock but heterotrophic respiration is under estimated by ESMs for soils rich in organic carbon (> 3,000 g C m^{-2}) (Fig. 6a). The model residuals for heterotrophic respiration flux are partially explained by the model residuals on net primary production with a slight increase from model underestimation to model overestimation when model residual on net primary production increase from -1,000 to 400 g C m^{-2} yr^{-1}. We noted that when net primary production fits well with satellite products (i.e. model residuals close to 0 g C m^{-2} yr^{-1}), the ESM residuals on the heterotrophic respiration flux are also close to 0 g C m^{-2} yr^{-1}. For a few grid cells where ESMs largely overestimate net primary production (i.e. model residuals higher than 400 g C m^{-2} yr^{-1}), the ESMs residuals on heterotrophic respiration flux tend to be negative suggesting that ESM underestimate heterotrophic respiration flux. The clearest tendency we obtained was with mean annual precipitation, the more it increases the more the models overestimate the heterotrophic respiration flux (Fig. 6c). The median ESMs residual was also partially controlled by mean annual temperature (Fig. 6d) with a relatively low overestimation by the models for cold temperatures such as those recorded in polar climate zones and in some continental climate zones (e.g. subarctic climate), a relatively good fit for temperature between 0 and 20°C corresponding to temperate and some continental climate zones (e.g. Hot summer continental climates) and then a sudden underestimation for warm temperatures above 20°C corresponding to tropical and dry climate zones. This sudden underestimation might be explained by an arbitrary maximum respiration level observed in this dataset and identified as the result of the temperature-dependence of soil respiration used by Hashimoto et al., (2015) (Varney et al., 2020).

315 Such bias may therefore be a consequence of the observation-based products used here rather than a real bias in ESMs. Similar results were obtained when performing the same analysis with means rather than medians (Fig. 7).

4 Discussion

In this study we evaluated, for the first time, the ability of the ESMs to reproduce heterotrophic respiration fluxes. Indeed, previous dataset were not gridded and so far spatial patterns of heterotrophic respiration in ESMs could only be 320 constrained indirectly by constraining other C fluxes including heterotrophic respiration such as net ecosystem exchange fluxes or through ecosystem respiration in which heterotrophic respiration is just one component the other being the autotrophic respiration (Stoy et al., 2013). We showed that only four of 13 of the CMIP models produce global-scale estimates that are consistent with observation-derived products. However, we also showed that this consistency was due to spatial bias compensations driven by different environmental variables. Heterotrophic respiration represents a carbon flux that is roughly 325 five times that of anthropogenic emissions (Friedlingstein et al., 2022) and, as such, it is vital that work is done to improve the ability of ESMs to reproduce this flux. Nevertheless, we also observed large discrepancies between observation-based products showing that our ability to provide heterotrophic flux based on observations is not optimal. To better constrain ESMs projections, some efforts are needed to reduce residuals between observation-based products.

However, working only on heterotrophic respiration may not be sufficient to improve the entire soil organic carbon module of 330 the ESMs (Table 2). ESM capacities to reproduce observed soil organic carbon stocks also need to be improved (Ito et al., 2020; Varney et al., 2022). To improve both soil organic carbon stocks and heterotrophic respiration fluxes soil organic carbon decomposition rates must be better constrained. The ESM residual analysis we performed here suggests some new research avenues relating to the response of the major drivers. First, most of the boundary conditions of the soil organic carbon modules of an ESM are calculated by the ESM itself. Thus, if soil moisture, soil temperature or litter production are incorrect, the soil 335 organic carbon dynamic cannot be correct. We observed that when the residual of NPP was close to zero the residual on heterotrophic respiration is also close to zero. Thus, improving the plant functioning scheme may ultimately improve the capacities of the ESMs to reproduce the heterotrophic respiration flux. Our study also showed that mean annual temperature is an important driver of the ESM residuals in particular for hot regions with large underestimations of the flux. It probably corresponds to very arid regions since for most of the ESMs, heterotrophic respiration fluxes from regions like Australia, the 340 Middle East or Northern Africa tend to be underestimated. Nevertheless, the underestimation observed in these regions may be also due to reduced C inputs and low SOC stocks reducing the heterotrophic respiration fluxes.

The response of soil organic decomposition by microorganisms is likely to be temperature dependent, with lower rates of decomposition seen in cold regions and higher rates in hot regions (Wang et al., 2010; Zhou et al., 2009). In contrast, the response of soil organic decomposition to temperature in ESMs is generally controlled by Q10 equations (Davidson and 345 Janssens, 2006) with fixed parameters not dynamic and not spatially distributed (Ito et al., 2020). Previous studies suggested

8

that a spatially distributed Q10 constrained by observations are an important step to improve ESMs (Koven et al., 2017; Varney et al., 2020). Our results support this and hint that having more flexible Q10 parameters may help to improve ESMs capacities to reproduce observation-derived products of heterotrophic respiration fluxes. Moreover, land surface scheme of ESMs are known to be very sensitive to Q10 values (Jones et al., 2003; Todd-Brown et al., 2018).

365 Finally, we observed a relatively linear, positive relationship between mean annual precipitation and the ESMs' residuals (Fig. 6c). This response is probably driven by soil moisture because it is a key driver of microbial activity and therefore of heterotrophic respiration fluxes (Moyano et al., 2012). ESMs use three main groups of soil moisture response function (Falloon et al., 2011): i) some models do not represent soil moisture effect, ii) some models increase soil organic decomposition when soil moisture increases assuming less water limitation for microbial activity and iii) some models assume a humped relationship

370 between soil moisture and soil organic decomposition, with high decomposition at intermediate soil moisture and low decomposition in very wet soils where microbial activity is reduced because of limitation by oxygen availability and in dry soils where microbial activity is reduced because of limitation by water. As with Q10, the land surface models are highly sensitive to which soil moisture response function chosen and most of the ESMs use option ii) (Varney et al., 2022). Soil incubations have repeatedly shown that the response of heterotrophic respiration fluxes to soil moisture is approximated by a

375 bell-shaped function with parameters depending on soil organic carbon, soil clay content, and soil bulk density (Moyano et al., 2012). Thus, for wet soils, heterotrophic respiration fluxes are probably reduced because of oxygen limitation. Implementing this bell-shaped function approach is necessary to accurately represent the soil organic carbon stock of peatland in some land surface schemes used by ESMs (Qiu et al., 2019). The approach proposed by Moyano et al., (2012) seems well adapted to constrain ESMs since the author proposed several versions of the bell-shaped function and defined a function using drivers

380 that are included in ESMs (the model 2 in Moyano et al., (2012)). The model including bulk density might perform better but bulk density is not calculated by ESMs and consequently such approach is hardly implementable in ESMs. Other approaches have been proposed in the literature (Davidson et al., 2014; Sierra et al., 2014) but the solutions proposed are mostly based on O_2 diffusion which is more mechanistic but more difficult to implement in an ESM compared to a more empirical solution as proposed by Moyano et al. (2012). Gas diffusion implementation at the spatial resolution of ESMs is quite challenging because

385 it depends on drivers highly variables at small scales. Not considering the possible oxygen limitation effect on wet soils may explain why ESMs tends to overestimate the heterotrophic respiration flux when mean annual precipitation is high. Changing soil moisture function to better represent this effect should be relatively easy and may substantially improve the capacities of ESMs to reproduce the heterotrophic respiration flux.

Another important parameter controlling heterotrophic respiration flux is carbon use efficiency defined as the ratio between

390 the carbon remaining in a system and the carbon entering that system (Manzoni et al., 2018). In our context this is the ratio between the carbon mineralized through microbial heterotrophic respiration and the carbon incorporated into the microbial biomass. The heterotrophic respiration flux therefore results from two processes in ESMs, the soil organic carbon decomposition and its allocation to other soil carbon pools or to heterotrophic respiration. Carbon use efficiency is highly

9

variable and depends on several biotic and abiotic factors (Sinsabaugh et al., 2013; Manzoni, 2017; Manzoni et al., 2012). In ESMs, carbon use efficiency is neither dynamic nor spatially distributed, and thus having flexible carbon use efficiency control may help to reproduce observations (Zhang et al., 2018). A simple approach that may aid a better representation of heterotrophic respiration fluxes is optimizing the carbon use efficiency parameters of the ESMs using a Bayesian approach as is done for other land fluxes (Kuppel et al., 2012). This would result in a spatially distributed set of parameters for carbon use efficiency but this approach would not be dynamic. Another option that might benefit the current large carbon use efficiency measures existing in the literature (Manzoni et al., 2012) to define statistical functions predicting carbon use efficiency based on explanatory variables that could themselves be dynamic (soil temperature, pH, soil C:N ratio, etc.). Thus, carbon use efficiency might be spatialized and dynamic.

A better representation of the heterotrophic respiration flux is also important for other biogeochemical variables in particular in ESMs with explicit nitrogen cycle representation in their land surface scheme. Indeed, heterotrophic respiration fluxes are indicators of soil organic carbon decomposition but when nitrogen is explicitly represented it also becomes an indicator of soil N mineralization (Vuichard et al., 2018). In the field, the soil organic matter is composed by complex molecules made of carbon and nitrogen among others (Cleveland and Liptzin, 2007). Microorganisms decompose soil organic matter releasing CO_2 to the atmosphere and mineral nitrogen to the soil solution. Microbial activity is therefore a major driver of mineral nitrogen availability and partially control nitrogen limitation on primary production and therefore on land carbon sink (Bragazza et al., 2013). Since more and more ESMs explicitly represent the nitrogen cycle in their land surface models (Varney et al., 2022; Davies-Barnard et al., 2020) resulting in well constrained heterotrophic respiration fluxes that may help to constrain the nitrogen mineralization flux as they both come from the soil organic matter decomposition by extracellular enzymes. A better representation of the mineral N release flux would probably, in turn, improve the simulation of NPP.

5 Conclusion

Our study showed that despite previous ESMs evaluation of heterotrophic respiration (Shao et al., 2013), few current ESMs represented total heterotrophic respiration flux well and all failed at representing its spatial distribution. Since heterotrophic fluxes are large and are a major determinant of whether land surfaces represent a carbon sink or source it is of major importance to better constrain these fluxes and how they will be impacted by climate and land use changes. We showed that current ESMs failed to reproduce heterotrophic respiration fluxes where precipitation is important probably because heterotrophic respiration responses to soil moisture are poor representations of reality. Nevertheless, it is important to note that soil moisture is not only driven by precipitation. Other water fluxes like runoff, drainage and evapotranspiration affect the water balance in soils. In this study we did not directly consider soil moisture because it was not available for all the ESMs. Another limitation of our study is that we did not account for other important drivers of heterotrophic respiration in our model residual analysis like pH, microbial biomass, nitrogen availability, etc. We decided to focus on explanatory variables calculated by all the models because

10

we aimed to identify biases due to feedbacks between ESMs' variables rather than identifying missing mechanisms. We propose several options to improve the ESM without extensive modifications of the current schemes. We believe that our proposals would be relatively easy to implement in the next generation of ESMs resulting in possible improvements. Finally, we observed important discrepancies between the observation products we used. We decided to use the median as the best estimate here but to better constraint ESM in the future, observation-based products need to be more accurate. Increasing the number of observations used to create such gridded products is an obvious recommendation but this can be complicated. New satellite products like GOSAT or OCO-2 are measuring net ecosystem exchange (Palmer et al., 2019). These satellite observations represent good candidates to better constrain the observation products in the future.

Data availability

All data are available in the main text.

Author contributions

BG, LC, PB and LB designed the study, BG, JO, OT and LS performed the analysis. All the authors participated to the results interpretation and to the writing.

Competing interests

All other authors declare they have no competing interests.

Acknowledgments

References

Bond-Lamberty, B. and Thomson, A.: A global database of soil respiration data., Biogeosciences, 2010.

Bond-Lamberty, B., Wang, C., and Gower, S. T.: A global relationship between the heterotrophic and autotrophic components

of soil respiration?, Glob. Chang. Biol., 10, 1756–1766, https://doi.org/10.1111/j.1365-2486.2004.00816.x, 2004.

Bragazza, L., Parisod, J., Buttler, A., and Bardgett, R. D.: Biogeochemical plant–soil microbe feedback in response to climate

500 warming in peatlands, Nat. Clim. Chang., 3, 273–277, https://doi.org/10.1038/nclimate1781, 2013.

Chabbi, A., Lehmann, J., Ciais, P., Loescher, H. W., Cotrufo, M. F., Don, A., SanClements, M., Schipper, L., Six, J., Smith, P., and Rumpel, C.: Aligning agriculture and climate policy, Nat. Clim. Chang., 7, 307–309, https://doi.org/10.1038/nclimate3286, 2017.

Ciais, P., Yao, Y., Gasser, T., Baccini, A., Wang, Y., Lauerwald, R., Peng, S., Bastos, A., Li, W., Raymond, P. A., Canadell,

505 J. G., Peters, G. P., Andres, R. J., Chang, J., Yue, C., Dolman, A. J., Haverd, V., Hartmann, J., Laruelle, G., Konings, A. G., King, A. W., Liu, Y., Luyssaert, S., Maignan, F., Patra, P. K., Peregon, A., Regnier, P., Pongratz, J., Poulter, B., Shvidenko, A., Valentini, R., Wang, R., Broquet, G., Yin, Y., Zscheischler, J., Guenet, B., Goll, D. S., Ballantyne, A.-P., Yang, H., Qiu, C., and Zhu, D.: Empirical estimates of regional carbon budgets imply reduced global soil heterotrophic respiration, Natl. Sci. Rev., 2, https://doi.org/10.1093/nsr/nwaa145, 2021.

510 Cleveland, C. C. and Liptzin, D.: C:N:P stoichiometry in soil: is there a '"Redfield ratio"' for the microbial biomass, Biogeochemistry, 85, 235–252, 2007.

Davidson, E. a and Janssens, I. a: Temperature sensitivity of soil carbon decomposition and feedbacks to climate change., Nature, 440, 165–73, https://doi.org/10.1038/nature04514, 2006.

Davidson, E. a, Savage, K. E., and Finzi, A. C.: A big-microsite framework for soil carbon modeling., Glob. Chang. Biol.,

515 3610–3620, https://doi.org/10.1111/gcb.12718, 2014.

Davies-Barnard, T., Meyerholt, J., Zaehle, S., Friedlingstein, P., Brovkin, V., Fan, Y., Fisher, R. A., Jones, C. D., Lee, H., Peano, D., Smith, B., Wårlind, D., Wiltshire, A. J., Wärlind, D., Wiltshire, A. J., Wårlind, D., and Wiltshire, A. J.: Nitrogen cycling in CMIP6 land surface models: Progress and limitations, Biogeosciences, 17, 5129–5148, https://doi.org/10.5194/bg-17-5129-2020, 2020.

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

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, Geosci. Model Dev., 9, 1937–1958,

525 https://doi.org/10.5194/gmd-9-1937-2016, 2016.

Falloon, P., Jones, C. D., Ades, M., and Paul, K.: Direct soil moisture controls of future global soil carbon changes: An important source of uncertainty, Global Biogeochem. Cycles, 25, 1–14, https://doi.org/10.1029/2010GB003938, 2011.

Friedlingstein, P., O'sullivan, M., Jones, M. W., Andrew, R. M., Gregor, L., Hauck, J., Le Quéré, C., Luijkx, I. T., Olsen, A., Peters, G. P., Peters, W., Pongratz, J., Schwingshackl, C., Sitch, S., Canadell, J. G., Ciais, P., Jackson, R. B., Alin, S. R.,

530 Alkama, R., Arneth, A., Arora, V. K., Bates, N. R., Becker, M., Bellouin, N., Bittig, H. C., Bopp, L., Chevallier, F., Chini, L. P., Cronin, M., Evans, W., Falk, S., Feely, R. A., Gasser, T., Gehlen, M., Gkritzalis, T., Gloege, L., Grassi, G., Gruber, N.,

Gürses, Ö., Harris, I., Hefner, M., Houghton, R. A., Hurtt, G. C., Iida, Y., Ilyina, T., Jain, A. K., Jersild, A., Kadono, K., Kato, E., Kennedy, D., Klein Goldewijk, K., Knauer, J., Korsbakken, J. I., Landschützer, P., Lefèvre, N., Lindsay, K., Liu, J., Liu, Z., Marland, G., Mayot, N., Mcgrath, M. J., Metzl, N., Monacci, N. M., Munro, D. R., Nakaoka, S. I., Niwa, Y., O'brien, K.,

535   Ono, T., Palmer, P. I., Pan, N., Pierrot, D., Pocock, K., Poulter, B., Resplandy, L., Robertson, E., Rödenbeck, C., Rodriguez, C., Rosan, T. M., Schwinger, J., Séférian, R., Shutler, J. D., Skjelvan, I., Steinhoff, T., Sun, Q., Sutton, A. J., Sweeney, C., Takao, S., Tanhua, T., Tans, P. P., Tian, X., Tian, H., Tilbrook, B., Tsujino, H., Tubiello, F., Van Der Werf, G. R., Walker, A. P., Wanninkhof, R., Whitehead, C., Willstrand Wranne, A., et al.: Global Carbon Budget 2022, Earth Syst. Sci. Data, 14, 4811–4900, https://doi.org/10.5194/essd-14-4811-2022, 2022.

540   Hartmann, J. and Moosdorf, N.: The new global lithological map database GLiM: A representation of rock properties at the Earth surface, Geochemistry, Geophys. Geosystems, 13, 1–37, https://doi.org/10.1029/2012GC004370, 2012.

Hashimoto, S., Carvalhais, N., Ito, A., Migliavacca, M., Nishina, K., and Reichstein, M.: Global spatiotemporal distribution of soil respiration modeled using a global database, Biogeosciences, 12, 4121–4132, https://doi.org/10.5194/bg-12-4121-2015, 2015.

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

Huang, Y., Guenet, B., Wang, Y. L., and Ciais, P.: Global Simulation and Evaluation of Soil Organic Matter and Microbial

550   Carbon and Nitrogen Stocks Using the Microbial Decomposition Model ORCHIMIC v2.0, Global Biogeochem. Cycles, 35, 1–20, https://doi.org/10.1029/2020GB006836, 2021.

Ito, A., Hajima, T., Lawrence, D. M., Brovkin, V., Delire, C., Guenet, B., Jones, C. D., Malyshev, S., Materia, S., McDermid, S. P., Peano, D., Pongratz, J., Robertson, E., Shevliakova, E., Vuichard, N., Wårlind, D., Wiltshire, A., and Ziehn, T.: Soil carbon sequestration simulated in CMIP6-LUMIP models: Implications for climatic mitigation, Environ. Res. Lett., 15,

555   124061, https://doi.org/10.1088/1748-9326/abc912, 2020.

Jones, C. D. D., Cox, P., and Huntingford, C.: Uncertainty in climate–carbon-cycle projections associated with the sensitivity of soil respiration to temperature, Tellus B, 55, 642–648, 2003.

Konings, A. G., Anthony Bloom, A., Liu, J., Parazoo, N. C., Schimel, D. S., and Bowman, K. W.: Global satellite-driven estimates of heterotrophic respiration, Biogeosciences, 16, 2269–2284, https://doi.org/10.5194/bg-16-2269-2019, 2019.

560   Kuppel, S., Peylin, P., Chevallier, F., Bacour, C., Maignan, F., and Richardson, a. D.: Constraining a global ecosystem model with multi-site eddy-covariance data, Biogeosciences, 9, 3757–3776, https://doi.org/10.5194/bg-9-3757-2012, 2012.

Manzoni, S.: Flexible carbon-use efficiency across litter types and during decomposition partly compensates nutrient imbalances-results from analytical stoichiometric models, Front. Microbiol., 8, 1–15, https://doi.org/10.3389/fmicb.2017.00661, 2017.

565   Manzoni, S., Taylor, P., Richter, A., Porporato, A., and Ågren, G. I.: Environmental and stoichiometric controls on microbial

13

carbon-use efficiency in soils, New Phytol., 196, 79–91, https://doi.org/10.1111/j.1469-8137.2012.04225.x, 2012.

Manzoni, S., Capek, P., Porada, P., Thurner, M., Winterdahl, M., Beer, C., Vico, G., Way, D., Čapek, P., Porada, P., Thurner, M., Winterdahl, M., Beer, C., Brüchert, V., Frouz, J., Herrmann, A. M., Lindahl, B. D., Lyon, S. W., Šantrůčková, H., Vico, G., Way, D., Capek, P., Porada, P., Thurner, M., Winterdahl, M., Beer, C., Vico, G., Way, D., Čapek, P., Porada, P., Thurner,
570 M., Winterdahl, M., Beer, C., Brüchert, V., Frouz, J., Herrmann, A. M., Lindahl, B. D., Lyon, S. W., Šantrůčková, H., Vico, G., Way, D., Capek, P., Porada, P., Thurner, M., Winterdahl, M., Beer, C., Vico, G., and Way, D.: Reviews and syntheses : Carbon use efficiency from organisms to ecosystems – definitions , theories , and empirical evidence, Biogeosciences, 15, 5929–5949, https://doi.org/10.5194/bg-2018-275, 2018.

Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gommis,
575 M. I. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R. B. R., Maycock, T. K. K., Waterfield, T., Yelekçi, O., Yu, R., Zhou, B., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gommis, M. I. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R. B. R., Maycock, T. K. K., Waterfield, T., Yelekçi, O., Yu, R., Zhou, B., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gommis, M. I. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R. B. R., Maycock, T. K. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B.: Climate change 2021: The physical science basis summary for policymakers. Contribution of
580 Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge Univerisyt Press, https://doi.org/10.1260/095830507781076194, 2021.

Moyano, F. E., Vasilyeva, N., Bouckaert, L., Cook, F., Craine, J., Curiel Yuste, J., Don, a., Epron, D., Formanek, P., Franzluebbers, a., Ilstedt, U., Kätterer, T., Orchard, V., Reichstein, M., Rey, a., Ruamps, L., Subke, J. -a., Thomsen, I. K., and Chenu, C.: The moisture response of soil heterotrophic respiration: interaction with soil properties, Biogeosciences, 9,
585 1173–1182, https://doi.org/10.5194/bg-9-1173-2012, 2012.

Le Noë, J., Manzoni, S., Abramoff, R. Z., Bruni, E., Cardinael, R., Ciais, P., Chenu, C., Clivot, H., Derrien, D., Ferchaud, F., Garnier, P., Goll, D., Lashermes, G., Martin, M., Rasse, D. P., Rees, F., Sainte-Marie, J., Salmon, E., Schiedung, M., Schimel, J., Wieder, W. R., Abiven, S., Barré, P., Cécillon, L., and Guenet, B.: Soil organic carbon models need more independent time-series validation for reliable predictions, Commun. Earth Environ., 1–8, https://doi.org/10.1038/s43247-023-00830-5, 2023.

590 Palmer, P. I., Feng, L., Baker, D., Chevallier, F., Bösch, H., and Somkuti, P.: Net carbon emissions from African biosphere dominate pan-tropical atmospheric CO2 signal, Nat. Commun., 10, 1–9, https://doi.org/10.1038/s41467-019-11097-w, 2019.

Qiu, C., Zhu, D., Ciais, P., Guenet, B., Peng, S., Krinner, G., Tootchi, A., Ducharne, A., and Hastie, A.: Modelling northern peatlands area and carbon dynamics since the Holocene with the ORCHIDEE-PEAT land surface model (SVN r5488), Geosci. Model Dev., 12, 2961–2982, https://doi.org/10.5194/gmd-2018-256, 2019.

595 Raich, J. W. W., Potter, C. S. S., and Bhagawati, D.: Interannual variability in global soil respiration, 1980–94, Glob. Chang. Biol., 8, 800–812, https://doi.org/10.1046/j.1365-2486.2002.00511.x, 2002.

Scharlemann, J. P., Tanner, E. V., Hiederer, R., and Kapos, V.: Global soil carbon: understanding and managing the largest terrestrial carbon pool, Carbon Manag., 5, 81–91, https://doi.org/10.4155/cmt.13.77, 2014.

Shao, P., Zeng, X., Moore, D. J. P., and Zeng, X.: Soil microbial respiration from observations and Earth System Models,

600 Environ. Res. Lett., 8, https://doi.org/10.1088/1748-9326/8/3/034034, 2013.

Sierra, C., Trumbore, S., Davidson, E., Vicca, S., and Janssens, I.: Sensitivity of soil organic matter decomposition to simultaneous changes in temperature and moisture, 16, 10086, 2014.

Sinsabaugh, R. L., Manzoni, S., Moorhead, D. L., and Richter, A.: Carbon use efficiency of microbial communities: stoichiometry, methodology and modelling., Ecol. Lett., 16, 930–939, https://doi.org/10.1111/ele.12113, 2013.

605 Stoy, P. C., Dietze, M. C., Richardson, A. D., Vargas, R., Barr, A. G., Anderson, R. S., Arain, M. A., Baker, I. T., Black, T. A., Chen, J. M., Cook, R. B., Gough, C. M., Grant, R. F., Hollinger, D. Y., Izaurralde, R. C., Kucharik, C. J., Lafleur, P., Law, B. E., Liu, S., Lokupitiya, E., Luo, Y., Munger, J. W., Peng, C., Poulter, B., Price, D. T., Ricciuto, D. M., Riley, W. J., Sahoo, A. K., Schaefer, K., Schwalm, C. R., Tian, H., Verbeeck, H., and Weng, E.: Evaluating the agreement between measurements and models of net ecosystem exchange at different times and timescales using wavelet coherence: An example using data from

610 the North American Carbon Program Site-Level Interim Synthesis, Biogeosciences, 10, 6893–6909, https://doi.org/10.5194/bg-10-6893-2013, 2013.

Subke, J. A., Inglima, I., and Cotrufo, M. F.: Trends and methodological impacts in soil CO2 efflux partitioning: A metaanalytical review, https://doi.org/10.1111/j.1365-2486.2006.01117.x, 1 June 2006.

Todd-Brown, K., Zheng, B., and Crowther, T. W.: Field-warmed soil carbon changes imply high 21st-century modeling

615 uncertainty, Biogeosciences, 15, 3659–3671, https://doi.org/10.5194/bg-15-3659-2018, 2018.

Todd-Brown, K. E. O., Randerson, J. T., Hopkins, F., Arora, V., Hajima, T., Jones, C., Shevliakova, E., Tjiputra, J., Volodin, E., Wu, T., Zhang, Q., and Allison, S. D.: Changes in soil organic carbon storage predicted by Earth system models during the 21st century, Biogeosciences, 11, 18969–19004, https://doi.org/10.5194/bgd-10-18969-2013, 2014.

Varney, R. M., Chadburn, S. E., Friedlingstein, P., Koven, C. D., Hugelius, G., Cox, P. M., and Burke, E. J.: soil carbon

620 turnover to global warming, Nat. Commun., 4–11, https://doi.org/10.1038/s41467-020-19208-8, 2020.

Varney, R. M., Chadburn, S. E., Burke, E. J., and Cox, P. M.: Evaluation of soil carbon simulation in CMIP6 Earth System Models, Biogeosciences, 6, 4671–4704, 2022.

Venables, W. N. and Ripley, B. D.: Modern Applied Statistics with S, Springer New York, New York, NY, https://doi.org/10.1007/978-0-387-21706-2, 2002.

625 Vuichard, N., Messina, P., Luyssaert, S., Guenet, B., Zaehle, S., Ghattas, J., Bastrikov, V., and Peylin, P.: Accounting for Carbon and Nitrogen interactions in the Global Terrestrial Ecosystem Model ORCHIDEE (trunk version, rev 4999): multi-scale evaluation of gross primary production, Geosci. Model Dev., 12, 1–41, https://doi.org/10.5194/gmd-2018-261, 2018.

Wang, X., Piao, S., Ciais, P., Janssens, I. a., Reichstein, M., Peng, S., and Wang, T.: Are ecological gradients in seasonal Q10 of soil respiration explained by climate or by vegetation seasonality?, Soil Biol. Biochem., 42, 1728–1734,

630 https://doi.org/10.1016/j.soilbio.2010.06.008, 2010.

Warner, D. L., Bond-Lamberty, B., Jian, J., Stell, E., and Vargas, R.: Spatial Predictions and Associated Uncertainty of Annual Soil Respiration at the Global Scale, Global Biogeochem. Cycles, 33, 1733–1745, https://doi.org/10.1029/2019GB006264, 2019.

Wieder, W. R., Allison, S. D., Davidson, E. a, Georgiou, K., Hararuk, O., He, Y., Hopkins, F., Luo, Y., Smith, M., Sulman,
635   B. N., Todd-Brown, K. E. O., Wang, Y., Xia, J., and Xu, X.: Explicitly representing soil microbial processes in Earth system
models, Global Biogeochem. Cycles, 29, https://doi.org/10.1002/2015GB005188, 2015.

Zhang, H., Goll, D. S., Manzoni, S., Ciais, P., Guenet, B., and Huang, Y.: Modeling the effects of litter stoichiometry and soil
mineral N availability on soil organic matter formation using CENTURY-CUE (v1.0), Geosci. Model Dev., 11, 4779–4796,
https://doi.org/10.5194/gmd-11-4779-2018, 2018.
640   Zhao, M., Heinsch, F. A., Nemani, R. R., and Running, S. W.: Improvements of the MODIS terrestrial gross and net primary
production global data set, Remote Sens. Environ., 95, 164–176, https://doi.org/10.1016/j.rse.2004.12.011, 2005.

Zhou, T., Shi, P., Hui, D., and Luo, Y.: Global pattern of temperature sensitivity of soil heterotrophic respiration (Q_{10}) and
its implications for carbon-climate feedback, J. Geophys. Res., 114, G02016, https://doi.org/10.1029/2008JG000850, 2009.
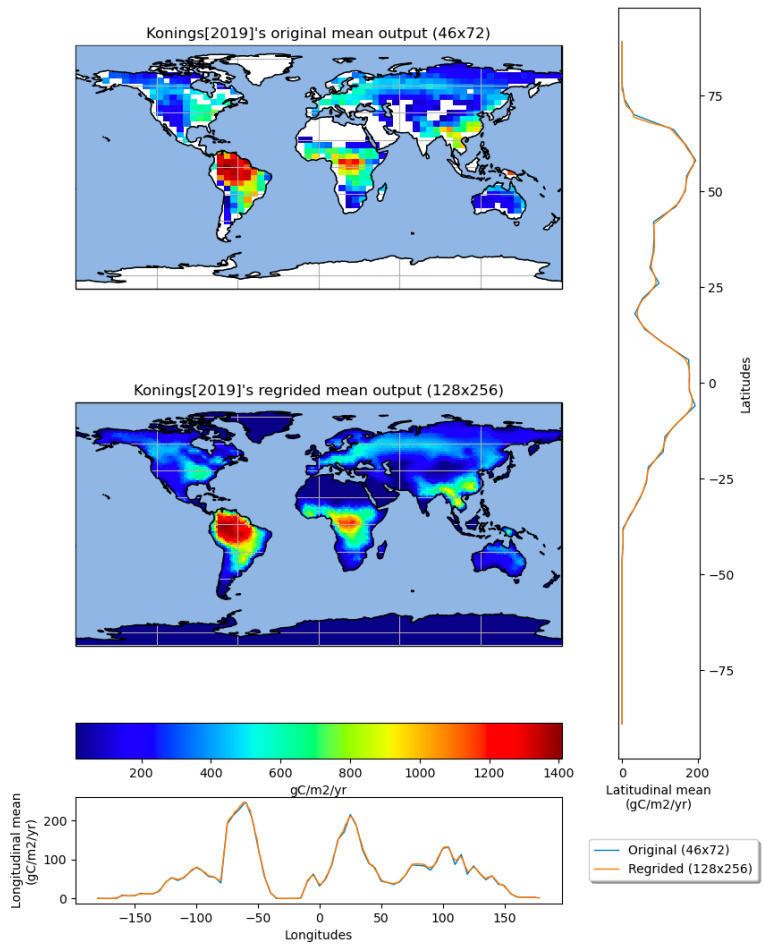
645

Fig. 1. Mean Rh spatial distribution over 2010-2012 from the Konings et al., (2019) product –original (46x72, top panel) vs regrided (128x256, bottom panel).
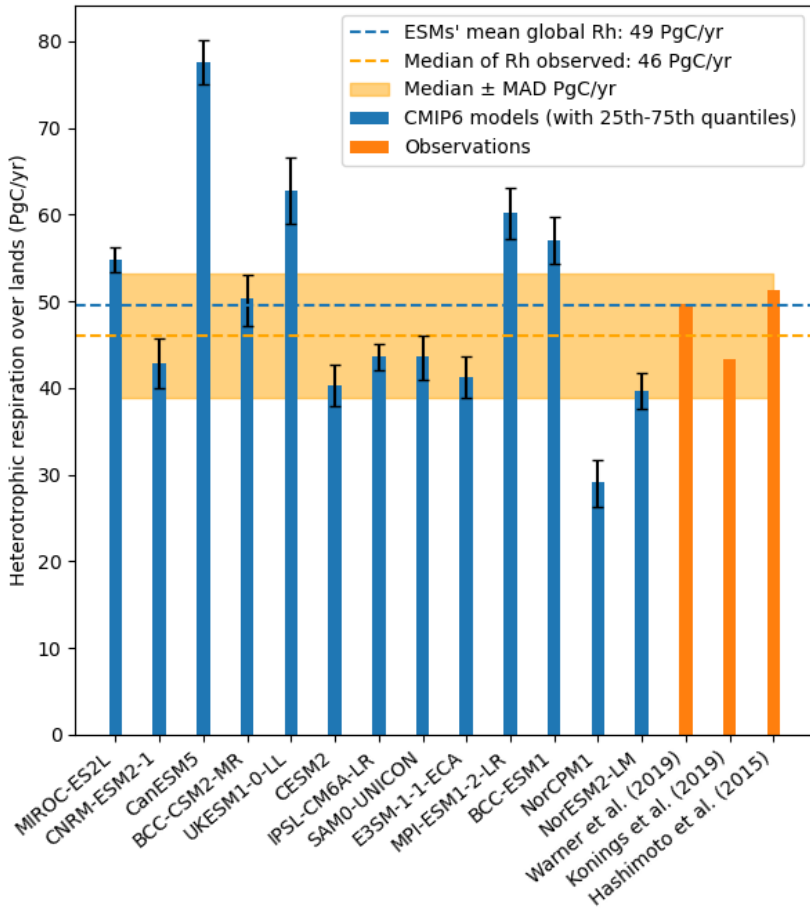
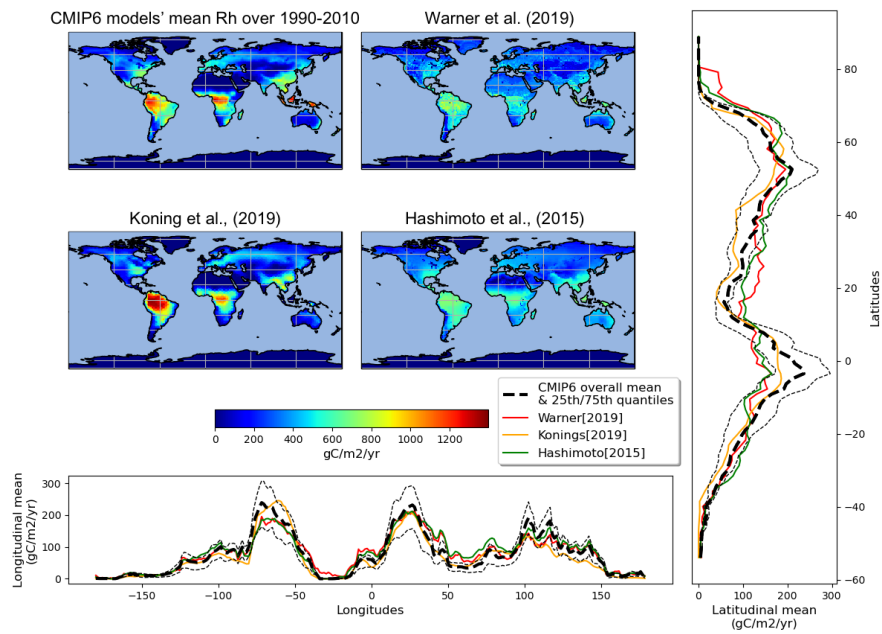**Fig. 2. Global estimations of soil heterotrophic respiration mean over 1990-2010 period.**

650

Fig. 3. Comparison of mean soil heterotrophic respiration spatial distribution among mean CMIP6 outputs and observation data.
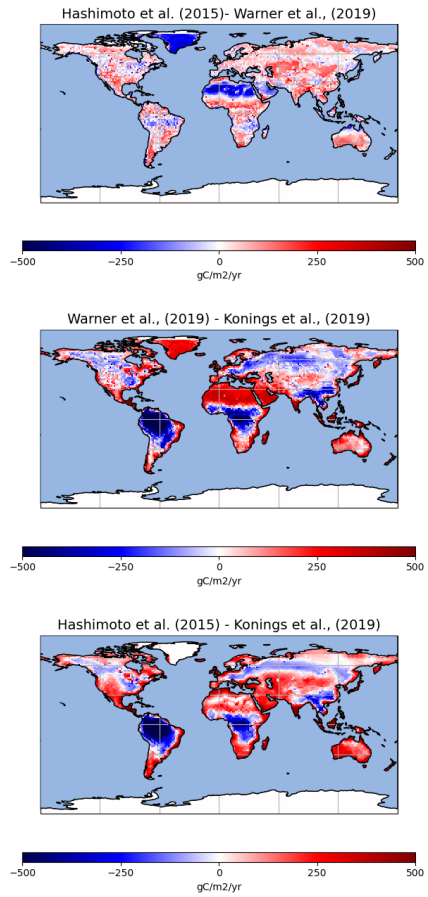
655

Fig. 4. Maps of difference between the observation based product used in this study (Hashimoto et al., (2015) – Warner et al., (2019) in the top panel, Warner et al., (2019) – Konings et al. (2019) in the middle panel and Hashimoto et al., (2015) – Konings et al. (2019) in the bottom panel).
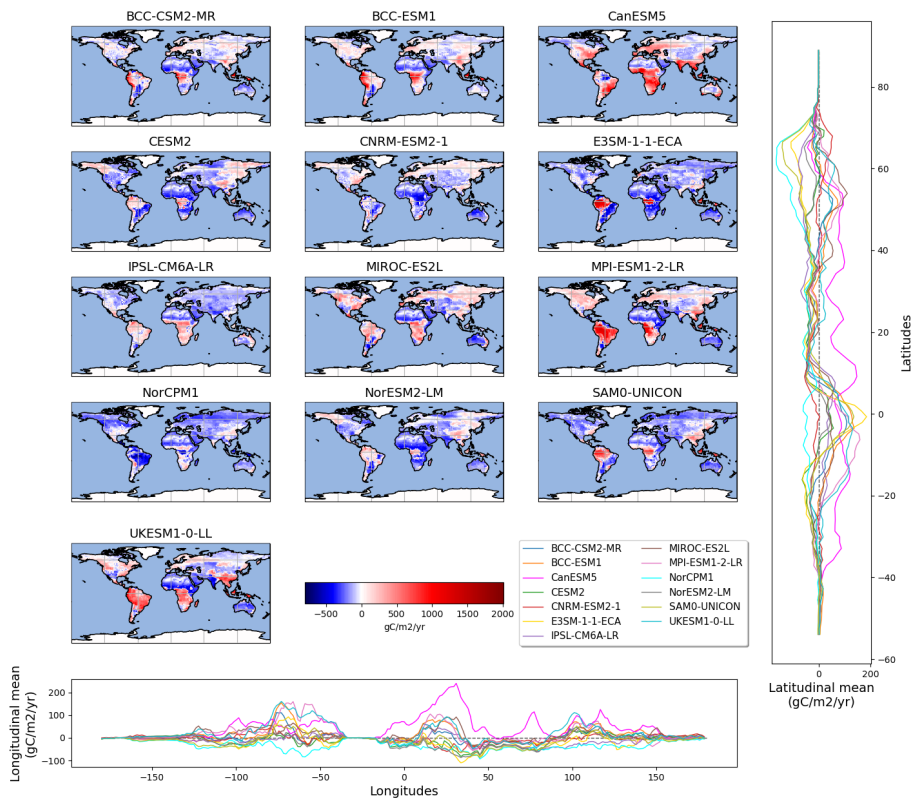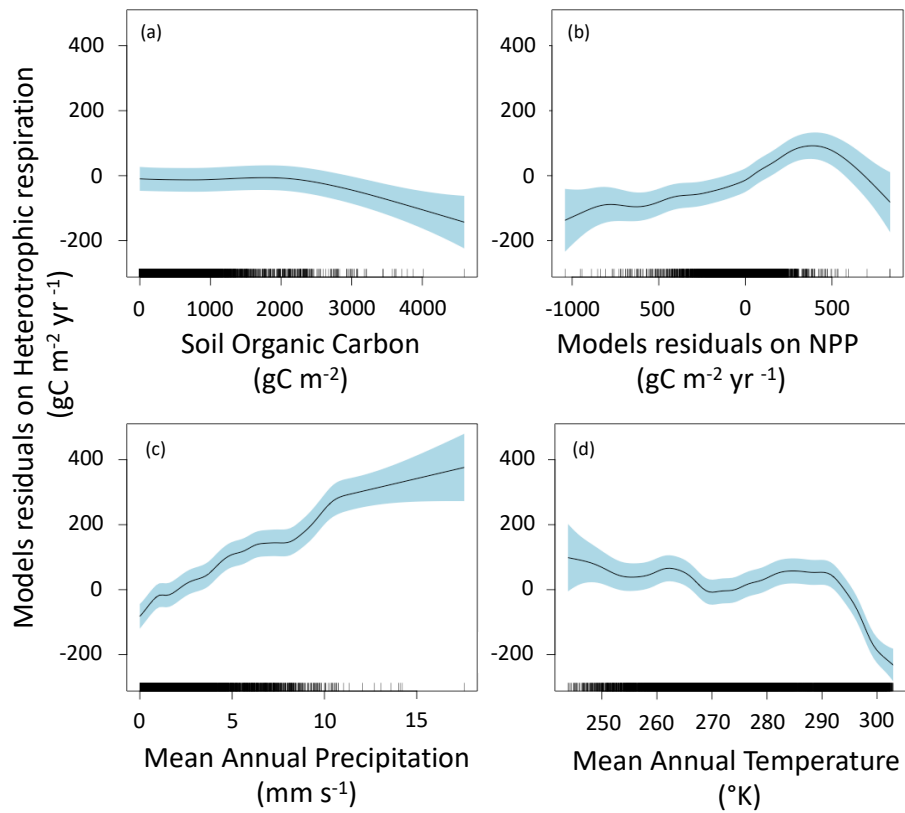
660

Fig. 5. Spatially distributed residuals of CMIP6 ESMs predictions over the period 1990-2010 with respect to median observation products.

Fig. 6. Median of ESMs residuals on soil heterotrophic respiration. The residuals are explained by soil organic carbon (a), median of NPP residuals (b), mean annual precipitation (c) and mean annual temperature (d). Negative values mean model underestimation.
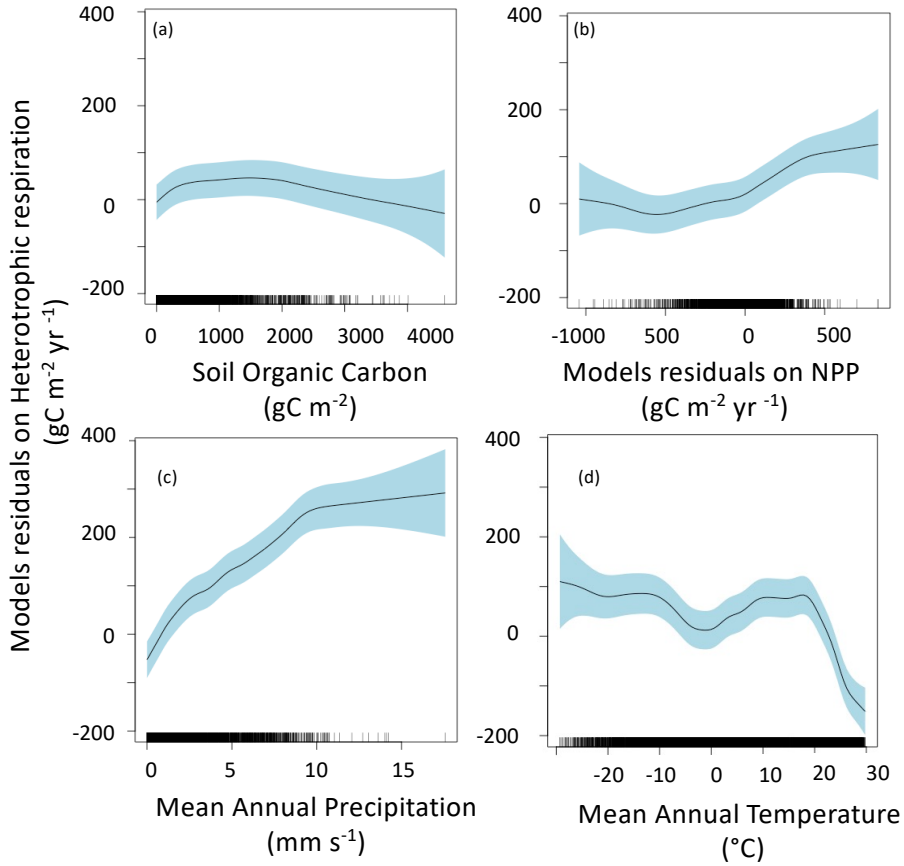
Fig. 7. Mean of ESMs residuals on soil heterotrophic respiration. The residuals are explained by soil organic carbon (a), mean of NPP residuals (b), mean annual precipitation (c) and mean annual temperature (d). Negative values mean model underestimation.

Table 1. Evaluation metrics for the different Earth system models.

Models	BCC-CSM2-MR	BCC-ESM1	CanESM5	CESM2	CNRM-ESM2-1	E3SM-1-1-ECA	IPSL-CM6A-LR	MIROC-ES2L	MPI-ESM1-2-LR	NorCPM1	NorESM2-LM	SAM0-UNICON	UKESM1-0-LL
% of grid cells within median ± mad	43.9	44.2	30.4	32.6	42.0	29.9	36.2	25.0	30.2	27.1	36.0	25.1	27.7
RMSE	224.4	229.0	345.1	199.2	171.1	281.4	170.9	229.3	314.0	212.3	187.7	231.9	302.1
R2	0.75	0.78	0.80	0.72	0.79	0.57	0.79	0.82	0.73	0.68	0.74	0.69	0.75

675

Table 2. Summary of the main features proposed in this study to improve the heterotrophic respiration fluxes in Earth system models.

	Main features to improve in the next ESM generation
NPP residues	Improving plant inputs through NPP is key to improve heterotrophic respiration by implementing N cycle for instance.
MAT	Dynamic and/or spatialized temperature sensitivity parameters such as Q10.
MAP	Improving the soil moisture functions using bell shape functions for instance.
SOC	More constrained parameters such as CUE and/or residence times.

680

24