Answer to comments from the reviewer #1.

We thank reviewer for the constructive evaluation of the manuscript. Please find below our answers to questions/comments. Comments from the reviewer were left intentionally in this document and written in roman font. Our answers are written in italics.

This study provides useful analysis on Rh within CMIP6 ESMs against observational datasets and provides useful direction in future development of ESMs. It is a study that is well suited for publication in Biogeosciences. Overall, the study is clear and well written, however some improved integration with existing studies and increased detail on the caveats will improve the study.

Thank you for the positive comment. The revised version of the manuscript will be improved including your suggestions

Major comments

Throughout the study it is referred to as being the 'first' to investigate heterotrophic respiration (Rh) in Earth system models (ESMs), though this is not the case. This study is still novel, however wording needs to be addressed here to include how this study fits in with the existing literature.

For example, in the abstract: "*capacity of Earth System Models (ESMs) to reproduce this flux has never been evaluated*" and "*for the first time*". Also, Line 182 in Discussion.

Relevant existing studies include:

1. Shao et al., 2013. This study evaluates Rh in CMIP5 ESMs against observational datasets (*Soil microbial respiration from observations and Earth System Models*).
2. Varney et al., 2022. This study focuses on soil carbon and has been cited, however spatial evaluation of soil carbon turnover (Cs / Rh) is included, and tables of global Rh values in CMIP6 and CMIP5 ESMs against observational dataset (Tables A1 and A2).

Yes, this true we have been a bit clumsy in the wording, this has been corrected in several part of the revised version (see below for some examples).

Line 183 – It is unclear to the reader what is meant here and there is no citation to back up this statement or to add clarity. Why is it that previously Rh in ESMs could only be constrained by NEE or ecosystem respiration? If the reason is lack of observational datasets, there are older soil respiration datasets (such as Raich et al., 2002 mentioned)? Plus, existing evaluation study on Rh in CMIP5 ESMs? Please expand on why this is the case or change the motivation behind the sentence.

We clarified our point as following: "Indeed, previous dataset were not gridded and so far spatial pattern of heterotrophic respiration in ESMs could only by constraint 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)*."*

Line 253 – Similar point here.

We also rephrase to clarify : "Our study showed that despite previous ESMs evaluation on heterotrophic respiration (Shao et al., 2013)*, a few current ESMs are fairly representing the total heterotrophic respiration flux but …"*

Line 36 – This sentence states that Rh has not been well incorporated into ESMs. If this is the case but this is the first study to evaluate this, how do we know? References need to be included here to back up this statement.

We rephrase to clarify : "Despite the importance of heterotrophic respiration fluxes, the scheme representing this flux in ESMs, which aim to simulate the most important drivers of the earth's climate system, are currently challenged because important drivers are missing (Huang et al., 2021; Wieder et al., 2015) *but the proposed new schemes lacks of sufficient evaluation on long term time series (Le Noë et al., 2023). Thus, how accurate are the prediction of ESMs for heterotrophic respiration fluxes is a key question to well constraint the carbon climate feedbacks in ESMs."*

Line 189 – The study notes large discrepancies in the observational datasets and is presented as an issue which needs to be addressed in the future. It would be beneficial to see more direct comparisons of the observational datasets. I think a useful addition to either an Appendix or Supplementary material would be comparing the observational datasets, potentially a correlation coefficient between them? I know maps of each are included in Fig. 3, but a quantification or difference map would be useful to see where there is more agreement or less agreement between them.

In the revised version we have added difference maps between the products in the supplementary materials.

It has previously been shown that the Hashimoto et al., 2015 dataset has an arbitrary maximum respiration level (see Supplementary Fig. 4 in Varney et al., 2020), which was shown in the same figure to not appear in additional respiration datasets. I think this point found here should be acknowledged and think about whether this could impact your residual results. Potentially the underestimation of Rh at high temperatures (Fig. 4)?

1. Varney, R.M., Chadburn, S.E., Friedlingstein, P. et al., A spatial emergent constraint on the sensitivity of soil carbon turnover to global warming. Nature Communications. **11**, 5544 (2020). https://doi.org/10.1038/s41467-020-19208-8.

Very good point, indeed this is an interesting suggestion and we add information on the revised version: "… and then a sudden underestimation for warm temperatures above 290K corresponding to tropical and dry climate zones. This sudden underestimation might be

explained by the 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)*. Such bias can therefore should be a consequence of the observation-based products used here rather than a real ESMs bias."*

Line 139 states that the observational data and ESM data Rh means are close in Boreal regions. However, on line 163 it is stated that Rh is underestimated by ESMs for soils rich in carbon (which tend to be boreal regions). Any idea why this is the case?

Indeed it might sounds surprising but two points may explain this. First, some peatlands are also in the tropic and because of the temperature conditions soil heterotrophic respiration may be higher and therefore impact more the results. Secondly, in boreal regions soils are carbon rich but temperature is cold and the bias explained by soil organic carbon can be compensated by the bias due to temperature that goes in the opposite direction (Fig. 4).

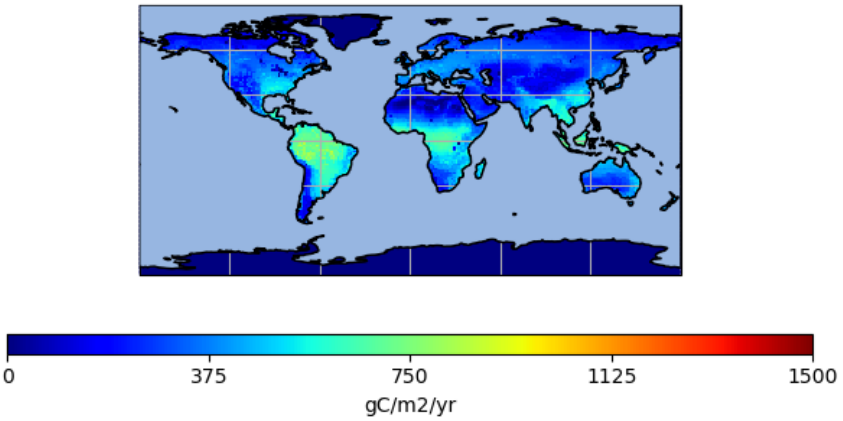
Paragraph from Line 140 – only tropics and temperate regions mentioned, what about the northern latitudes?

We added information in the revised version. "Models perform relatively well in temperate regions with for instance 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 largely underestimating the heterotrophic respiration fluxes (e.g. NorCPM1 or SAM0-UNICON) and other overestimating the fluxes (MPI-ESM1-2-LR). The BCC models (BCC-CSM2-MR and BCC-ESM1) were performing quite well over this region.

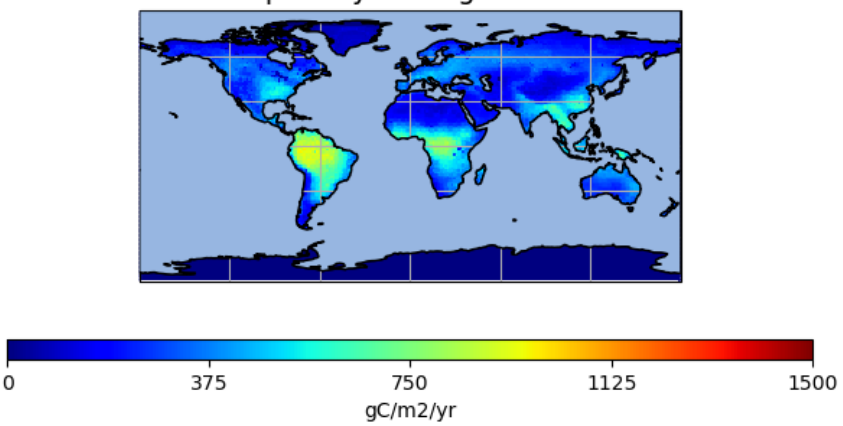
The use of the ESM and observational median is used throughout this study. I was wondering whether as it is not known which dataset or model is 'better', a mean value would give equal waiting to each, so could be a fairer metric. Does redoing the analysis with the mean instead affect the results? Especially spatially in regions where the datasets disagree more (Fig. 3)? If it does make a difference, it might be worth thinking about which is better for what you are trying to show or including in the Supplementary Material.

We tried earlier to work with means and it did not change drastically the results. We decided to present medians instead of means because it was more adapted to small size populations. In the supplementary material of the revised version, we will present similar analysis with means instead of medians. You can find below for instance a comparison of the products means and medians. The mean and the median have similar patterns but the heterotrophic respiration is higher in the tropic with the mean because the weight of the Konnings et al products is higher when calculating the mean.

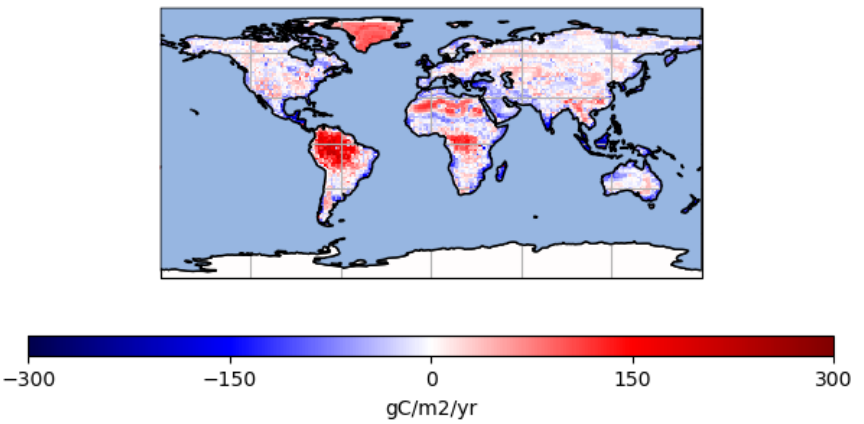
Median of the three spatially averaged observations datasets



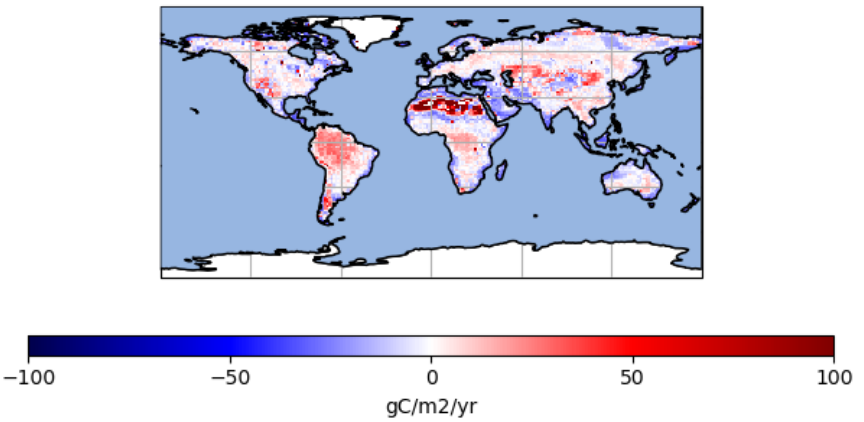
Mean of the three spatially averaged observations datasets



Absolute difference mean - median of observations



Relative difference (%) mean - median of observations



Line 202 – The temperature sensitivity of soil carbon turnover time (Cs / Rh) has been previously investigated in similar ESMs, including discussion on variable Q10s spatially and a constraint on effective Q10 in ESMs (Koven et al., 2017 and Varney et al., 2020). This might link with some of the discussion in this paragraph.

1. Koven, C., Hugelius, G., Lawrence, D. et al., Higher climatological temperature sensitivity of soil carbon in cold than warm climates. Nature Climate Change. **7**, 817–822 (2017). https://doi.org/10.1038/nclimate3421.
2. Varney, R.M., Chadburn, S.E., Friedlingstein, P. et al,. A spatial emergent constraint on the sensitivity of soil carbon turnover to global warming. Nature Communications. **11**, 5544 (2020). https://doi.org/10.1038/s41467-020-19208-8.

We modified the text in the revised version: "… with fixed parameters not dynamic and not spatially distributed (Ito et al., 2020). Previous studies suggested that a spatially distributed Q10 constrained on observations would be an important step to improve ESMs (Koven et al., 2017; Varney et al., 2020)*. Our results are online with this statement and suggest that having more flexible Q10 parameters may help to improve ESMs capacities to reproduce observation-derived products of heterotrophic respiration fluxes.*

Line 203 – Could the underestimation in these regions be due to little or no soil carbon in these regions within ESMs (Varney et al. 2022)?

Exact, we added this information : "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, Middle East or Northern Africa tend to be underestimated. Nevertheless, the underestimation observed in these regions can be also due to reduced C inputs and low SOC stocks reducing mechanically the heterotrophic respiration fluxes."

Line 209 – I would also include a more recent reference, for example, Todd-Brown et al., 2018 (*Field-warmed soil carbon changes imply high 21st-century modeling uncertainty*). In this study Q10 values are derived and the sensitivity of ESMs to this parameter is investigated.

The Todd-Brown et al. reference will be added in the revised version.

Minor Comments

All the minor comments will be considered in the revised version.

Abstract – I would include that you are looking at CMIP6 ESMs here as I had to skim to the end of the introduction to check this, and it is useful to know upfront.

Line 31 / Line 188 – Update Friedlingstein et al., 2020 reference to Friedlingstein et al., 2022. As this is the most up to date Global Carbon Budget paper.

Line 66 – I don't think this sentence makes sense "*, which were used to derived two observation products we used.*" I think it should be "*dervive the*", rather than "*derived*".

Line 114 – The acronym AIC is used in this study, but it is not defined. I would at least add a sentence in the Methods to describe what this term measures.

Line 155 – Ito et al., 2020 is cited here, however the first order kinetics of decomposition is not discussed in this study that I can see. Todd-Brown et al. 2013 and Varney et al. 2022 include information and discussion about ESM decomposition dependencies to temperature and precipitation.

Line 160 – "*Since the drivers are ...*" might be worth changing to "*Since the **main** drivers are ...*" as many factors affecting respiration, as stated in your conclusions (Schmidt et al., 2011).

1. Schmidt, M., Torn, M., Abiven, S. et al., Persistence of soil organic matter as an ecosystem property. Nature. **478**, 49–56 (2011). https://doi.org/10.1038/nature10386.

Line 178 – Maybe better to present temperatures in degrees C rather than K in European journal, and better relates to how 1.5C / 2C targets are often presented.

Line 183 – This sentence might change due to an above comment, but there is a typo. "*by constrtaint*" should read "*be constrained*".

Line 210 – Do you mean Figure 4c here? I would include this in brackets so reader can be reminded where this result came from.

References cited

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.
Huang, Y., Guenet, B., Wang, Y. L., and Ciais, P.: Global Simulation and Evaluation of Soil Organic Matter and Microbial 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.
Koven, C. D., Hugelius, G., Lawrence, D. M., and Wieder, W. R.: Higher climatological temperature sensitivity of soil carbon in cold than warm climates, Nat. Clim. Chang., 7, 817–822, https://doi.org/10.1038/nclimate3421, 2017.
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.
Shao, P., Zeng, X., Moore, D. J. P., and Zeng, X.: Soil microbial respiration from observations and Earth System Models, Environ. Res. Lett., 8, https://doi.org/10.1088/1748-9326/8/3/034034, 2013.
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 the North American Carbon Program Site-Level Interim Synthesis, Biogeosciences, 10, 6893–6909, https://doi.org/10.5194/bg-10-6893-2013, 2013.
Varney, R. M., Chadburn, S. E., Friedlingstein, P., Koven, C. D., Hugelius, G., Cox, P. M., and Burke, E. J.: soil carbon turnover to global warming, Nat. Commun., 4–11, https://doi.org/10.1038/s41467-020-19208-8, 2020.
Wieder, W. R., Allison, S. D., Davidson, E. a, Georgiou, K., Hararuk, O., He, Y., Hopkins, F., Luo, Y., Smith, M., Sulman, 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.