We thank the reviewer for their helpful comments, and agree that more detailed sensitivity analyses can clarify the uncertainty of the method. A point-by-point response is included below.

Konings and colleagues aimed to derive global, satellite-driven estimates of het-erotrophic respiration.

Reviewer: Here already lies the problem with the manuscript: Konings and colleagues focus too much on deriving the individual ecosystem fluxes that make up Rh top-down. GPP is derived from sun-induced fluorescence (top-down), but the uncertainty from using bottom-up estimates such as FLUXCOM is not evaluated. To my mind it should not matter if all fluxes that can be used to derive Rh top-down are also top-down estimates. Instead of using GPP from SIF also FLUXCOM-GPP (bottom-up) could be used –would that make a difference regarding spatial patterns?

Response: As we discuss in Sec 4.1, bottom-up estimates are generally based on sparse samples that are often not representatives. For example, FLUXCOM products are highly undersampled in the tropics. For example, of the 225 sites used to train FLUXCOM, only 17 were in the Southern Hemisphere, and of those 4 were in Australia. With such a low number of training data in the wet and dry tropics, the model is likely overfitting in the climatic conditions of these regions. This is why we chose the more top-down approach here. We explicitly contrast this to the best-available bottom-up dataset for R_h and show the uncertainties are comparable.

Nevertheless, we will include a sensitivity analysis using FLUXCOM GPP in the revised manuscript. While using FLUXCOM GPP affects the spatial patterns of the resulting R_h more than the CUE sensitivity analysis in Fig. 4 of the current manuscript (which is also consistent with the uncertainty analysis in Fig. 5 of the current manuscript), the difference in the spatial difference of R_h between different GPP assumptions is still less than the difference between the top-down and bottom-up products. The absolute differences are largest in the mid-latitudes and boreal regions.

Reviewer: For NEP the authors should discuss the effect of different products, for example Jena CarboScope NEP (http://www. bgc-jena.mpg.de/CarboScope/) or Chevallier et al. (2010) or FLUXCOM (Zscheischler et al., 2017) (how problematic this may be).

Response: An exhaustive discussion of propagated effects of different NEP and GPP product combinations is beyond the scope of this manuscript. However, in the revised version, we will add an uncertainty analysis for NEP with a value of zero everywhere, as suggested by this reviewer in a later comment. Doing so changes the spatial distribution of mean R_h less than changing the GPP assumptions, but does create a greater (normalized) root-mean-square-difference with the baseline R_h. We will discuss these results in the revised manuscript.

Reviewer: On a similar note, one can get an estimate of Rh from CARDAMOM: this should be very much dictated by data. How does Rh from CARDAMOM compare to the satellite-driven estimates and Hashimoto's approach?

Response: While CARDAMOM uses model-data fusion to incorporate information from remote sensing products, its soil carbon pools and fluxes are much less well constrained than the aboveground carbon pools and fluxes. The version of CARDAMOM used in this manuscript (that most consistent with the published literature, e.g Bloom et al PNAS 2016) predicts heterotrophic respiration only as a function of pixel-dependent base respiration rates, turnover time, and temperature, but does not account for water. As such, while CARDAMOM R_h and our newly derived R_h are largely consistent in the mid-latitudes and boreal regions, they actually have opposite seasonality in the dry tropics, where soil moisture limitations largely drive the seasonal cycle of R_h and the CARDAMOM R_h is unrealistic.

We emphasize that this does not mean CARDAMOM CUE cannot be used in our top-down method. As we have previously discussed in the manuscript, the allocation fractions that influence the CUE are particularly well-constrained. Indeed, the spatial variability in CARDAMOM is consistent with a recent meta-analysis compiling carbon use efficiency across 188 sites that has been submitted to Biogeosciences Discussions since the time we originally submitted our manuscript (Tang et al, in review). We will include this reference in the revised manuscript. We have also since updated CARDAMOM to account for the effects of soil water limitations on heterotrophic respiration, with little qualitative effect on the resulting global vegetation CUE maps, but have not included this set of maps in the current work because they have not been published elsewhere and the validity of changes to CARDAMOM is beyond the scope of this research.

Reference:

Tang et al (2019): Global variability of carbon use efficiency in terrestrial ecosystems. *Biogeosciences Discussions*, in review. https://doi.org/10.5194/bg-2019-37

Reviewer: How different would global numbers be if NEP was 0 globally? Would spatial patterns change a lot? It seems like that due to the coarse NEP estimates you cannot achieve reasonable resolutions for Rh.

Response: The 4^{o} by 5^{o} resolution is not much coarser than other atmospheric inversions. For example, the Jena CarboScope has the same resolution, and the Chevallier et al (2010) data mentioned by the reviewer in an earlier comment have a resolution of 3.75^{o} by 2^{o}. As we discuss in Section 4.3, these resolutions may also become finer in the future when the spatial resolution of remotely sensed xCO2 measurements improves with OCO-3 and GeoCarb measurements. Furthermore, as we discussion in Section 4.2, one could also use the approach presented in this paper using higher-resolution regional-scale (instead of global-scale) atmospheric inversions for particular applications where resolution is a significant concern.

In the revised manuscript, we will include a sensitivity analysis for NEP with a global value of zero. As we mentioned above, this changes the spatial distribution of mean R_h less than changing the GPP assumptions, but does create a greater (normalized) root-mean-square-difference with the baseline R_h. This suggests the temporal variability of R_h is affected by the temporal variability of NEP, and that inclusion of an accurate atmospheric inversion can help constrain R_h.

Reviewer: Overall, I cannot follow why we need such a coarse estimate of Rh. On page 14 line 7-8, the authors state that estimates of Rh can be helpful as a validation for ESMs. Using Ecosystem respiration as a validation would be enough to my mind. One evaluates temporal and spatial patterns of Reco to deduce if the representation of Ra and Rh can reproduce these patterns. In the approach presented here one ends up with partitioned Rh, but this heavily depends on the prescribed CUE.

Response: We acknowledge that our method depends on the assumed constancy of the CUE, and have tried to be transparent about this uncertainty, including in the sensitivity analysis in Figure 4, and in the Discussion in Section 4.3, where we write:

> "However, because the assumption of constant CUE employed here has a particularly strong effect on the seasonal cycle of R_h in the wet tropics (Fig. 4b) care should be taken in assessing how this assumption propagates to other studies of top-down R_h variations"

In the revised manuscript, we will expand on this text to clarify that such care should be taken everywhere, not just in the tropics.

In terms of why even coarse top-down R_h data can be useful: global R_h remains a highly uncertain flux, as discussed in detail in Tian et al. (2015) and Bond-Lamberty et al. (2016). Furthermore, in the 4 years since it has been published, the bottom-up global Hashimoto et al. dataset (2015) has been cited 73 times. While not all of these citations focused on R_h, we show that the Hashimoto et al. (2015) approach is sensitive to the overfitting that occurred in that dataset for both R_s and R_h. Furthermore, our manuscript introduces not just an alternative top-down dataset, but also a method that could be applied at a variety of scales and (at smaller scales) resolutions.

Lastly, as to why R_h is preferable to R_{eco} for validating Earth System Models: R_a and R_h can have differential sensitivity to drought and other climatic variations; see, for example, two papers cited below (Sun et al, 2019; Zhang et al, 2019) that have been published only in the last few months making this point, though of course there are others. Thus, while R_{eco} can be used as an indirect constraint on R_h and R_a, knowing only R_{eco} is not enough to unambiguously determine which process representation in the models needs the most improvement, and to test possible alternatives (particularly since, for example, a change to the Ra formulation could affect how much carbon is left for allocation to foliar carbon pools, and thus eventually for R_h). As we recommend in Section 4, when using data derived from our proposed method, alternative CUE assumptions can be easily tested to ensure model evaluations are not affected. We will expand on this discussion point in the revised manuscript.

References

Bond-Lamberty, B., Epron, D., Harden, J., Harmon M.E., Hoffman, F., Kumar, J., McGuire, A.D., and R. Vargas (2016): Estimating heterotrophic respiration at large scales: challenges, approaches, and next steps. *Ecosphere* 7(6):e01380, doi.org/10.1002/ecs2.1380.

Sun, S., Lie, H., and S.X. Chang (2019): Drought differentially affects autotrophic and heterotrophic soil respiration rates and their temperature sensitivity. *Biology and Fertility of Soils.* doi.org/10.1007/s00374-019-01347-w

Tian, H., Lu, C., Yang, J., Banger, K., Huntzinger, D. N., Schwalm, C. R., Michalak, A. 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., Ren, W., Ricciuto, D., Schaefer, K., Shi, X., Tao, B., Wang, W., Wei, Y., Yang, Q., Zhang, B. and Zeng, N. (2015): Global patterns and controls of soil organic carbon dynamics as simulated by multiple terrestrial biosphere models: Current status and future directions, *Glob. Biogeochem*. 10 Cycles, 29, 775–792, doi:10.1002/2014GB005021.

Zhang F., Quan Q., Ma, F., Tian D., Zhou Q., and S. Niu (2019): Differential responses of ecosystem carbon flux components to experimental precipitation gradient in an alpine meadow. *Functional Ecology* 1-12. doi.org/10.1111/1365-2435.13300

Technical and other comments:

Page 7, line 13: Hashimoto et al. (2002), I think this should be 2015.

Response: We will fix this, thank you

Figure 5: In the map there are yellow colors. In the RGB legend, however, yellow cannot be seen. Please correct.

Response: The yellow is between green and red on the bottom axis. We will make it more prominent in the revised version.

References

Chevallier F, Ciais P, Conway TJ et al. (2010) CO 2 surface fluxes at grid point scale estimated from a global 21 year reanalysis of atmospheric measurements. 115.

Zscheischler J, Mahecha MD, Avitabile V et al. (2017) Reviews and syntheses: An empirical spatiotemporal description of the global surface–atmosphere carbon fluxes: opportunities and data limitations. Biogeosciences, 14, 3685-3703.