Major comment responses

Major comment #1

- In the Results, section 4, the spatial analysis is gradually reduced as the paper progresses, from a grid-cell-scale analysis to a more regional-scale aggregation, and finally to being almost absent in the later sections. It would be helpful if the authors could include a short paragraph about this change in spatial scale. If the reason the coarser resolution is used is due to the availability of the data to resolve features at smaller scales, comparability of models, the concept of observational uncertainty, or to facilitate interpretability then this should be stated. This might help the reader understand whether the reduction in scale is due to the nature of the data or some other factor.

Author: Thank you for this comment. We agree, and, as requested, the following paragraph has been added at the end of the Methodology to explain these changes:

Line 193: ***The spatial scale of analysis varies across the results (Section 4), reflecting the nature of the data and the message of each analysis. Fire emissions are presented at grid-cell scale because the GFED benchmarks and model output are available at this resolution and spatial biases are clearly visible. Fire combustion rates are analysed regionally, because observational uncertainty in the benchmark datasets may be substantial at the grid-cell level, and regional aggregation makes the inter-model spread, magnitudes and trends easier to interpret. Land sink estimates are evaluated globally and regionally because top-down inversion products have high spatial uncertainty and regional and global biases and uncertainties are more relevant to carbon budgets. Finally, the JULES sensitivity experiments are presented globally because the aim is to show the impact of historical burned area trends on the contemporary global land sink.***

Major comment #2

- Figure 2 provides important comparison of the global fire-emission estimates, yet it is not explicitly cited or analyzed in the main text. The authors could consider to cite it in the Results section and to provide a brief analysis of its main implications. In particular, it would be important to analyze the differences across the prognostic and diagnostic simulations, the reduction of the fire-emission estimate using the diagnostic burned area, and the large difference between the GFED4.1s and GFED5 reference estimates, with implications for assessing how well the simulations compare to observational constraints.

Author: Good points. We have now referenced the figure and elaborated the analysis (see below). Based on other reviewer comments, we have shifted to focus more on fire combustion rate estimates over fire emissions, which removes the issue of the difference in fire emission magnitude between GFED4.1s and GFED5 (text added below):

> *When diagnosing BA across the five DGVM models (JULES, EDv3, VISIT-UT, ORCHIDEE and OCN), improvements to global fire emissions (fFire) were made. Specifically, we reduced the inter-model spread by 59% (1.8 PgC year−1 to 0.7 PgC year−1, **Fig. 2**) and aligned well with GFED4.1s "natural" mean emissions (where global total burned area is similar, Fig. 1).*
>
> ***On average, fire emission estimates fell by 0.3 PgC year−1 under diagnostic BA compared to prognostic, however this change in magnitude was highly variable across models, since it is driven by differences in prognostic burned areas (ranging from a decrease of 3.2 PgC year−1 on average over 2001–2020 in OCN, to an increase of 1.2 PgC year−1 in EDv3). In JULES, VISIT-UT and ORCHIDEE, global total emissions were broadly similar between the diagnostic and prognostic burned area simulations (all differences of <0.3 PgC year−1 on average over 2001–2020).***
>
> *…* ***Therefore, we focus more on fire combustion rates (fFire per unit BA, Section 4.1.2), which better isolate combustion processes, allowing for assessment of DGVM fire emission performance when BA is constrained. This also enables us to better compare the simulations to the observation constrained GFED emission products, since fire emission trends and magnitudes vary notably between product versions (Fig. 2), whereas trends and magnitudes in fire combustion rates are broadly similar globally and regionally (Table S2, Table S3).***

Minor comment responses

Minor comment #1

- In the Abstract, the manuscript uses the phrase "dynamic global vegetation (DGVM) models." The authors may consider moving the abbreviation to the end of the full term, i.e., "dynamic global vegetation models (DGVMs)," which would also be consistent with the terminology used throughout the rest of the manuscript.

Author: Thanks for noting this, the following update has been made as suggested:

dynamic global vegetation models (DGVMs)