Major comment responses

Major comment #1

- The manuscript places substantial emphasis on the reduction in fire-emission bias and inter-model spread after prescribing satellite-constrained BA. This is useful but also expected, since BA is a direct driver of fire emissions and a major source of disagreement among prognostic fire models. In addition, the GFED emission benchmarks are not fully independent of the prescribed BA, so part of the apparent improvement in "realism" is imposed by construction. In my view, the manuscript contains two more scientifically important results that deserve much greater emphasis. First, once BA is controlled, substantial disagreement remains in fire emissions per unit BA, revealing structural uncertainty in fuel, combustion, mortality, and related processes. Second, the JULES experiments show that historical BA differences can propagate into modern NBP through NEP, not only direct fire emissions. I recommend reorganizing the Results and Discussion around this hierarchy, with the BA to emission improvement presented more concisely as an expected baseline and the other two having more emphasis. This framing would make the manuscript's genuine scientific contribution much clearer and more compelling.

Author: We thank the reviewer for this very constructive suggestion. The framing has been changed to focus on these novel highlight-results of uncertainty in carbon combustion rates and the influence of historic BA on the modern NBP. The following structural and wording changes have been made:

Abstract:

Line 7, added: *By diagnosing BA, we reduced the inter-model standard deviation in global fire emissions by 59% compared to the prognostic BA simulations, and significantly increased estimate agreement with observations spatially and regionally,* ***as expected given that BA is a key driver of fire emissions.***

Line 10, changed *"Additionally, regional simulated fire combustion rate (fire carbon emissions per unit BA) magnitudes and trends did not align with observations under diagnostic BA …"* to increase emphasis*:*

Importantly, even when BA was constrained regional simulated fire combustion rate (fire carbon emissions per unit BA) uncertainty remained high and magnitudes and trends did not align with observations. This reveals uncertainty in fuel loads, combustion completeness, fire-induced

mortality, and vegetation recovery processes, and influenced simulated land sink trends

Line 18, changed: "~~*Using a counterfactual simulation, we show that this uncertainty likely stems from differences in reconstructed pre-satellite era burned area dynamics having legacy impacts on both fire emission and ecosystem productivity estimates.*~~" To increase emphasis on ecosystem productivity:

A counterfactual simulation further indicates that differences in reconstructed pre-satellite era burned area dynamics may propagate onto contemporary land sink estimates through ecosystem productivity changes, not only direct fire emissions.

Line 21, changed: *"improve fire emission realism" to* ***"reduce fire emission biases"***

Methodology:

Section 3.3

Line 162, changed: *… meaning this dataset was used primarily when assessing the* ~~*realism*~~ ***performance*** *of DS_CCI fire emission estimates spatially.*

Results:

Section 4.1.1

Line 199, removed the following trend discussion: *Global emission trends varied between benchmark datasets (weak but significant negative trend in GFED4.1s, non-significant trend in GFED5, Fig. 2) due to differences in burned area trends. Therefore, we focus on trends in fire combustion rates (fFire per unit BA, Section 4.1.2), which better isolate combustion processes.*

Line 213, text and NME table removed (Table 3): As stated, *"emission improvements not fully independent of BA",* so spatial NME improvements may be misleading.

Added paragraph at end of results Section 4.1.1: ***The reductions to fire emission inter-model spread and spatial and regional improvements under diagnostic BA are partly expected, since fire emissions are not independent of BA. Specifically, BA biases and inter-model uncertainty in the prognostic simulations propagate onto fire emission estimates, so constraining BA removes much of this error by construction. Additionally, the BA data used to produce the GFED emission products is not fully independent of the diagnostic BA, as both FireCCI5.1 and GFED use MODIS satellite data in their burned area estimates. 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.***

Discussion:

Section 5.1.1

Line 306, removed: *Although global burned area and fire emission estimates were similar on average under diagnostic burned area, the reduction in spatial and regional biases were substantial.*

Line 309, changed the following show fire emission improvements expected: *"These improvements could be especially relevant to regional carbon budgets, where regional fire disturbance and fire emission biases may have significant impacts on DGVM carbon cycle estimates. Additionally, the inter-model spread in global fire emissions is reduced by more than a factor of 2 under diagnostic burned area compared to prognostic, meaning by implementing diagnostic burned area in DGVMs, we can notably reduce the inter-model uncertainty around the short-term impacts of fire on the carbon cycle"*

Although these improvements were expected since BA is a key driver of fire emissions, they still indicate *that diagnosing burned areas in DGVMs may be especially relevant to regional carbon budgets, where regional fire disturbance and fire emission biases may have significant impacts on DGVM carbon cycle estimates. Additionally,* ***the expected reduction in inter-model spread in global fire emissions when BA is constrained (of more than a factor of 2)*** *means that by implementing diagnostic burned area in DGVMs, we can significantly reduce the inter-model uncertainty around the short term impacts of fire on the carbon cycle.*

Conclusions:

Line 469, changed: *In conclusion, by applying long-term diagnostic burned area to a subset of five DGVMs in TRENDY, this study shows that fire emission estimate* ~~*realism*~~ ***regional performance*** *and constraint …*

Major comment #2

- The combustion-rate analysis is one of the strongest parts of the manuscript because, after BA is constrained, substantial inter-model disagreement remains in fire emissions per unit burned area, including large regional biases and failure to reproduce the observed positive trend. However, the discussion sometimes moves too quickly from this diagnosis to specific causal explanations involving fuel loads, combustion completeness, fire intensity, mortality, and SOC combustion. These are plausible mechanisms, but the current analysis does not isolate their individual contributions. For example, ORCHIDEE explicitly represents fire intensity but still produces a Boreal Asia trend substantially weaker than GFED, suggesting that the bias cannot be attributed simply to the absence of fire-intensity representation. I therefore

recommend tempering the causal language and presenting these processes as plausible sources of disagreement unless additional process-level or sensitivity analyses are provided.

Author: We agree that this concern is valid, so have tempered the wording around the causal nature of these processes

Abstract:

Line 12, changed: *This indicates poor representation of fire dynamics, ~~specifically around~~* ***potentially linked to*** *the representation of fuel loads and combustion completeness, fire-induced mortality, fire intensity, and vegetation recovery processes.*

Discussion:

Section 5.1.1

Lines 343-355 softened as follows: *In all our participating models below-surface soil organic carbon is unavailable to fires, meaning this significant component of the fuel load in boreal forests is unaccounted for in these models. Additionally, in four of the five participating models (JULES, EDv3, VISIT-UT, OCN), vegetation killed by fire forms a component of direct fire emissions (Table 2). In these models, fire intensity is not calculated and fire-induced mortality is kept as a constant proportion across all forest PFTs and biomes. In EDv3, mortality does vary based on tree height (an observed relationship, Balch et al., 2011), but this does not differ based on forest type or location. Therefore, most of our participating models* ***may*** *underestimate the level of fire-induced mortality in boreal forests, in turn impacting their fire combustion rate estimates. This potential underestimation of fire-induced mortality and lack of SOC combustion* ***may contribute*** *to the underestimation of fire combustion rates (and fire emissions) in boreal regions, where fire intensity and severity is generally high and below-surface SOC may become available to fire.*

Lines 356-370 softened as follows: *In addition, GFED estimates, along with other bottom-up estimates (Jones et al., 2024), and top-down estimates (Zheng et al.,2021), show an increasing global trend in fire combustion rates, attributed to an increase in forest fire intensity and burned area. Specifically, this increase is linked to climate change induced changes to forest fire regimes in the extratropics (Jones et al.,2024; Fan et al., 2022), partly associated with increases in extreme fire events (Cunningham et al., 2024). Under diagnostic burned area, our models failed to replicate the magnitude of this positive trend, especially in Boreal Asia. In fire-prone boreal regions, woody land cover dominates (Flannigan et al., 2009; Helbig et al., 2024).*

Therefore, burned area in these regions is mainly attributed to woody classes, so uncertainties in how burned area is partitioned between land cover classes (present both in our methodology and in burned area products (Chuvieco et al., 2019; Lizundia-Loiola et al., 2020)) have a lower impact on combustion rate estimates in these regions. Also, in the GFED5 product (where burned area land cover partitions are provided), no significant trend in boreal forest burned area is observed in Boreal Asia, while a significant positive trend in boreal forest fire combustion rate is present. ***This suggests that the trend is not driven by burned area partitioning, and instead reflects changes in fuel consumption within boreal forests, potentially associated with variations in fire intensity, combustion completeness, or fuel composition.*** *As stated previously, four of the five participating models (JULES, EDv3, VISIT-UT, OCN) do not estimate fire intensity.* ***In ORCHIDEE, fire intensity (fireline intensity, kWm−1, estimated based on fire rate of spread and fuel loads) is used to estimate scorch height, from which both the level of crown fuel consumption and the level of fire-induced mortality through crown damage and cambium death are calculated. However, ORCHIDEE also produces a much weaker positive trend in fire combustion rate in Boreal Asia compared to the GFED products. A possible explanation for this may be that fire intensity only impacts direct fire emissions through live biomass combustion in ORCHIDEE, so intensity-induced variations in litter or soil carbon consumption are missed (Boby et al., 2010). These results suggest that representing fire intensity and linking it to combustion completeness and fuel consumption may improve simulated combustion rate trends in boreal regions, but the ORCHIDEE result indicates that representing fire intensity alone is unlikely to be sufficient. Specifically, the way in which intensity is linked to emissions may also be important, especially in terms of soil and litter fuel consumption.***

Conclusion:

Line 477 updated: This **may** arise from **a mixture** of the poor representation of fuel loads and fuel composition, combustion completeness, fire-induced tree mortality, and the influence of fire intensity on fire emissions.

Major comment #3

- The 0.8 PgC yr^{-1} difference in modern NBP between the two JULES diagnostic BA simulations is important, but it is a single-model result. Its magnitude depends on JULES-specific representations of post-fire mortality, decomposition, and carbon-pool dynamics, which the manuscript itself acknowledges may exaggerate the NEP response. The result supports a potentially important historical-BA legacy effect in JULES, but not a 0.8 PgC

yr^{-1} uncertainty across DGVMs. This model dependence should be made much more explicit.

Author: We agree and thank you for highlighting this, it was not our intention to suggest that other models would be as sensitive as JULES to historic BA dynamics. To amend this, the following text updates are made:

Abstract:

Line 17, added/changed*: This shows that, **in JULES**, the underlying uncertainty in BA mapping and reconstructed BA trends can add a large source of uncertainty to ~~modern day~~ **contemporary** land sink estimates, **but the magnitude of this uncertainty may vary substantially between DGVMs**.*

Conclusions:

Line 483, added/changed*: "… However, land sink estimates varied substantially between the diagnostic burned area simulations (differing by 0.8 PgC year-1 on average over 2001-2020), ~~probably due to~~ **in part due to** differences in burned area dynamics across the 20th century between the two reconstructions having legacy effects on both fire emission and NEP estimates (evidenced by the use of a counterfactual simulation). **This is a single model result which relies on JULES specific representations of post-fire mortality, decomposition, and carbon-pool dynamics, meaning the true magnitude of BA legacy effects on NEP is highly uncertain.***

***Nonetheless**, this finding still indicates that a potentially high level of uncertainty around modern day land sink estimates in DGVMs ~~is~~ **may be** attributable to pre-satellite era burned area dynamics, highlighting the need for further investigation into the effects that burned area dynamics have on the carbon cycle and the uncertainty around past burned area trends."*

Major comment #4

- The manuscript interprets the 0.8 PgC yr^{-1} difference between the DSCCI and DSGFED JULES simulations primarily as a legacy effect of differing pre-satellite BA dynamics. Historical BA dynamics clearly matter, as demonstrated by the counterfactual experiment. However, this experiment shows that one historical BA trajectory can produce a legacy effect; it does not demonstrate that differences in historical temporal trends explain the full DSCCI–DSGFED NBP difference. The two reconstructions also differ in mean BA, spatial distribution, and treatment of land-use-associated fires. All of these differences could contribute to the NBP divergence. The attribution should therefore be softened, or additional sensitivity experiments should separate temporal-history effects from differences in BA magnitude and spatial pattern.

Author: We agree that more may be needed to isolate the impact of the historic BA in the DSGFED simulation. Illustrating the attribution via an additional counterfactual experiment with GFED5 is difficult since BA dynamics differ across 1901-1980 between the GFED5 and FireCCI5.1 reconstructions. Therefore, we have chosen to soften the language on attribution, as follows:

Discussion, *Section 5.1.3:*

Line 404, changed: *These differences in past BA dynamics between the reconstructions ~~are the most plausible explanation for~~* ***are likely a major contributor*** *to the lower modern day land sink magnitude in this simulation.*

Added: ***However, the two reconstructions also differ in, total BA, BA spatial distribution, and treatment of land-use associated fires, which may all also contribute to this difference in the land sink.***

Section 5.2:

Line 432, changed: *Specifically, the unrealistically low JULES NBP estimate under the GFED5 reconstruction, ~~probably driven by~~* ***influenced by*** *pre-satellite burned area dynamics, suggests that the land sink improvements observed under diagnostic BA may not be robust to differing burned area reconstructions and mappings*

Conclusions:

Line 483, added/changed*: "… However, land sink estimates varied substantially between the diagnostic burned area simulations (differing by 0.8 PgC year-1 on average over 2001-2020), ~~probably due to~~* ***in part due to*** *differences in burned area dynamics across the 20th century between the two reconstructions having legacy effects on both fire emission and NEP estimates (evidenced by the use of a counterfactual simulation).*

Minor comment responses

Minor comment #1

- Lines 239: "Diagnosing BA reduced the multi-model mean land sink by 0.3±0.4 PgC year^{-1} over 2001–2020 compared to the prognostic simulations, despite the concurrent reduction in fire emissions." The reason for this seemingly counterintuitive response is not clearly explained. Since lower fire emissions alone should increase NBP, the reduced land sink must be driven by changes in NEP that more than offset this effect. Please clarify and, if possible, quantify the contributions from NPP and HR.

Author: Agreed, the following text has been added:

This was due to a higher magnitude reduction in NEP of 0.9 ± 1.0 PgC year−1, associated with a reduction in NPP of 0.5 ± 1.4 PgC year−1 and an increase in HR of 0.4 ± 1.6 PgC year−1.

Minor comment #2

- Line 270 “This NBP estimate in the simulation with GFED5 diagnostic BA fell below all GCB inversion and land-use dataset change combinations (minimum of 1.2 PgC year−1).” Please also report the mean of the GCB inversion estimates, not only the minimum. From Fig. 6, the PS simulation appears to align more closely with the inversion mean than both diagnostic simulations, which should also be discussed.

Author: We agree, and have amended the text as follows:

…, ***mean of 2.6 PgC year-1***

Added: ***Under prognostic BA, the JULES land sink estimate aligned closely with both the GCB residual and inversion datasets, whereas alignment was worse under diagnostic BA, especially post 2000 where the behaviour differs notably.***

Minor comment #3

- Line 292: “The magnitude of year-to-year differences in NBP between the counterfactual and original simulations were aligned much more closely with year to year differences in forest BA rather than total BA” This wording is unclear because NBP and BA have different units. Please clarify whether “aligned” refers to temporal correlation or another quantitative relationship, perhaps referring directly to the R^2 analysis reported immediately afterward.

Author: We agree that this wording is unclear, so we have calculated a new correlation coefficient and updated the text to the following with ‘magnitude’ removed:

Year-to-year differences in NBP between the counterfactual and original simulations were more strongly correlated with year-to-year differences in forest BA (R^2 = 0.94) than with differences in total BA (R^2 = 0.87, Fig. 7).

Minor comment #4

- Line 298: “Therefore, the counterfactual NBP was 2.7 PgC $year^{-1}$ over this period which is similar to the PS average of 2.6 PgC $year^{-1}$ and lies well within 1 S.D of the mean of the GCB inversions and close to the GCB residual sink.” The implication of this comparison is unclear. Is the agreement with the GCB residual intended to suggest that the counterfactual represents a more

realistic historical BA trajectory? If not, please clarify what conclusion should be drawn from the closer agreement.

Author: We agree the implication is unclear, thank you for noting this. The conclusion we want drawn from this is not that the counterfactual represents a more realistic historical BA trajectory, but that model sensitivity to historic BA trajectories may lead to poor agreement with contemporary land sink observations, meaning this effect needs to be better handled by models.

Line 402, Added in discussion*: This suggests that without this increase in burned area the FireCCI5.1 diagnostic BA simulation modern day land sink estimate would be similar to the prognostic simulation average, and align well with the GCB inversions and GCB residual,* ***highlighting the sensitivity of JULES NBP estimates to burned area dynamics.***

Minor comment #5

- Line 350: “However, vegetation killed by fire does not form a component of fire emissions in ORCHIDEE, so variations in fire-induced mortality do not impact fire emission estimates.” I do not think vegetation killed by fire should necessarily be considered part of direct fire emissions. GFED also primarily represents carbon emitted during combustion rather than subsequent mortality losses.

Author: Thanks for noting this, we agree that vegetation killed by fire should not necessarily be part of direct emissions. This sentence has been removed, and further changes relating to this comment are the same as those completed for major comment #2

Minor comment #6

- Lines 361: "Although uncertainties exist around the partitioning of burned area across different land cover classes … suggesting the partitioning of burned area across different land cover classes is less uncertain in these regions" is difficult to follow. Please rephrase to clarify why dominance of woody land cover in boreal regions reduces uncertainty in the partitioning of BA among land-cover classes.

Author: We agree this wording is convoluted, the following changes were made:

In fire-prone boreal regions, woody land cover dominates (Flannigan et al., 2009; Helbig et al., 2024). Therefore, burned area in these regions is mainly attributed to woody classes, so uncertainties in how burned area is partitioned between land cover classes (present both in our methodology and in burned area products (Chuvieco et al., 2019; Lizundia-Loiola et al., 2020)) have a lower impact on combustion rate estimates in these regions

Minor comment #7

- Line 367 “This further emphasizes the need for models to produce estimates of fire intensity and relate these to emissions in order to better replicate observed magnitudes and trends in fire combustion rates.” Does ORCHIDEE, which explicitly calculates fire intensity, actually reproduce the Boreal Asia trend better than the other models? In Lines 224–226, ORCHIDEE is grouped with VISIT-UT and OCN as producing significant but substantially weaker trends than GFED. If including fire intensity does not clearly improve the trend in ORCHIDEE, the argument that missing fire intensity is a primary explanation for the model bias should be qualified.

Author: Thanks for this comment, good points. Changes relating to this comment are the same as those completed for major comment #3, specifically the following text has been added:

> … ***In ORCHIDEE, fire intensity (fireline intensity, kWm-1, estimated based on fire rate of spread and fuel loads) is used to estimate scorch height, from which both the level of crown fuel consumption and the level of fire-induced mortality through crown damage and cambium death are calculated. However, ORCHIDEE also produces a much weaker positive trend in fire combustion rate in Boreal Asia compared to the GFED products. A possible explanation for this may be that fire intensity only impacts direct fire emissions through live biomass combustion in ORCHIDEE, so intensity-induced variations in litter or soil carbon consumption are missed (Boby et al., 2010).***
>
> ***These results suggest that representing fire intensity and linking it to combustion completeness and fuel consumption may improve simulated combustion rate trends in boreal regions, but the ORCHIDEE result indicates that representing fire intensity alone is unlikely to be sufficient. Specifically, the way in which intensity is linked to emissions may also be important, especially in terms of soil and litter fuel consumption.***