Reviewer #1:

Reza et al. improved the fire module in the SEIB-DGVM model and used the model to project Siberia wildfire dynamics, and emissions under various climate scenarios. I appreciate the valuable model development and analysis, and it provide us important implications on how future climate change affects Siberia boreal forest. The paper is informative and has a lot of important results to discuss, but it is not well-written in terms of smoothness and readability. There are confusions on methodology and data usage. Most importantly, this manuscript over-emphasizes future RCP projections. However, historical validation is not clear and convincing. A more reasonable approach is to first thoroughly demonstrate the model performance during the historical period by comparing it with observations. Then discuss future projections and caveats. Below are my major comments.

> **Response:**
>
> Thank you for your valuable feedback on our manuscript. We appreciate your understanding of importance in Siberia's boreal forests under future climate scenarios. We understand your concerns regarding the clarity and readability of the manuscript, as well as the need for a stronger emphasis on historical validation. We will revise the manuscript to improve its smoothness and address any confusions regarding methodology and data usage. Additionally, we will prioritize demonstrating the model's performance during the historical period before discussing future projections. Once again, thank you for your constructive comments and we have replied to your comments and also adjusted our manuscript accordingly.

1. There is no quantitative evidence of the model's performance. During the historical period, how did the model simulate burned area, and fire emissions, compared with existing datasets (e.g., GFED) at gridcell level? Plot scatter plots of modeled versus observed variables (burned area, emissions, biomass) will be super helpful. Showing R2 and RMSE of gridcell level comparison in the abstract is highly encouraged. As it is stated in the abstract the motivation for integrating SPITFIRE into DGVM is to improve the accuracy of fire modeling. Therefore, it is also important to show how much improvement has been achieved by SPITFIRE compared with DGVM's default fire model.

> **Response:**
>
> Thank you for your valuable feedback on our manuscript. We have validated the model using some benchmark data at longitude average grid cell level. GFED4s: burned area 1996-2016 (Figure S14.b in supplement). GFED4s: burned fraction 1997-2016 (Figure S14.a in supplement), dry matter 1997-2016 (Figure S19 in supplement), CO_2 emissions 1997-2016 (Figure 9.a: L445). GBEI: CO_2 emissions 2001-2020 (Figure 9.b: L445). ESA Biomass CCI: Aboveground biomass 2010, 2017-2018 (Figure S16 in supplement).
>
> We have added the R^2 and RMSE value on the abstract.

In the model, the historical simulation starts from 1850-2005, and the future simulation starts from 2006-2100. We extracted the relevant years from both simulation periods when needed to compare with the benchmark datasets.

We updated the comparison of AGB model and ESA Biomass CCI data: previous figures are based on the longitude average, while the new figures are based on the latitude average. We added double error marks to identify the model uncertainty and the discrepancy with the benchmark datasets. We will replace all of the previous plot with the new plot below.

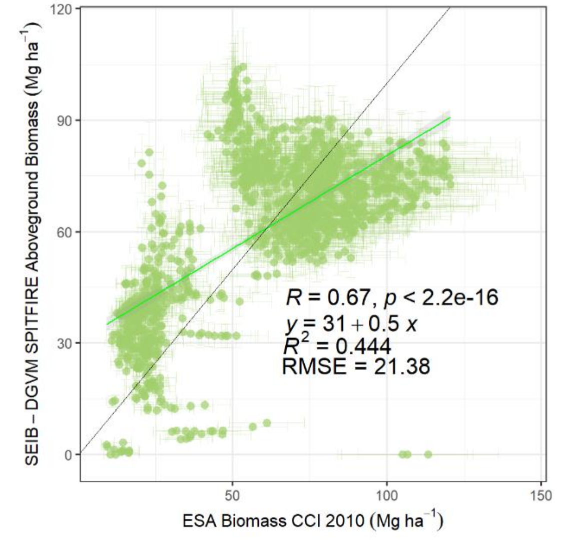
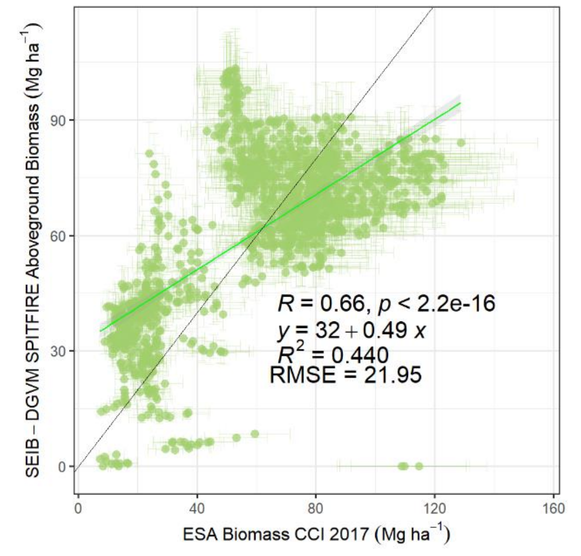
	

Figure S16.a (adjusted) AGB 2010	Figure S16.b (adjusted) AGB 2017

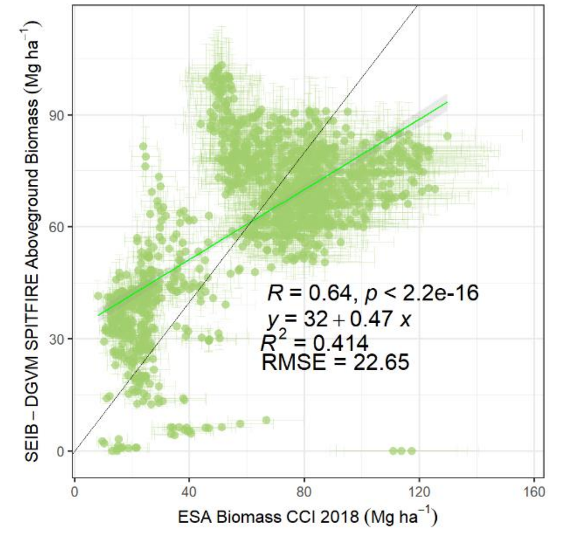
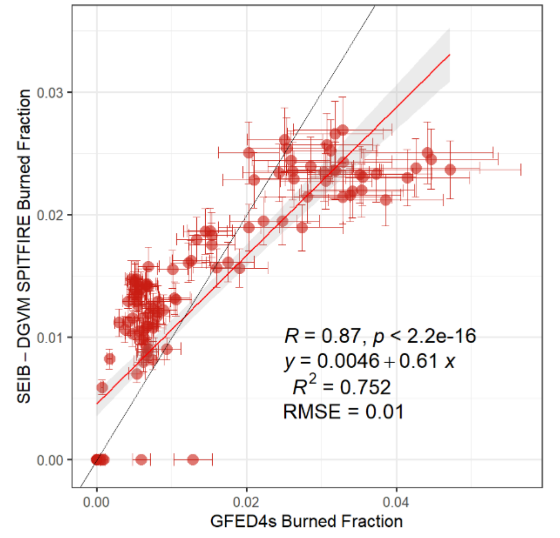
	

Figure S16.c (adjusted) AGB 2018	Figure S14.a Burned Fraction (1997-2016)

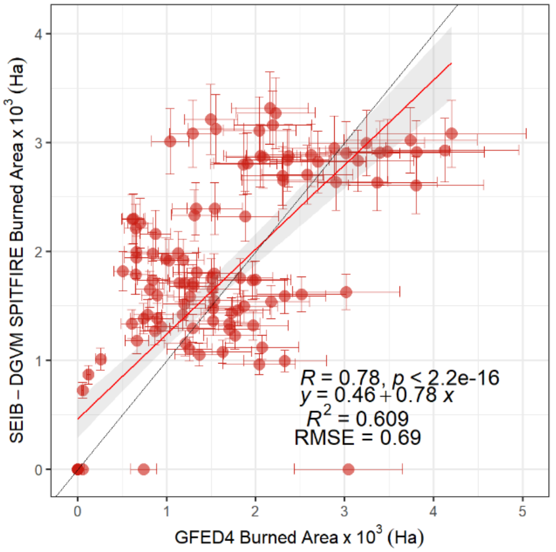


Figure S14.b Burned Area (1996-2016)

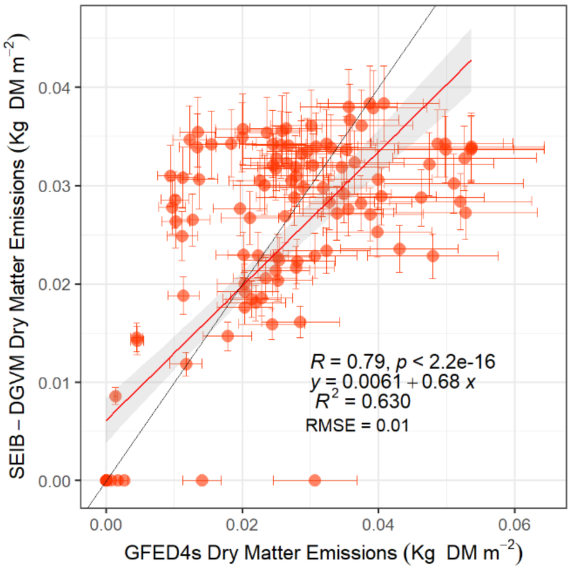


Figure S19 Dry Matter (1997-2016)

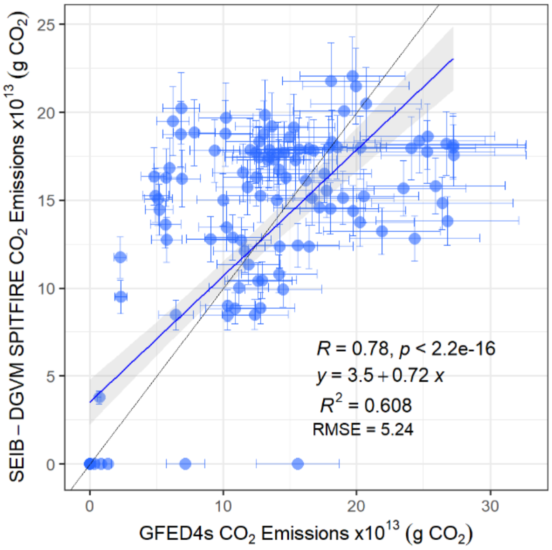


Figure 9.a GFED4s CO$_2$ (1997-2016)

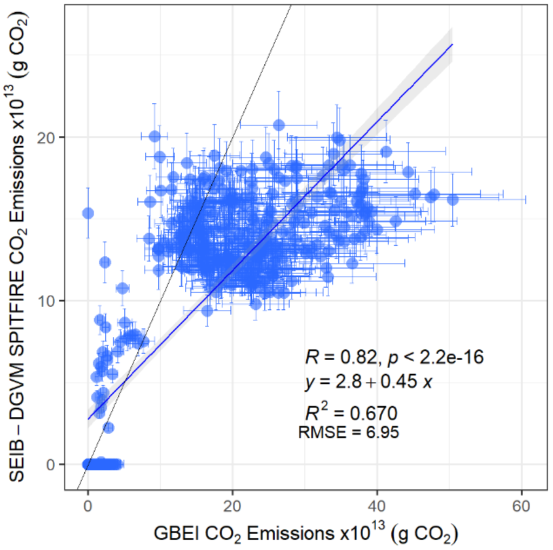
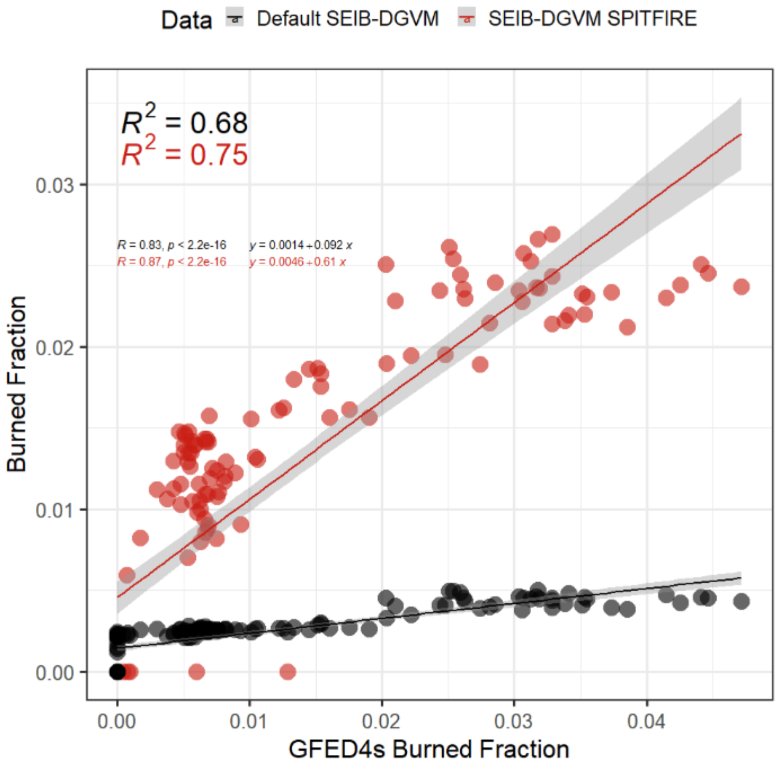
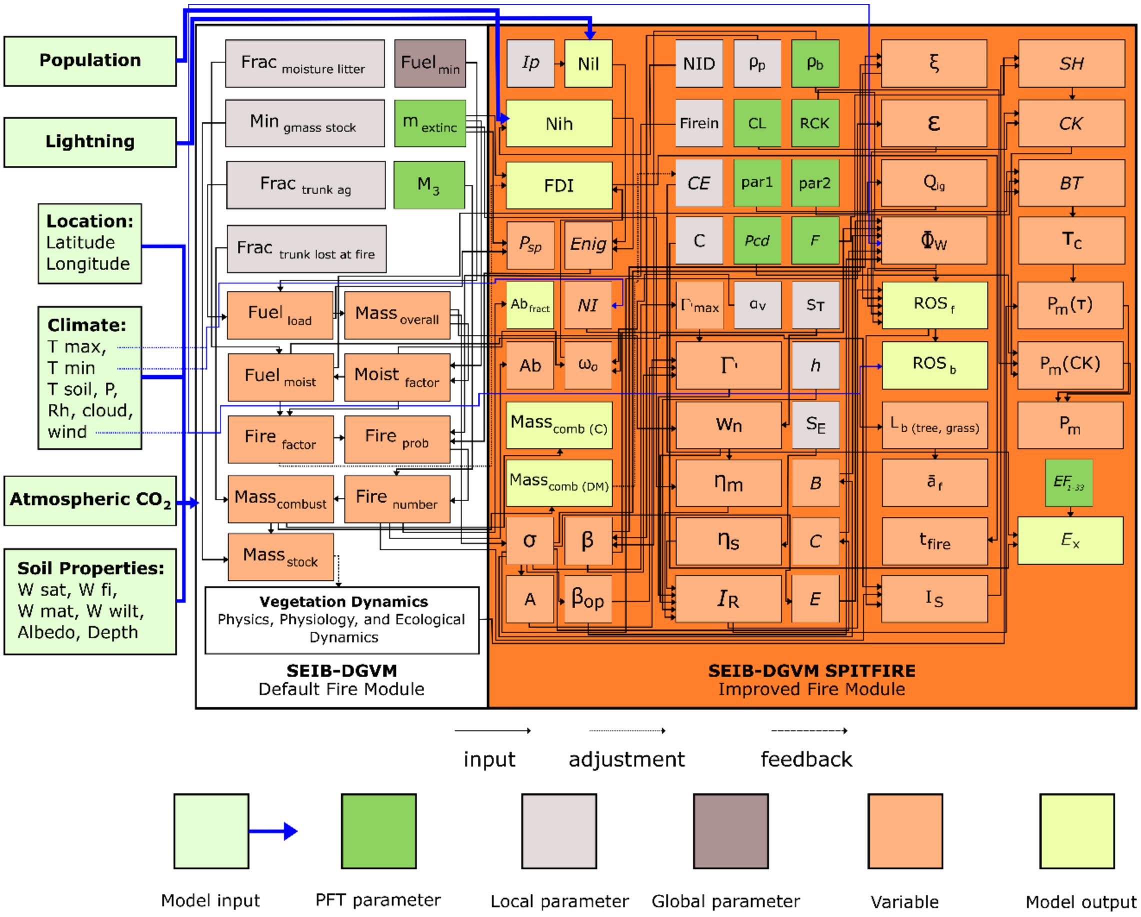


Figure 9.b GBEI CO$_2$ (2001-2020)

we added a comparison of the main variables generated in the default model and the improved model in sections 3.1 to 3.4.

Related to the fire estimation improvements of the improved model with the default model, we have added this explanation in sub section 3.2. and figure S5. in the supplement.





We replaced the previous Figure 2 with new Figure 2 above to describe in more detail about the difference and the integration between the default SEIB-DGVM and the SEIB-DGVM SPITFIRE.

2. Improvement from annual time step estimate to monthly is an important change. However, most of the fires last less than one month. How does the monthly time step fire module resolve a process that lasts much shorter than a month?

Response:

Estimating fire occurrence at a daily timestep would likely decrease the model accuracy (there are just too many factors to expect the model fit perfectly with reality). However, rather than knowing the exact day of the fire, knowing in which month(s) to expect higher fire activity would help to adjust prevention measures. Furthermore, we have calibrated our variable with the benchmark datasets (related to the fire: we used GFED4s and GFED4 datasets) which have the daily time-step data, thus the monthly output is an accumulation of the daily process.

3. Human ignition is considered in E term (equation 2), how about human suppression on fire spread?

Response:

Human suppression on the fire ignition is considered in the equation 8 (in the supplement). The human ignition factor (Nih) value will increase as the population increases, but if at a certain population density, the ignition factor will decrease steadily to a value of 0, indicating that there is human suppression on fire. Equation (8) has a maximum value at a population density of 16 km^{-2}.

PD is the population density (individuals km^{-2}), and aND (ignitions individual^{-1} d^{-1}) is a parameter expressing the propensity of people to produce ignition events. aND is and user-definable parameter, with a scale of 0.0 – 1.0, and in this study we adjust the value to 0.7 or 70% (total of human and unknown caused fire), as the recent research in eastern Siberia Xu *et al.*, (2022), shown that fires over Yakutia, 31.4 ± 6.8% caused by lightning ignitions, 51.0 ± 6.9% caused by anthropogenic ignitions, and the last 14.4% unknown cause.

Human suppression on fire is considered/ calculated in the model at the ignition stage only. While at the fire spread, it is only influenced by fuel availability, moisture, and climatic factors (wind). However, the model performs calculations on each individual (0.5 x 0.5) grid cell. Therefore, if observed more broadly, humans have suppression of ignition and fire spread.

4. Methodology, between 2.2 model description and 2.3 model application, there seems missing a section about model calibration. How was the SPITFIRE module calibrated for burned areas or emissions? How was the DGVM model calibrated to capture observed AGB?

Response:

Before validating with benchmarks and observational data, we calibrated all major output variables (Burned Fraction, Burned Area, Dry Matter, Aboveground Biomass, CO_2 Emissions, $PM_{2.5}$ Emissions, and some forest ecology variables). The emission calibration process is only carried out on CO_2 and $PM_{2.5}$ emissions as representative major emissions.

We have added the explanation of the model calibration process in 2.3 Model Calibration, the other subchapters will be renumbered accordingly, 2.4. Model application, and so on.

"We calibrate the improved model by using all of the benchmark datasets (Table 3). The calibration process is done sequentially, from burned fraction, burned area, dry matter, aboveground biomass, burned biomass emissions, and the forest ecology variables (Figure 3). The process is sequential because one variable is used for the calculation of another variable (such as burned fraction and burned area affecting aboveground biomass, forest structure, dry matter, and emissions). One calibration process is performed with multiple iterations until the output variable has similar numerical values and spatial distribution to the benchmark data, and the process is repeated for other variables once the previous variable has been calibrated.” (L211)

We have adjusted the “Figure 3. Workflow of improving the SEIB-DGVM fire module” by adding a model calibration scheme with benchmark data. In the next manuscript, we will replace Figure 3 with the latest one.

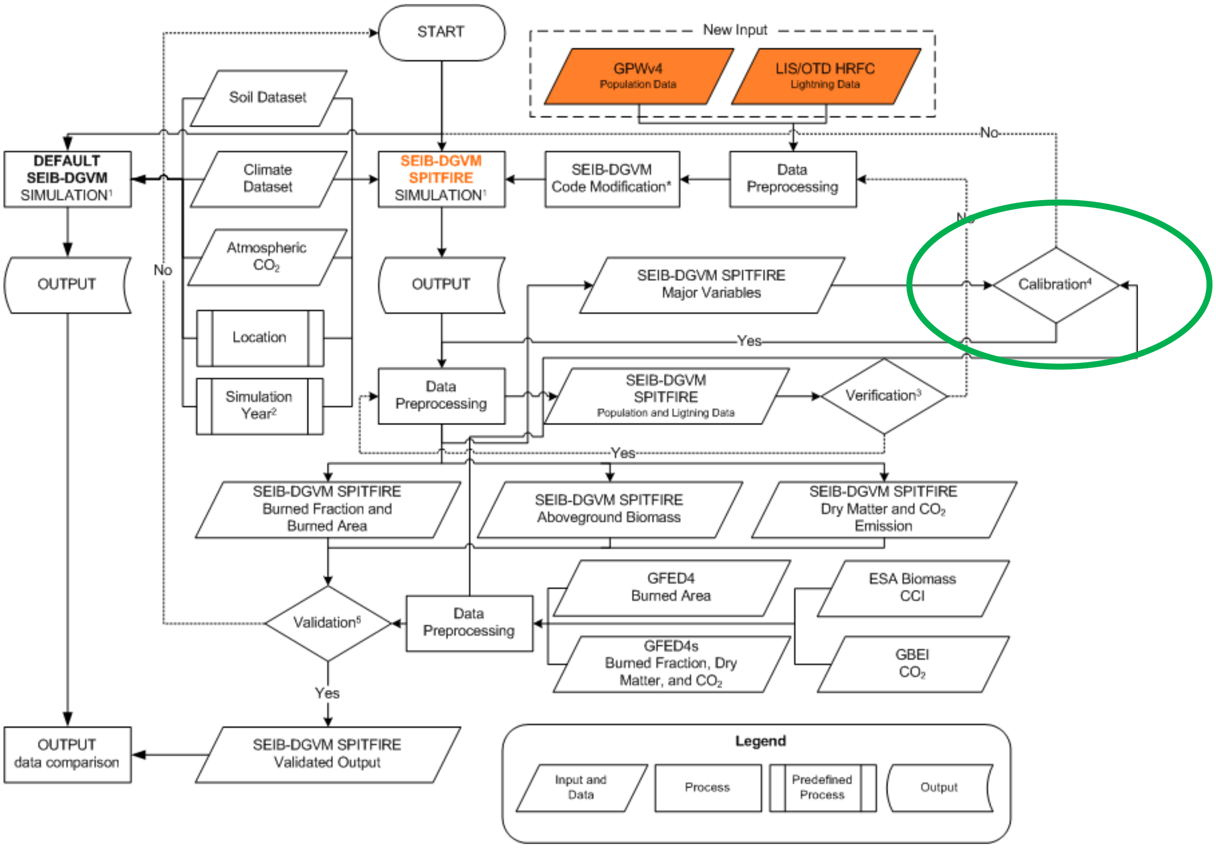


Figure 3. Workflow of improving the SEIB-DGVM fire module.

5. For future projection, are population density changes and lightning flash changes considered under future scenarios?

Response:

In future simulations, the variables of population density and lightning flash remain the same as in historical simulations (constant). This is because the temporal coverage of the datasets are not long enough (LIS/OTD High-Resolution Full Climatology (HRFC) V2.3.2015 (2000-2020) and Gridded Population of the World (GPWv4) (2015) (Table 2).

6. Section 2.4, the model is validated by GFED4 burned area and GFED4s burned fraction. However, GFED4s and GFED4 are different because GFED4S include smaller fires. How to reconcile GFED4s and GFED4 during validation?

Response:

Yes, GFED4s burned fraction and GFED4 burned area are different because GFED4s includes small fires. We have mentioned this in section 3.1.1 Fire products (L 305: revised manuscript). The validation process is carried out separately between variables (burned fraction and burned area), so there is no reconciliation process needed between benchmark datasets (GFED4s and GFED4).

Based on the validation results, in the comparison of distribution patterns, the results of the burned fraction variable comparison model with GFED4s have a higher value than the burned area variable model with GFED4. Because GFED4s data has a slightly wider distribution pattern because it including small fires.

7. Section 3.1-3.4. Historical validation is an important component of this study. I would like to suggest before discussing RCP results thoroughly compare the simulated burned area, emission, and biomass with observations. Discuss model performance and biases at gridcell level, and the regional level.

Response:

We have validated the model in section 3.7. We will move this section to the first section.

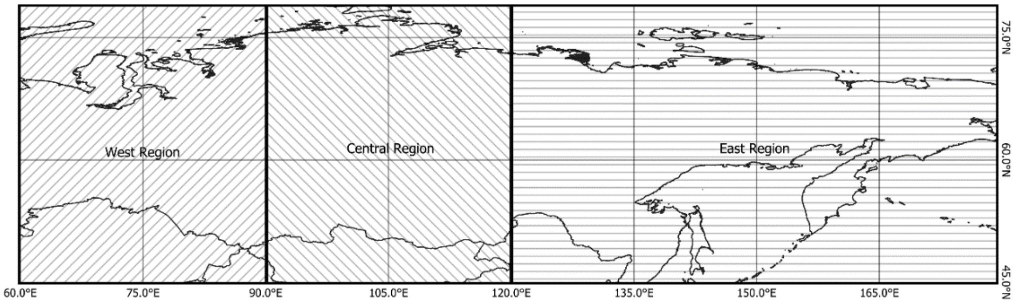
Regarding the historical validation, there is a difference in that the historical time in the model starts from 1850-2005 (according to the climate input data) and the 2006-2100 data is included in the future simulation, while the benchmark data varies from 1996-2020. (This topic has been mentioned in the 1st comment response). Thus, we extracted the relevant years from both simulation periods when needed to compare with the benchmark datasets.

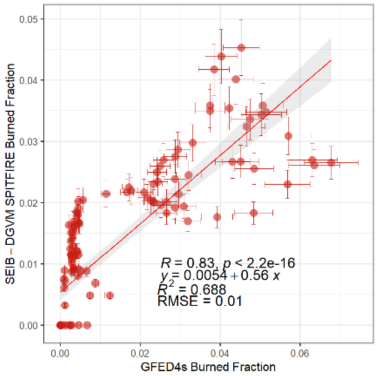
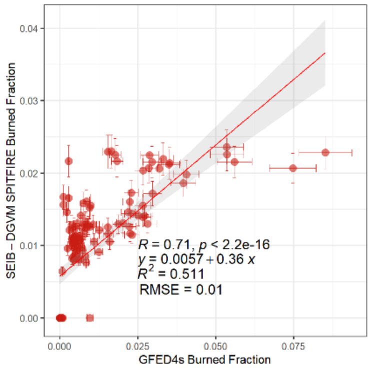
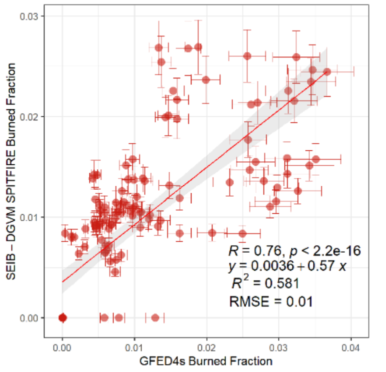
Instead of grid level validation, we used latitude average comparison to determine the average value, pattern and dynamics of variables at longitude point of view. Furthermore, in accordance of previous studies, model projections are not validated spatially at grid cell level but they validated the model output with observational data in a numerical comparison (specific area, temporal average, and only few variables). e.g. Fig. 3 Verification of the simulated length of fire season using LPJ-DGVM against observations in the sample regions (Thonicke et al., 2001), Fig. 4 Fire return intervals for the period 1987–96 derived from the national fire statistics of forest services and simulated for the same period by the LPJ-DGVM (Thonicke et al., 2001), Fig. 6. Observed (MODIS) versus simulated fire season lengths for biomes (Thonicke et al., 2010), and Fig. 8. (a) Comparison of observed (Ni, 2004) and simulated net primary production in northern China. (b) Comparison of observed (Sukhinin et al., 2004) and simulated area burnt for 1997–2002 (Thonicke et al., 2010).

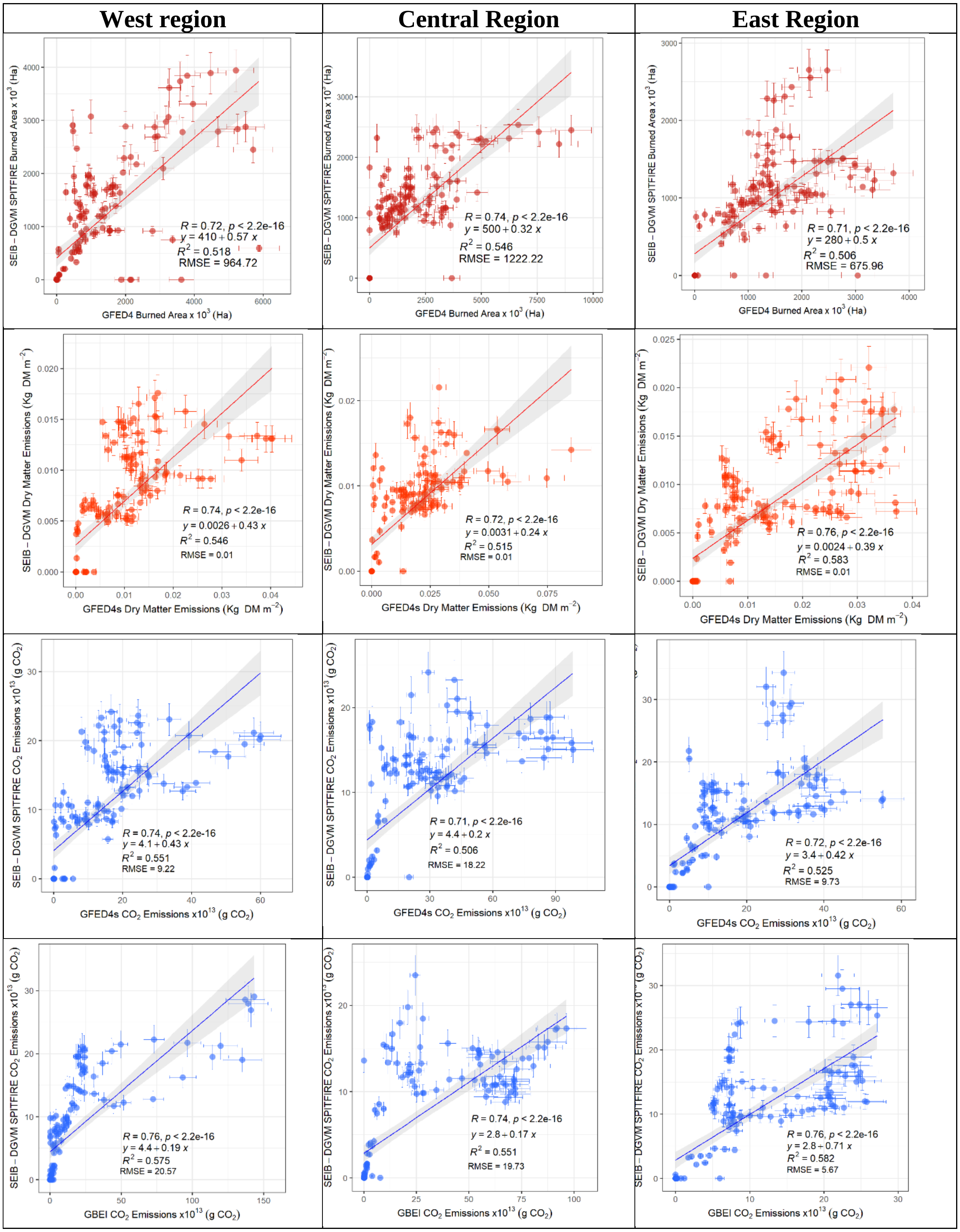
In this study we compare both (numerical comparison and spatial comparison: average latitude at Siberian level and regional level) for all major variables. Therefore, our validation process is better because it covers all major variables and uses two types of validation. (section 4.1. Paragraph 3).

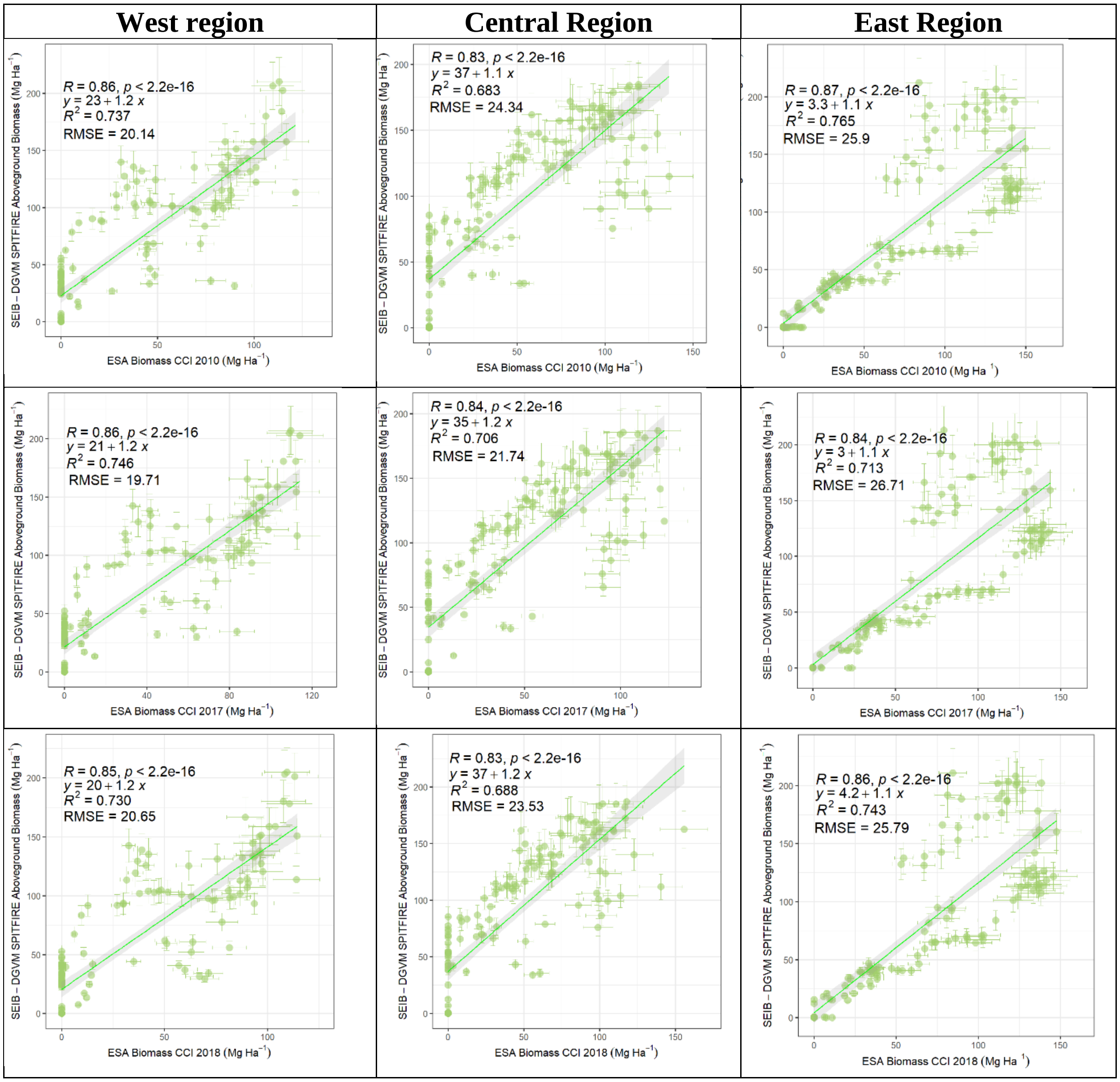
Below are the validation results for all major variables at the regional level. We divide Siberia into three areas: west region (60°-90°E and 45°-80°N), central region (90°-120°E and 45°-80°N), and east region (120°-180°E and 45°-80°N).

We added this explanation in the Section 2.5. (L258) and all of the results section related to each variables.



West region	Central Region	East Region
		

West region
Central Region
East Region
SEIB – DGVM SPITFIRE Burned Area x 10^3 (Ha)
GFED4 Burned Area x 10^3 (Ha)
R = 0.72, p < 2.2e-16
y = 410 + 0.57 x
R^2 = 0.518
RMSE = 964.72
R = 0.74, p < 2.2e-16
y = 500 + 0.32 x
R^2 = 0.546
RMSE = 1222.22
R = 0.71, p < 2.2e-16
y = 280 + 0.5 x
R^2 = 0.506
RMSE = 675.96
SEIB – DGVM Dry Matter Emissions (Kg DM m^-2)
GFED4s Dry Matter Emissions (Kg DM m^-2)
R = 0.74, p < 2.2e-16
y = 0.0026 + 0.43 x
R^2 = 0.546
RMSE = 0.01
R = 0.72, p < 2.2e-16
y = 0.0031 + 0.24 x
R^2 = 0.515
RMSE = 0.01
R = 0.76, p < 2.2e-16
y = 0.0024 + 0.39 x
R^2 = 0.583
RMSE = 0.01
SEIB – DGVM SPITFIRE CO2 Emissions x10^13 (g CO2)
GFED4s CO2 Emissions x10^13 (g CO2)
R = 0.74, p < 2.2e-16
y = 4.1 + 0.43 x
R^2 = 0.551
RMSE = 9.22
R = 0.71, p < 2.2e-16
y = 4.4 + 0.2 x
R^2 = 0.506
RMSE = 18.22
R = 0.72, p < 2.2e-16
y = 3.4 + 0.42 x
R^2 = 0.525
RMSE = 9.73
GBEI CO2 Emissions x10^13 (g CO2)
R = 0.76, p < 2.2e-16
y = 4.4 + 0.19 x
R^2 = 0.575
RMSE = 20.57
R = 0.74, p < 2.2e-16
y = 2.8 + 0.17 x
R^2 = 0.551
RMSE = 19.73
R = 0.76, p < 2.2e-16
y = 2.8 + 0.71 x
R^2 = 0.582
RMSE = 5.67



8. Section 3.5 Fire-off simulation is not mentioned in the methodology section. When was the fire turned off? Does the model run an extended spinup with fire-off? One would expect much larger aboveground biomass when the fire is turned off.

Response:

We have explained the fire-off simulation in the section 2.4. Model application (L244). We ran the fire-off simulation separately from the fire-on simulation (starting from spinup until future simulation using the same protocols: model settings and climate input dataset). The aboveground biomass was indeed larger when the fire was off.

We explained the comparison result in the section 3.5 Forest ecological variables under fire-on and fire-off simulation.

9. Figure 7 showing spatial coverage is confusing because GFED4s provide burned fraction, which should be directly compared with SEIB-DGVM at gridcell level. Figure 8, a and b are duplicated, both showed long-term monthly average dry matter emission. Figure 9, why dry matter emission is perfectly simulated by SEIB-DGVM (shown in Figure 8), but the performance of CO2 emission is much worse. It seems inconsistent. What does it look like, if plot a spatial average comparison of DM emissions between SEIB-DGVM with GFED4s?

Thank you for the the comments. We will address each point separately:

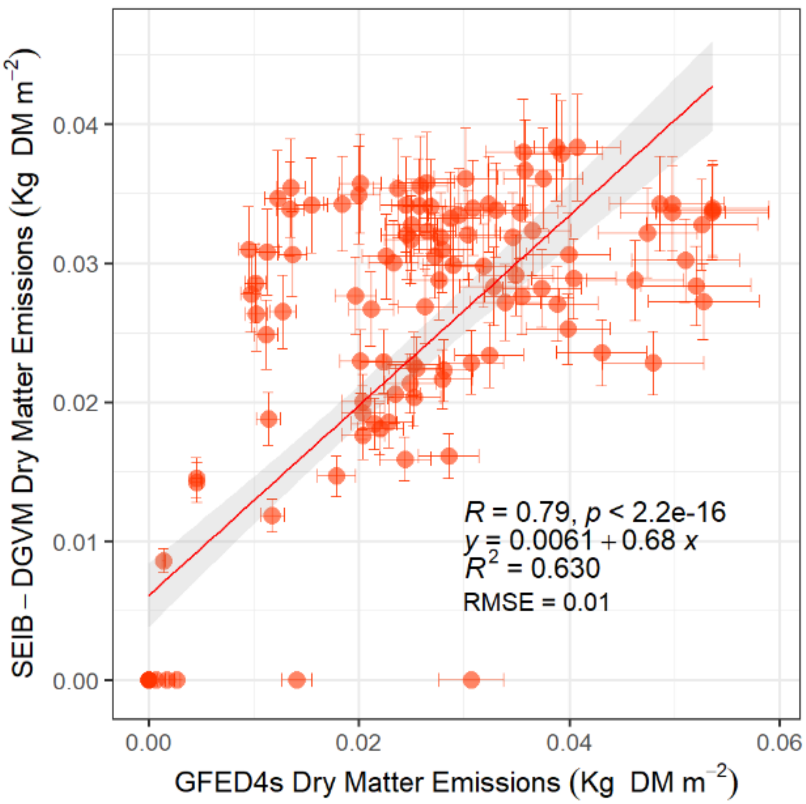
a. Our purpose of showing Figure 7 is only to determine the agreement of burned fraction model variables with GFED4s data (ignoring the value of each grid), we did as in Fig. (b) Comparison of observed (Sukhinin et al., 2004) and simulated burnt area for 1997-2002 (Thonicke et al., 2010).
   We have also compared the burned fraction variable (and all variables) with benchmark data at the grid level (mean latitude: Figure S14.a and updated plots are in the response to first comment).
b. Yes, figure 8 a. and b. show the same data, we will remove figure 8.b in the adjusted version of the manuscript.

   Figure 8 is a comparison of annual average extracted values of modeled dry matter emissions with GFED4s data. Numerically and on a large scale (entire study site: Siberia) and long-time horizon (monthly average), the model is able to produce values that are very close to the benchmark data. The model is able to simulate well because each variable has been calibrated with benchmark data with several iterations to produce good data.

   In contrast, Figure 9 is a spatial comparison, where the spatial distribution pattern of emissions (all fire-related variables), has a similar distribution pattern because it comes from fire variables (burned fraction and burned area) which are calculated with specific equations to produce emission variables.

   The above is also explained as a limitation of the model in section 4.1 (L 574), that the model is able to simulate numerically well, but the distribution pattern is highly dependent on the initial variable (fire), so adjusting ignition and related factors can produce a better distribution pattern (closer to benchmark data), which will affect other variables (dry matter emissions, agb, forest ecology, and all burning biomass emissions).

   As to the last point, we have compared the modeled dry matter emission variables with GFED4s (Figure S19 in the supplement, and the new plot is in the response to the first comment).



SEIB – DGVM Dry Matter Emissions (Kg DM m^{-2})
0.04
0.03
0.02
0.01
0.00
0.00
0.02
0.04
0.06
GFED4s Dry Matter Emissions (Kg DM m^{-2})
$R = 0.79$, $p < 2.2e\text{-}16$
$y = 0.0061 + 0.68\,x$
$R^2 = 0.630$
RMSE = 0.01