Response to the comments of Anonymous Referee #2 (RC2)

We thank the reviewers and the editor for their suggestions and comments on the manuscript. We have considered all their comments and hope that the revised draft properly addresses their suggestions. Please find our point-by-point replies below (colored in blue). A revised manuscript with tracked changes will be uploaded. Line numbers in our texts refer to the no-markup version.

-This paper, submitted to the journal, Geoscientific Model Development (GMD), by Q. Li, Serbin, Lamour, Davidson, Ely, and Rogers, entitled "Implementation and evaluation of the unified stomatal optimization approach in the Functionally Assembled Terrestrial Ecosystem Simulator (FATES)", is a well-written paper that could be accepted after mild revisions. The topic is important for understanding climate and land-surface processes better, and the modeling exhibited here is first rate. I detail my comments below.

I am impressed by the fact that the authors have started the FATES model runs with real-world forest inventory data, as stated on Line 115.

We appreciate the reviewer taking the time to review and providing valuable feedbacks to this study. We are very excited to contribute to the development and evaluation of FATES.

-Line 136: "we set the precipitation to 1.47^10-5 mm/s" = 1mm/day? So it is always raining? Is this consistent with the humidity or VPD values of the model experiments? Is it consistent with the PAR values of the model experiments?

We have varied precipitation but did not find it was related to the change of humidity in the model. It is the specific humidity that determines the humidity and influences VPD. In our sensitivity runs with synthetic climate forcing, precipitation does not need to be consistent with humidity or PAR. The value 1.47×10^{-5} mm/s (i.e., 1.3 mm/day) was calculated as the annual averaged precipitation and used for the standard condition when we explored model responses to other climate variables. Because our sensitivity runs were only conducted on a short time scale, the precipitation was not able to influence soil moisture. We changed the unit of this value to "mm/day" at Line 165 to make it more concise.

-Fig 5.: There is not much difference in A_net or g_sw for the 3 days for either BWB_mean or MED_mean, even though the average-peak PAR increases from 700 to 1200 to 1500 mol/m2/s for the 3 successive days. This approximate independence of g_sw on PAR is what can be expected from Fig. 1a, for PAR > 500 mol/m2/s. But from Fig 1a, it might be expected that MED−default_mean and BWB_mean would differ by a factor of 2 in Fig. 5a. Is this Figure 5 actually for MED-B_mean instead of being for MED-default_mean? If it is, then the lack of difference between the modeled values for A_net or g_sw would make more sense.

We appreciate the reviewer's idea of linking the evaluation results with previous sensitivity analysis. Yes, the slope parameter g_1 in MED_mean and BWB_mean in Fig. 5 is calibrated based on the same field measurements, therefore equivalent to MED-B and BWB in Fig. 1. The reason why we did not name the MED simulation as "MED-B" in Fig. 5 is that we tried to differentiate the parameterization for the evaluation simulations from that for the sensitivity simulations. To increase the connection between the figures, we added the following sentences at Lines 189-190:" Because g_1 was estimated for BWB and MED models based on the same measurements, g_1 was equivalent for the two models and the simulations resemble MED-B and BWB in section 2.3".

-Or should we be comparing to the ecosystem dependence shown in Fig. 2a, which shows little difference? I would expect the LICOR measurements to be done on a single leaf, instead of measuring over a larger ecosystem.

Yes, the LICOR measurements were conducted on a single leaf.

-The case of PAR < 500 mol/m2/s seems to be handled robustly for the date of May 25, in Fig. 8, where both BWB_mean and MED_mean are lower than the previous 2 days in May, particularly later in the afternoon on May

25. In this case of May 25, BWB_mean does seem to be 2 times higher than MED_mean, even in the morning, which might make a bit more sense if is for MED-default_mean instead of MED-B_mean, this time. On May 23 and on May 24, BWB_mean is 50% greater than MED_mean in the morning, but by mid-day, the models don't differ much. Maybe the higher VPD that is reached by mid-day on May 23 and May 24 effectively closes the pores, causing the models not to differ? May seems different than (the dry season of) February - April, in that VPD is 0 kPa at night for May.

We guess that the reviewer intended to express that MED_mean is two times higher than BWB_mean on May 25, and MED_mean is 50% greater than BWB_mean in the morning of May 23 and 24 based on previous Fig. 8c. As we stated above, we used the comparable g_1 parameters for BWB and MED models in the evaluation. The reason for higher MED_mean under those conditions could be that the modeled VPD at the leaf surface (VPD_s) is very low due to the low VPD in the air (VPD_a). See Fig. 5c in Franks et al. (2017), the MED model predicts markedly higher g_{sw} than the BWB model when VPD_s is very low.

-Lines 376-378: "Our method in keeping VPD in the air constant when studying model response to varying T_air by adjusting specific humidity concurrently is inspiring for other modelers." Such future inspiration of other modelers may indeed happen, but the language in this sentence is a bit presumptuous.

We modified this sentence to: "By adjusting the specific humidity concurrently with air temperature, we were able to isolate the model response to changing air temperature from typically concurrent change in VPD_a" at Lines 418-419.

-Line 619: citation for Pachauri et al. should have 51 authors instead of 10 authors.

We corrected this reference in the revised manuscript.

-Fig S2b: The r^2 value for the MED model is quite a bit lower than for the BWB model. Is this a real effect? Maybe the fit can be improved by removing a single outlier for MED at a value of Modeled g_sw = 0.24? It's ok sometimes to remove outliers when doing fits. And that outlier seems unusual, too, since it is a MED point that doesn't have a corresponding nearby BWB point like most of the other points do.

In that figure we combined results from the four campaigns to indicate the model overall performance in capturing the means of observations across all the species. The lower R^2 for g_{sw} fitting when using the MED compared with the BWB is mainly contributed by the results in May, when MED model overestimated mean g_{sw} whereas BWB model captured mean g_{sw} relatively well. However, given the small number of measurements and large uncertainty range of both measurements and model results, we could not tell which model is superior simply based on the fitting of mean responses. We believe the evaluation would be more informative as more observations are available in the future.

References

Franks, P. J., Berry, J. A., Lombardozzi, D. L., and Bonan, G. B.: Stomatal function across temporal and spatial scales: Deep-time trends, land-atmosphere coupling and global models, Plant Physiol., 174, 583–602, https://doi.org/10.1104/pp.17.00287, 2017.