We thank the authors for their helpful clarifications. Our more detailed comments can be found below:

1. **Timing of model updates on page 7**
   Thanks for the clarifications on this, but it seems like we phrased our initial comment in a way that was not clear. The main question was why this should be updated annually at Oct 1st, instead of immediately after the fire. If real conditions change, so should the model right (in other words, isn't getting changes to the simulated fluxes the reason for updating the model)? Is it some model constraint that vegetation can only be changed once a year? It probably doesn't matter a ton given that the fire happened in September, but it's still worth clarifying.

2. **Calibration of Correction Factors**
   Thanks for outlining the Bayesian perspective on the calibration of forcing data correction factors. A couple of thoughts:

   Specifically related to the statement (in the response document): *"Compensation between unknown errors in the meteorology data, model structure and calibration, and reconstructed streamflow can potentially contribute to spurious goodness-of-fit metrics with hidden physical deficiencies."* -- This is presented as an argument in favor of calibrating correction factors (i.e., if input data are flawed, any resulting calibration of the model will be trying to correct for this and thus the resulting parameters will be flawed too) but it works as much in the other direction: allowing input data to be freely "corrected" introduces another source of uncertainty. In such cases, the calibration process gains additional degrees of freedom, potentially adjusting inputs (e.g., precipitation) to achieve desired outputs, without necessarily improving the physical realism of the resulting parameters. This may still produce strong goodness-of-fit metrics while concealing underlying model deficiencies.

   Related to calling a priori bias correction "calibration by a different name" – this is of course one way to frame this, but that may be missing the point. Our initial concern is not so much with changing (calibrating) the forcing data, as it is with doing this simultaneously as calibrating the model parameters:

(1) By bias-correcting the forcing a priori, the forcing needs to be matched to some expectation of the real forcing instead of being turned into another dial to get the streamflow that comes out of the model looking right.

(2) By keeping the forcing constant across the different model trials, any differences between the trials therefore originate in a more constrained part of the problem: the model parameters. This should lead to cleaner insight about the model's capability to model pre- and post-disturbance situations with a given set of inputs. See, e.g. any papers on hypothesis testing in hydrology and isolating different modelling decisions to assess their impact.

That said, there is clearly a difference in point of view between the authors (forcing + model is the system of interest) and the reviewer (model is the system of interest). Either side can be argued, and re-running the experiment seems computationally expensive (l275-277).

To avoid this turning into a yes/no back and forth, can the authors add some information that shows to what values these forcing correction parameters were calibrated across the 30 final parameter sets, and discuss accordingly? If the values are approximately the same, then the reviewers' point is moot. If there's wide dispersion in calibrated correction factors, there's some reason to believe that what these parameters are doing is not unambiguously pushing incorrect forcing data closer to their true values. The third discussion paragraph could be a good spot to add a few words about this.

3. **Leaf Area Index (LAI) Clarification**

Clarifications helped to better understand Figure 4, but it would be good to add some extra words that explain why spatially averaging LAI covers such a large range. If LAI is derived empirically from fractional tree cover and tree cover is derived from vegetation maps (p7), wouldn't LAI be a static value that doesn't change between the different parameter sets?

4. **Citation Suggestions for Equifinal Parameter Sets Producing Divergent Predictions**

A cursory literature search suggest various papers that may support this statement in a general sense (e.g., Kelleher et al., 2017: demonstrating that parameter sets equifinal with respect to streamflow and SWE produced markedly different annual and seasonal simulations of water table depth), in the context of changing conditions (e.g., Melsen et al., 2018: showing that ostensibly equally plausible models can have markedly divergent predictions on future conditions), and specifically in the context of disturbances (e.g., Seibert & McDonnell, 2010: showing the impact on streamflow of forest clearing practices in a paired watershed experiment).

Refererence 1: Kelleher, C., McGlynn, B., and Wagener, T.: Characterizing and reducing equifinality by constraining a distributed catchment model with regional signatures, local observations, and process understanding, Hydrol. Earth Syst. Sci., 21, 3325–3352, https://doi.org/10.5194/hess-21-3325-2017, 2017.

Reference 2: Melsen, L. A., Addor, N., Mizukami, N., Newman, A. J., Torfs, P. J. J. F., Clark, M. P., Uijlenhoet, R., and Teuling, A. J.: Mapping (dis)agreement in hydrologic projections, Hydrol. Earth Syst. Sci., 22, 1775–1791. https://doi.org/10.5194/hess-22-1775-2018, 2018.

Reference 3: Seibert, J., & McDonnell, J. J. (2010). Land-cover impacts on streamflow: a change-detection modelling approach that incorporates parameter uncertainty. *Hydrological Sciences Journal*, *55*(3), 316–332. https://doi.org/10.1080/02626661003683264.