Reviewer #2

General comments

This manuscript presents the development and evaluation of WULFFSS, a novel stochastic monthly gridded model for simulating large forest fires in the western United States. The model operates at a 12-km resolution and leverages interpretable statistical methods to estimate fire probability and size. Its key strengths include the integration of spatial, annual cycle and temporal anomaly components that can interact, high computational efficiency enabling large ensembles, and its design for coupling with the DYNAFFOREST ecosystem model to simulate vegetation-fire feedbacks. The model demonstrates considerable skill in capturing frequency and extent of western US forest fires, as validated through rigorous temporal and spatial cross-validation. The manuscript is thorough, well-structured, and describes the model, data, and validation procedures in significant detail. The work represents a valuable contribution to the field of fire modeling. However, the text is somewhat lengthy, and there is room for improvement. Therefore, I recommend that this manuscript be accepted with minor revisions.

We thank the reviewer for the thorough and constructive review. We are especially glad that this review motivated us to add a diagram of the general model structure and also to look into the effects of including 3-way interactions between the spatial, mean annual climate cycle, and temporal climate anomaly components of the model. Although including this term did not improve model performance and we thus chose to not keep it, we suspect this would have been a common question among readers and we now explicitly note that the 3-way interaction was considered but not included because it did not contribute additional skill. We also appreciate the reviewer's concern shared above regarding the length of the paper. In light of that comment we carefully went over the writing with conciseness in mind. The clearest opportunities for streamlining were in the Introduction and the revised Introduction is approximately 200 words shorter than the original. We also appreciate the reviewer's suggestion to include a diagram illustrating the general model framework and we now include such a figure. Notably, most reviewer requests were for additional details, so ultimately the revised paper is longer than the original despite our work to shorten the original text. We have revised the paper in light of the reviewer's comments and suggestions and we provide point-by-point responses in blue font below.

Specific comments

1. The model operates at a 12-km resolution, and it is possible that within a grid cell, the actual burned area constitutes only a small fraction. A key concern is the issue of intra-grid-cell heterogeneity and the representativeness of the predictor variables. It would be valuable if the authors could discuss the potential implications of this scaling issue.

We agree that the probability and size of a fire occurring in a given 12-km grid cell should be sensitive to heterogeneity at the sub-12-km scale. We do use landcover at sub-12-km resolution to calculate the 12-km predictor values for many of the spatial variables. For example, fractional forest coverage and forest biomass comes from our 1-km resolution forest-ecosystem model (DYNAFFOREST) and the fractional cover of other landcover types such as barren, water, and grass/shrub are calculated from 30-m maps. We have added a sentence in this vein to the end of the methods section 3.4 about landcover:

L343–345: "*Likewise, our use of sub-12-km landcover data to produce landcover predictors allows our modelling to include the effects of within-grid heterogeneity of fuel conditions, which is important given that most fires are smaller than 144 km^2.*"

In addition, we have added the following note to the paragraph about fire spread in the Discussion section:

Finally, the previous version of the Discussion did indicate how fire-spread is simulated within our forest-ecosystem model at the sub-12-km level as well as opportunities for more improvements of sub-grid processes. We have modified that text and it now reads:

L1009–1014: "*Another opportunity for improvement is to explicitly simulate fire spread. Currently, WULFFSS only estimates the final forest area burned by each simulated fire. When coupled within DYNAFFOREST, the ignition of a given simulated fire is assigned to a random 1-km forested-grid cell within the 12-km grid cell of WULFFSS and the fire spirals through adjoining or nearby forested areas until the pre-determined fire size is achieved or no nearby forested grid cells remain. Future improvements to WULFFSS should include estimating ignition location at sub-12-km resolution and modelling fire spread while maintaining computational efficiency.*"

2. The authors have considered a comprehensive list of potential predictors for constructing the Sp, Cp, and Tp components. However, the rationale or guiding principles for assigning specific variables to each of these three domains (Spatial, Annual-cycle, Temporal anomaly) is not sufficiently elaborated. A clearer explanation of the criteria used to categorize predictors into S, C, or T would significantly enhance the methodological transparency and reproducibility.
This comment is similar to the second comment from Reviewer #1 and we agree that the original submission fell short in describing the logic behind splitting the variables into the three components and we expanded the second paragraph of section 4 (Model description) to state:

L395–401: "*The S component is constructed first to capture the how variations in fire activity are driven by factors that are far more variable in space than in time, as these factors (e.g., forest biomass, lightning frequency, variables related to human population and fire suppression) are likely to modulate the sensitivity of fire activity to temporal variables. The C component is then constructed to account for variations in fire activity that are due to the mean annual climate cycle. Finally, the T component is constructed to account for effects of interannual climate variability, which are likely to be strongly modulated by the effects of the S and C variables already accounted for.*"

We also note that the original submission did describe the guiding principles in terms of how it was determined how it was determined which domain a given variable would be assigned to. In section 4, this is what was stated for each domain, now with slight revisions to clarify that the mean annual cycle (*C*) and temporal anomaly (*T*) components are composed of climate variables.

Spatial (*S*): L403–405: "*The S component represents drivers of forest-fire occurrence or size that are most variable in the spatial domain, such as topographic slope, fuel availability, human* *population, mean annual lightning frequency, and long-term mean aridity, all of which may directly influence fire occurrence and also modulate the effects of C and T.*"

Mean annual cycle (*C*): L411–415: "*The C component represents climatological drivers of forest-fire occurrence or size that are most variable in the domain of the mean annual cycle, such as long-term means of each month's lightning frequency as well as variables that influence the seasonality of fuel moisture such as prec, solar, and VPD.*"

Temporal anomalies (*T*): L422–429: "*The T component represents climatological drivers of forest-fire occurrence or size that are most variable in the temporal domain of interannual and longer. … Because T is meant to represent climate variability beyond the annual cycle, T variables are standardized so that for a given variable in a given grid cell, values have a mean of 0 and standard deviation of 1 for each of the 12 months during the calibration period.*"

3. The model commendably incorporates interaction terms between the S, C, and T composite predictors, but these are limited to pairwise (two-way) interactions. The potential three-way interaction (S×C×T) is not considered. Could the authors please justify this methodological choice?

This comment motivated us to produce an alternative version of the model that includes 3-way interactions between *S*, *C*, and *T* variables. Our choice to not do so originally was simply to limit the computational cost of parameterizing the fire-probability model as well as some worry that *S*x*C*x*T* interactions could produce occasional extreme simulation outcomes that are unrealistic. When we introduced the 3-way interactions in light of this reviewer comment, we found the performance of the new model was virtually identical to that of the old. Below in Fig. R2 we provide alternative versions of Figures 17 and 18, where we show how the new simulations that include 3-way interactions compare to observations in terms of annual and monthly fire frequencies and areas burned at the scales of the western US and regionally. Comparing these to Figures 17 and 18 in the resubmitted paper indicates that the alternative model with 3-way interactions performs virtually identically to, and not better than, the original model. Notably, we evaluated the behavior and performance of the alternative model with more depth than simply generating the alternative versions of Figs. 17 and 18 below. We found that allowing for 3-way interactions does generally lead to slightly more variables being included in the construction of the *S*, *C*, and *T* predictors, but these additional variables have minimal impact because the most impactful variables were consistently selected in the same order and with the same relationships to fire probability/size regardless of whether the 3-way interactions were included or excluded. Thus, we have decided to not include 3-way interactions in the final version of the model. We have added the following sentence to section 4.1 (Model framework):

L466–467: "*Notably, we considered including three-way interactions between the S, C, and T predictors in the P and A models but doing so did not improve model skill.*"

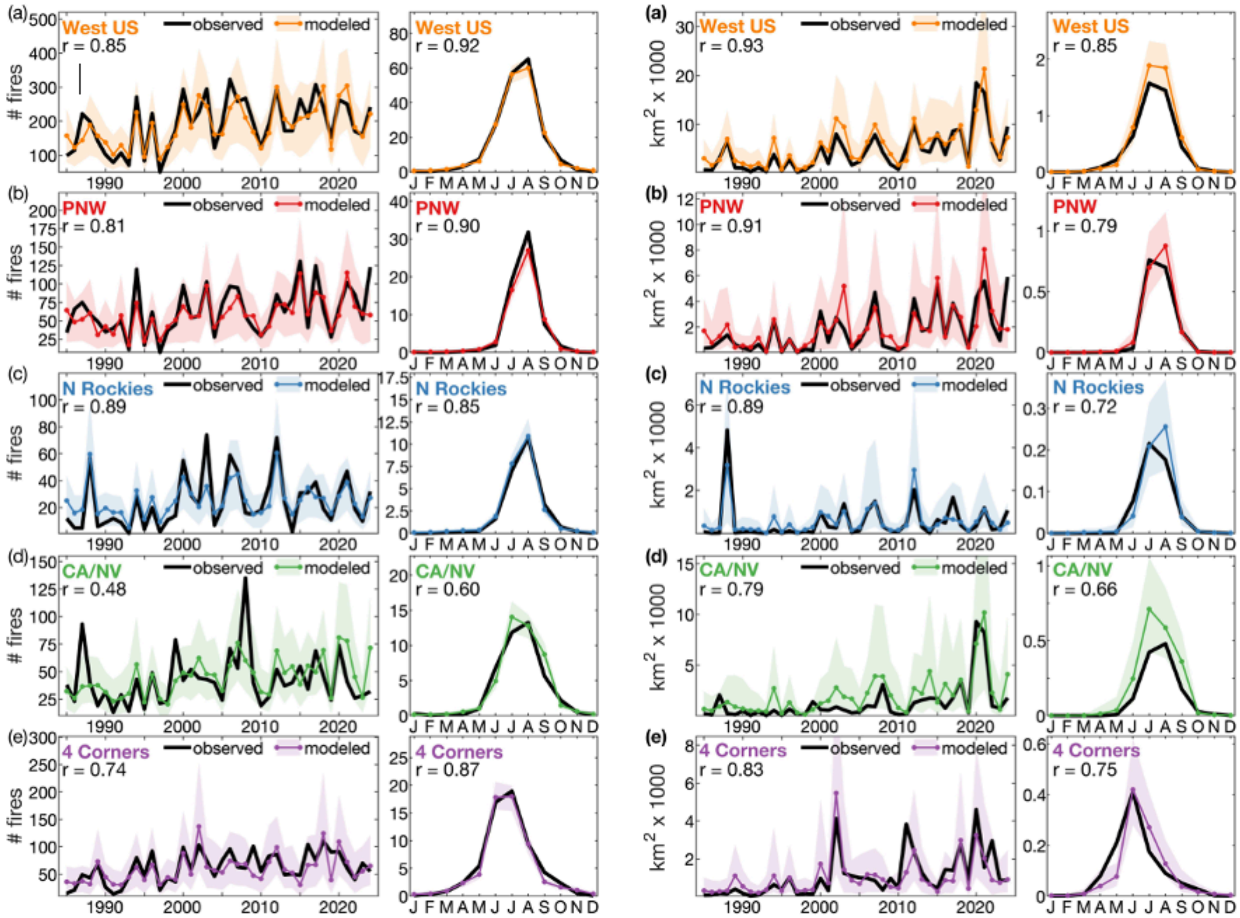

Fig R2. Alternative versions of Fig. 17 (left) and Fig. 18. (right) where the probability and size models each include a 3-way interaction term between the spatial, seasonal, and temporal predictors. Comparison to Figs. 17 and 18 in the resubmitted paper indicates that model performance is minimally influenced, and not systematically positively, but inclusion of the additional interaction term.

4. I recommend including a schematic diagram illustrating the general framework of WULFFSS. This figure should visually depict the relationships between the three core statistical models (P, N, A), their required input data streams (landcover, topography, climate, etc.), the key data processing steps (e.g., resampling), and the bidirectional coupling relationship between WULFFSS and the DYNAFFOREST model.

We agree that a diagram illustrating the general model framework would be helpful. We have added such a figure (now Figure 4 in the manuscript) and we provide it below as Figure R3.

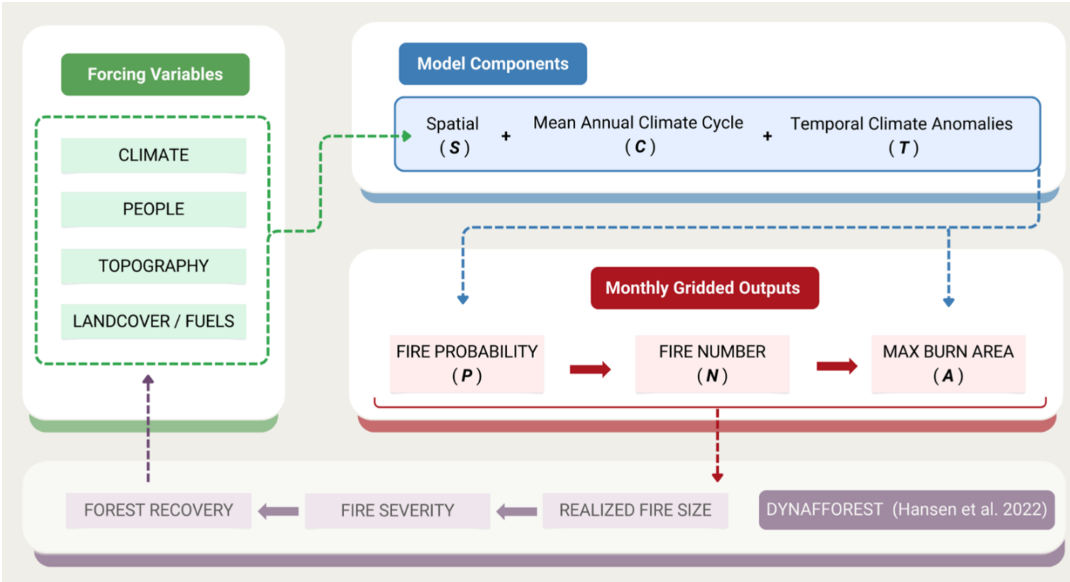

Figure R3: Flowchart outlining the general framework of the WULFFSS.

Figure 3: Adding text labels (e.g., "Pacific Northwest" or “PNW”) onto the map for the four quadrant regions would improve its immediacy and clarity.
We have added a legend to Figure 3 to clarify the names of the regions.

Figure 4: Providing quantitative goodness-of-fit metrics for the curve fits in each panel would be welcome. Furthermore, the relationship for some predictors (e.g., wetds_mean_1mo in panel b) appears weak or poorly captured by the fitted curve.
We now include the AICc and p-values associated with each curve fits shown in the figure showing P model predictors (now Fig. 5) as well as the figure showing A model predictors (now Fig. 8).

Line 37: “fire” to “fires”
We fixed this typo and thank the reviewer for pointing it out.