(Note: The reviewer's comments are in gray and the author's responses are in blue. Unless specified otherwise, the line numbers quoted in our responses are with reference to the revised manuscript.)

Jatan Buch et al. developed a SMLFire model based on Mixture Density Networks and monthly climate, land surface, and atmospheric conditions. This work focused on both fire frequency and total burnt area over Western US. In general, this study is timely and important. The high performance of SMLFire is exciting for both fire frequency and burnt area. The presentation is smooth and well-done. Congratulations. Below are my comments and recommendations.

We thank the reviewer for their positive feedback on our manuscript. A detailed response to their individual comments is given below:

1. Fuel load seems missing in the input variable list, which is an important predictor for fire spread thus burnt area. Also GDP (missing) is often considered as an important indicator of human effects on fire management and firefighting efforts. Some others are also potentially useful to consider, e.g. road density. I would suggest creating a new table with a full list of input variables and explaining how these variables possibly affect fire frequency and burnt area.

We agree with the reviewer's comments that fuel load is an important input variable for fire activity. Ideally, we would like to include dynamic fuel load variables that track changes in fuel density from climate perturbations as well as previous fires. However, we could not find any fuel load products solely based on observations in the literature, so we used the fractional land cover outputs from the National Land Cover Database (NLCD) as input predictors instead. We have also mentioned (lines 148-150 in the original manuscript) that a promising future direction of research is to precisely accomplish what the reviewer suggests: include fuel load variables such as dead and live biomass from a dynamic vegetation model.

Since our fire model is based only on the western US, it is unclear to us how GDP, typically a national level economic indicator, will be a helpful predictor. We would appreciate helpful references from the reviewer on this point.

We have now included the following table as Table S3 in our supplementary information section as per the reviewer's suggestion to clarify the qualitative effect of individual variables on fire frequency and burned area:

Predictors	Qualitative effect		Comments
	Fire frequency	Fire size	
VPD, AntVPD_*M*mon, VPD^{maxX}	↑	↑	VPD on multiple timescales, from weekly to seasonal, is positively correlated with both fire frequency and size.
Tmax,	↑	↑	Tmax on multiple timescales, from weekly to

AntTmax_*M*mon $Tmax^{maxX}$			seasonal, is positively correlated with both fire frequency and size.
Tmin, $Tmin^{maxX}$	/	↑	Both extreme Tmin and monthly mean Tmin are positively correlated with fire size. Tmin is not a significant predictor for fire frequency.
Prec, AntPrec_*M*mon	↓	↓	Prec on multiple timescales, from monthly to seasonal, is negatively correlated with both fire frequency and size.
AntPrec_lag1, AntPrec_lag2	↑	↑	Annual mean of Prec in lagging years, a proxy for biomass growth, is positively correlated with fire frequency and size.
SWE_mean, SWE_max AvgSWE_*M*mon	↓	↓	Snow water equivalent on multiple timescales, from monthly to seasonal, is negatively correlated with both fire frequency and size.
FM1000	↓	↓	1000-hour dead fuel moisture is negatively correlated with fire frequency and size.
FFWI, $FFWI^{maxX}$	↑	↑	Mean and extreme values of FFWI are positively correlated with fire frequency and size.
$Wind^{maxX}$	/	↑	Monthly maxima of X-day mean wind speed is positively correlated with fire size. Wind speed is not a significant predictor for fire frequency.
Biomass	↕	↑	Spatial variance in biomass is positively correlated with fire size, however its effect on fire frequency is ambiguous with potential confounding by human action predictors.
Grassland, Shrubland	↑	↑	Fraction of grassland and shrubland cover increases fuel flammability and continuity over a landscape, and is thus positively correlated with fire frequency and fire size.
Lightning	↑	/	Increased lightning strike density contributes additional ignitions, and is positively correlated with fire frequency. Lightning is not a significant predictor of fire size.
Slope	↑	↑	Slope is positively correlated with fire frequency and size since the rate of fire spread is proportional to the degree of slope.
Southness	↑	↑	Southness, or mean south-facing degree of slope, dictates the level of solar insolation and is positively correlated with fire frequency and size.

Pop10_dist	↕	↑	Increased distance from areas with population density greater than 10 km^2 is a correlate of remoteness leading to a larger fire size. Its effect on fire frequency is ambiguous since these areas experience fewer ignitions while also having reduced access to early fire containment efforts.

Table 1. Summary of the qualitative effect and physical meaning of important model predictors on fire frequency and burned area. The symbols ↑, ↓, ↕, and / refer to positive, negative, ambiguous, and insignificant correlations between a predictor and fire response variable respectively.

2. Spatial evaluation of SMLFire simulation is limited to regions, but evaluation on gridcell scale is also important because the model is gridcell-based and spatially-explicit. Suggest showing spatial maps of simulated vs observed western US fire frequency and burnt area statistics for long-term mean, decadal trend etc. It is interesting to see 12-km scale spatial hot spots of trends and variability as well.

Thank you for raising an important point. We think spatial validation of our model at the 12-km scale is difficult for two main reasons:

i) Most 12 km grid cells (≳ 99%) do not experience any fires in the study period of ~40 years, and only ~5% of the grid cells with fires experience more than 1 fire in any month. Even on decadal scales, fires are incredibly rare, so unless the model is trained on 1-2 degree (~50-100 km) grid size scale, it will not contain enough fires for a meaningful trend. Thus, we choose aggregate EPA Level III (L3) ecoregions to demonstrate the monthly, interannual, and decadal variability of fire frequency and burned area across the western United States.

ii) Moreover, the stochasticity of monthly scale climate predictors that our model is trained on also contributes to the lack of spatial precision over longer timescales. Most papers in the literature that simulate long-term trends in fire probability (for example, Parisien and Moritz, 2009; Chen et. al., 2021) end up relying on climate normals (*i.e.* long-term averages) as predictors. We believe that one of the strengths of our models is its ability to leverage this stochasticity for projecting a range of possible outcomes, which is more useful than accuracy for planning fire mitigation.

3. Human vs natural ignited fires have clear differences in ignition location and background climate. I understand that SMLFire does not distinguish human vs natural fire, but it is worth exploring or discussion on how that might bias SMLFire in simulating spatial-temporal distribution of fires as well as the interpretation of underlying control factors for fire frequency and burnt area.

We appreciate the reviewer's comment about accounting for the difference in human vs natural ignitions. As a preliminary exploration of potential biases in SMLFire, we performed the following experiment: based on Fig. 1 of Balch et al., 2017, we identified that a large proportion of human-started fires (≳ 60% of all fires in that ecoregion) in our western United States (WUS) study region occur in Mediterranean California (CA) ecoregions as well as coastal parts of the Pacific Northwest (PNW). Next, we trained SMLFire on fires from all 5 L3 ecoregions from this area and performed the SHAP analysis for fire frequency predictors.

In Fig. 1 below, we compared the SHAP values for all fires in the CA+PNW ecoregions from the above experiment with the results for SMLFire trained on fires across the WUS (as shown in Fig. 6 of the manuscript, which is also reproduced as the right panel of Fig. 1). We find that relative to the WUS case, the SHAP values for Lightning are more important for the CA+PNW case. However, unlike the WUS case, we find that Campnum, or the mean number of campsites in a grid cell, emerges as an important predictor of fire frequency. None of the other human related predictors are selected in our experiment.

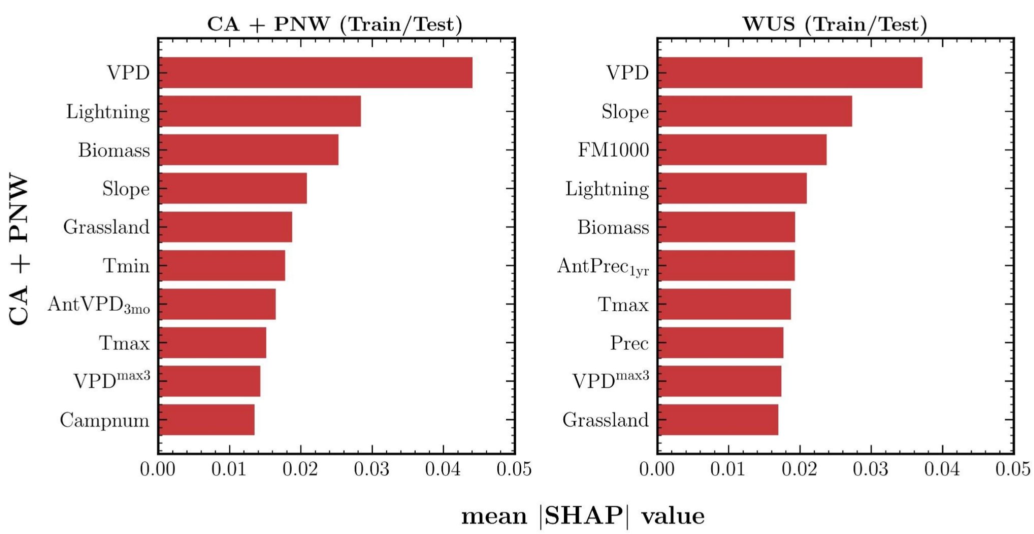

Figure 1. Mean absolute SHAP values for all fires in the CA+PNW ecoregions with the frequency MDN model trained on: (Left) only fires from the 5 L3 ecoregions in the CA+PNW area; (Right) fires from across the WUS as shown in Fig. 6 of the manuscript.

Based on the experiment outlined above, we have added the following sentences to our discussion clarifying the potential bias in SMLFire while modeling fires with human vs natural ignitions:

(lines 421-423) *Given that a large fraction of fires in parts of the WUS, especially Mediterranean California and coastal PNW (Balch et. al. 2017), are human ignited, this result could stem from a skewed sampling of fires while training SMLFire1.0 as well as the lack of correlation between our chosen human predictors and fire occurrences.*

(lines 557-559) *Alternatively, we could leverage the seasonal differences between human and lightning started fires to account for potential selection biases in training data for SMLFire1.0.*

4. It's not clear how uncertainty quantification is done. Does it only consider parametric uncertainty? How about model structure (how many layers of hidden layer, number of neurons for each layer), other hyperparameters? How about forcing data uncertainties?

The mixture density network used in our analysis learns the function map between input predictors and the parameters of a mixture distribution for each individual fire. This may be interpreted as approximating the likelihood of fire occurrence or size given the input predictors. We estimate the parametric model uncertainty by performing Monte Carlo simulations on a frequency or size MDN model with parameters fixed to their optimal values and calculating the variance of the samples.

Our framework does not account for uncertainties due to hyperparameter or forcing data uncertainties in this analysis. However, given that we are approximating the likelihood function, we can easily embed it in a hierarchical Bayesian model to account for the hyperparameter and data forcing uncertainties. Typically, we expect the data forcing uncertainties to be a much bigger factor while using SMLFire in conjunction with seasonal and subseasonal-to-seasonal (S2S) climate model forecasts.

We have edited the following sentence in the Methods section to clarify the reviewer's point (lines 223-224):

> *We treat the variance as an estimate of the parametric model uncertainty, or equivalently the uncertainty in modeled frequency due to different realizations of a parametric model.*

We have also lightly edited the first paragraph of the Conclusions section (lines 603-608) to make the point about parametric model uncertainty estimation clearer.

References:

1. Parisien, M. and Moritz, M.A. (2009), Environmental controls on the distribution of wildfire at multiple spatial scales. Ecological Monographs, 79: 127-154. https://doi.org/10.1890/07-1289.1

2. Chen, B., Jin, Y., Scaduto, E., Moritz, M. A., Goulden, M. L., & Randerson, J. T. (2021). Climate, fuel, and land use shaped the spatial pattern of wildfire in California's Sierra Nevada. *Journal of Geophysical Research: Biogeosciences*, 126, e2020JG005786. https://doi.org/10.1029/2020JG005786

3. Balch, J. K., Bradley, B. A., Abatzoglou, J. T., Nagy, R. C., Fusco, E. J., & Mahood, A. L. (2017). Human-started wildfires expand the fire niche across the United States. Proceedings of the National Academy of Sciences, 114(11), 2946–2951. https://doi.org/10.1073/pnas.1617394114