*(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.)*

Overall, this study has interesting components and would be a nice contribution to the literature applying ML approach on wildfire prediction. The paper is well-written, and I think that the authors' are in a strong position to introduce the stochastic ML method, which is a relatively new concept, to the wildfire modeling community. There are a few aspects that could be further addressed before the paper is suitable for publication.

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

1a. This study used SHAP values to compare importance between predictors for the entire period as a global perspective and for each Ecoregion. But, how about the importance changes for temporal aspects (e.g. dry/wet season or extreme fire events)?

We have split the reviewer's original point into two parts for clarity. We agree that the temporal aspects of SHAP are an important diagnostic given that previous works (in particular, Wang and Wang, 2020) have shown subtle differences in the fire behavior between the dry and wet seasons in south central United States (US). Shown in Fig. 1a and Fig. 1b are the western US SHAP importance plots for the dry (May - September) and wet (October - March) seasons respectively for a MDN frequency model trained on all fires. The most significant differences between the two seasons are: the number of fires, with the dry season experiencing a factor of ~10 more fires than the wet season, and the increased importance of extreme weather variables such as $VPD^{max3}$ and $FFWI^{max3}$.

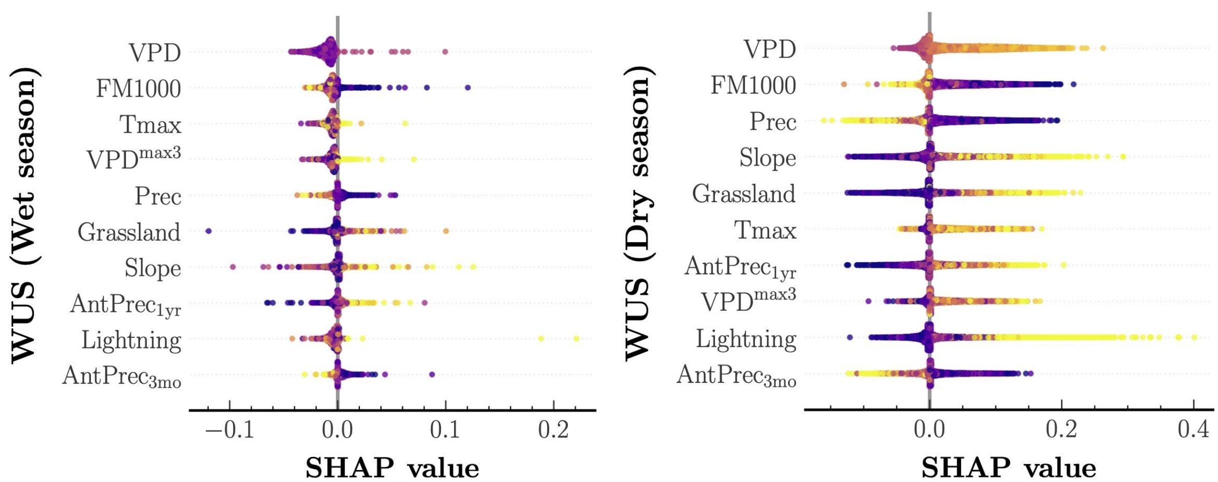



**Figure 1a.** SHapley Additive exPlanation (SHAP) analysis of the fire frequency MDN model outputs across the western United States for the wet (left) and dry (right) seasons. Input predictors sorted in descending order of their mean SHAP values aggregated over the entire

study period. Each colored point along the x-axis represents an individual prediction with the color corresponding to high (yellow) or low (indigo) values of the respective input predictor.

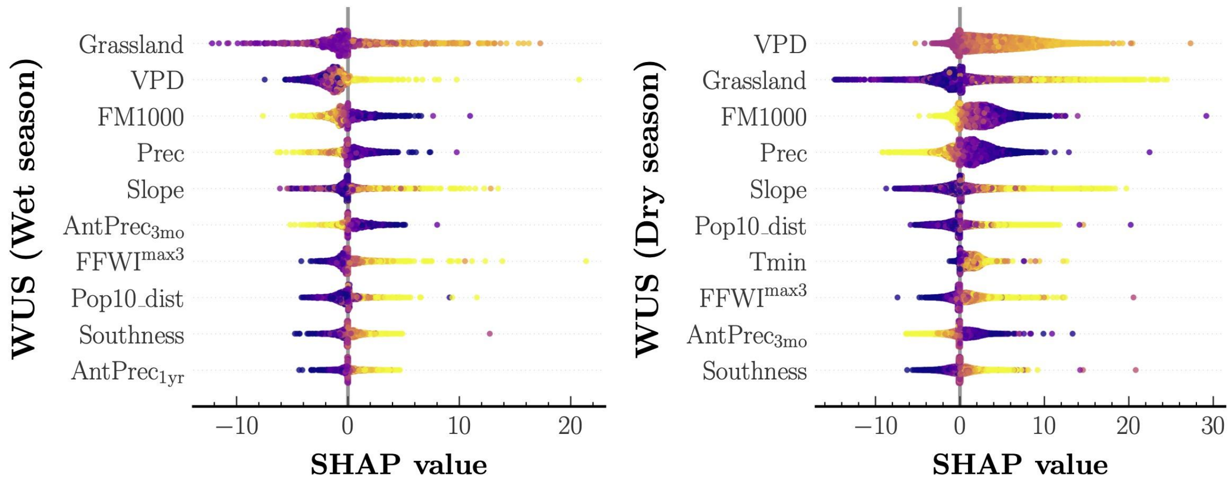



**Figure 1b.** Same as Figure 1a. but for the fire size MDN model.

1b. It is interesting that none of the results in this study actually show significant importance for wind speed, although it is a key factor of fire spread.

The reviewer is correct in pointing out that wind speed is a key driver of fire spread. Indeed our results in the manuscript corroborate this fact since the monthly mean of 3-day maximum Fosberg Fire Weather Index ($FFWI^{max3}$) is an important predictor in our SHAP plots for the fire size model in almost all the ecoregions and their aggregations. The FFWI is an index of fire severity calculated using temperature, humidity, and wind speed, which has been shown to be an important correlate of wind-driven fires (Moritz et al., 2010; Barbero et al., 2014).

As such, the SHAP importance values shown in the manuscript are for a model trained with both the monthly mean of 3 day maximum wind speed (Wind) and $FFWI^{max3}$, with the $FFWI^{max3}$ being more important in all ecoregions. When we repeated our analysis after removing $FFWI^{max3}$ as a predictor, we obtained SHAP importance plots with Wind as one of the important predictors as shown in Fig. 2.

To clarify the properties of the FFWI as a correlate of hot, dry windy conditions, we have also included the following sentence in our revised manuscript (lines 124-126):

> *The FFWI, which is calculated using temperature, humidity, and wind speed (Fosberg, 1978), has been shown to be an important correlate of dry, windy conditions associated with fire weather (Moritz et al., 2010).*

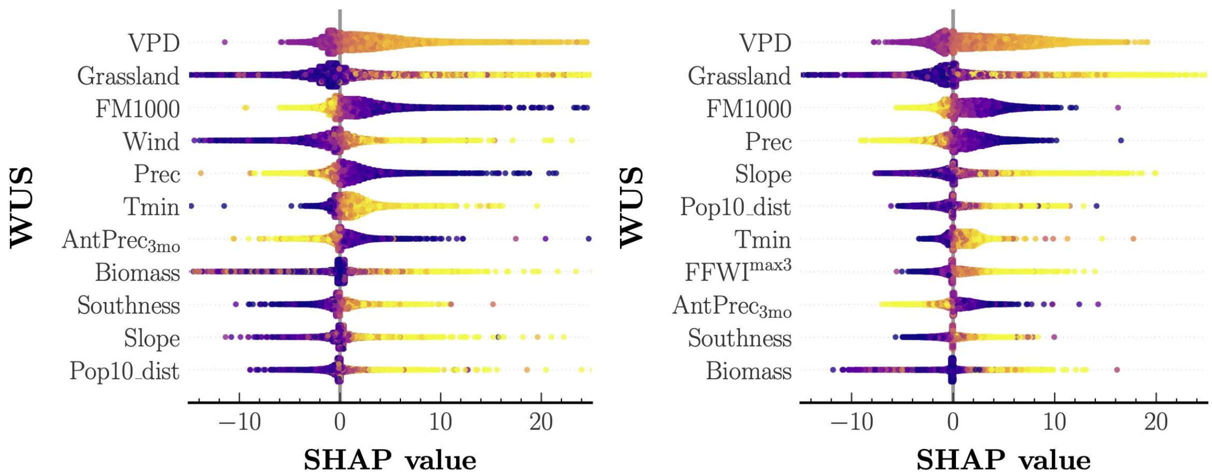



**Figure 2.** SHapley Additive exPlanation (SHAP) analysis of the fire size MDN model outputs across the western United States with wind speed as a predictor instead of FFWI (left), and both wind speed as well as FFWI (right).

2. The discrepancy of the year 2020 (Figure 11) can be further analyzed with input predictors. Although the scale of AAB 2020 is out of the range during the training period, it can be associated with abnormal patterns in climate/vegetation or sudden changes in human induced predictors. The authors may consider this further.

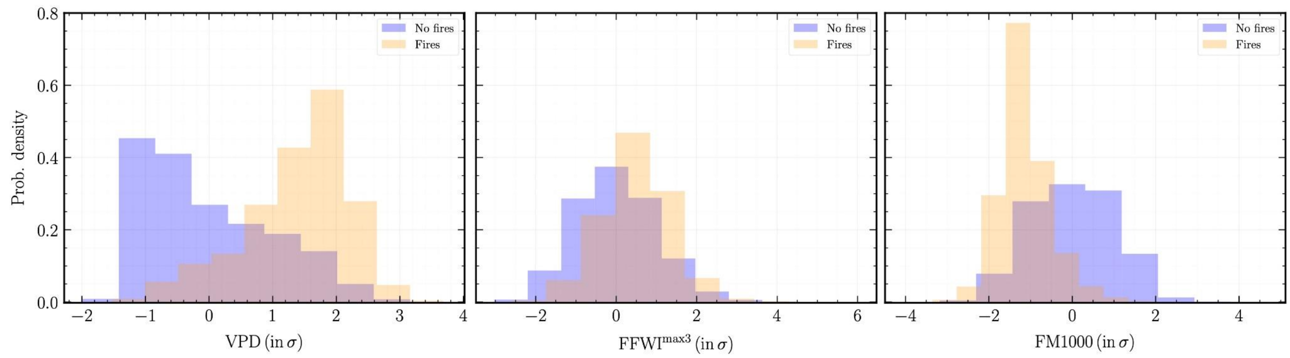
The reviewer raises an interesting point that we considered while preparing our manuscript, but never explicitly addressed in its writing. Essentially, the question can be posed as follows: "Are extreme sizes for individual fires a result of extreme or anomalous values of various predictors? And since such conditions are rare, by definition, how do we train our statistical or machine learning model to correctly assign extreme responses to extreme predictor values?"

The second question is admittedly harder, but as the plots in Fig. 3 indicate: the largest fires in 2020 did *not* correspond to extreme fire weather conditions. In fact, there is a significant difference in the fire sizes of the largest fire and the fire corresponding to the most extreme value of an important predictor, potentially highlighting the role of prompt human action in containing the growth of large fires.

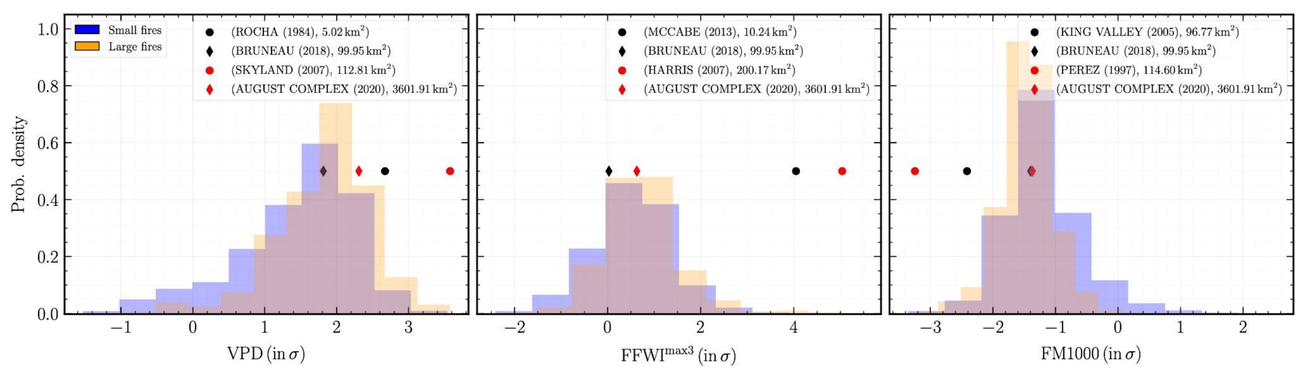
We consider four different types of plots to argue the above point. First, we contrast the distribution of three important fire weather predictors: VPD, $FFWI^{max3}$, and FM1000 in grid cells with and without fires. As shown in Fig. 3a, we observe a clear shift in the distributions of VPD and FM1000 in the presence and absence of fires, but there is also a sizable overlap for months with moderate fire weather. This overlap is even more significant when contrasting the predictor distributions for small and large fires in Fig. 3b. Moreover, while the most extreme predictor values correspond to large fires (red circle), the largest fire size (red diamond) occurred at moderately high but not extreme fire weather. Most strikingly, while comparing the predictor distributions for large fires that occurred between 1984-2019 and those that occurred in 2020, we find in Fig. 3c that: a) the most extreme weather conditions led to smaller fires (red, black circles) relative to the largest fire sizes (red, black diamonds) in the respective time period; b) with the exception of $FFWI^{max3}$ values for several fires in 2020, the distributions of other predictors in grid cells with large fires were on average *less* extreme than in 1984-2019.

In summary, there appears to be only a very weak monotonic relationship between extreme predictor values and fire sizes. An important caveat being that the predictor values used in our analysis are at coarse spatial and temporal resolutions relative to the physical scales of fire front propagation. We are exploring different ways to bridge the two regimes in ongoing work.

**a)**




**b)**




**c)**

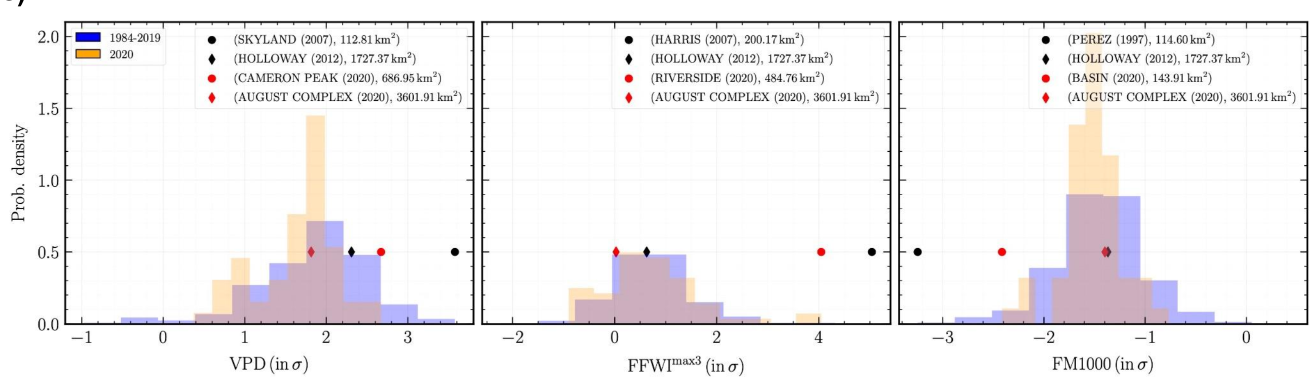


**Figure 3.** Contrasting probability distributions of VPD, $FFWI^{max3}$, and FM1000 values (in $\sigma$) for grid cells with: a) no fires and fires; b) small fires and large fires; c) all large fires in 1984-2019 and 2020. In b) and c), the circles indicate fires that occurred in grid cells with the most extreme predictor value, whereas the diamonds indicate the largest fire by burned area.

3. Why is 'Southness' selected rather than other directions? Also, interesting since it is included in the top 10 important predictors for the size model (Figure 12 and 13). It would be nice to further describe the role of 'Southness' in this study domain.

We appreciate the reviewer raising this comment. Mean south-facing degree of slope (also referred to as slope aspect in the fire ecology literature), or Southness, is associated with higher insolation in the Northern Hemisphere, resulting in drier conditions and low fuel moisture than all other slope directions (Rollins et al., 2002; Dillon et al., 2011). We think it is one of the strengths of our model that it is able to simultaneously assess the relative importance of climatic predictors such as VPD and topographic predictors like Slope and Southess simultaneously.

We included the following sentence (lines 161-163) in the Data section of our revised manuscript to clarify the reviewer's comment:

> *In the Northern Hemisphere, Southness is associated with higher insolation which results in drier conditions and low fuel moisture relative to other slope directions (Rollins et al., 2002; Dillon et al., 2011).*

4. A typo in L361 : 'are modeled'

We have fixed this typo in the revised manuscript.

## References:


1. Barbero, R., Abatzoglou, J. T., Steel, E. A., & Larkin, N. K. (2014). Modeling very large-fire occurrences over the continental United States from weather and climate forcing. Environmental Research Letters, 9(12), 124009, doi:10.1088/1748-9326/9/12/124009

2. Moritz, M. A., Moody, T. J., Krawchuk, M. A., Hughes, M., and Hall, A. (2010), Spatial variation in extreme winds predicts large wildfire locations in chaparral ecosystems, Geophys. Res. Lett., 37, L04801, doi:10.1029/2009GL041735

3. Dillon, G. K., Z. A. Holden, P. Morgan, M. A. Crimmins, E. K. Heyerdahl, and C. H. Luce. (2011). Both topography and climate affected forest and woodland burn severity in two regions of the western US, 1984 to 2006. Ecosphere 2(12):130, doi:10.1890/ES11-00271.1

4. Wang, S. S.-C., & Wang, Y. (2020). Quantifying the effects of environmental factors on wildfire burned area in the south central US using integrated machine learning techniques. *Atmospheric Chemistry and Physics*, *20*(18), 11065–11087. https://doi.org/10.5194/acp-20-11065-2020