Peer-Review of "SMLFire1.0: a stochastic machine learning (SML) model for wildfire activity in the western United States"

Title: SMLFire1.0: a stochastic machine learning (SML) model for wildfire activity in the western United States
Authors: J. Buch, A. P. Williams, C. Juang, W. D. Hansen, and P. Gentine

Summary

This manuscript introduces a stochastic ML fire model, SMLFire1.0, to estimate the probability distribution of monthly fire frequencies and sizes across the WUS over 1984-2020. The SMLFire1.0 model captures the interannual variability and the distinct multidecade increases in the annual area burned over different ecoregions over the WUS. The SHAP is applied to evaluate the importance of predictors and the authors find that vapor pressure deficit (VPD) is the dominant driver of fire frequencies and sizes across the WUS, followed by 1000-hour dead fuel moisture (FM1000), total monthly precipitation (Prec), mean daily maximum temperature (Tmax), and the fraction of grassland cover in a grid cell.

Overall Feedback

This is another work applying ML to predict fires, but this time more focusing on the WUS where fires are more connected to human activities. This take on the ML-based fire model is unique and interesting. The ML framework used here is standard, though the advantage of SMLFire1.0 compared to the gradient-boosted tree model needs more explicitly described and discussed in the main text. For instance, one of the advantages of SMLFire1.0 is uncertainty quantification. What source of uncertainty is evaluated? If it is only statistical uncertainty, would it be model-dependent? How does this model account for the spatiotemporal variability of the predictors and their non-linear interactions and why the other models couldn't? The cited work Wang et al., 2021 also considered those features in their XGBoost model and SHAP analysis. The interpolation of SHAP values also needs to be more carefully reviewed. Such as human predictors can introduce contrast impacts even at the same location. Higher population density can cause both fire ignition and suppression, a small SHAP value would overlook the impact of this predictor. For these reasons, and others mentioned below, **major revisions** to the manuscript are needed before possible publication.

Major Remarks

Selection of predictors

1. The meteorological predictors were obtained from multiple data sources. The inconsistency between the data sources may introduce additional uncertainties. Why are the daily and X-day minimum and maximum temperatures extracted from different sources? The fire weather index is calculated from relative humidity and wind speed from gridMET. Both variables from UCLA-ERA5 were used as individual predictors. Why not use the same data source to derive FFWI?

2. In addition to monthly mean daily maximum and minimum temperature, what physical information can the X-day mean variables add in? Is X-day mean calculated from the running average?

3. The predictors are selected with physical meanings. Could the authors elaborate on why a variable is chosen?

4. **Line 162**: Table S2 lists 30 predictors. This number of predictors is more close to the one after iteratively dropping off predictors that do not improve overall performance and are highly correlated as the authors mentioned in the results section. If two variables are highly correlated, which one will be kept? I would also move this to the method section.

5. Since the authors found that the antecedent precipitation is one of most the important drivers affecting plant growth, would it be helpful to include vegetation predictors that more closely connect to fuel conditions? Meanwhile, the results show low importance of the spatial variability of vegetation predictors, however, the temporal variability is important, which appears to conflict with the justification of using time-invariant biomass.

6. Surprisingly, human predictors are not among the top 10 predictors for fire frequency, while over 90 % of the California fire ignitions were associated with human activities (e.g. Balch et al., 2017).

Explanation of SHAP values

1. **Line 303-304** As far as I know, the Kernel SHAP also makes an assumption about feature independence. If two features are highly correlated, one value will be replaced with random ones from the background dataset, and then SHAP will generate predictions based on the new datasets while making the SHAP value estimation less reliable.

2. **Figure 11-Forests** shows increases in both grassland fraction and above-ground biomass increase the burned area. In the forests, the above-ground biomass is mainly contributed by wooden biomass. Therefore, I am wondering if grassland fraction and biomass will increase simultaneously. Could the authors plot a partial dependence plot for those two predictors? How to understand the higher T_{max} would suppress fire spread? When replacing VPD with relative humidity relevant variables, the relationship between T_{max} and burned area becomes positive (Figure S10-lower left panel). Would the correlation between the predictors affect the results?

Specific Remarks

1. A stochastic ML fire model is introduced. I am wondering if any physical processes (e.g., lighting and human ignition) are included in this model?

2. Mapping the Ecoregions also on Figure 1 would much help the reader to understand their locations. The “desert” is not a place closely related to fires, considering using an alternative name for this division.

3. **Line 125**: What’s the spatial resolution of ERA5-WRF? Suggest adding spatial resolution to Table S2.

4. **Line 153**: The variable Popdensity actually measures the distance to human settlements. However, the Popdensity sounds like population density which is inversely proportional to the distance. Suggest changing to Pop_dist or similar terms.

5. **Line 221-224**: While I am okay with this approximating, the MTBS dataset provides the extent of fires 1000 acres or greater in the WUS, which might be more precise than approximating each fire as a circle.

6. **Line 235-238**: Can the MDN predict burned area at each grid cell directory or for individual fires? If yes, why not sum up burned area at each grid cell or individual fires to get the Ecoregion level burned area?

7. **Line 235**: What does the spatialtemporal scale mean here? How can the CCDF plot support the breakpoint selection based on this definition?

8. **Line 242**: Should this be “consecutive time period”? Please describe how a breakpoint will be determined in the method.

9. **Line 310-315**: As stated in **Line 189**, 17489 grid cells correspond to active fires. Is it associated with the same value as stated here “ ˜20000” test points?

10. **Line 361**: “are modeled”.

11. **Line 451**: Which fire frequency is used in the follow on analysis?

12. **Line 449-500**: Was this statement based on all the 28 predictors or the top 10 shown in Figure 12? Increasing the distance to the camp ground does not seem to increase the burned area.

13. **Line 591-592**: The slope variables are also invariant with time.

14. **Line 603-604**: What does the “fire month” stands for?

Reference

Balch, Jennifer K., Bethany A. Bradley, John T. Abatzoglou, R. Chelsea Nagy, Emily J. Fusco, and Adam L. Mahood. "Human-started wildfires expand the fire niche across the United States." Proceedings of the National Academy of Sciences 114, no. 11 (2017): 2946-2951.