Earth Syst. Sci. Data Discuss.,
https://doi.org/10.5194/essd-2017-136-AC1, 2018

Open Access Earth System Science Data Discussions

Interactive comment on "A weekly, near real-time dataset of the probability of large wildfire across western US forests and woodlands" *by* Miranda E. Gray et al.

Miranda E. Gray et al.

miranda@csp-inc.org

Received and published: 12 April 2018

R = Reviewer Comment, A = Author Response

Review of "A weekly, near real-time dataset of the probability of large wildfire across western US forests and woodlands" by Gray et al. https://doi.org/10.5194/essd-2017-136

R: This is a very ambitious effort to collate numerous disparate databases, develop a model of large fire probability, and create a real-time system for distributing and updating data. I applaud the authors for taking on this challenge. That said, there is

Printer-friendly version

Discussion paper

substantial room for improvement with respect to the documentation of the modeling effort that should be undertaken before the paper is accepted. I also believe that a more compelling set of figures could accompany the figure to elucidate its utility.

Specific comments/considerations 1. The modeling framework is poorly described. The authors talk a lot about the data inputs, but it's not transparent how the model was created (other than RF), whether a single model was used for the entire region, how exactly the 10-different models compare, which variables of those selected were significant predictors.

A: We've added text to make it more transparent how the models were created (Starting pg 4, line 1). We've also added more to the 'Dataset Evaluation' section to address how the 10 different models compared and variable importance. Briefly, using all training data from 2005-2015, we compared models and extracted variable importance. Across the 10 models, overall accuracy was consistently from 0.7 - 0.72 and AUC was consistently from 0.78-0.79. More importantly, we examined the main differences across models in the top 20 important variables. Human modification or distance to urban development, the mean BI, and the lowest percentiles of NDWI were consistently in the top 5 variables. Topographic roughness and mean ERC were consistently in the top 10 variables. Slope, LST, mean 1000-hr FM, the highest percentiles of NDWI, precipitation of the warmest month, and EVI (both short- and long-term) were consistently in the top 20 variables. Other variables showed up inconsistently in the top 20 variables. Please see pg 7 and 8, lines 29-31 and 1-6.

R: The validation of the data/model is poorly described making it challenging to assess whether the model is "good". This needs to be improved upon for the credibility of the dataset.

A: As indicated, we added some of the above to the 'Dataset Evaluation' section of the manuscript, and also clarified the evaluation protocol on an independent testing dataset of 2015-2016 fires. See pg 8 lines 7-19. We also revised Figure 3, which we

believe will add to the credibility of the dataset.

R: It is a bit challenging to understand the mismatch in spatial scales that define a large fire (400ha) and the spatial scale of output (250m). Technically, a large fire would consists on a continuous group of at least 64 250m x x250m pixels. The configuration of fuels and topography would appear to be important in determining if a given pixel could be part of a large fire. This would require not only considering the fuel/topography/weather at each individual pixel, but neighboring pixels. I don't know how this is dealt with in the model.

A: A moving window analysis was used to take the mean values of predictor variables in a circular kernel with a radius of 1135 meters. This results in a window size that is approximately the size of a large fire (i.e., 405 ha), and assumes that this entire area has an influence on whether the focal pixel is likely to burn (e.g., fire may spread from a distant source). Please see pg 5, lines 13-15.

R: Technical Corrections/Considerations

Line 25, I am unsure whether it is appropriate to provide links to cloud storage or GEE image collections that may be temporary and not available in a few years time. I think the link to the doi is appropriate, and would suggest keeping the links to the other sources in the main text. This is for the journal to decide though.

A: Our expectation is that these links will remain up-to-date as long as the doi is current, and so have retained them in the abstract and main text. We will leave it up to the journal/editor to also comment on this issue.

R: The first paragraph is a bit rough and could use revision. Specifically you are contrasting the research and operational needs that operate at different spatiotemporal resolutions and highlighting the different data needs. I think the point should be made that while many datasets exist to support such efforts, and effort to synthesize and model large fire risk using an empirical framework at such fine-scales is a gap.

A: We agree and have revised the first paragraph to make the distinction between research and operational needs clearer.

R: Line 9, Page 2, insert “moderate temporal resolution (weekly)” to emphasize that you are not only modeling at high-spatial resolution.

A: Done. Pg 2, lines 7-8

R: Line 16, Page 2, replace “misspecification of these parameters” to “assumed model parameterizations”

A: Done. Pg 2, line 23

R: Line 19, Page 2, I believe that models of fire spread potential are run operationally on fires with FlamMap. It might be best to specifically address the upshot of the proposed modeling platform here over FlamMap (e.g., different goals, spatial extent).

A: Yes, fire spread potential is run operationally on fires with FlamMap and other programs (e.g., Farsite and FSPro). We’ve specifically addressed this and any differences with our data product on pg 2, lines 11-19.

R: Line 29, Page 2, “show that both long-term normals and variability in climate and vegetation. . .”

A: Done. Pg 3, line 4

R: Line 31, Page 2, It might be useful to specify how flammability is being used here and throughout. It could refer to fuel dryness, but also fuel abundance, or their combination. Long term climate exerts an influence on biomass production and the biogeography of vegetation and hence sets the stage for fuel in addition to average fuel-moisture which limits fire. Also, it might be useful somewhere in the text to specify the timescales of “short-term weather”. To some, this is on the matter of hours-to-days in terms of wind-driven fire. In this paper it likely refers to sub monthly timescales related to vegetation dryness.

A: Thank you for catching this. We have clarified our use of the term 'flammability' and have operationally defined our use of the term 'short-term weather' (weekly timescales), on pg 3, lines 6 and 11.

R: Line 32, Page 2, Rather than say flammability here, I think you really mean biomass or fuel abundance.

A: That is correct. We have clarified on pg 3 line 8

R: Line 8, Page 3, Does ignition refer to anything that would potential spread fire into a pixel (ignition, or fire spreading into region)?

A: Yes, because the response variable included both small-fire ignitions and pixels that were part of a large fire, we meant ignition to include both ignition and fire spreading into a region, and have clarified on pg 3 line 18.

R: Line 12, Page 3, It is probably worth caveating somewhere that ignitions are not random, but adhere to specific spatial patterns tied to anthropogenic activity or lightning. It is fine to state that this is not part of the modeling framework.

A: We have added this caveat and also a citation on pg 3, lines 22-23.

R: Line 8, Page 4, How are prescribed fires excluded here? MODIS will pick up large prescribed fires. You could use the Short FPA data to exclude any large fires that might have been prescribed. Note that your seasonal window April-October will not eliminate prescribed fires as they occur in many regions in late Sep/Oct. That said, I doubt they are a sizable number given the size thresholds used for large fires.

A: Large prescribed fires were excluded from the data used in our analyses by only selecting large fire samples from the MODIS BA dataset that were within MTBS wildfires. Please see pg 4, lines 25-27.

R: Line 16, Page 4, Is this NLCD data specific to a certain year. Ideally, this could be land cover pre-fire.

A: We used the NLCD from 2001. Given our focus on fires from 2005-present, it reflects the best (and most contemporaneous) pre-fire land cover. We clarified this on pg 4, lines 30-31.

R: 13. Line 18, Page 4, Please define size threshold for small fires.

A: Done. Pg 4, line 2

R: Line 25, Page 4, I understand wanting a balanced database. But what consequences do you think there are by selecting small fires in the same month/year from ecoregions where a large fire occurred. There is a lot of spatiotemporal autocorrelation in your primary drivers of temporal variability in your model that could weaken your relationships. Presumably, you'd want random samples of small fires from ecoregions.

A: We agree that the way we had set up the sampling was overly restrictive and may have weakened some of the real temporal signal. As suggested, we only took equal samples from within ecoregions to maintain the spatial balance.

R: Line 30, Page 4, I don't complete understand how EVI and NDWI were used here. Long term averages implies averaging over years and months. But in many regions that are snow covered, you would have poor EVI data that is unrepresentative. It would also be useful to provide a basis for using these variables for live fuel moisture.

A: We have added text to justify our use of long-term EVI and NDWI, and clarified that both of these variables were masked for snow and ice before taking long-term percentile values. We were incorrect in stating that NDWI can be used directly for live fuel moisture without also accounting for the amount of vegetation, so we've provided a revised basis (with appropriate references) for using the NDWI to get at water content per pixel, as well as live fuel moisture when coupled with EVI. Please see pg 5, lines 18-32.

R: Line 20, Page 5, I think you used 1981-2010 climate normals, as this is the standard 30-year period. I am confused by the variables listed. I believe none of these are

interannual. Furthermore, if you used monthly temperature and precipitation, what additional value is there in having annual sum/averages, and CV.

A: Yes, we used 1981-2010 climate normals and have revised the text. We've also winnowed this list of bioclimatic predictors down to ones that are least correlated and have been used in previous fire-climate studies. Please see pg 6, lines 20-26.

R: Line 27, Page 5, See Boer et al. (2017) for LST as a proxy for fire danger.

A: Yes, and also Nolan et al. 2016 which we've also included in the text on pg 7, lines 5-7.

R: Line 7, Page 6, The NFDRS typically involves a fuel model. What fuel model was used and was is consistent spatially?

A: Fuel model G was assumed in the GRIDMET dataset and was spatially consistent. Please see pg 7, line 20.

R: Line 19, Page 6, How many large fires occurred in total in 2015/16? I think that by "randomly" using 400 large fires, and pulling from a number not much larger than that is a limiting factor (e.g., not very independent).

A: Approximately 400 large fires occurred across our study extent in 2015/16, and we did intend to sample all of these fires (but from at most one fire) for testing purposes. Please see clarifications on pg 8, lines 7-9.

R: Line 25, Page 8, How do you suspect the model handles non-stationarity? It was built with historical conditions.

A: We meant to imply that the models themselves can be easily updated with new historical data, without having to change the input variables. Underlying relationships may be changing– for instance precipitation of the wettest month or the average early May EVI values – but the models would simply need to be re-trained on updated datasets to integrate such nonstationarities. We have clarified this point on Pg 10, lines 21-25.

R: Table 1: Missing ERC from section 4. Table 1 figure is a bit awkward and might not be needed.

A: We have added ERC to Table 1 and also removed the figure.

R: Figure 1 caption: Use SI units. Figure 1 itself shows MTBS. Why introduce MTBS here? Also, do you require a different training seed within each large fire? Note that these really won't be independent points due to autocorrelation.

A: We've switched the units in Figure 1 to SI. We believe that with revisions to the manuscript, it is clearer why we introduced MTBS here, and also how the different training seeds were used to build 10 different models.

R: Figure 3, I think it is best to just show a single map, but specify the exact date, and perhaps show any large fires that actually occurred.

A: Agreed - we have created a new Figure 3.

R: Figure 5, This is a poor figure. I don't understand what the 10 different points refer to, why the 0.45 cutoff is shown, and didn't realize that there was a model for small fires in addition to large fires. It also looks like this includes data through 2017. If so, please update the caption.

A: We have removed this figure from the manuscript.

Please also note the supplement to this comment:
https://www.earth-syst-sci-data-discuss.net/essd-2017-136/essd-2017-136-AC1-supplement.pdf

---

Interactive comment on Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-136, 2018.

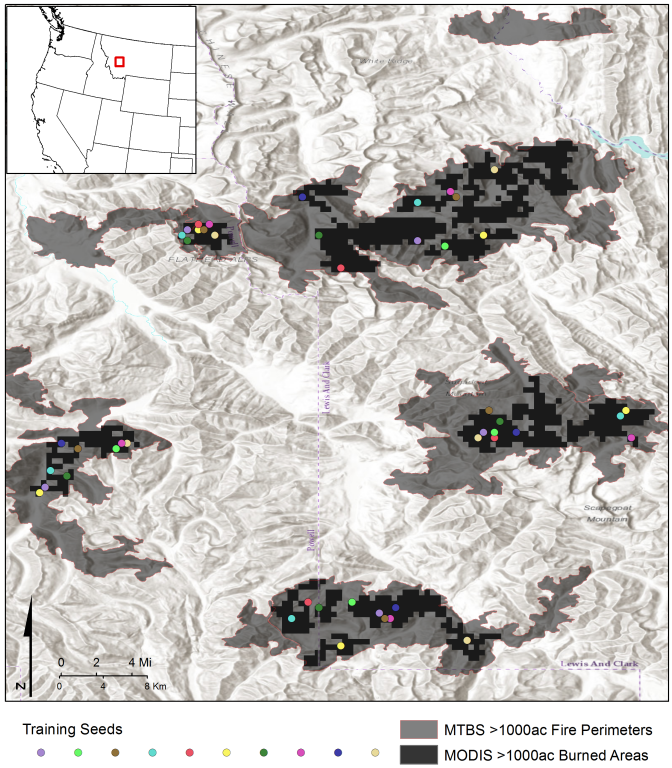

Fig. 1. Example of how the MODerate-resolution Imaging Spectroradiometer (MODIS) Burned Area (BA) dataset was used to draw 10 random sample seeds from within large fires. Each seed, across all large fires in

Printer-friendly version

Discussion paper

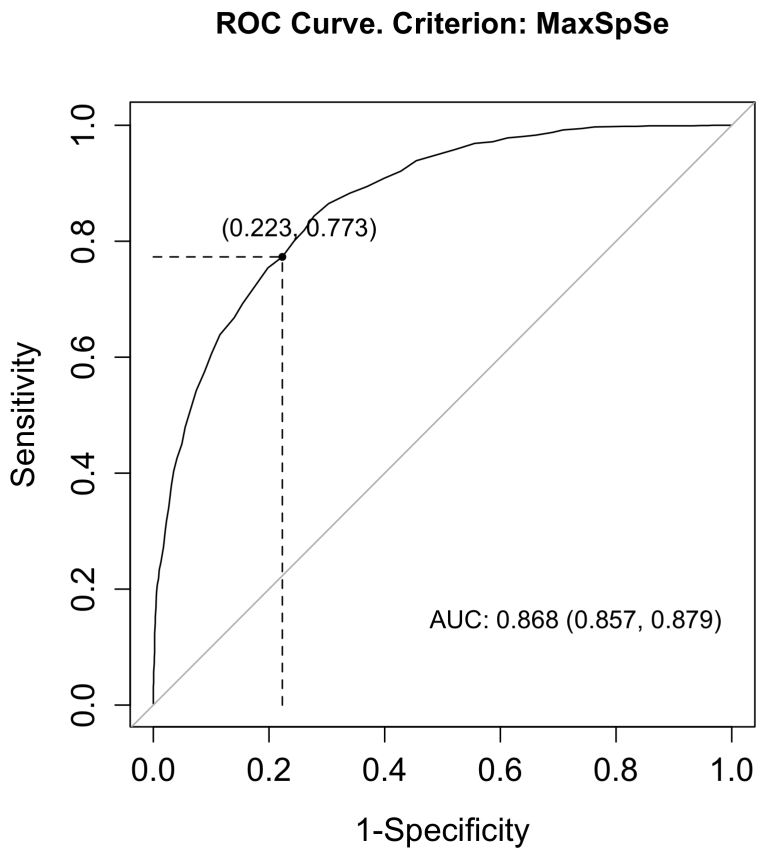

Fig. 2. Receiver Operating Curve (ROC) for an independent testing dataset of small and large fires that occurred from 2015-2016. Sensitivity and (1-Specificity) values are shown for the point where large fire

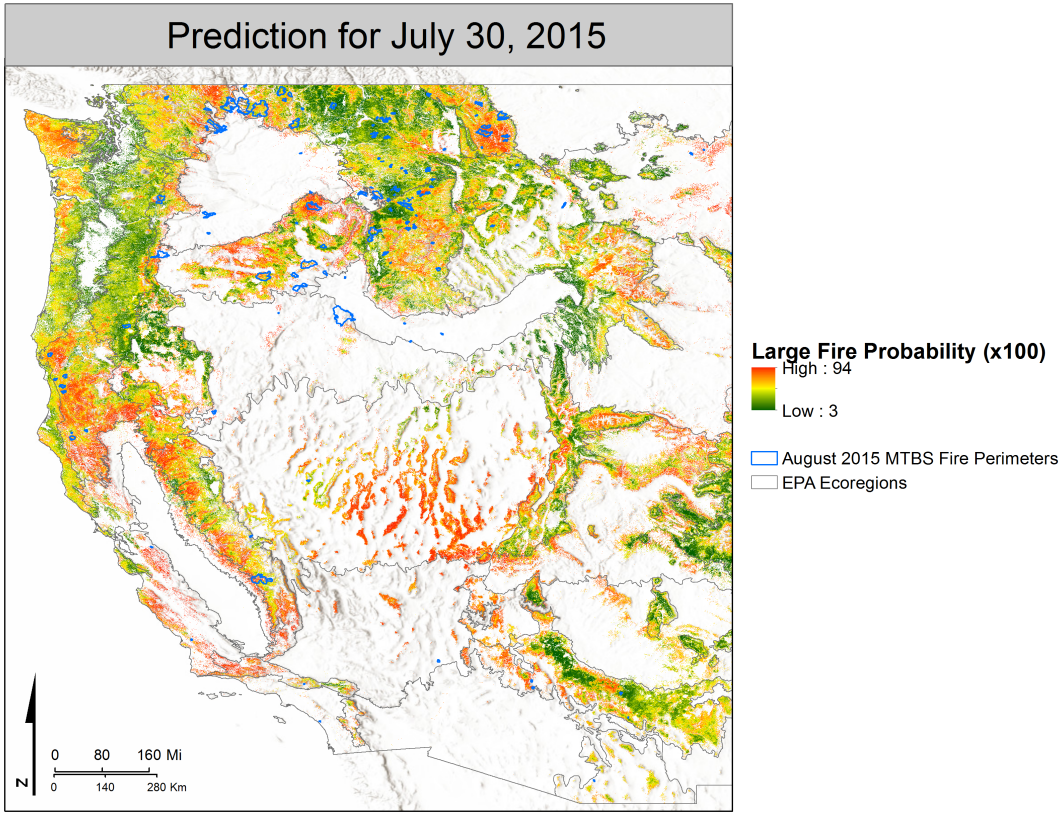

Fig. 3. Large fire probability for the week of July 30, 2015. MTBS fires greater than 405 hectares, and that started in August 2015, are overlaid on the map.

Printer-friendly version

Discussion paper

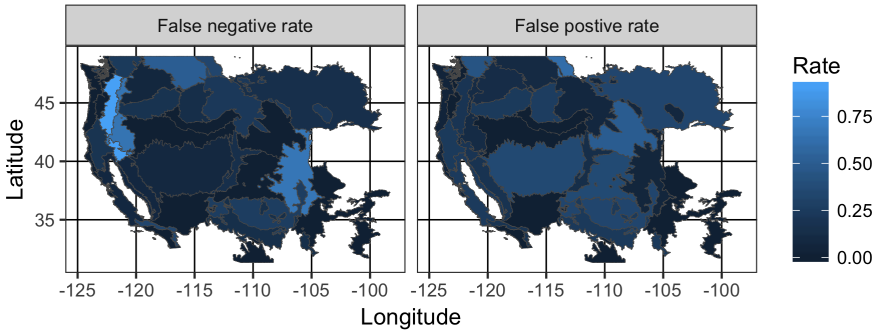

Fig. 4. False positive (FP) and false negative (FN) rates of an independent testing dataset of small and large fires from 2015-2016, mapped across EPA level three ecoregions. No testing data was available for

Printer-friendly version

Discussion paper