Biogeosciences Discuss.,
https://doi.org/10.5194/bg-2020-409-RC3, 2021

Biogeosciences
Discussions
Open Access

Interactive comment on "Quantifying the Importance of Antecedent Fuel-Related Vegetation Properties for Burnt Area using Random Forests" *by* Alexander Kuhn-Régnier et al.

Rene Orth (Referee)

rene.orth@bgc-jena.mpg.de

Received and published: 18 March 2021

Review of Kuhn-Regnier et al., bg-2020-409 "Quantifying the Importance of Antecedent Fuel-Related Vegetation Properties for Burnt Area using Random Forests"

This study investigates the role of antecedent fuel and moisture conditions for global temporal and spatial fire patterns represented by burnt area. Using a suite of random forest models with different sets of explanatory variables, the authors show that both antecedent moisture and fuel conditions are relevant for accurately modelling/predicting observed burnt areas. Thereby, the time scales extend over a few

Printer-friendly version

Discussion paper

months prior to the fire, with pronounced variations across biomes.

——————-

Recommendation: I think the paper requires moderate revisions.

This is an interesting analysis that is both relevant to the readership of Biogeosciences and a timely contribution to the ecohydrology-fire community. Legacy effects undoubtedly affect wildfire dynamics and can be a potential source of difficulties of state-of-the-art models to accurately capture fire dynamics across time scales. The machine learning approach in concert with various ecological and meteorological datasets is therefore well suited to study the underlying relationships without prior assumptions to finally provide valuable insights for the development of physically-based models. However, I have some concerns regarding the robustness of the analysis with respect to the gap filling strategy, the employed fire dataset and the relatively short analysis time period, which should to be addressed before the paper is published in Biogeosciences.

——————

General comments:

(1) While I recognize the necessity to perform gap filling for the random forest approach in this study, I do not really like the strategy. Persistent gaps are filled using minimum values which in the case of soil water index would produce artifical droughts. While I actually do not fully understand the difference between transient and persistent gaps I agree with the authors that applying a regression-based can be suitable to fill short gaps of a few months. Nevertheless, and especially for the longer gaps extending across several consecutive months I think at least the role of the gap filling for the final conclusions needs to be tested. This could be done by additionally using an alternative gap filling strategy, or by adding random noise to the gap filled values which could be scaled by the typical inter-annual dynamics of the respective month-of-year or season of the concerned metric.

(2) It is known that there are differences between fire datasets. To illustrate the robustness of the findings of this study, it would be helpful to re-compute selected key figures with the MODIS-based ESA CCI fire dataset (Chuvieco et al. 2018).

(3) I agree with the authors that the relatively short analysis time period could have a impact on the results, particularly with respect to the long legacy effects. In this context, as lightning data which limits the available time period is not employed in all experiments with 15 predictors, they could be performed with more input data covering a longer time period.

(4) I really like that different metrics are jointly used to quantify the importance of the predictors in a robust way. It would be great if the authors could add some information in the (dis?)agreement of the results between the individual importance metrics, also to inform similar future analyses.

I do not wish to remain anonymous - Rene Orth.

Specific comments:

line 8: this should be "simulated burnt area" I guess

lines 25/26: Here you could cite O et al. (2020) and/or other previous studies on related topics.

line 66: What do you mean here with "visualization techniques"?

line 117-125: So this means you are using anomalies in the case of antecedent values for DD and the vegetation productivity proxies but absolute values in the case of the current variables? While I can understand the motivation for the removal of the seasonal cycle, I feel this is inconsistent. Why not give it all the random forest model in the ALL analysis, i.e. absolute and anomaly versions of DD and the vegetation productivity proxies at current and antecedent times, and let the model decide which of these are most relevant? This would seem more objective to me.

line 139-141: Wouldn't it be more straightforward to use the OOB score for this, instead of dividing the dataset into training and validation parts while it is relatively short anyway?

line 142-143: Please add a comment why the random forest model is not re-calibrated for each experiment where different (numbers of) predictors are used.

lines 180-181: Why only the first 300'000?

line 228: I agree with the approach to focus on the most relevant predictors, but why did you decide on using 15 rather even fewer which could probably reduce overfitting even more while still preserving most of the model skill?

line 252: I think the FAPAR impact is strongest at high levels rather than intermediate levels.

lines 273-277: I think this is a particularly nice result which could be more highlighted in the abstract or conclusions.

line 324-326: Couldn't it be a solution to test the inclusion of antecedent BA as a predictor in the random forest model?

Figures 2 and 3: More colors are needed for the color bars to enable a finer distinction of the spatial patterns.

Figures 4 and 5: It could be informative to add uncertainty ranges to the curves, for example by re-running the random forest models many times.

Figure 5, caption: "LAI" should probably be removed in "First-order LAI ALEs"; furthermore it is not explained what is meant by first-order and second-order.

Figures 2-6: Please adapt the value labels of the axes and color bars to avoid the use of the exponent term to improve readability.

Figure 6: Why are there darker colors surrounding the gray area in the upperleft corner?

Figure 7: Why is there no data in the southern half of Australia?

References:

Chuvieco, E. et al. Generation and analysis of a new global burned area product based on MODIS 250m reflectance bands and thermal anomalies, Earth Syst. Sci. Data 10, 2015–2031 (2018).

O, S. et al. Observational evidence of wild fire-promoting soil moisture anomalies, Sci. Rep. 10, 11008 (2020).

Interactive comment on Biogeosciences Discuss., https://doi.org/10.5194/bg-2020-409, 2020.