Biogeosciences Discuss.,
https://doi.org/10.5194/bg-2020-409-AC1, 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.

Alexander Kuhn-Régnier et al.

ahk114@ic.ac.uk

Received and published: 8 April 2021

We thank the reviewer for their constructive feedback which will greatly help to improve the quality of the paper.

Referee comments are cited in *italics* and author's responses in normal font. Responses are separated by horizontal lines.

The study tries to quantify the importance of antecedent vegetation status as drivers *of global burnt area. The study builds on previous research using random forest modelling to understand the drivers of burnt area. While the importance of antecedent fuel load and type as drivers have been indicated previously, the study is well conducted and provides a deeper understanding towards the importance of these variables. The manuscript is well written. I have a couple of suggestions which I hope will improve the manuscript further.*

The paper shows clearly that including indicators of fuel quantity and properties improves the representation of monthly burnt area. However, monthly burnt area includes the spatial pattern, seasonality and interannual variability all together. One thing I would have like to see was a figure trying to split these factors apart, so showing whether the inclusion of antecedent fuel indicators also improves the seasonality and/or IAV of burnt area. Now, everything is mixed together, and it is hard to know whether the results are caused by an overall improvement of the spatial pattern in burnt area, the improved representation of the seasonality or improved representation of IAV. The problems with representing seasonality and IAV of burnt area is one of the topics discussed in the introduction, and seemingly partly why the authors conducted the study, so it is a bit strange that no detailed results are presented on this topic.

I was surprised to read that the study only used data for 2010-2015, while the MODIS record now covers 20 years. Knowing that in fuel limited semi-arid regions wet events can be very sporadic, I wondered whether this short timespan does not limit the study too much, especially with regard to representing IAV. The authors indicate that they use the time period for which all variables are present, but a slightly more restrictive set of variables (e.g. the least important ones) might allow for a much longer timeseries to be used and hence present more robust results. This might also help to extract results regarding seasonality and IAV.

Mentioning the IAV in the introduction is a motivation for our study, but we acknowledge that our study did not allow us to look at it previously.

In our revised manuscript, we make an attempt to disentangle the improvement to the representation of IAV using an analysis of monthly data (as opposed to climatological averages) for the time period 11-2000âĂŤ12-2019 (230 months), which also expands upon the limited temporal range as requested by the reviewer. A different burnt area dataset, MODIS MCD64CMQ BA, was also used for this run. This experiment, the 15VEG_FAPAR_MON experiment, otherwise uses the same variables as the 15VEG_FAPAR experiment.

These results show decreased R2 values, which is to be expected given the higher variance of this data:

- Using random cross-validation; Test R2: 0.45, OOB R2: 0.45
- Excluding the years 2009-2012 (including 2012); Test R2: 0.38, OOB R2: 0.45
- Excluding the years 2016-2019 (including 2019); Test R2: 0.41, OOB R2: 0.45

For this temporal cross-validation either period 2009-2012 or period 2016-2019 is left out for testing.

The main relationships identified by the model did not change appreciably, however, showing that there is no temporal change that is important. The spatial patterns are dominant, since the models behave very similarly when comparing climatological and monthly data, and the main commonality between those data is the geographical pattern.

For example, the FAPAR ALE plots for the 15VEG_FAPAR (climatological) and 15VEG_FAPAR_MON (monthly, expanded temporal range) models are shown in Figures 1 and 2, respectively.

These ALEs are highly consistent. Note that the smaller variability in the second plot (monthly analysis) arises from the fact that a constant proportion of bootstrap samples is used to construct the shaded region, and since the monthly data has more samples, these random samples are larger.

Other variables are similarly consistent, e.g. DD 3M as shown in Figures 3 and 4.

While I think it is nice that the authors present the training and validation model performance results (Figure 1), I cannot deny that I am a bit worried regarding the relatively large differences in model performance between the training and the validation data. One indeed always expect some difference, but (at least for the models I have been building) the differences never get so big except when there is some important overfitting going on.

We recognise that looking just at the training (in-the-bag) and validation scores is not valid, which is why we also include the OOB scores in our revised manuscript. These provide a good indication of how the model is expected to generalise to unseen data (Fox et al., 2017). Considering the construction of the random forest models – a set of decision trees which, in principle, can fit all training data perfectly – it is expected that some overfitting will occur even after cross-validation and given that the algorithm mitigates overfitting by averaging over the predictions of many decision trees (i.e. the forest). Since our OOB scores and the validation scores are similar (see Fig. 5), this indicates that the model is not overfitting.

Furthermore, there is no agreed standard for considering whether a model is robust or not. However, values of explained variance above 60% (i.e. an R2 score of 0.6) have been used to indicate this (e.g. Thomas et al., 2014). Both the OOB and validation set R2 scores are above this threshold for the relevant models.

As an additional assurance of model robustness, we present the PFI importances computed separately on the training and test sets in Fig. 6. The fact that there is no appreciable difference between the two is an indicator of the lack of overfitting by the model, since the model has not learnt to consider certain variables more important than they actually are during training (Dankers and Pfisterer, 2020).

Some minor comments:

Title: the title reads a bit weird, at least the "for Burnt Area" part. Maybe "as drivers of burnt area", or something similar might sound slightly better?

The title has been changed to "The Importance of Antecedent Vegetation and Drought Conditions as Global Drivers of Burnt Area" to reflect the key findings more closely.

L13: "are more sensitive to current conditions" are you still talking about the length of the period which needs to be considered to account for fuel build-up, or more about fuel dryness?

We have updated the abstract to clarify that the current conditions refer to fuel dryness instead of fuel build-up.

Table 1: "End" date seems pretty arbitrary, e.g. GFED4 burnt area has been updated up to the present.

As far as we are aware, the GFED4 dataset without small fires has not been updated since we conducted our original study. Therefore, we have used the MODIS MCD64CMQ burnt area dataset to carry out modelling studies up to and including 12-2019. We have updated other variables in the table to reflect their evolving nature (e.g.

MODIS data, which is continually being updated), instead of only reflecting the state of our processed data.

Figure 2: There seems to be a couple of issues with the figure: 1) there is no separation between areas without data and no fire (e.g. Sahara compared to S-Australia). 2) There seems to be an artefact in the Iran/Afghanistan area, with a block-shape present. 3) for plot c the colors blend into each other so that it is hard to see any pattern (if it is present).

As shown in Fig. 7, we have added grey shading to indicate regions with fire data availability, but where one or more of the other datasets is not available. Light grey indicates regions where mean BA is 0, with dark grey representing regions with nonzero mean BA. The 'block-shapes' originate from the AGB dataset used, which we have indicated in all relevant captions in the updated manuscript. Note that in addition to the shown changes in this figure, the colour scheme for plot c will also be revised to make the differences more apparent, and the number of divisions adjusted in order to make the plots clearer.

L215: there are a lot of abbreviations used in the manuscript. Many are pretty obvious (e.g. MaxT), but I would suggest writing out completely the less obvious ones like DD.

We include many abbreviations, but since it is difficult to distinguish common from uncommon abbreviations, we previously included a table with definitions (Table 1). We have now improved the presentation of abbreviations in Table 1 by listing each variable along with its abbreviation on a separate line. This should hopefully make it easier to discern which abbreviation corresponds to which variable and dataset.

References

Dankers, Cord, and Florian Pfisterer. 2020. Chapter 11 PFI: Training vs. Test Data | Limitations of Interpretable Machine Learning Methods. https://compstat-lmu.github.io/iml_methods_limitations/pfi-data.html#pfi-data.

Fox, Eric W., Ryan A. Hill, Scott G. Leibowitz, Anthony R. Olsen, Darren J. Thornbrugh, and Marc H. Weber. 2017. 'Assessing the Accuracy and Stability of Variable Selection Methods for Random Forest Modeling in Ecology'. Environmental Monitoring and Assessment 189 (7): 316. https://doi.org/10.1007/s10661-017-6025-0.

Thomas, P. B., P. J. Watson, R. A. Bradstock, T. D. Penman, and O. F. Price. 2014. 'Modelling Surface Fine Fuel Dynamics across Climate Gradients in Eucalypt Forests of South-Eastern Australia'. Ecography 37 (9): 827–37. https://doi.org/10.1111/ecog.00445.

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

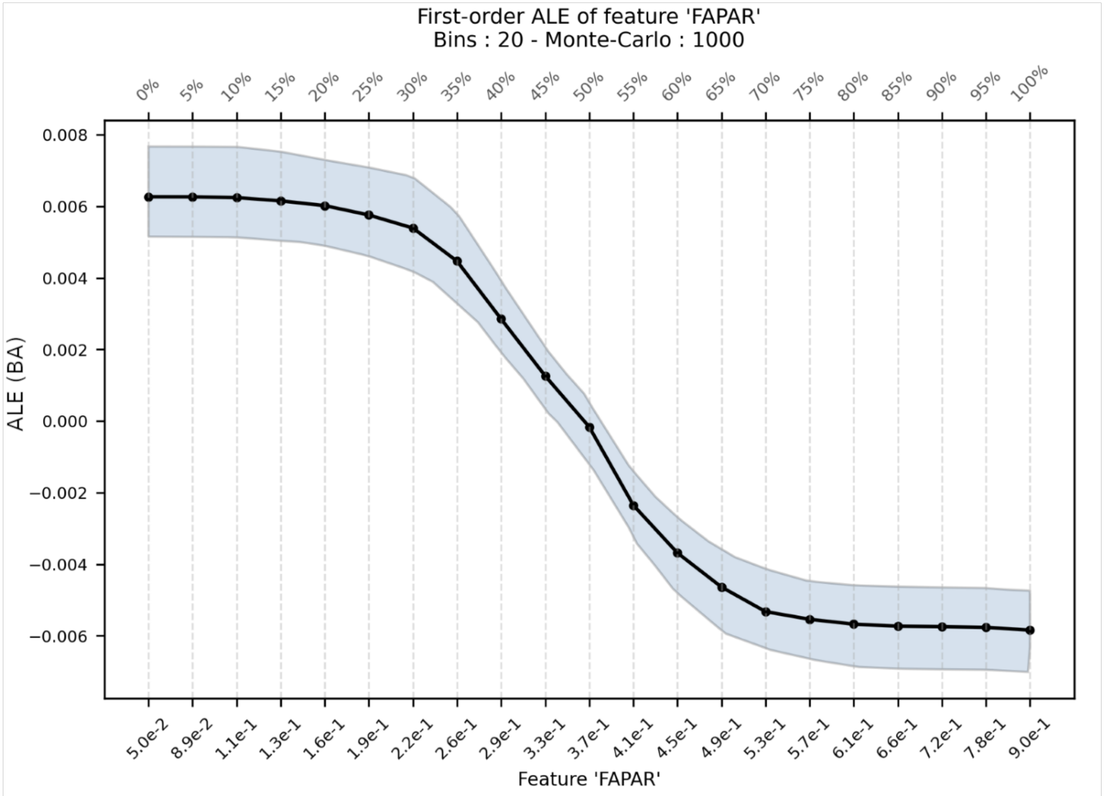

Fig. 1. First-order ALE plot showing the effect of instantaneous FAPAR on burnt area (BA) in the 15VEG_FAPAR model after accounting for all other variables.

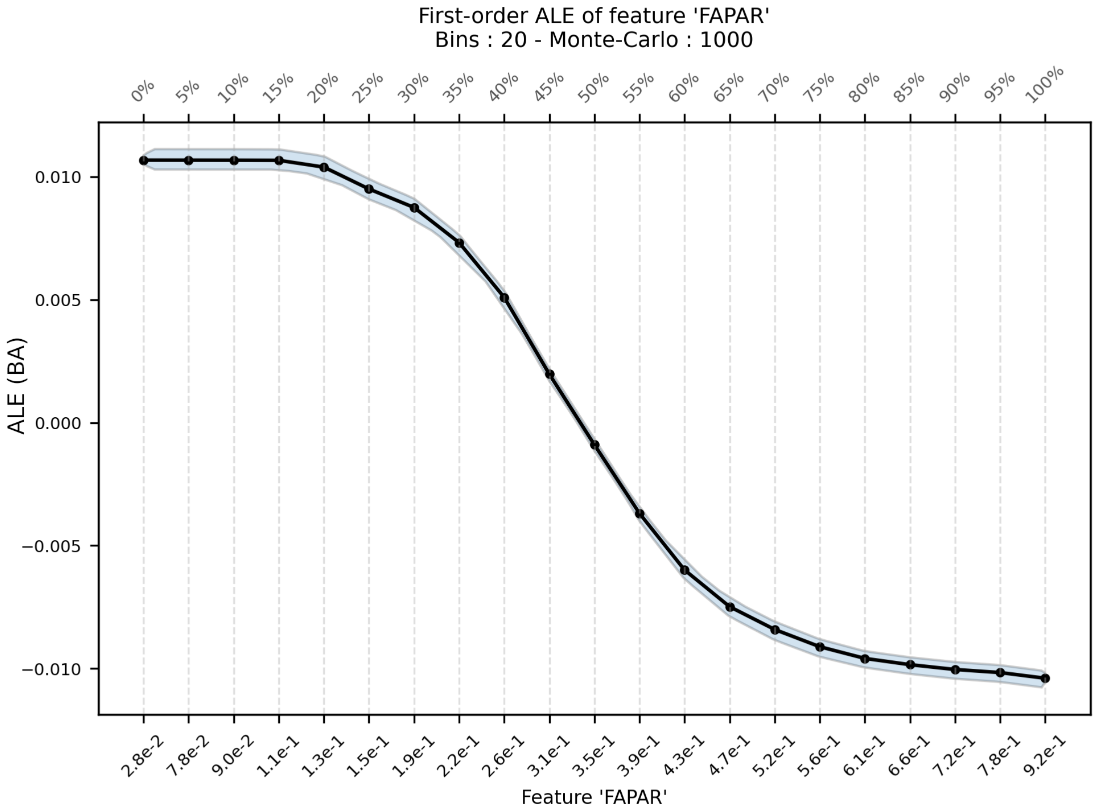

Fig. 2. First-order ALE plot showing the effect of instantaneous FAPAR on burnt area (BA) in the 15VEG_FAPAR_MON model after accounting for all other variables.

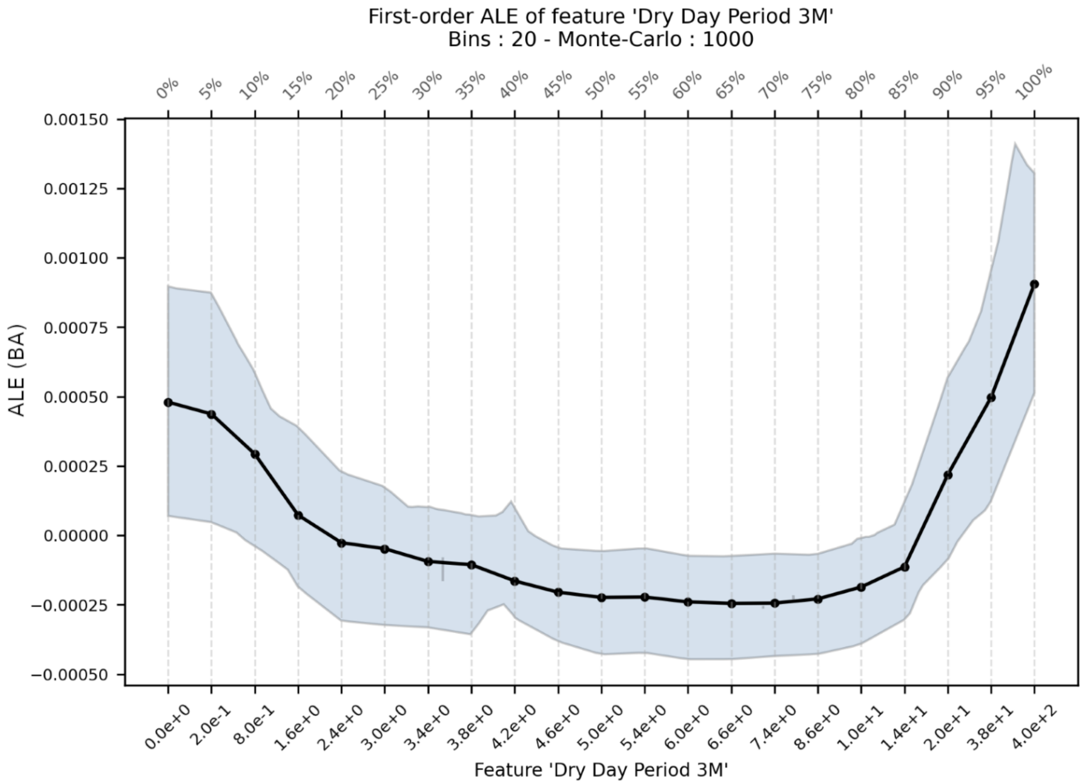

Fig. 3. First-order ALE plot showing the effect of the 3-month antecedent dry-day period (DD 3M) on burnt area (BA) in the 15VEG_FAPAR model after accounting for all other variables.

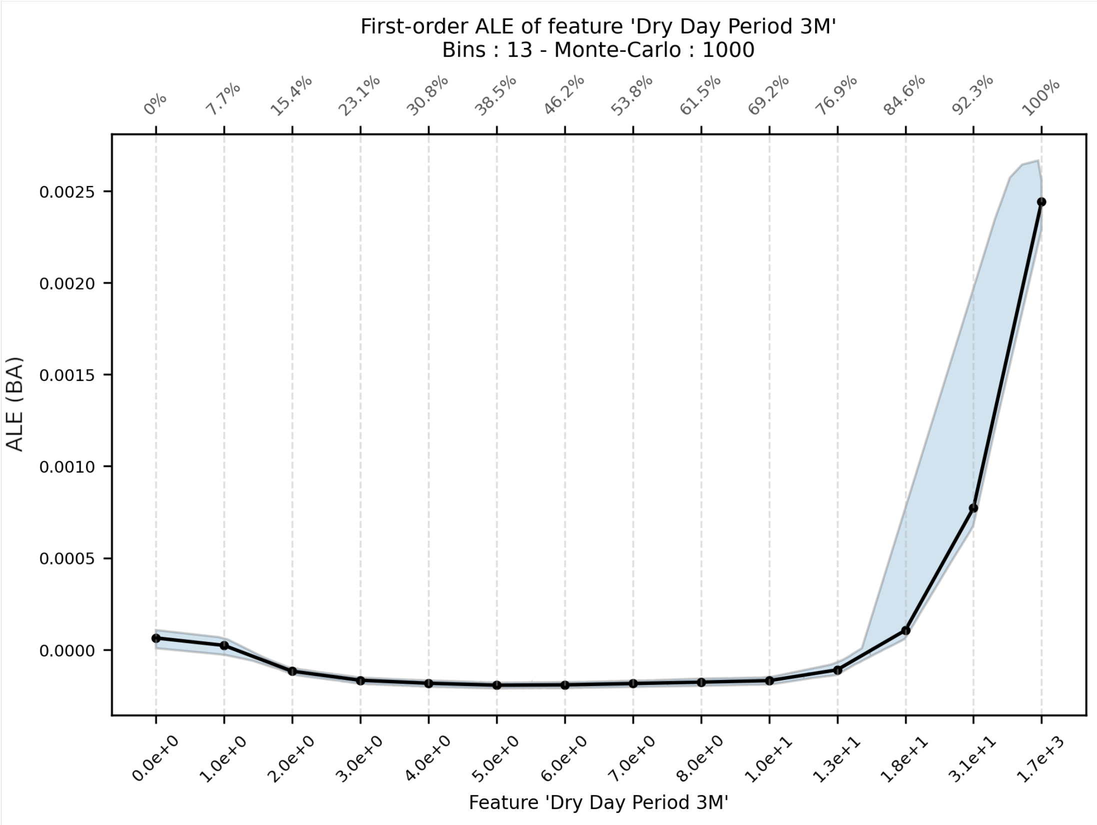

Fig. 4. First-order ALE plot showing the effect of the 3-month antecedent dry-day period (DD 3M) on burnt area (BA) in the 15VEG_FAPAR_MON model after accounting for all other variables.

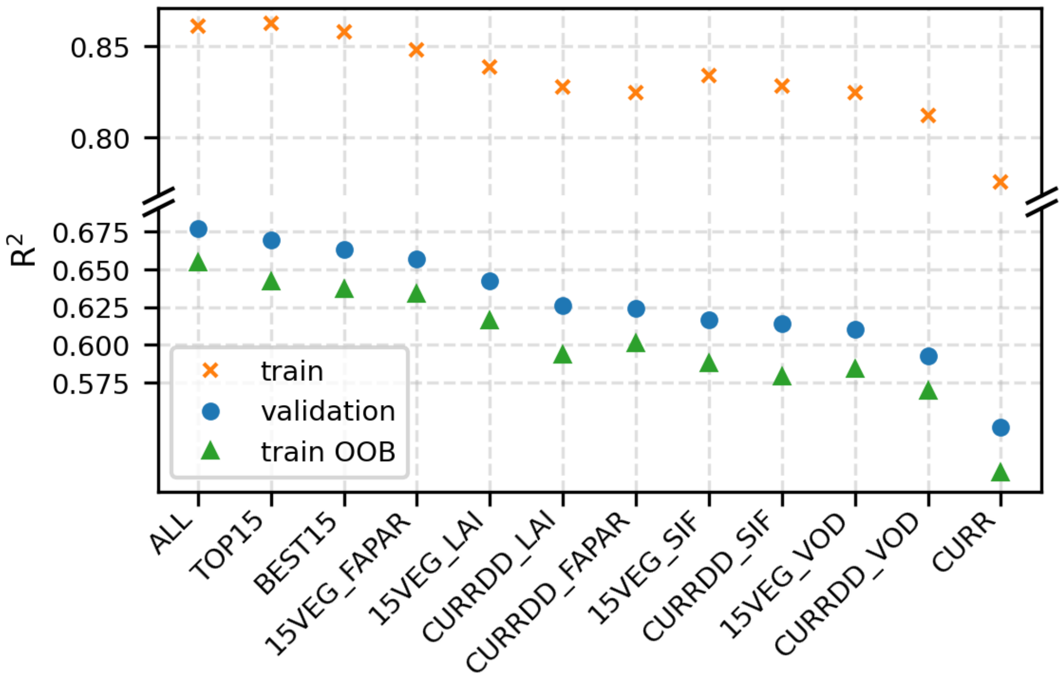

Fig. 5. Global climatological R2 scores for the different experiments. Note that although train R2 scores are shown here, train OOB scores are more indicative of model performance on unseen data.

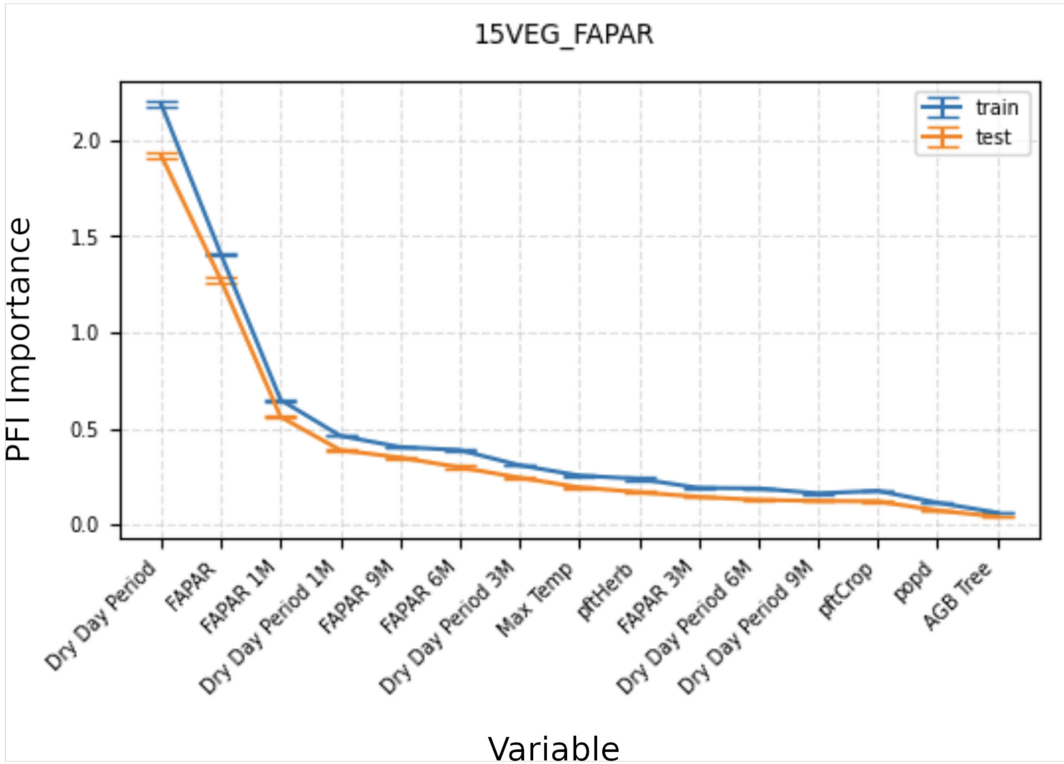

Fig. 6. PFI importances for the 15VEG_FAPAR model for the training and test sets. Note that the error bars originate from repeated shuffling of the investigated variable as opposed to different datasets.

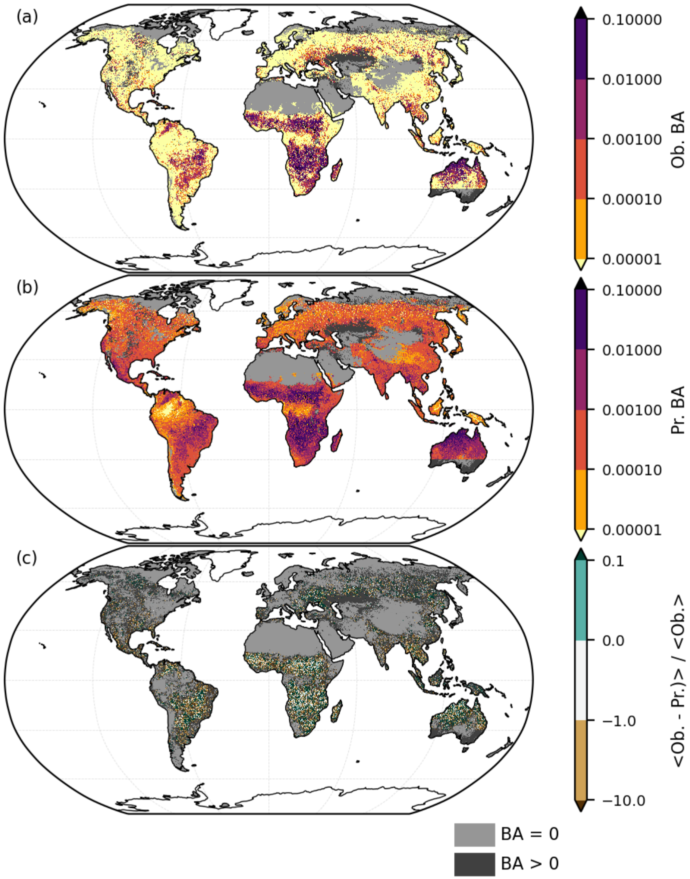

Fig. 7. (a) Average observed (Ob.) BA derived from the GFED4 BA dataset. (b) Out-of-sample predictions (Pr.) by the ALL model. (c) Relative prediction error of the ALL model.