Biogeosciences Discuss.,
https://doi.org/10.5194/bg-2020-409-AC4, 2021
© Author(s) 2021. CC BY 4.0 License.

Biogeosciences
Discussions

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

Dear René Orth,

Thank you for your thorough review which will surely enhance the quality of this work.

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

This study investigates the role of antecedent fuel and moisture conditions for global

Printer-friendly version

Discussion paper

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 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

Printer-friendly version

Discussion paper

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.

(1) The algorithm used to discern the location of 'permanent' as opposed to 'transient' gaps utilises the amount of missing information for a specific month at each location. For example, if a certain grid cell was missing data for more than 50% of all Decembers in the record, these gaps in December would be treated as a permanent gap and therefore subject to filling by minimum values. Remaining gaps are treated as transient and therefore filled using the regression model outlined in the text.

Use of a different gap filling mechanism (Fig. 1 on the left; temporal nearest-neighbour gap-filling) yields very similar results (this applies to ALE plots, too). This simple nearest-neighbour gap-filling approach works by considering timeseries at each location and filling gaps by using the temporally closest valid sample at that location. Of the two approaches, we have decided to use the season-trend model with minima filling because it represents a more physical solution, which is based on an approach previously used for vegetation variables (see Beck et al. 2006), for which one would expect minima to occur during winter.

Additionally, we have quantified how many samples were filled using minimum values outside of winter, and as can be seen in Fig. 2, virtually no samples are filled using minimum values outside of winter in the northern extreme latitudes.

Regarding the filling of SWI, this should not have a big influence on the final results because we are not using antecedent values of SWI in any case. Since we don't expect fires during the winter, having (by necessity of gap filling) unphysical values of SWI in the winter should not affect results where relevant for our analysis.

(2 & 3) In our revised manuscript, we present an analysis of monthly data (as opposed to climatological averages) for the time period 11-2000–12-2019 (230 months), which also expands upon the previously limited temporal range. 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 3 and 4, 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 5 and 6.

We expect results for the MODIS-based ESA CCI fire dataset to be consistent with the above, since it is based on the same underlying product as the MODIS MCD64CMQ dataset.

(4) The results shown in Fig. 1 are also indicative of the disagreement between the different importance metrics that were used. We will include an analogous figure (sorted by combined importance instead of Gini importance) in the updated supplementary materials, which will show the level of (dis-)agreement between the different metrics.

Specific comments:

---

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

Correct, this has now been rectified in the abstract.

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

This part of the introduction has been rewritten to include more references to previous studies of the effect of antecedent productivity, moisture, and fuel build-up on fire activity:

“A number of regional and global studies have indicated the importance of antecedent fuel build-up for BA. For example, links between fire activity and antecedent productivity have been found in South Africa (Van Wilgen et al., 2000), central Australia (Griffin et al., 1983), grass and shrublands in the western US (Littell et al., 2009; Westerling et al., 2003; Swetnam and Betancourt, 1998), NSW Australia (for bushfire fuel) (Jenkins et al., 2020), and southern Africa (Archibald et al., 2009). Global studies have identified similar relationships (a positive relationship between pre-season productivity and fire activity in the following dry season) in some dry areas. By studying the correlation between growing period (i.e. antecedent) soil moisture and fire activity, Krawchuk and Moritz (2011) found fire activity in dry regions to be related to antecedent productivity. Similarly, van der Werf et al. (2008) found a similar relationship for arid ecosystem (e.g. N. AUS), where antecedent wet conditions coupled with instantaneous drying were important. Other global studies have identified northern Australia as obeying this relationship too (Randerson et al., 2005; Spessa et al., 2005). In a more recent global analysis, O et al. (2020) found that for arid regions, wet anomalies (soil moisture) lead to increased fire later in the year by increasing fuel loads and biomass. Thus, it is clear that a better understanding of the timescales of fuel accumulation, the interaction between biophysical drivers and fuel build-up, and the effects of antecedent weather conditions on both fuel loads and fuel drying is needed in order to improve predictions of BA.”

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

Visualisation techniques here refers to the use of ALE plots and SHAP value plots in order to visualise the relationships between the predictor variables and BA and the interactions between the different predictor variables. This has been made clearer in the Introduction.

---

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.

We are only using anomalies for very large antecedent shifts $\geq$12 months. The reasoning for this was heuristically based on our desire to keep results interpretable, as opposed to improving the model's performance. This is because we expect anomalies with respect to the current year's seasonal cycle to be more interpretable as opposed to the absolute values, which are bound to be quite similar to the current year's value due to their large autocorrelation (globally, on average). Since the model is able (in theory) to capture arbitrary, non-linear, relationships between predictor variables, the inclusion of absolute values or anomalies should not affect the performance of the model (or, consequently, the feature importances) if the model has access to both the values for $\geq$12 months and current year's values (i.e. <12 months). The model should be able to compute the anomalies itself if these were indeed significant predictors (which, as it turns out, for the most part they are not). The same also applies in reverse, i.e. the model should be able to reconstruct the absolute values from the anomalies since it has access to the current year's variables which were originally used to compute the anomalies in the first place.

---

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?

We recognise that looking just at the training (in-the-bag) and validation scores is not valid, which is why we are also including the OOB scores in the updated manuscript.

We are showing OOB results (see Fig. 7) because they are a good indication of how the model is expected to generalise to unseen data (Fox et al., 2017). However, a separate validation dataset is still needed in analyses like ours (see also e.g. Joshi and Sukumar (2021)) because we are optimising the model using the training data. Therefore, even the OOB data which the model 'technically' has never seen influence the choice of variables, for example. Thus, a truly unseen dataset, the validation dataset, is reserved for a more unbiased evaluation of how the model should generalise.

---

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.

The random forest models are not individually calibrated because this would be too computationally demanding. A comment to this end has been added to the manuscript where we introduce the models and when talking about potential future work.

---

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

The SHAP interaction values were only calculated for a limited number of samples because of the very large computational effort required to do so. However, since a discussion of the SHAP interaction values is beyond the scope of the current work, we will omit this statement from the manuscript for clarity.

---

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?

While we agree that the choice of 15 predictors is somewhat arbitrary, this is heuristically based on the slope of the feature importance plots (see Fig. 8). Looking at these feature importances, where the importance change is ‘flat’ (by inspection) after 15 variables, no additional information was being conveyed. Therefore, we decided to use this as our threshold.

Given this choice of 15 variables (and some constraints as discussed in the manuscript), we also carry out an iterative approach where we investigate different sets of variables to determine which combination of vegetation variables yields the best results. This was done to balance the computationally intensive nature of the CV calculations with the ability to answer the question “Which vegetation proxy is the most important?”.

---

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

That is correct. This sentence has been rephrased to clarify the original intention: the importance of FAPAR changes most rapidly at intermediate levels of FAPAR. This is important to clarify, since it establishes that the relationship of current FAPAR and BA is described by two plateaus at extreme values, combined with a rapid transition through intermediate values.

---

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

The empirically discovered interactions have now been mentioned both in the abstract and 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?

We expect antecedent burning in previous months to be reflected in the vegetation variables used in our studies, e.g. a decrease in FAPAR following a fire. The problem we discuss in lines 324–326, however, would not be addressed by knowledge of the burnt area in previous months, because here we are concerned with burnt area within the current month and its effects on our analysis. As we discuss in the manuscript, the monthly timescales we are working on prevent us from clearly resolving the effects of fires in the current month.

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

More colours will be added to the colour bar.

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.

We originally felt that inclusion of such plots might introduce too many visual elements. However, we have now added uncertainty ranges to Fig. 4 (in the manuscript), generated by repeated bootstrapping of the training data before generating the plots. This is shown in Fig. 9 (in the current document).

However, we believe that due to the large number of curves present in Fig. 5 (in the manuscript), adding yet more elements to this plot is infeasible without an undue enlargement of the plot.

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.

That is correct and has now been rectified. The distinction between first-order and second-order ALE plots had previously been defined in the Methods section, albeit using the terms 1D and 2D instead of first-order and second-order, respectively. We have now removed this inconsistency to make this clearer.

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.

We will remove the exponent terms in the labels (see Fig. 9).

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

These darker colours are the result of the overlap between the grey areas representing missing data and the colours representing the ALE plot itself. This overlap results from the fact that ALEs are calculated for each bin edge (thus being plotted as $\frac{1}{2}$ in one cell and $\frac{1}{2}$ in the next), while the missing data applies to entire bins (i.e. between the two bin edges, which is what the tick marks on the plot correspond to). Therefore, the rectangles indicating missing data end up being shifted by $\frac{1}{2}$ of the bin size relative to the ALE data. This partial overlap is also meant to indicate that the affected ALE values are less reliable than others.

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

This spatial artifact is a result of our chosen AGB dataset. This has now been stated for all figures exhibiting these artifacts.

References

Archibald, Sally, David P. Roy, Brian W. van Wilgen, and Robert J. Scholes. 2009. ‘What Limits Fire? An Examination of Drivers of Burnt Area in Southern Africa’. Global Change Biology 15 (3): 613–30. https://doi.org/10.1111/j.1365-2486.2008.01754.x.

Beck, Pieter S. A., Clement Atzberger, Kjell Arild Høgda, Bernt Johansen, and Andrew K. Skidmore. 2006. ‘Improved Monitoring of Vegetation Dynamics at Very High Latitudes: A New Method Using MODIS NDVI’. Remote Sensing of Environment 100 (3): 321–34. https://doi.org/10.1016/j.rse.2005.10.021.

Chuvieco, Emilio, Joshua Lizundia-Loiola, Maria Lucrecia Pettinari, Ruben Ramo, Marc Padilla, Kevin Tansey, Florent Mouillot, et al. 2018. ‘Generation and Analysis of a New Global Burned Area Product Based on MODIS 250 m Reflectance Bands and Thermal Anomalies’. Earth System Science Data 10 (4): 2015–31. https://doi.org/10.5194/essd-10-2015-2018.

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.

Griffin, GF, NF Price, and HF Portlock. 1983. 'Wildfires in the Central Australian Rangelands, 1970-1980.' Journal of Environmental Management 17 (4): 311–23.

Jenkins, Meaghan E., Michael Bedward, Owen Price, and Ross A. Bradstock. 2020. 'Modelling Bushfire Fuel Hazard Using Biophysical Parameters'. Forests 11 (9): 925. https://doi.org/10.3390/f11090925.

Joshi, Jaideep, and Raman Sukumar. 2021. 'Improving Prediction and Assessment of Global Fires Using Multilayer Neural Networks'. Scientific Reports 11 (1): 3295. https://doi.org/10.1038/s41598-021-81233-4.

Krawchuk, Meg A., and Max A. Moritz. 2011. 'Constraints on Global Fire Activity Vary across a Resource Gradient'. Ecology 92 (1): 121–32. https://doi.org/10.1890/09-1843.1.

Littell, Jeremy S., Donald McKenzie, David L. Peterson, and Anthony L. Westerling. 2009. 'Climate and Wildfire Area Burned in Western U.S. Ecoprovinces, 1916–2003'. Ecological Applications 19 (4): 1003–21. https://doi.org/10.1890/07-1183.1.

O, Sungmin, Xinyuan Hou, and Rene Orth. 2020. 'Observational Evidence of Wildfire-Promoting Soil Moisture Anomalies'. Scientific Reports 10 (1): 11008. https://doi.org/10.1038/s41598-020-67530-4.

Randerson, J. T., G. R. van der Werf, G. J. Collatz, L. Giglio, C. J. Still, P. Kasibhatla, J. B. Miller, J. W. C. White, R. S. DeFries, and E. S. Kasischke. 2005. 'Fire Emissions from C3 and C4 Vegetation and Their Influence on Interannual Variability of Atmospheric CO2 and Δ13CO2'. Global Biogeochemical Cycles 19 (2). https://doi.org/10.1029/2004GB002366.

Spessa, Allan, Bevan McBeth, and Colin Prentice. 2005. 'Relationships among Fire Frequency, Rainfall and Vegetation Patterns in the Wet–Dry Tropics of Northern Australia: An Analysis Based on NOAA-AVHRR Data'. Global Ecology and Biogeography 14 (5): 439–54. https://doi.org/10.1111/j.1466-822x.2005.00174.x.

Swetnam, Thomas W., and Julio L. Betancourt. 1998. 'Mesoscale Disturbance and Ecological Response to Decadal Climatic Variability in the American Southwest'. Journal of Climate 11 (12): 3128–47. https://doi.org/10.1175/1520-0442(1998)011<3128:MDAERT>2.0.CO;2.

Van Wilgen, B. W., H. Biggs, S. P. O'Regan, and N. Mare. 2000. 'Fire History of the Savanna Ecosystems in the Kruger National Park, South Africa, between 1941 and 1996'. South African Journal of Science 96 (April). https://researchspace.csir.co.za/dspace/handle/10204/1890.

Werf, Guido R. van der, James T. Randerson, Louis Giglio, Nadine Gobron, and A. J. Dolman. 2008. 'Climate Controls on the Variability of Fires in the Tropics and Subtropics'. Global Biogeochemical Cycles 22 (3). https://doi.org/10.1029/2007GB003122.

Westerling, A. L., A. Gershunov, T. J. Brown, D. R. Cayan, and M. D. Dettinger. 2003. 'Climate and Wildfire in the Western United States'. Bulletin of the American Meteorological Society 84 (5): 595–604. https://doi.org/10.1175/BAMS-84-5-595.

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

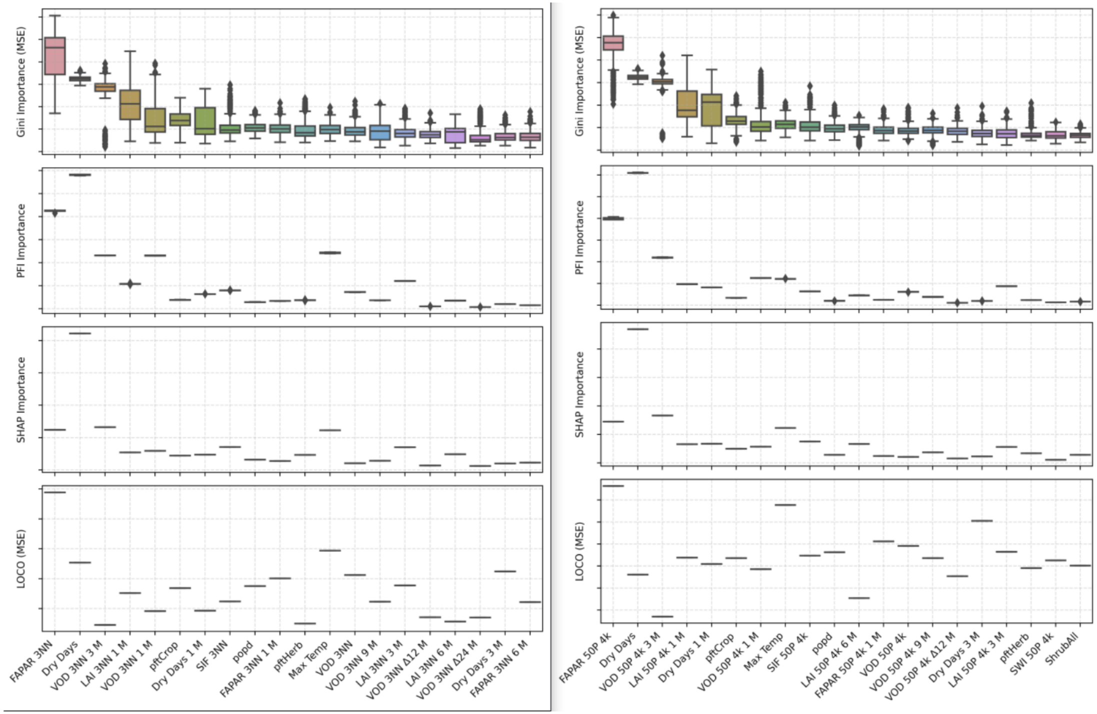

Fig. 1. Comparison of feature importances using four metrics (Gini, PFI, SHAP, and LOCO) for the ALL model using nearest-neighbour filling on the left, and season-trend model filling on the right.

Printer-friendly version

Discussion paper

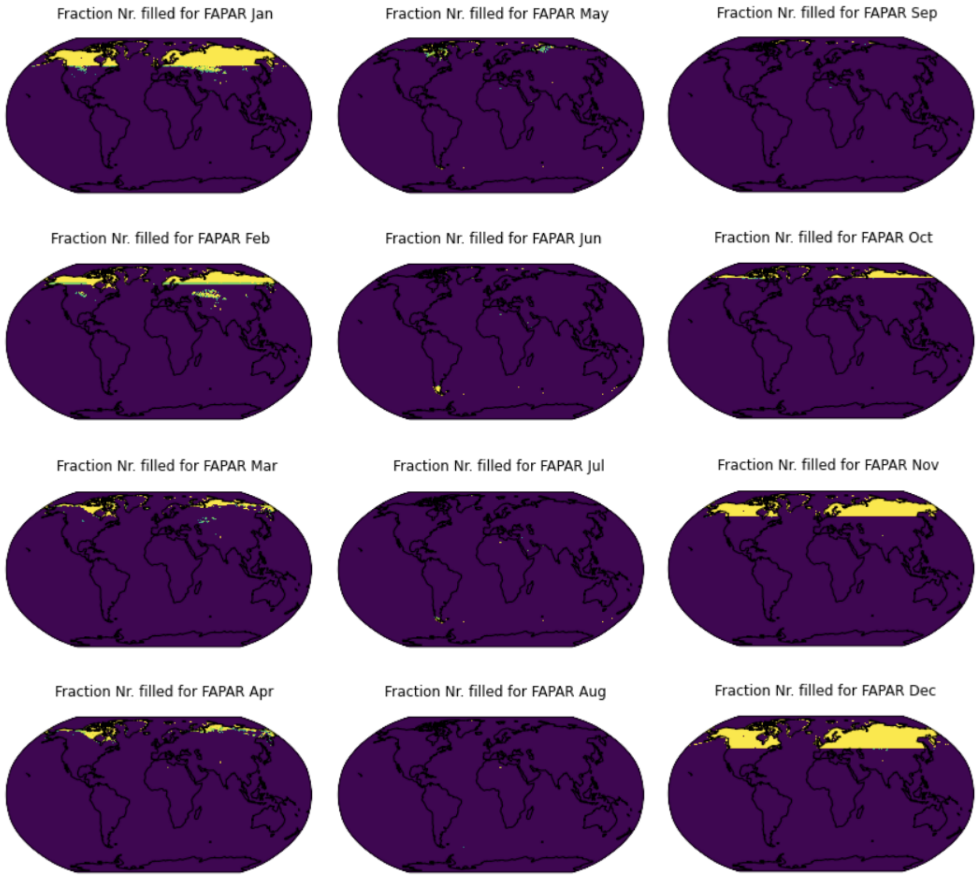

Fig. 2. The proportion of filled samples for FAPAR, with yellow indicating that all occurrences of a given month at a given location were filled and purple indicating no filling was done.

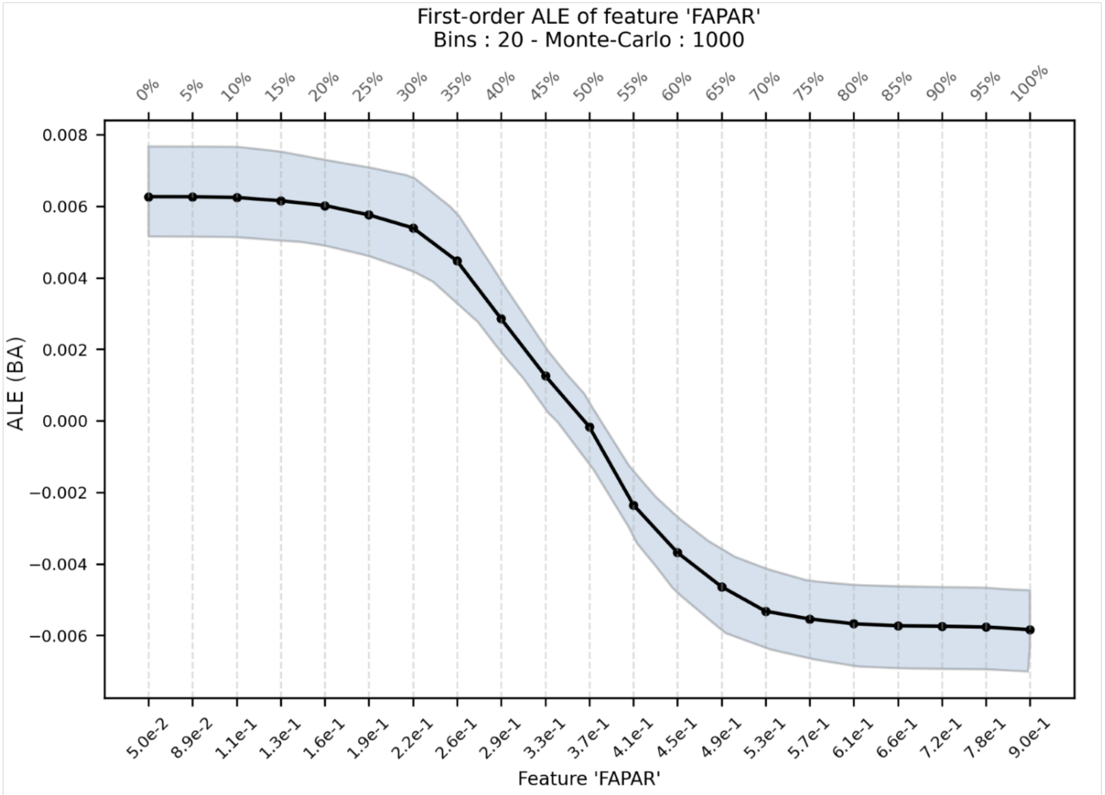

Fig. 3. First-order ALE plot showing the effect of instantaneous FAPAR on burnt area (BA) in the 15VEG_FAPAR model after accounting for all other variables.

Printer-friendly version

Discussion paper

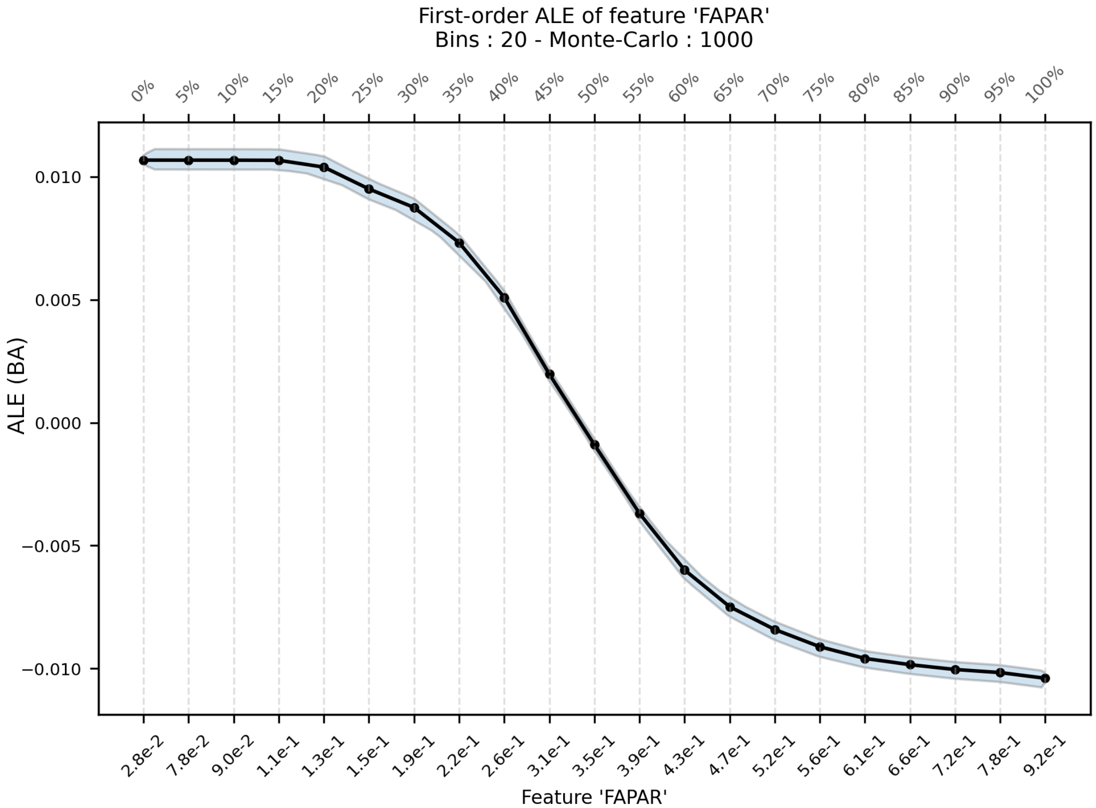

Fig. 4. First-order ALE plot showing the effect of instantaneous FAPAR on burnt area (BA) in the 15VEG_FAPAR_MON model after accounting for all other variables.

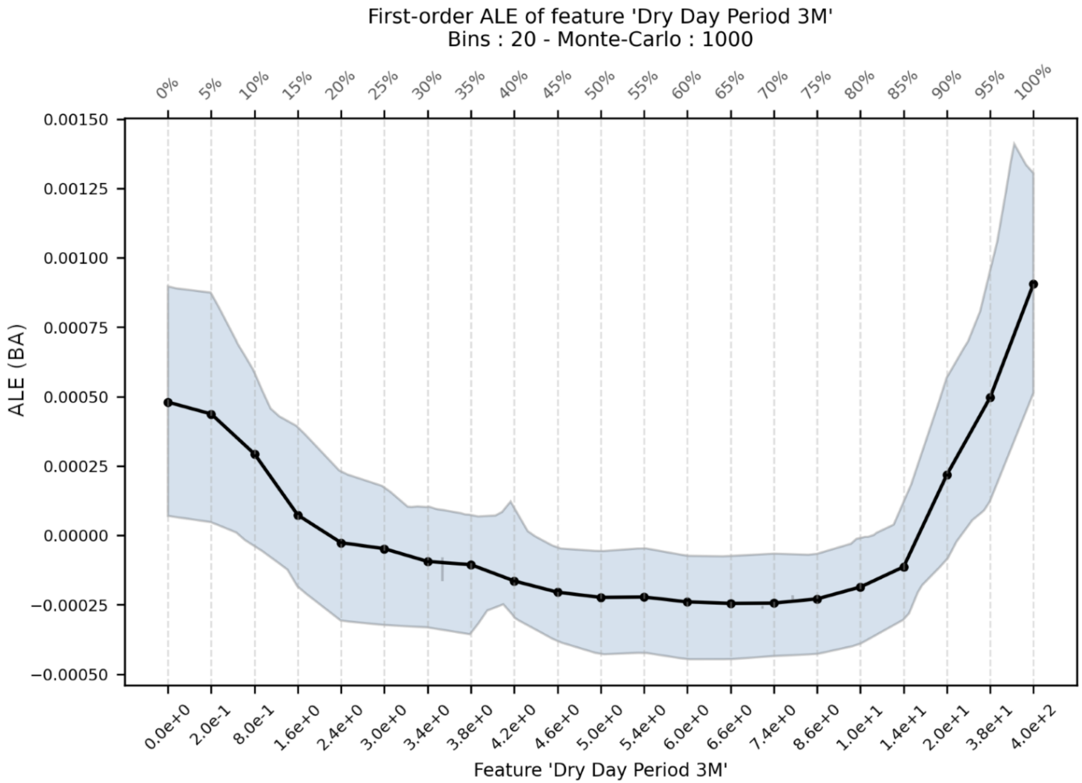

Fig. 5. First-order ALE plot showing the effect of the 3-month antecedent dry-day period (DD 3M) on burnt area (BA) in the 15VEG_FAPAR model after accounting for all other variables.

Printer-friendly version

Discussion paper

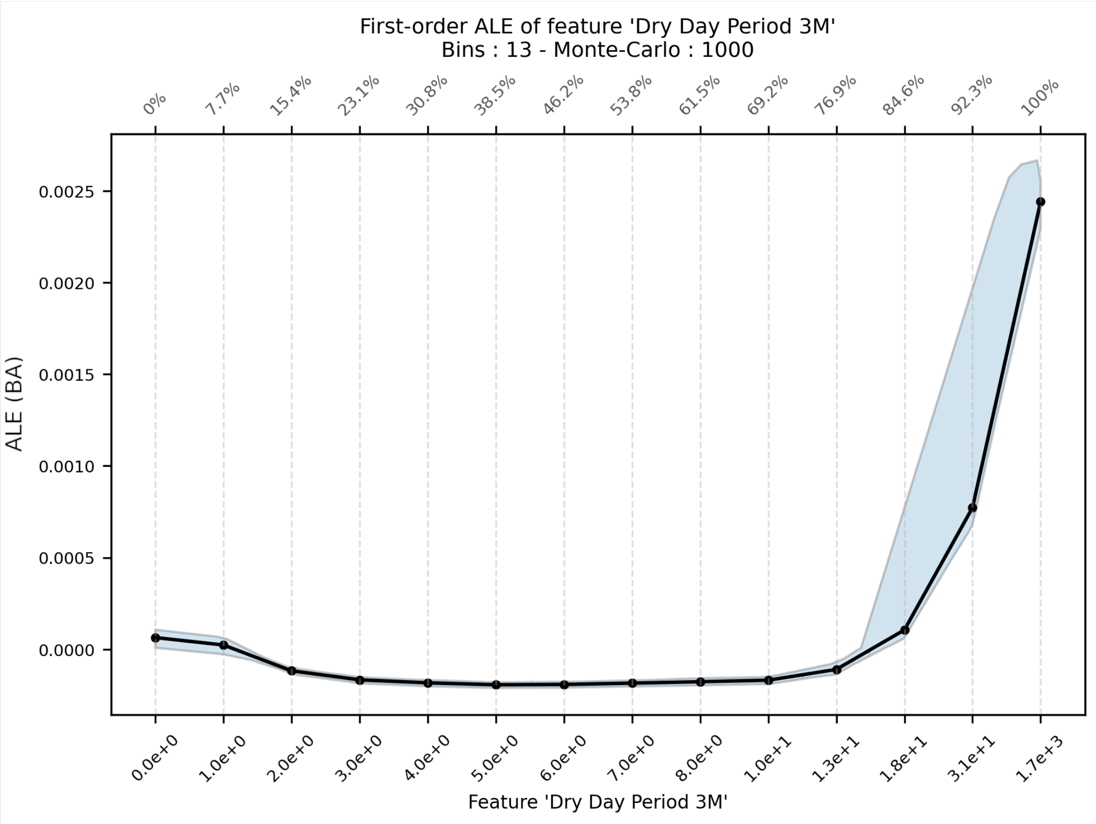

Fig. 6. First-order ALE plot showing the effect of the 3-month antecedent dry-day period (DD 3M) on burnt area (BA) in the 15VEG_FAPAR_MON model after accounting for all other variables.

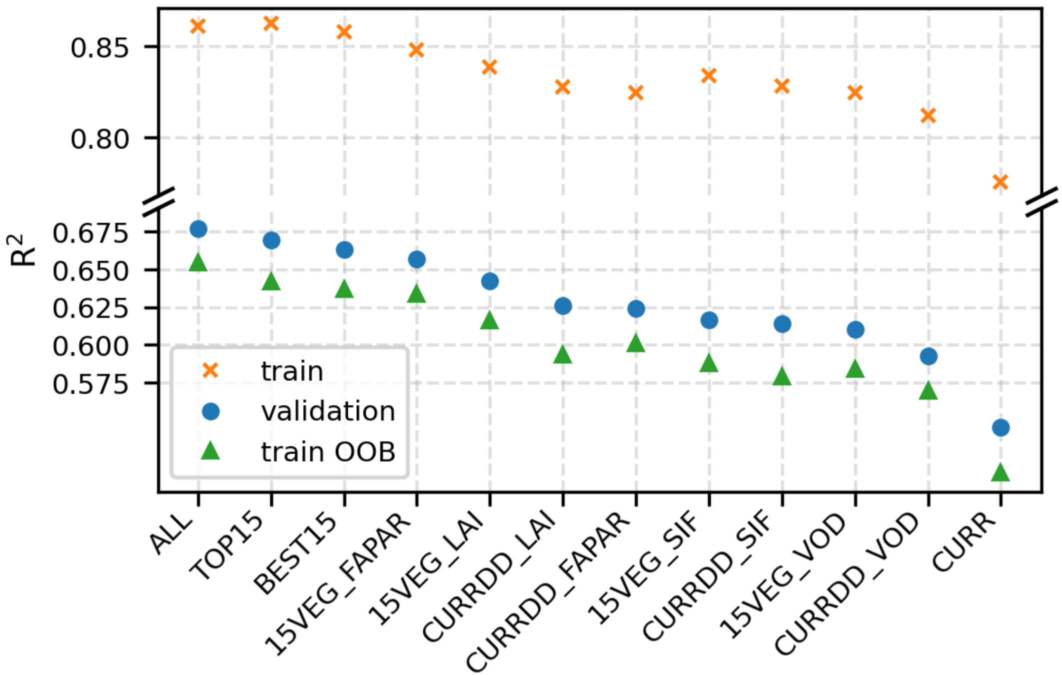

Fig. 7. Global climatological R2 scores for the different experiments. Note that although train R2 scores are shown here, train OOB scores are more indicative of model performance on unseen data.

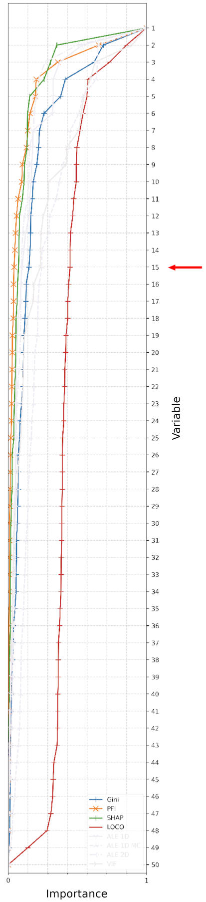

Fig. 8. Sorted variable importance metrics (Gini, PFI, SHAP, and LOCO) for the ALL model, with the highest variable importance according to each metric at the top. Note alternative metrics are greyed out.

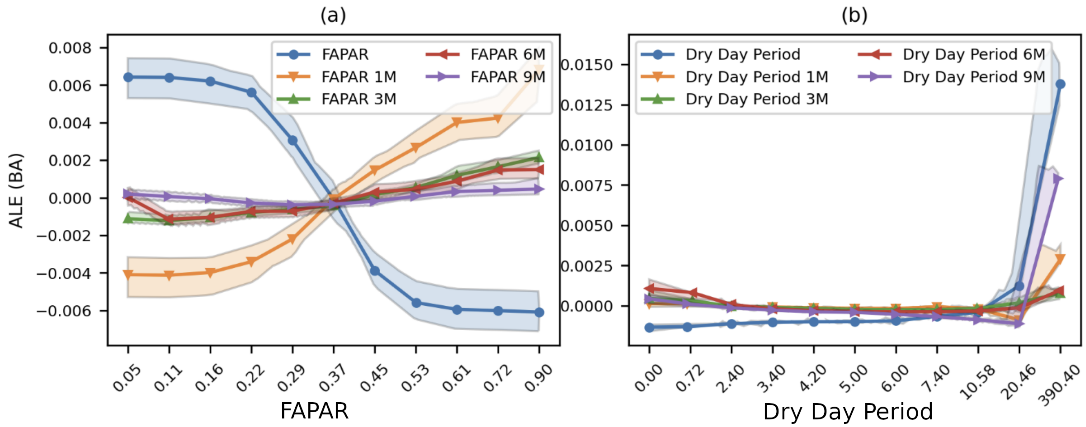

Fig. 9. First-order ALEs for different antecedent (< 1 yr) relationships with (a) FAPAR and (b) dry-day period (DD) in the 15VEG_FAPAR model, showing the underlying relationships with BA.

Printer-friendly version

Discussion paper