Response to reviews and proposed changes regarding bg-2020-409

We thank all reviewers and the associate editor for their constructive feedback.

Previous responses to reviewers' comments can be found here: `https://bg.copernicus.org/preprints/bg-2020-409/`, but some comments will additionally be treated in more depth below.

As before, referee comments are cited in *italics* and author's responses in normal font. Responses are separated by horizontal lines. However, here we only discuss comments raised by the reviewers that warrant further discussion. All other comments will be treated as set out in our previous responses (AC1, AC2, AC3, and AC4, which can be found at the link above). Further, a list of major changes concludes this part of the document, followed by the tracked changes between the manuscript versions and the new supplementary materials for reference. Figure numbers and section references in our responses below refer to the updated documents which are included at the end.

Additional responses to RC2

I am concerned about the cross-validation strategy employed here (L137. By randomly choosing the validation dataset the potential for spatial autocorrelation issues arises. This is well known in the literature (see Roberts et al. 2017; Ploton et al. 2020; Kühn and Dormann, 2012; Meyer et al. 2019). Here is a snippet from the abstract to the Roberts article: 'Ecological data often show temporal, spatial, hierarchical (random effects), or phylogenetic structure. Modern statistical approaches are increasingly accounting for such dependencies. However, when performing cross-validation, these structures are regularly ignored, resulting in serious underestimation of predictive error. One cause for the poor performance of uncorrected (random) cross-validation, noted often by modellers, are dependence structures in the data that persist as dependence structures in model residuals, violating the assumption of independence. Even more concerning, because often overlooked, is that structured data also provides ample opportunity for overfitting with non-causal predictors.'. Because the authors devote considerable space to discussion of these predictors, I think this issue is worth consideration. The authors also argue [that] the gap in R2 of the training-validation simulations gives an idea of the generalizability of the model - but that breaks down if there are spatial autocorrelation issues. Also there is some spatial structure in their biases (Fig S2) that could be coming from this issue. I would suggest adopting other CV strategies as outlined in the papers I list above.

As I am not yet convinced by their CV strategy, which is important as it impacts the results quite heavily, I suggest major revisions as I assume it will take a bit of work to demonstrate that the chosen CV strategy doesn't give misleading results.

In addition to our previous response to RC2 and as already indicated therein, we have made it clear in our updated manuscript how we assert the robustness of our model. We have included the following in the updated methods section, which expands upon our previous response:

"However, this is only valid if there is no autocorrelation between the samples. We investigated the degree of spatial autocorrelation using a variogram of global GFED4 BA which informed a buffered leave-one-out (B-LOO) CV procedure following Ploton et al. [2020]. This was carried out to determine how much the autocorrelation that may be present influences the amount of potential overfitting. We did not employ the extrapolation-prevention procedure used in Ploton et al. [2020] because it led to the exclusion of significant areas like northern Australia and west Africa. The B-LOO CV was executed as follows, where $r_{max} = 50$ pixels and N_t was chosen such that the number of potential training samples was guaranteed to be equal to or above N_t for all $r \leq r_{max}$:

1. Randomly choose a single location. The 12 monthly samples at this test location constitute the test set.
2. Exclude samples from the potential training set in a circular region of radius r pixels around the test location, such that no potential training sample is closer than r pixels to the test location. This limits the influence of spatial autocorrelation.
3. Randomly choose N_t training samples from all remaining potential training samples.
4. Using a model trained on the above training samples, predict BA for the test location.
5. Increment r and repeat steps 1–4 until r has reached r_{max}.

This process was repeated 4000 times for each of the eight linearly spaced investigated radii, with the lowest radius being equal to 0. Due to computational constraints, the B-LOO CV was only carried out for a single set of variables."

In the results and discussion section, we go on to explain:

"Using a combination of regional neural networks trained on fewer variables at a coarser spatial resolution of $1^\circ \times 1^\circ$, Joshi and Sukumar [2021] found a global R^2 score for BA prediction of 0.36. An earlier study by Thomas et al. [2014] considered an R^2 score of 0.6 as indicating a robust prediction. Our results compare favourably to both. To further ensure model robustness, we also compared the PFI importances computed separately on the training and validation sets in Fig. S6. There is no appreciable difference between the two, which is indicative of a lack of overfitting, since the model training has not unduly prioritised certain variables based on the training set [Dankers and Pfisterer, 2020]. Additionally, using the variogram shown in Fig. S7, we carried out the B-LOO CV as detailed in Sect. 2.3 in order to investigate the influence of spatial autocorrelation on the 15VEG_FAPAR model (see Fig. S8). The performance of the model drops as a larger region around the test samples is excluded (with 30 pixels corresponding to $\sim$900 km at the equator, which is the scale of autocorrelation identified using Fig. S7). However, as opposed to the case study in Ploton et al. [2020], the R^2 score plateaus at around 0.1–0.4 beyond $\sim$900 km instead of dropping to 0, thereby indicating the robustness of our model. Certain regions and extreme events are poorly captured by the model, accounting for the lower end of this range. Furthermore, the model is potentially forced to extrapolate to a larger extent as the exclusion radius is increased, leading to an overly pessimistic performance estimate. The extrapolation-prevention procedure in Ploton et al. [2020] was not used here because it led to the exclusion of certain key regions."

- L105 - I am concerned about the gap-filling approach.So for SWI doesn't this mean that it would assume drought conditions? How often would you have this condition applied (outside of winter, L100)?

We have expanded our previous response by including a more detailed explanation of our gap-filling approach and the implications for SWI in the manuscript:

"There are gaps in the SWI, FAPAR, LAI, SIF, and VOD datasets in winter months at latitudes above $\sim$60°N, and in the austral winter for southern South America, due to high solar zenith angles for FAPAR, LAI, and SIF and because of snow cover and frozen soil for SWI and VOD [see e.g. Moesinger et al., 2020]. There are also sporadic missing values in these datasets caused by e.g. cloud cover. Unfortunately, simple exclusion of the times lacking data is not possible for our analysis because we commonly rely on antecedent samples throughout. Thus, data gaps were filled using a two-step approach as in Forkel et al. [2017] in order to allow analysis of summer months at the affected locations. This approach differentiates between two gap types based on the amount of missing information for a specific month at each location.

First, 'persistent' gaps, defined as months for which 50% or more of the observations across all years are missing, were filled using the minimum value observed at that location for the given predictor variable. We assume that this indicates missing data during the winter, since other causes for data gaps (e.g. cloud cover) are predominantly 'transient'. 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 persistent and therefore filled using minima.

Second, the remaining transient gaps were filled using season-trend regression models with four harmonic terms ($k = 4$) and without breakpoints. These models were fitted using ordinary least squares regression to the entire timeseries obtained during the first step, as mentioned before using data from January 2008 to April 2015 (or November 2000 to December 2019 for the monthly analysis). Cloud cover, which also affects detection in tropical and subtropical regions, is usually transient, and therefore filled using the regression models. Locations where no observations were available for > 52 months out of the total 88 months (regardless of whether such data gaps always occur in the same month, as for persistent gaps, or at any point throughout the year) were discarded in a trade-off between data quality and geographic extent. For the monthly analysis, locations were discarded given > 138 unavailable months (out of 239 months).

Use of a different gap filling mechanism (Fig. S1b; temporal nearest-neighbour gap-filling) yielded very similar results. This simple nearest-neighbour gap-filling approach used for the eventual ALL_NN model processes timeseries at a given location, filling gaps by using the temporally closest available samples at that location. Of the two approaches, we decided to use the season-trend model with minima filling because it

represents a more physical solution; it 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. Indeed, as can be seen in Fig. S2, virtually no samples are being filled with minima outside of winter, and predominantly in the northern extreme latitudes. While our gap-filling methodology may yield unphysical values for non-vegetation variables like SWI, we do not expect the filling of SWI to have a big influence on the final results because we do not use antecedent values of SWI. Since we do not anticipate fires during the winter, having (by necessity of gap filling) potentially unphysical values of SWI in the winter should not affect results where relevant for our analysis."

Additional response to RC3

(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 artificial 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.

We have updated our manuscript in order to further explain the gap-filling procedure and demonstrate its robustness:

"There are gaps in the SWI, FAPAR, LAI, SIF, and VOD datasets in winter months at latitudes above $\sim$60°N, and in the austral winter for southern South America, due to high solar zenith angles for FAPAR, LAI, and SIF and because of snow cover and frozen soil for SWI and VOD [see e.g. Moesinger et al., 2020]. There are also sporadic missing values in these datasets caused by e.g. cloud cover. Unfortunately, simple exclusion of the times lacking data is not possible for our analysis because we commonly rely on antecedent samples throughout. Thus, data gaps were filled using a two-step approach as in Forkel et al. [2017] in order to allow analysis of summer months at the affected locations. This approach differentiates between two gap types based on the amount of missing information for a specific month at each location.

First, 'persistent' gaps, defined as months for which 50% or more of the observations across all years are missing, were filled using the minimum value observed at that location for the given predictor variable. We assume that this indicates missing data during the winter, since other causes for data gaps (e.g. cloud cover) are predominantly 'transient'. 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 persistent and therefore filled using minima.

Second, the remaining transient gaps were filled using season-trend regression models with four harmonic terms ($k = 4$) and without breakpoints. These models were fitted using ordinary least squares regression to the entire timeseries obtained during the first step, as mentioned before using data from January 2008 to April 2015 (or November 2000 to December 2019 for the monthly analysis). Cloud cover, which also affects detection in tropical and subtropical regions, is usually transient, and therefore filled using the regression models. Locations where no observations were available for > 52 months out of the total 88 months (regardless of whether such data gaps always occur in the same month, as for persistent gaps, or at any point throughout the year) were discarded in a trade-off between data quality and geographic extent. For the monthly analysis, locations were discarded given > 138 unavailable months (out of 239 months).

Use of a different gap filling mechanism (Fig. S1b; temporal nearest-neighbour gap-filling) yielded very similar results. This simple nearest-neighbour gap-filling approach used for the eventual ALL_NN model processes timeseries at a given location, filling gaps by using the temporally closest available samples at that location. Of the two approaches, we decided to use the season-trend model with minima filling because it represents a more physical solution; it 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. Indeed, as can be seen in Fig. S2, virtually no samples are being filled with minima outside of winter, and predominantly in the northern extreme latitudes. While our gap-filling methodology may yield unphysical values for non-vegetation variables like SWI, we do not expect the filling of SWI to have a big influence on the final results because we do not use antecedent values of SWI. Since we do not anticipate fires during the winter, having (by necessity of gap filling)

potentially unphysical values of SWI in the winter should not affect results where relevant for our analysis."

Also, we wish to clarify that transient gaps (as defined above, e.g. due to cloud cover) will always be filled using the season-trend regression model. This has been made clearer as in the response above.

Changes to the manuscript

Line numbers and other references in brackets below (e.g. L100) refer to the tracked changes document that is included further below.

1. Typographical errors and grammatical issues have been corrected.

2. (Fig. 6) Fig. 6 will remain in the manuscript.

3. Figure captions have been updated in order to pronounce key findings. The following sentences have been added to the captions of:

   (a) Fig. 1: "The CURR model is the only model that does not include antecedent conditions, and it performs much worse as a result. Despite the fact that the ALL model contains 50 predictors, while all other models contain just 15, it does not perform significantly better than the best models containing just 15 predictors (e.g. 15VEG_FAPAR). Note that although train R^2 scores are shown here, they are not indicative of model performance on unseen data, for which the shown train OOB scores should be used instead."

   (b) Fig. 2: "Despite the visual exaggeration of the errors, which are generally small, there is no overall pattern."

   (c) Fig. 4: "A clear difference between instantaneous and antecedent relationships can be seen in both cases, with instantaneous FAPAR limiting BA while antecedent FAPAR promotes BA, and vice versa for the dry-day period." Note that the enhancement of BA due to extreme droughts (extreme dry-day period) is apparent across time periods.

   (d) Fig. 5: "Notably, the relationship between LAI and BA is not modelled consistently by the CURR model (b), but relationships with BA are generally consistent across models otherwise."

   (e) Fig. 6: "It can be seen that the combined effect of FAPAR and FAPAR 1M on BA is positive if FAPAR is low while FAPAR 1M is high."

4. (L236) Regarding the choice of 15 predictor variables, we have updated the methods section:

   "The choice of 15 predictors was heuristically based on the slope of the feature importance plots (see Fig. S3), where, by inspection, the importance change is minimal after 15 variables. Thereafter, no additional information was being conveyed, so we decided to use this as our threshold. While use of the more rigorous recursive feature elimination with cross-validation (RFECV) would be possible in principle, this commonly makes use of the Gini importance owing to its ease of calculation, as it only considers data already seen during training. Unfortunately, this also means that RFECV fails to account for overfitting, as it only considers the training data when calculating feature importance [Meyer et al., 2019]. In contrast, the different approaches we jointly utilised to calculate a more robust feature importance metric are much more computationally demanding, making RFECV infeasible."

5. The title has been changed to "The Importance of Antecedent Vegetation and Drought Conditions as Global Drivers of Burnt Area".

6. The abstract has been updated to further pronounce the empirically discovered interactions and the fact that the current conditions relevant for the moisture-limited regions relate to fuel drying. It has also generally been rephrased.

7. Numerical output (and importance ranks) have been updated throughout the paper in response to renewed runs of the model resulting in different hyperparameters.

8. (L25) The introduction has been rephrased as detailed in AC2.

9. (L67) We have added the following paragraphs to the introduction which further detail previous relevant studies:

   "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 United States [Littell et al., 2009, Westerling et al., 2003, Swetnam and Betancourt, 1998], New South Wales, 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. northern Australia), where antecedent wet conditions coupled with instantaneous drying were found to be important. Other global studies have also identified northern Australia as obeying this relationship [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.

   While other studies have used machine learning to explore fire drivers including the effect of antecedent productivity [e.g. Archibald et al., 2009, Forkel et al., 2017, Joshi and Sukumar, 2021], they have not explored the relationship between antecedent conditions (fuel load and drying) and fire in detail. Here we quantify the roles that antecedent vegetation productivity and aridity play relative to instantaneous conditions, the critical number of months that are most important for each, the shape of their relationships to BA, and the interactions between them. While the (relative) importance of antecedent variables has been investigated before [Bessie and Johnson, 1995], we aim to quantify this on a global scale. Since other climate factors, ignitions, and human activities also influence BA, we necessarily include these factors in our analysis. The use of a machine learning approach enables us to identify non-linear relationships and interactions between the drivers. This is then combined with analysis and visualisation techniques that provide insights into the modelled relationships while mitigating the effects of correlations among variables. Such insights include the effect of a particular driver on BA and the interactions between pairs of drivers."

10. (L90) The visualisation techniques used in our work have been described in more detail at the end of the introduction:

    "The use of a machine learning approach enables us to identify non-linear relationships and interactions between the drivers. This is then combined with analysis and visualisation techniques that provide insights into the modelled relationships while mitigating the effects of correlations among variables. Such insights include the effect of a particular driver on BA and the interactions between pairs of drivers."

11. (L98) While introducing the datasets used, we now mention that "A longer time period from November 2000 to December 2019 was also considered in an analysis using fewer variables.", in order to introduce the new monthly analysis that was undertaken.

12. (L107) We have added an example dry-day calculation:

    "A period contiguous with the previous month's dry-day period was concatenated such that the sum of both (number of days) was used to determine the longest period. For example, consider a 30-day long month with a 10-day long dry-day period at the beginning of the month, followed by a wetting precipitation event on day 11, and then a dry-day period for the following 19 days. This month has a dry-day period of 19 days. However, if the previous month were to terminate in a 10-day long dry-day period, these 10 days would be added to the initial 10-day dry-day period of the current month, thereby making this combined dry-day period the longest."

13. (L114) We have clarified that the WGLC data we are using (which is based on but not equivalent to WWLLN data) mainly detect cloud-to-ground strikes:

    "We used the WGLC dataset [Kaplan and Lau, 2019] which provides counts of monthly lightning strikes. It is based on the World Wide Lightning Location Network (WWLLN) dataset, which mainly detects cloud-to-ground strikes [Rodger et al., 2004, Abarca et al., 2010], as opposed to LIS lightning data [Bürgesser, 2017]."

14. (L122) We have clarified how the AGB datasets we used were combined in regions where they overlap (which is by taking the mean):

    "Tree AGB was obtained by mosaicking AGB datasets for the tropics [Avitabile et al., 2016, 1 km resolution] and northern forests [Thurner et al., 2014, $0.01°$ resolution] using the mean after resampling each to a common spatial resolution of $0.25°$."

15. (L125) We mention our usage of an updated version of the HYDE 3.2 dataset. This is required for the new longer monthly analysis.

16. (Table 1) We have indicated datasets that are continually being updated in the updated Table 1.

17. (Table 1) We have added the MCD64CMQ BA dataset to the updated Table 1.

18. (Table 1) We have updated the table to reflect updated datasets.

19. (L185) We have used X 0M instead of X when discussing the calculation of antecedent anomalies to make it clearer that we are referring to the current variable.

20. (L199) We have updated our description of the hyperparameters used to reflect the updated state of our models.

21. (L204) We have mentioned that the number of split levels were limited both in order to limit overfitting and to enable SHAP value calculation.

22. (L259) We have included the following description of our monthly analysis in the methods section:

    "In addition to the above climatological experiments, we also investigate monthly data for the time period 11-2000–12-2029 (230 months) using the 15VEG_FAPAR_MON model. To avoid the temporal limits of the GFED4 dataset (see Table 1) the MODIS MCD64CMQ [Giglio et al., 2018] BA dataset was used. Otherwise, this experiment uses the same variables as the 15VEG_FAPAR experiment with the exception of lightning, which was replaced with the similarly significant variable AGB (see Fig. S3) in order to enable processing of a longer time period. In addition to five-fold random CV as for the other models, the performance and generalisability of this model was also measured using temporal CV. Here, the model was trained on all samples excluding either all months within the years 2009–2012 (including 2012) or 2016–2019 (including 2019). Thereafter, the R^2 was measured on whichever years were excluded for training. Note that unless explicitly specified, all following methodological descriptions will relate to the climatological experiments as opposed to the monthly 15VEG_FAPAR_MON experiment."

23. (Table 2) We have added the 15VEG_FAPAR_MON and ALL_NN experiments to Table 2.

24. (Sect. 2.4) We have rephrased Sect. 2.4 in order to explain our calculations more clearly.

25. (L314) Instead of introducing 1D and 2D ALEs, we now uniformly refer to these as first-order and second-order ALEs.

26. (L320 onwards) We have combined the results and discussion sections.

27. (Fig. 2) We have updated all maps to include grey shading that indicates regions where BA data is available, but other datasets are not.

28. (L336) We have added an explanation regarding the absence of 0 predictions by the model:

    "The model may struggle to predict 0 BA because the random forest model consists of many smaller decision trees. All 500 individual models would have to predict 0 to yield this value overall, which does not appear to occur given the stochastic nature of the training process. Failure to capture extreme events well is likely due to their rarity, resulting in the absence of comparable training data [see also e.g. Joshi and Sukumar, 2021]."

29. (L340) As indicated further above, we have included an explanation of the B-LOO CV procedure and its results.

30. (L366) We have updated our discussion of why antecedent conditions $\geq$ 1 yr. may not be important predictors.

31. (L381) Paragraphs from the previous discussion section have been moved into the combined results and discussion section.

32. (Figure 4) Figure 4 has been updated to include uncertainties using shaded regions. An inset axis has also been added to (b) to further highlight the relationships between DD and BA. The updated figure caption makes it clear that we used $\sim 10\%$ of training data 100 times to construct the uncertainty ranges.

33. (L450) We have added a reference to a recent relevant study ([Joshi and Sukumar, 2021]) in our discussion of poor predictability in boreal regions (amongst others).

34. (L455) Our description of the relationship between FAPAR and BA has been corrected to indicate that it changes most rapidly at intermediate levels of FAPAR.

35. (L457) This sentence has been moved up in order to make the paragraph more coherent.

36. (L474) The results and discussion sections have been merged.

37. (L492) Results and discussion of our new monthly analysis have been added, in order to illustrate the robustness of our modelling approach given a variety of CV scenarios:

    "Relationships between predictors and BA were also stable when considering the 15VEG_FAPAR_MON model (Fig. S12) which not only uses monthly instead of climatological data, but also a different BA dataset. Using random CV, a test R^2 of 0.501 and an OOB train R^2 of 0.498 were measured. Excluding the years 2009–2012, a test R^2 of 0.403 and an OOB train R^2 of 0.507 were measured. Excluding the final years 2016–2019, a test R^2 of 0.435 and an OOB train R^2 of 0.505 were measured. While these R^2 scores are lower than those observed for the previously discussed climatological analyses, they demonstrate that the model is able to robustly predict BA under multiple CV scenarios. Lower R^2 scores are also expected given the higher variance of this data. Additionally, the relationships identified by the model are highly consistent with the previous climatological analyses, showing that there is no temporal change that is important. The spatial patterns are dominant as the models behave very similarly when fit on climatological and monthly data; and the main commonality between those data is the geographical pattern. Note also that while lightning is omitted from this experiment in contrast to the climatological 15VEG_FAPAR experiment, lightning is also not present in the TOP15 model which performs similarly to the ALL and BEST15 models. Furthermore, as shown in Fig. S3, the importance of lightning and its replacement, AGB, are very similar."

38. (L515) A reference to van der Werf et al. [2008] has been added to contextualise the empirically discovered interactions.

39. (L535) The conclusions section has been expanded to further highlight previous studies, make the novel contributions in our paper clearer, and introduce potential future work.

40. (L626) An additional acknowledgement has been added due to the contribution of an updated HYDE dataset by Kees Klein Goldewijk.

References

Pierre Ploton, Frédéric Mortier, Maxime Réjou-Méchain, Nicolas Barbier, Nicolas Picard, Vivien Rossi, Carsten Dormann, Guillaume Cornu, Gaëlle Viennois, Nicolas Bayol, Alexei Lyapustin, Sylvie Gourlet-Fleury, and Raphaël Pélissier. Spatial validation reveals poor predictive performance of large-scale ecological mapping models. *Nat. Commun.*, 11(1):4540, September 2020. ISSN 2041-1723. doi: 10.1038/s41467-020-18321-y.

Jaideep Joshi and Raman Sukumar. Improving prediction and assessment of global fires using multilayer neural networks. *Sci. Rep.*, 11(1):3295, February 2021. ISSN 2045-2322. doi: 10.1038/s41598-021-81233-4.

P. B. Thomas, P. J. Watson, R. A. Bradstock, T. D. Penman, and O. F. Price. Modelling surface fine fuel dynamics across climate gradients in eucalypt forests of south-eastern Australia. *Ecography*, 37(9):827–837, 2014. ISSN 1600-0587. doi: 10.1111/ecog.00445.

Cord Dankers and Florian Pfisterer. *Chapter 11 PFI: Training vs. Test Data | Limitations of Interpretable Machine Learning Methods*. October 2020.

Leander Moesinger, Wouter Dorigo, Richard de Jeu, Robin van der Schalie, Tracy Scanlon, Irene Teubner, and Matthias Forkel. The global long-term microwave Vegetation Optical Depth Climate Archive (VODCA). *Earth Syst. Sci. Data*, 12(1):177–196, January 2020. ISSN 1866-3508. doi: 10.5194/essd-12-177-2020.

Matthias Forkel, Wouter Dorigo, Gitta Lasslop, Irene Teubner, Emilio Chuvieco, and Kirsten Thonicke. A data-driven approach to identify controls on global fire activity from satellite and climate observations (SOFIA V1). *Geosci. Model Dev.*, 10(12):4443–4476, December 2017. ISSN 1991-9603. doi: 10.5194/gmd-10-4443-2017.

Pieter S. A. Beck, Clement Atzberger, Kjell Arild Høgda, Bernt Johansen, and Andrew K. Skidmore. Improved monitoring of vegetation dynamics at very high latitudes: A new method using MODIS NDVI. *Remote Sensing of Environment*, 100(3):321–334, February 2006. ISSN 0034-4257. doi: 10.1016/j.rse.2005.10.021.

Hanna Meyer, Christoph Reudenbach, Stephan Wöllauer, and Thomas Nauss. Importance of spatial predictor variable selection in machine learning applications – Moving from data reproduction to spatial prediction. *Ecological Modelling*, 411:108815, November 2019. ISSN 0304-3800. doi: 10.1016/j.ecolmodel.2019.108815.

B. W. Van Wilgen, H. Biggs, S. P. O'Regan, and N. Mare. Fire history of the savanna ecosystems in the Kruger National Park, South Africa, between 1941 and 1996. *South Afr. J. Sci.*, 96, April 2000. ISSN 0038-2353.

GF Griffin, NF Price, and HF Portlock. Wildfires in the central Australian rangelands, 1970-1980. *J. Environ. Manage.*, 17(4):311–323, 1983.

Jeremy S. Littell, Donald McKenzie, David L. Peterson, and Anthony L. Westerling. Climate and wildfire area burned in western U.S. ecoprovinces, 1916–2003. *Ecol. Appl.*, 19(4):1003–1021, 2009. ISSN 1939-5582. doi: 10.1890/07-1183.1.

A. L. Westerling, A. Gershunov, T. J. Brown, D. R. Cayan, and M. D. Dettinger. Climate and Wildfire in the Western United States. *Bull. Am. Meteorol. Soc.*, 84(5):595–604, May 2003. ISSN 0003-0007, 1520-0477. doi: 10.1175/BAMS-84-5-595.

Thomas W. Swetnam and Julio L. Betancourt. Mesoscale Disturbance and Ecological Response to Decadal Climatic Variability in the American Southwest. *J. Clim.*, 11(12):3128–3147, December 1998. ISSN 0894-8755, 1520-0442. doi: 10.1175/1520-0442(1998)011<3128:MDAERT>2.0.CO;2.

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

Sally Archibald, David P. Roy, Brian W. van Wilgen, and Robert J. Scholes. What limits fire? An examination of drivers of burnt area in Southern Africa. *Glob. Change Biol.*, 15(3):613–630, March 2009. ISSN 13541013, 13652486. doi: 10.1111/j.1365-2486.2008.01754.x.

Meg A. Krawchuk and Max A. Moritz. Constraints on global fire activity vary across a resource gradient. *Ecology*, 92(1):121–132, January 2011. ISSN 1939-9170. doi: 10.1890/09-1843.1.

Guido R. van der Werf, James T. Randerson, Louis Giglio, Nadine Gobron, and A. J. Dolman. Climate controls on the variability of fires in the tropics and subtropics. *Glob. Biogeochem. Cycles*, 22(3), 2008. ISSN 1944-9224. doi: 10.1029/2007GB003122.

J. T. Randerson, 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. Fire emissions from C3 and C4 vegetation and their influence on interannual variability of atmospheric CO2 and δ13CO2. *Glob. Biogeochem. Cycles*, 19(2), 2005. ISSN 1944-9224. doi: 10.1029/2004GB002366.

Allan Spessa, Bevan McBeth, and Colin Prentice. Relationships among fire frequency, rainfall and vegetation patterns in the wet–dry tropics of northern Australia: An analysis based on NOAA-AVHRR data. *Glob. Ecol. Biogeogr.*, 14(5):439–454, 2005. ISSN 1466-8238. doi: 10.1111/j.1466-822x.2005.00174.x.

Sungmin O, Xinyuan Hou, and Rene Orth. Observational evidence of wildfire-promoting soil moisture anomalies. *Sci. Rep.*, 10(1):11008, July 2020. ISSN 2045-2322. doi: 10.1038/s41598-020-67530-4.

W. C. Bessie and E. A. Johnson. The Relative Importance of Fuels and Weather on Fire Behavior in Subalpine Forests. *Ecology*, 76(3):747–762, 1995. ISSN 0012-9658. doi: 10.2307/1939341.

Jed O. Kaplan and Hong-Kiu Lau. The WGLC global gridded monthly lightning stroke density and climatology, July 2019.

C. J. Rodger, J. B. Brundell, R. L. Dowden, and N. R. Thomson. Location accuracy of long distance VLF lightning locationnetwork. *Ann. Geophys.*, 22(3):747–758, March 2004. ISSN 0992-7689. doi: 10.5194/angeo-22-747-2004.

Sergio F. Abarca, Kristen L. Corbosiero, and Thomas J. Galarneau. An evaluation of the Worldwide Lightning Location Network (WWLLN) using the National Lightning Detection Network (NLDN) as ground truth. *J. Geophys. Res.*, 115(D18):D18206, September 2010. ISSN 0148-0227. doi: 10.1029/2009JD013411.

Rodrigo E. Bürgesser. Assessment of the World Wide Lightning Location Network (WWLLN) detection efficiency by comparison to the Lightning Imaging Sensor (LIS): WWLLN Detection Efficiency Relative to LIS. *Q.J.R. Meteorol. Soc*, 143(708):2809–2817, October 2017. ISSN 00359009. doi: 10.1002/qj.3129.

Valerio Avitabile, Martin Herold, Gerard B. M. Heuvelink, Simon L. Lewis, Oliver L. Phillips, Gregory P. Asner, John Armston, Peter S. Ashton, Lindsay Banin, Nicolas Bayol, Nicholas J. Berry, Pascal Boeckx, Bernardus H. J. de Jong, Ben DeVries, Cecile A. J. Girardin, Elizabeth Kearsley, Jeremy A. Lindsell, Gabriela Lopez-Gonzalez, Richard Lucas, Yadvinder Malhi, Alexandra Morel, Edward T. A. Mitchard, Laszlo Nagy, Lan Qie, Marcela J. Quinones, Casey M. Ryan, Slik J. W. Ferry, Terry Sunderland, Gaia Vaglio Laurin, Roberto Cazzolla Gatti, Riccardo Valentini, Hans Verbeeck, Arief Wijaya, and Simon Willcock. An integrated pan-tropical biomass map using multiple reference datasets. *Glob. Change Biol.*, 22(4):1406–1420, April 2016. ISSN 13541013. doi: 10.1111/gcb.13139.

Martin Thurner, Christian Beer, Maurizio Santoro, Nuno Carvalhais, Thomas Wutzler, Dmitry Schepaschenko, Anatoly Shvidenko, Elisabeth Kompter, Bernhard Ahrens, Shaun R. Levick, and Christiane Schmullius. Carbon stock and density of northern boreal and temperate forests. *Glob. Ecol. Biogeogr.*, 23(3):297–310, 2014. ISSN 1466-8238. doi: 10.1111/geb.12125.

Louis Giglio, Luigi Boschetti, David P. Roy, Michael L. Humber, and Christopher O. Justice. The Collection 6 MODIS burned area mapping algorithm and product. *Remote Sensing of Environment*, 217:72–85, November 2018. ISSN 00344257. doi: 10.1016/j.rse.2018.08.005.

~~Quantifying the~~ The Importance of Antecedent ~~Fuel-Related~~ Vegetation ~~Properties for~~ and Drought Conditions as Global Drivers of Burnt Area~~using Random Forests~~

Alexander Kuhn-Régnier[1,2], Apostolos Voulgarakis[1,2,3], Peer Nowack[2,4,5], Matthias Forkel[6], I. Colin Prentice[1,7], and Sandy P. Harrison[1,8]

[1]Leverhulme Centre for Wildfires, Environment, and Society
[2]Department of Physics, Imperial College London, UK
[3]School of Environmental Engineering, Technical University of Crete, Chania, Greece
[4]Grantham Institute and the Data Science Institute, Imperial College London, UK
[5]Climatic Research Unit, School of Environmental Sciences, University of East Anglia, UK
[6]Environmental Remote Sensing Group, TU Dresden, Dresden, Germany
[7]Department of Life Sciences, Imperial College London, UK
[8]Geography and Environmental Science, University of Reading, UK

Correspondence: Alexander Kuhn-Régnier (alexander.kuhn-regnier14@imperial.ac.uk)

Abstract. The seasonal and longer-term dynamics of fuel accumulation affect fire seasonality and the occurrence of extreme wildfires. Failure to account for their influence may help to explain why state-of-the-art fire models do not simulate the length and timing of the fire season or interannual variability in burnt area well. We investigated the impact of accounting for different timescales of fuel production and accumulation on burnt area using a suite of random forest regression models that included the immediate impact of climate, vegetation, and human influences in a given month, and tested the impact of various combinations of antecedent conditions in four productivity-related vegetation indices and in antecedent moisture conditions. Analyses were conducted for the period from 2010 to 2015 inclusive. ~~We showed that the inclusion of antecedent vegetation conditions on timescales > 1 yr had no impact on burnt area, but inclusion~~ Inclusion of antecedent vegetation conditions representing fuel build-up led to an improvement of the global, climatological out-of-sample R^2 from ~~0.567 to 0.686. The~~ 0.579 to 0.701, but the inclusion of antecedent ~~moisture conditions also improved the simulation of burnt areathrough its~~ vegetation conditions on timescales ≥1 yr had no impact on simulated burnt area. Current moisture levels were the dominant influence on fuel ~~build-up, which is additional to the influence of current moisture levels on fuel~~ drying. Additionally, antecedent moisture levels were important for fuel build-up. The models also enabled the visualisation of interactions between variables, such as the importance of antecedent productivity coupled with instantaneous drying. The length of the period which needs to be considered ~~to account for fuel build-up~~ varies across biomes; fuel-limited regions are sensitive to antecedent conditions that determine fuel build-up over longer time periods (∼4 months)~~and~~ , while moisture-limited regions are more sensitive to current conditions that regulate fuel drying.

1 Introduction

Wildfires are an important natural disturbance of the Earth System. They have extensive socio-economic impacts as well as profound effects on vegetation, atmospheric composition, and climate (Bowman et al., 2011; Voulgarakis and Field, 2015; Andela et al., 2017; Lasslop et al., 2019). How fire regimes may change in the future, and how fire-related feedbacks may influence climate and global environmental changes are growing concerns.

The factors that influence the occurrence and intensity of fire are well-known: the presence of an ignition source, vegetation properties that determine the availability of fuel, and weather conditions that promote fuel drying and thereby the rate of fire spread. ~~Human activities are also implicated, either promoting or suppressing fire through ignitions, fuel management, and landscape modification.~~ However, these factors are strongly coupled to one another. Climate conditions influence the incidence of lightning and the nature of the vegetation, while wind strength and the impact of atmospheric conditions on drying are modulated by vegetation cover. Furthermore, the relationships among ignitions, vegetation, and climate may depend on the timescales involved~~. Short-term~~ ; short-term drought promotes fuel drying and hence increases ~~the risk of fire~~ fire risk, but in the longer term, drought conditions reduce vegetation cover and fuel loads. This complexity makes it challenging to disentangle the causes of observed changes in fire ~~regimes. Recent~~ activity.

Furthermore, recent declines in burnt area (BA) in some regions have been explained as a consequence of human activity, through indirect and direct intervention (Martínez et al., 2009; Andela et al., 2017) albeit modulated by climate and vegetation (Forkel et al., 2019b). ~~However, it has been argued that climate may become increasingly important in the future (???Barbero et al., 2015; Goss et al., 2020), especially in the extratropics (?). Indeed, a~~ Such human intervention can promote or suppress fire through ignitions, fuel management, and landscape modification. A mainly temperature-driven increase in conditions conducive to wildfires was suggested by a number of regional studies (e.g. Westerling, 2006; van Oldenborgh et al., 2020; Goss et al., 2020; Barbero et al., 2015). At the global scale, Abatzoglou et al. (2019) showed that anthropogenic climate change had led to an increase in fire weather over 22% of the global burnable area by 2019, while Jolly et al. (2015) found that anthropogenic climate change has led to a lengthening of the fire season across more than a quarter of global vegetated land in recent decades. Increases in fire weather are predicted under different assumptions about levels of future warming (e.g. Burton et al., 2018; Turco et al., 2018; Bedia et al., 2015).

Understanding the interplay among the different present-day controls of fire is also a key requirement for the prediction of future fire-regime shifts and impacts on the land biosphere and human activities. Coupled fire-vegetation models can be used to predict changes in large-scale fire regimes in response to future climate change scenarios (see e.g. Knorr et al., 2016; Kloster et al., 2012) and to explore how these changes are affected by and will affect regional vegetation patterns and climate. Although these models are reasonably good at simulating modern geographical fire patterns in BA, they are poor at reproducing observed fire-season length and inter-annual variability (IAV) in BA (Hantson et al., 2020). Furthermore, there are large differences in their predictions of both historical (Teckentrup et al., 2019) and future (Kloster and Lasslop, 2017; Sanderson and Fisher, 2020) trends.

Studies have pinpointed the relationship between simulated vegetation properties and BA as a cause for concern ~~(e.g. ?Kelley et al., 2019; .?~~ (e.g. Forkel et al., 2019a; Kelley et al., 2019; Teckentrup et al., 2019; Hantson et al., 2020). Forkel et al. (2019a) analysed satellite data to show that while state-of-the-art ~~fire–vegetation~~ fire-vegetation models reproduce the emergent relationships with climatic variables, they do not correctly represent the relationship between vegetation and BA. Hantson et al. (2020) highlighted the need for improved understanding of vegetation drivers of fire season length and ~~inter-annual variability (IAV)~~ IAV of BA. Both ~~?~~ Forkel et al. (2019a) and Hantson et al. (2020) argued for a better understanding of how vegetation properties control fuel build-up, and therefore fire occurrence and intensity.

Fuel is organic matter that is available for ignition (Keane et al., 2001). The type, amount, and spatial arrangement of fuel affect its tendency to burn (Archibald et al., 2009). These properties, dictated by vegetation, in turn affect fuel connectivity and hence fire spread in addition to how rapidly fuel dries out and becomes combustible. ~~The accumulation of fuel over time is expected to influence fire intensity.~~ Antecedent weather conditions in the weeks to years before fire events can ~~also~~ determine fuel availability (van Oldenborgh et al., 2020) and hence fire occurrence. The effect of antecedent weather conditions on BA may depend on the types of vegetation present (which influences whether fuel drying or accumulation is most important): antecedent precipitation will increase BA in fuel-limited regions, for example, but decrease BA in regions where fuel drying is the major control ~~(Alvarado et al., 2020; Abatzoglou and Kolden, 2013). A~~ (Alvarado et al., 2020; Abatzoglou and Kolden, 2013; Littell et al., .

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 United States (Littell et al., 2009; Westerling et al., 2003; Swetnam and , New South Wales, 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. northern Australia), where antecedent wet conditions coupled with instantaneous drying were found to be important. Other global studies have also identified northern Australia as obeying this relationship (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.

~~In this study, we investigate the importance of current and antecedent conditions on BA, focusing particularly on the link of BA to~~ While other studies have used machine learning to explore fire drivers including the effect of antecedent productivity (e.g. Archibald et al., 2009; Forkel et al., 2017; Joshi and Sukumar, 2021), they have not explored the relationship between antecedent conditions (fuel load and drying) and fire in detail. Here we quantify the roles that antecedent vegetation productivity and aridity play relative to instantaneous conditions, the critical number of months that are most important for each, the shape of their relationships to BA, and the interactions between them. While the (relative) importance of antecedent variables has been investigated before (Bessie and Johnson, 1995), we aim to quantify this on a global scale. Since other climate factors, ignitions, and human activities also influence BA, we necessarily include these factors in our analysis. ~~We use~~ The use of a machine learning approach enables us to identify non-linear relationships and interactions between the drivers. This is ~~supported by~~ then combined with analysis and visualisation techniques ~~designed to mitigate~~ that provide insights into the modelled relationships while mitigating the effects of correlations among variables~~, and to provide insights into the modelled interactions~~. Such insights include the effect of a particular driver on BA and the interactions between pairs of drivers.

2 Methods

2.1 Data

The predictor and BA datasets are available for different but overlapping time periods (Table 1). We pre-processed each dataset separately and conducted random forest analyses based on the common period from January 2010 to April 2015. Monthly fractional BA for this period was obtained from the GFED4 dataset (Giglio et al., 2013) (data were retrieved from https://www.globalfiredata.org/data.html). A longer time period from November 2000 to December 2019 was also considered in an analysis using fewer variables.

Diurnal temperature range (DTR), maximum temperature (MaxT), dry-day period (DD), and soil moisture are important climate factors influencing BA ~~(Archibald et al., 2009; Bistinas et al., 2014; Forkel et al., 2017; ?; Abatzoglou et al., 2018; Kelley et al., 2019~~ (Archibald et al., 2009; Bistinas et al., 2014; Forkel et al., 2017, 2019a; Abatzoglou et al., 2018; Kelley et al., 2019) and are thus considered as predictors in our analyses. DTR was calculated by taking the monthly average of the difference between the daily maximum and minimum ERA5 (Copernicus Climate Change Service (C3S), 2017) 2 m temperatures.

The dry-day period was defined as the longest contiguous period of ERA5 mean daily precipitation below 0.1 mm day^{-1} (wetting rainfall; Harris et al., 2014; Jolly et al., 2015) within each month. ~~If this period was contiguous with a preceding~~ A period contiguous with the previous month's dry-day period was concatenated such that the sum of both (number of days) was used to determine the longest period. For example, consider a 30-day long month with a 10-day long dry-day period ~~, these were concatenated.~~ at the beginning of the month, followed by a wetting precipitation event on day 11, and then a dry-day period for the following 19 days. This month has a dry-day period of 19 days. However, if the previous month were to terminate in a 10-day long dry-day period, these 10 days would be added to the initial 10-day dry-day period of the current month, thereby making this combined dry-day period the longest.

Soil moisture was taken from the Copernicus soil water index (SWI) dataset (Albergel et al., 2008; Wagner et al., 1999). We used the WGLC dataset (Kaplan and Lau, 2019) which provides counts of monthly lightning strikes. It is based on the World Wide Lightning Location Network (WWLLN) dataset ~~(Kaplan and Lau, 2019) which provides counts of monthly lightning ground strikes~~, which mainly detects cloud-to-ground strikes (Rodger et al., 2004; Abarca et al., 2010), as opposed to LIS lightning data (Bürgesser, 2017).

Land cover was shown in previous studies to be another important influence on BA. We included several alternative representations of land cover including above-ground tree biomass (AGB) and the fractional cover of trees (TREE), shrubs (SHRUB), herbaceous vegetation (HERB), and crops (CROP) in our predictor set. Tree AGB was obtained by ~~combining~~ mosaicking AGB datasets for the tropics ~~and for~~ (Avitabile et al., 2016, 1 km resolution) and northern forests ~~(Avitabile et al., 2016; Thurner et al., 2014)~~ (Thurner et al., 2014, 0.01° resolution) using the mean after resampling each to a common spatial resolution of 0.25°. Yearly land cover values were obtained from the ESA CCI Land Cover dataset (Li et al., 2018). Land cover types were converted to fractional cover according to Poulter et al. (2015) using the conversion table as in Forkel et al. (2017). Global population density (POPD) from an updated version of the HYDE 3.2 dataset ~~(Klein Goldewijk, 2017)~~ (Klein Goldewijk, 2017, Kees Klein Goldewijk; person was used as a measure of human influence on vegetation and fire regimes.

Field data on fuel loads is sparse and the only global dataset (Pettinari and Chuvieco, 2016) is based on extrapolating scattered field measurements by biome. We therefore used four remotely sensed vegetation properties related to total biomass or leaf cover that could be regarded as indices for fuel load in our predictor set: solar-induced fluorescence (SIF), vegetation optical depth (VOD), fraction of absorbed photosynthetically active radiation (FAPAR), and leaf area index (LAI). All four properties have previously been used as productivity indices (e.g. Mohammed et al., 2019; Ryu et al., 2019; Teubner et al., 2018; Ogutu et al., 2014) and we use all four because it is uncertain which would be most closely related to fuel loads. Monthly SIF was obtained from the GlobFluo SIF dataset (Köhler et al., 2015). Ku-band VOD was obtained from the VODCA dataset (Moesinger et al., 2020). FAPAR and LAI were obtained from the MOD15A2H dataset (Myneni et al., 2015). ~~For~~

pre-processing, To pre-process data for the period from January 2010 to April 2015 we used data from January 2008 to April 2015, for which period all four datasets are available. Similarly, relevant data from February 2000 to December 2019 was pre-processed to enable analysis of the period from November 2000 to December 2019.

Table 1. Characteristics of the datasets. End times as applicable to the processed data are indicated in brackets.

Variable	Abbreviation	Dataset	Start	End	Time	Reference
burnt area	BA	GFED4	06-1995	12-2016	monthly	Giglio et al. (2013)
diurnal temperature range ~~, maximum temperature ,~~	DTR	ERA5	1950	present (12-2020)	monthly	Copernicus Climate Change Service (C3S) (
maximum temperature	MaxT	ERA5	1950	present (12-2020)	monthly	Copernicus Climate Change Service (C3S) (
dry-day period	~~DTR, MaxT,~~ DD	ERA5	~~01-1990~~ 1950	~~12-2018~~ present (11-2020)	monthly	Copernicus Climate Change Service (C3S) (2017)
soil moisture	SWI	Copernicus SWI	01-2007	11-2018	monthly	Albergel et al. (2008); Wagner et al. (1999)
lightning	Lightning	~~WWLLN~~ WGLC Lightning	01-2010	12-2018	monthly	Kaplan and Lau (2019)
above-ground tree biomass	AGB	Tropical AGB: Avitabile, Northern AGB: Thurner	static	static	static	Avitabile et al. (2016); Thurner et al. (2014)
land cover (fractional cover per grid cell)	CROP, SHRUB, TREE, HERB	ESA CCI Land Cover	1992	~~2015~~ present (2019)	yearly	Li et al. (2018)
solar-induced fluorescence	SIF	GlobFluo SIF	01-2007	04-2015	monthly	Köhler et al. (2015)
vegetation optical depth	VOD	VODCA (Ku-band)	12-1997	12-2018	monthly	Moesinger et al. (2019)
fraction of absorbed photosynthetically active radiation ~~,~~	FAPAR	MOD15A2H	02-2000	present (03-2021)	monthly	Myneni et al. (2015)
leaf area index	~~FAPAR,~~ LAI	MOD15A2H	02-2000	present (11-2018)	monthly	Myneni et al. (2015)
population density	POPD	HYDE 3.2 (updated)	2000	~~2017~~ present (2020)	yearly	Klein Goldewijk (2017), Kees Klein Goldewijk; personal communication (February 2021)
MCD64CMQ burnt area	MCD64 BA	MCD64CMQ	11-2000	present (06-2020)	monthly	Giglio et al. (2018)

2.2 Data processing

2.2.1 Gap filling

There are gaps in the SWI, FAPAR, LAI, SIF, and VOD datasets in winter months at latitudes above ~60°N, and in the austral winter for southern South America, due to high ~~sun~~ solar zenith angles for FAPAR, LAI, and SIF and because of snow cover and frozen soil for SWI and VOD ~~(e.g. Moesinger et al., 2020)~~(see e.g. Moesinger et al., 2020). There are also sporadic missing values in these datasets caused by e.g. cloud cover. ~~Missing values affect the calculation of antecedent conditions. These gaps were therefore~~ Unfortunately, simple exclusion of the times lacking data is not possible for our analysis because we commonly rely on antecedent samples throughout. Thus, data gaps were filled using a two-step approach as in Forkel et al. (2017) ~~for~~ in order to allow analysis of summer months at the affected locations. This approach differentiates between two gap types based on the amount of missing information for a specific month at each location. ~~First, persistent~~

First, 'persistent' gaps, defined as months for which ~~50%~~ 50% or more of the observations across all years are missing, were filled using the minimum value observed at that location ~~,~~ for the given predictor variable. ~~Second,~~ We assume that this indicates missing data during the winter, since other causes for data gaps (e.g. cloud cover) are predominantly 'transient'. 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 persistent and therefore filled using minima.

Second, the remaining transient gaps were filled using ~~a~~ season-trend regression ~~model~~ models with four harmonic terms ($k = 4$) and without breakpoints. These models were fitted using ordinary least ~~square~~ squares regression to the entire time-series obtained during the first step, as mentioned before using data from January 2008 to April ~~2015.~~ 2015 (or November 2000 to December 2019 for the monthly analysis). Cloud cover, which also affects detection in tropical and subtropical regions, is usually transient, and therefore filled using the regression models. Locations where no observations were available for > 52 months out of the total 88 months (regardless of whether such data gaps always occur in the same month, as for persistent gaps, or at any point throughout the year) were discarded in a trade-off between data quality and geographic extent. ~~Missing land cover data, which mostly occurred in fire-free regions, were replaced by global minimum area fractions: (HERB: 6.171491×10^{-6}; CROP: 4.936834×10^{-5})~~ For the monthly analysis, locations were discarded given > 138 unavailable months (out of 239 months).

Use of a different gap filling mechanism (Fig. S1b; temporal nearest-neighbour gap-filling) yielded very similar results. This simple nearest-neighbour gap-filling approach used for the eventual ALL_NN model processes timeseries at a given location, filling gaps by using the temporally closest available samples at that location. Of the two approaches, we decided to use the season-trend model with minima filling because it represents a more physical solution; it 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. Indeed, as can be seen in Fig. S2, virtually no samples are being filled with minima outside of winter, and predominantly in the northern extreme latitudes. While our gap-filling methodology may yield unphysical values for non-vegetation variables like SWI, we do not expect the filling of SWI to have a big influence on the final results because we do not use antecedent values of SWI.

Since we do not anticipate fires during the winter, having (by necessity of gap filling) potentially unphysical values of SWI in the winter should not affect results where relevant for our analysis.

2.2.2 Interpolation

All datasets were interpolated to a common $0.25°$ spatial grid. Datasets where the original spatial resolution was higher than this were averaged; the other datasets were interpolated using nearest-neighbour interpolation to avoid smoothing local extrema (Forkel et al., 2017). Datasets that were only available at yearly time resolution (i.e. land cover, POPD) were linearly interpolated to monthly intervals. Temporally static data (i.e. AGB) were recycled. Processing was carried out before averaging to provide monthly climatological time series where applicable.

2.2.3 Antecedent predictor variables

The influence of antecedent conditions that might affect fuel loads or fuel dryness, specifically vegetation properties and DD, on BA was investigated by using antecedent FAPAR, LAI, VOD, SIF, and DD data from up to two years before any given month (1M, 3M, 6M, 9M, 12M, 18M, 24M, where M denotes months). The large autocorrelation between predictor variables could impede the visual interpretation of the impacts of antecedent periods ≥ 1 yr. Thus, anomalies were computed by subtracting the seasonal cycle relative to the designated month, resulting in the following transformations:

$$(\mathrm{X\ 12M}) - \mathrm{X}(\mathrm{X\ 0M}) \rightarrow \mathrm{X}\ \Delta 12\mathrm{M},$$

$$(\mathrm{X\ 18M}) - (\mathrm{X\ 6M}) \rightarrow \mathrm{X}\ \Delta 18\mathrm{M},$$

$$\text{and } (\mathrm{X\ 24M}) - \mathrm{X}(\mathrm{X\ 0M}) \rightarrow \mathrm{X}\ \Delta 24\mathrm{M},$$

where $\mathrm{X} \in \{\text{FAPAR, LAI, VOD, SIF, DD}\}$ and X 0M refers to the variable X in the current month. For example, the 12-month antecedent X 12M was transformed by subtracting the instantaneous (month 0) value of X, thereby yielding the anomaly in X, X Δ12M, that may be easier to interpret.

2.3 Machine learning experiments

We used random forest (RF) regression to model the relationships between BA and the driver variables (predictors). RF is an ensemble learning approach in which multiple decision trees are constructed using a randomly sampled subset of training observations. The final model is the average result from all of the individual decision trees. RF regression is highly suited to investigating the emergent controls on fire because it is able to learn non-linear relationships in high-dimensional space (Archibald et al., 2009). By averaging over multiple decision trees, RFs also mitigate overfitting (Breiman, 2001). We used the scikit-learn version ~~0.23.0~~0.24.1 (Pedregosa et al., 2011) RF regression implementation in Python, with hyperparameters determined using five-fold ~~cross validation of the final~~ random cross validation (CV) of the eventual ALL model: `n_estimators`: 500, `max_depth`: 18, ~~`min_samples_leaf`: 3, `ccp_alpha`: 2×10^{-9},~~ and default values for all other parameters. The number of estimators (`n_estimators`) determines the number of trees whose predictions are averaged. The maximum depth

(max_depth) limits the number of split levels~~. Setting min_samples_leaf to three requires that all final splits contain at least three samples. Finally, the ccp_alpha parameter controls the cost-complexity pruning algorithm that penalises a large number of splits if this does not increase performance sufficiently. By limiting the number of splits according to different metrics, these measures reduce overfitting.~~ , which can reduce overfitting. We also found that a limited number of split levels was necessary for the computation of SHAP values, although we expect this to be a limitation of the specific software we used as opposed to the SHAP method itself. The hyperparameters were only estimated once for the model containing all variables due to computational constraints.

The validation dataset was randomly sampled across space and time and comprised ~~30%~~ 30% of the data. ~~The gap between the~~ To estimate how the model will perform on unseen data, the out-of-bag (OOB) R^2 ~~values obtained on the training and validation datasets provides a~~ for the training dataset can be used (Fox et al., 2017). However, since this data still belongs to the training dataset, the R^2 for the validation dataset, which has not been used for variable selection or hyperparameter tuning, is also used to provide an alternative, independent, measure of the generalisability of a given model~~, where a reduction in the training–validation gap indicates a greater capacity toidentify underlying relationships accurately~~.

However, this is only valid if there is no autocorrelation between the samples. We investigated the degree of spatial autocorrelation using a variogram of global GFED4 BA which informed a buffered leave-one-out (B-LOO) CV procedure following Ploton et al. (2020) . This was carried out to determine how much the autocorrelation that may be present influences the amount of potential overfitting. We did not employ the extrapolation-prevention procedure used in Ploton et al. (2020) because it led to the exclusion of significant areas like northern Australia and west Africa. The B-LOO CV was executed as follows, where $r_{max} = 50$ pixels and N_t was chosen such that the number of potential training samples was guaranteed to be equal to or above N_t for all $r \leq r_{max}$:

1. Randomly choose a single location. The 12 monthly samples at this test location constitute the test set.
2. Exclude samples from the potential training set in a circular region of radius r pixels around the test location, such that no potential training sample is closer than r pixels to the test location. This limits the influence of spatial autocorrelation.
3. Randomly choose N_t training samples from all remaining potential training samples.
4. Using a model trained on the above training samples, predict BA for the test location.
5. Increment r and repeat steps 1–4 until r has reached r_{max}.

This process was repeated 4000 times for each of the eight linearly spaced investigated radii, with the lowest radius being equal to 0. Due to computational constraints, the B-LOO CV was only carried out for a single set of variables.

We trained a number of different RF regression models to test explicit hypotheses about the importance of antecedent conditions on BA (see Table 2) using the defined hyperparameters on the climatological timeseries. The initial experiment (ALL) was run using the basic set of 15 predictor variables related to climate, vegetation, and human influences on fire (Table 1)

and included both current and antecedent values of the four vegetation indices and DD, giving 50 predictor variables. A second experiment (TOP15) used only the 15 most important predictors from the ALL model, as a way of testing whether all the predictors were necessary and whether including so many predictors resulted in overfitting.

The choice of 15 predictors was heuristically based on the slope of the feature importance plots (see Fig. S3), where, by inspection, the importance change is minimal after 15 variables. Thereafter, no additional information was being conveyed, so we decided to use this as our threshold. While use of the more rigorous recursive feature elimination with cross-validation (RFECV) would be possible in principle, this commonly makes use of the Gini importance owing to its ease of calculation, as it only considers data already seen during training. Unfortunately, this also means that RFECV fails to account for overfitting, as it only considers the training data when calculating feature importance (Meyer et al., 2019). In contrast, the different approaches we jointly utilised to calculate a more robust feature importance metric are much more computationally demanding, making RFECV infeasible.

All of the remaining experiments used combinations of 15 predictor variables. The CURR experiment only used current-month values of each predictor. ~~Comparison~~ Therefore, comparison of the CURR ~~, ALL, and TOP15~~ and ALL experiments allowed the impact of including antecedent vegetation and moisture conditions to be evaluated. ~~Some~~ However, some of the vegetation predictors are highly correlated with one another ~~and this~~ which could artificially decrease their importance~~in the ALL model~~. To test this, we ran four further experiments (15VEG_FAPAR, 15VEG_LAI, 15VEG_VOD, 15VEG_SIF) that included the 10 most important non-vegetation predictors from the ALL model, potentially including current and antecedent values of DD. In addition, each of these experiments contained one of the four vegetation predictors represented by both current (0M) and antecedent values ~~for~~ (1M, 3M, 6M, and 9M~~. These four experiments all included current and antecedent values of DD. To separate out~~). To disentangle the effects of antecedent DD and antecedent vegetation properties, we ran a second set of vegetation experiments (CURRDD_FAPAR, CURRDD_LAI, CURRDD_VOD, CURRDD_SIF) where each vegetation predictor was represented by both current (0M) and antecedent values ~~for~~ (1M, 3M, 6M, and 9M) but only using current DD and the next nine most important ~~other~~ non-vegetation factors from the CURR model. Finally, five-fold ~~cross validation~~ random CV was used to isolate the best combination of the vegetation predictors under the constraint that each of the ~~four antecedent states (1M–9M~~five states (0M–9M) must be represented exactly once (using any of the four vegetation predictors), resulting in the BEST15 model.

In addition to the above climatological experiments, we also investigate monthly data for the time period 11-2000–12-2029 (230 months) using the 15VEG_FAPAR_MON model. To avoid the temporal limits of the GFED4 dataset (see Table 1) the MODIS MCD64CMQ (Giglio et al., 2018) BA dataset was used. Otherwise, this experiment uses the same variables as the 15VEG_FAPAR experiment with the exception of lightning, which was replaced with the similarly significant variable AGB (see Fig. S3) in order to enable processing of a longer time period. In addition to five-fold random CV as for the other models, the performance and generalisability of this model was also measured using temporal CV. Here, the model was trained on all samples excluding either all months within the years 2009–2012 (including 2012) or 2016–2019 (including 2019). Thereafter, the R^2 was measured on whichever years were excluded for training. Note that unless explicitly specified, all following

Table 2. The modelling experiments. Except for the ALL ~~experiment~~experiments, the other experiments included 15 predictor variables for comparability. Differences in number of antecedent variables included in each experiment meant that different numbers of variables from the basic set were used in these experiments. For the TOP15, 15VEG_X(_MON), and BEST15 models we used the most important variables from the ALL experiment up to the required number of 15. For the CURRDD_X models, we used the most important non-vegetation variables from the CURR model. Table S1 provides a detailed list of the variables included in each experiment.

Name	Nr.	Variables
ALL & ALL_NN	50	Basic set of current variables + current month and antecedent (1, 3, 6, 12, 18, 24M) values for Dry Days and vegetation indices (FAPAR, LAI, VOD, SIF)
TOP15	15	Top 15 predictors from the ALL model
CURR	15	Only current values of the basic set of 15 variables
15VEG_X (e.g. 15VEG_FAPAR)	15	Top 10 non-vegetation variables from the ALL experiment, plus current and antecedent (1, 3, 6, 9 months) vegetation index X (e.g. FAPAR)
CURRDD_X (e.g. CURRDD_FAPAR)	15	Current and antecedent (1, 3, 6, 9 months) versions of the vegetation index X (e.g. FAPAR), current DD, top 9 other variables from the CURR experiment
BEST15	15	Current and antecedent DD, one current, 1M, 3M, 6M, 9M vegetation index (drawn from the four potential vegetation indices) and 5 most important other variables from the basic set ~~.~~
15VEG_FAPAR_MON	15	Same as the 15VEG_FAPAR experiment with monthly data instead of climatological data and lightning data replaced by the next-most important non-vegetation variable

methodological descriptions will relate to the climatological experiments as opposed to the monthly 15VEG_FAPAR_MON experiment.

2.4 Measuring predictor importance and relationships

Our goal is to determine the contribution of individual predictors (including antecedent states of these predictors) to model skill at predicting BA ~~,~~and to examine the relationships between predictors and BA.

There is no unique way to measure the importance of a given predictor on model skill in predicting BA and it is particularly difficult to assign importance to individual predictors when there is a high degree of collinearity between them ~~(Dormann et al., 2013; Nowa~~ (Dormann et al., 2013; Nowack et al., 2018; Mansfield et al., 2020). We use four techniques to ~~assess the robustness of the inferred~~ infer the importance of individual predictors: Gini impurity, permutation feature importance (PFI), leave-one-column-out (LOCO), and SHapley Additive exPlanations (SHAP) values. The Gini importance aggregates the decrease in mean squared error (MSE) for each split involving a given predictor variable over the individual decision trees making up the RF. The PFI was calculated from five permutations of each predictor variable ~~using~~ (validation set) using the ELI5 ~~0.10.1~~ 0.11.0 PermutationImportance (https://eli5.readthedocs.io/en/latest/index.html). While this provides an alternative assessment of the prediction score, the permutations may result in unlikely or impossible combinations of predictors and thus the PFI approach has a known tendency to overemphasise the importance of individual variables (Hooker and Mentch, 2019). The LOCO importance measure is estimated by repeatedly retraining the RF models, each time without one particular predictor variable. The relative impor-

tance of this predictor variable is then measured as the change in MSE on the ~~training~~ validation dataset, where a larger drop in MSE signifies a larger significance for the variable within the dataset. The importance of correlated predictor variables may be under-emphasised in this approach since the model is retrained, and thus some of the importance associated with the removed variable may be transferred to the correlated variables during the re-training process. The SHAP value (Lundberg and Lee, 2017; Lundberg et al., 2020) is the average of the marginal contributions from a series of perturbations of the predictor variables. In a similar way to the PFI, this method shares the importance amongst correlated predictor variables, which may make them appear less significant than if they were included on their own. ~~As for the other metrics,~~ SHAP values were computed ~~on all training samples. However, interaction strengths were only calculated for the first 300000 training samples. We created a~~ for all validation samples. In order to create a robust composite importance metric for each predictor variable~~by dividing~~, we divided the Gini, PFI, LOCO, and SHAP metrics by the sum of their absolute values and then ~~summing~~ summed them.

~~For predictor variable x, at location ℓ on the latitude–longitude grid, the SHAP value with the largest~~ Maximally significant timescales were calculated by weighting antecedent months using the largest SHAP value magnitude out of all 12 months ~~, m, was calculated as~~ in the climatological data. The maximum SHAP value magnitudes were calculated for a predictor variable x at location ℓ on the latitude-longitude grid as follows:

$$\mathrm{SHAP}_{x,\ell} := \mathrm{SHAP}_{x,\ell,\mathrm{m}_{\max}}, \quad \text{where } \mathrm{m}_{\max} = \underset{\mathrm{m}\in\{1,2,\ldots,12\}}{\arg\max} |\mathrm{SHAP}_{x,\ell,\mathrm{m}}|. \tag{1}$$

~~Maximally significant timescales were then calculated by weighting the antecedent months, a, using SHAP magnitudes:~~

$$t_{\max,\mathrm{X},\ell} = \frac{\sum_{\mathrm{a},x} \mathrm{a}|\mathrm{SHAP}_{x,\ell}|}{\sum_{\mathrm{a},x} |\mathrm{SHAP}_{x,\ell}|} \quad \text{for a}, x \in \{(0,\mathrm{X}), (1,\mathrm{X}\ 1\mathrm{M}), (3,\mathrm{X}\ 3\mathrm{M}), (6,\mathrm{X}\ 6\mathrm{M}), (9,\mathrm{X}\ 9\mathrm{M})\},$$

~~where X denotes~~ These were then used to calculate the maximally significant timescales for the basis predictor variable ~~for which to carry out the calculation~~ X (e.g. FAPAR) ~~.~~ using an average over antecedent months, a, weighted by maximum SHAP value magnitudes:

$$t_{\max,\mathrm{X},\ell} = \frac{\sum_{\mathrm{a},x} \mathrm{a}|\mathrm{SHAP}_{x,\ell}|}{\sum_{\mathrm{a},x} |\mathrm{SHAP}_{x,\ell}|} \quad \text{for a}, x \in \{(0,\mathrm{X}\ 0\mathrm{M}), (1,\mathrm{X}\ 1\mathrm{M}), (3,\mathrm{X}\ 3\mathrm{M}), (6,\mathrm{X}\ 6\mathrm{M}), (9,\mathrm{X}\ 9\mathrm{M})\}. \tag{2}$$

Locations with too many significant antecedent months were ignored in order to visualise resulting relationships more reliably; for example, if both the current (~~|SHAP$_{\mathrm{X},\ell}$|~~|SHAP$_{\mathrm{X}\,0\mathrm{M},\ell}$|) and nine-month antecedent (|SHAP$_{\mathrm{X}\,9\mathrm{M},\ell}$|) magnitudes are dominant, the weighted mean month (according to Eq. (2)) would lie in between, which is physically meaningless. We designed an algorithm to detect SHAP values that differ significantly from the baseline in order to mitigate this. Additionally, we also ~~employed~~ applied a range-based threshold, whereby locations were ignored ~~using the mean BA~~ if the variability of the SHAP values at location ℓ was below a threshold heuristically related to the mean BA, $\overline{\mathrm{BA}}_\ell$ (based on all BA samples)~~, if~~ :

$$\max_x(\mathrm{SHAP}_{x,\ell}) - \min_x(\mathrm{SHAP}_{x,\ell}) < 0.12 \times \overline{\mathrm{BA}}_\ell. \tag{3}$$

We further used Accumulated Local Effects (ALE) plots (Apley and Zhu, 2020) to examine and interpret the coupled relationships fitted by the RF models. ALE plots are a more robust alternative to partial dependence plots (PDP) or individual

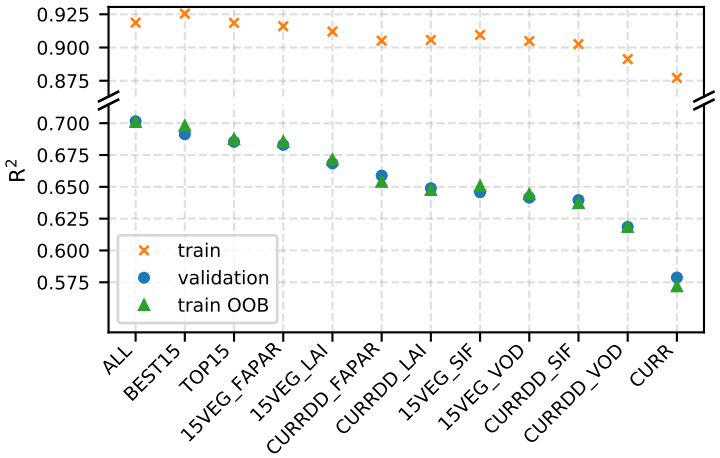

Figure 1. Global climatological R^2 scores for the different experiments. The CURR model is the only model that does not include antecedent conditions, and it performs much worse as a result. Despite the fact that the ALL model contains 50 predictors, while all other models contain just 15, it does not perform significantly better than the best models containing just 15 predictors (e.g. 15VEG_FAPAR). Note that although train R^2 scores are shown here, they are not indicative of model performance on unseen data, for which the shown train OOB scores should be used instead.

conditional expectation (ICE) plots (Apley and Zhu, 2020; Molnar, 2020). We assessed the impact of each of the predictor variables on BA in isolation using ~~1D ALEs , taking~~ first-order ALEs which take into account the effect of all other predictor variables. ~~Underlying~~ However, underlying inhomogeneities may appear when the model fits different relationships for different locations ~~. We~~ or times. We therefore tested for inhomogeneities by subsampling the dataset prior to ALE ~~plot construction, allowing~~ plotting to enable the visualisation of underlying relationships for a subset of locations and times. The causes of ~~the~~ these inhomogeneities were explored using ~~2D~~ second-order ALE plots, which show the combined effect of two predictor variables on BA.

3 Results and discussion

In general, all models are able to predict BA using the given biophysical predictors. However, the inclusion of antecedent predictors significantly improves model performance. Below, we discuss the performance of the different models, the importance of the different predictor variables, and their relationships with BA.

3.1 Model performance

The ALL model, which includes all 50 variables, achieves an in-the-bag R^2 of ~~0.884~~ 0.919 and out-of-bag (OOB) R^2 of 0.701 for the training dataset and ~~of 0.686~~ an R^2 of 0.701 for the validation dataset (Fig. 1). ~~Out-of-sample predictions~~ Predictions

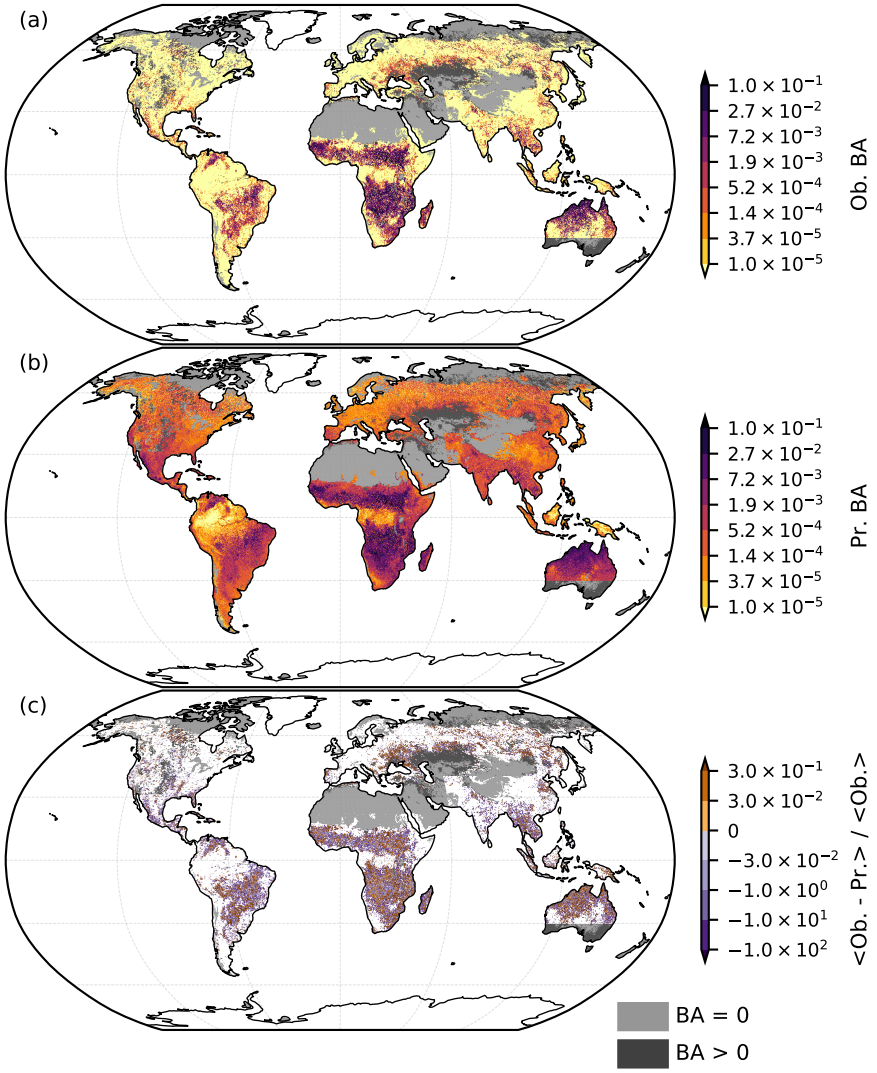

Figure 2. (a) Average observed (Ob.) BA derived from the GFED4 BA dataset (Giglio et al., 2013). (b) Out-of-sample predictions (Pr.) by the ALL model on the validation dataset. (c) Relative prediction error of the ALL model calculated by taking the mean of the difference between observations and predictions divided by the mean observations. While the predictions in (b) are qualitatively very similar to the observations in (a), there is an overestimation of low BA. Areas with very low or 0 observed BA (a) are omitted from (c) to avoid division by (nearly) 0. Despite the visual exaggeration of the errors, which are generally small, there is no overall pattern. Note that sharp data availability boundaries (e.g. in western Asia, southern Australia) are introduced by the AGB dataset. Grey shading indicates 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 non-zero mean BA.

on the validation set (Fig. 2b) show a similar geographic pattern to observed BA (Fig. 2a). However, overprediction in the validation set relative to observed BA is more widespread than underprediction (Fig. 2c). Nonetheless, there is no bias; the ALL model predicts a mean out-of-sample BA of ~~2.49×10^{-3}~~ 2.54×10^{-3} compared to the expected 2.48×10^{-3}. The apparent overprediction is the result of plotting relative (as opposed to absolute) errors, which amplifies the fact that the ALL model does not predict very low BA accurately~~.~~; out-of-sample BA predictions are no lower than ~~2.22×10^{-5}~~ 7.39×10^{-7}, while the observed BA is ~~0 for 85.7%~~ 0 for 85.7% of samples. Generally, ~~extreme BA is captured more poorly by the model than intermediate BA~~ the model captures intermediate BA better than extreme BA, leading to overprediction at low and underprediction at high BA values. ~~More~~ Thus, more samples are over-predicted because there are more values with low BA than high BA, leading to many instances of slight overprediction balanced by few instances of comparatively large underprediction (Fig. ~~S1, S2).~~ S4, S5). The model may struggle to predict 0 BA because the random forest model consists of many smaller decision trees. All 500 individual models would have to predict 0 to yield this value overall, which does not appear to occur given the stochastic nature of the training process. Failure to capture extreme events well is likely due to their rarity, resulting in the absence of comparable training data (see also e.g. Joshi and Sukumar, 2021).

Using a combination of regional neural networks trained on fewer variables at a coarser spatial resolution of $1° \times 1°$, Joshi and Sukumar (2021) found a global R^2 score for BA prediction of 0.36. An earlier study by Thomas et al. (2014) considered an R^2 score of 0.6 as indicating a robust prediction. Our results compare favourably to both. To further ensure model robustness, we also compared the PFI importances computed separately on the training and validation sets in Fig. S6. There is no appreciable difference between the two, which is indicative of a lack of overfitting, since the model training has not unduly prioritised certain variables based on the training set (Dankers and Pfisterer, 2020). Additionally, using the variogram shown in Fig. S7, we carried out the B-LOO CV as detailed in Sect. 2.3 in order to investigate the influence of spatial autocorrelation on the 15VEG_FAPAR model (see Fig. S8). The performance of the model drops as a larger region around the test samples is excluded (with 30 pixels corresponding to ~900 km at the equator, which is the scale of autocorrelation identified using Fig. S7). However, as opposed to the case study in Ploton et al. (2020), the R^2 score plateaus at around 0.1–0.4 beyond ~900 km instead of dropping to 0, thereby indicating the robustness of our model. Certain regions and extreme events are poorly captured by the model, accounting for the lower end of this range. Furthermore, the model is potentially forced to extrapolate to a larger extent as the exclusion radius is increased, leading to an overly pessimistic performance estimate. The extrapolation-prevention procedure in Ploton et al. (2020) was not used here because it led to the exclusion of certain key regions.

3.2 Importance of predictors

Climate variables and fuel-related vegetation indices have the strongest influence on BA in the ALL model (Table S2). Both current and antecedent conditions are important. Current DD and MaxT are ranked ~~1st and 3rd~~ first and fifth respectively, but antecedent DD also has a ~~strong~~ moderate influence (DD 1M and DD 3M are ranked ~~6th and 7th in importance~~ 10th and 13th in importance, respectively). Similarly, although current FAPAR is the most important vegetation index (2nd), with both current SIF and VOD occurring in the top ~~15~~ 15 predictors, antecedent vegetation state also has a strong influence on BA. However,

antecedent conditions > 9M are unimportant in the ALL model. Vegetation characteristics such as the cover of specific plant types (TREE, SHRUB, HERB) ~~or~~ and AGB are only moderately important in determining BA (all ranked below the top 15 predictors). Human impacts, as represented by CROP and POPD, are also only moderately important globally for BA, ranked respectively ~~13th~~ 8th and 15th. Natural ignitions ~~,~~ as represented by lightning are only ranked ~~26th~~21st, suggesting that at a global climatological scale burning is not limited by lightning.

The finding that fuel build-up on timescales longer than a year is not an important predictor of BA may initially be surprising given that fuel build-up as a result of fire suppression has been linked to large and catastrophic fires (e.g. in the USA; Marlon et al., 2012; Pa . The failure to detect an influence of longer-term fuel build-up on BA probably reflects the short time interval (1–2 yrs) considered for antecedent fuel build-up, far shorter than the timescales of coarse fuel build-up in these ecosystems. The seasonal differences captured by our analyses may also be unimportant in regions where long fire return times (or fire suppression) allow fuel build-up over longer periods. Wetter forests with long fire-return intervals may also be more affected by longer term moisture deficits (Abatzoglou and Kolden, 2013) that are not captured in the limited time period analysed. However, van der Werf et al. (2008) used 13 months of fuel accumulation before the peak of the fire season to investigate herbaceous fuels, which supports our findings somewhat. It would be worthwhile to re-examine the influence of longer timescales on BA when longer datasets are available, as, even when considering the ~20-year long MODIS record (which we do using the 15VEG_FAPAR_MON model), we are strongly limited by the data available to us. Predictability in boreal ecosystems is expected to remain very limited because the return times are many times longer than the time series, so there is a very large stochastic component. The lower performance of our monthly 15VEG_FAPAR_MON model with over 19 years of data, presented further below, in addition to the limited predictability of boreal regions found by Joshi and Sukumar (2021) despite their use of 14 years of data, both support this.

Our analyses are also impacted by the influence of previous fires on current vegetation conditions. Burnt grid cells could have a lower FAPAR, for example, as a result of prior burning within the current month. This is a problem because we are solely interested in how pre-fire vegetation conditions affect BA. The temporal and spatial scales of the analysis are responsible for this: a monthly analysis cannot resolve processes that occur on the order of days. Further, the impact of previous fires on spatially averaged vegetation properties is expected to be proportional to the burnt fraction of the affected grid cell. In savannah regions of Africa and northern Australia, where on the order of 10% of a 0.25° grid cell may burn in a given month, this could have a significant effect on the averaged values of vegetation properties used in our analyses. Analysis using a finer spatial scale would counteract this spatial smoothing by allowing burnt pixels to be ignored so that predictor values may be estimated only from unburnt cells. Using a finer temporal resolution would allow the calculation of predictor variables only up to the time of burning. In practice, however, while many variables (e.g. MODIS-based vegetation variables) are available at finer resolutions, the lack of accurate, reliable fire statistics at finer scales (Abatzoglou and Kolden, 2013) limits the temporal and spatial resolution that can usefully be achieved.

Although the limitation of the spatial (and temporal) resolution of the observations could impact the realism of our models, as could the omission of variables that affect fuel build-up, the consistency of the vegetation relationships shown by all the models (as detailed below) including antecedent conditions indicates that processes related to fuel build-up are adequately

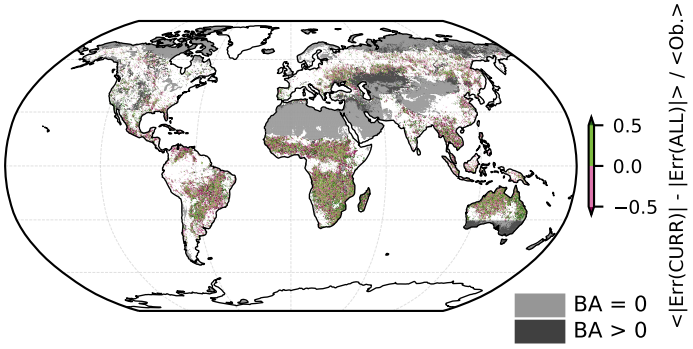

Figure 3. ~~Global climatological scores~~ Mean change in out-of-sample prediction error between the CURR and ALL models, relative to mean observations (<Ob.>). Green regions have decreased prediction error using the ALL model compared to the CURR model, and vice versa for the ~~different experiments~~purple regions. Areas with high BA (see Fig. 2a) tend to experience lower changes in relative prediction error. Areas with very low or 0 observed BA (see Fig. 2a) are omitted to avoid division by (nearly) 0. Note that sharp data availability boundaries (e.g. in western Asia, southern Australia) are introduced by the AGB dataset. Grey shading indicates 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 non-zero mean BA.

represented by the chosen set of predictors. The different importance metrics used are also in broad agreement, especially regarding the most important predictors like FAPAR and DD (see Fig. S1).

3.3 Models with fewer predictors

The model using the top 15 predictors from the ALL model (TOP15) performs only marginally worse than the ALL model, with an in-the-bag R^2 of ~~0.876~~ 0.919 and out-of-bag (OOB) R^2 of 0.688 for the training dataset and ~~of 0.676~~ an R^2 of 0.685 for the validation dataset. This nearly equivalent performance reflects the fact that there is a high degree of correlation between several of the variables (Fig. ~~S3~~S9) included in the ALL model~~. The absolute values and difference between the for the training and validation datasets in the TOP15 and ALL models are effectively the same,~~ , while also implying that the inclusion of extra predictors in the ALL model does not improve predictive capability. Therefore, this shows that it is not necessary to include multiple fuel-related vegetation variables in order to predict BA, provided that both current and antecedent conditions are taken into consideration. The removal of predictor variables is however likely to reduce overfitting in the TOP15 model compared to the ALL model ~~(Runge et al., 2019; Nowack et al., 2020)~~ (Runge et al., 2019; Nowack et al., 2020; Joshi and Sukumar, 2021) .

3.3.1 Importance of antecedent fuel-related predictors

The importance of antecedent fuel-related vegetation indices for predicting BA is corroborated by the results from the model that only includes predictors for the current month (CURR), where there is a large decrease in the R^2 for the validation dataset

compared to either the ALL (~~-0.119~~−0.123) or TOP15 (~~-0.116~~−0.107) model. ~~Comparison~~ The decrease in the R^2 for the training dataset is smaller (−0.042) than for the validation dataset, indicating that overfitting may be more of a problem in the CURR model than the ALL model. Analysis of the mean out-of-sample prediction error shows that ~~56%~~ 54.7% of grid cells are better predicted in the ALL model compared to the CURR model (Fig. 3). ~~However, there does not appear to be any regional patterning in the improvement caused by including antecedent vegetation conditions. The decrease in the~~ The performance improvements (from the CURR to the ALL model) also tend to have a larger magnitude than the performance decreases, contributing to the improvement in the global R^2 ~~for the training dataset is smaller (−0.064) than for the validation dataset, indicating that overfitting may be more of a problem in the CURR model than the ALL model~~score. Compared to the ALL model, fuel-related vegetation properties are less important in the CURR model: VOD is the highest-ranked vegetation variable but is only ~~fifth~~ fourth in importance (Table S2).

~~(a) Average observed (Ob.) BA derived from the GFED4 BA dataset (Giglio et al., 2013). (b) Out-of-sample predictions (Pr.) by the ALL model. (c) Relative prediction error of the ALL model calculated by taking the mean of the difference between observations and predictions divided by the mean observations. Areas with very low (or non-existent) observed BA (a) are omitted from (c) due to the effect of dividing by (nearly) zero.~~

~~Mean change in out-of-sample prediction error between the CURR and ALL models, relative to mean observations (<Ob.>). Green regions have decreased prediction error using the ALL model compared to the CURR model, and vice versa for the purple regions. Areas with high BA (see Fig. 2a) tend to experience lower changes in relative prediction error.~~

~~The~~ The four fuel-related vegetation variables included in the TOP15 model are correlated with one another (Fig. ~~S3), and there are high correlations between these four variables~~ S9), especially on specific antecedent timescales. This suggests it may be unnecessary to include all these variables to capture the influence of fuel build-up on BA. Comparison of the models which only include current and antecedent conditions for one fuel-related vegetation variable (15VEG_FAPAR, 15VEG_LAI, 15VEG_VOD, 15VEG_SIF) confirms this. ~~These models all~~ However, while all these models perform better than the CURR model (Fig. 1)~~. The~~ , only the 15VEG_FAPAR model performs ~~better than~~ similarly to the TOP15~~model and almost as well as the ALL model. LAI appears to be a better predictor~~, BEST15, and ALL models. Thus, considered on its own, ~~than either SIFor~~ FAPAR is the best fuel-related vegetation predictor, followed by LAI, SIF, and then VOD (Fig. 1). ~~The 15VEG_FAPAR also has the lowest training–validation gap (0.192) of all the four models, which suggests that it has the most generalisable relationships.~~ However, all four fuel-related vegetation predictors produce reasonable results, and other predictors (e.g. VOD) have been found to be important in other studies (e.g. Forkel et al., 2017).

The importance of including antecedent DD is borne out by the comparison of these four experiments and the experiments which only included current DD (CURRDD_FAPAR, CURRDD_LAI, CURRDD_VOD, CURRDD_SIF). In each case, the predictions for the same vegetation predictor variable are worse (Fig. 1). ~~The BEST 15~~

The BEST15 model contains the best combination of the fuel-related vegetation predictors (current FAPAR, FAPAR 1M, LAI 3M, SIF 6M, ~~VOD~~ SIF 9M), determined by optimising their timescales. This suggests that FAPAR is most important on short timescales (current, 1M) with the other vegetation properties appearing to be more useful on longer timeframes. The performance of this model, with a training R^2 of ~~0.878~~ 0.925 and a validation R^2 of ~~0.683~~0.691, is only bettered by the ALL

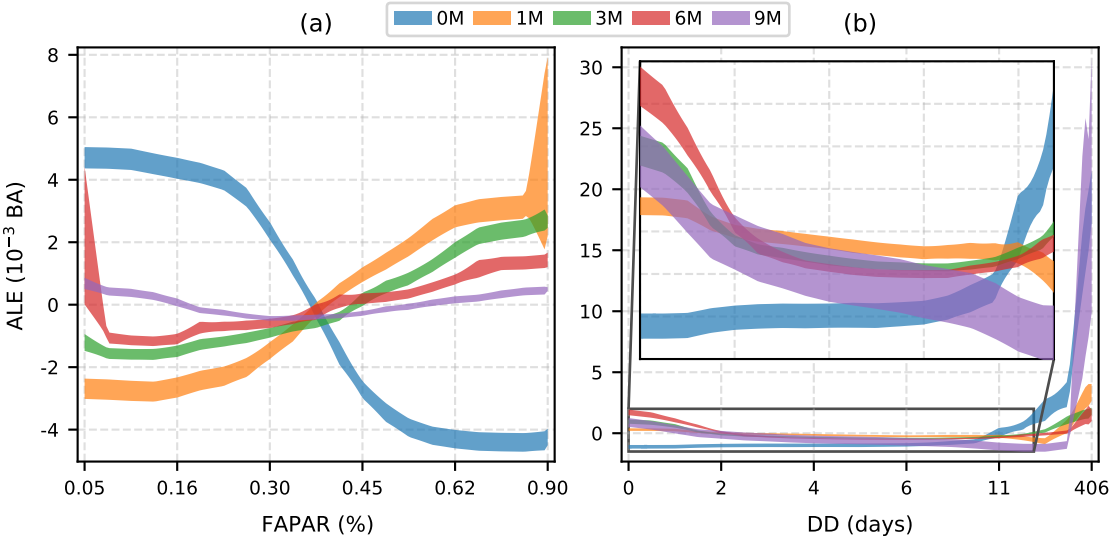

Figure 4. First-order ALEs for different antecedent (< 1 yr) relationships with (a) FAPAR and (b) DD in the 15VEG_FAPAR model, showing the underlying relationships with BA after accounting for all other variables. The shaded regions represent the standard deviation around the mean of 100 ALEs each using 122567 random samples of the training data (~10%). Evenly spaced quantiles were used in the construction of the plots. Labels were calculated using the averaged quantiles of all the variables used. A clear difference between instantaneous and antecedent relationships can be seen in both cases, with instantaneous FAPAR limiting BA while antecedent FAPAR promotes BA, and vice versa for the dry-day period. Note that the enhancement of BA due to extreme droughts (extreme dry-day period) is apparent across time periods.

model. The good performance of models including FAPAR is due to the fact that the parts of the world responding most strongly to FAPAR 0M and FAPAR 1M tend to be fuel-limited, dry biomes accounting for the majority of global BA (Giglio et al., 2013). Therefore, globally averaged model performance metrics will tend to favour predictor variables which best represent these dominant fire regimes. This is supported by previous analyses which have found predictability in regions with infrequent fires like boreal regions or Europe to be poor in contrast to regions with more frequent fires (e.g. Joshi and Sukumar, 2021).

3.4 Current and antecedent relationships with BA

Current and antecedent states of both fuel-related vegetation properties and DD have different impacts on BA (Fig. 4). Current FAPAR has a negative effect on BA, while antecedent FAPAR has a positive effect on BA (Fig. 4a), ~~which is strongest~~. The importance of current FAPAR changes most rapidly at intermediate levels of FAPAR~~, while antecedent FAPAR has a positive effect on BA~~. The impact of antecedent FAPAR is strongest for the preceding 1 month but persists for up to 6 months; longer lags tend to produce results more similar to the current relationship because of autocorrelation at the yearly scale. These relationships make intuitive sense: whereas high antecedent levels of FAPAR suggest that fuel availability is not a limiting

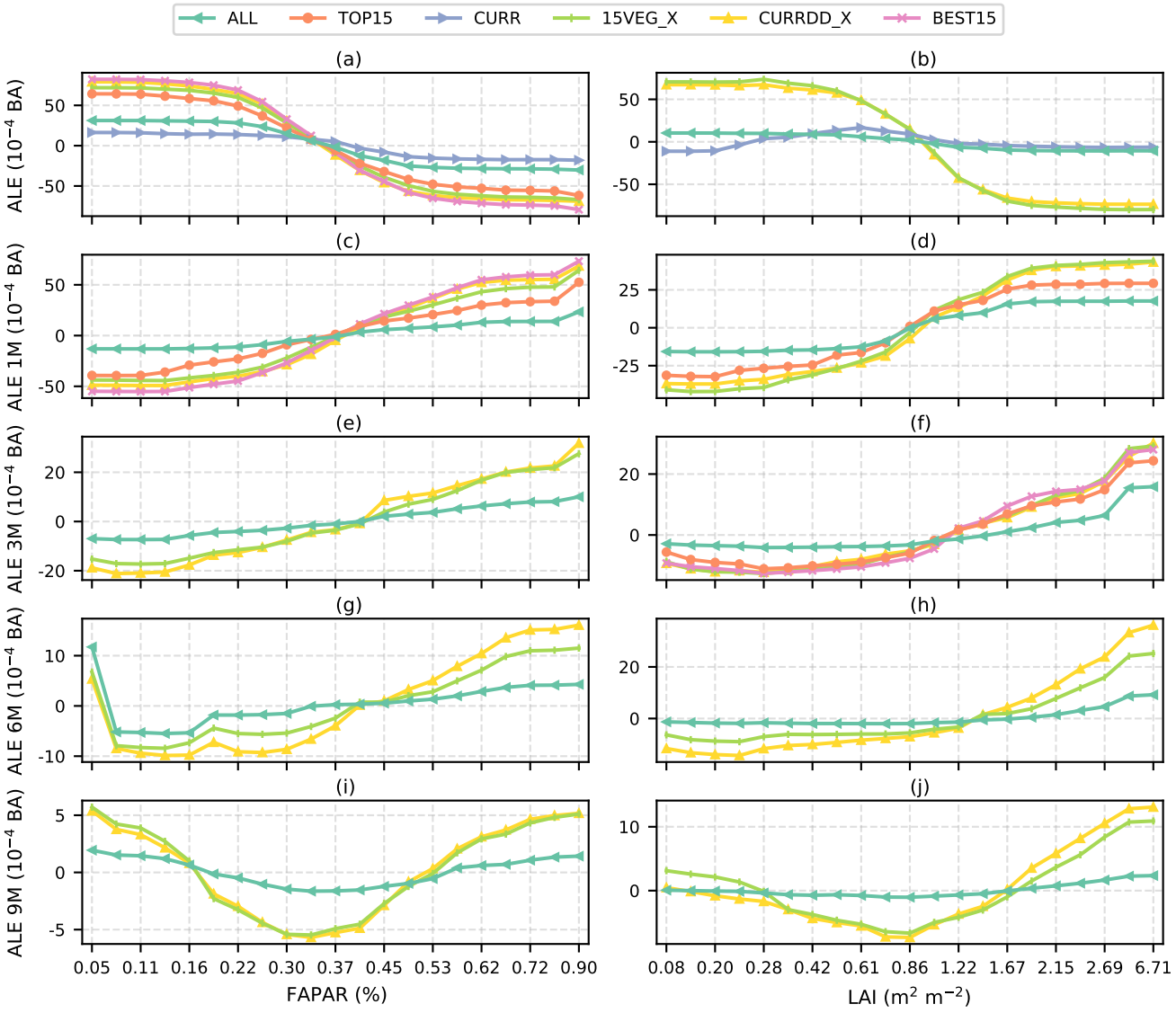

Figure 5. First-order ALEs for different lags (< 1 yr) from all relevant modelling experiments for the relationships between BA and FAPAR (left hand column) and LAI (right hand column). Evenly spaced quantiles were used in the construction of the plots. Notably, the relationship between LAI and BA is not modelled consistently by the CURR model (b), but relationships with BA are generally consistent across models otherwise.

factor, high FAPAR in the current month indicates that the vegetation has sufficient moisture to be actively growing and is therefore less likely to burn.

Current DD has a positive effect on BA (Fig. 4b), while antecedent DD has a generally negative effect except if DD is very high when the effect becomes positive again. Whereas the positive antecedent effect of FAPAR on BA is strongest for the preceding 1M relationship and then gets weaker, the negative impact of antecedent DD becomes gradually stronger ~~with antecedent DD~~ up to 9M. ~~These relationships make intuitive sense: whereas high antecedent levels of FAPAR suggest that fuel availability is not a limiting factor, high FAPAR in the current month indicates that the vegetation has sufficient moisture to be actively growing and therefore is less likely to burn. In contrast~~ In contrast to FAPAR, dry conditions in the current month promote fire whereas dry conditions in preceding months reduce vegetation growth and hence fuel build-up. Although prolonged droughts might be expected to reduce the availability of fuel, the (on average) positive relationship between BA and DD at very high levels of DD across all antecedent states does not support this expectation. The positive impact of drought in the current month becomes apparent for dry days $> \sim 10$ days, whereas the threshold is higher for antecedent months: BA only increases when the number of dry days is $> \sim 20$ days for the preceding month (DD 1M) and requires $> \sim 40$ days for DD 9M. This suggests that positive large antecedent DD may reflect prolonged droughts extending into the current month, thereby increasing fuel flammability and promoting fire.

~~First-order ALEs for different antecedent (< 1 yr) relationships with (a) FAPAR and (b) DD in the 15VEG_FAPAR model, showing the underlying relationships with BA after accounting for all other variables. Evenly spaced quantiles were used in the construction of the plots. Labels were calculated using the averaged quantiles of all the datasets used.~~ From Fig. 7a it can be seen that the negative effects of current FAPAR are most important in wet biomes, where limitation of fire activity due to instantaneous moisture conditions would also be expected to be strongest. This panel also has strong similarities with the results of Boer et al. (2021), with moisture-limited (dryness-limited) regions they identified corresponding broadly to the regions where instantaneous FAPAR is dominant due to the limitation imposed by high FAPAR on BA (see Fig. S10a). In these regions, instantaneous conditions reduce fuel available to burn due to moisture and antecedent conditions are less important due to the lack of a seasonal fuel build-up pattern. This then shows that in moisture-limited regions, dry events are important for enabling fire. From Fig. 7b it is apparent that on average, antecedent FAPAR is most important on a $\sim$4 month timescale.

~~The~~

3.4.1 Consistency of relationships

Consistent relationships between current or antecedent conditions and BA are generally reproduced in all of the RF models (Fig. 4, 5, ~~S4~~S11). However, exclusion of antecedent vegetation predictors can lead to counter-intuitive relationships between the current vegetation state and BA. Although the CURR model produces the expected relationship ~~between~~ with current FAPAR (and SIF; Fig. 5a, ~~S4a~~S11a), the relationship between current LAI (and VOD) and BA is initially positive and then flat (Fig. 5b, ~~S4b). This~~ S11b); this model does not show the expected strong negative relationship between current LAI and BA that occurs when antecedent moisture and vegetation conditions are included.

Relationships between predictors and BA were also stable when considering the 15VEG_FAPAR_MON model (Fig. S12) which not only uses monthly instead of climatological data, but also a different BA dataset. Using random CV, a validation R^2 of 0.501 and an OOB train R^2 of 0.498 were measured. Excluding the years 2009–2012, a validation R^2 of 0.403 and

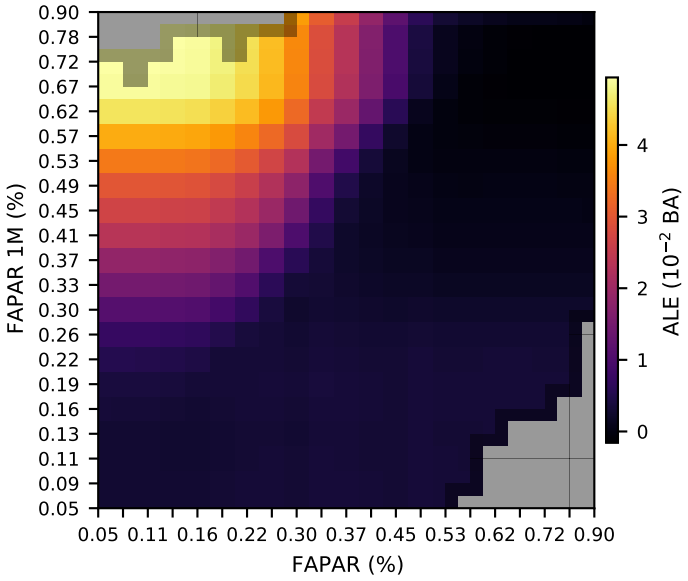

Figure 6. ~~First-order LAI ALEs for different lags~~ Second-order ALE plot showing the combined zeroth order (~~< 1 yr~~mean), first order, and second order modelled effects of FAPAR and FAPAR 1M on BA from the 15VEG_FAPAR model, taking into account all ~~relevant modelling experiments~~ other variables. Grey boxes indicate missing data. See Fig. S13 for the ~~relationships~~ sample count matrix which demonstrates the correlation between ~~BA and FAPAR (left hand column)~~ the variables and ~~LAI (right hand column)~~thus shows that samples are unlikely to fall into the top-left or bottom-right bins. Evenly spaced quantiles were used in the construction and labelling of the plots. ~~Labels were calculated using~~ It can be seen that the ~~averaged quantiles~~ combined effect of ~~all the datasets used~~FAPAR and FAPAR 1M on BA is positive if FAPAR is low while FAPAR 1M is high.

an OOB train R^2 of 0.507 were measured. Excluding the final years 2016–2019, a validation R^2 of 0.435 and an OOB train R^2 of 0.505 were measured. While these R^2 scores are lower than those observed for the previously discussed climatological analyses, they demonstrate that the model is able to robustly predict BA under multiple CV scenarios. Lower R^2 scores are also expected given the higher variance of this data. Additionally, the relationships identified by the model are highly consistent with the previous climatological analyses, showing that there is no temporal change that is important. The spatial patterns are dominant as the models behave very similarly when fit on climatological and monthly data; and the main commonality between those data is the geographical pattern. Note also that while lightning is omitted from this experiment in contrast to the climatological 15VEG_FAPAR experiment, lightning is also not present in the TOP15 model which performs similarly to the ALL and BEST15 models. Furthermore, as shown in Fig. S3, the importance of lightning and its replacement, AGB, are very similar.

3.4.2 Interactions

Although it is informative to consider the impact of individual predictor variables on BA, the expression of these relationships in the real world is likely to be conditioned by interactions with other variables. For example, low values of current FAPAR are associated with high BA (Fig. 4a), but this association occurs only when antecedent FAPAR is high (Fig. 6). Low FAPAR in the current month reflects unsuitable conditions for plant growth, for example during the dry season~~, and~~. Therefore, fuel build-up during the preceding months is ~~therefore~~ a prerequisite for fire to occur. The strong autocorrelation between current and preceding FAPAR values means that the occurrence of low current FAPAR coupled with high antecedent FAPAR is not widespread, being largely confined to shrublands in Africa. However, there is a significant interaction between current DD and current FAPAR (Fig. ~~S5~~S14), with positive reinforcement of their mutual influence on BA when DD is high and FAPAR is low and a negative influence on BA when DD is high and FAPAR is high. Increased BA for high DD and low FAPAR is consistent with strong drought-induced fire in low productivity environments of sub-Saharan Africa, northern Australia, and isolated regions bordering the African tropical rainforests. These findings therefore support previous results, e.g. by van der Werf et al. (2008) for Australia, where it was found that antecedent precipitation coupled with instantaneous drying was important for fire activity. Decreased BA as a result of increased DD and increased FAPAR is likely a sign of high-productivity environments that are not fire-prone, despite occasional drought.

3.4.3 Geographically varying timescales of importance

The timescales of both fuel build-up and fuel drying are influenced by fuel type and are therefore expected to vary across biomes. Current fuel-related vegetation properties, such as FAPAR (Fig. 7a), have an important effect in tropical regions, particularly dry tropical regions, but are less important in temperate forest regions. Antecedent FAPAR (Fig. 7b) is important in most regions, with the strongest influence from the antecedent 3–6 months. Current DD (Fig. 7c) is generally more important than antecedent DD, although the impact on BA varies geographically: tropical and boreal regions show decreased BA as a result of low current DD, while northwestern Australia, extra-tropical Africa, the Cerrado of Brazil, and the western USA experience increased burning as a result of ~~low~~ high current DD (Fig. ~~S6~~S10c). The timescale on which antecedent drought affects BA (Fig. 7d) is more variable than that for fuel-related vegetation properties, ranging from 1–3 months in boreal forests, parts of sub-Saharan Africa, and northern Australia, to ∼4 months in the tropics, and ∼6 months or longer in more arid regions.

~~Timescales of influence of FAPAR and DD on BA. The plots show the period that is most important for determining BA from the 15VEG_FAPAR model for (a) FAPAR and (c) DD. Plots (b) and (d) show which of the antecedent periods is most important (i.e. disregarding the influence of current conditions) in these experiments. Note that the maps give no indication about the sign of the influence of the predictor on BA (see Fig. S6).~~

4 ~~Discussion~~

4 Conclusions

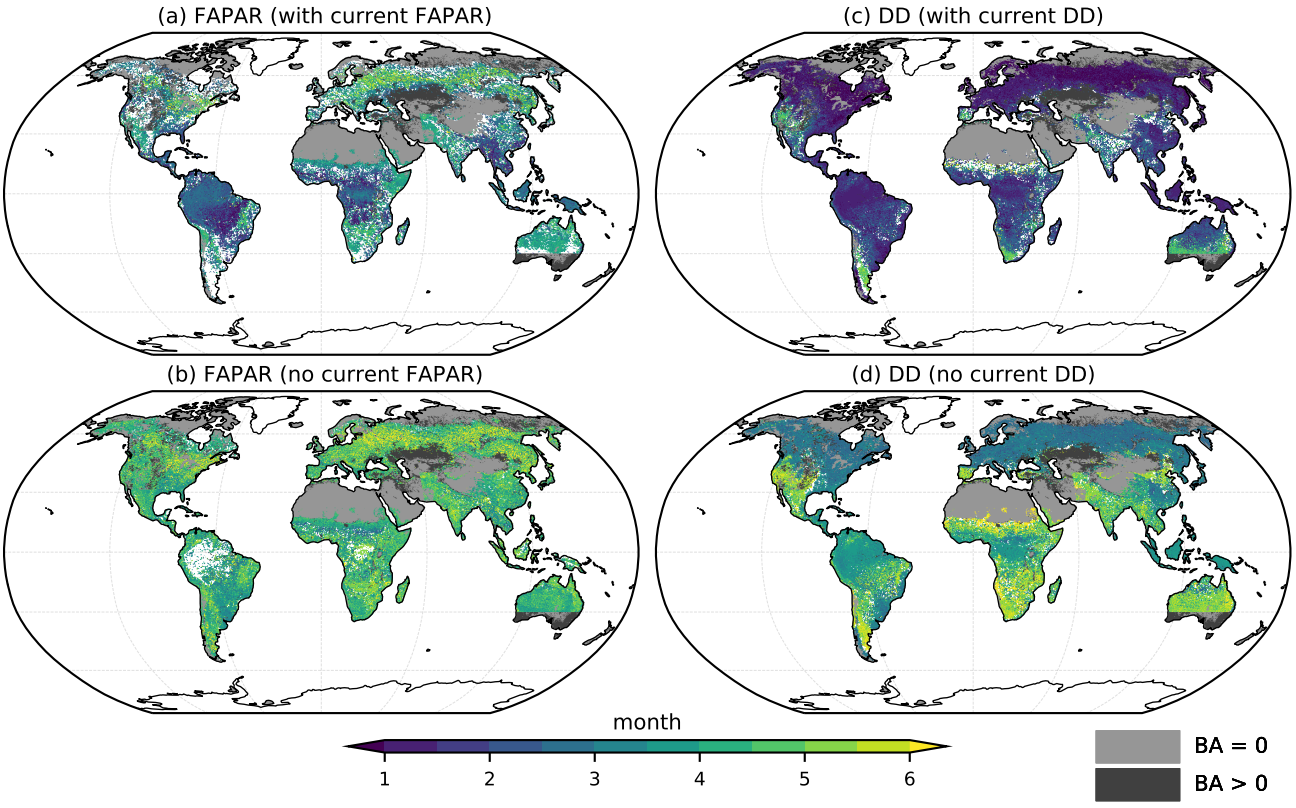

Figure 7. ~~Second-order ALE plot showing the combined zeroth order (mean), first order, and second order modelled effects~~ Timescales of influence of FAPAR and ~~FAPAR 1M~~ DD on BA. The plots show the period that is most important for determining BA from the 15VEG_FAPAR model ~~, taking into account all other variables~~for (a) FAPAR and (c) DD. ~~Grey boxes indicate missing data~~Plots (b) and (d) show which antecedent period is most important by disregarding the influence of current conditions during plotting. ~~See Fig~~Moister biomes are seen to be more influenced by current FAPAR (a), while current DD has a large influence globally (c). ~~S7 for~~ Note that the ~~sample count matrix which demonstrates~~ maps give no indication about the ~~correlation between~~ sign of the ~~variables and thus shows that samples are unlikely to fall into~~ influence of the ~~top-left bins~~predictor on BA (see Fig. S10). Note also that sharp data availability boundaries (e.g. ~~Evenly spaced quantiles are used~~ in western Asia, southern Australia) are introduced by the ~~construction and labelling~~ AGB dataset. Grey shading indicates regions with fire data availability, but where one or more of the ~~plots~~other datasets is not available. Light grey indicates regions where mean BA is 0, with dark grey representing regions with non-zero mean BA.

~~Many studies~~ By using random forest algorithms to model the dependence of BA on multiple climatic and biophysical variables, we have shown that antecedent vegetation conditions that influence fuel build-up and antecedent conditions that influence fuel drying significantly improved model performance when predicting BA in a given month. FAPAR was shown to be the most significant vegetation variable, and only a single vegetation variable is required for accurate BA prediction if antecedent conditions are included. Dry-day period and maximum temperature were the most significant climatic variables influencing

BA. This supports previous studies which have shown that ~~climate, vegetation properties , and human factors are all important predictors~~ current climate and vegetation properties are important overall determinants of BA ~~(e.g. Aldersley et al., 2011; Bistinas et al., 201~~ ~~. Although spatial variation in BA is predominantly the result of spatial variability in moisture and fuel-load (Archibald et al., 2009)~~ ~~, there is still poor understanding of the critical timescales for fuel build-up and fuel drying~~(e.g. Aldersley et al., 2011; Bistinas et al., 2014;) . The influence of antecedent climate conditions on both fuel buildup and fuel drying has also been identified as crucial in many regions (e.g. Van Wilgen et al., 2000; Griffin et al., 1983; Westerling et al., 2003; Swetnam and Betancourt, 1998; Jenkins et al., 2020; Arc . Indeed, the geographical patterning of BA can be linked to the spatial variability of fuel loads and fuel moisture (Archibald et al., 2009; Boe .

Our model-based analyses allowed us to distinguish between the immediate and antecedent impacts of fuel loads and fuel dryness on BA, while also allowing their relative contributions to be determined. We have ~~shown that the inclusion of both antecedent vegetation conditions that influence fuel build-up and antecedent conditions that influence fuel drying are necessary for the accurate prediction of BA. While conditions in the year before a fire are important, antecedent~~ further shown that current and antecedent conditions ~~for intervals $>$ 1 yr do not have a significant impact on BA in our analyses~~can influence BA in opposite but intuitively understandable ways: for example, wet conditions in antecedent months lead to more fuel buildup in fuel limited regions and promote increased BA whereas wet conditions during any given month reduce fuel dryness and thus limit BA. Furthermore, we have demonstrated that antecedent conditions $\geq$1 year are not important on a global scale. A similar conclusion was reached by Forkel et al. (2017). ~~The critical timescale for fuel build-up varies~~ Important and intuitive interactions between instantaneous and antecedent variables were captured by the models, for example supporting previous findings that increased antecedent productivity (FAPAR) coupled with instantaneous drying (Dry Days) promotes fire activity (e.g. van der Werf et al., 2008).

A clear contrast between fuel- and moisture-limited regions was also identified using the spatial variation of the relationship between antecedent FAPAR and BA. The critical timescales involved varied with vegetation type~~, with longer timescales being more important~~ ; longer timescales (~4 months) were more important for fuel build-up in temperate regions ~~and recent conditions being more important~~ while recent conditions were more important for fuel drying in the tropics. The effect of vegetation variables is also biome-dependent because of differing climatic constraints. The length of the dry-day period in the current month ~~has~~ had the largest impact on BA but antecedent DD ~~can also be~~ was also important, particularly in temperate regions.

~~The finding that fuel build-up on timescales longer than a year is not an important predictor of BA is surprising, given that fuel build-up as a result of fire suppression has been linked to large and catastrophic fires (e.g. in the United States; Marlon et al., 2012; Parks et . The failure to detect an influence of longer-term fuel build-up on BA probably reflects the short time interval (~5 yrs) used for the analyses because of the requirement that all of the datasets were available for the same time period. The seasonal differences captured by our analyses may be unimportant in regions where long fire return times (or fire suppression) allow fuel build-up over longer periods. Wetter forests with long fire-return intervals may also be more affected by longer term moisture deficits (Abatzoglou and Kolden, 2013) that are not captured in the limited time period analysed. It would be worthwhile re-examining the influence of longer timescales on BA in the future, when longer datasets are available.~~

We have shown that it is not necessary to include multiple fuel-related vegetation variables in order to predict BA, provided that both current and antecedent conditions are taken into consideration. According to our analyses, FAPAR is the best fuel-related vegetation predictor. However, all four fuel-related vegetation predictors produce reasonable results, and other predictors (e.g. VOD) have been found to be important in other studies (e.g. Forkel et al., 2017). Optimising the timescales of different fuel-related vegetation predictors suggests that FAPAR is most important on short timescales (current, 1M) and the other vegetation properties appear to be more useful on longer timeframes. The good performance of models including FAPAR is due to the fact that the parts of the world responding most strongly to FAPAR and FAPAR 1M tend to be fuel-limited, dry biomes accounting for the majority of global BA (Giglio et al., 2013). Therefore, globally averaged model performance metrics will tend to favour model predictor variables which best represent these dominant fire regimes.

Our analyses are impacted by the influence of previous fires on current vegetation conditions. Burnt grid cells could have a lower FAPAR, for example, as a result of prior burning in the current month. This is a problem, related to the temporal and spatial scales of the analysis, because we are solely interested in how pre-fire vegetation conditions affect BA. A monthly analysis cannot resolve processes that occur on the order of only a few days, while the impact of previous fires on spatially averaged vegetation properties is proportional to the fraction of the grid cell that is burnt. In savannah regions of Africa and northern Australia, where on average up to 10% of a 0.25° grid cell burns in any given month, this could have a significant effect on the averaged values of vegetation properties used in our analyses. Using a finer spatial scale for analysis would counteract the impact of spatial smoothing, allowing burnt pixels to be ignored and predictor values to be estimated only from unburnt cells. Using a finer temporal resolution would allow the calculation of predictor variables only up to the time of burning. In practice, however, the lack of accurate fire statistics at finer scales (Abatzoglou and Kolden, 2013) limits the temporal and spatial resolution that could usefully be achieved.

Although the limitation of the spatial (and temporal) resolution of the observations could impact the realism of the RF model, as could the omission of variables that affect fuel build-up, the consistency of the vegetation relationships shown by all the models including antecedent conditions indicates that processes related to fuel build-up are adequately represented by the chosen set of predictors.

The dependence of BA on multiple climatic and biophysical variables has been modelled using a random forest algorithm. We showed that FAPAR was the most significant vegetation variable, and dry-day period and maximum temperature the most significant climatic variables, influencing BA. The inclusion of antecedent relationships with FAPAR and dry-day period significantly improved model performance. Antecedent relationships to BA were consistently different from instantaneous relationships, with the spatiotemporal variation of fuel build-up processes apparent via the relative importance of these different lagged relationships. A clear contrast between fuel- Future work could re-calibrate the models for each set of variables to potentially improve their performance. Approaches like B-LOO CV or our combined variable importance metric could also be used more extensively to select the predictor variables and moisture-limited regions was identified using the spatial variation of the relationship between antecedent FAPAR and BA. Moisture-limited regions were more strongly affected by suppression of fire at instantaneous timescales, while fuel-limited regions were most affected by fuel build-up over the previous ~4 months. The instantaneous dry-day period was found to have the biggest effect on BA, suppressing or promoting firebased on the

~~climate~~potentially tune the model hyperparameters, given that these methods are currently very computationally intensive. Finally, the findings presented herein have the potential to improve the modelling of fire at a global scale in order to improve the way that Earth System Models depict the interactions between climate, vegetation, and fire.

Code availability. Computer code can be found in the empirical-fire-modelling package (Kuhn-Régnier, 2021a). ALE plots were generated using the ALEPython package (Kuhn-Régnier et al., 2021). Data analysis was carried out using the Python 3.7 (Van Rossum and Drake, 2009) packages SciPy (Virtanen et al., 2020), Matplotlib (Hunter, 2007), NumPy (Oliphant, 2006), Iris (Met Office, 2010), Dask (Dask Development Team, 2016), Jupyter notebooks (Kluyver et al., 2016), wildfires (Kuhn-Régnier, 2021b), and era5analysis (Kuhn-Régnier, 2020). GFED4 data was read using pyhdf (https://github.com/fhs/pyhdf, wraps NCSA HDF version 4).

Author contributions. Conceptualization, AKR, AV, ICP, SPH; Methodology, AKR; Software, AKR; Validation, AKR; Formal Analysis, AKR; Investigation, AKR, SPH; Data Curation, AKR, MF; Writing – Original Draft, AKR; Writing – Review & Editing, AKR, AV, PN, MF, ICP, SPH; Visualization, AKR; Supervision, AV, PN, ICP, SPH.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. AKR acknowledges support from the NERC Centre for Doctoral Training in Quantitative and Modelling skills in Ecology and Evolution (QMEE, grant reference NE/P012345/1). ICP acknowledges support from the ERC-funded project REALM (grant number 787203). SPH acknowledges support from the ERC-funded project GC2.0 (Global Change 2.0: Unlocking the past for a clearer future, grant number 694481). This research was partially funded by the Leverhulme Centre for Wildfires, Environment, and Society through the Leverhulme Trust, grant number RC-2018-023. The authors would like to thank Kees Klein Goldewijk for contributing an updated HYDE dataset.

References

Abarca, S. F., Corbosiero, K. L., and Galarneau, T. J.: An Evaluation of the Worldwide Lightning Location Network (WWLLN) Using the National Lightning Detection Network (NLDN) as Ground Truth, J. Geophys. Res., 115, D18 206, https://doi.org/10.1029/2009JD013411, 2010.

Abatzoglou, J. T. and Kolden, C. A.: Relationships between Climate and Macroscale Area Burned in the Western United States, Int. J. Wildland Fire, 22, 1003–1020, https://doi.org/10.1071/WF13019, 2013.

~~Abatzoglou, J. T. and Williams, A. P.: Impact of Anthropogenic Climate Change on Wildfire across Western USForests, Proc Natl Acad Sci USA, 113, 11 770–11 775, , 2016.~~

Abatzoglou, J. T., Williams, A. P., Boschetti, L., Zubkova, M., and Kolden, C. A.: Global Patterns of Interannual Climate-Fire Relationships, Glob Change Biol, 24, 5164–5175, https://doi.org/10.1111/gcb.14405, 2018.

Abatzoglou, J. T., Williams, A. P., and Barbero, R.: Global Emergence of Anthropogenic Climate Change in Fire Weather Indices, Geophys. Res. Lett., 46, 326–336, https://doi.org/10.1029/2018GL080959, 2019.

Albergel, C., Rüdiger, C., Pellarin, T., Calvet, J.-C., Fritz, N., Froissard, F., Suquia, D., Petitpa, A., Piguet, B., and Martin, E.: From Near-Surface to Root-Zone Soil Moisture Using an Exponential Filter: An Assessment of the Method Based on in-Situ Observations and Model Simulations, Hydrol. Earth Syst. Sci., 12, 1323–1337, https://doi.org/10.5194/hess-12-1323-2008, 2008.

Aldersley, A., Murray, S. J., and Cornell, S. E.: Global and Regional Analysis of Climate and Human Drivers of Wildfire, Science of The Total Environment, 409, 3472–3481, https://doi.org/10.1016/j.scitotenv.2011.05.032, 2011.

Alvarado, S. T., Andela, N., Silva, T. S. F., and Archibald, S.: Thresholds of Fire Response to Moisture and Fuel Load Differ between Tropical Savannas and Grasslands across Continents, Global Ecol Biogeogr, 29, 331–344, https://doi.org/10.1111/geb.13034, 2020.

Andela, N., Morton, D. C., Giglio, L., Chen, Y., van der Werf, G. R., Kasibhatla, P. S., DeFries, R. S., Collatz, G. J., Hantson, S., Kloster, S., Bachelet, D., Forrest, M., Lasslop, G., Li, F., Mangeon, S., Melton, J. R., Yue, C., and Randerson, J. T.: A Human-Driven Decline in Global Burned Area, Science, 356, 1356–1362, https://doi.org/10.1126/science.aal4108, 2017.

Apley, D. W. and Zhu, J.: Visualizing the Effects of Predictor Variables in Black Box Supervised Learning Models, J. R. Stat. Soc. Ser. B Stat. Methodol., 82, 1059–1086, https://doi.org/10.1111/rssb.12377, 2020.

Archibald, S., Roy, D. P., van Wilgen, B. W., and Scholes, R. J.: What Limits Fire? An Examination of Drivers of Burnt Area in Southern Africa, Glob. Change Biol., 15, 613–630, https://doi.org/10.1111/j.1365-2486.2008.01754.x, 2009.

Avitabile, V., Herold, M., Heuvelink, G. B. M., Lewis, S. L., Phillips, O. L., Asner, G. P., Armston, J., Ashton, P. S., Banin, L., Bayol, N., Berry, N. J., Boeckx, P., de Jong, B. H. J., DeVries, B., Girardin, C. A. J., Kearsley, E., Lindsell, J. A., Lopez-Gonzalez, G., Lucas, R., Malhi, Y., Morel, A., Mitchard, E. T. A., Nagy, L., Qie, L., Quinones, M. J., Ryan, C. M., Ferry, S. J. W., Sunderland, T., Laurin, G. V., Gatti, R. C., Valentini, R., Verbeeck, H., Wijaya, A., and Willcock, S.: An Integrated Pan-Tropical Biomass Map Using Multiple Reference Datasets, Glob. Change Biol., 22, 1406–1420, https://doi.org/10.1111/gcb.13139, 2016.

Barbero, R., Abatzoglou, J. T., Larkin, N. K., Kolden, C. A., and Stocks, B.: Climate Change Presents Increased Potential for Very Large Fires in the Contiguous United States, Int. J. Wildland Fire, 24, 892–899, https://doi.org/10.1071/WF15083, 2015.

Beck, P. S. A., Atzberger, C., Høgda, K. A., Johansen, B., and Skidmore, A. K.: Improved Monitoring of Vegetation Dynamics at Very High Latitudes: A New Method Using MODIS NDVI, Remote Sensing of Environment, 100, 321–334, https://doi.org/10.1016/j.rse.2005.10.021, 2006.

Bedia, J., Herrera, S., Gutiérrez, J. M., Benali, A., Brands, S., Mota, B., and Moreno, J. M.: Global Patterns in the Sensitivity of Burned Area to Fire-Weather: Implications for Climate Change, Agricultural and Forest Meteorology, 214-215, 369–379, https://doi.org/10.1016/j.agrformet.2015.09.002, 2015.

Bessie, W. C. and Johnson, E. A.: The Relative Importance of Fuels and Weather on Fire Behavior in Subalpine Forests, Ecology, 76, 747–762, https://doi.org/10.2307/1939341, 1995.

Bistinas, I., Harrison, S. P., Prentice, I. C., and Pereira, J. M. C.: Causal Relationships versus Emergent Patterns in the Global Controls of Fire Frequency, Biogeosciences, 11, 5087–5101, https://doi.org/10.5194/bg-11-5087-2014, 2014.

Boer, M. M., Dios, V. R. D., Stefaniak, E., and Bradstock, R. A.: A Hydroclimatic Model for the Distribution of Fire on Earth, Environ. Res. Commun., https://doi.org/10.1088/2515-7620/abec1f, 2021.

Bowman, D. M. J. S., Balch, J., Artaxo, P., Bond, W. J., Cochrane, M. A., D'Antonio, C. M., DeFries, R., Johnston, F. H., Keeley, J. E., Krawchuk, M. A., Kull, C. A., Mack, M., Moritz, M. A., Pyne, S., Roos, C. I., Scott, A. C., Sodhi, N. S., and Swetnam, T. W.: The Human Dimension of Fire Regimes on Earth, J. Biogeogr., 38, 2223–2236, https://doi.org/10.1111/j.1365-2699.2011.02595.x, 2011.

~~Bowman, D. M. J. S., Williamson, G. J., Abatzoglou, J. T., Kolden, C. A., Cochrane, M. A., and Smith, A. M. S.: Human Exposure and Sensitivity to Globally Extreme Wildfire Events, Nat. Ecol. Evol., 1, 1–6, , 2017.~~

Breiman, L.: Random Forests, Machine Learning, 45, 5–32, https://doi.org/10.1023/A:1010933404324, 2001.

Bürgesser, R. E.: Assessment of the World Wide Lightning Location Network (WWLLN) Detection Efficiency by Comparison to the Lightning Imaging Sensor (LIS): WWLLN Detection Efficiency Relative to LIS, Q.J.R. Meteorol. Soc, 143, 2809–2817, https://doi.org/10.1002/qj.3129, 2017.

Burton, C., Betts, R. A., Jones, C. D., and Williams, K.: Will Fire Danger Be Reduced by Using Solar Radiation Management to Limit Global Warming to 1.5 °C Compared to 2.0 °C?, Geophys. Res. Lett., 45, 3644–3652, https://doi.org/10.1002/2018GL077848, 2018.

Copernicus Climate Change Service (C3S): ERA5: Fifth Generation of ECMWF Atmospheric Reanalyses of the Global Climate . Copernicus Climate Change Service Climate Data Store (CDS), https://cds.climate.copernicus.eu/cdsapp#!/home, 2017.

Dankers, C. and Pfisterer, F.: Chapter 11 PFI: Training vs. Test Data | Limitations of Interpretable Machine Learning Methods, 2020.

Dask Development Team: Dask: Library for Dynamic Task Scheduling, 2016.

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., and Lautenbach, S.: Collinearity: A Review of Methods to Deal with It and a Simulation Study Evaluating Their Performance, Ecography, 36, 27–46, https://doi.org/10.1111/j.1600-0587.2012.07348.x, 2013.

Forkel, M., Dorigo, W., Lasslop, G., Teubner, I., Chuvieco, E., and Thonicke, K.: A Data-Driven Approach to Identify Controls on Global Fire Activity from Satellite and Climate Observations (SOFIA V1), Geosci. Model Dev., 10, 4443–4476, https://doi.org/10.5194/gmd-10-4443-2017, 2017.

Forkel, M., Andela, N., Harrison, S. P., Lasslop, G., ~~van Marle~~van Marle, M., Chuvieco, E., Dorigo, W., Forrest, M., Hantson, S., Heil, A., Li, F., Melton, J., Sitch, S., Yue, C., and Arneth, A.: Emergent Relationships ~~on~~ with Respect to Burned Area in Global Satellite Observations and Fire-Enabled Vegetation Models, Biogeosciences~~Discuss., pp. 1–31,~~ , 16, 57–76, https://doi.org/10.5194/bg-16-57-2019, 2019a.

Forkel, M., Dorigo, W., Lasslop, G., Chuvieco, E., Hantson, S., Heil, A., Teubner, I., Thonicke, K., and Harrison, S. P.: Recent Global and Regional Trends in Burned Area and Their Compensating Environmental Controls, Environ. Res. Commun., 1, 051 005, https://doi.org/10.1088/2515-7620/ab25d2, 2019b.

Fox, E. W., Hill, R. A., Leibowitz, S. G., Olsen, A. R., Thornbrugh, D. J., and Weber, M. H.: Assessing the Accuracy and Stability of Variable Selection Methods for Random Forest Modeling in Ecology, Environ Monit Assess, 189, 316, https://doi.org/10.1007/s10661-017-6025-0, 2017.

Giglio, L., Randerson, J. T., and van der Werf, G. R.: Analysis of Daily, Monthly, and Annual Burned Area Using the Fourth-Generation Global Fire Emissions Database (GFED4), J. Geophys. Res. Biogeosciences, 118, 317–328, https://doi.org/10.1002/jgrg.20042, 2013.

Giglio, L., Boschetti, L., Roy, D. P., Humber, M. L., and Justice, C. O.: The Collection 6 MODIS Burned Area Mapping Algorithm and Product, Remote Sensing of Environment, 217, 72–85, https://doi.org/10.1016/j.rse.2018.08.005, 2018.

Goss, M., Swain, D. L., Abatzoglou, J. T., Sarhadi, A., Kolden, C., Williams, A. P., and Diffenbaugh, N. S.: Climate Change Is Increasing the Risk of Extreme Autumn Wildfire Conditions across California, Environ. Res. Lett., https://doi.org/10.1088/1748-9326/ab83a7, 2020.

Griffin, G., Price, N., and Portlock, H.: Wildfires in the Central Australian Rangelands, 1970-1980., J. Environ. Manage., 17, 311–323, 1983.

Hantson, S., Kelley, D. I., Arneth, A., Harrison, S. P., Archibald, S., Bachelet, D., Forrest, M., Hickler, T., Lasslop, G., Li, F., Mangeon, S., Melton, J. R., Nieradzik, L., Rabin, S. S., Prentice, I. C., Sheehan, T., Sitch, S., Teckentrup, L., Voulgarakis, A., and Yue, C.: Quantitative Assessment of Fire and Vegetation Properties in Historical Simulations with Fire-Enabled Vegetation Models from the Fire Model Intercomparison Project, Preprint, Biogeosciences, https://doi.org/10.5194/gmd-2019-261, 2020.

Harris, I., Jones, P., Osborn, T., and Lister, D.: Updated High-resolution Grids of Monthly Climatic Observations – the CRU TS3.10 Dataset, Int. J. Climatol., 34, 623–642, https://doi.org/10.1002/joc.3711, 2014.

Higuera, P. E., Abatzoglou, J. T., Littell, J. S., and Morgan, P.: The Changing Strength and Nature of Fire-Climate Relationships in the Northern Rocky Mountains, U.S.A., 1902-2008, PLOS ONE, 10, e0127 563, https://doi.org/10.1371/journal.pone.0127563, 2015.

Hooker, G. and Mentch, L.: Please Stop Permuting Features: An Explanation and Alternatives, ArXiv190503151 Cs Stat, 2019.

Hunter, J. D.: Matplotlib: A 2D Graphics Environment, Comput. Sci. Eng., 9, 90–95, 2007.

Jenkins, M. E., Bedward, M., Price, O., and Bradstock, R. A.: Modelling Bushfire Fuel Hazard Using Biophysical Parameters, Forests, 11, 925, https://doi.org/10.3390/f11090925, 2020.

Jolly, W. M., Cochrane, M. A., Freeborn, P. H., Holden, Z. A., Brown, T. J., Williamson, G. J., and Bowman, D. M. J. S.: Climate-Induced Variations in Global Wildfire Danger from 1979 to 2013, Nat. Commun., 6, 7537, https://doi.org/10.1038/ncomms8537, 2015.

Joshi, J. and Sukumar, R.: Improving Prediction and Assessment of Global Fires Using Multilayer Neural Networks, Sci. Rep., 11, 3295, https://doi.org/10.1038/s41598-021-81233-4, 2021.

Kaplan, J. O. and Lau, H.-K.: The WGLC Global Gridded Monthly Lightning Stroke Density and Climatology, https://doi.org/10.1594/PANGAEA.904253, 2019.

Keane, R. E., Burgan, R., and van Wagtendonk, J.: Mapping Wildland Fuels for Fire Management across Multiple Scales: Integrating Remote Sensing, GIS, and Biophysical Modeling, Int. J. Wildland Fire, 10, 301, https://doi.org/10.1071/WF01028, 2001.

Kelley, D. I., Bistinas, I., Whitley, R., Burton, C., Marthews, T. R., and Dong, N.: How Contemporary Bioclimatic and Human Controls Change Global Fire Regimes, Nat. Clim. Chang., 9, 690–696, https://doi.org/10.1038/s41558-019-0540-7, 2019.

Klein Goldewijk, C.: Anthropogenic Land-Use Estimates for the Holocene; HYDE 3.2, https://doi.org/10.17026/DANS-25G-GEZ3, 2017.

Kloster, S. and Lasslop, G.: Historical and Future Fire Occurrence (1850 to 2100) Simulated in CMIP5 Earth System Models, Global and Planetary Change, 150, 58–69, https://doi.org/10.1016/j.gloplacha.2016.12.017, 2017.

Kloster, S., Mahowald, N. M., Randerson, J. T., and Lawrence, P. J.: The Impacts of Climate, Land Use, and Demography on Fires during the 21st Century Simulated by CLM-CN, Biogeosciences, 9, 509–525, https://doi.org/10.5194/bg-9-509-2012, 2012.

Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J., Grout, J., Corlay, S., Ivanov, P., Avila, D., Abdalla, S., and Willing, C.: Jupyter Notebooks – a Publishing Format for Reproducible Computational Workflows, in: Positioning and Power in Academic Publishing: Players, Agents and Agendas, edited by Loizides, F. and Schmidt, B., pp. 87–90, IOS Press, 2016.

Knorr, W., Jiang, L., and Arneth, A.: Climate, CO_2 and Human Population Impacts on Global Wildfire Emissions, Biogeosciences, 13, 267–282, https://doi.org/10.5194/bg-13-267-2016, 2016.

Köhler, P., Guanter, L., and Joiner, J.: A Linear Method for the Retrieval of Sun-Induced Chlorophyll Fluorescence from GOME-2 and SCIAMACHY Data, Atmospheric Meas. Tech., 8, 2589–2608, https://doi.org/10.5194/amt-8-2589-2015, 2015.

Krawchuk, M. A. and Moritz, M. A.: Constraints on Global Fire Activity Vary across a Resource Gradient, Ecology, 92, 121–132, https://doi.org/10.1890/09-1843.1, 2011.

~~Kuhn-Regnier~~Kuhn-Régnier, A.: era5analysis (Version 0.2.1), Zenodo, https://doi.org/10.5281/zenodo.4173493, ~~2020.~~ 2020.

~~Kuhn-Regnier~~Kuhn-Régnier, A.: ~~fuel-build-up~~empirical-fire-modelling (Version 0.1.1), Zenodo, ~~,~~ ~~2020~~https://doi.org/10.5281/zenodo.4739596, 2021a.

~~Kuhn-Regnier~~Kuhn-Régnier, A.: wildfires (Version 0.10.2), Zenodo, ~~, 2020~~https://doi.org/10.5281/zenodo.4739193, 2021b.

~~Kuhn-Regnier~~Kuhn-Régnier, A., Jumelle, M., and Rajaratnam, S.: ALEPython (Version 0.5.5), Zenodo, ~~, 2020.~~ https://doi.org/10.5281/zenodo.4739201, 2021.

Lasslop, G., Coppola, A. I., Voulgarakis, A., Yue, C., and Veraverbeke, S.: Influence of Fire on the Carbon Cycle and Climate, Curr Clim Change Rep, 5, 112–123, https://doi.org/10.1007/s40641-019-00128-9, 2019.

Li, W., MacBean, N., Ciais, P., Defourny, P., Lamarche, C., Bontemps, S., Moreau, I., Houghton, R. A., and Peng, S.: Gross and Net Land Cover Changes in the Main Plant Functional Types Derived from the Annual ESA CCI Land Cover Maps (1992–2015), Earth Syst. Sci. Data, 10, 219–234, https://doi.org/10.5194/essd-10-219-2018, 2018.

Littell, J. S., McKenzie, D., Peterson, D. L., and Westerling, A. L.: Climate and Wildfire Area Burned in Western U.S. Ecoprovinces, 1916–2003, Ecol. Appl., 19, 1003–1021, https://doi.org/10.1890/07-1183.1, 2009.

Lundberg, S. and Lee, S.-I.: A Unified Approach to Interpreting Model Predictions, in: Advances in Neural Information Processing Systems, edited by Guyon, I., Fergus, R., Wallach, H., von Luxburg, U., Garnett, R., Vishwanathan, S., and Bengio, S., vol. 2017-December, pp. 4766–4775, Neural information processing systems foundation, 2017.

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., and Lee, S.-I.: From Local Explanations to Global Understanding with Explainable AI for Trees, Nat. Mach. Intell., 2, 56–67, https://doi.org/10.1038/s42256-019-0138-9, 2020.

Mansfield, L. A., Nowack, P. J., Kasoar, M., Everitt, R. G., Collins, W. J., and Voulgarakis, A.: Predicting Global Patterns of Long-Term Climate Change from Short-Term Simulations Using Machine Learning, Npj Clim. Atmospheric Sci., 3, 1–9, https://doi.org/10.1038/s41612-020-00148-5, 2020.

Marlon, J. R., Bartlein, P. J., Gavin, D. G., Long, C. J., Anderson, R. S., Briles, C. E., Brown, K. J., Colombaroli, D., Hallett, D. J., Power, M. J., Scharf, E. A., and Walsh, M. K.: Long-Term Perspective on Wildfires in the Western USA, PNAS, 109, E535–E543, https://doi.org/10.1073/pnas.1112839109, 2012.

Martínez, J., Vega-Garcia, C., and Chuvieco, E.: Human-Caused Wildfire Risk Rating for Prevention Planning in Spain, Journal of Environmental Management, 90, 1241–1252, https://doi.org/10.1016/j.jenvman.2008.07.005, 2009.

Met Office: Iris: A Python Library for Analysing and Visualising Meteorological and Oceanographic Data Sets, Exeter, Devon, v2.4 edn., 2010.

Meyer, H., Reudenbach, C., Wöllauer, S., and Nauss, T.: Importance of Spatial Predictor Variable Selection in Machine Learning Applications – Moving from Data Reproduction to Spatial Prediction, Ecological Modelling, 411, 108 815, https://doi.org/10.1016/j.ecolmodel.2019.108815, 2019.

Moesinger, L., Dorigo, W., De Jeu, R., Van der Schalie, R., Scanlon, T., Teubner, I., and Forkel, M.: The Global Long-Term Microwave Vegetation Optical Depth Climate Archive VODCA (Version 1.0) [Data Set], https://doi.org/10.5281/zenodo.2575599, 2019.

Moesinger, L., Dorigo, W., de Jeu, R., van der Schalie, R., Scanlon, T., Teubner, I., and Forkel, M.: The Global Long-Term Microwave Vegetation Optical Depth Climate Archive (VODCA), Earth Syst. Sci. Data, 12, 177–196, https://doi.org/10.5194/essd-12-177-2020, 2020.

Mohammed, G. H., Colombo, R., Middleton, E. M., Rascher, U., van der Tol, C., Nedbal, L., Goulas, Y., Pérez-Priego, O., Damm, A., Meroni, M., Joiner, J., Cogliati, S., Verhoef, W., Malenovský, Z., Gastellu-Etchegorry, J.-P., Miller, J. R., Guanter, L., Moreno, J., Moya, I., Berry, J. A., Frankenberg, C., and Zarco-Tejada, P. J.: Remote Sensing of Solar-Induced Chlorophyll Fluorescence (SIF) in Vegetation: 50 years of Progress, Remote Sensing of Environment, 231, 111 177, https://doi.org/10.1016/j.rse.2019.04.030, 2019.

Molnar, C.: Interpretable Machine Learning, 2020.

Myneni, R., Knyazikhin, Y., and Park, T.: MOD15A2H MODIS/Terra Leaf Area Index/FPAR 8-Day L4 Global 500m SIN Grid V006, https://doi.org/10.5067/MODIS/MOD15A2H.006, 2015.

Nowack, P., Braesicke, P., Haigh, J., Abraham, N. L., Pyle, J., and Voulgarakis, A.: Using Machine Learning to Build Temperature-Based Ozone Parameterizations for Climate Sensitivity Simulations, Environ. Res. Lett., 13, 104 016, https://doi.org/10.1088/1748-9326/aae2be, 2018.

Nowack, P., Runge, J., Eyring, V., and Haigh, J. D.: Causal Networks for Climate Model Evaluation and Constrained Projections, Nat. Commun., 11, 1415, https://doi.org/10.1038/s41467-020-15195-y, 2020.

O, S., Hou, X., and Orth, R.: Observational Evidence of Wildfire-Promoting Soil Moisture Anomalies, Sci. Rep., 10, 11 008, https://doi.org/10.1038/s41598-020-67530-4, 2020.

Ogutu, B. O., Dash, J., and Dawson, T. P.: Evaluation of the Influence of Two Operational Fraction of Absorbed Photosynthetically Active Radiation (FAPAR) Products on Terrestrial Ecosystem Productivity Modelling, Int. J. Remote Sens., 35, 321–340, https://doi.org/10.1080/01431161.2013.871083, 2014.

Oliphant, T. E.: A Guide to NumPy, vol. 1, Trelgol Publishing USA, 2006.

Parks, S. A., Miller, C., Parisien, M.-A., Holsinger, L. M., Dobrowski, S. Z., and Abatzoglou, J.: Wildland Fire Deficit and Surplus in the Western United States, 1984–2012, Ecosphere, 6, 1–13, https://doi.org/10.1890/ES15-00294.1, 2015.

~~Pechony, O. and Shindell, D. T.: Driving Forces of Global Wildfires over the Past Millennium and the Forthcoming Century, Proc. Natl. Acad. Sci., 107, 19 167–19 170, , 2010.~~

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E.: Scikit-Learn: Machine Learning in Python, J. Mach. Learn. Res., 12, 2825–2830, 2011.

Pettinari, M. L. and Chuvieco, E.: Generation of a Global Fuel Data Set Using the Fuel Characteristic Classification System, Biogeosciences, 13, 2061–2076, https://doi.org/10.5194/bg-13-2061-2016, 2016.

Ploton, P., Mortier, F., Réjou-Méchain, M., Barbier, N., Picard, N., Rossi, V., Dormann, C., Cornu, G., Viennois, G., Bayol, N., Lyapustin, A., Gourlet-Fleury, S., and Pélissier, R.: Spatial Validation Reveals Poor Predictive Performance of Large-Scale Ecological Mapping Models, Nat. Commun., 11, 4540, https://doi.org/10.1038/s41467-020-18321-y, 2020.

Poulter, B., MacBean, N., Hartley, A., Khlystova, I., Arino, O., Betts, R., Bontemps, S., Boettcher, M., Brockmann, C., Defourny, P., Hagemann, S., Herold, M., Kirches, G., Lamarche, C., Lederer, D., Ottlé, C., Peters, M., and Peylin, P.: Plant Functional Type Classification for Earth System Models: Results from the European Space Agency's Land Cover Climate Change Initiative, Geosci. Model Dev., 8, 2315–2328, https://doi.org/10.5194/gmd-8-2315-2015, 2015.

Randerson, J. T., van der Werf, G. R., Collatz, G. J., Giglio, L., Still, C. J., Kasibhatla, P., Miller, J. B., White, J. W. C., DeFries, R. S., and Kasischke, E. S.: Fire Emissions from C3 and C4 Vegetation and Their Influence on Interannual Variability of Atmospheric CO2 and δ13CO2, Glob. Biogeochem. Cycles, 19, https://doi.org/10.1029/2004GB002366, 2005.

Rodger, C. J., Brundell, J. B., Dowden, R. L., and Thomson, N. R.: Location Accuracy of Long Distance VLF Lightning Locationnetwork, Ann. Geophys., 22, 747–758, https://doi.org/10.5194/angeo-22-747-2004, 2004.

Runge, J., Nowack, P., Kretschmer, M., Flaxman, S., and Sejdinovic, D.: Detecting and Quantifying Causal Associations in Large Nonlinear Time Series Datasets, Sci. Adv., 5, eaau4996, https://doi.org/10.1126/sciadv.aau4996, 2019.

Ryu, Y., Berry, J. A., and Baldocchi, D. D.: What Is Global Photosynthesis? History, Uncertainties and Opportunities, Remote Sensing of Environment, 223, 95–114, https://doi.org/10.1016/j.rse.2019.01.016, 2019.

Sanderson, B. M. and Fisher, R. A.: A Fiery Wake-up Call for Climate Science, Nat. Clim. Chang., 10, 175–177, https://doi.org/10.1038/s41558-020-0707-2, 2020.

Spessa, A., McBeth, B., and Prentice, C.: Relationships among Fire Frequency, Rainfall and Vegetation Patterns in the Wet–Dry Tropics of Northern Australia: An Analysis Based on NOAA-AVHRR Data, Glob. Ecol. Biogeogr., 14, 439–454, https://doi.org/10.1111/j.1466-822x.2005.00174.x, 2005.

Swetnam, T. W. and Betancourt, J. L.: Mesoscale Disturbance and Ecological Response to Decadal Climatic Variability in the American Southwest, J. Clim., 11, 3128–3147, https://doi.org/10.1175/1520-0442(1998)011<3128:MDAERT>2.0.CO;2, 1998.

Teckentrup, L., Harrison, S. P., Hantson, S., Heil, A., Melton, J. R., Forrest, M., Li, F., Yue, C., Arneth, A., Hickler, T., Sitch, S., and Lasslop, G.: Response of Simulated Burned Area to Historical Changes in Environmental and Anthropogenic Factors: A Comparison of Seven Fire Models, Biogeosciences, 16, 3883–3910, https://doi.org/10.5194/bg-16-3883-2019, 2019.

Teubner, I. E., Forkel, M., Jung, M., Liu, Y. Y., Miralles, D. G., Parinussa, R., van der Schalie, R., Vreugdenhil, M., Schwalm, C. R., Tramontana, G., Camps-Valls, G., and Dorigo, W. A.: Assessing the Relationship between Microwave Vegetation Optical Depth and Gross Primary Production, Int. J. Appl. Earth Obs. Geoinformation, 65, 79–91, https://doi.org/10.1016/j.jag.2017.10.006, 2018.

Thomas, P. B., Watson, P. J., Bradstock, R. A., Penman, T. D., and Price, O. F.: Modelling Surface Fine Fuel Dynamics across Climate Gradients in Eucalypt Forests of South-Eastern Australia, Ecography, 37, 827–837, https://doi.org/10.1111/ecog.00445, 2014.

Thurner, M., Beer, C., Santoro, M., Carvalhais, N., Wutzler, T., Schepaschenko, D., Shvidenko, A., Kompter, E., Ahrens, B., Levick, S. R., and Schmullius, C.: Carbon Stock and Density of Northern Boreal and Temperate Forests, Glob. Ecol. Biogeogr., 23, 297–310, https://doi.org/10.1111/geb.12125, 2014.

Turco, M., Rosa-Cánovas, J. J., Bedia, J., Jerez, S., Montávez, J. P., Llasat, M. C., and Provenzale, A.: Exacerbated Fires in Mediterranean Europe Due to Anthropogenic Warming Projected with Non-Stationary Climate-Fire Models, Nat Commun, 9, 3821, https://doi.org/10.1038/s41467-018-06358-z, 2018.

van der Werf, G. R., Randerson, J. T., Giglio, L., Gobron, N., and Dolman, A. J.: Climate Controls on the Variability of Fires in the Tropics and Subtropics, Glob. Biogeochem. Cycles, 22, https://doi.org/10.1029/2007GB003122, 2008.

van Oldenborgh, G. J., Krikken, F., Lewis, S., Leach, N. J., Lehner, F., Saunders, K. R., van Weele, M., Haustein, K., Li, S., Wallom, D., Sparrow, S., Arrighi, J., Singh, R. P., van Aalst, M. K., Philip, S. Y., Vautard, R., and Otto, F. E. L.: Attribution of the Australian Bushfire Risk to Anthropogenic Climate Change, Nat. Hazards Earth Syst. Sci. Discuss., pp. 1–46, https://doi.org/10.5194/nhess-2020-69, 2020.

Van Rossum, G. and Drake, F. L.: Python 3 Reference Manual, CreateSpace, Scotts Valley, CA, 2009.

Van Wilgen, B. W., Biggs, H., O'Regan, S. P., and Mare, N.: Fire History of the Savanna Ecosystems in the Kruger National Park, South Africa, between 1941 and 1996, South Afr. J. Sci., 96, 2000.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Jarrod Millman, K., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C., Polat, İ., Feng, Y., Moore, E. W., Vand erPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and Contributors, S. . .: SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python, Nat. Methods, https://doi.org/10.1038/s41592-019-0686-2, 2020.

Voulgarakis, A. and Field, R. D.: Fire Influences on Atmospheric Composition, Air Quality and Climate, Curr Pollution Rep, 1, 70–81, https://doi.org/10.1007/s40726-015-0007-z, 2015.

Wagner, W., Lemoine, G., and Rott, H.: A Method for Estimating Soil Moisture from ERS Scatterometer and Soil Data, Remote Sensing of Environment, 70, 191–207, https://doi.org/10.1016/S0034-4257(99)00036-X, 1999.

Westerling, A. L.: Warming and Earlier Spring Increase Western U.S. Forest Wildfire Activity, Science, 313, 940–943, https://doi.org/10.1126/science.1128834, 2006.

~~Yang, J., Tian, H., Tao, B., Ren, W., Kush,~~

Westerling, A. L., Gershunov, A., Brown, T. J., ~~Liu, Y~~Cayan, D. R., and ~~Wang, Y.: Spatial and Temporal Patterns of Global Burned Area in Response to Anthropogenic and Environmental Factors:~~ Dettinger, M. D.: Climate and ~~Reconstructing~~Wildfire ~~Global Fire History for the 20th and Early 21st Centuries, J. Geophys. Res.Biogeosciences, 119, 249–263, , 2014.~~ in the Western United States, Bull. Am. Meteorol. Soc., 84, 595–604, https://doi.org/10.1175/BAMS-84-5-595, 2003.

Table S1. Variables used in the experiments. 'C' denotes current-month variables, 'all A' represents all antecedent months (1M–24M), and 1M represents one-month antecedent variables, with similar notation for other antecedent months.

	DD	SWI	MaxT	DTR	Light-ning	CROP	POPD	HERB	SHRUB	TREE	AGB	VOD	FAPAR	LAI	SIF
ALL	C & all A	C	C	C	C	C	C	C	C	C	C	C & all A	C & all A	C & all A	C & all A
ALL_NN	C & all A	C	C	C	C	C	C	C	C	C	C	C & all A	C & all A	C & all A	C & all A
CURR	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
BEST15	C, 1M, 3M, 6M, 9M		C	C	C	C	C						C, 1M	3M	6M, 9M
TOP15	C, 1M, 3M, 9M		C			C	C					C, 1M, 3M	C, 1M	1M, 3M	C
15VEG_FAPAR	C, 1M, 3M, 6M, 9M		C	C	C	C	C						C, 1M, 3M, 6M, 9M		
15VEG_FAPAR-_MON	C, 1M, 3M, 6M, 9M		C	C		C	C				C		C, 1M, 3M, 6M, 9M		
15VEG_LAI	C, 1M, 3M, 6M, 9M		C	C	C	C	C							C, 1M, 3M, 6M, 9M	
15VEG_SIF	C, 1M, 3M, 6M, 9M		C	C	C	C	C								C, 1M, 3M, 6M, 9M
15VEG_VOD	C, 1M, 3M, 6M, 9M		C	C	C	C	C					C, 1M, 3M, 6M, 9M			
CURRDD_FAPAR	C	C	C	C	C	C		C	C	C	C		C, 1M, 3M, 6M, 9M		
CURRDD_LAI	C	C	C	C	C	C		C	C	C	C			C, 1M, 3M, 6M, 9M	
CURRDD_SIF	C	C	C	C	C	C		C	C	C	C				C, 1M, 3M, 6M, 9M
CURRDD_VOD	C	C	C	C	C	C		C	C	C	C	C, 1M, 3M, 6M, 9M			

Table S2. Ranked importance of variables in the RF experiments according to the composite importance measure introduced in Sect. 2.4.

	ALL	ALL_NN	TOP15	CURR	15VEG_FAPAR	15VEG_LAI	15VEG_SIF	15VEG_VOD	CURRDD_FAPAR	CURRDD_LAI	CURRDD_SIF	CURRDD_VOD	BEST15
1	DD	DD	FAPAR	DD	FAPAR	LAI	SIF	VOD 1M	FAPAR 1M	LAI	SIF	VOD 1M	FAPAR
2	FAPAR	FAPAR	DD	MaxT	DD	DD	DD	DD	FAPAR	LAI 1M	DD	VOD	DD
3	VOD 1M	VOD 3M	MaxT	TREE	FAPAR 1M	LAI 1M	MaxT	VOD	DD	DD	MaxT	VOD 3M	FAPAR 1M
4	VOD 3M	VOD 1M	VOD 3M	VOD	MaxT	LAI 3M	CROP	VOD 3M	MaxT	LAI 3M	SIF 3M	DD	LAI 3M
5	MaxT	MaxT	SIF	SWI	CROP	MaxT	SIF 6M	MaxT	FAPAR 6M	MaxT	SIF 6M	MaxT	CROP
6	SIF	SIF	DD 9M	LAI	DD 1M	CROP	DD 1M	VOD 9M	FAPAR 3M	LAI 6M	SIF 1M	VOD 9M	MaxT
7	FAPAR 1M	VOD	LAI 1M	SIF	DD 3M	LAI 6M	DD 3M	DD 9M	CROP	HERB	SIF 9M	VOD 6M	SIF 9M
8	CROP	FAPAR 1M	VOD	FAPAR	FAPAR 6M	DD 1M	SIF 9M	CROP	FAPAR 9M	LAI 9M	TREE	AGB	DD 1M
9	VOD	LAI 1M	DD 1M	HERB	FAPAR 9M	DD 3M	SIF 3M	DD 3M	HERB	CROP	CROP	DTR	POPD
10	DD 1M	DD 1M	FAPAR 1M	DTR	FAPAR 3M	DD 9M	SIF 1M	VOD 6M	Lightning	Lightning	Lightning	Lightning	DD 9M
11	LAI 1M	LAI 3M	CROP	Lightning	DD 6M	LAI 9M	DD 6M	DD 1M	DTR	DTR	DTR	HERB	DD 6M
12	LAI 3M	CROP	LAI 3M	AGB	DD 9M	DD 6M	DD 9M	DD 6M	SWI	SHRUB	SWI	SHRUB	Lightning
13	DD 3M	POPD	VOD 1M	CROP	POPD	POPD	DTR	POPD	SHRUB	SWI	SHRUB	SWI	SIF 6M
14	DD 9M	DD 9M	DD 3M	SHRUB	Lightning	Lightning	Lightning	DTR	TREE	TREE	HERB	TREE	DD 3M
15	POPD	DD 3M	POPD	POPD	DTR	DTR	POPD	Lightning	AGB	AGB	AGB	CROP	DTR
16	SIF 9M	FAPAR 6M											
17	FAPAR 6M	SIF 9M											
18	LAI	DD 6M											
19	DD 6M	VOD 9M											
20	VOD 9M	Lightning											
21	Lightning	LAI											
22	DTR	FAPAR 9M											
23	AGB	FAPAR 3M											
24	SHRUB	DD Δ12M											
25	FAPAR 9M	DTR											
26	SIF 6M	AGB											
27	LAI 6M	LAI 6M											
28	SWI	SWI											
29	FAPAR 3M	SHRUB											
30	VOD Δ12M	SIF 6M											
31	DD Δ12M	LAI 9M											
32	SIF 3M	SIF 3M											
33	LAI 9M	VOD 6M											
34	VOD 6M	VOD Δ12M											
35	VOD Δ24M	SIF 1M											
36	VOD Δ18M	HERB											
37	SIF 1M	TREE											
38	LAI Δ24M	VOD Δ24M											
39	TREE	LAI Δ12M											
40	FAPAR Δ24M	SIF Δ24M											
41	HERB	DD Δ24M											
42	SIF Δ24M	LAI Δ24M											
43	FAPAR Δ12M	FAPAR Δ24M											
44	SIF Δ18M	SIF Δ18M											
45	SIF Δ12M	SIF Δ12M											
46	DD Δ24M	FAPAR Δ12M											
47	DD Δ18M	VOD Δ18M											
48	LAI Δ12M	LAI Δ18M											
49	FAPAR Δ18M	DD Δ18M											
50	LAI Δ18M	FAPAR Δ18M											

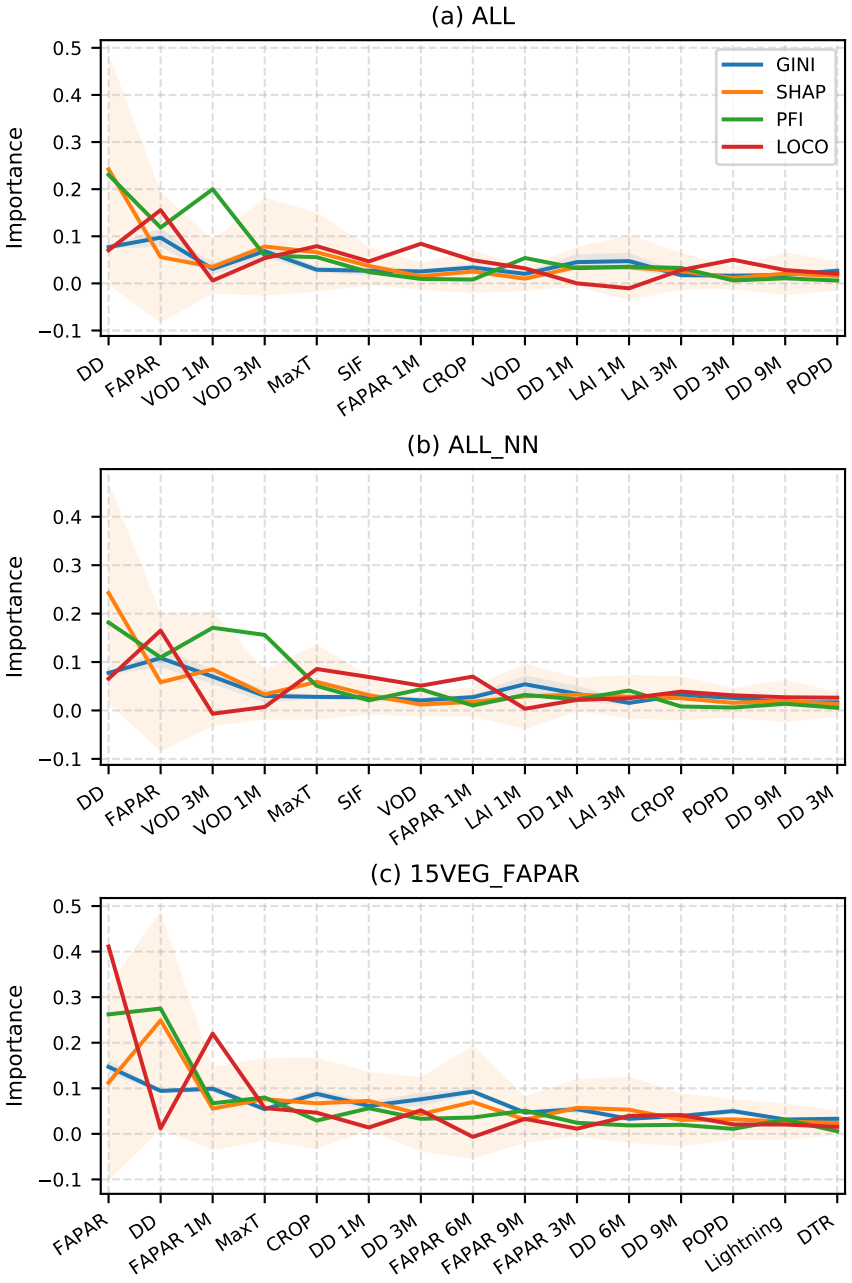

Figure S1. Transformed variable importance metrics (Gini, PFI, SHAP, and LOCO) for the (a) ALL, (b) ALL_NN, and (c) 15VEG_FAPAR models. The 15 most important variables (with others omitted for clarity) are sorted by their combined importance with the most important on the left. Uncertainties using the standard deviation are indicated using shaded regions. The uncertainty magnitudes differ between the metrics due to the way they are calculated; SHAP values are calculated for every sample, Gini importances are calculated based on splits for individual decision trees, PFI calculations are repeated after permuting the original dataset, and LOCO importances are only calculated once. Therefore, based on the number of samples used for their calculation, the SHAP importances are expected to have the highest variance, followed by the Gini and then PFI importances and lastly the LOCO importances without any quantification of the error.

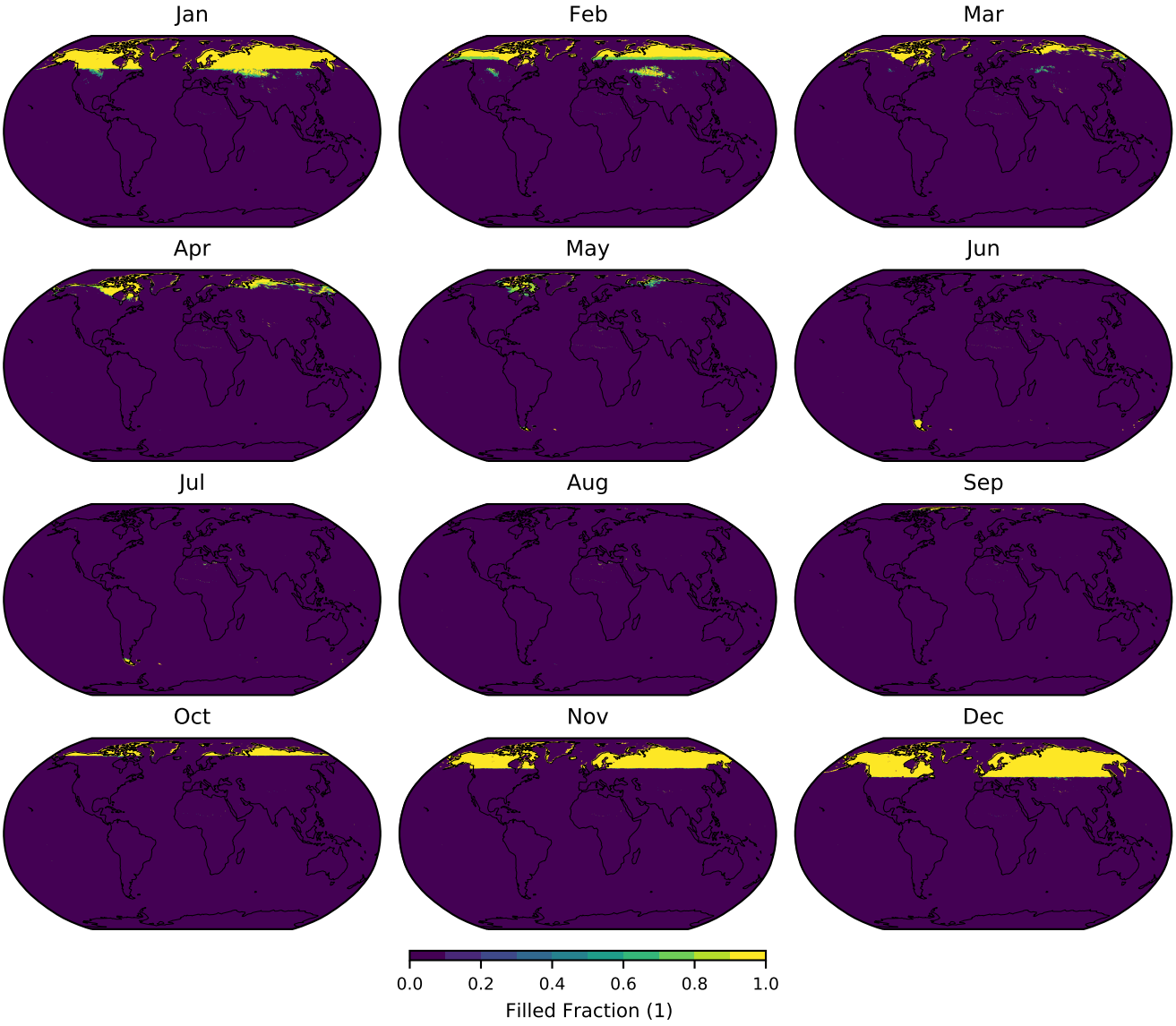

Figure S2. The fraction of filled samples for FAPAR (January 2008 to April 2015) at a given location for each month, with yellow indicating that all occurrences of a given month at a given location were filled and purple indicating no filling was done. Filling is mostly carried out in winter in northern latitudes.

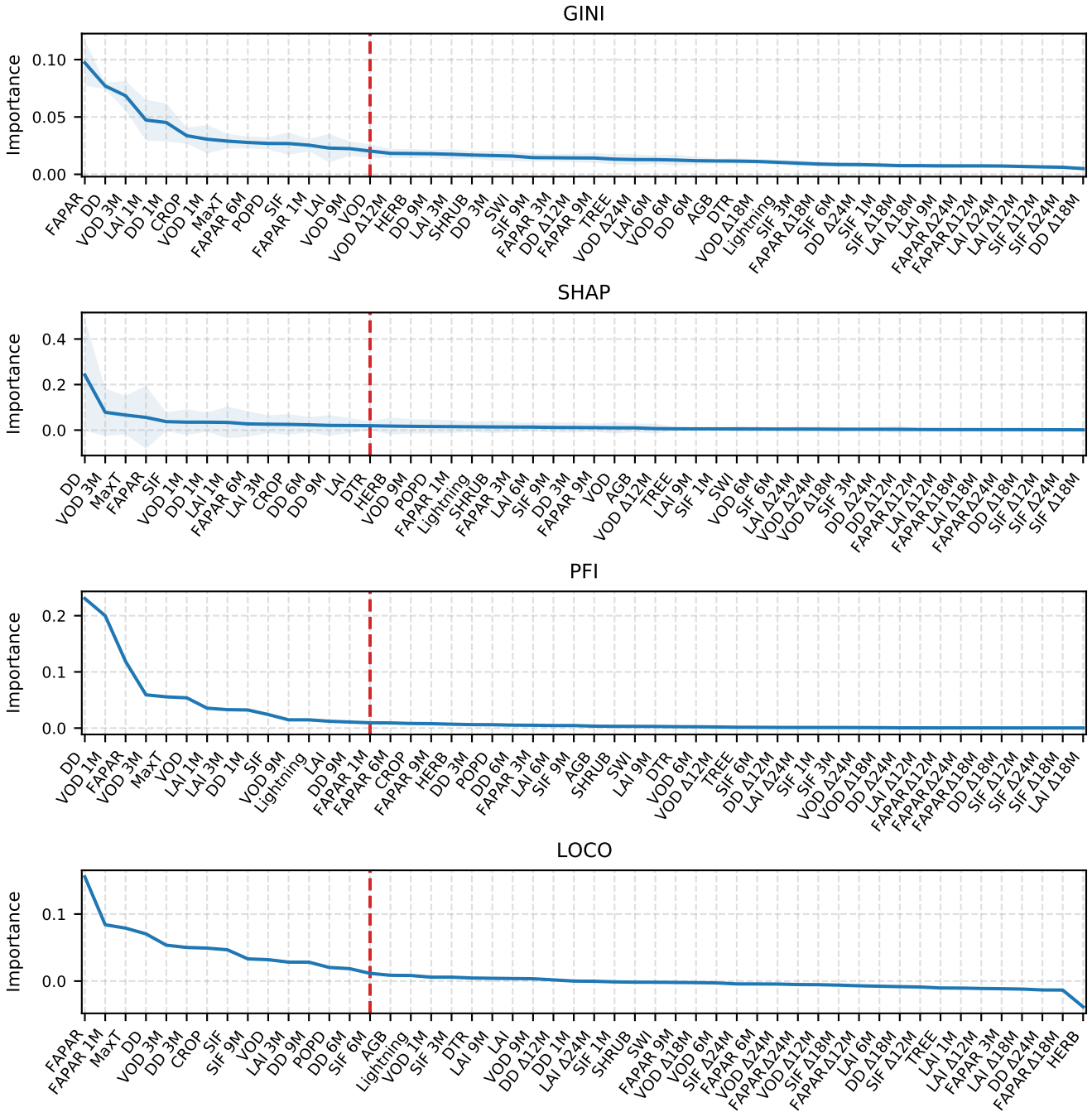

Figure S3. Sorted variable importance metrics (Gini, SHAP, PFI, and LOCO) for the ALL model, with the highest variable importance according to each metric on the left. The dotted red line indicates the 15th variable. Uncertainties are calculated using the standard deviation and indicated using the shaded regions.

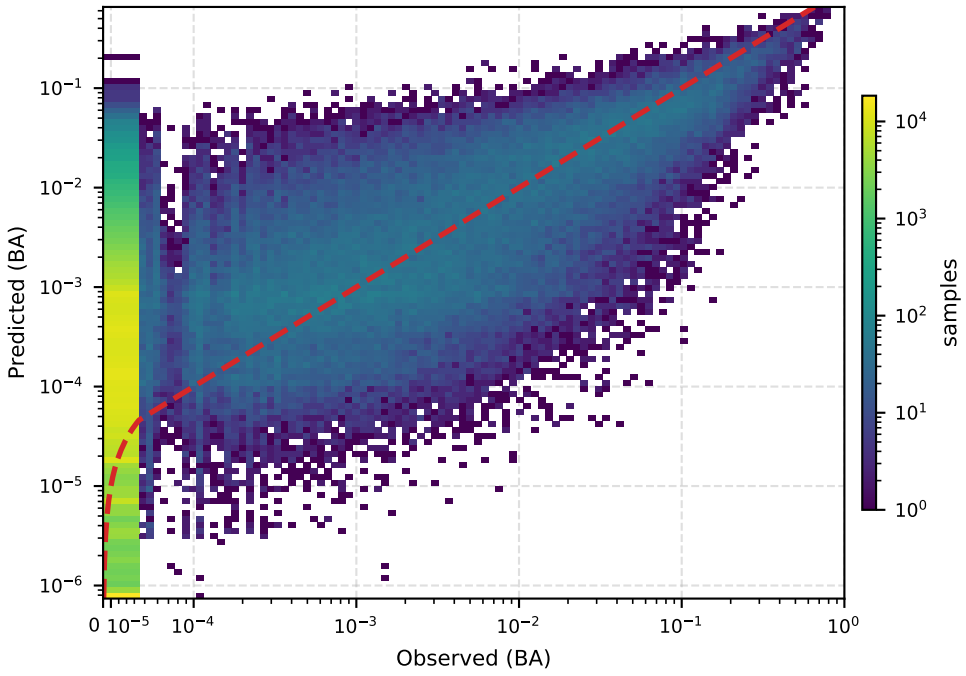

Figure S4. Out-of-sample BA predictions by the ALL model and corresponding observations. Note that logarithmic scales are used throughout except for the lower end of the x-axis, where a linear scale is used.

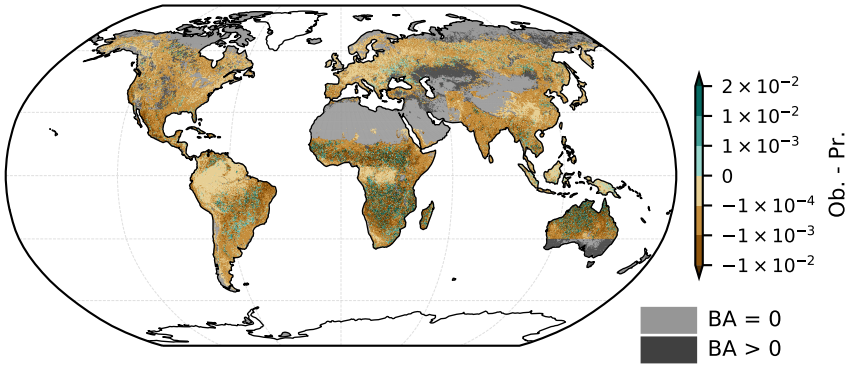

Figure S5. Mean difference between the out-of-sample observed (Ob.) and predicted (Pr.; by the ALL model) BA. The major spatial patterns follow the magnitude of mean BA (see Fig. 2a), with (on average) underprediction most prevalent in regions with large mean BA and vice versa. Note that sharp data availability boundaries (e.g. in western Asia, southern Australia) are introduced by the AGB dataset. Grey shading indicates 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 non-zero mean BA.

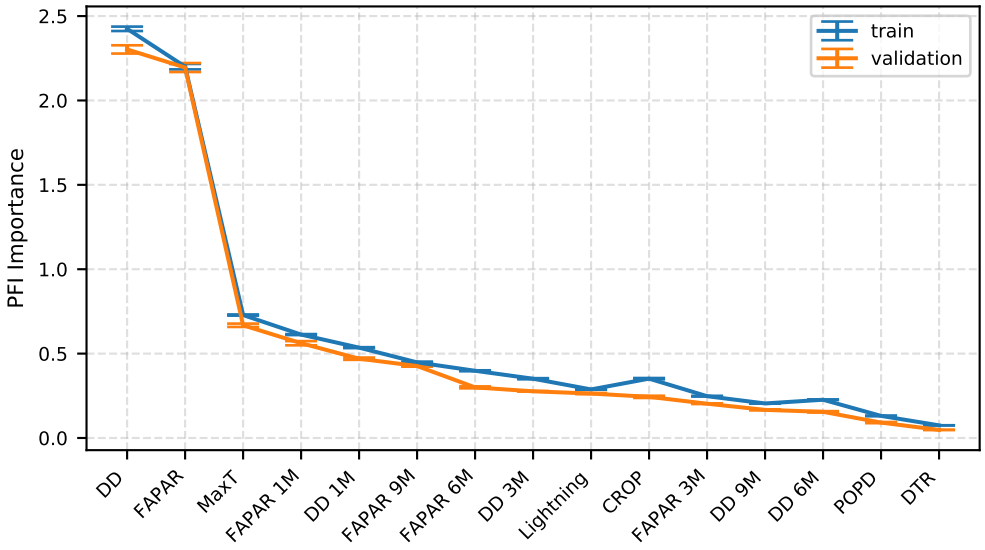

Figure S6. PFI importances for the 15VEG_FAPAR model computed separately on the training and validation sets. The error bars originate from repeated shuffling of the investigated variable.

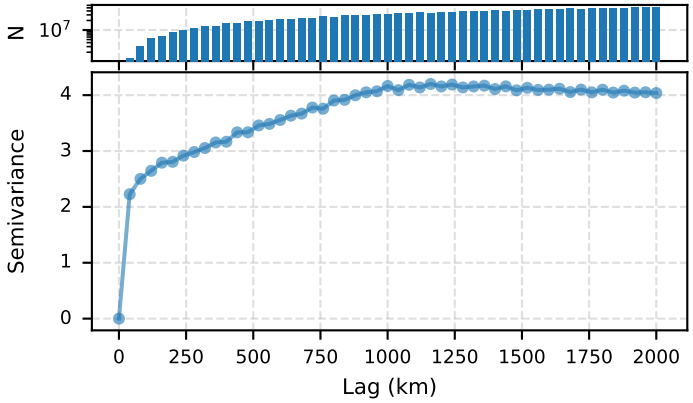

Figure S7. Variogram of mean GFED4 BA (June 1995 to December 2016) using all 237373 available samples. Semivariance can be seen to increase until ∼1000 km. Note that a logarithmic scale is used for the sample counts at the top.

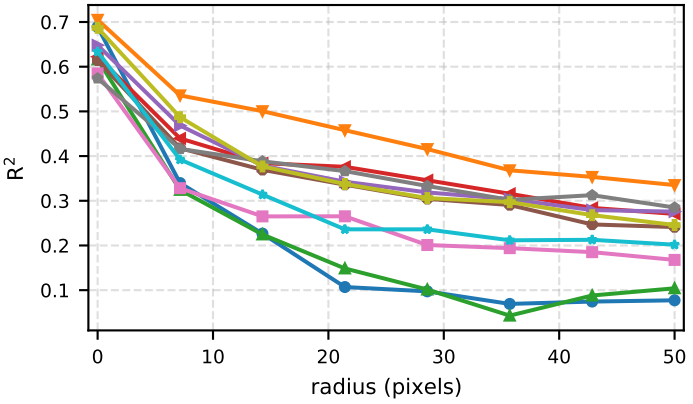

Figure S8. R^2 scores for burnt area (BA) prediction on test samples for different exclusion radii around the test samples using the 15VEG_FAPAR model. Each of the 10 lines represents the R^2 score for 400 test samples computed for the shown radii, where each individual test sample is chosen randomly and surrounded by a circular region of ignored data that is not used for training, with varying radii as shown. The disagreement between the lines is indicative of the statistical uncertainty, regional variability, and potentially different degrees of model extrapolation.

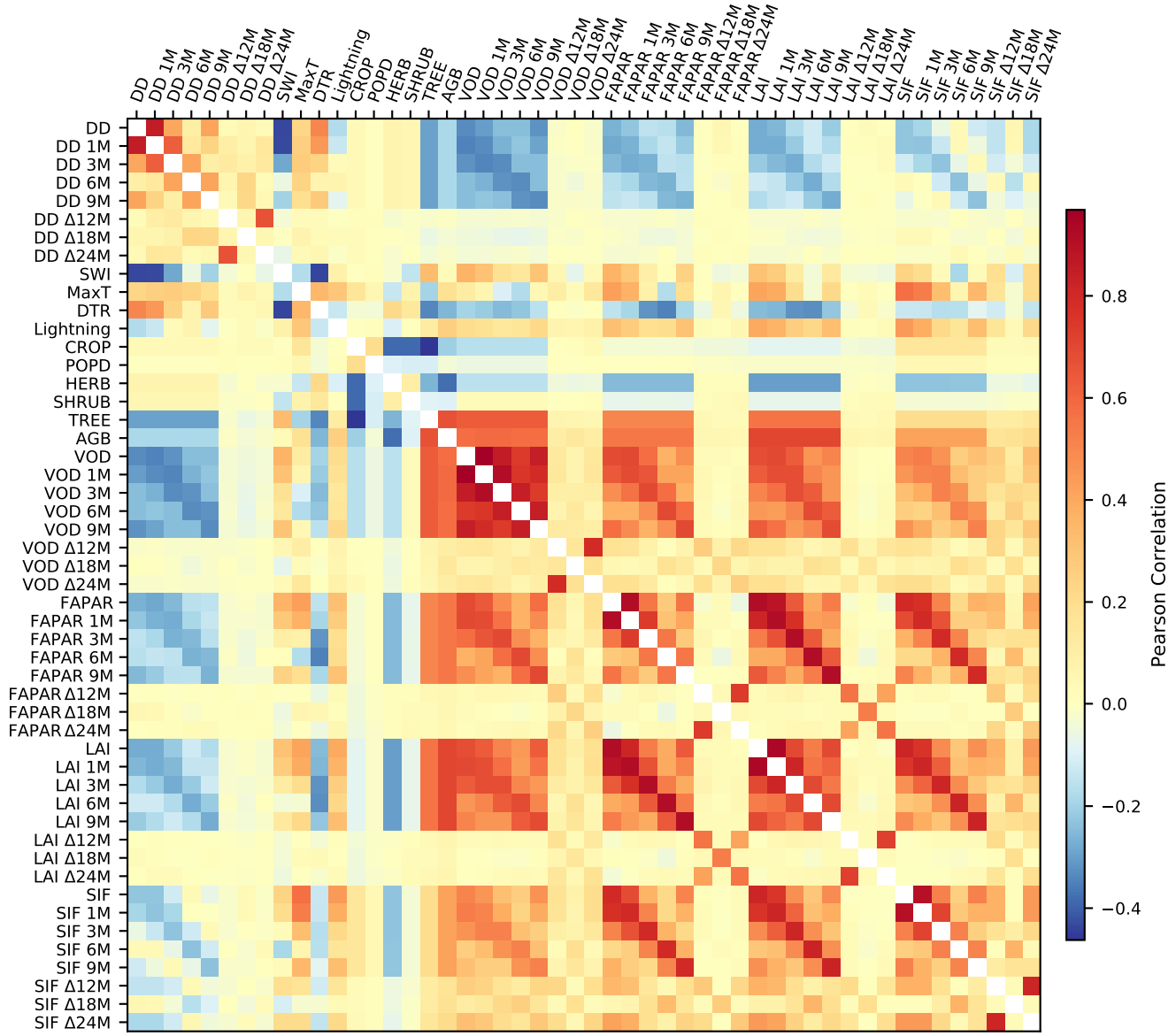

Figure S9. Pearson correlations between all variables used in the analysis for the time period from January 2010 to April 2015. Especially large positive correlations exist between variable pairs separated by multiples of 12 months and between FAPAR, LAI, SIF, and VOD. The largest negative correlations are found between SWI and DD (instantaneous, 12 month, 24 month), SWI and DTR, and CROP and TREE.

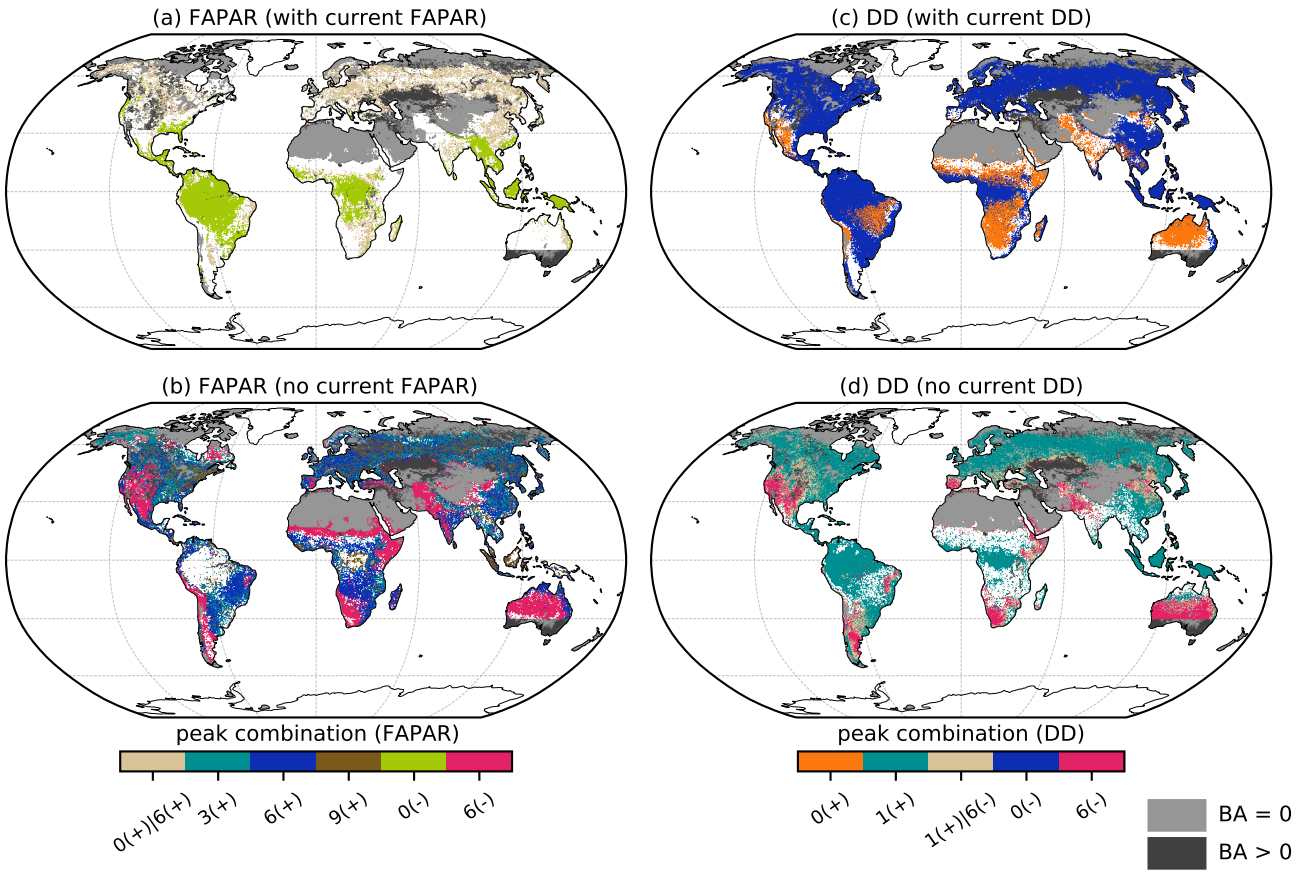

Figure S10. Spatial distribution of individual peak combinations for SHAP values as in Fig. 7. The sign of the maximum effect on BA at a certain antecedent month is indicated in parentheses after each month. The peak combinations are shown here such that their ordering has no significance (i.e. $0(+)|3(+)$ equals $3(+)|0(+)$). Dominant antecedent periods are apparent from Fig. 7. Most clearly, the general limitation of BA by instantaneous DD in tropical and boreal regions is seen in (c), combined with the positive effect of instantaneous DD on burning in the remaining regions. The limitation of BA by instantaneous DD shown in (c) generally agrees with the enhancement of BA by three-month antecedent FAPAR shown in (b), as well as the enhancement by one-month antecedent DD in (d). Note that sharp data availability boundaries (e.g. in western Asia, southern Australia) are introduced by the AGB dataset. Grey shading indicates 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 non-zero mean BA.

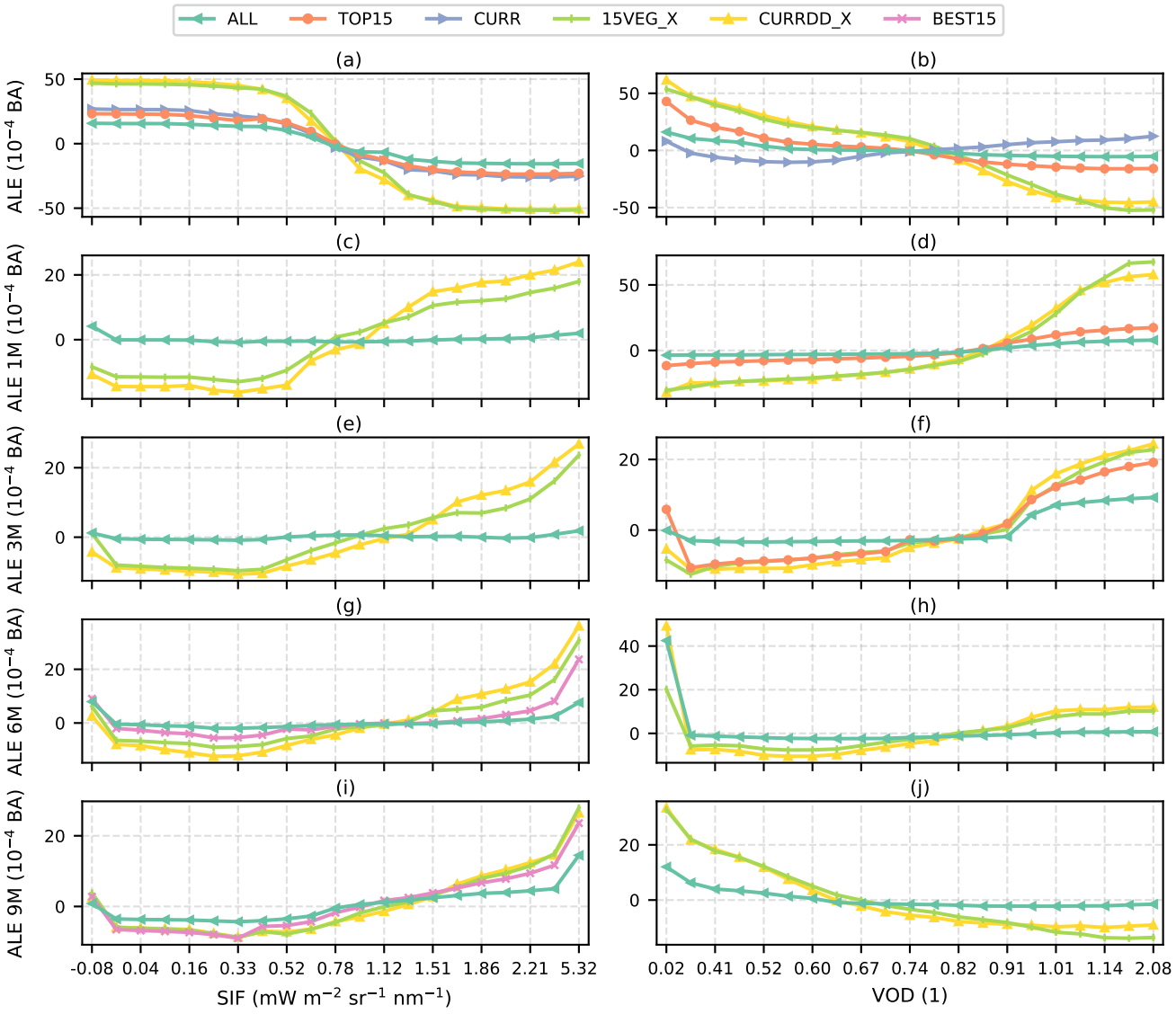

Figure S11. First-order LAI ALEs for different lags (< 1 yr) from all relevant modelling experiments for the relationships between BA and SIF (left hand columns) and VOD (right hand columns). Evenly spaced quantiles were used in the construction of the plots.

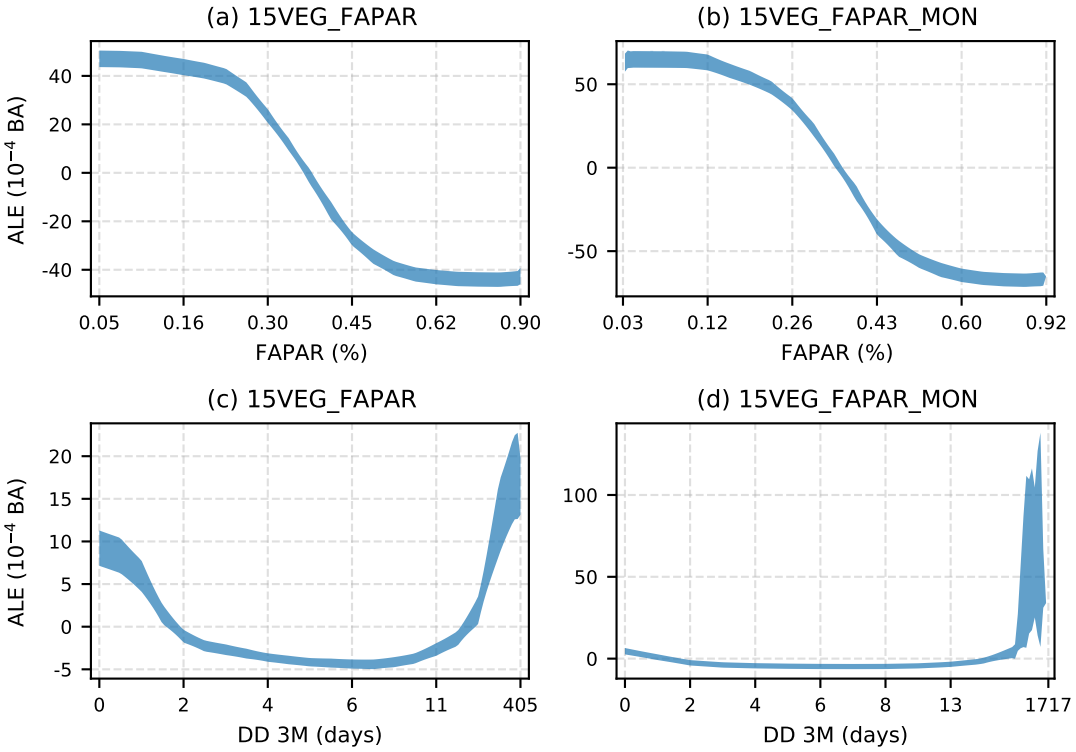

Figure S12. First-order ALE plots showing the effect of FAPAR (a, b) and the 3-month antecedent dry-day period (DD 3M; c, d) on burnt area (BA) in the 15VEG_FAPAR model (a, c) and the 15VEG_FAPAR_MON model (b, d) after accounting for all other variables. The shaded regions represent the standard deviation around the mean of 100 ALEs each using 122567 random samples of the training data.

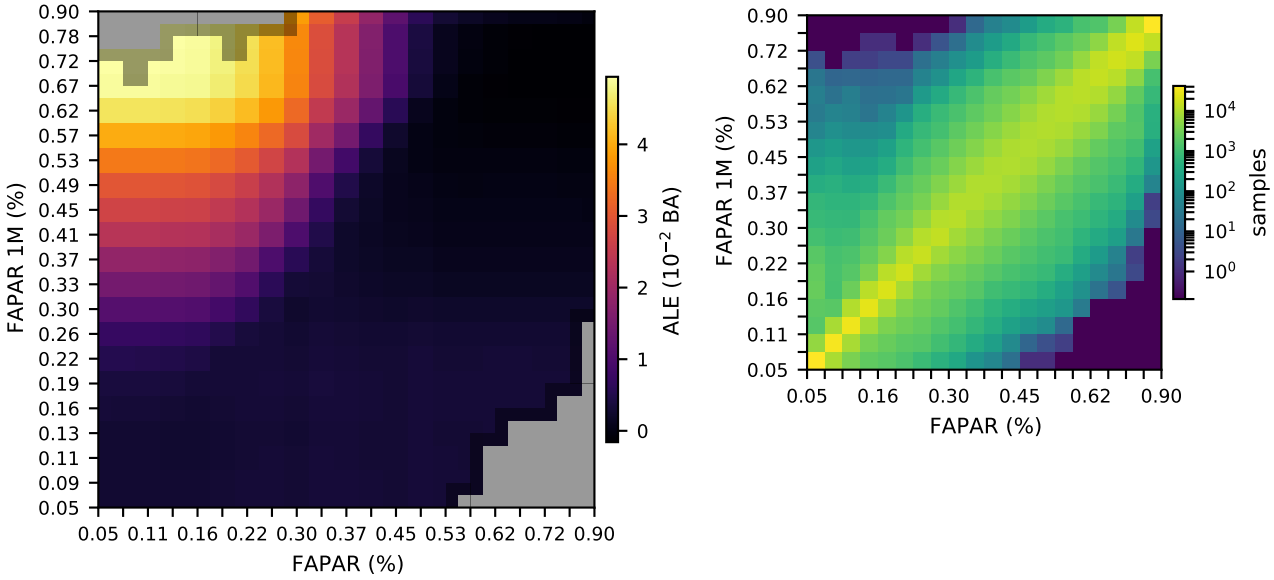

Figure S13. Second-order ALE plot showing the combined zeroth order (mean), first order, and second order modelled effects of FAPAR and FAPAR 1M on BA from the 15VEG_FAPAR model, taking into account all other variables. Grey boxes indicate missing data. The diagonal structure of the sample count matrix demonstrates the correlation between these variables. Evenly spaced quantiles were used in the construction and labelling of the plots.

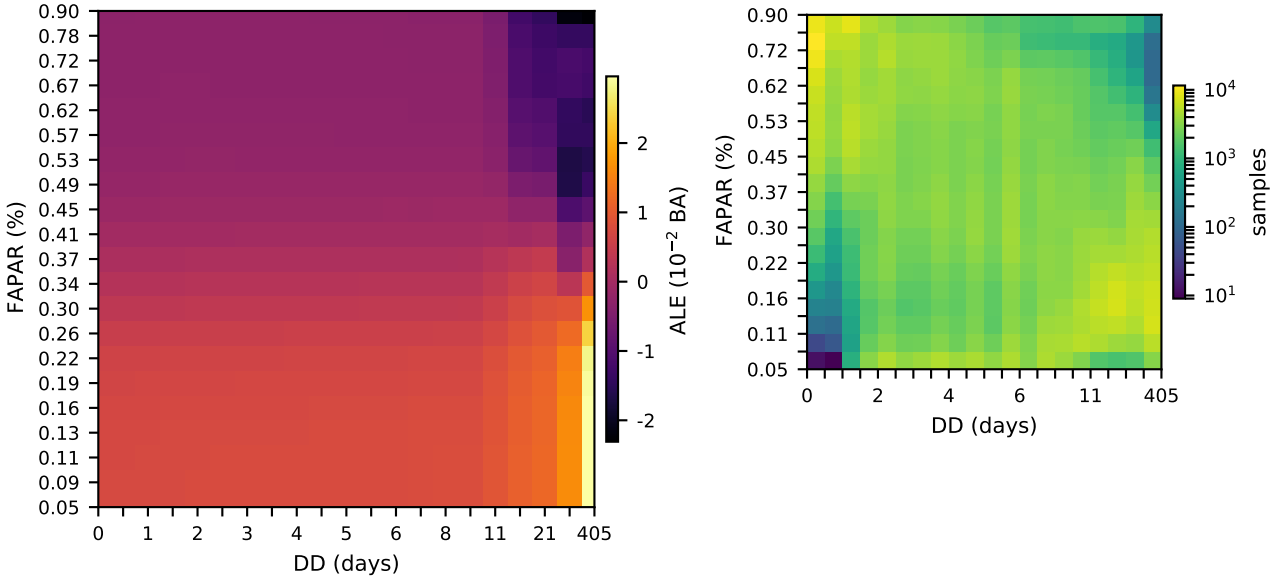

Figure S14. Second-order ALE plot showing the combined zeroth order (mean), first order, and second order modelled effects of DD and FAPAR on BA from the 15VEG_FAPAR model, taking into account all other variables. The diagonal structure of the sample count matrix demonstrates the anticorrelation between these variables. Evenly spaced quantiles were used in the construction and labelling of the plots.