Supplement of Biogeosciences, 18, 3861–3879, 2021
https://doi.org/10.5194/bg-18-3861-2021-supplement

Biogeosciences
Open Access

Supplement of

The importance of antecedent vegetation and drought conditions as global drivers of burnt area

Alexander Kuhn-Régnier et al.

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

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	Lightning	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											

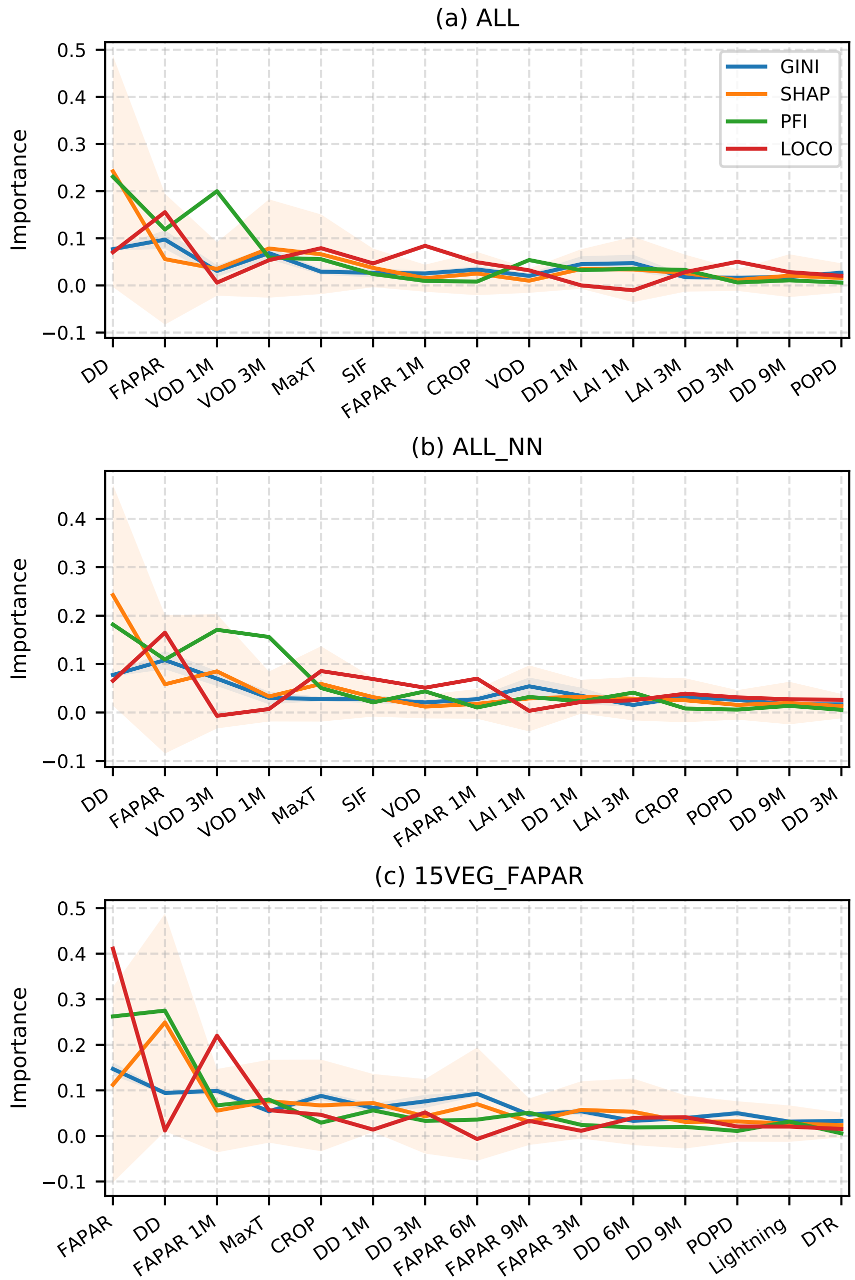

Figure S1. Transformed variable importance metrics (Gini, PFI, SHAP, and LOCO) for the (a) ALL, (b) ALL_NN, and (c) 15VEG_FAPAR models. The 15 most important variables (with others omitted for clarity) are sorted by their combined importance with the most important on the left. Uncertainties using the 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.

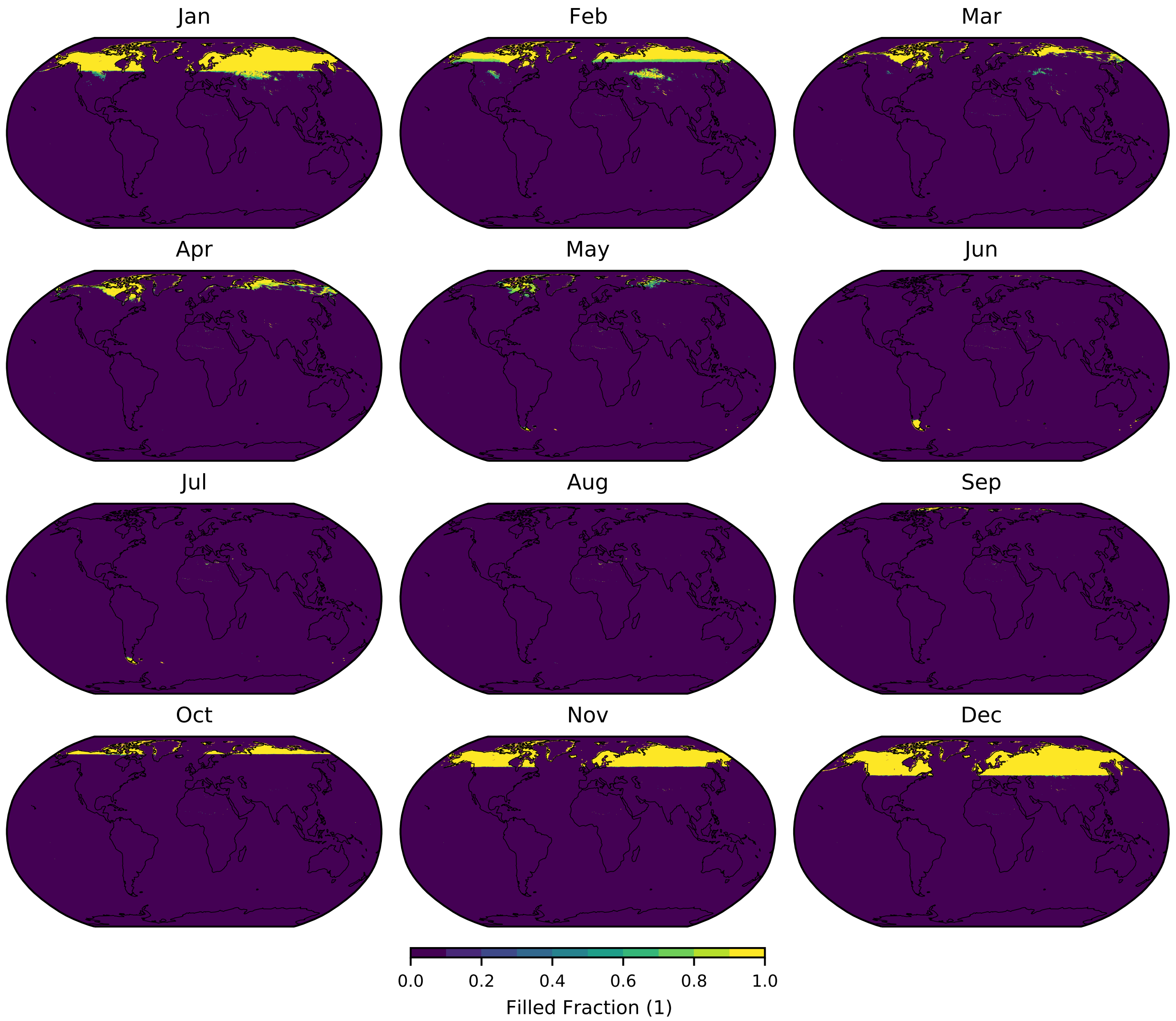

Figure S2. The fraction of filled samples for FAPAR (January 2008 to April 2015) at a given location for each month, with yellow indicating that all occurrences of a given month at a given location were filled and purple indicating no filling was done. Filling is mostly carried out in winter in northern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

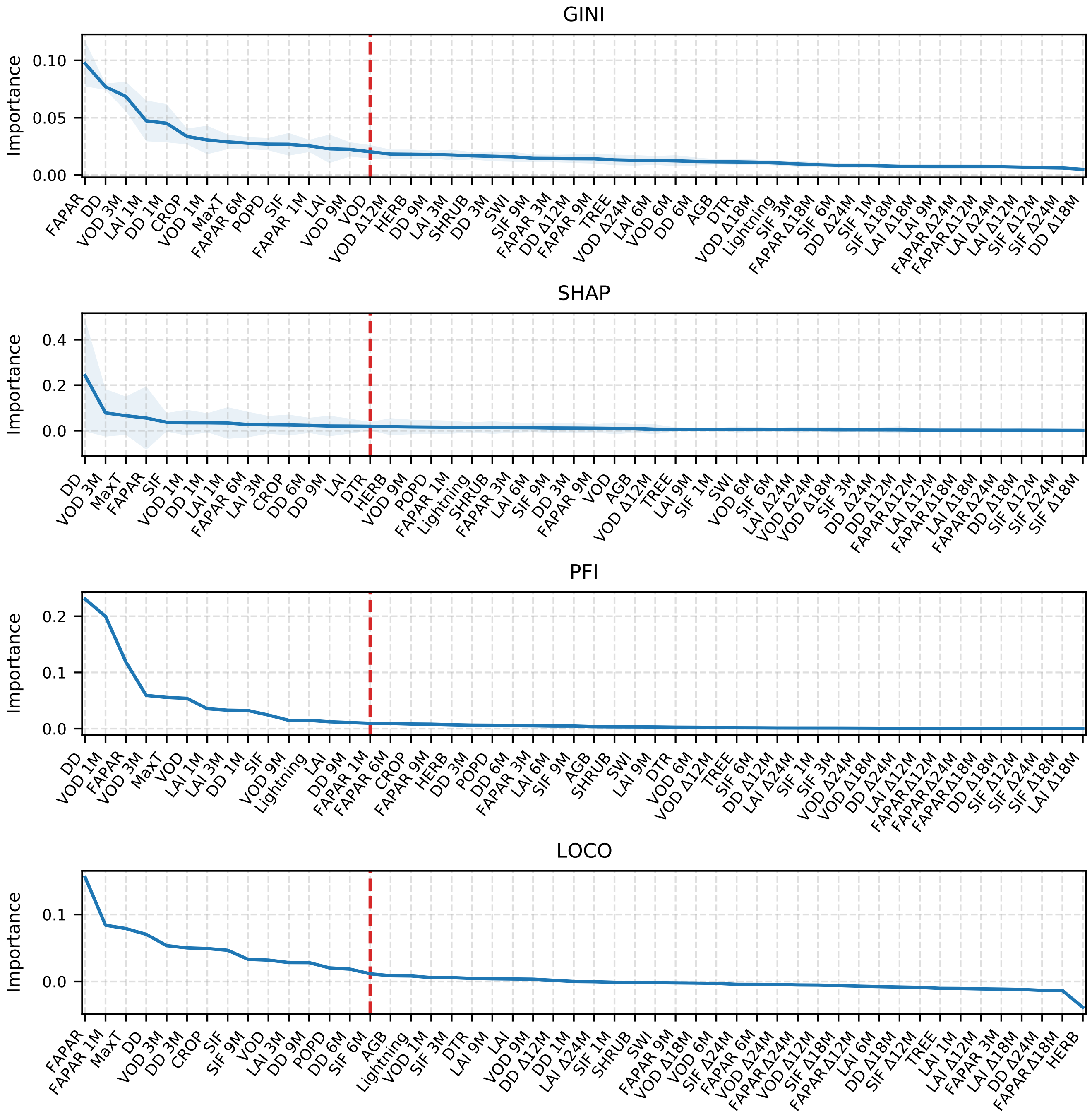

Figure S3. Sorted variable importance metrics (Gini, SHAP, PFI, and LOCO) for the ALL model, with the highest variable importance according to each metric on the left. The dotted red line indicates the 15th variable. Uncertainties are calculated using the standard deviation and indicated using the shaded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

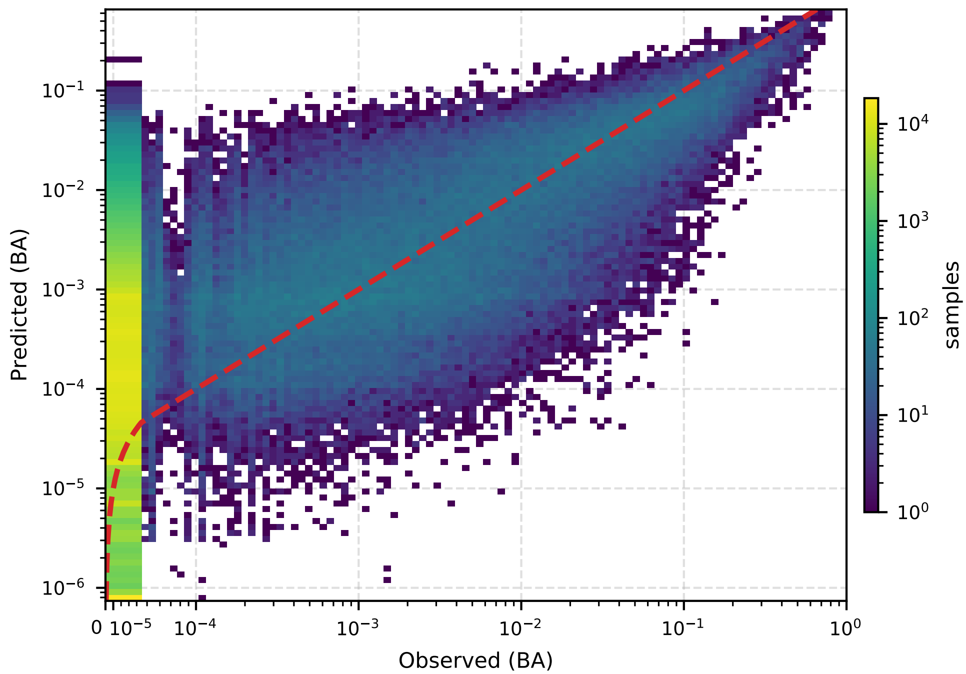

Figure S4. Out-of-sample BA predictions by the ALL model and corresponding observations. Note that logarithmic scales are used throughout except for the lower end of the x-axis, where a linear scale is used.

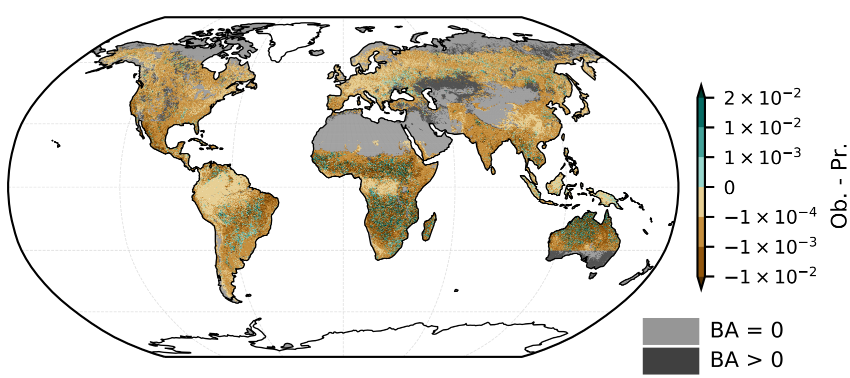

Figure S5. Mean difference between the out-of-sample observed (Ob.) and predicted (Pr.; by the ALL model) BA. The major spatial patterns follow the magnitude of mean BA (see Fig. 2a), with (on average) underprediction most prevalent in regions with large mean BA and vice versa. Note that sharp data availability 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.

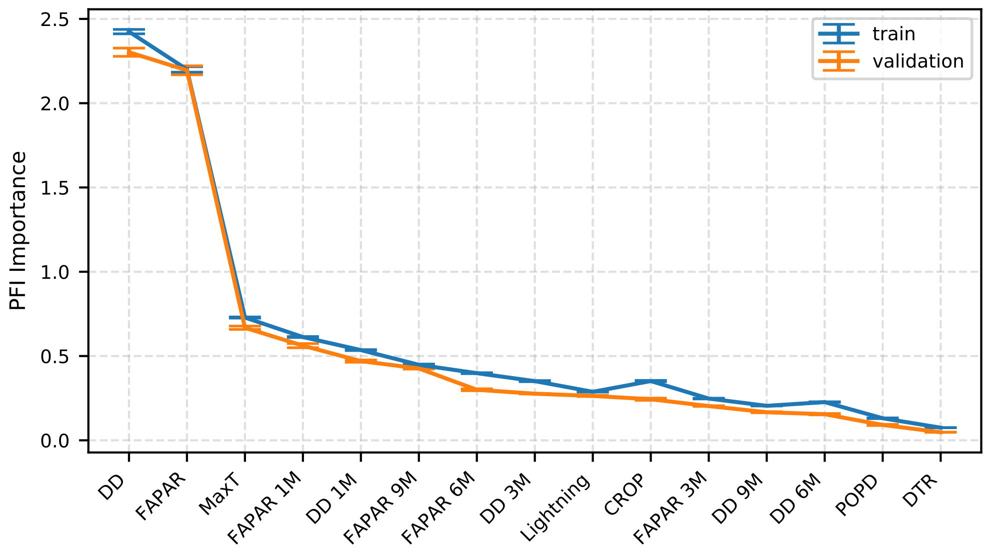

Figure S6. PFI importances for the 15VEG_FAPAR model computed separately on the training and validation sets. The error bars originate from repeated shuffling of the investigated variable.

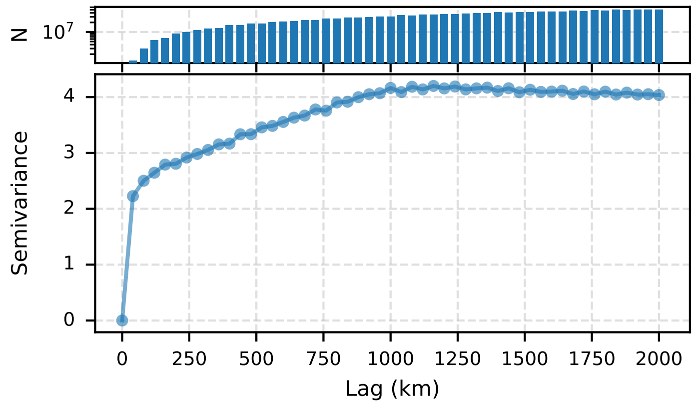

Figure S7. Variogram of mean GFED4 BA (June 1995 to December 2016) using all 237373 available samples. Semivariance can be seen to increase until $\sim$1000 km. Note that a logarithmic scale is used for the sample counts at the top.

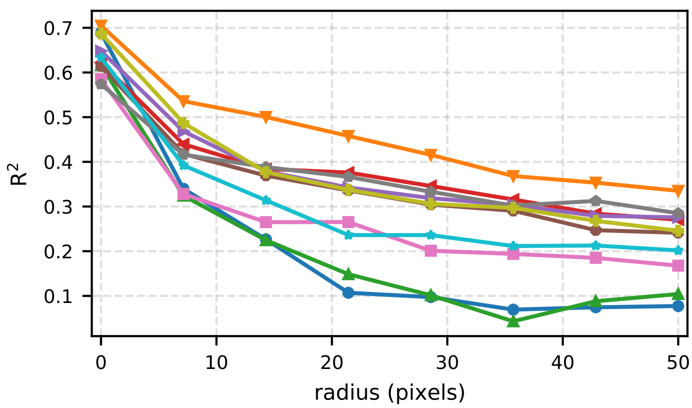

Figure S8. R^2 scores for burnt area (BA) prediction on test samples for different exclusion radii around the test samples using the 15VEG_FAPAR model. Each of the 10 lines represents the R^2 score for 400 test samples computed for the shown radii, where each individual test sample is chosen randomly 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.

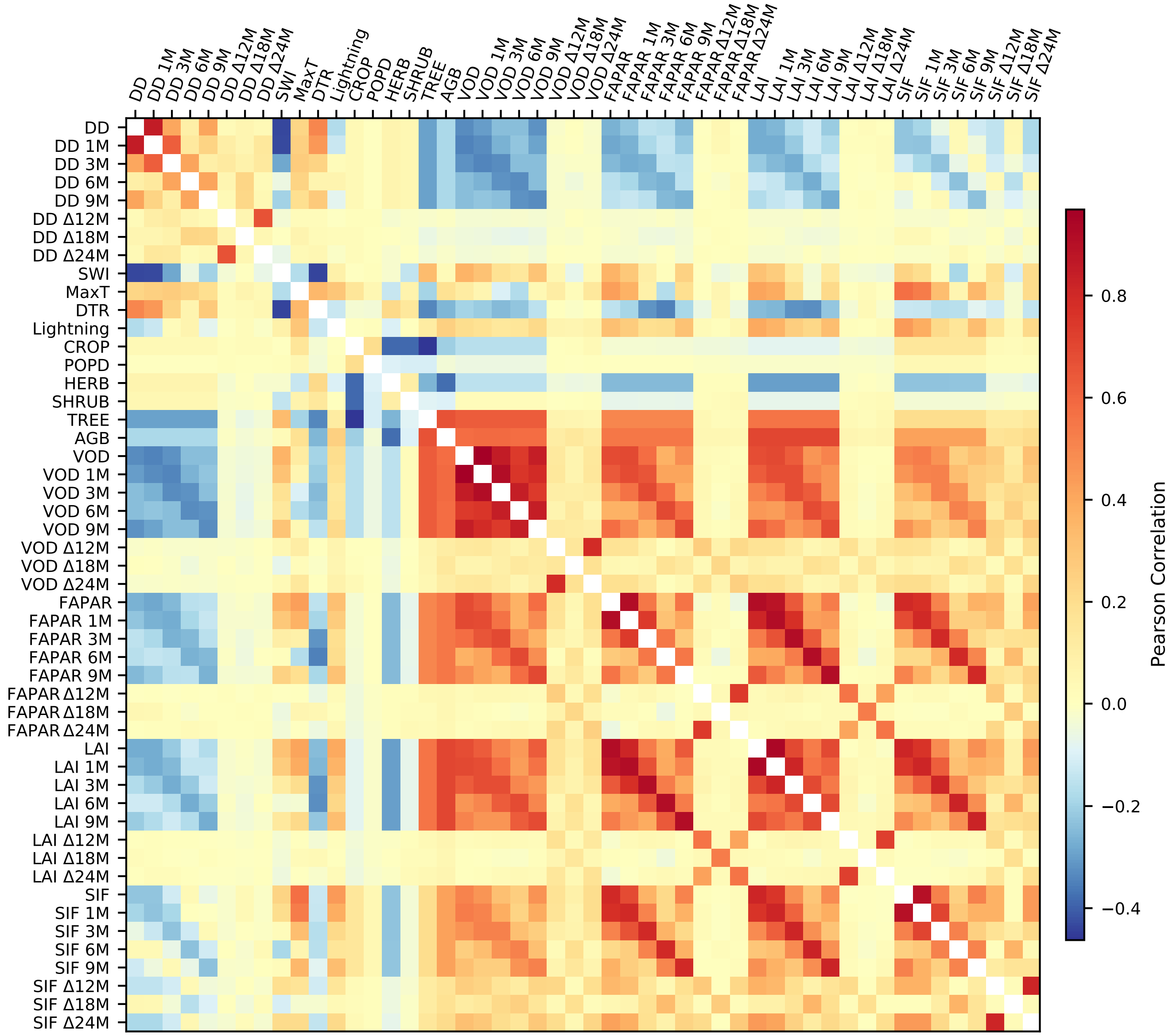

Figure S9. Pearson correlations between all variables used in the analysis for the time period from January 2010 to April 2015. Especially large positive correlations exist between variable pairs separated by multiples of 12 months and between FAPAR, LAI, SIF, and VOD. The largest negative correlations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

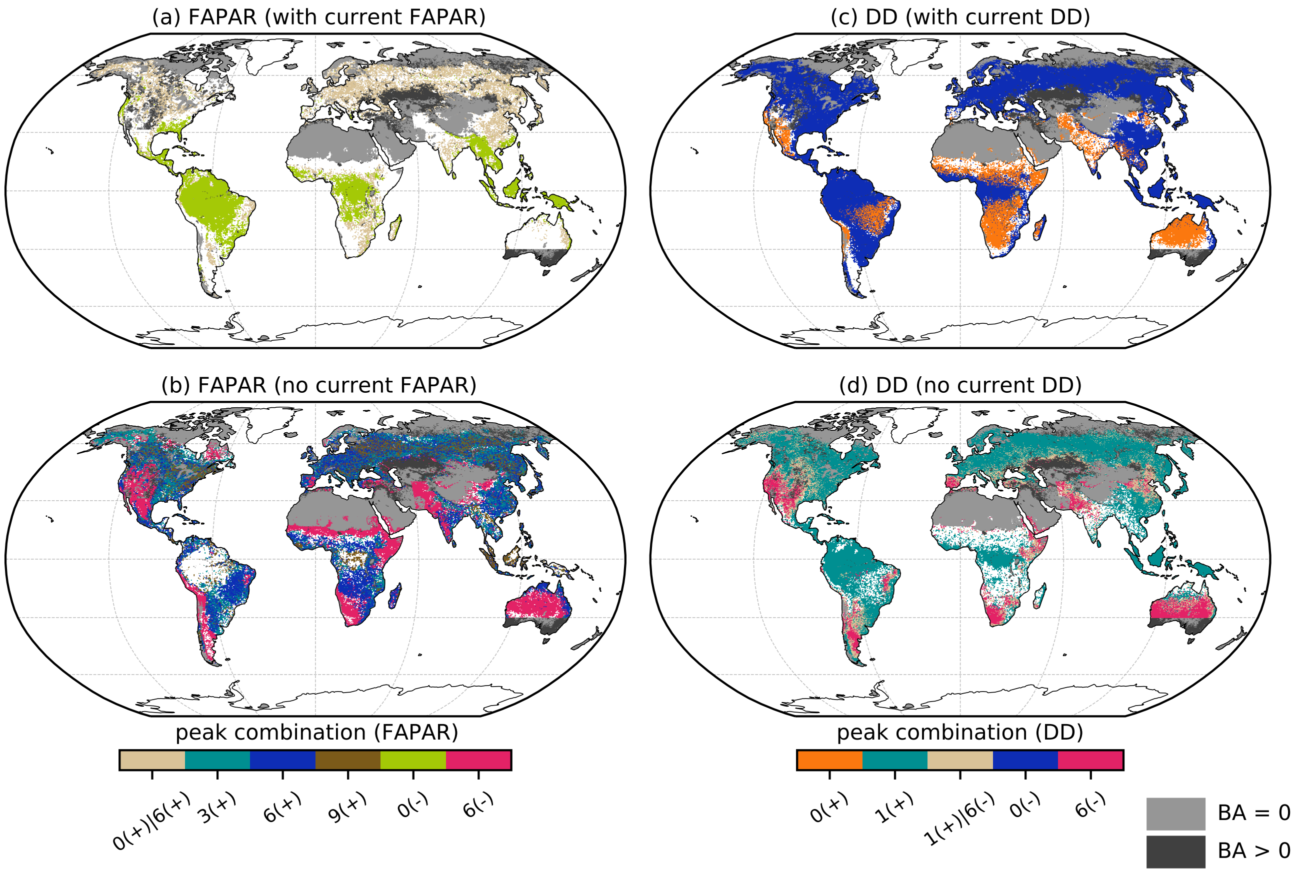

Figure S10. Spatial distribution of individual peak combinations for SHAP values as in Fig. 7. The sign of the maximum effect on BA at a certain antecedent month is indicated in parentheses after each month. The peak combinations are shown here such that their ordering has no significance (i.e. $0(+)|3(+)$ 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.

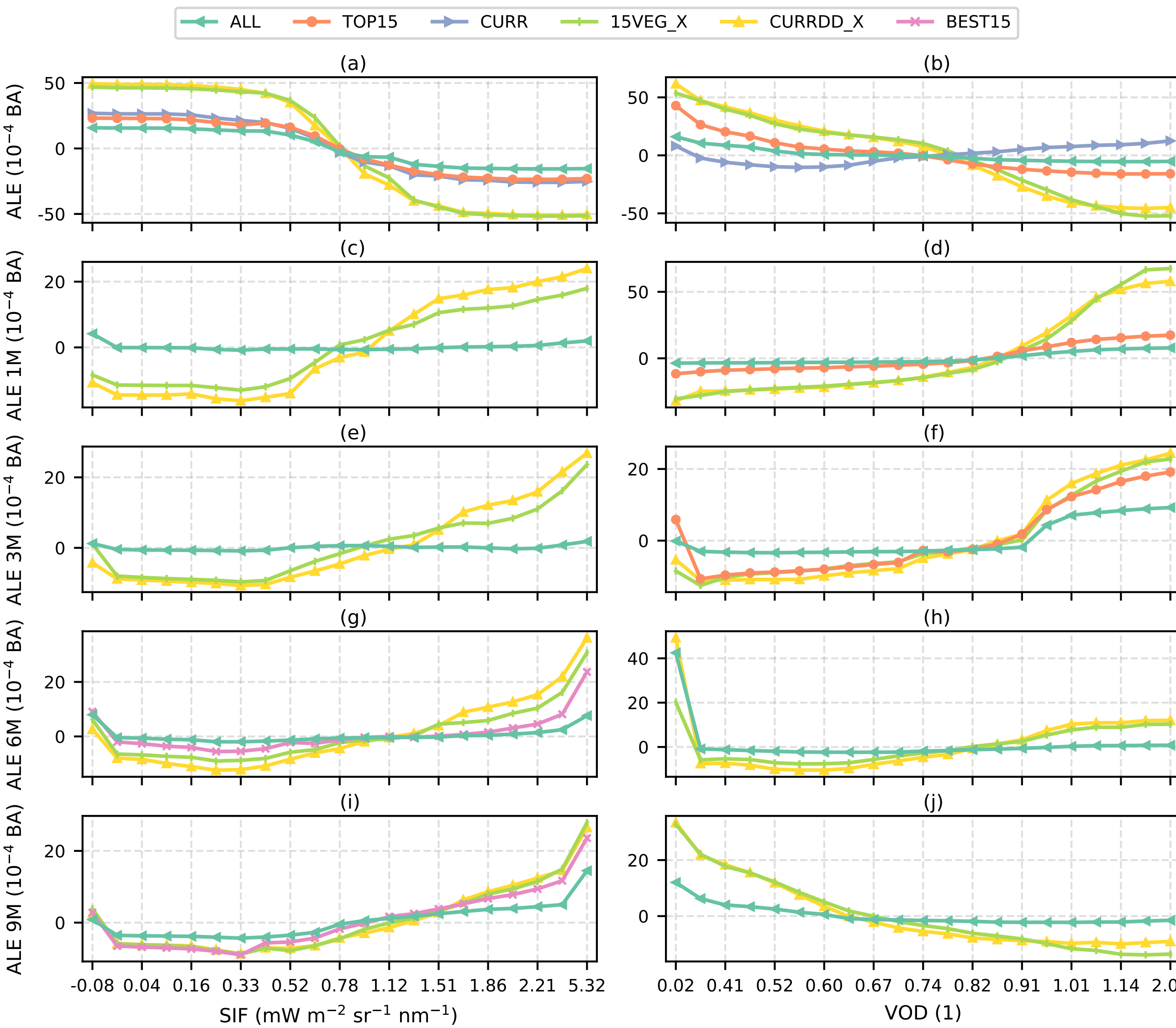

Figure S11. First-order LAI ALEs for different lags (< 1 yr) from all relevant modelling experiments for the relationships between BA and SIF (left hand columns) and VOD (right hand columns). Evenly spaced quantiles were used in the construction of the plots.

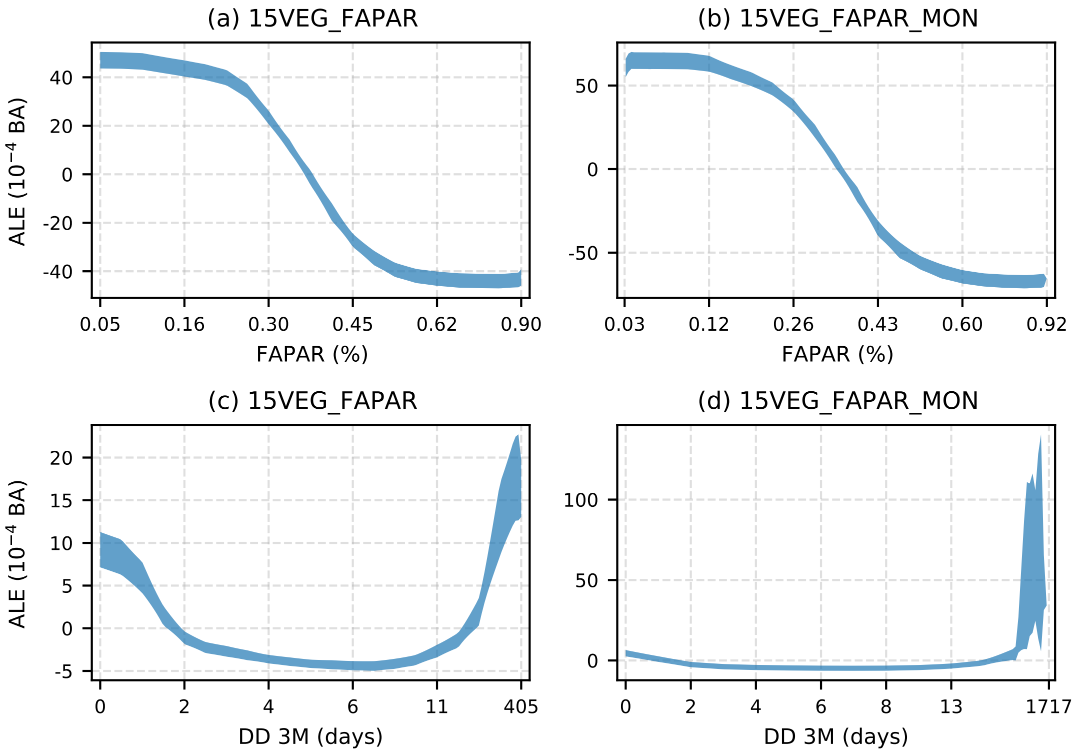

Figure S12. First-order ALE plots showing the effect of FAPAR (a, b) and the 3-month antecedent dry-day period (DD 3M; c, d) on burnt area (BA) in the 15VEG_FAPAR model (a, c) and the 15VEG_FAPAR_MON model (b, d) after accounting for all other variables. The shaded regions represent the standard deviation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

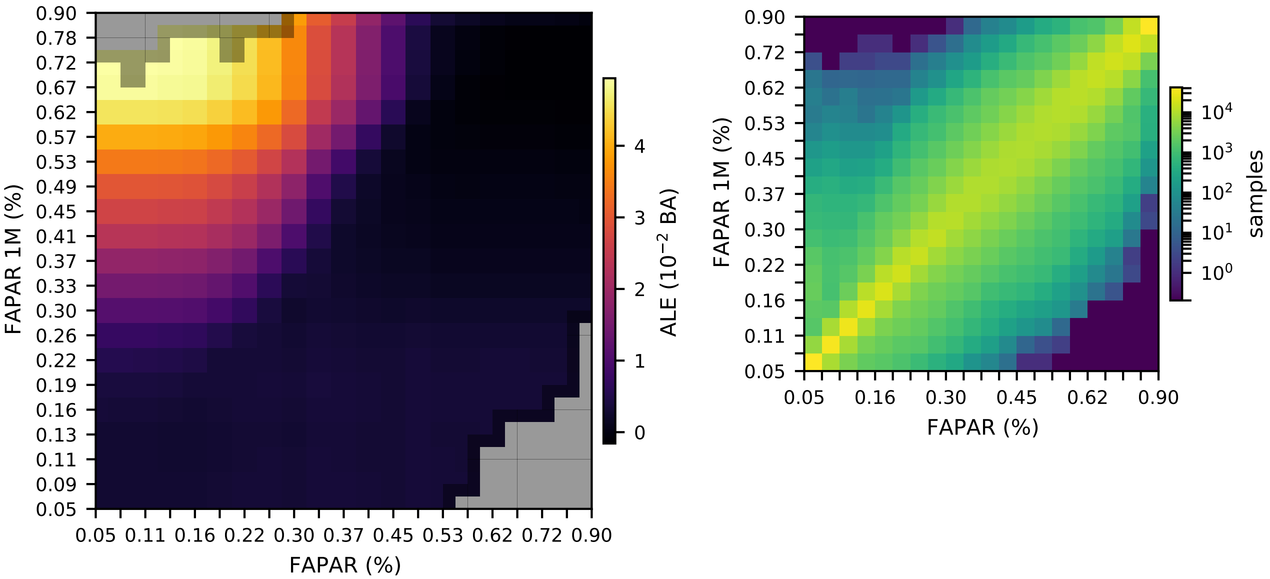

Figure S13. Second-order ALE plot showing the combined zeroth order (mean), first order, and second order modelled effects of FAPAR and FAPAR 1M on BA from the 15VEG_FAPAR model, taking into account all other variables. Grey boxes indicate missing data. The diagonal structure of the sample count matrix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

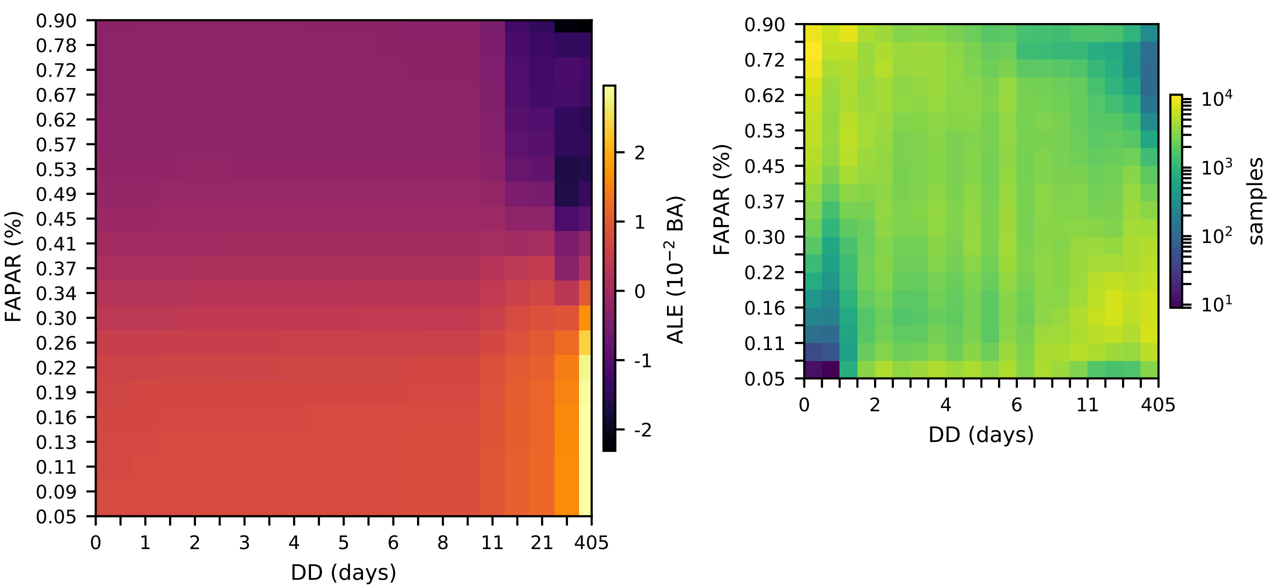

Figure S14. Second-order ALE plot showing the combined zeroth order (mean), first order, and second order modelled effects of DD and FAPAR on BA from the 15VEG_FAPAR model, taking into account all other variables. The diagonal structure of the sample count matrix demonstrates the anticorrelation between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