**Supporting Information**

**SI 1: Methods**

**Table S 1: Aggregation table to convert the PFT classes of each FireMIP model into a set of common PFTs. Note that JSBACH does
5  not include a PFT for needle-leaved deciduous trees. Therefore, the ExtD (extratropical deciduous tree) PFT was spilt into Tree.ND
for regions in north-east Siberia and to Tree.BD for all other regions.**

| Model | Common PFT classes used in analysis | | | | | |
|---|---|---|---|---|---|---|
| | **Tree.NE** | **Tree.ND** | **Tree.BE** | **Tree.BD** | **Herb** | **Crop** |
| | Needle-leaved evergreen trees | Needle-leaved deciduous trees | Broadleaved evergreen trees | Broadleaved deciduous trees | Herbaceous vegetation | Croplands |
| CLM | TeNE + BNE | BNS | TrBE + TeBE + BE_Shb | TrBR + TeBS + BBS + BBS_Shb | C3G_arc + C3G + C4G | Crop1 + Crop2 |
| CTEM | NDL-EVG | NDL-DCD | BDL-EVG | BDL-DCD-COLD + BDL-DCD-DRY | C3-GRASS + C4-GRASS | C3-CROP + C4-CROP |
| JSBACH-SPITFIRE | ExtE | ExtD (if lon > 95°E & lat > 48°N & ExtD > 0.2) | TrE | TrD + Rg_Shb + De_Shb + ExtD (if not classified as Tree.ND) | C3G + C4G + C3G_pas + C4G_pas | Crop |
| JULES-INFERNO | NE + Ev_Shb | ND | TrBE + TeBE | BD + De_Shb | C3G + C4G | -- |
| LPJ-GUESS-SIMFIRE | BNE + BINE | BNS | TeBE + TrBE + TrIBE | TeBS + IBS + TrBR | C3G + C4G + C3G_pas + C4G_pas | TeSW + TeSWirr + TeWW + TeWWirr + TeCo + TeCoirr |
| LPJ-GUESS-SPITFIRE | BNE + BINE + TeNE | BNS | TeBE + TrBE + TrIBE | BIBS + TeBS + TeIBS | C3G + C4G | -- |
| ORCHIDEE-SPITFIRE | TeNE + BNE | BNS | TrBE + TeBE | TrBR + TeBS + BBS | C3G + C4G | C3_agr + C4_agr |

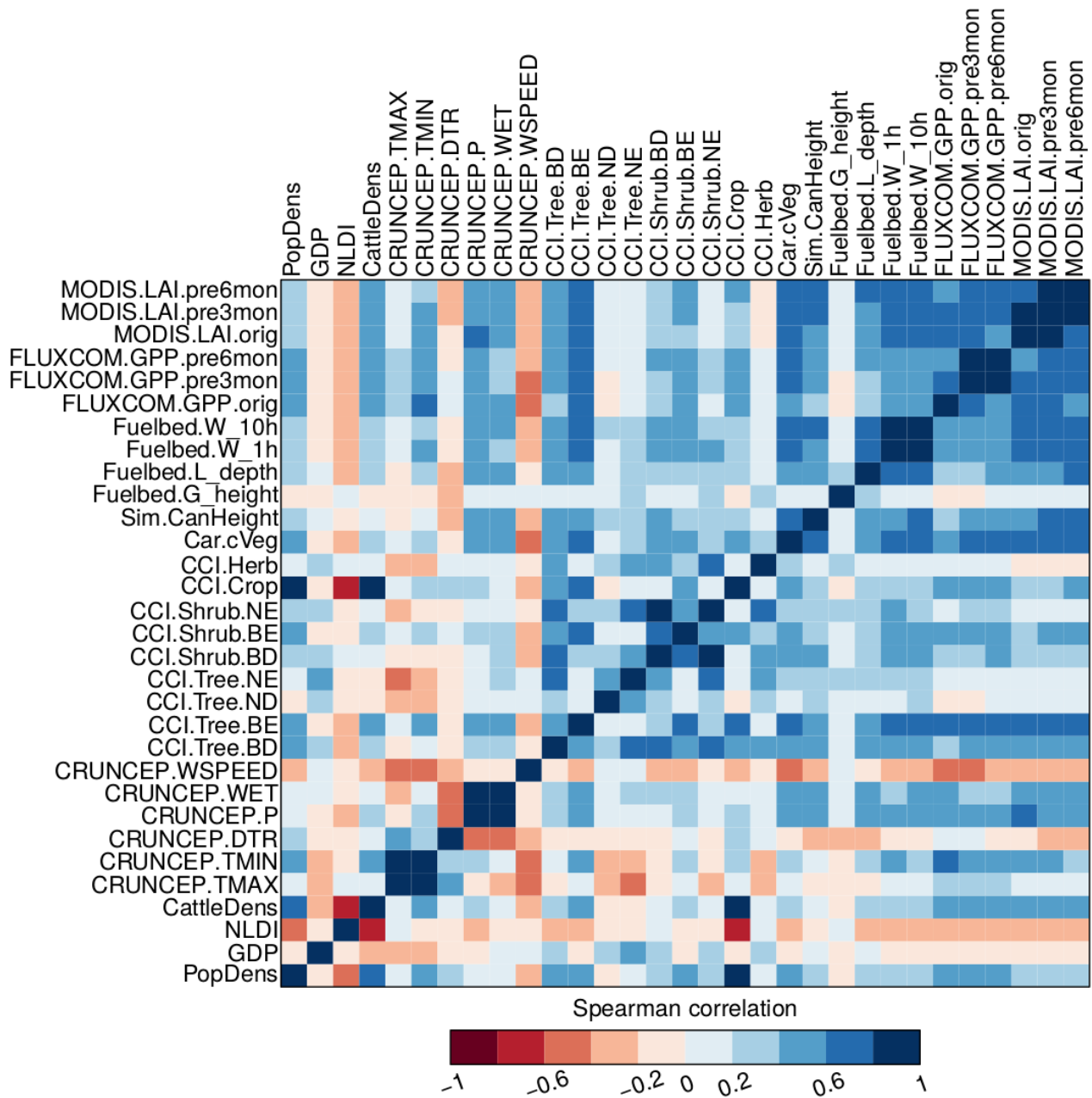



**Figure S 1: Pair-wise correlations of observation-based predictor variables.** Correlations are based on the global dataset that includes monthly observations on 2.5° x 1.89° for the period 2005-2011. Values from annual datasets were repeated to match monthly observations. Some predictor variables were not used in random forest models because of high correlations (r >= 0.8) with other variables, i.e. night-light development index (with CCI.Crop), cattle density (with CCI.Crop), woody litter for the 10 h fuel size class (with Fuelbed.W_1h), precedent 3-monthly GPP (with FLUXCOM.GPP.pre6month), and precedent 3-monthly LAI (with MODIS.LAI.orig and MODIS.LAI.pre6month).

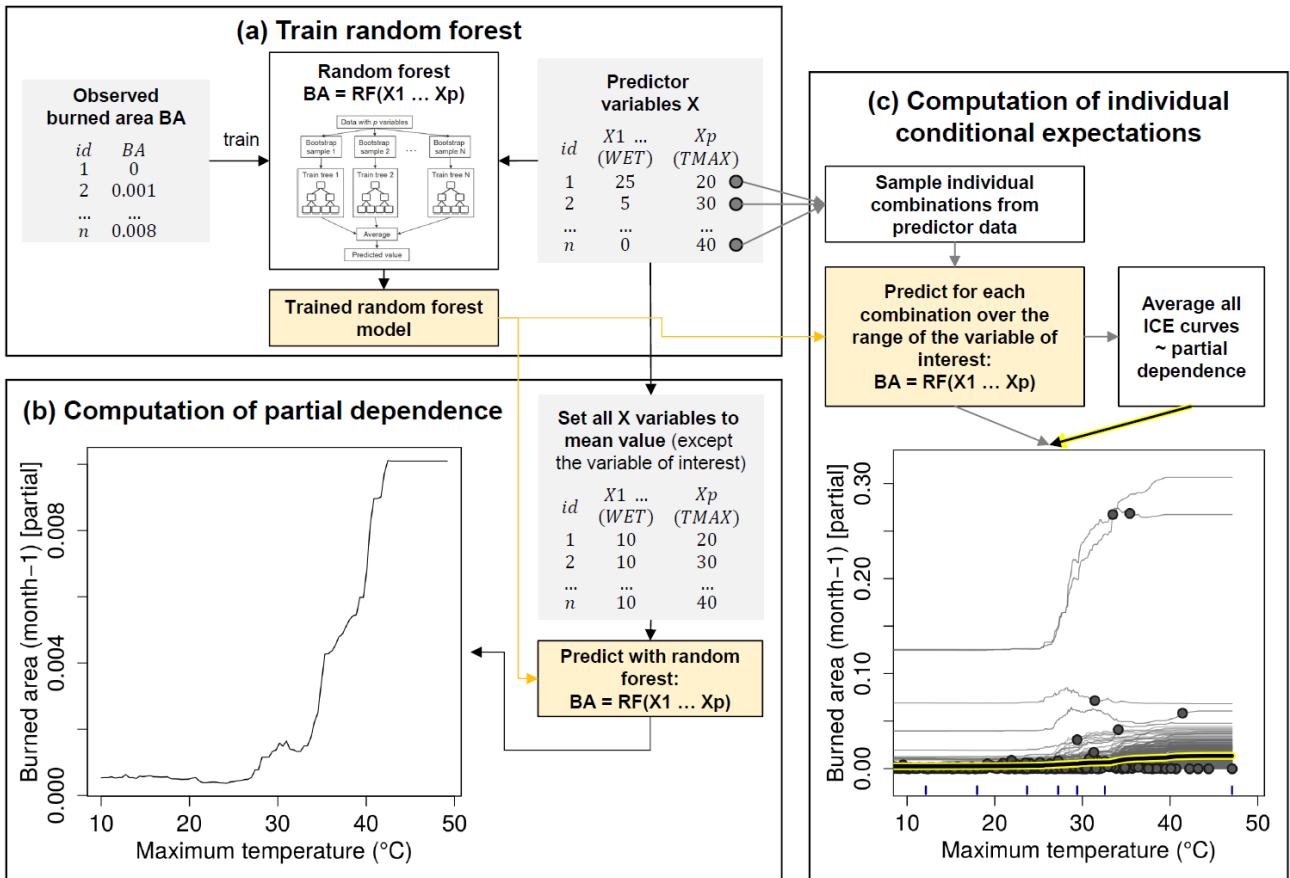



**Figure S 2: Deriving partial dependencies and individual conditional expectation curves from trained random forest models. (a) A random forest is trained against the target variable (observed burned area) based on a set of predictor variables X. (b) The partial dependency (e.g. to maximum temperature, TMAX) is derived from the trained random forest experiment by setting all other predictor variables (e.g. number of wet days, WET) to its mean values. Random forest predictions are then done over the range of the variable of interest (TMAX). (c) For the computation of individual conditional expectation curves, individual combinations of predictor variables are sampled from the predictor data (grey dots in a and c). Predictions are then done for each case over the range of the variable of interest (grey lines in c). The average over all ICE curves approximates then the partial dependence (yellow highlighted line in c). The average ICE curve agrees with the partial dependence when no sampling was performed (i.e. ICE curves were computed for all cases in the input data). The figures in (b) and (c) were computed from the random forest experiment RF.CCI_MERIS.fm by using the global predictor dataset and hence show the global sensitivity of burned area to TMAX. ICE curves in (c) show that in most cases burned area increases with increasing TMAX. This behaviour is then also reflected in the average ICE curve and the partial dependency, respectively. However, some ICE curves show a stable or non-monotonic response of burned area to TMAX which indicates that certain cases burned area does not increase with TMAX.**

# SI 2: Evaluation of temporal burned area dynamics

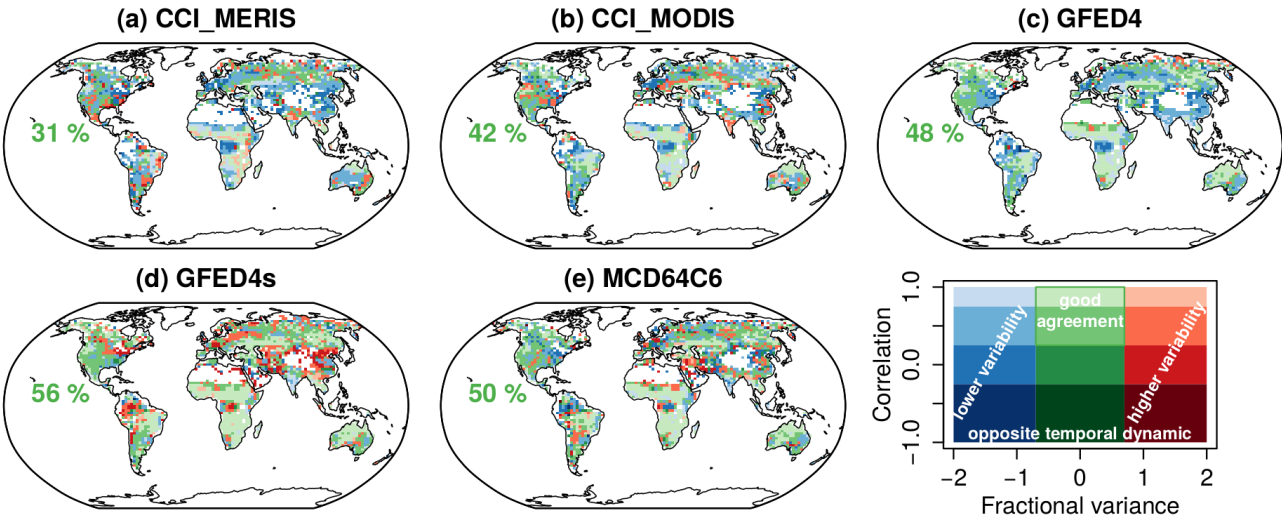



**Figure S 3: Comparison of burned area datasets with each other.** Shown is the Spearman rank-correlation coefficient and fractional variance of the monthly burned area in 2005-2011 from a single satellite dataset in comparison to the four other datasets. See Figure 2 for a detailed description. Individual datasets show a weaker agreement than the agreement of all datasets (Figure 2) because low FV or correlations of a single dataset at grid cell-level are averaged-out (median), resulting in larger areas with "good" agreement.

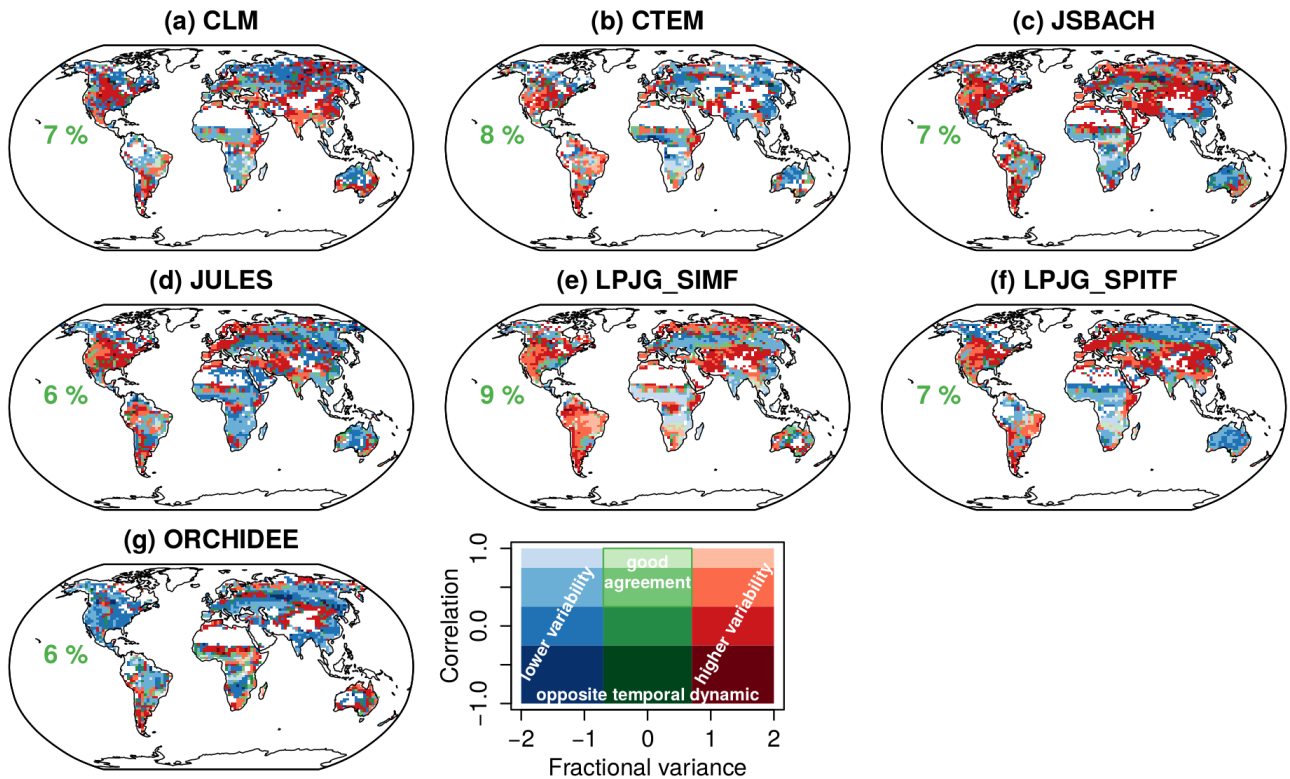



**Figure S 4: Evaluation of simulated burned area from FireMIP models against satellite datasets. Shown is the Spearman rank-correlation coefficient and fractional variance of the monthly burned area in 2005-2011 from one FireMIP model in comparison to the five satellite datasets. See Figure 2 for a detailed description.**

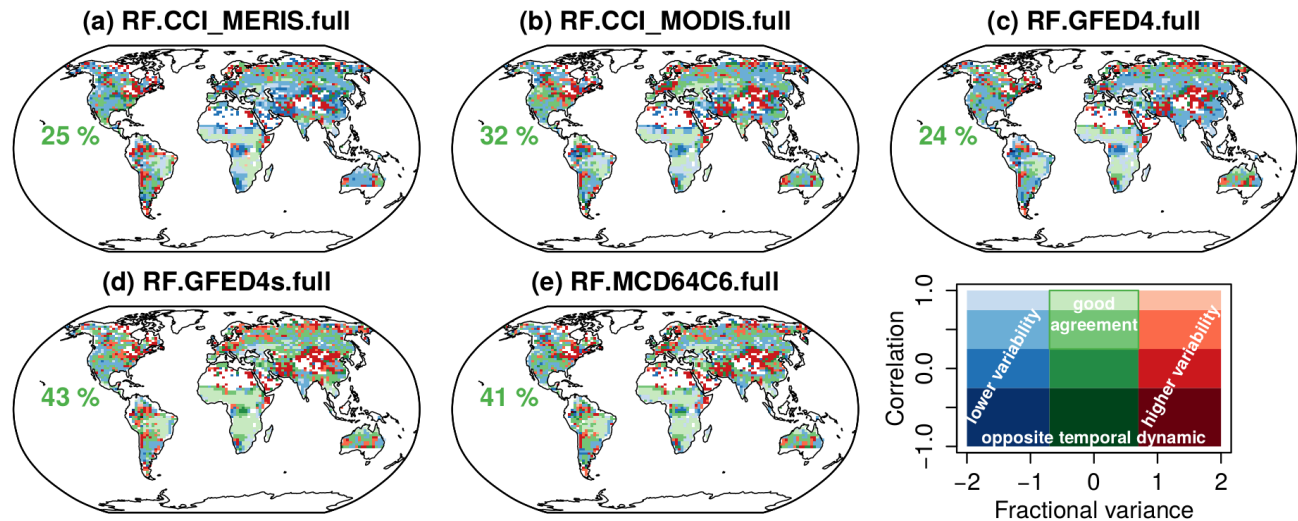
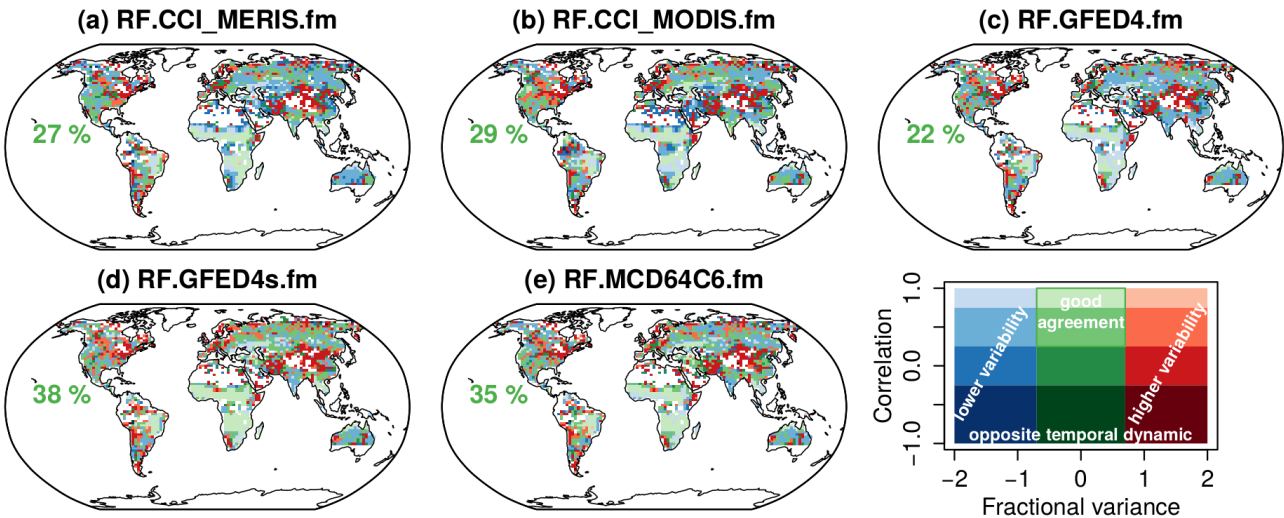



**Figure S 5: Evaluation of predicted burned area from the "full" random forest experiments with the "full" set of predictor variables against satellite datasets. Shown is the Spearman rank-correlation coefficient and fractional variance of the monthly burned area in 2005-2011 from one random forest experiment in comparison to the five satellite datasets. Each random forest experiment was**

trained against a single burned area dataset based on the "full" set of predictor variables and the predicted burned area was then evaluated against the five burned area datasets. See Figure 2 for a detailed description.

5    Figure S 6: Evaluation of predicted burned area from the "fm" random forest experiments against satellite datasets. Shown is the Spearman rank-correlation coefficient and fractional variance of the monthly burned area in 2005-2011 from one random forest experiment in comparison to the five satellite datasets. Each random forest experiment was trained against a single burned area dataset based on "fm" predictor variables that are also available for FireMIP models and the predicted burned area was then evaluated against the five burned area datasets. See Figure 2 for a detailed description.

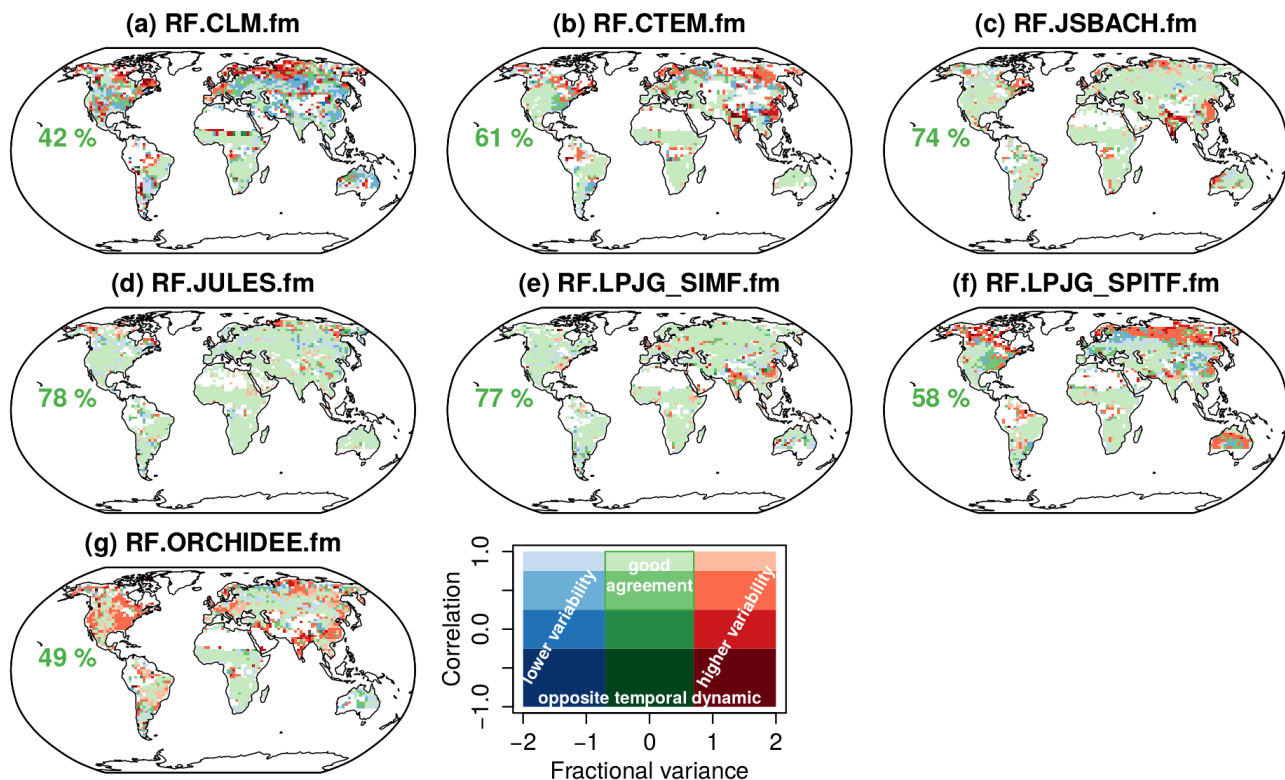



**Figure S 7: Performance of random forest in reproducing the simulated burned area of each FireMIP model. Shown is the Spearman rank-correlation coefficient and fractional variance of the monthly burned area in 2005-2011 from one random forest experiment in comparison to the simulated burned area of the FireMIP model that was used to train the random forest. Each random forest experiment was trained based on "fm" predictor variables and the predicted burned area was then evaluated against the original burned area from this FireMIP model. See Figure 2 for a detailed description.**

# SI 3: Importance of predictor variables

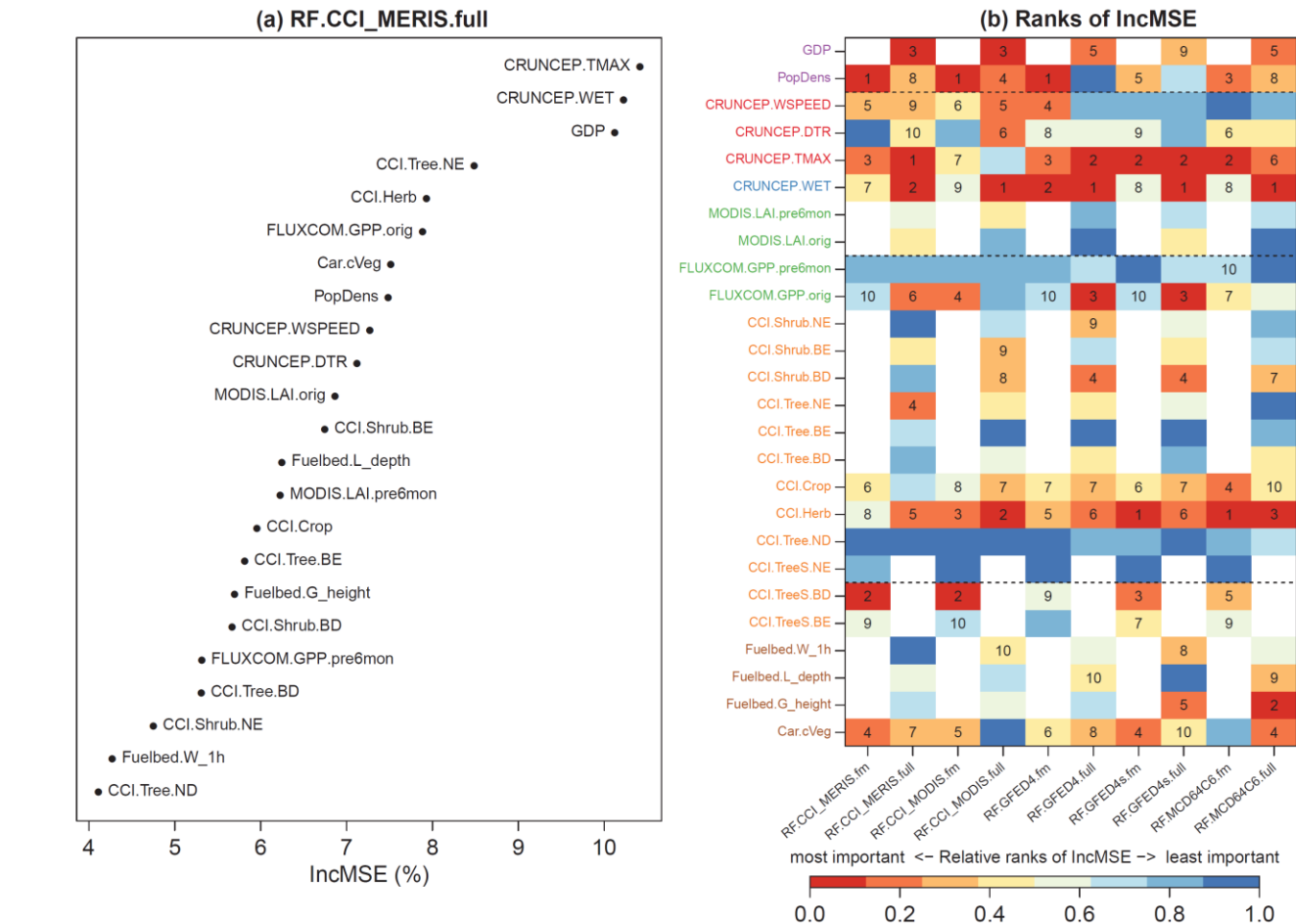



**Figure S 8: Importance of predictor variables in satellite-derived RF experiments. Panel (a) shows the percentage increment in the out-of-bag prediction mean squared error (MSE) if a variable is permuted for the RF.CCI_MERIS.full experiment. The most important variable (CRUNCEP.TMAX) has the highest increment in MSE. Panel (b) compares the ranked increment in MSE for different RF experiments. For example, the rank of variables in panel (a) is shown in the second column in panel (b) and the most important variable is coloured in red and denoted with 1. White fields denote predictor variables that were not used in the respective RF experiment. The importance of variables depends on the burned area dataset that was used to train RF and differs between the "full" and "fm" set of predictor variables.**

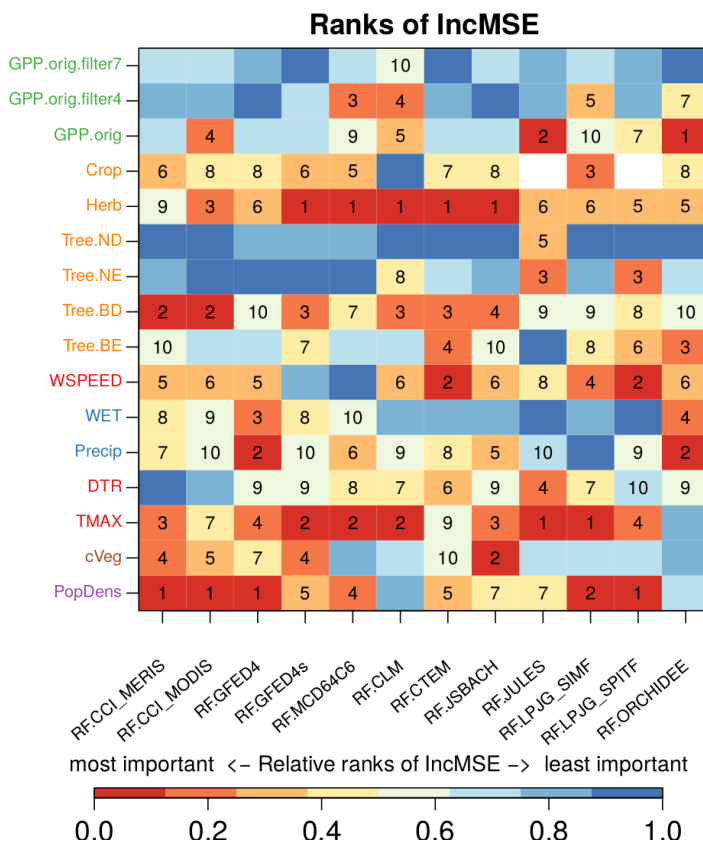



**Figure S 9: Importance of predictor variables in satellite- and FireMIP-based "fm" RF experiments. Further explanations of this figure are provided in Figure S 8.**

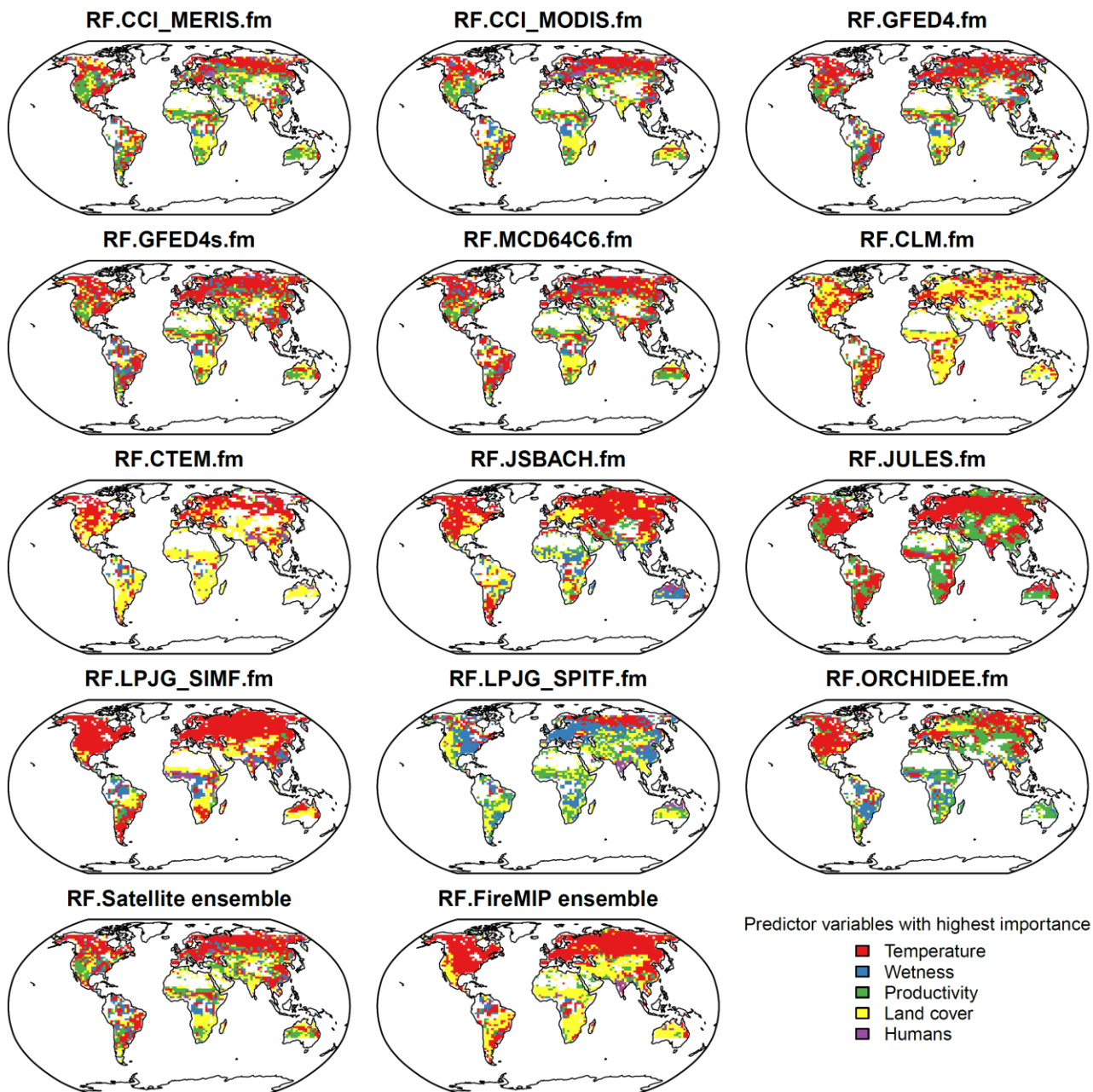



**Figure S 10: Groups of predictor variables with the highest importance for RF predictions. See Figure 3 for further details.**

## SI 4: Satellite-derived global relationships

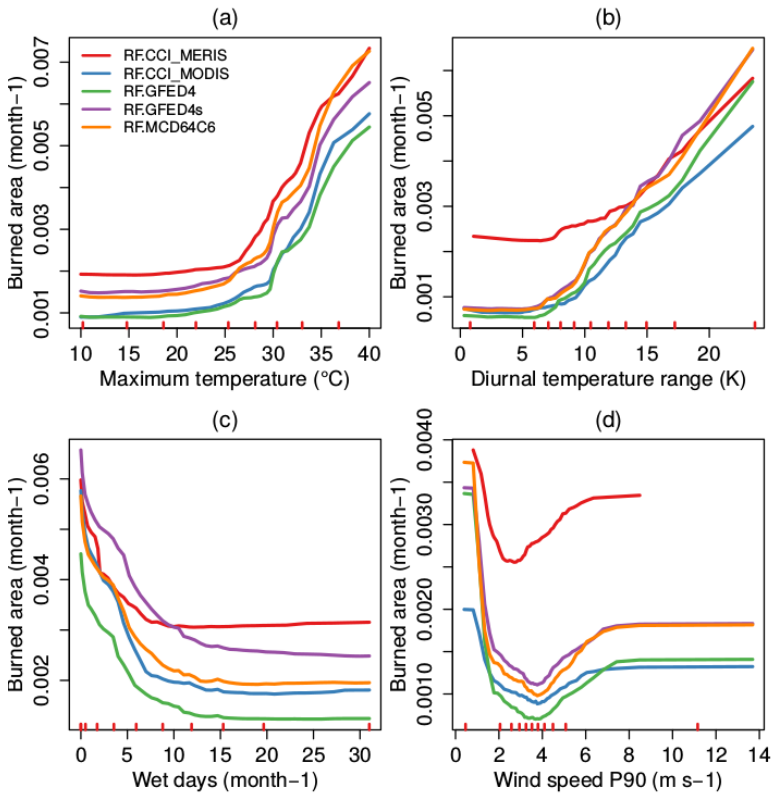



**Figure S 11: Global relationships of the fractional burned area per month to climate-related predictor variables as learned by the "full" random forest experiments.**

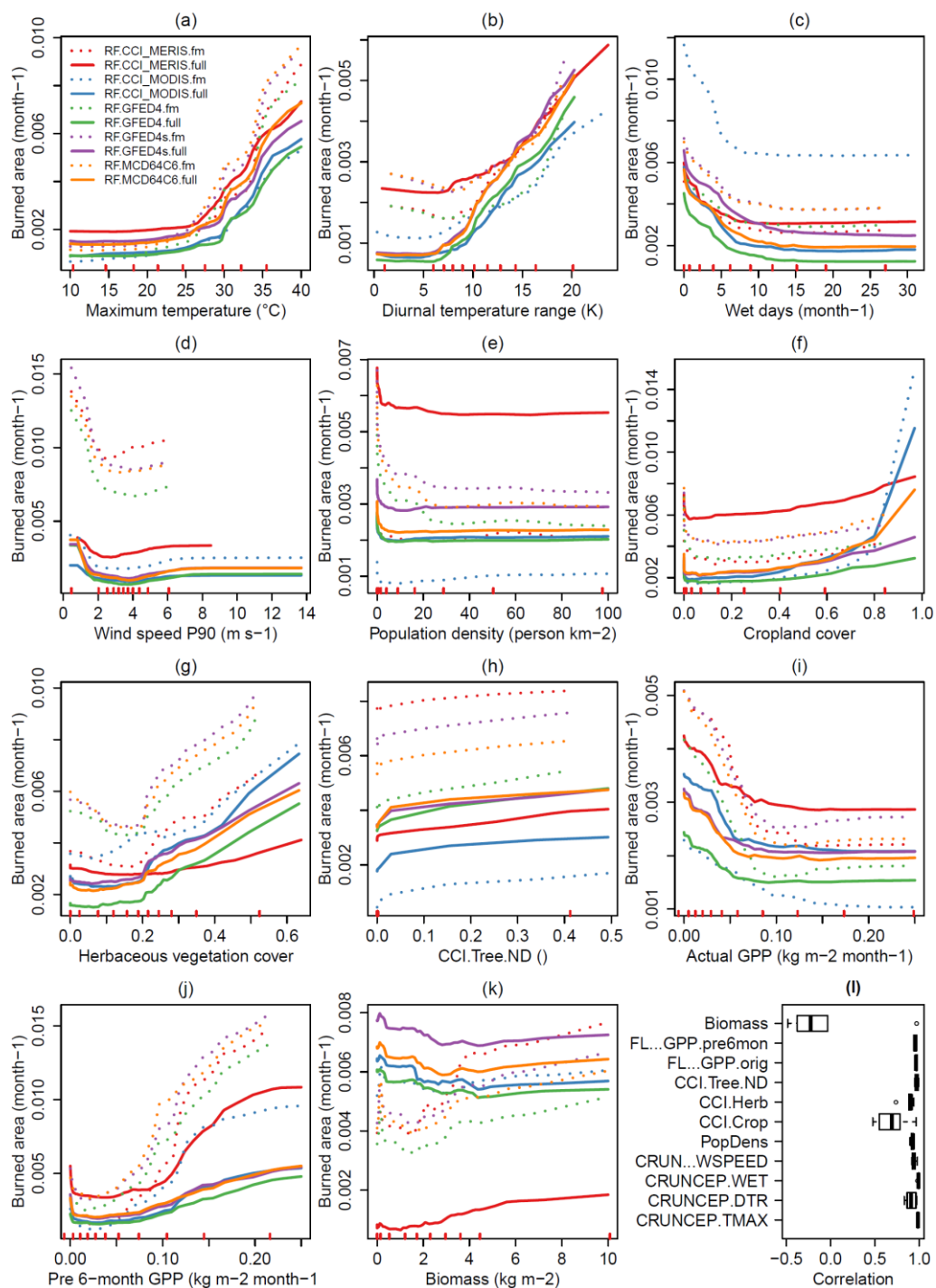



**Figure S 12: Comparison of global relationships of the fractional burned area per month to predictor variables as learned by the "full" and "fm" random forest experiments. Panel (l) shows the distribution of correlations between the global relationships from the "full" and "fm" random forest experiments for each predictor variable.**

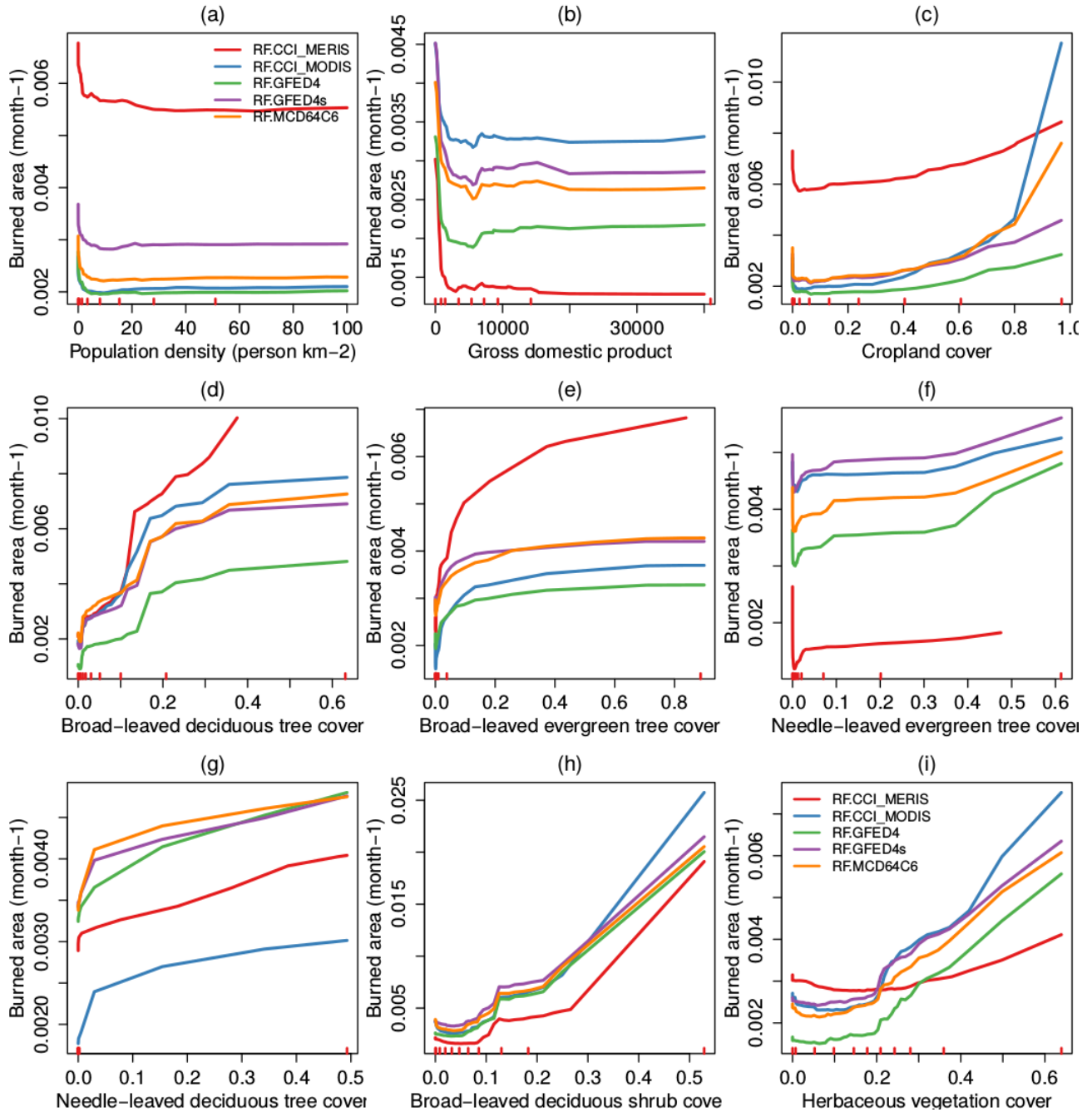



**Figure S 13: Global relationships of the fractional burned area per month to socioeconomic and land cover-related predictor variables as learned by the "full" random forest experiments.**

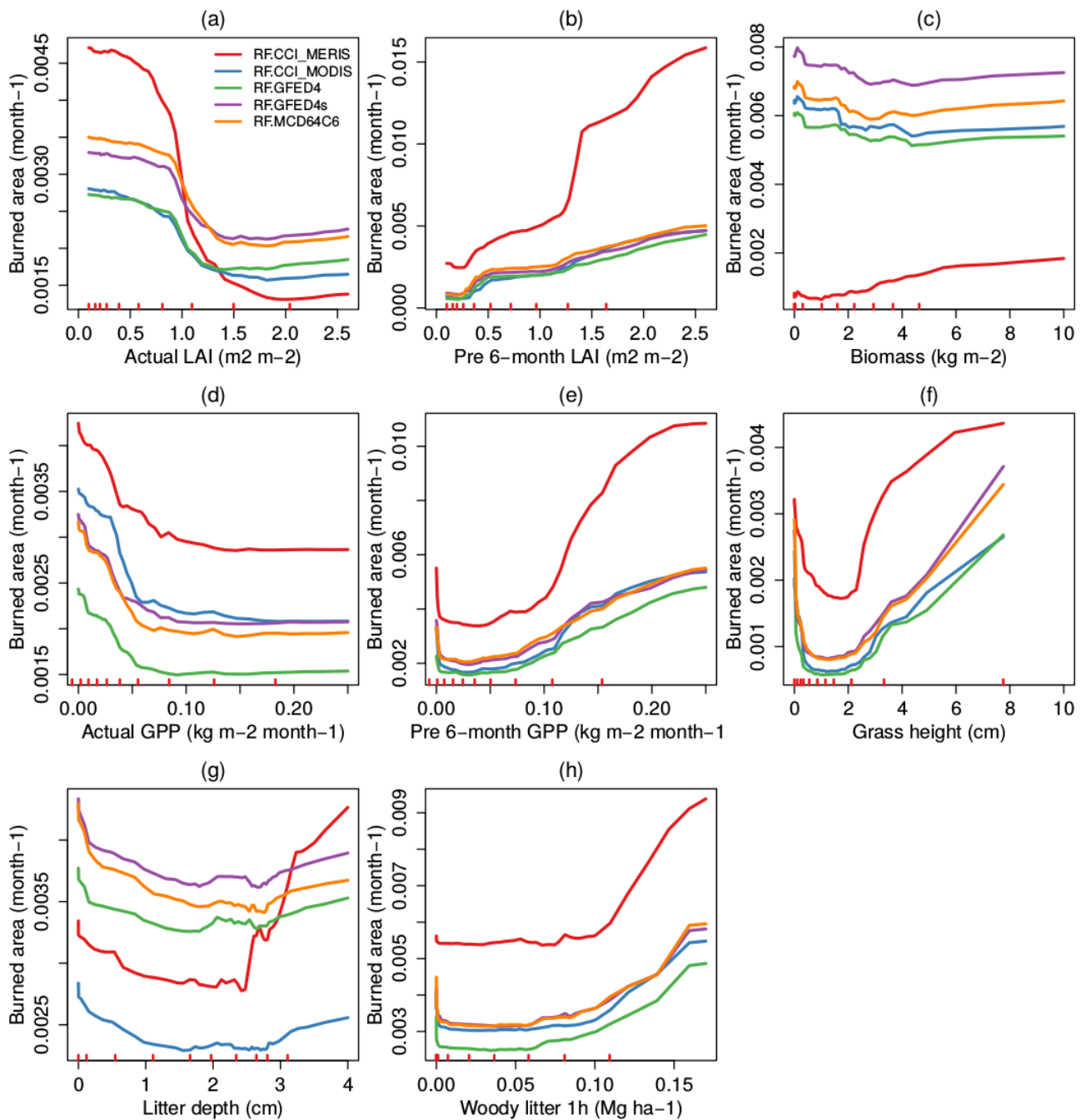



**Figure S 14: Global relationships of the fractional burned area per month to vegetation and fuel-related predictor variables as learned by the "full" random forest experiments.**

**SI 5: Comparison of satellite- and model-derived global relationships**

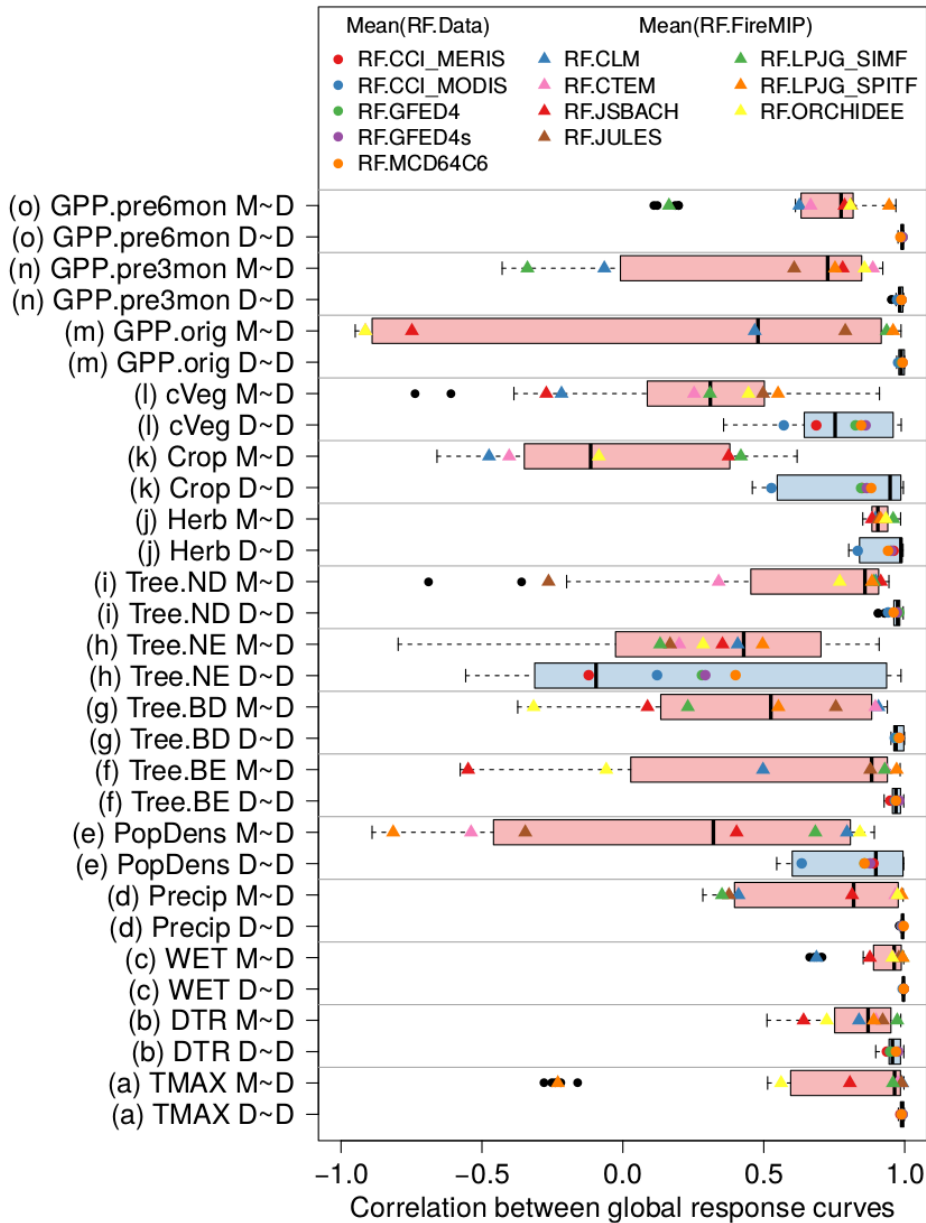



**Figure S 15: Correlations between global sensitivity functions from satellite-derived against other satellite-derived random forests (D~D) and from model-derived against satellite-derived random forests (M~D). Pearson correlations were computed from the partial dependencies as shown in Figure 2. Boxes show the distribution of all data-data (blue, 5 datasets) and model-data correlations (red, 5 datasets x 7 models), respectively. Coloured dots and triangles show the mean correlation for each satellite-derived and model-derived sensitivity function, respectively.**

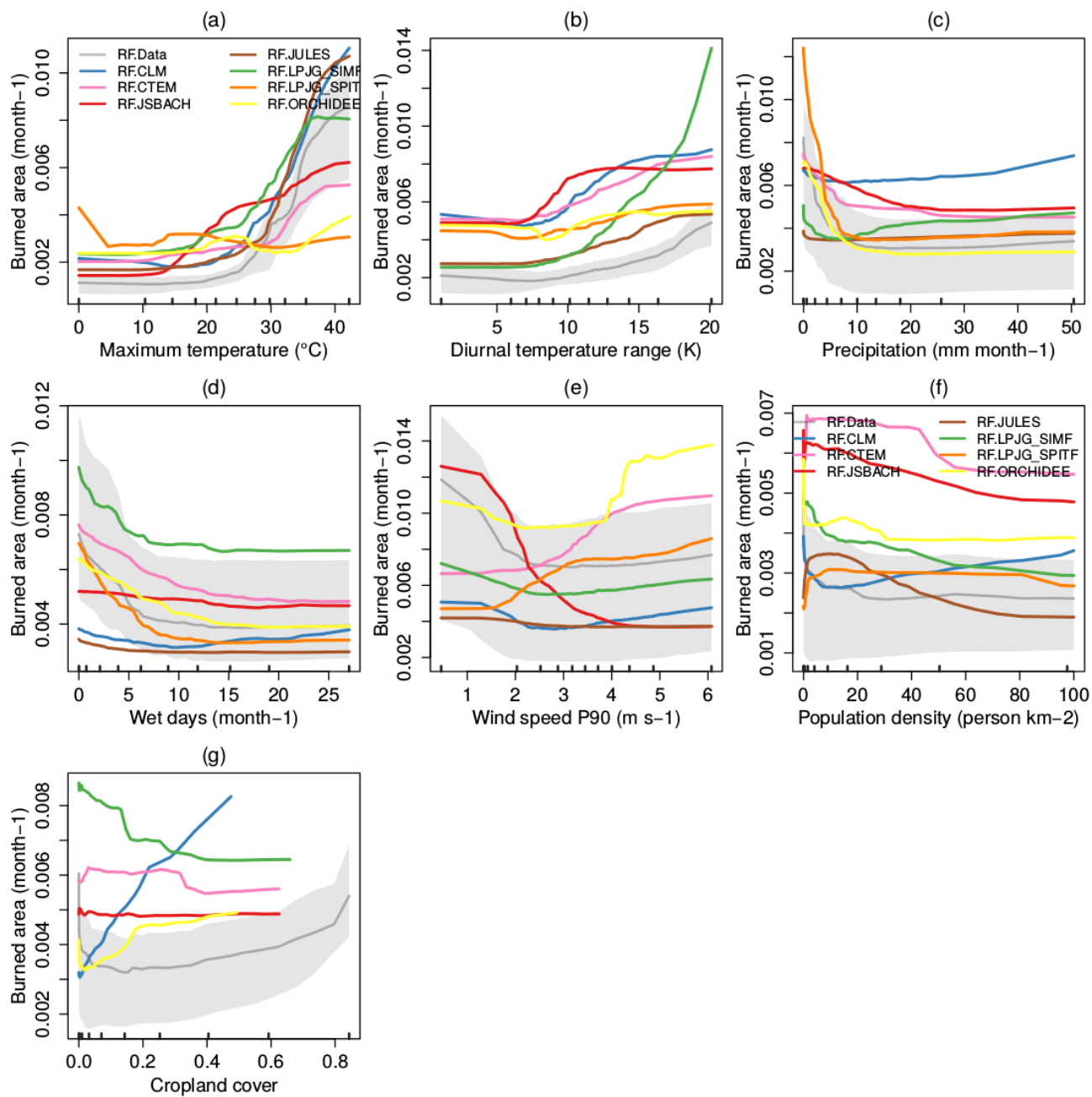



**Figure S 16: Global sensitivities of the fractional burned area per month to predictor variables as learned by the "fm" random forest experiments (Part 1).**

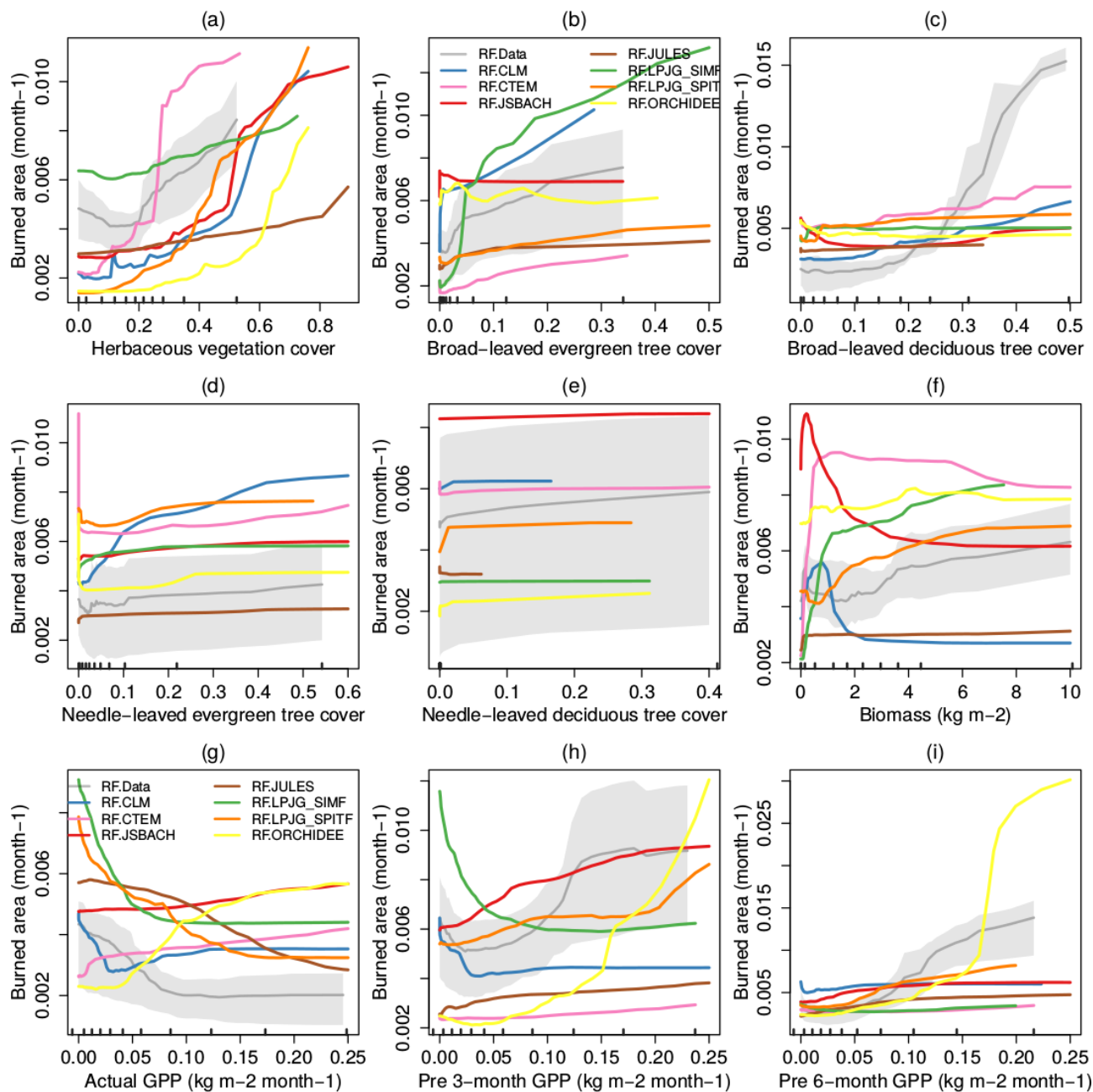



**Figure S 17: Global sensitivities of the fractional burned area per month to predictor variables as learned by the "fm" random forest experiments (Part 2).**

## SI 6: Regional sensitivities

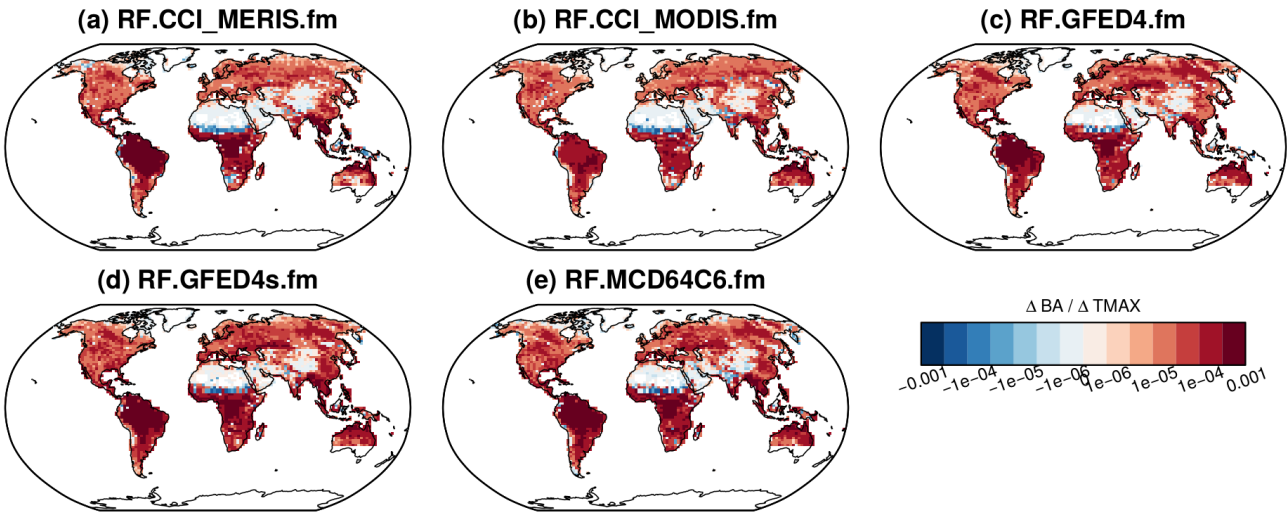



**Figure S 18: Regional sensitivities of the partial fractional burned area per month to monthly maximum temperature from satellite-derived "fm" RF models.**

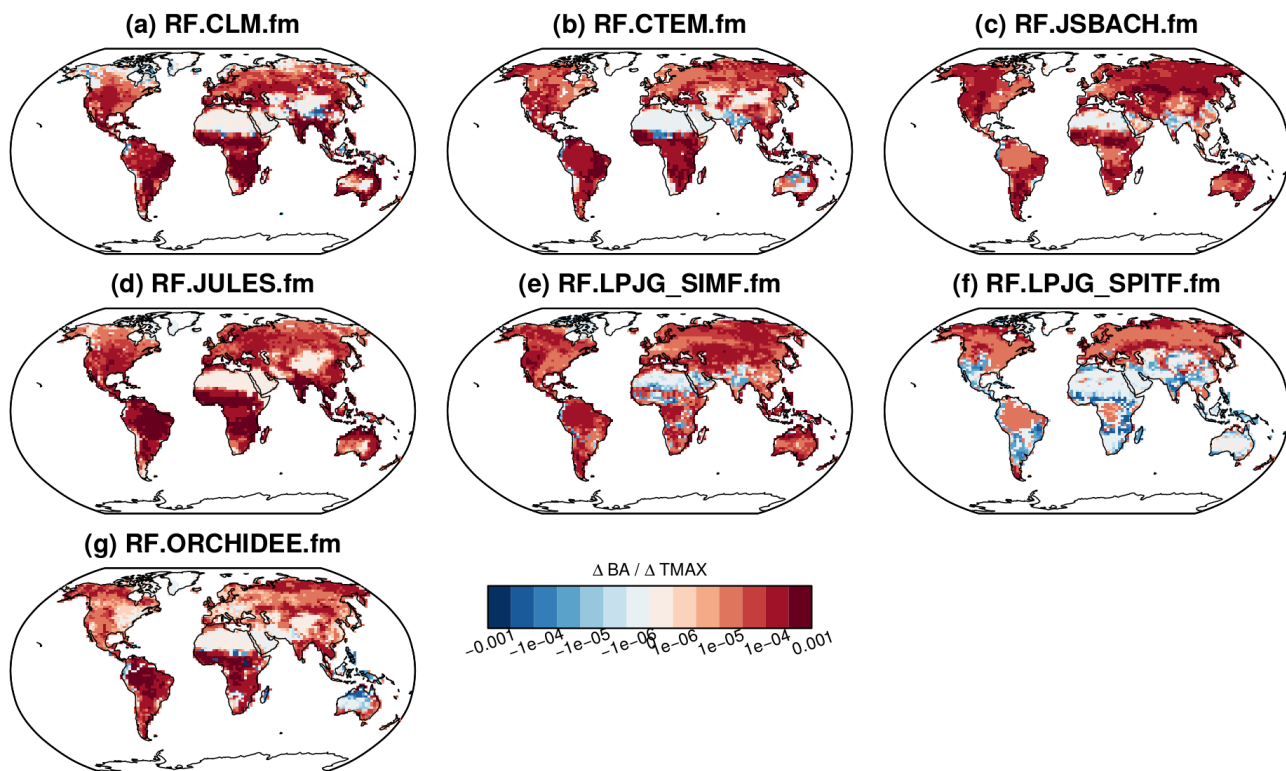



**Figure S 19: Regional sensitivities of the partial fractional burned area per month to monthly maximum temperature from model-derived "fm" RF models.**

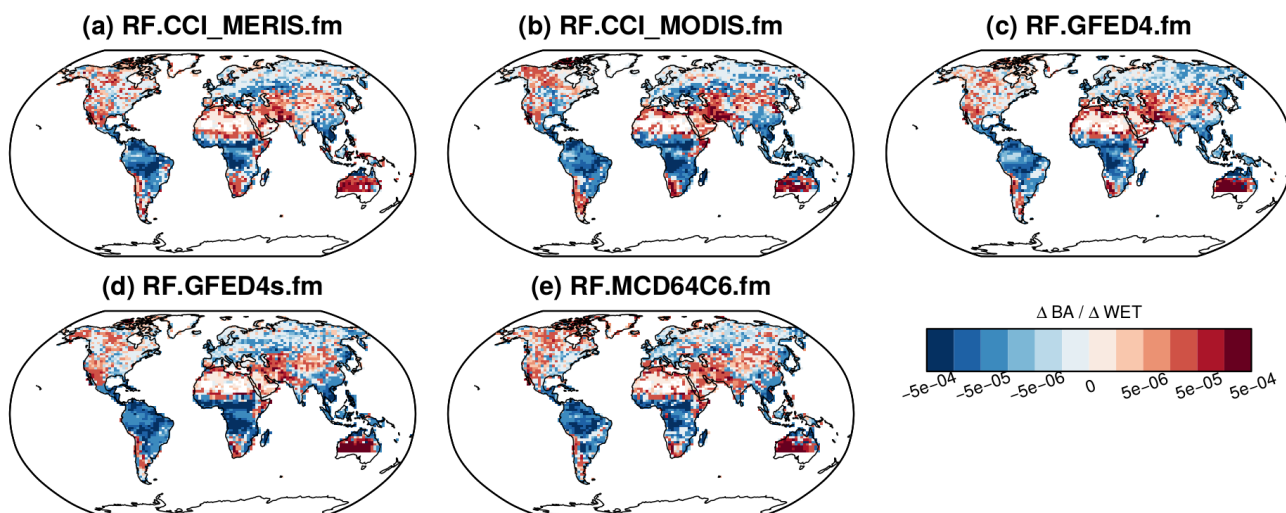



**Figure S 20: Regional sensitivities of the partial fractional burned area per month to the monthly number of wet days from satellite-derived "fm" RF models.**

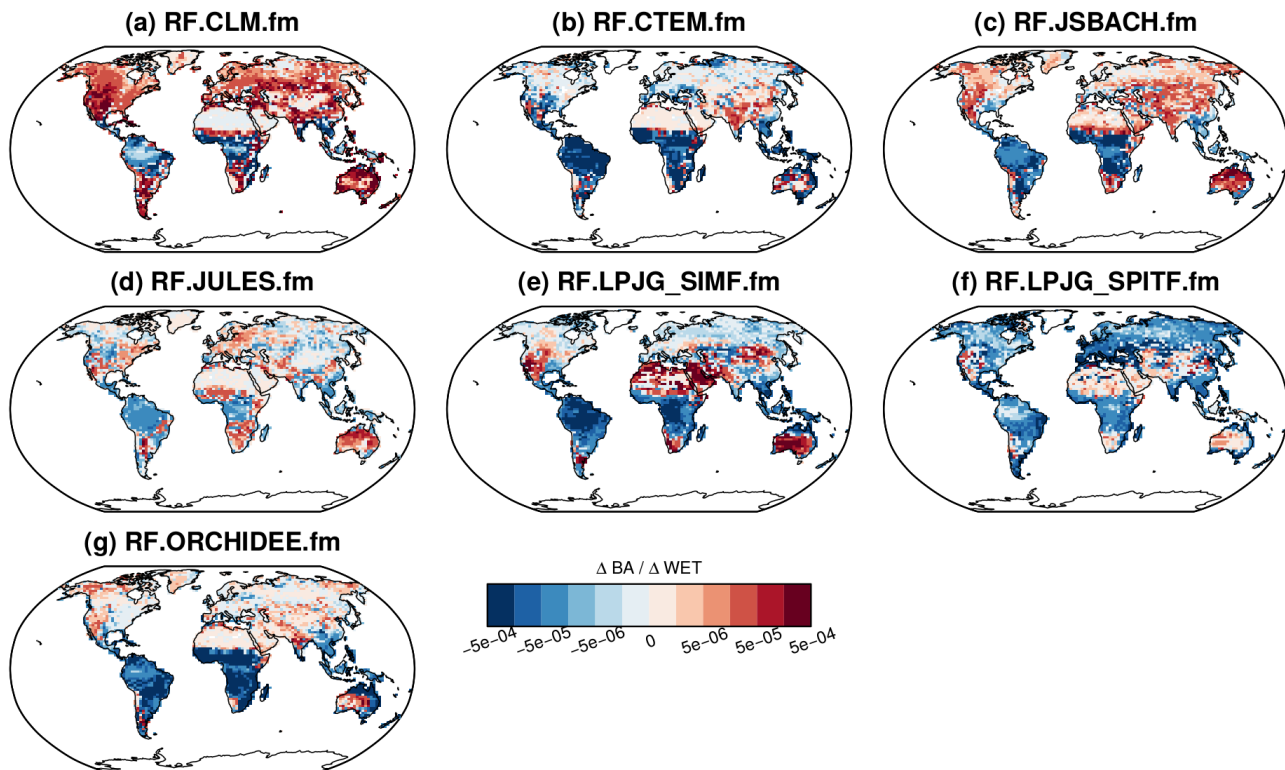



**Figure S 21: Regional sensitivities of the partial fractional burned area per month to the monthly number of wet days from model-derived "fm" RF models.**

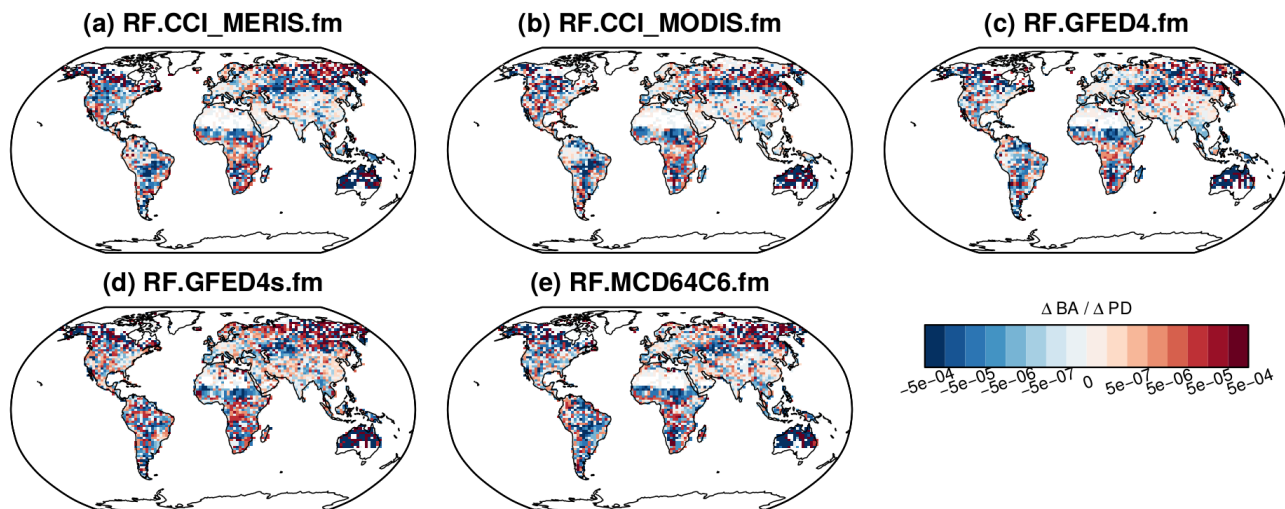



**Figure S 22: Regional sensitivities of the partial fractional burned area per month to the population density from satellite-derived "fm" RF models.**

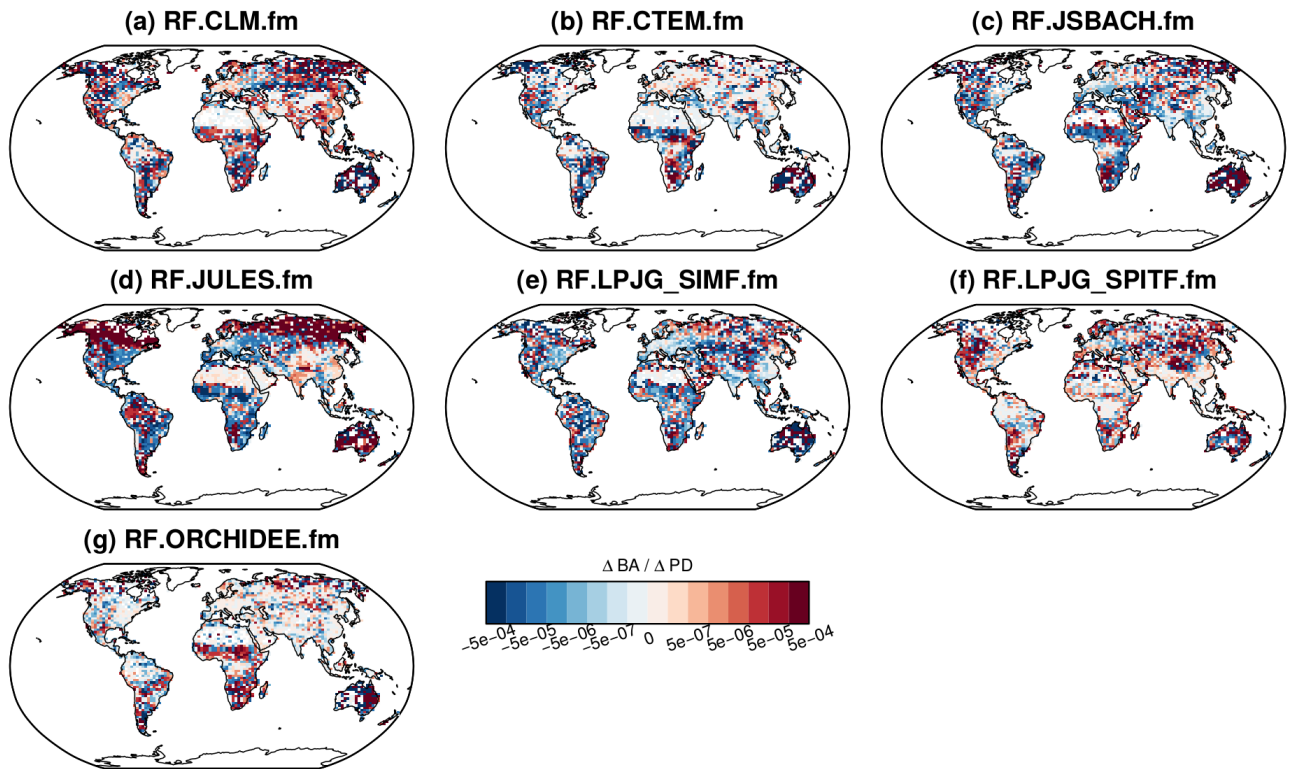



**Figure S 23: Regional sensitivities of the partial fractional burned area per month to the population density from model-derived "fm" RF models.**

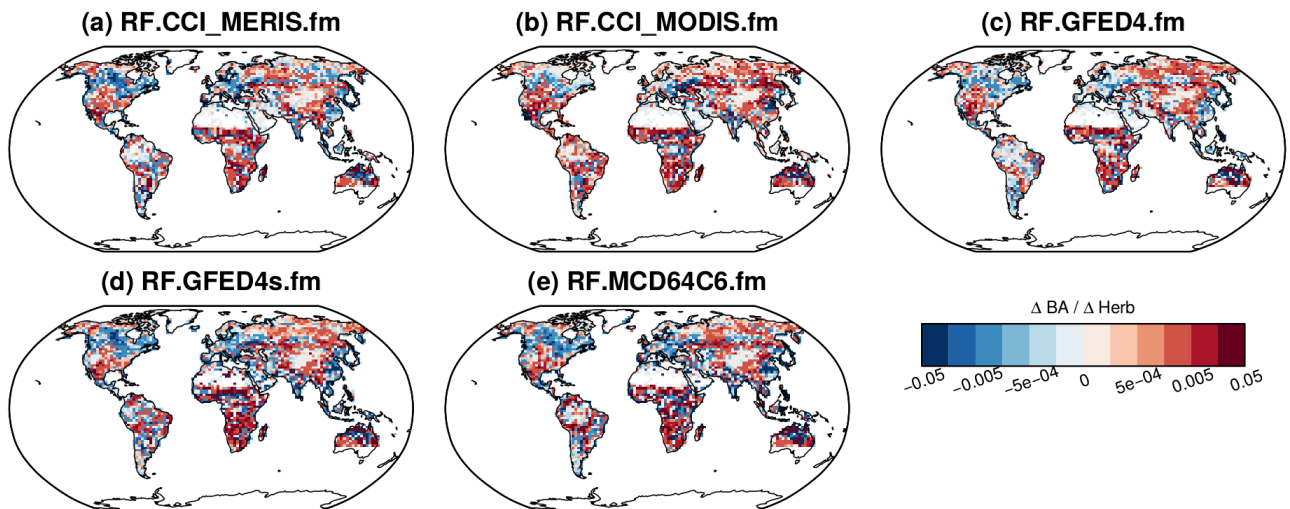



**Figure S 24: Regional sensitivities of the partial fractional burned area per month to the herbaceous vegetation cover from satellite-derived "fm" RF models.**

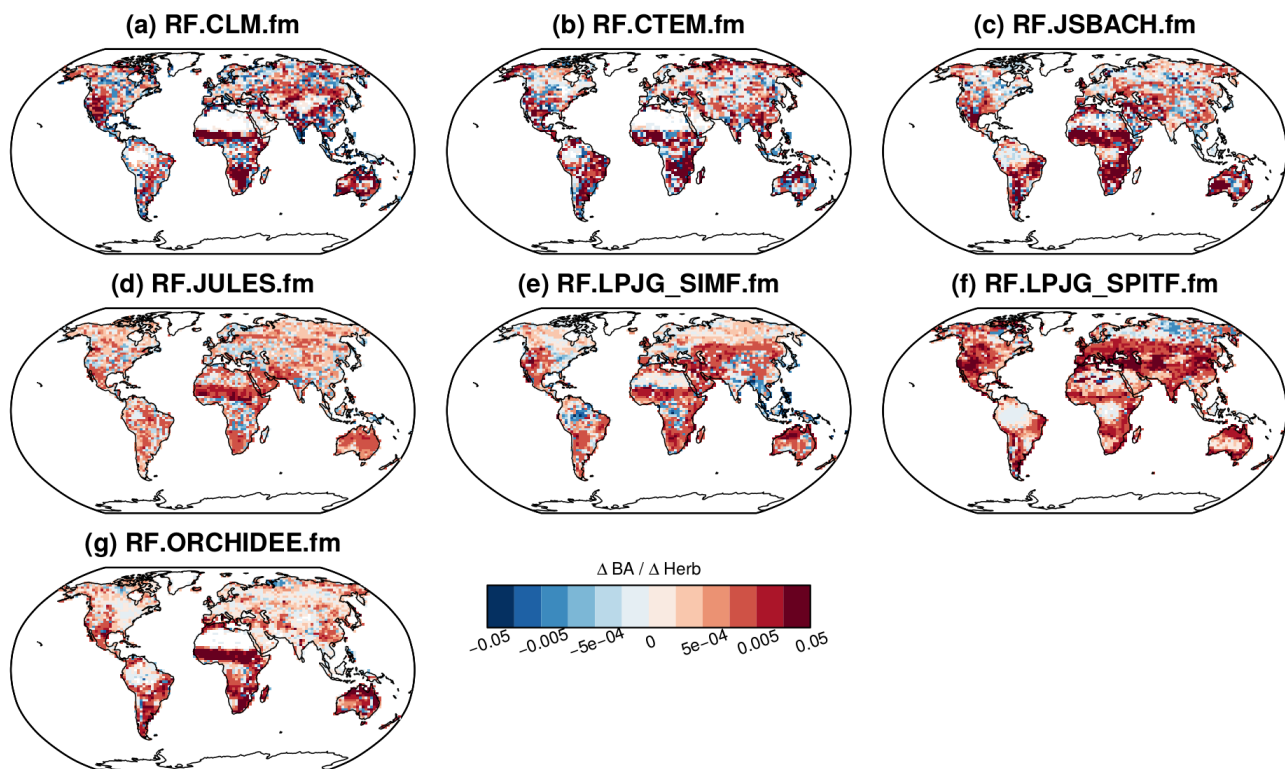



**Figure S 25: Regional sensitivities of the partial fractional burned area per month to the herbaceous vegetation cover from model-derived "fm" RF models.**

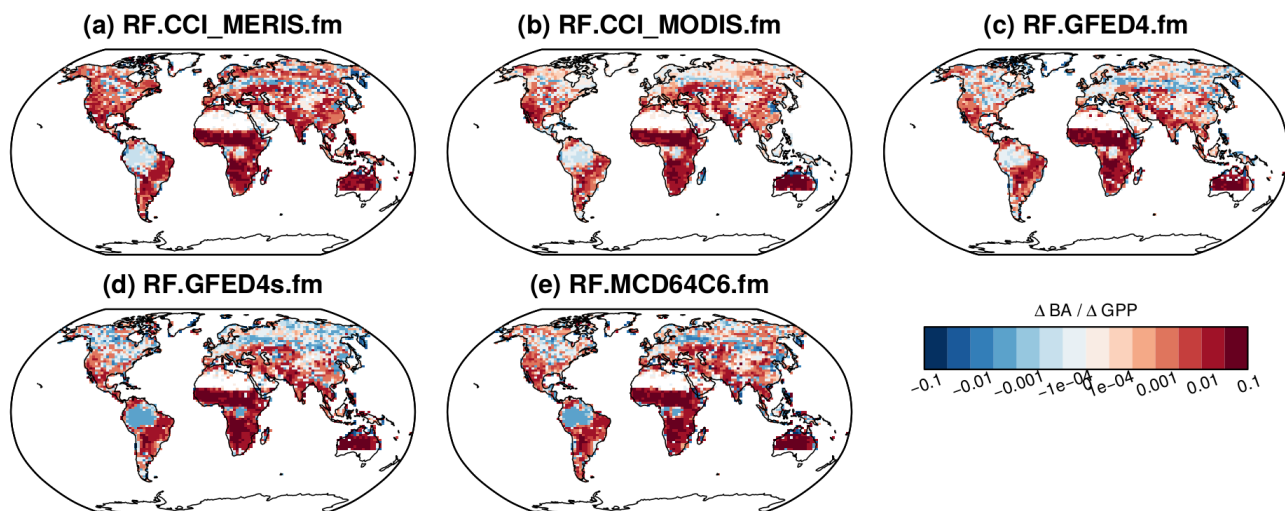



**Figure S 26: Regional sensitivities of the partial fractional burned area per month to precedent 6-monthly GPP from satellite-derived "fm" RF models.**

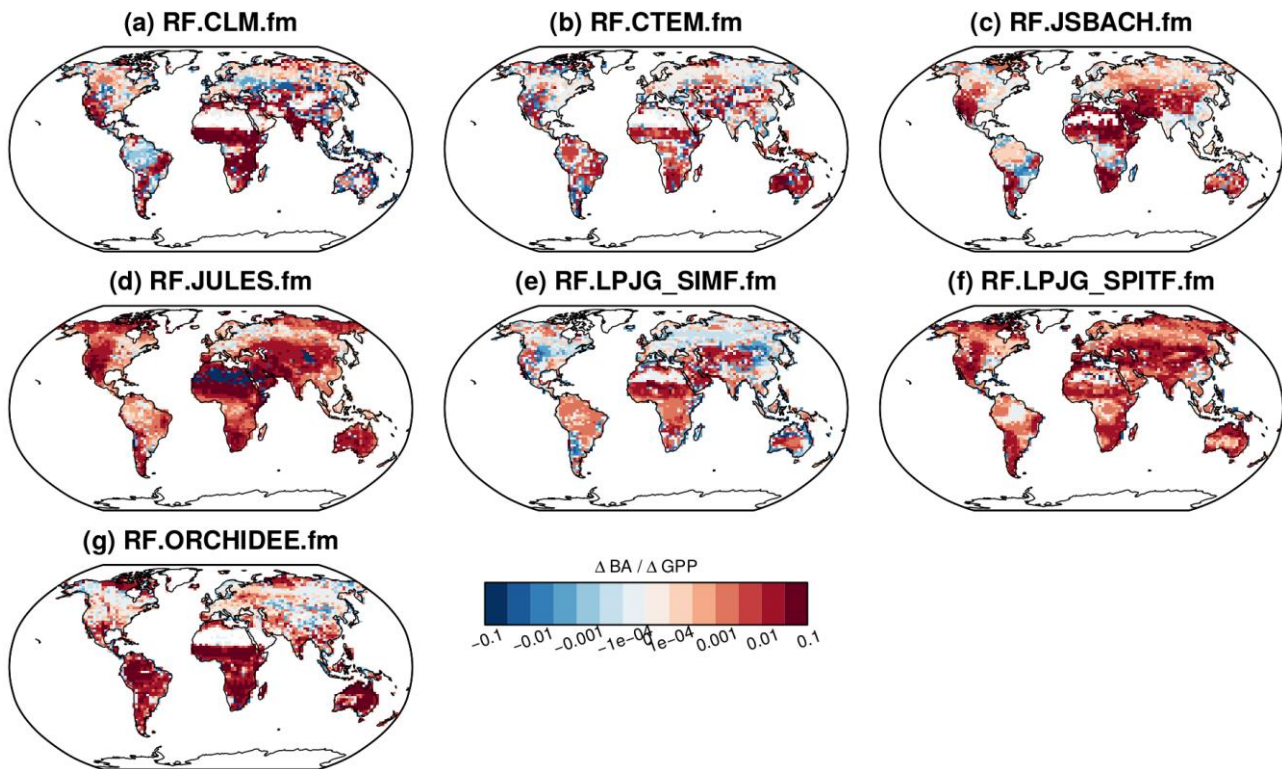



**Figure S 27: Regional sensitivities of the partial fractional burned area per month to precedent 6-monthly GPP from model-derived "fm" RF models.**