

## **Supplementary Material**

### **Strong impact of groundwater on long-term photosynthesis**

Francesco Giardina<sup>1,2</sup>, Jiangong Liu<sup>2</sup>, Sonia I. Seneviratne<sup>1</sup>, Benjamin D. Stocker<sup>3,4</sup>, Pierre Gentine<sup>2,5</sup>

<sup>1</sup>Institute for Atmospheric and Climate Science, Department of Environmental Systems Science, ETH Zurich, CH-8092 Zürich, Switzerland

<sup>2</sup>Department of Earth and Environmental Engineering, Columbia University, New York, New York 10027, USA

<sup>3</sup>Institute of Geography, University of Bern, Hallerstrasse 12, 3012 Bern, Switzerland

<sup>4</sup>Oeschger Centre for Climate Change Research, University of Bern, Falkenplatz 16, 3012 Bern, Switzerland

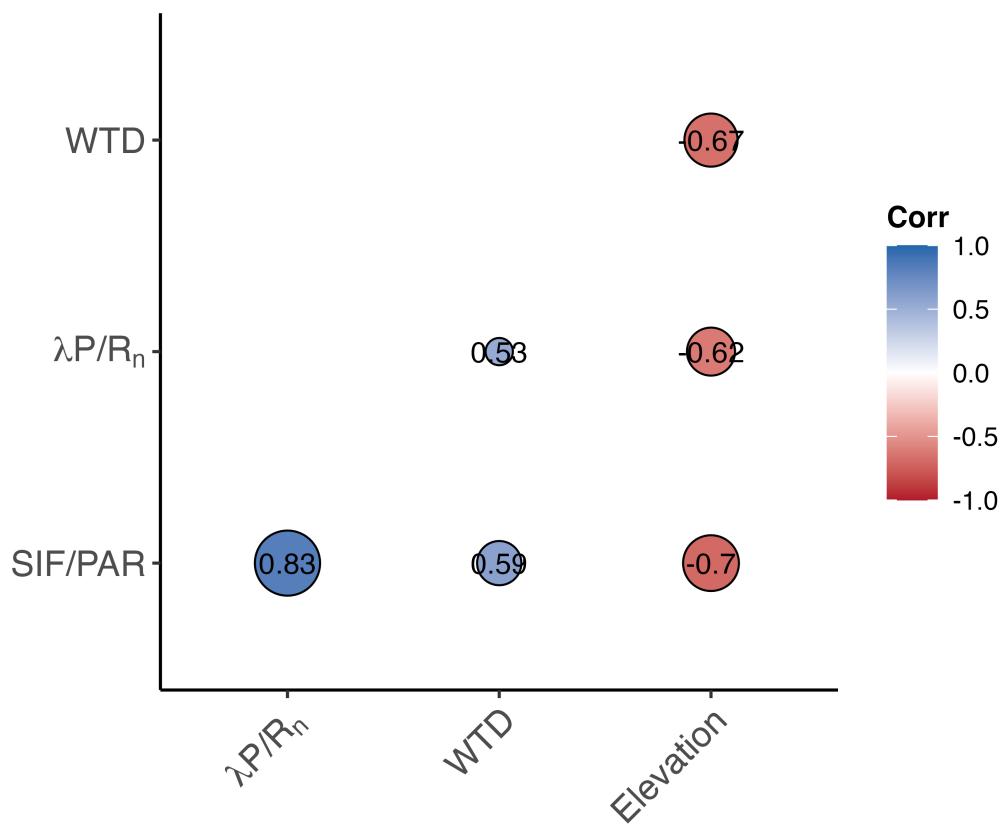
<sup>5</sup>Center for Learning the Earth with Artificial intelligence and Physics (LEAP), Columbia University, New York, New York 10027, USA

Authors for correspondence:

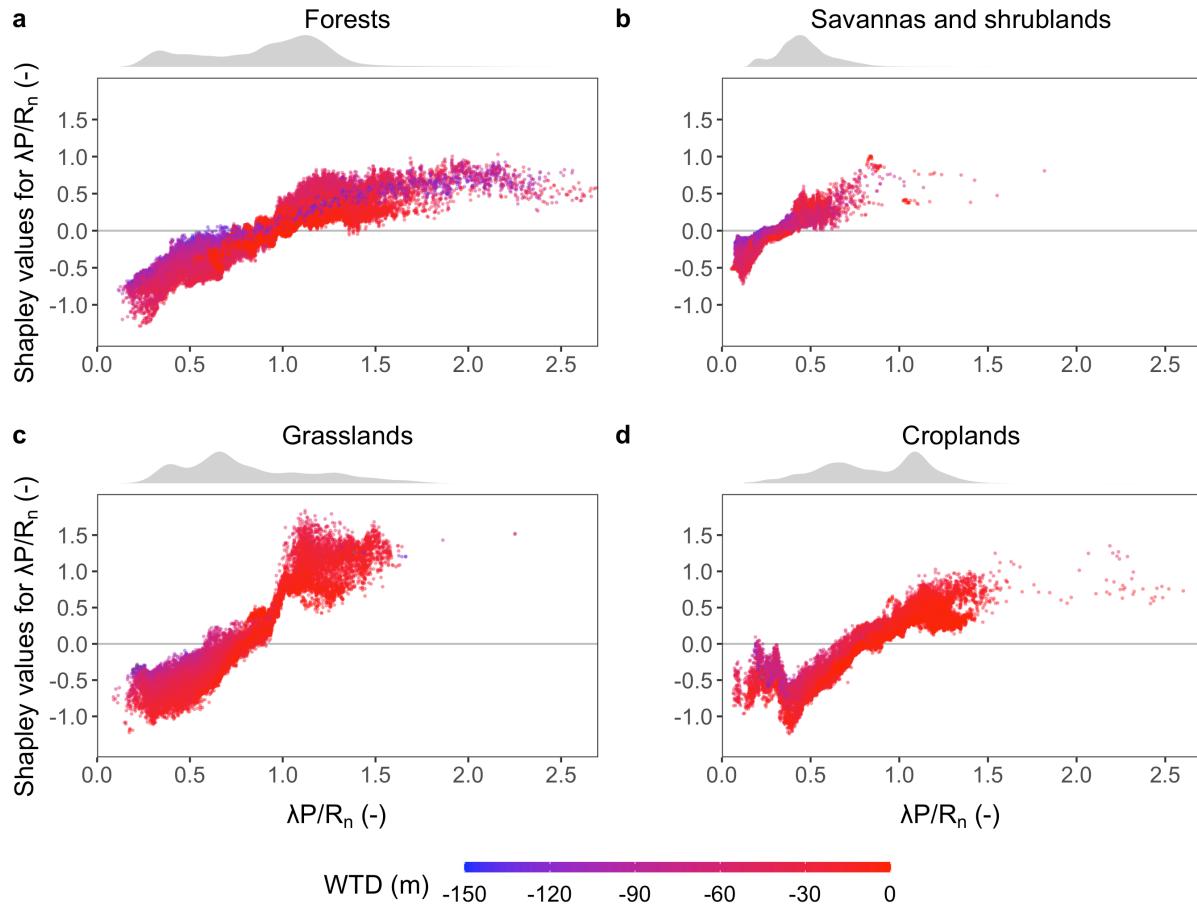
Francesco Giardina, Email: [fgiardina@ethz.ch](mailto:fgiardina@ethz.ch)

Jiangong Liu, Email: [jl6314@columbia.edu](mailto:jl6314@columbia.edu)

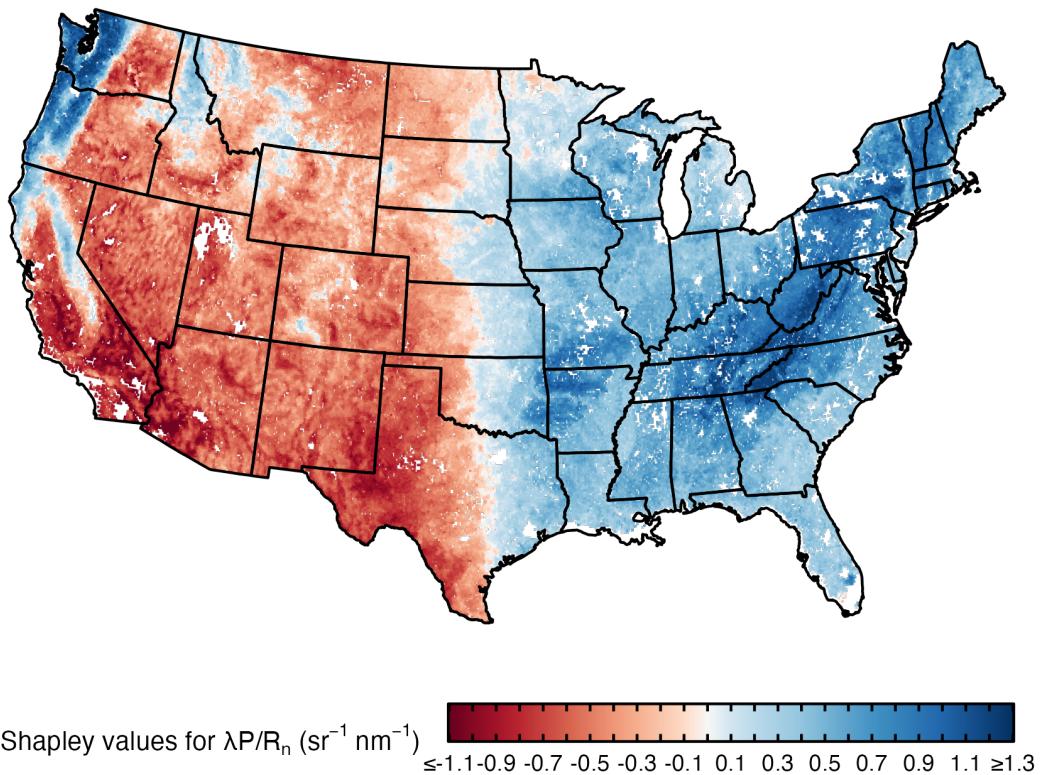
## Supplementary Figures



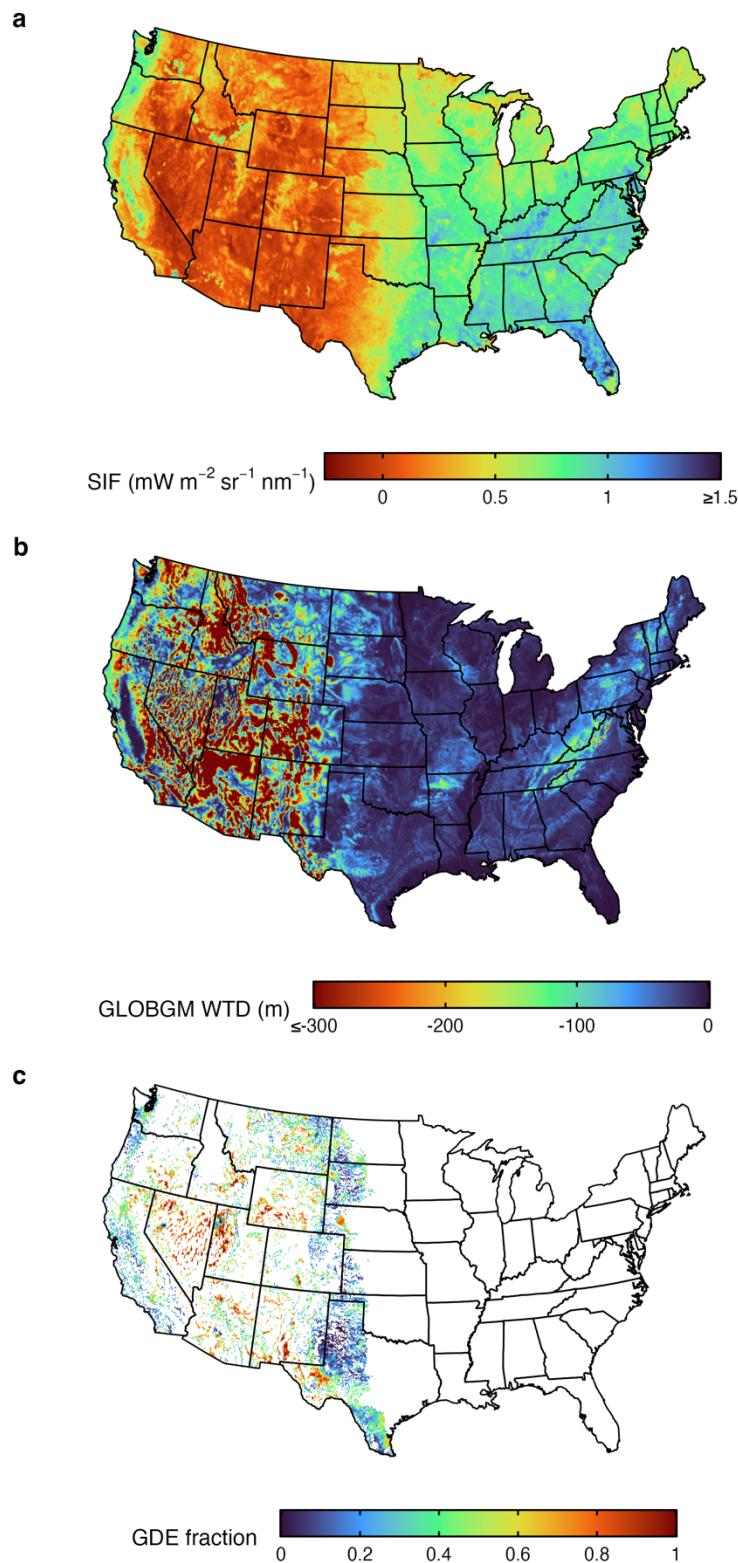
**Supplementary Fig. 1 | Pearson correlation coefficients between the target variable (SIF/PAR) and predictors used in the Extreme Gradient Boosting models across the contiguous United States for all vegetation groups combined.** SIF/PAR: Solar-induced fluorescence to photosynthetically active radiation ratio, proxy of photosynthetic activity. WTD: Water-table depth,  $\lambda P/R_n$ : Moisture index.



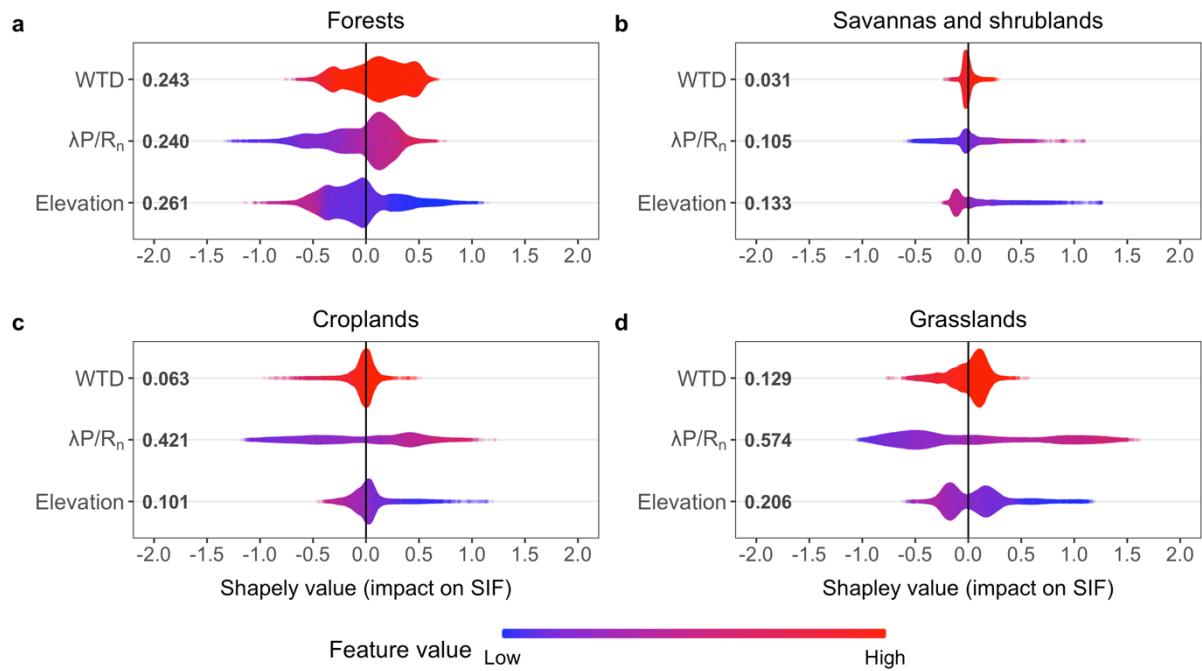
**Supplementary Fig. 2 | Shapley dependence plots of  $\lambda P/R_n$  versus its Shapley value along WTD gradients across plant functional types.** Shapley dependence plot show how a specific predictor affects model outcomes while accounting for interaction effects between predictors. Each dot represents the long-term mean of an individual pixel from global datasets. The colour of a dot indicates the magnitude of the moisture index at that location. WTD: water table depth,  $\lambda P/R_n$ : moisture index. **a**, Forests (evergreen and deciduous, needle-leaved and broadleaved, and mixed forests). **b**, Savannahs and shrublands (savannahs and woody savannahs, open and closed shrublands). **c**, Grasslands. **d**, Croplands. Grey histograms on top of each panel depict marginal density plots for  $\lambda P/R_n$ .



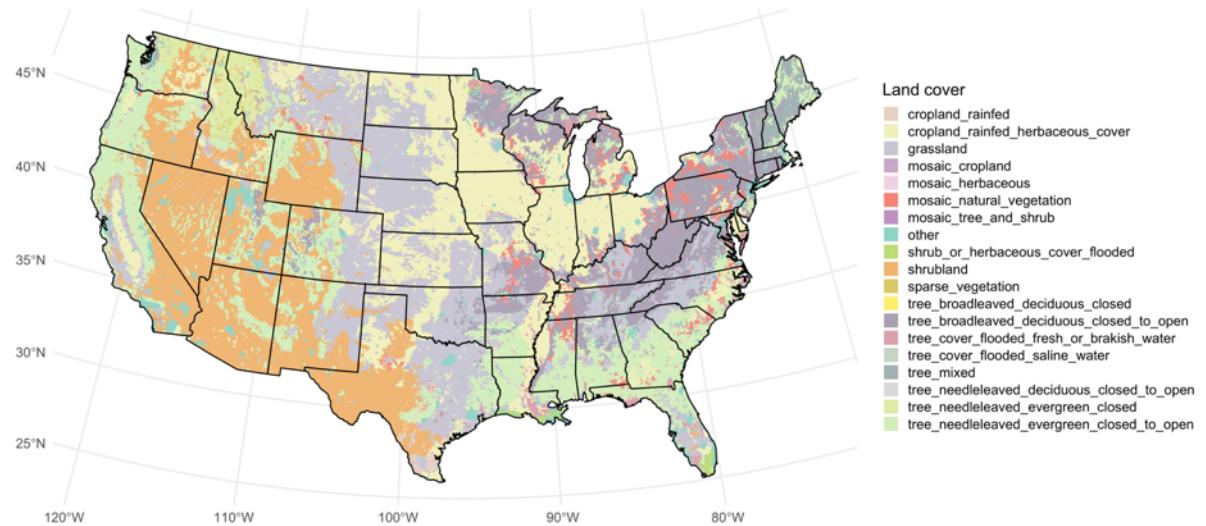
**Supplementary Fig. 3 | Spatial distribution of Causal Shapley values of the Moisture index across the contiguous United States.** Shapley values illustrate the impact of the Moisture index ( $\lambda P/R_n$ ) on SIF/PAR. In this figure, the calculation of the Causal Shapley values is based on a single XGBoost model that includes all vegetation groups combined (see Methods).



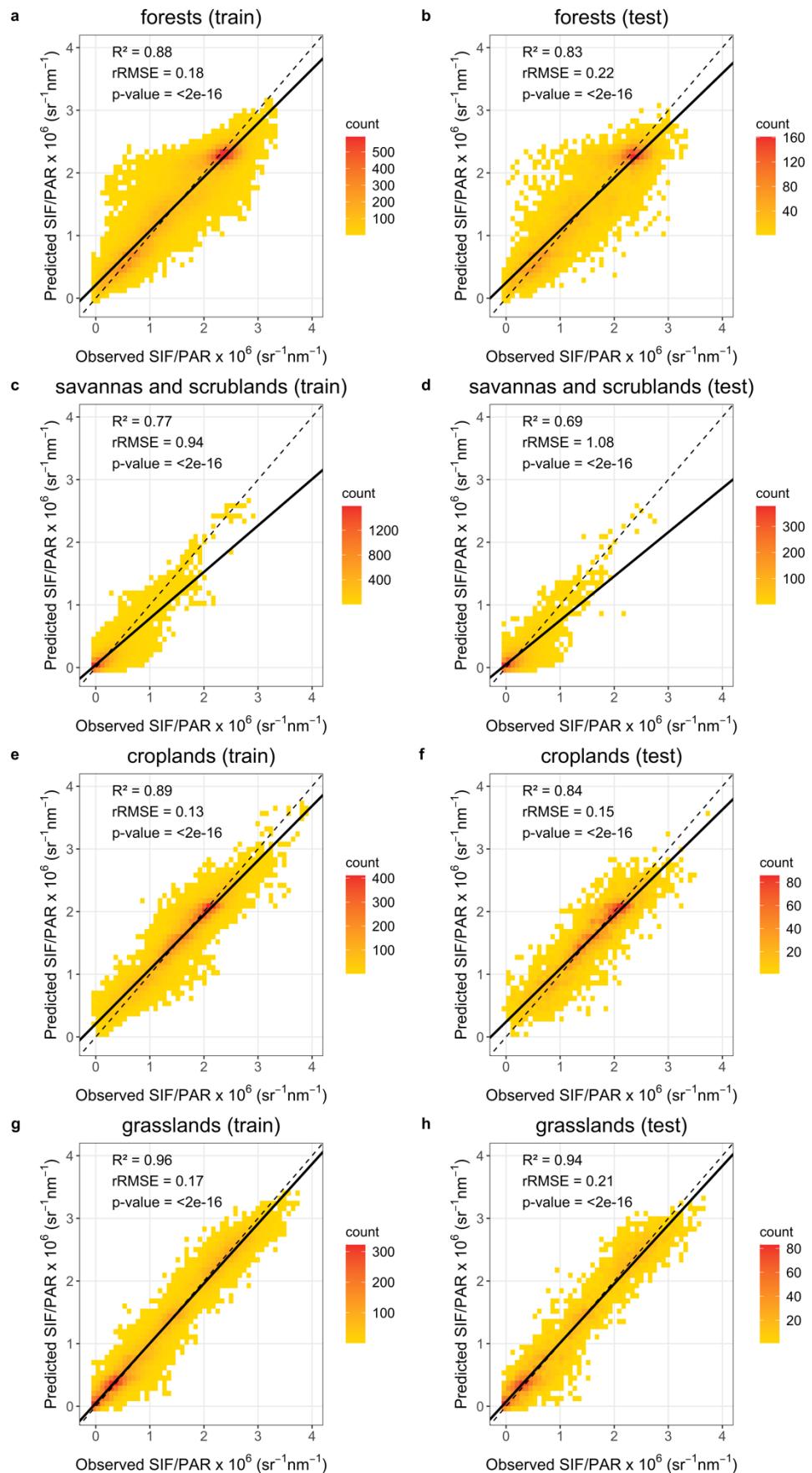
**Supplementary Fig. 4 | Additional spatial patterns of plant-functional types and long-term means of environmental variables in the United States.** **a**, Long-term mean of solar-induced fluorescence (SIF/PAR). **b**, Water table depth (WTD) from GLOBGM dataset. **c**, Fraction of Groundwater-dependent ecosystems (GDE).



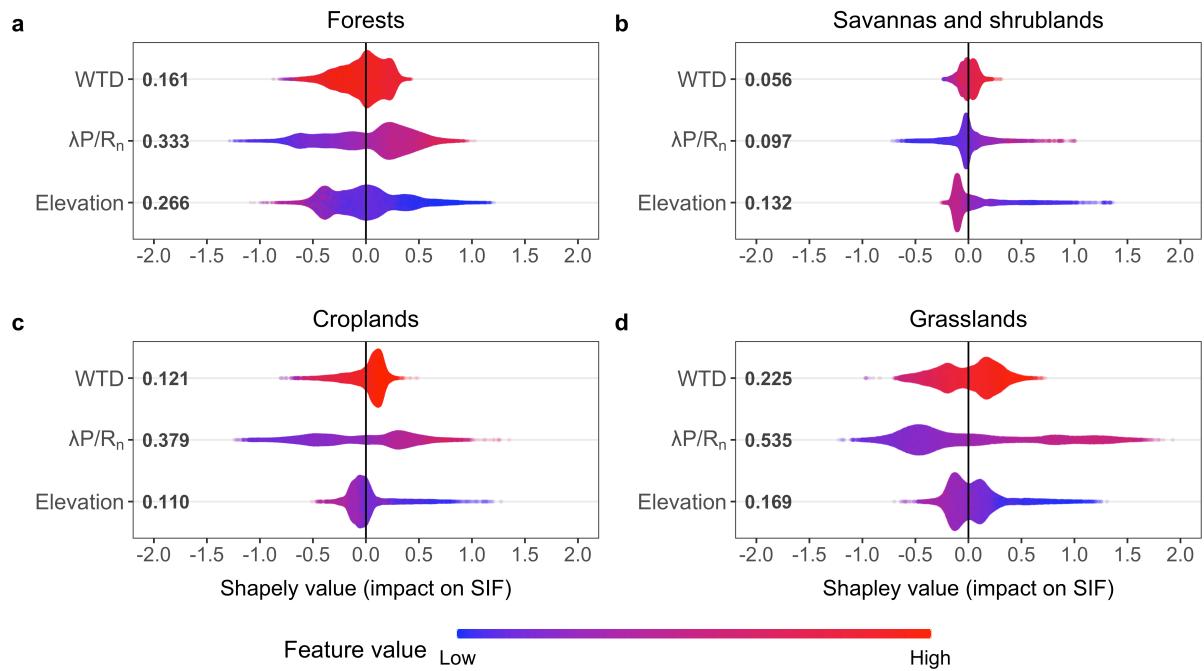
**Supplementary Fig. 5 | Causal Shapley values of ecosystem productivity drivers, based on gradient boosting decision tree models trained on a dataset of solar-induced fluorescence. Calculated using WTD from GLOBGM dataset.** **a-d**, Shapley summary plots, which show the effect of different predictors on each model outcome. Shapley values are in the same units as SIF/PAR. Each dot corresponds to the long-term mean of an individual pixel from global datasets. The mean absolute Shapley value, displayed to the left of each plot near the name of each predictor, represents the mean contribution of that predictor. The colour of a dot indicates the magnitude of each predictor at that location. The x-axis position of a dot represents the local Shapley value of the predictor, indicating how a predictor affects model outcome at the level of a single data point. A positive Shapley value suggests an increase in SIF/PAR while a negative one indicates the opposite. Overlapping dots at an x-coordinate denote higher density, suggesting similar effects across multiple points. WTD: water table depth,  $\lambda P/R_n$ : moisture index. **a**, Forests (evergreen and deciduous, needle-leaved and broadleaved, and mixed forests). **b**, Savannahs and shrublands (savannahs and woody savannahs, open and closed shrublands). **c**, Croplands. **d**, Grasslands.



**Supplementary Fig. 6 | Spatial patterns of land cover types in the United States.**

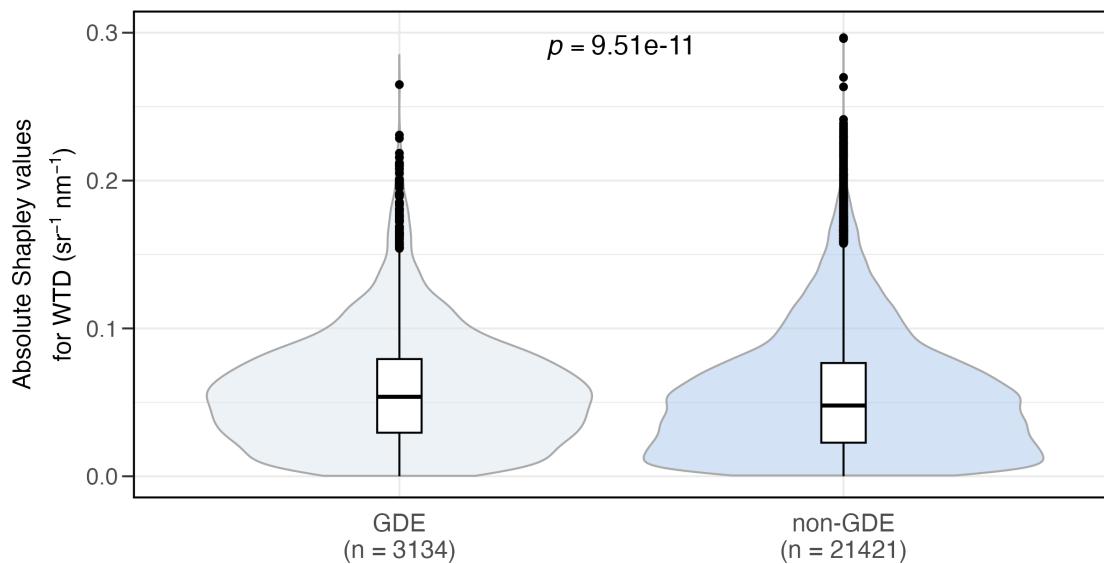


**Supplementary Fig. 7 | Performance of Extreme Gradient Boosting (XGB) models**

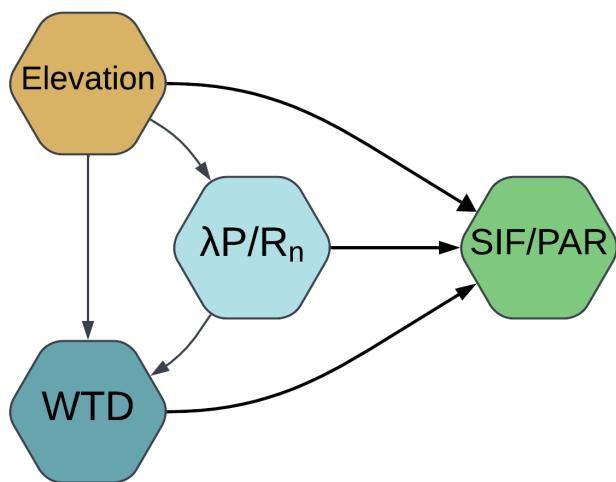


**Supplementary Fig. 8 | Casual Shapley values of ecosystem productivity drivers, based on a gradient boosting decision tree model trained on a dataset of solar-induced fluorescence.**

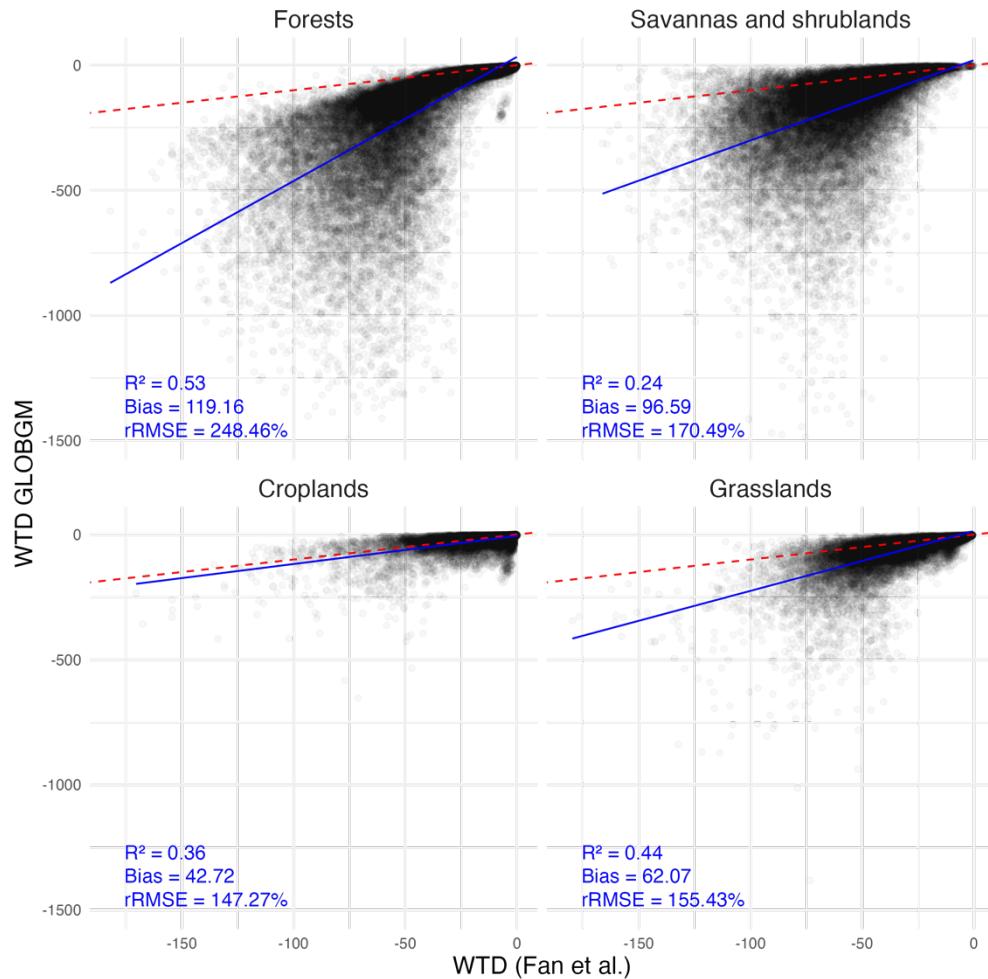
**Including results for Elevation.** **a-d**, Shapley summary plots, which show the effect of different predictors on each model outcome. Each dot corresponds to the long-term mean of the studied variable. The average Shapley value, displayed to the left of each plot near the name of each predictor, represents the mean contribution of that predictor. The colour of a dot indicates the magnitude of each predictor at that location. The x-axis position of a dot represents the local Shapley value of the predictor, indicating how a predictor affects model outcome at the level of a single data point. A positive Shapley value suggests an increase in SIF/PAR while a negative one indicates the opposite. Overlapping dots at an x-coordinate denote higher density, suggesting similar effects across multiple points. WTD: water table depth,  $\lambda P/R_n$ : moisture index. **a**, Forests (evergreen and deciduous, needle-leaved and broadleaved, and mixed forests). **b**, Savannas and shrublands (savannahs and woody savannahs, open and closed shrublands). **c**, Croplands. **d**, Grasslands.



**Supplementary Fig. 9 | Comparison of absolute Shapley values for Water Table Depth (WTD) between GDE and non-GDE pixels within the 'Savannahs and shrublands' group.** Boxplots and violin plots (shaded areas) illustrate the distributions of absolute Shapley values for water table depth (WTD) at pixels corresponding to groundwater-dependent ecosystems (GDE) and non-GDE. A one-tailed Mann-Whitney U test was performed to assess whether the absolute Shapley values for WTD at GDE pixels are significantly greater than those at non-GDE pixels. The test yielded a highly significant result ( $p = 9.51 \times 10^{-11}$ ), confirming that the Shapley values of WTD at GDE pixels are higher compared to non-GDE pixels. We focus on the 'Savannahs and Shrublands' vegetation group, as most GDE pixels fall within this category (Supplementary Fig. 5c and Fig. 2).



**Supplementary Fig. 10 | Directed Acyclic Graph (DAG) illustrating the causal relationships among the drivers of photosynthetic activity selected in this study.** SIF/PAR: Solar-induced fluorescence to photosynthetically active radiation ratio, proxy of photosynthetic activity. WTD: Water-table depth,  $\lambda P/R_n$ : Moisture index. Elevation influences WTD through topographical effects. It also impacts the moisture index by altering precipitation patterns and net radiation; specifically, incoming shortwave radiation increases and incoming longwave radiation decreases with elevation due to atmospheric changes associated with altitude. The moisture index controls WTD because overall climatic moisture conditions influence the level of the water table. Elevation directly affects photosynthesis through temperature, which is why we included it as a predictor.



**Supplementary Fig. 11 | Scatterplots comparing the two WTD datasets used in this study.** The mean standardized bias ('Bias') and the relative root mean square error (rRMSE) were calculated by dividing the mean bias and RMSE by the combined mean of both datasets. Although the moderate  $R^2$  values indicate some correlation, they do not reflect the scale of differences between the datasets. High rRMSE values reveal large discrepancies in magnitude, suggesting that while the datasets may trend similarly, their actual values differ significantly. The combination of high mean standardized bias and high rRMSE demonstrates poor agreement between the datasets.