

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

2 Francesco Giardina^{1,2}, Jiangong Liu², Sonia I. Seneviratne¹, Benjamin D. Stocker^{3,4}, Pierre
3 Gentine^{2,5}

4

5 ¹Institute for Atmospheric and Climate Science, Department of Environmental Systems Science, ETH
6 Zurich, CH-8092 Zürich, Switzerland

7 ²Department of Earth and Environmental Engineering, Columbia University, New York, New York
8 10027, USA

9 ³Institute of Geography, University of Bern, Hallerstrasse 12, 3012 Bern, Switzerland

10 ⁴Oeschger Centre for Climate Change Research, University of Bern, Falkenplatz 16, 3012 Bern,
11 Switzerland

12 ⁵Center for Learning the Earth with Artificial intelligence and Physics (LEAP), Columbia University,
13 New York, New York 10027, USA

14

15 Authors for correspondence:

16 Francesco Giardina, Email: fgiardina@ethz.ch

17 Jiangong Liu, Email: jl6314@columbia.edu

18 **Abstract**

19 Plants can access underground water reserves to sustain their activity, releasing moisture into
20 the atmosphere—a critical survival mechanism during drought. Understanding the role of
21 groundwater in regulating photosynthesis is thus key for predicting land-surface processes.
22 However, the impact of groundwater on terrestrial ecosystem productivity remains poorly
23 quantified, particularly when compared to well-known factors like aridity. Here, we use
24 satellite observations of solar-induced fluorescence as a proxy for photosynthesis, together
25 with model estimates of water table depth and aridity, quantified by the moisture index with
26 reanalysis data, to investigate the relationship between groundwater and photosynthesis.
27 Using causality-guided explainable machine learning, we demonstrate that groundwater plays
28 a crucial role in determining spatial patterns of photosynthesis, with varying importance
29 across ecosystem types, and that its effect is comparable to aridity. We show that the relative
30 importance of groundwater accounts for 48 to 101% of the effect attributed to aridity in
31 modulating forest photosynthesis across the contiguous USA. The relative importance of
32 groundwater compared to the aridity remains substantial in savannahs and shrublands (30-
33 58%), grasslands (22-42%), and croplands (15-32%). Our findings highlight the key role of
34 groundwater in driving ecosystem long-term productivity.

35 **Main text**

36 Groundwater is a critical component of terrestrial water storage¹, yet its impact on
37 photosynthesis remains poorly constrained. Fluctuations in water table levels exhibit distinct
38 spatiotemporal patterns compared to surface soil moisture^{2,3}, as they can be related to long-
39 range landscape topography, soil components, or both, in addition to local climate.
40 Understanding how groundwater modulates photosynthesis and its impact on the global water
41 and energy fluxes is essential in the context of the rising occurrence of dry extremes in
42 several land regions⁴. Water stored below the surface becomes indispensable during dry
43 spells^{5–7}, which are projected to become more frequent and intense under climate change⁴.
44 Deep-rooted plants can tap into underground water pools stored in a saturated zone^{8–10} and
45 then release water into the atmosphere through leaf stomata, regulating atmospheric humidity
46 and influencing regional climate^{11,12}. Determining the control of groundwater on
47 photosynthesis is thus key to accurately predicting land-surface processes and their effects on
48 climate^{13,14}. There has been a growing effort to incorporate groundwater into Earth System
49 Models in recent years^{14–17}, motivated in part by evidence that water tables can influence
50 vegetation over up to 32% of the Earth's surface^{10,14}. However, to improve the representation
51 of groundwater-surface interactions in models, it is necessary to develop a solid empirical
52 basis for detecting, quantifying, and mapping the control of groundwater on plant
53 physiological processes and ecosystem carbon and water fluxes.

54 Systematically assessing the contribution of groundwater to photosynthesis poses
55 significant challenges due to data sampling limitations and the representativity of the in-situ
56 data used, given the complex nature of groundwater across diverse landscapes. Recent
57 research based on inverse modelling has shown that, globally and at the annual scale, the
58 primary source of plant-available water is recent precipitation stored in shallow soil layers
59 (e.g., the top 30 cm)^{2,18}. Concurrently, further studies have documented extensive plant water
60 uptake from deeper soil^{19–22}, rock moisture^{23–25} (water stored in bedrock above the water
61 table)^{21–23} or groundwater^{26–29} (water present in saturated aquifers beneath soil and bedrock).
62 Global syntheses of stable isotope measurements have revealed that groundwater use is
63 widespread across various biomes, with an increased relative contribution in arid or
64 seasonally dry regions^{26,27}. Yet, the presence of deep-rooted plants worldwide suggests that
65 groundwater usage may extend beyond those sparse sampled sites³⁰.

66 Investigations into the uptake of groundwater by plants have thus either focused
67 regionally on syntheses of measurements from stable isotope techniques^{2,26,27}, and at larger
68 scales, they have relied on indirect methods such as geospatial data-based techniques^{31,32},

69 land-surface model^{15,33} and inverse modelling approaches^{2,10,34}. However, there is a lack of
70 research at the multi-year scale that incorporates consistent large-scale data while at the same
71 time accounting for plant physiologies and aridity gradients. Using a remotely sensed dataset
72 of solar-induced fluorescence (SIF) as a proxy for photosynthesis^{35–37}, coupled with maps of
73 plant functional types (PFTs)³⁸ and water table depth (WTD) estimates^{34,39}, we evaluate the
74 contribution of groundwater to long-term photosynthesis with a specific focus on the United
75 States. SIF is nearly proportional with gross primary production (GPP) over large temporal
76 and spatial scales^{36,37}. Applying causality-guided explainable machine learning, we show that
77 groundwater plays a role essentially equivalent to aridity in regulating photosynthesis across
78 PFTs around the globe.

79

80 **Conceptual framework: groundwater as a regulator of the influence of 81 aridity on long-term photosynthesis**

82 An important contribution to the conceptual understanding of regional dryness
83 patterns is the Budyko relationship, which can also illuminate the role of long-term water
84 stress on evapotranspiration and photosynthesis. We present it here in one of its formulations
85 most aligned with our study^{40,41}:

86

$$87 \frac{ET}{P} = f\left(\frac{R_n}{\lambda P}\right) \quad (1)$$

88

89 Where ET is the actual evapotranspiration (typically in mm/year), R_n is net radiation
90 (W/m²) and P is precipitation (mm/year) multiplied by λ, the latent heat of evaporation
91 (J/Kg)^{40,41}. The relationship outlines the partitioning of water across an aridity gradient, from
92 water-limited to energy-limited regimes. It conceptualizes the fact that in drier climates, a
93 larger fraction of the available energy is partitioned into evapotranspiration per unit
94 precipitation^{40,41} (Fig. 1). However, a major limitation of the traditional Budyko framework is
95 that it assumes that the main regulator of ET on long-term equilibrium is local climate^{42,43}
96 and it does not explicitly account for the effect of long-term water storage changes by
97 groundwater, despite its significant role in the water balance. Recent studies, however, have
98 suggested that groundwater can alter this relationship^{42–46}. The inclusion of groundwater in
99 the Budyko framework influences its shape^{42,43}. In particular, a positive groundwater
100 contribution leads to higher evapotranspiration for a given aridity index (or "radiative index
101 of dryness", as referred to by Budyko), thus resulting in points above the Budyko curve in its

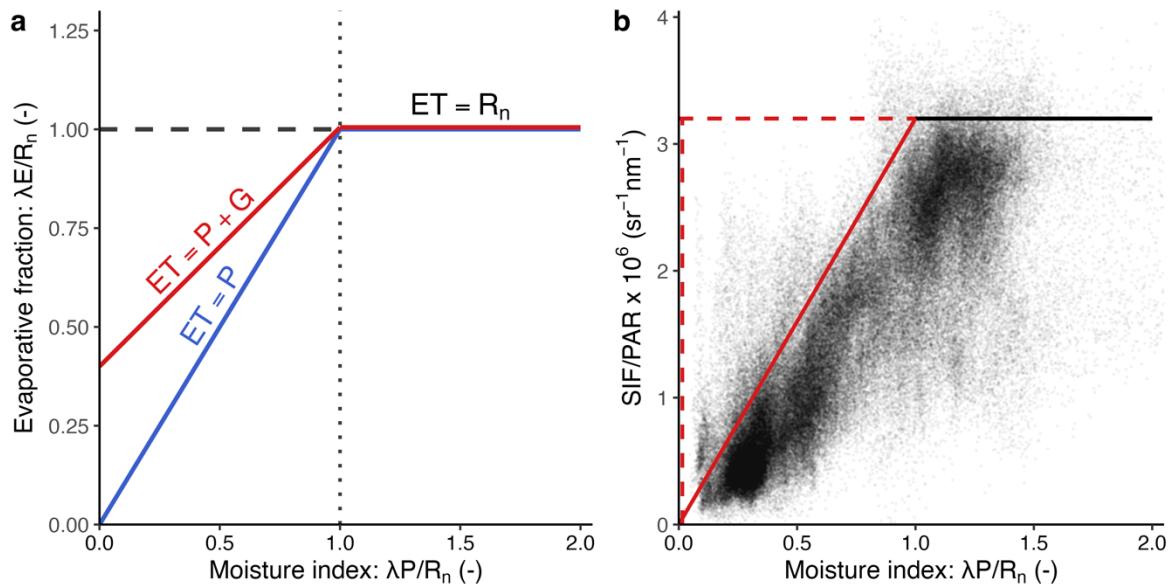
102 water-limited region^{42,43} (i.e. $\lambda P/R_n < 1$ in Fig. 1a). This indicates that groundwater access
103 supports additional evapotranspiration through plant transpiration or direct soil evaporation,
104 enhancing ET beyond the constraints of the local water balance represented by the Budyko
105 curve—which assumes precipitation is the only water input^{42,43}. We here use a similar
106 conceptual approach but for photosynthesis, as photosynthesis and ET are tightly coupled via
107 the water use efficiency, which in turn varies across vegetation communities and species^{47–49}.

108 In our study, we want first to estimate the extent of groundwater use across an aridity
109 gradient. Given the near-linear relationship between SIF and GPP^{36,37} at long time scales, we
110 use SIF based on the high-resolution global TROPOMI sensor⁵⁰ as a proxy for ecosystem-
111 level photosynthesis. To assess variations in SIF beyond climate forcings, we normalize SIF
112 by photosynthetically active radiation (PAR) from ERA5-Land⁵¹ (SIF/PAR, see Methods).
113 SIF has been shown to accurately quantify the impact of environmental stress on ecosystem
114 transpiration (T)^{49,52} and shows a closer relation to T than any other space-based
115 measurements⁴⁷. We thus use SIF/PAR as an observational proxy of the evaporative fraction
116 (defined as the ratio of the latent heat flux to net radiation) on the y-axis in the Budyko
117 framework (Fig. 1b). We define aridity as the ratio of MSWEP precipitation⁵³ (P) to ERA5-
118 Land net radiation⁵¹ (R_n), denoted as $\lambda P/R_n$, where λ is the latent heat of vaporization ($J\ kg^{-1}$)
119 for units consistency. This metric provides a simple yet effective way to characterize the
120 water availability of a region in relation to the available energy and, since P and R_n are based
121 on long-term averages, it is an index of the mean climatic conditions of the area^{42,43,54–56}.
122 Given that higher $\lambda P/R_n$ values indicate lower aridity (the inverse of Budyko's formulation)
123 and more water availability, we adopt the term moisture index rather than aridity index to
124 avoid confusion, in line with many previous studies^{5,57}. It is important to note that, while
125 commonly used in hydroclimatological studies^{42,43,54–56}, the index is just one of the possible
126 formulations of aridity⁵⁸.

127 As large-scale remote sensing datasets do not provide the exact variables used by
128 Budyko (e.g. ET), we describe our approach as 'Budyko-like', indicating an adaptation of the
129 original framework rather than a direct application (Fig. 1). This is similar to a moisture
130 limitation model previously formulated by Manabe based on Budyko ("bucket model")^{12,59},
131 but on long-time scales. In this framework, we expect that a positive contribution of
132 groundwater results in points higher than the Budyko curve in the water-limited regime (Fig.
133 1a, red versus blue lines). We identified pixels deviating from the moisture index as those
134 potentially located in zones where plants might access groundwater. Note that this figure is

135 not a quantitative assessment of groundwater access; rather, it offers a qualitative overview of
 136 the issue to contextualize our study and motivates the subsequent sections.

137
 138

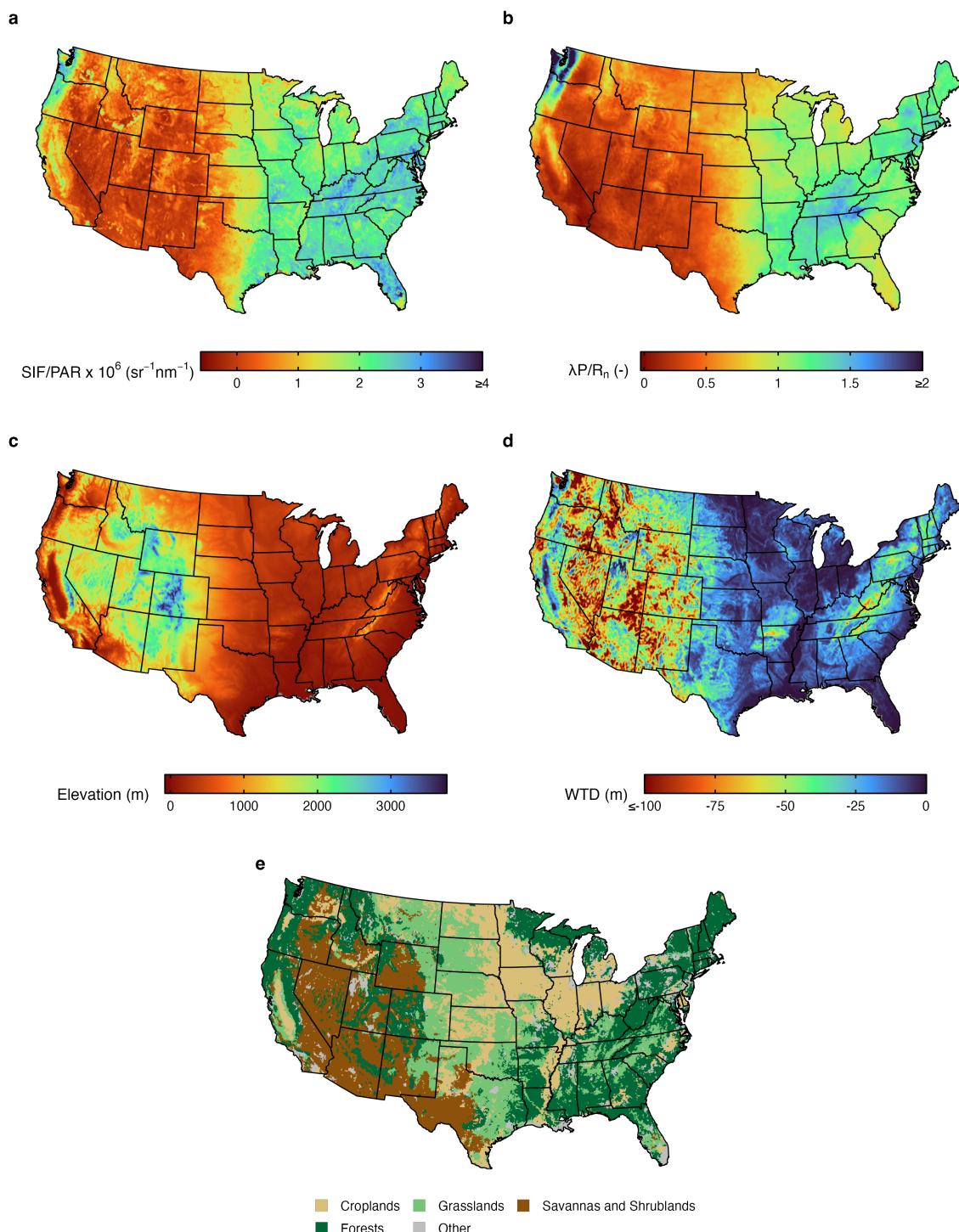


139
 140 **Fig. 1 | Groundwater in the Budyko-like framework.** The x-axes represent the moisture index
 141 ($\lambda P/R_n$), where higher values indicate increasingly humid conditions. **a**, Conceptual Budyko-like
 142 framework. The area between the blue and red lines in the water-limited regime ($\lambda P/R_n < 1$) indicates
 143 groundwater access due to ET values exceeding expected limits from the moisture index⁴². ET:
 144 evaporation, P: precipitation, G: groundwater. R_n : net radiation. λ : latent heat of vaporization. **b**,
 145 SIF/PAR distribution in the Budyko-like space, with each point representing a pixel across the
 146 CONUS. SIF/PAR serves as an observational proxy for the evaporative fraction. Continuous red and
 147 black segments illustrate the Budyko-like framework as per framework in **a**. Points within the red
 148 triangle indicate areas where plants potentially access groundwater.
 149

150 Local contributions of groundwater table and moisture index

151 To quantitatively study the relative contributions of groundwater and aridity on
 152 photosynthesis, we trained Extreme Gradient Boosting (XGBoost) models using long-term
 153 means of SIF/PAR (Fig. 2a) as the target variable with WTD, $\lambda P/R_n$, and elevation as
 154 predictors (Fig. 2b–d; see Methods). Thanks to their ability to encode non-linear
 155 relationships, machine learning models based on decision trees, such as XGBoost, have
 156 gained increasing popularity in ecological applications^{60–62}. Consistent with previous
 157 studies^{63,64}, we aggregated PFTs into four groups: forests, shrublands and savannahs,
 158 grasslands, and croplands, and trained one XGBoost model per group (Fig. 2e, see Methods).
 159 Since TROPOMI was launched in 2018, we used a three-year record of SIF data (March
 160 2018–March 2021) and calculated the mean value over this period for each variable at each
 161 0.083° pixel. Our models show good predictive performance for the United States, with $R^2 \geq$

162 0.69 on the test set for all vegetation groups ($R^2=0.83$ for forests, $R^2=0.69$ for savannahs and
 163 shrublands, $R^2=0.94$ for grasslands and $R^2=0.84$ for croplands, Supplementary Fig. 7). This
 164 demonstrates that the XGBoost models are effective in predicting long-term SIF/PAR. To
 165 assess the individual contributions of WTD and $\lambda P/R_n$ to the sensitivity of SIF/PAR, we
 166 calculated Causal Shapley values⁶⁵ for each XGBoost model, i.e. for each vegetation group
 167 (Fig. 2e).



168

169 **Fig. 2 | Spatial patterns of long-term means of climatic and environmental variables in the**
170 **United States.** **a**, Solar-induced fluorescence to photosynthetically active radiation ratio (SIF/PAR). **b**,
171 Moisture index represented as the ratio of precipitation to net radiation ($\lambda P/R_n$). **c**, Elevation above sea
172 level. **d**, Water-table depth (WTD) from Refs^{10,34}. **e**, Classification of vegetation types.
173

174 Machine learning models are based on nonlinear algorithms that typically yield higher
175 accuracy than linear models, especially on large datasets. Yet, they often lack interpretability
176 ('Black box' model predictions⁶⁶). Shapley values have been used in ecology for interpreting
177 model predictions⁶³. They are based on an algorithm rooted in game theory and describe the
178 influence of each input feature (or predictor, i.e. WTD, $\lambda P/R_n$ and elevation) on a specific
179 prediction (i.e. SIF/PAR value predicted by the XGBoost models). In our analysis, it is
180 essential to include elevation as a predictor due to its control over long-term photosynthesis⁶⁷.
181 However, elevation acts as a confounder because it also influences WTD and the moisture
182 index¹⁴. We thus use Causal Shapley Values⁶⁵ to constrain the relative contributions of WTD
183 and $\lambda P/R_n$ to SIF/PAR and factor out the confounding effect of elevation (Supplementary
184 Fig. 1).

185 Shapley values directly quantify local (i.e., for a single data point) effects of
186 individual predictors, while at the same time allowing to account for the global structure of
187 the model, thus unveiling meaningful patterns that might otherwise remain undetected^{66,68}.
188 More positive Shapley values indicates that a particular predictor leads to an increase in the
189 target variable (i.e. SIF/PAR) above the average outcome predicted by the XGBoost model
190 (i.e. the baseline value). More negative Shapley value denotes the opposite. Shapley values
191 that approach zero thus indicate that SIF/PAR is nearing its average value across the dataset.
192 This, however, should not be interpreted as lack of sensitivity with the corresponding
193 predictor. Traditional Shapley values are valid under the assumption that predictors are
194 independent. In contrast, the Causal Shapley value approach used here is based on causal
195 chain diagrams informed by expert knowledge about cause-and-effect relationships to better
196 constrain the relative feature contributions to model prediction⁶⁵. This approach effectively
197 addresses the independence assumption by explicitly considering the causal relationships
198 among variables (see Methods and Supplementary Fig. 10). The absolute Shapley value of a
199 predictor represents its specific effect on an individual model prediction (local contribution),
200 whereas averaging these values across all data samples of a specific predictor yields a
201 measure of the global importance of that predictor (average of absolute Shapley values)⁶⁵.

202 We compute the average of absolute Causal Shapley values for WTD and $\lambda P/R_n$
203 across all pixels in CONUS (Fig. 3, numbers shown on the left of each plot). Our analysis

204 indicates that in forested areas, the relative importance of groundwater (average of absolute
205 Shapley values of $0.161 \times 10^{-6} \text{ sr}^{-1} \text{ nm}^{-1}$, Fig. 3a) is 48% of the effect linked to the moisture
206 index (average of absolute Shapley values of $0.333 \times 10^{-6} \text{ sr}^{-1} \text{ nm}^{-1}$, Fig. 3a) in determining
207 ecosystem productivity. In savannahs and shrublands, the relative importance of groundwater
208 compared to the moisture index is 58% (average of absolute Shapley values of 0.056 and
209 $0.097 \times 10^{-6} \text{ sr}^{-1} \text{ nm}^{-1}$, respectively, Fig. 3b) decreasing to 42% in grasslands (average of
210 absolute Shapley values of 0.225 and $0.535 \times 10^{-6} \text{ sr}^{-1} \text{ nm}^{-1}$, respectively, Fig. 3d), and to 32%
211 in croplands (average of absolute Shapley values of 0.121 and $0.379 \times 10^{-6} \text{ sr}^{-1} \text{ nm}^{-1}$, Fig. 3c).
212 This outcome is consistent with the fact that forests, savannahs and shrublands typically have
213 deeper-rooted and longer-lived plants compared to other PFTs, allowing them to access water
214 sources at greater depths and sustain productivity over longer periods^{30,34}. As a result, their
215 productivity is more affected by changes in WTD. Note that, especially in the Western USA,
216 croplands are often irrigated with water extracted from aquifers, which could affect the
217 estimate of the relative importance of groundwater in these areas⁶⁹.

218 We recognize that the accuracy of our findings depends on the quality of the input
219 data, and notably the WTD estimates. To have a sense of the uncertainty in our approach, we
220 repeated the analysis using an alternative WTD dataset³⁹. Despite the limited agreement
221 between the two modelled WTD datasets (Supplementary Figs 11 and 4b, Fig. 2), our results
222 with the alternative dataset confirm that WTD exerts a significant control on SIF/PAR
223 (Supplementary Fig. 5). Specifically, the relative importance of WTD versus the moisture
224 index using this alternative WTD dataset is 101% for forests, 30% for savannahs and
225 shrublands, 15% for croplands, and 22% for grasslands. These values should be considered as
226 indicators of the general control exerted by WTD relative to aridity on photosynthesis,
227 demonstrating for the first time that both factors operate within the same order of magnitude.
228 To account for the uncertainty in estimating these relative importances, we present our results
229 as intervals derived from both WTD datasets.

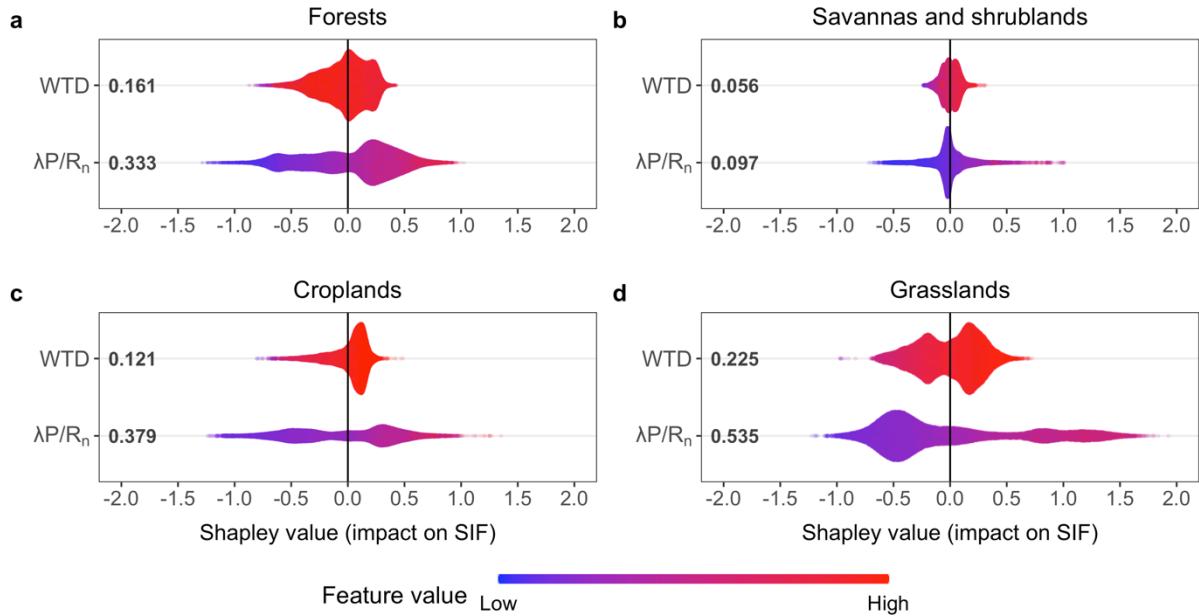
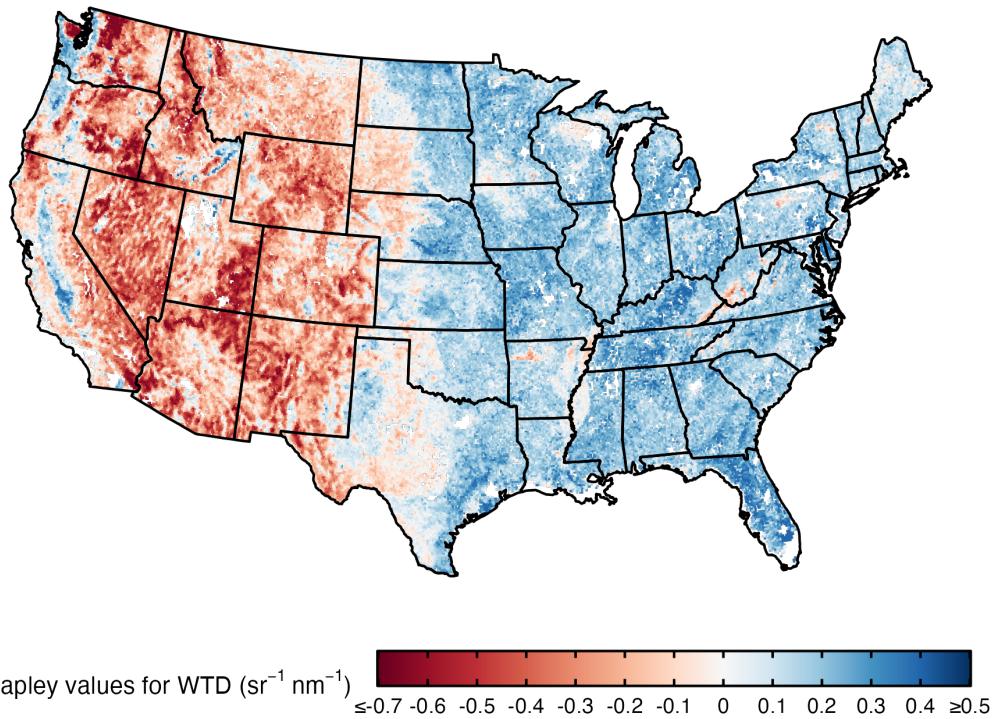


Fig. 3 | Causal Shapley values of ecosystem productivity drivers, based on gradient boosting decision tree models trained on a dataset of solar-induced fluorescence. a-d, Shapley summary plots, which show the effect of different predictors on each model outcome. Shapley values are in the same units as SIF/PAR. Each dot corresponds to the long-term mean of the studied variable. 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. 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.

Shapley summary plots identify high-magnitude effects that would be difficult to discern otherwise (Fig. 3, bee-swarm plots). The long tails depicted in Fig. 3 demonstrate that a feature with relatively low global importance can still hold significance for an individual sample, i.e. at a specific pixel or group of pixels. Overall, pixels exhibiting a higher moisture index (indicated by higher $\lambda P/R_n$, coloured in pink in the lower halves of each panel in Fig. 3) tend to yield higher local output values of SIF/PAR, as demonstrated by the corresponding positive Shapley values (Fig. 3). The same can be said concerning pixels with a high WTD value: they tend to have a positive effect on the model outcome.

The spatial distribution of the Causal Shapley values for WTD (Fig. 4) highlight that regions like the Pacific Northwest (e.g. coastal Washington and Oregon), the Central Valley of California, Florida and in general the Eastern USA emerge as zones where WTD has a positive effect on photosynthesis (i.e. positive Shapley values, blue areas in Fig. 4). In contrast, areas such as the arid southwestern USA (e.g. Arizona, New Mexico, Nevada, Utah)

257 and in general the area of the Rocky Mountains show a negative effect of WTD on
258 photosynthesis (red areas in Fig. 4). The spatial distribution of the Causal Shapley values for
259 $\lambda P/R_n$ (Supplementary Fig. 3) aligns with the long-term means of $\lambda P/R_n$ (Fig. 2b).



260
261 **Fig. 4 | Spatial distribution of Causal Shapley values of Water-table depth across the contiguous**
262 **United States.** Shapley values illustrate the impact of Water-table depth (WTD) on SIF/PAR. In this
263 figure, the calculation of the Causal Shapley values is based on a single XGBoost model that includes
264 all vegetation groups combined (see Methods).

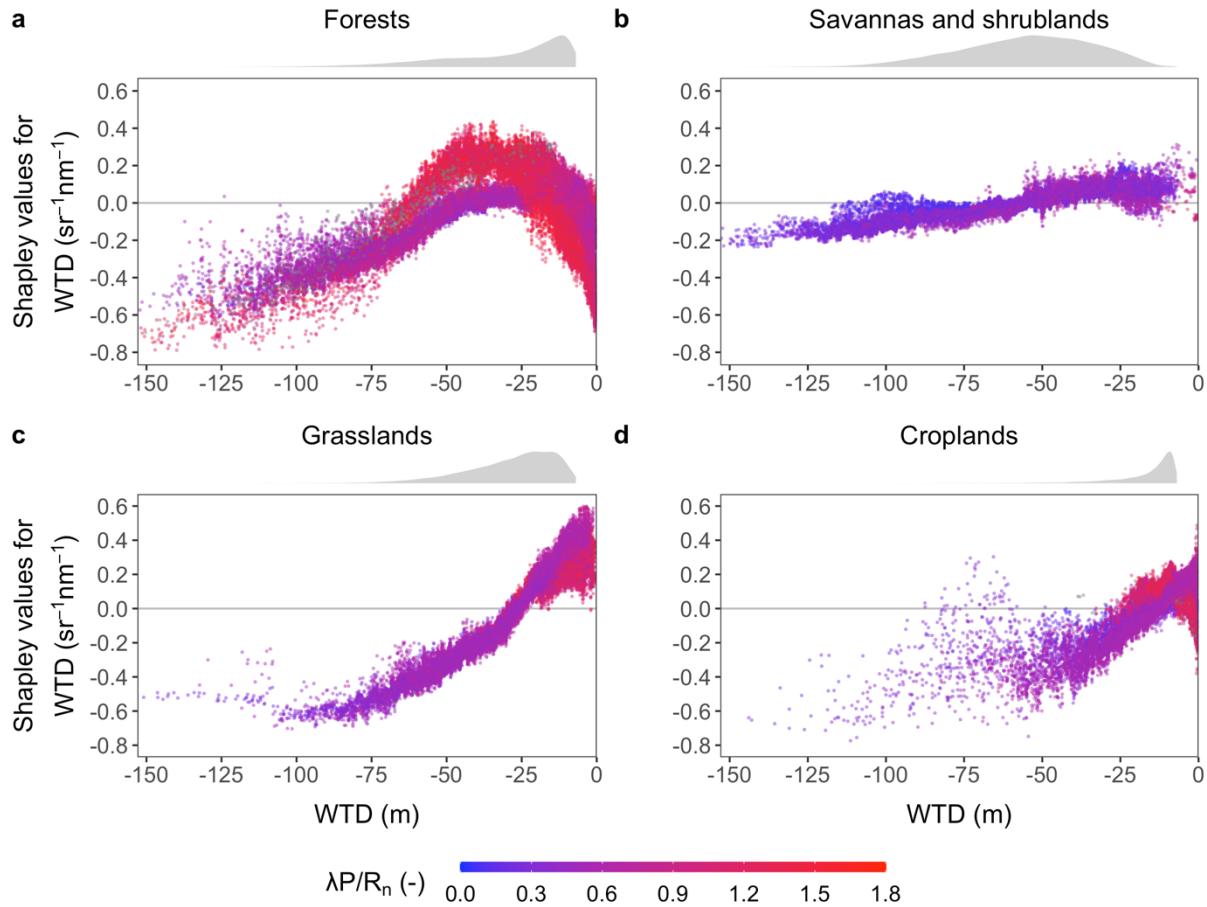
265
266 To further support our findings, we compared the absolute Shapley values of WTD
267 between groundwater-dependent ecosystems (GDEs) and non-GDE pixels, using a recently
268 published map of GDEs⁷⁰ (Supplementary Fig. 4c). We use absolute Shapley values to
269 capture the magnitude of their impact, irrespective of their sign. Our results consistently show
270 that the distribution of absolute Shapley values for WTD at GDE pixels is significantly
271 greater than at non-GDE pixels (Supplementary Fig. 9).

272

273 **Coupling effect between water table depth and moisture index**

274 Shapley dependence plots reveal the impact of a particular variable on the prediction,
275 as well as the coupling or dependence effect between two factors (Fig. 5). A coupling effect
276 between WTD and the moisture index is discernible in Fig. 5, characterized by a vertical
277 dispersion in the Shapley dependence plots.

278



279
280 **Fig. 5 | Shapley dependence plots of water table depth versus its Shapley value along aridity**
281 **gradients across vegetation types.** Shapley dependence plots show how a specific predictor (x -axis)
282 affects model outcomes (y -axis) for each data-point, while accounting for interaction effects between
283 predictors. Each dot represents the long-term mean of the studies variable. The colour of a dot
284 indicates the magnitude of the moisture index at that location. WTD: water table depth, $\lambda P/R_n$:
285 moisture index. **a**, Forests (evergreen and deciduous, needle-leaved and broadleaved, and mixed
286 forests). **b**, Savannahs and shrublands (savannahs and woody savannahs, open and closed shrublands).
287 **c**, Grasslands. **d**, Croplands. Grey histograms on top of each panel depict marginal density plots for
288 WTD.

289 The dependence plot of forests reveals a range of intermediate WTD values associated
290 with positive Shapley values, indicative of stress-free and positive conditions for vegetation
291 photosynthesis (Fig. 5a). Deviating from this range, either towards higher or lower WTD
292 values, respectively corresponds to water-logging or highly arid conditions, resulting in
293 vegetation stress, as evidenced by the corresponding negative Shapley values. These results
294 align with the negative effects of perched water tables on photosynthesis observed under
295 water-logging conditions, which leads to poor oxygenation of the roots^{71,72}. Laboratory
296 experiments have shown that GPP and stomatal conductance exhibit negative responses
297 shortly after the onset of water-logging conditions^{71,72}. Moreover, long-term observations in
298 boreal forests have demonstrated that persistent water-logging conditions reduce surface

conductance, evapotranspiration (ET), and GPP⁷³. Conversely, low WTD values produce a negative effect on SIF sensitivity, as indicated by the corresponding negative Shapley values. This finding aligns with previous literature, which suggests that deep water tables, typically found in arid zones, negatively impact vegetation productivity^{10,12,34}. Distinct WTD dependence curves corresponding to different $\lambda P/R_n$ conditions (i.e. curves with different colours in Fig. 5) are observed. This potentially indicates a relationship between a specific response to WTD within the same vegetation group along the prevailing climatic gradient. In grasslands, the absence of negative effects for shallow WTD values (i.e. positive Shapley even at very shallow WTD in Fig. 5c) aligns with their greater adaptation to water-logging conditions⁷⁴. Only the negative effects resulting from very low WTD values are evident (i.e. left tail of Fig. 5c). In the case of croplands, the negative effect of waterlogging is also present in croplands (Fig. 5c), indicated by the negative Shapley values at shallow WTD. Consistently, very high values of $\lambda P/R_n$ associated with moist climates have minimal impact on vegetation productivity, as indicated by levelling-off observed in the dependence plot of $\lambda P/R_n$ (Supplementary Fig. 2).

In forests, WTD has a positive effect on SIF/PAR up to approximately -25 m, after which there is a levelling-off where WTD exerts little to no influence up to around -45 m. Beyond this point, the Shapley values decrease, suggesting a negative impact on SIF. In grasslands, savannahs and shrublands, and to a certain extent in Croplands too, the plot shows a consistent downward trend, indicating that lower WTD levels are associated with progressively lower effects on photosynthesis. It is important to note that in Fig. 5, we present on the x-axis the depths of water tables potentially capable of impacting photosynthesis. These water table depths might exceed plant rooting depths, i.e. water tables situated deeper than plant roots can still play a role in photosynthesis, for instance by indirectly regulating soil moisture in the unsaturated zone³.

324

325 Conclusion

326 In this study, we focused on disentangling the effects of aridity versus groundwater on 327 photosynthesis. A key point in our analysis is the Causal Shapley values approach, which 328 accounts for the causal relationships between predictors, thus avoiding the confounding 329 effects of cross-correlation. The high R^2 of our XGBoost models ($R^2 \geq 0.69$ on the test set for 330 all vegetation groups, Supplementary Fig. 7) contributes to the robustness of our explainable 331 machine-learning approach. In line with a principle of simplicity, we focused on long-term 332 means, reflecting the perspective of the long-term partitioning of water and energy of the

333 Budyko-like framework (Fig. 1). Working with long-term means also has the advantage of
334 avoiding dealing with interannual variability, which can further complicate the relationship
335 with groundwater. Groundwater and aridity modulate the long-term mean of SIF, but it is
336 challenging to quantify this effect with a physical model. In this context, our data-driven
337 explainable machine-learning approach proves to be particularly useful. We acknowledge
338 that the reliability of our results is fundamentally linked to the quality of the input data,
339 notably the WTD estimates. This is why we repeated the analysis with two different WTD
340 datasets and report our results as intervals derived from both. Future research could focus on
341 developing WTD datasets that are as precise and robust as possible.

342 Our results demonstrate that groundwater is of the same order of magnitude as aridity
343 (quantified by the moisture index) in regulating photosynthesis in forests functional types in
344 CONUS (relative importance of 48 to 101%), with similar importance in savannahs and
345 shrublands (30-58%), grasslands (22-42%), and croplands (15-32%) compared to the
346 moisture index. Our findings highlight the crucial role of groundwater in modulating
347 ecosystem productivity across biomes, particularly in a world where escalating human
348 groundwater extraction adds further pressure on these ecosystems¹.

349 **Methods**

350 **Data sets**

351 SIF, PAR, net radiation, air temperature, precipitation, elevation, and land cover data were
352 derived for the CONUS from global data sets. We obtained quality-checked SIF from the
353 Sentinel-5 Precursor mission funded by the European Space Agency (ESA)⁵⁰, which is
354 available at a grid resolution of $0.083^\circ \times 0.083^\circ$ every eight days, for the period between
355 March 2018 and March 2021. We excluded grid cells that did not have at least one high-
356 quality data point per month over the period of availability. PAR was calculated based on the
357 monthly surface solar radiation downwards data from ERA5-Land at a resolution of $\sim 9\text{ km}^{51}$,
358 with a scaling factor of 2.04 ($\mu\text{mol J}^{-1}$)⁷⁵. Net radiation was calculated using surface net solar
359 and net thermal radiation from monthly ERA5-Land⁵¹ and converted to mm/year dividing it
360 by the latent heat of vaporization (J kg^{-1}) as a function of air temperature, also from ERA5-
361 Land⁵². Precipitation is from version 2.8 of MSWEP⁵³, which has been shown to outperform
362 other products for hydrological applications⁷⁶. This is consistent with the large positive biases
363 of ERA5-Land precipitation in tropical regions⁵¹. Elevation is obtained from the Global
364 Multi-resolution Terrain Elevation Data 2010 provided by the U.S. Geological Survey. For
365 each variable, we calculated the long-term mean over the available temporal record in every
366 cell⁷⁷.

367 We obtained pixel-wise information about the dominant plant functional types (PFTs) from
368 the 300-m ESA CCI land cover map³⁸ (Supplementary Fig. 6). We classified the four major
369 vegetation types according to the land cover map for the year 2020 (v2.1.1). The analysis was
370 only conducted for the areas which experience no land cover change for the period between
371 1992 and 2020, based on the annual ESA CCI land cover maps. The water table depth (WTD)
372 was derived from an inverse modelling study that accounted for the coupling between soil
373 water balance and groundwater recharge and discharge and was benchmarked against satellite
374 observations of Leaf-Area Index (LAI)³⁴. A limitation of using these estimates is that the
375 model was run over the 2004-2014 decade, which differs from the record availability of SIF
376 (February 2018 to October 2021). However, since WTD is derived from a statistical model,
377 the exact years used when running the model are not highly significant. For natural and
378 undisturbed lands, WTD tends to show limited year-to-year fluctuations⁷⁸. Additionally, we
379 use the mean over a 10-year period, which should provide a robust estimate. It is difficult to
380 estimate the uncertainties caused by the choice of the WTD product as the errors relevant to
381 the WTD estimate are unavailable. Instead, we used a recent global WTD dataset estimated
382 by the GLOBGM v1.0 model for cross-validation³⁹. To compare the contributions of WTD to

383 SIF/PAR between groundwater-dependent ecosystems (GDEs) and non-GDEs, we used a
384 GDE map of the CONUS from a recent study⁷⁰. GDEs areas were defined as grid cells with a
385 GDE area density greater than 0.6. For consistency, most datasets were converted to a
386 common grid resolution of $0.083^\circ \times 0.083^\circ$ using bilinear interpolation. WTD³⁴ was
387 averaged from $0.0083^\circ \times 0.0083^\circ$ to $0.083^\circ \times 0.083^\circ$.

388 TROPOMI Solar-Induced Fluorescence (SIF) was chosen for this study due to its direct link
389 to plant physiological activity. SIF is a by-product of the photosynthetic process, making it a
390 robust proxy for plant activity^{37,79}. This contrasts with vegetation indices (e.g. Landsat),
391 which primarily reflect canopy structure rather than direct physiological processes⁸⁰.
392 Additionally, TROPOMI SIF provides the highest spatial resolution among SIF satellite
393 products, with a grid resolution of $0.083^\circ \times 0.083^\circ$ (approximately 10 km at mid-
394 latitudes)^{50,81}. While TROPOMI SIF does suffer from data gaps, particularly in mountainous
395 areas and deserts, the focus of our analysis on vegetated land mitigates this potential source
396 of bias. Therefore, vegetation indices from other satellite products with higher resolution
397 were not considered in our analysis due to their indirect relationship with plant physiology.
398

399 Explainable tree-based models

400 To test the relative importance of water table depth vs the moisture index in explaining the
401 variance of SIF/PAR, we trained XGBoost models using SIF/PAR as the target variable, with
402 $\lambda P/R_n$, WTD, and elevation as predictors. We trained one model per vegetation group, which
403 were defined in accordance with previous studies^{63,64}: forests (open and closed evergreen and
404 deciduous, needle-leaved and broadleaved, and mixed forests), shrublands and savannahs
405 (savannahs and woody savannahs, open and closed shrublands), grasslands, and croplands.
406 For the Spatial representation of Causal Shapley values (Fig. 4 and Supplementary Fig. 3) we
407 trained one model for all vegetation groups combined. This approach allows the direct
408 comparison of feature importances across the entire spatial area, thus avoiding the misleading
409 juxtaposition of Shapley values from different models with different baseline values in the
410 same map.

411 XGBoost models have recently been employed within the tree-based Shapley framework
412 across different scientific fields to quantify feature contribution^{63,66,82,83}. These models tend to
413 converge faster compared to neural networks, particularly in regression problems, where each
414 feature holds individual significance⁶⁶. The XGBoost algorithm uses an iterative decision tree
415 model composed of multiple decision trees^{84,85} and incorporates shrinkage to prevent
416 overfitting, and column subsampling to accelerate the training process⁸⁵. We focused on the

CONUS, where WTD data are constrained by more observations and are thus more reliable³⁴. The original global data sets were thus cropped between 24°N to 50°N and 125°W to 65°W. A smaller area also avoids training the model with data from very different environmental conditions within the same PFT, possibly reducing the explanatory capacity of the model. We used 80% data for model training and the remaining 20% for model performance testing. The XGBoost model hyperparameters were tuned using grid search and tenfold cross-validation and yield $R^2 \geq 0.69$ on the test set for all vegetation groups (Supplementary Fig. 7). Intermediary data and all code are available in our shared Zenodo repository (see the 'Data Availability' section).

Here, we applied the Causal Shapley framework⁶⁵ to calculate Shapley values, mean absolute Shapley values and dependence plots to disentangle how our predictors contribute to determining the outcome of SIF/PAR. The framework offers a more nuanced approach to understanding feature contributions to SIF/PAR by explicitly incorporating causal relationships between variables (Supplementary Fig. 10). Unlike the conventional SHAP framework, which assumes independence between features, causal Shapley values rely on Pearl's do-calculus to account for both direct and indirect effects of features on the model output. This is particularly useful in scenarios where features are correlated, as causal Shapley values can distinguish between the immediate contribution of a feature to the target variable (direct effect) and its mediated influence through other features (indirect effect). In other words, Causal Shapley values allow for clearer interpretations of feature relevance in systems where causal structures are known or hypothesized. This is achieved by conditioning on interventions rather than observations, capturing both causal effects and the dependencies between variables. As a result, causal Shapley values provide a more robust framework for feature attribution in complex models, improving the interpretability of machine learning models in a causally meaningful way.

In this study, we restricted our predictors to WTD, $\lambda P/R_n$, and elevation to analyse the relative effects of aridity and groundwater on photosynthesis, in line with our central hypothesis. WTD and $\lambda P/R_n$ are the primary variables of interest (Fig. 3), with results involving elevation provided in the supplementary material (Supplementary Fig. 8). Since we focus on long-term photosynthesis, other predictors (e.g., temperature) were excluded due to their strong correlation with the local climate over such timescales. Our objective was not to identify all possible predictors of SIF/PAR for perfect prediction but to focus on the main drivers. The XGB models demonstrate robust predictive capability ($R^2 \geq 0.69$ on the test set

450 for all vegetation groups), particularly given the inherent challenges in modelling ecosystem
451 processes.

452 **Data availability**

453 All data used in this study are openly available.

454 - TROPOMI SIF: <ftp://fluo.gps.caltech.edu/data/tropomi/gridded/SIF740/>

455 - Water table depth product from Ref.³⁴: [http://thredds-](http://thredds-gfnl.usc.es/thredds/catalog/GLOBALWTDFTP/annualmeans/catalog.html)

456 [gfnl.usc.es/thredds/catalog/GLOBALWTDFTP/annualmeans/catalog.html](http://thredds-gfnl.usc.es/thredds/catalog/GLOBALWTDFTP/annualmeans/catalog.html)

457 - GLOBGM v1.0 water table depth product: <https://github.com/UU-Hydro/GLOBGM>

458 - ERA5-Land 2-m air temperature, surface solar radiation downwards, surface net solar and

459 net thermal radiation net radiation: <https://cds.climate.copernicus.eu/cdsapp#!/home>

460 - MSWEP precipitation: <http://www.gloh2o.org/mswep/>

461 - Global terrain elevation data:

462 https://topotools.cr.usgs.gov/gmted_viewer/gmted2010_global_grids.php

463 - ESA CCI land cover map: <https://maps.elie.ucl.ac.be/CCI/viewer/download.php>

464 - Global GDE map: <https://zenodo.org/records/11062894>

465 **Code availability**

466 All analyses were performed using R Statistical Software⁸⁶. All intermediary data and code
467 supporting this study are available in the Zenodo Digital Repository:

468 <https://zenodo.org/doi/10.5281/zenodo.10498084> (Giardina et al. 2024).

469

470 **Acknowledgements**

471 The authors thank the providers of the data sets used in this study. We would also like to
472 acknowledge funding from the Swiss National Science Foundation (grant no.
473 PCEFP2_181115), as well as support from the LEMONTREE project (Land Ecosystem
474 Models based on New Theory, Observation, and Experiments), funded through the generosity
475 of Eric and Wendy Schmidt, as recommended by the Schmidt Futures program. Additionally,
476 we acknowledge the USMILE European Research Council Synergy Grant and the Learning
477 the Earth with Artificial Intelligence and Physics (LEAP) Science and Technology Center,
478 funded by the National Science Foundation (Award #2019625-STC).
479

480 **Author Contributions**

481 F.G. wrote the main manuscript in collaboration with P.G. F.G., P.G. and J.L. designed the
482 study. F.G. performed the analysis and prepared the figures in collaboration with J.L. F.G.,
483 P.G., S.I.S, J.L., and B.D.S reviewed and edited the manuscript.

484

485 **Competing Financial Interests**

486 The authors declare no competing financial interests.

487 **References**

- 488 1. Famiglietti, J. S. The global groundwater crisis. *Nature Climate Change* vol. 4 945–
489 948 Preprint at <https://doi.org/10.1038/nclimate2425> (2014).
- 490 2. Miguez-Macho, G. & Fan, Y. Spatiotemporal origin of soil water taken up by
491 vegetation. *Nature* **598**, 624–628 (2021).
- 492 3. Salvucci, G. D. & Entekhabi, D. Equivalent steady soil moisture profile and the time
493 compression approximation in water balance modeling. *Water Resour Res* **30**, 2737–
494 2749 (1994).
- 495 4. Seneviratne, S. I. *et al.* Weather and Climate Extreme Events in a Changing Climate.
496 in *Climate Change 2021: The Physical Science Basis. Contribution of Working Group
497 I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*
498 1513–1766 (Cambridge University Press, Cambridge, United Kingdom and New York,
499 NY, USA, 2021). doi:<https://doi.org/10.1017/9781009157896.013>.
- 500 5. Giardina, F., Gentile, P., Konings, A. G., Seneviratne, S. I. & Stocker, B. D.
501 Diagnosing evapotranspiration responses to water deficit across biomes using deep
502 learning. *New Phytologist* (2023) doi:10.1111/nph.19197.
- 503 6. Famiglietti, J. S. *et al.* Satellites measure recent rates of groundwater depletion in
504 California's Central Valley. *Geophys Res Lett* **38**, 2010GL046442 (2011).
- 505 7. Giardina, F. The mediation of terrestrial vegetation in water-carbon coupling: from
506 ecophysiology to land-atmosphere interactions. (2024) doi:10.3929/ETHZ-B-
507 000666415.
- 508 8. Oshun, J., Dietrich, W. E., Dawson, T. E. & Fung, I. Dynamic, structured
509 heterogeneity of water isotopes inside hillslopes. *Water Resour Res* **52**, 164–189
510 (2016).
- 511 9. Schwinnning, S. The ecohydrology of roots in rocks. *Ecohydrology* **3**, 238–245 (2010).
- 512 10. Fan, Y., Li, H. & Miguez-Macho, G. Global patterns of groundwater table depth.
513 *Science (1979)* **339**, 940–943 (2013).
- 514 11. Lee, J. E., Oliveira, R. S., Dawson, T. E. & Fung, I. Root functioning modifies
515 seasonal climate. *Proc Natl Acad Sci U S A* **102**, 17576–17581 (2005).
- 516 12. Seneviratne, S. I. *et al.* Investigating soil moisture-climate interactions in a changing
517 climate: A review. *Earth Sci Rev* **99**, 125–161 (2010).
- 518 13. Goedhart, C. M. & Pataki, D. E. Ecosystem effects of groundwater depth in Owens
519 Valley, California. *Ecohydrology* **4**, 458–468 (2011).
- 520 14. Fan, Y. Groundwater in the Earth's critical zone: Relevance to large-scale patterns and
521 processes. *Water Resour Res* **51**, 3052–3069 (2015).
- 522 15. Clark, M. P. *et al.* Improving the representation of hydrologic processes in Earth
523 System Models. *Water Resour Res* **51**, 5929–5956 (2015).
- 524 16. Maxwell, R. M. & Condon, L. E. Connections between groundwater flow and
525 transpiration partitioning. *Science (1979)* **353**, 377–380 (2016).
- 526 17. Swenson, S. C., Clark, M., Fan, Y., Lawrence, D. M. & Perket, J. Representing
527 Intrahillslope Lateral Subsurface Flow in the Community Land Model. *J Adv Model
528 Earth Syst* **11**, 4044–4065 (2019).
- 529 18. Weil, R. R. & Brady, N. C. *The Nature and Properties of Soils*, Global Edition. 10105
530 (2017).
- 531 19. Dawson, T. E. & Pate, J. S. Seasonal water uptake and movement in root systems of
532 Australian phreatophytic plants of dimorphic root morphology: A stable isotope
533 investigation. *Oecologia* **107**, 13–20 (1996).
- 534 20. Voltas, J., Lucabaugh, D., Chambel, M. R. & Ferrio, J. P. Intraspecific variation in the
535 use of water sources by the circum-Mediterranean conifer *Pinus halepensis*. *New
536 Phytologist* **208**, 1031–1041 (2015).

- 537 21. Grossiord, C. *et al.* Prolonged warming and drought modify belowground interactions
538 for water among coexisting plants. *Tree Physiol* **39**, 55–63 (2018).
- 539 22. Stocker, B. D. *et al.* Global patterns of water storage in the rooting zones of vegetation.
540 *Nat Geosci* **16**, 250–256 (2023).
- 541 23. Rempe, D. M. & Dietrich, W. E. Direct observations of rock moisture, a hidden
542 component of the hydrologic cycle. *Proc Natl Acad Sci U S A* **115**, 2664–2669 (2018).
- 543 24. McCormick, E. L. *et al.* Widespread woody plant use of water stored in bedrock.
544 *Nature* **597**, 225–229 (2021).
- 545 25. Querejeta, J. I., Estrada-Medina, H., Allen, M. F. & Jiménez-Osornio, J. J. Water
546 source partitioning among trees growing on shallow karst soils in a seasonally dry
547 tropical climate. *Oecologia* **152**, 26–36 (2007).
- 548 26. Evaristo, J. & McDonnell, J. J. Prevalence and magnitude of groundwater use by
549 vegetation: A global stable isotope meta-analysis. *Sci Rep* **7**, 1–12 (2017).
- 550 27. Barbeta, A. & Peñuelas, J. Relative contribution of groundwater to plant transpiration
551 estimated with stable isotopes. *Sci Rep* **7**, 1–10 (2017).
- 552 28. Jobbágy, E. G., Nosetto, M. D., Villagra, P. E. & Jackson, R. B. Water subsidies from
553 mountains to deserts: their role in sustaining groundwater-fed oases in a sandy
554 landscape. *Ecological Applications* **21**, 678–694 (2011).
- 555 29. Stocker, B. , *et al.* Towards a better understanding of deep belowground water stores
556 and their influence on land-atmosphere exchange and drought impacts. *EGU General
557 Assembly 2023 EGU23-9503*, (2023).
- 558 30. Tumber-Dávila, S. J., Schenk, H. J., Du, E. & Jackson, R. B. Plant sizes and shapes
559 above and belowground and their interactions with climate. *New Phytologist* **235**,
560 1032–1056 (2022).
- 561 31. Elmore, A. J., Manning, S. J., Mustard, J. F. & Craine, J. M. Decline in alkali meadow
562 vegetation cover in California: The effects of groundwater extraction and drought.
563 *Journal of Applied Ecology* **43**, 770–779 (2006).
- 564 32. Howard, J. & Merrifield, M. Mapping Groundwater Dependent Ecosystems in
565 California. *PLoS One* **5**, e11249 (2010).
- 566 33. Sutanto, S. J. *et al.* A perspective on isotope versus non-isotope approaches to
567 determine the contribution of transpiration to total evaporation. *Hydrol Earth Syst Sci*
568 **18**, 2815–2827 (2014).
- 569 34. Fan, Y., Miguez-Macho, G., Jobbágy, E. G., Jackson, R. B. & Otero-Casal, C.
570 Hydrologic regulation of plant rooting depth. *Proc Natl Acad Sci U S A* **114**, 10572–
571 10577 (2017).
- 572 35. Joiner, J. *et al.* Global monitoring of terrestrial chlorophyll fluorescence from
573 moderate-spectral-resolution near-infrared satellite measurements: methodology,
574 simulations, and application to GOME-2. *Atmos Meas Tech* **6**, 2803–2823 (2013).
- 575 36. Frankenberg, C. *et al.* New global observations of the terrestrial carbon cycle from
576 GOSAT: Patterns of plant fluorescence with gross primary productivity. *Geophys Res
577 Lett* **38**, 1–6 (2011).
- 578 37. Green, J. K. *et al.* Regionally strong feedbacks between the atmosphere and terrestrial
579 biosphere. *Nat Geosci* **10**, 410–414 (2017).
- 580 38. Harper, K. L. *et al.* A 29-year time series of annual 300m resolution plant-functional-
581 type maps for climate models. *Earth Syst Sci Data* **15**, 1465–1499 (2023).
- 582 39. Verkaik, J., Sutanudjaja, E. H., Oude Essink, G. H. P., Lin, H. X. & Bierkens, M. F. P.
583 GLOBGM v1.0: A parallel implementation of a 30arcsec PCR-GLOBWB-
584 MODFLOW global-scale groundwater model. *Geosci Model Dev* **17**, 275–300 (2024).
- 585 40. Budyko, M. I. *Climate and Life*. (Academic Press, 1974).

- 586 41. Budyko, M. I. *Тепловой Баланс Земной Поверхности (Heat Balance of the Earth's*
587 *Surface)*. (Gidrometeoizdat, Leningrad, 1956).
- 588 42. Condon, L. E. & Maxwell, R. M. Systematic shifts in Budyko relationships caused by
589 groundwater storage changes. *Hydrol Earth Syst Sci* **21**, 1117–1135 (2017).
- 590 43. Istanbulluoglu, E., Wang, T., Wright, O. M. & Lenters, J. D. Interpretation of
591 hydrologic trends from a water balance perspective: The role of groundwater storage in
592 the Budyko hypothesis. *Water Resour Res* **48**, (2012).
- 593 44. Greve, P., Gudmundsson, L., Orlowsky, B. & Seneviratne, S. I. Introducing a
594 probabilistic Budyko framework. *Geophys Res Lett* **42**, 2261–2269 (2015).
- 595 45. Crosbie, R. S., Holland, K. L. & McVicar, T. R. Regional-scale partitioning of
596 transmission losses and groundwater recharge using satellite estimates of actual
597 evapotranspiration in an arid environment. *Ecohydrology* **16**, e2490 (2023).
- 598 46. O'Grady, A. P., Carter, J. L. & Bruce, J. Can we predict groundwater discharge from
599 terrestrial ecosystems using existing eco-hydrological concepts? *Hydrol Earth Syst Sci*
600 **15**, 3731–3739 (2011).
- 601 47. Maes, W. H. *et al.* Sun-induced fluorescence closely linked to ecosystem transpiration
602 as evidenced by satellite data and radiative transfer models. *Remote Sens Environ* **249**,
603 112030 (2020).
- 604 48. Pagán, B., Maes, W., Gentine, P., Martens, B. & Miralles, D. Exploring the Potential
605 of Satellite Solar-Induced Fluorescence to Constrain Global Transpiration Estimates.
606 *Remote Sens (Basel)* **11**, 413 (2019).
- 607 49. Jonard, F. *et al.* Value of sun-induced chlorophyll fluorescence for quantifying
608 hydrological states and fluxes: Current status and challenges. *Agric For Meteorol* **291**,
609 108088 (2020).
- 610 50. Köhler, P. *et al.* Global Retrievals of Solar-Induced Chlorophyll Fluorescence With
611 TROPOMI: First Results and Intersensor Comparison to OCO-2. *Geophys Res Lett* **45**,
612 10,456-10,463 (2018).
- 613 51. Hersbach, H. *et al.* The ERA5 global reanalysis. *Quarterly Journal of the Royal
614 Meteorological Society* **146**, 1999–2049 (2020).
- 615 52. Pagán, B., Maes, W., Gentine, P., Martens, B. & Miralles, D. Exploring the Potential
616 of Satellite Solar-Induced Fluorescence to Constrain Global Transpiration Estimates.
617 *Remote Sens (Basel)* **11**, 413 (2019).
- 618 53. Beck, H. E. *et al.* MSWEP V2 Global 3-hourly 0.1° Precipitation: Methodology and
619 Quantitative Assessment. *Bull Am Meteorol Soc* **100**, 473–500 (2019).
- 620 54. Gentine, P., D'Odorico, P., Lintner, B. R., Sivandran, G. & Salvucci, G.
621 Interdependence of climate, soil, and vegetation as constrained by the Budyko curve.
622 *Geophys Res Lett* **39**, 2–7 (2012).
- 623 55. Gudmundsson, L., Greve, P. & Seneviratne, S. I. The sensitivity of water availability
624 to changes in the aridity index and other factors-A probabilistic analysis in the Budyko
625 space. *Geophys Res Lett* **43**, 6985–6994 (2016).
- 626 56. Lintner, B. R., Gentine, P., Findell, K. L. & Salvucci, G. D. The Budyko and
627 complementary relationships in an idealized model of large-scale land-atmosphere
628 coupling. *Hydrol Earth Syst Sci* **19**, 2119–2131 (2015).
- 629 57. Prentice, I. C. *et al.* Evidence of a universal scaling relationship for leaf CO₂
630 drawdown along an aridity gradient. *New Phytologist* **190**, 169–180 (2011).
- 631 58. Greve, P., Roderick, M. L. & Seneviratne, S. I. Simulated changes in aridity from the
632 last glacial maximum to 4xCO₂. *Environmental Research Letters* **12**, 114021 (2017).
- 633 59. Manabe, S. *Climate and Ocean Circulation. Part I: The Atmospheric Circulation and
634 the Hydrology of the Earth's Surface*. vol. 97 (Mon. Weather Rev., 1969).

- 635 60. Cutler, D. R. *et al.* Random forests for classification in ecology. *Ecology* **88**, 2783–
636 2792 (2007).
- 637 61. Hutengs, C. & Vohland, M. Downscaling land surface temperatures at regional scales
638 with random forest regression. *Remote Sens Environ* **178**, 127–141 (2016).
- 639 62. Green, J. K. *et al.* Surface temperatures reveal the patterns of vegetation water stress
640 and their environmental drivers across the tropical Americas. *Glob Chang Biol* **28**,
641 2940–2955 (2022).
- 642 63. Wang, H. *et al.* Exploring complex water stress–gross primary production
643 relationships: Impact of climatic drivers, main effects, and interactive effects. *Glob
644 Chang Biol* **28**, 4110–4123 (2022).
- 645 64. Fernández-Martínez, M. *et al.* The role of climate, foliar stoichiometry and plant
646 diversity on ecosystem carbon balance. *Glob Chang Biol* **26**, 7067–7078 (2020).
- 647 65. Heskes, T., Sijben, E., Bucur, I. G. & Claassen, T. Causal Shapley Values: Exploiting
648 Causal Knowledge to Explain Individual Predictions of Complex Models. *Adv Neural
649 Inf Process Syst* **2020-December**, (2020).
- 650 66. Lundberg, S. M. *et al.* From local explanations to global understanding with
651 explainable AI for trees. *Nat Mach Intell* **2**, 56–67 (2020).
- 652 67. Wei, D. *et al.* Elevation-dependent pattern of net CO₂ uptake across China. *Nature
653 Communications* **2024 15:1** **15**, 1–10 (2024).
- 654 68. Lundberg, S. M. & Lee, S.-I. A Unified Approach to Interpreting Model Predictions.
655 *Adv Neural Inf Process Syst* **30**, 4768–4777 (2017).
- 656 69. Xie, Y., Lark, T. J., Brown, J. F. & Gibbs, H. K. Mapping irrigated cropland extent
657 across the conterminous United States at 30 m resolution using a semi-automatic
658 training approach on Google Earth Engine. *ISPRS Journal of Photogrammetry and
659 Remote Sensing* **155**, 136–149 (2019).
- 660 70. Rohde, M. M. *et al.* Groundwater-dependent ecosystem map exposes global dryland
661 protection needs. *Nature* **2024 632:8023** **632**, 101–107 (2024).
- 662 71. Terazawa, K., Maruyama, Y. & Morikawa, Y. Photosynthetic and stomatal responses
663 of Larix kaempferi seedlings to short-term waterlogging. *Ecol Res* **7**, 193–197 (1992).
- 664 72. Terazawa, K. & Kikuzawa, K. Effects of flooding on leaf dynamics and other seedling
665 responses in flood-tolerant Alnus japonica and flood-intolerant Betula platyphylla var.
666 japonica. *Tree Physiol* **14**, 251–261 (1994).
- 667 73. Ohta, T. *et al.* Effects of waterlogging on water and carbon dioxide fluxes and
668 environmental variables in a Siberian larch forest, 1998–2011. *Agric For Meteorol* **188**,
669 64–75 (2014).
- 670 74. Fan, Y. *et al.* Hillslope Hydrology in Global Change Research and Earth System
671 Modeling. *Water Resour Res* **55**, 1737–1772 (2019).
- 672 75. Meek, D. W., Hatfield, J. L., Howell, T. A., Idso, S. B. & Reginato, R. J. A
673 Generalized Relationship between Photosynthetically Active Radiation and Solar
674 Radiation1. *Agron J* **76**, 939–945 (1984).
- 675 76. Tarek, M., Brissette, F. P. & Arsenault, R. Large-scale analysis of global gridded
676 precipitation and temperature datasets for climate change impact studies. *J
677 Hydrometeorol* **21**, 2623–2640 (2020).
- 678 77. GMTED2010 Global Grids.
679 https://topotools.cr.usgs.gov/gmted_viewer/gmted2010_global_grids.php.
- 680 78. Duan, L., Liu, T., Wang, X. & Luo, Y. Spatio-Temporal Patterns of Water Table and
681 Vegetation Status of a Deserted Area. *Water* **2015**, Vol. 7, Pages 5788–5805 **7**, 5788–
682 5805 (2015).
- 683 79. Giardina, F. *et al.* Tall Amazonian forests are less sensitive to precipitation variability.
684 *Nat Geosci* **11**, 405–409 (2018).

- 685 80. Zeng, Y. *et al.* Optical vegetation indices for monitoring terrestrial ecosystems
686 globally. *Nature Reviews Earth & Environment* 2022 3:7 3, 477–493 (2022).
- 687 81. Guanter, L. *et al.* The TROPOSIF global sun-induced fluorescence dataset from the
688 Sentinel-5P TROPOMI mission. *Earth Syst Sci Data* 13, 5423–5440 (2021).
- 689 82. Yang, C., Chen, M. & Yuan, Q. The application of XGBoost and SHAP to examining
690 the factors in freight truck-related crashes: An exploratory analysis. *Accid Anal Prev*
691 158, 106153 (2021).
- 692 83. Guo, M. *et al.* Older Pedestrian Traffic Crashes Severity Analysis Based on an
693 Emerging Machine Learning XGBoost. *Sustainability* 13, 926 (2021).
- 694 84. Yan, F., Song, K., Liu, Y., Chen, S. & Chen, J. Predictions and mechanism analyses of
695 the fatigue strength of steel based on machine learning. *J Mater Sci* 55, 15334–15349
696 (2020).
- 697 85. Chen, T. & Guestrin, C. XGBoost: A scalable tree boosting system. in *Proceedings of
698 the ACM SIGKDD International Conference on Knowledge Discovery and Data
699 Mining* vols 13-17-August-2016 785–794 (Association for Computing Machinery,
700 New York, NY, USA, 2016).
- 701 86. R Core Team. R: A language and environment for statistical computing. R Foundation
702 for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>. (2023).
- 703