

1 **Strong impact of groundwater on long-term photosynthesis in the**
2 **contiguous United States**

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

5

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

8 ²Department of Earth and Environmental Engineering, Columbia University, New York, New York
9 10027, USA

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

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

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

15

16 Authors for correspondence:

17 Francesco Giardina, Email: fgiardina@ethz.ch

18 Jiangong Liu, Email: jl6314@columbia.edu

19 **Abstract**

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

36 **Main text**

37 Groundwater is a critical component of terrestrial water storage¹, yet its impact on
38 photosynthesis remains poorly constrained. Fluctuations in water table levels exhibit distinct
39 spatiotemporal patterns compared to surface soil moisture^{2,3}, as they can be related to long-
40 range landscape topography, soil components, or both, in addition to local climate.
41 Understanding how groundwater modulates photosynthesis and its impact on the water and
42 energy fluxes is essential in the context of the rising occurrence of dry extremes over land⁴.
43 Water stored in the saturated zone can become indispensable during dry spells^{1,5,6}, which are
44 projected to become more frequent and intense under climate change⁴. Deep-rooted plants
45 can tap into water reserves in the saturated zone^{7–9} and then release water into the atmosphere
46 through leaf stomata, regulating atmospheric humidity and influencing regional climate^{10,11}.
47 Determining the control of groundwater on photosynthesis is thus key to accurately
48 predicting land-surface processes and their effects on climate^{12,13}. There has been a growing
49 effort to incorporate groundwater into Earth System Models in recent years^{13–16}, motivated in
50 part by evidence that water tables can influence vegetation over up to 32% of the Earth's
51 surface^{9,13}. However, to improve the representation of groundwater-surface interactions in
52 models, it is necessary to develop a solid empirical basis for detecting, quantifying, and
53 mapping the control of groundwater on plant physiological processes and ecosystem carbon
54 and water fluxes.

55 Systematically assessing the long-term contribution of groundwater to photosynthesis
56 poses significant challenges due to data sampling limitations and the representativity of the
57 in-situ data used, given the complex nature of groundwater across diverse landscapes. Recent
58 research based on inverse modelling has shown that, globally and at the annual scale, the
59 primary source of plant-available water is recent precipitation stored in shallow soil layers
60 (e.g., the top 30 cm)^{2,17}. At the same time, other studies have documented extensive plant
61 water uptake from deeper soil^{18–21}, rock moisture^{22–24} (water stored in bedrock above the
62 water table)^{21–23} or groundwater^{25–27} (water in saturated subsurface zones, whether in shallow
63 alluvial deposits or deeper bedrock aquifers). Global syntheses of stable isotope
64 measurements have revealed that groundwater use is widespread across various biomes, with
65 an increased relative contribution in arid or seasonally dry regions^{25,26}. Yet, the presence of
66 deep-rooted plants worldwide suggests that groundwater usage may extend beyond those
67 sparse sampled sites²⁸.

68 Investigations into the uptake of groundwater by plants have thus either focused
69 regionally on syntheses of measurements from stable isotope techniques^{2,25,26}, and at larger

70 scales, they have relied on indirect methods such as geospatial data-based techniques^{29,30},
71 land-surface model^{14,31} and inverse modelling approaches^{2,9,32}. However, there is a lack of
72 research at the multi-year scale that incorporates consistent large-scale data while at the same
73 time accounting for plant physiologies and aridity gradients. Using a remotely sensed dataset
74 of solar-induced fluorescence (SIF) as a proxy for photosynthesis^{33–35}, coupled with maps of
75 plant functional types (PFTs)³⁶ and water table depth (WTD) estimates^{32,37}, we evaluate the
76 contribution of groundwater to long-term photosynthesis with a specific focus on the
77 contiguous United States (CONUS). SIF is nearly proportional with gross primary production
78 (GPP) over large temporal and spatial scales^{34,35}. Applying causality-guided explainable
79 machine learning, we show that groundwater plays a role essentially equivalent to aridity in
80 regulating photosynthesis across PFTs in the CONUS.

81

82 **Conceptual framework: groundwater as a regulator of the influence of 83 aridity on long-term photosynthesis**

84 An important contribution to the conceptual understanding of regional dryness
85 patterns is the Budyko relationship, which can also illuminate the role of long-term water
86 stress on evapotranspiration and photosynthesis^{38–40}. We present it here in one of its
87 formulations most aligned with our study^{39,40}:

88

$$89 \frac{ET}{P} = f\left(\frac{R_n}{\lambda P}\right) \quad (1)$$

90

91 Where ET is the actual evapotranspiration (typically in mm/year), R_n is net radiation
92 (W/m²) and P is precipitation (mm/year) multiplied by λ, the latent heat of evaporation
93 (J/Kg)^{39,40}. The relationship outlines the partitioning of water across an aridity gradient, from
94 water-limited to energy-limited regimes. It conceptualizes the fact that in drier climates, a
95 larger fraction of the available energy is partitioned into evapotranspiration per unit
96 precipitation^{39,40} (Fig. 1). However, a major limitation of the traditional Budyko framework is
97 that it assumes that the main regulator of ET on long-term equilibrium is local climate^{41,42}
98 and it does not explicitly account for the effect of long-term water storage changes by e.g.
99 groundwater, despite its significant role in the water balance. Recent studies have shown how
100 groundwater dynamics can alter this relationship^{41–45}. The inclusion of groundwater in the
101 Budyko framework influences its shape^{41,42}. In particular, a positive groundwater contribution
102 leads to higher evapotranspiration for a given aridity index (or "radiative index of dryness",

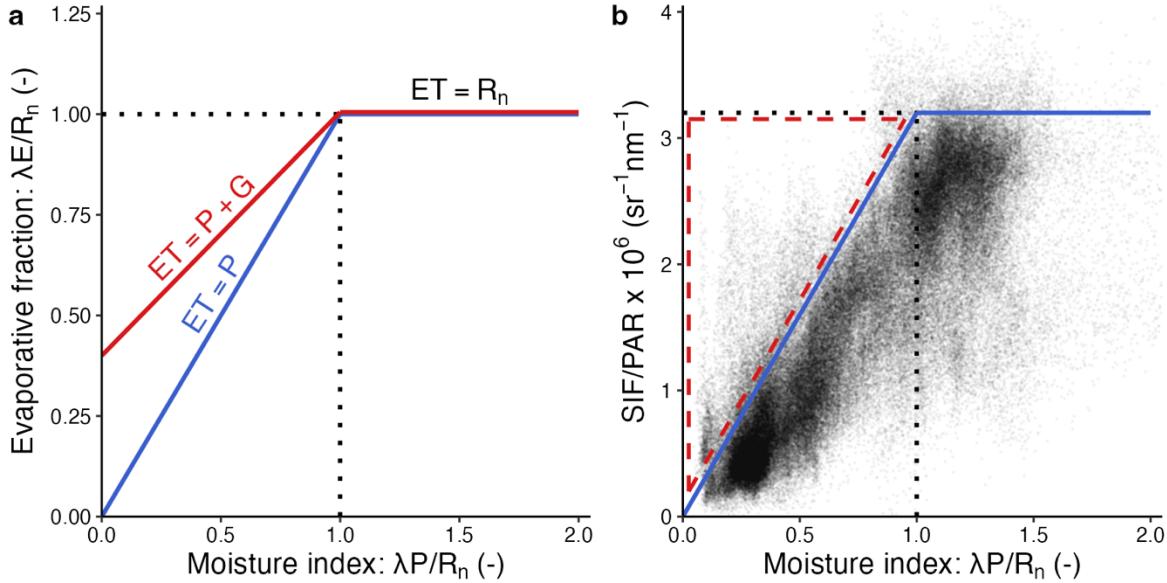
103 as referred to by Budyko), thus resulting in points above the Budyko curve in its water-
104 limited region^{41,42} (i.e. $\lambda P/R_n < 1$ in Fig. 1a). This indicates that groundwater access supports
105 additional evapotranspiration through plant transpiration or direct soil evaporation, enhancing
106 ET beyond the constraints of the local water balance represented by the Budyko curve—
107 which assumes precipitation is the only water input^{41,42}. We here use a similar conceptual
108 approach but for photosynthesis³⁸, as photosynthesis and ET are tightly coupled via the water
109 use efficiency, which in turn varies across vegetation communities and species^{46–48}.

110 In our study, we want first to estimate the extent of groundwater use across an aridity
111 gradient. Given the near-linear relationship between SIF and GPP^{34,35} at long time scales, we
112 use SIF based on the high-resolution global TROPOMI sensor⁴⁹ as a proxy for ecosystem-
113 level photosynthesis. To assess variations in SIF beyond climate forcings, we normalize SIF
114 by photosynthetically active radiation (PAR) from ERA5-Land⁵⁰ (SIF/PAR, see Methods).
115 SIF has been shown to accurately quantify the impact of environmental stress on ecosystem
116 transpiration (T)^{48,51} and shows a closer relation to T than any other space-based
117 measurements⁴⁶. We thus use SIF/PAR as an observational proxy of the evaporative fraction
118 (defined as the ratio of the latent heat flux to net radiation) on the y-axis in the Budyko
119 framework (Fig. 1b). We define aridity as the ratio of MSWEP precipitation⁵² (P) to ERA5-
120 Land net radiation⁵⁰ (R_n), denoted as $\lambda P/R_n$, where λ is the latent heat of vaporization ($J\ kg^{-1}$)
121 for units consistency. This metric provides a simple yet effective way to characterize the
122 water availability of a region in relation to the available energy and, since P and R_n are based
123 on long-term averages, it is an index of area-specific mean climate conditions^{41,42,53–55}. Given
124 that higher $\lambda P/R_n$ values indicate lower aridity (the inverse of Budyko's formulation) and
125 more water availability, we adopt the term 'moisture index' rather than 'aridity index' to avoid
126 confusion, in line with previous studies^{6,56}. Note that, while commonly used in
127 hydroclimatological studies^{41,42,53–55}, the index is just one of the possible formulations of
128 aridity⁵⁷.

129 As large-scale remote sensing datasets do not provide the exact variables used by
130 Budyko (e.g. ET), we describe our approach as 'Budyko-like', indicating an adaptation of the
131 original framework rather than a direct application (Fig. 1). This is similar to a moisture
132 limitation model previously formulated by Manabe based on Budyko ("bucket model")^{11,58},
133 but on long-time scales. In our Budyko-like framework, we expect that a positive
134 contribution of groundwater results in points higher than the Budyko curve in the water-
135 limited regime (Fig. 1a, red versus blue line). We identified pixels deviating from the
136 moisture index as those potentially located in zones where plants might access groundwater.

137 Note that this figure is not a quantitative assessment of groundwater access; rather, it offers a
 138 qualitative overview of the issue to contextualize our study and motivates the subsequent
 139 sections.

140



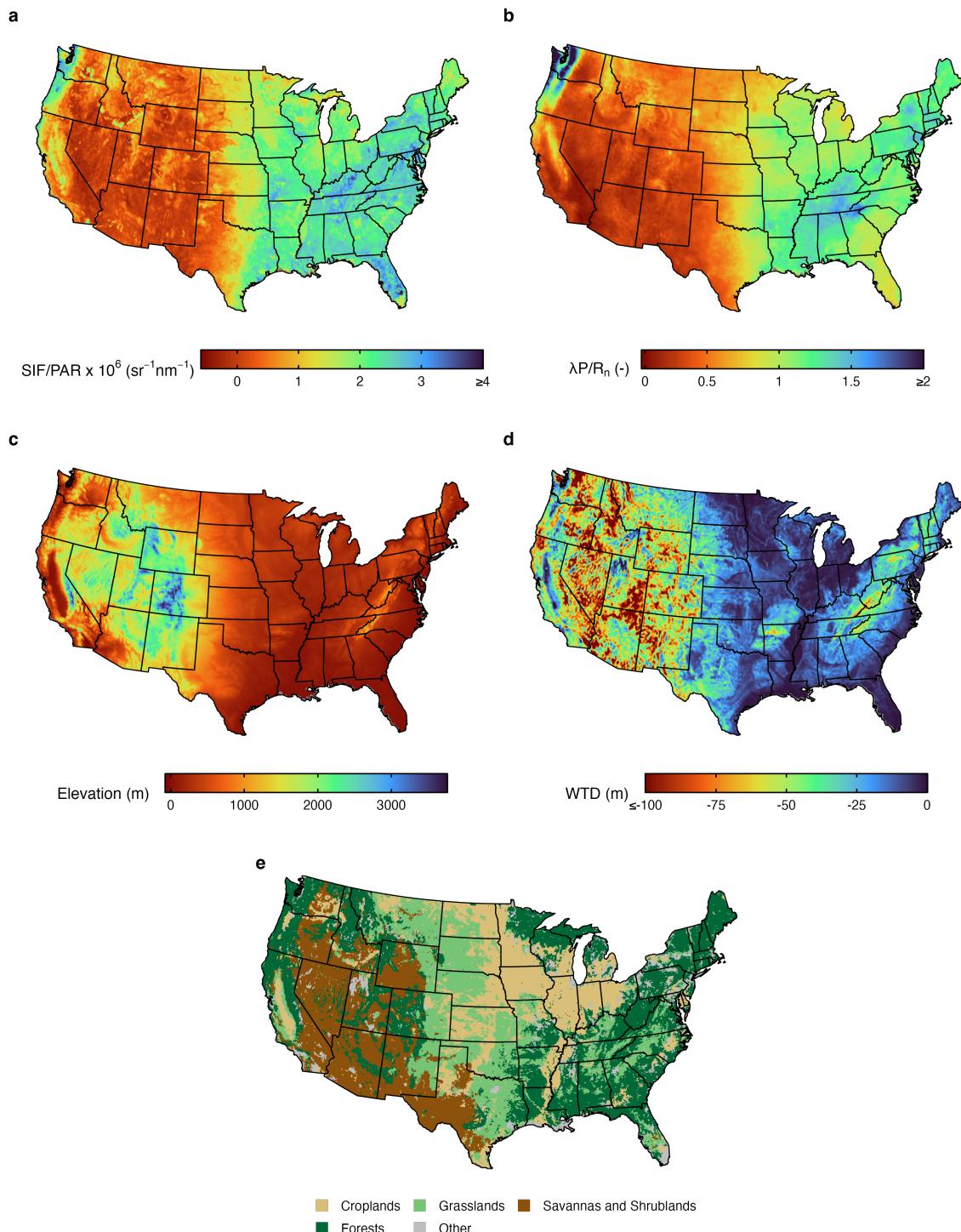
141
 142 **Fig. 1 | Groundwater in the Budyko-like framework.** The x-axes represent the moisture index
 143 ($\lambda P/R_n$), where higher values indicate increasingly humid conditions. **a**, Conceptual Budyko-like
 144 framework. The area between the blue and red lines in the water-limited regime ($\lambda P/R_n < 1$) indicates
 145 groundwater access due to ET values exceeding expected limits from the moisture index⁴¹. ET:
 146 evaporation, P: precipitation, G: groundwater. R_n : net radiation. λ : latent heat of vaporization. **b**,
 147 SIF/PAR distribution in the Budyko-like space, with each point representing a pixel across the
 148 CONUS. SIF/PAR serves as an observational proxy for the evaporative fraction. Continuous blue
 149 segments illustrate the Budyko-like framework as per framework in **a**. Points within the red triangle
 150 indicate areas where plants potentially access groundwater.

151

152 Local contributions of groundwater table and moisture index

153 To quantitatively study the relative contributions of groundwater and aridity on
 154 photosynthesis, we trained Extreme Gradient Boosting (XGBoost) models using long-term
 155 means of SIF/PAR (Fig. 2a) as the target variable with WTD, $\lambda P/R_n$, and elevation as
 156 predictors (Fig. 2b–d; see Methods). Thanks to their ability to encode non-linear
 157 relationships, machine learning models based on decision trees, such as XGBoost, have
 158 gained increasing popularity in ecological applications^{59–61}. Consistent with previous
 159 studies^{62,63}, we aggregated PFTs into four groups: forests, shrublands and savannahs,
 160 grasslands, and croplands, and trained one XGBoost model per group (Fig. 2e, see Methods).
 161 Since TROPOMI was launched in 2018, we used its three-year SIF record (March 2018–
 162 March 2021) to define the analysis period for all variables. Over this interval, we then
 163 calculated the mean values of SIF, WTD, $\lambda P/R_n$, and elevation at each 0.083° pixel. Our

models show good predictive performance for the CONUS, with $R^2 \geq 0.69$ on the test set for all vegetation groups ($R^2=0.83$ for forests, $R^2=0.69$ for savannahs and shrublands, $R^2=0.94$ for grasslands and $R^2=0.84$ for croplands, Supplementary Fig. 7). This demonstrates that the XGBoost models are effective in predicting long-term SIF/PAR. To assess the individual contributions of WTD and $\lambda P/R_n$ to the sensitivity of SIF/PAR, we calculated causal Shapley values⁶⁴ for each XGBoost model, i.e. for each vegetation group (Fig. 2e).



170

171 **Fig. 2 | Spatial patterns of long-term means of climatic and environmental variables in the**
172 **contiguous United States.** **a**, Solar-induced fluorescence to photosynthetically active radiation ratio
173 (SIF/PAR). **b**, Moisture index represented as the ratio of precipitation to net radiation ($\lambda P/R_n$). **c**,
174 Elevation above sea level. **d**, Water-table depth (WTD) from Refs^{9,32}. **e**, Classification of vegetation
175 types.

176

177 Machine learning models are based on nonlinear algorithms that typically yield higher
178 accuracy than linear models, especially on large complex datasets. Yet, they often lack
179 interpretability ('Black box' model predictions⁶⁵). Shapley values have been used in ecology
180 for interpreting model predictions⁶². They are based on an algorithm rooted in game theory
181 and describe the influence of each input feature (or predictor, i.e. WTD, $\lambda P/R_n$ and elevation)
182 on a specific prediction (i.e. SIF/PAR value predicted by the XGBoost models). In our
183 analysis, it is essential to include elevation as a predictor due to its control over long-term
184 photosynthesis⁶⁶. However, elevation acts as a confounder because it also influences WTD
185 and the moisture index¹³. Traditional Shapley values assume that predictors are independent,
186 but this assumption is violated in our case due to the causal relationships between elevation,
187 WTD, and $\lambda P/R_n$ (Supplementary Fig. 1). To account for this, we use causal Shapley
188 Values⁶⁴, which incorporate causal chain diagrams informed by expert knowledge, to better
189 constrain the relative contributions of WTD and $\lambda P/R_n$ to SIF/PAR while factoring out the
190 confounding effect of elevation (Supplementary Fig. 10).

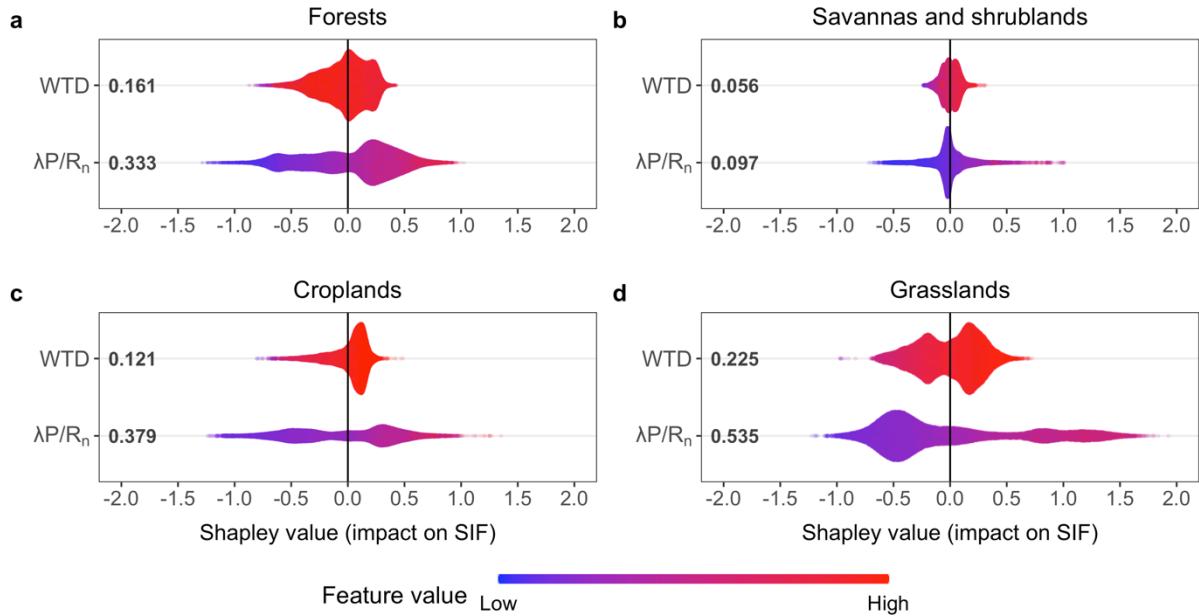
191 Shapley values directly quantify local (i.e. for a single data point) effects of individual
192 predictors, while at the same time allowing to account for the global structure of the model,
193 thus unveiling meaningful patterns that might otherwise remain undetected^{65,67}. More positive
194 Shapley values indicate that a particular predictor leads to an increase in the predicted target
195 variable (i.e. SIF/PAR) above the average outcome predicted by the XGBoost model (i.e. the
196 baseline value). A more negative Shapley value denotes the opposite. Shapley values that
197 approach zero thus indicate that SIF/PAR is nearing its baseline value. This, however, should
198 not be interpreted as lack of sensitivity with the corresponding predictor.

199 Because traditional Shapley values rely on an independence assumption that does not
200 hold in our case, we use the causal Shapley approach, which explicitly incorporates causal
201 relationships among variables to provide a more accurate assessment of feature
202 contributions⁶⁴ (see Methods and Supplementary Fig. 10). The absolute Shapley value of a
203 predictor represents its specific effect on an individual model prediction (local contribution),
204 whereas averaging these values across all data samples of a specific predictor yields a
205 measure of the global importance of that predictor (average of absolute Shapley values)⁶⁴.

We compute the average of absolute causal Shapley values for WTD and $\lambda P/R_n$ across all pixels in CONUS (Fig. 3, numbers shown on the left of each plot). Our analysis indicates that in forested areas, the relative importance of groundwater (average of absolute 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 index (average of absolute Shapley values of $0.333 \times 10^{-6} \text{ sr}^{-1} \text{ nm}^{-1}$, Fig. 3a) in determining ecosystem productivity. In savannahs and shrublands, the relative importance of groundwater compared to the moisture index is 58% (average of absolute Shapley values of 0.056 and $0.097 \times 10^{-6} \text{ sr}^{-1} \text{ nm}^{-1}$, respectively, Fig. 3b) decreasing to 42% in grasslands (average of absolute Shapley values of 0.225 and $0.535 \times 10^{-6} \text{ sr}^{-1} \text{ nm}^{-1}$, respectively, Fig. 3d), and to 32% in croplands (average of absolute Shapley values of 0.121 and $0.379 \times 10^{-6} \text{ sr}^{-1} \text{ nm}^{-1}$, Fig. 3c). This outcome is consistent with the fact that forests, savannahs and shrublands typically have deeper-rooted and longer-lived plants compared to other PFTs, allowing them to access water sources at greater depths and sustain productivity over longer periods^{28,32}. As a result, their productivity is more affected by changes in WTD.

We recognize that the accuracy of our findings depends on the quality of the input data, and notably the WTD estimates. To have a sense of the uncertainty in our approach, we repeated the analysis using an alternative WTD dataset³⁷. Despite the limited agreement between the two modelled WTD datasets (Supplementary Figs 11 and 4b, Fig. 2), our results with the alternative dataset confirm that WTD exerts a strong control on SIF/PAR (Supplementary Fig. 5). Specifically, the relative importance of WTD versus the moisture index using this alternative WTD dataset is 101% for forests, 30% for savannahs and shrublands, 15% for croplands, and 22% for grasslands. These values should be considered as indicators of the general control exerted by WTD relative to aridity on photosynthesis, demonstrating for the first time that both factors operate within the same order of magnitude. To account for the uncertainty in estimating the relative importance of groundwater compared to aridity, we present our results as intervals derived from both WTD datasets: 48–101% for forests, 30–58% for savannahs and shrublands, 22–42% for grasslands, and 15–32% for croplands.

234



235

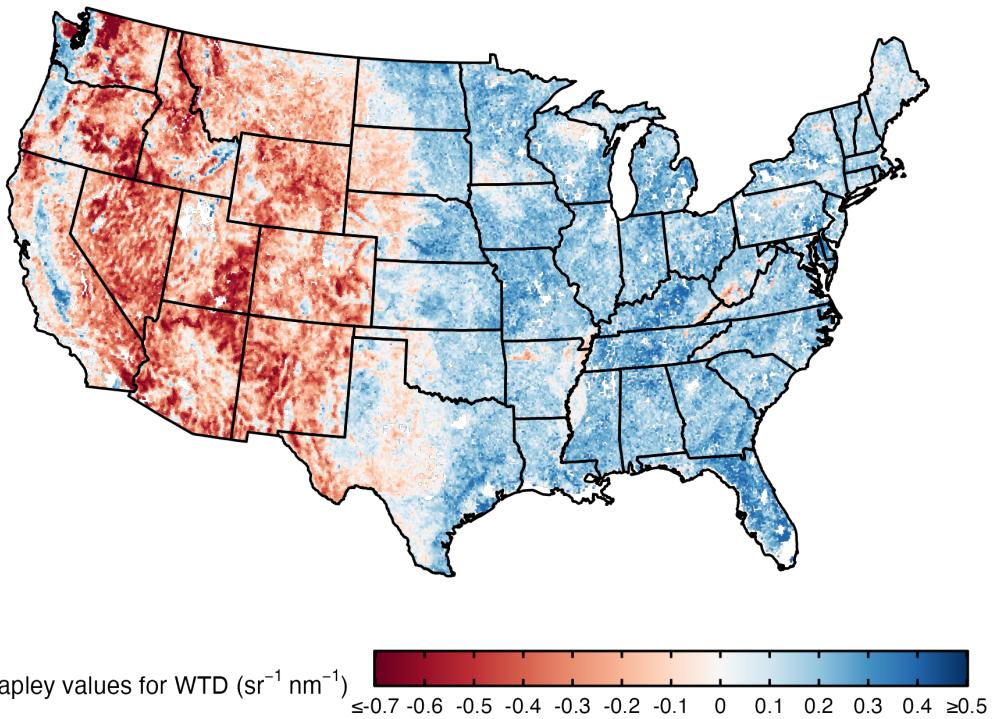
Fig. 3 | Causal Shapley values of ecosystem productivity drivers, based on gradient boosting decision tree models trained on a dataset of solar-induced fluorescence. **a-d**, Shapley summary plots, which show the effect of different predictors on each model outcome. Shapley values are in the same units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

248

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)

262 and in general the area of the Rocky Mountains show a negative effect of WTD on
263 photosynthesis (red areas in Fig. 4). The spatial distribution of the causal Shapley values for
264 $\lambda P/R_n$ (Supplementary Fig. 3) aligns with the long-term means of $\lambda P/R_n$ (Fig. 2b).



265
266 **Fig. 4 | Spatial distribution of causal Shapley values of Water-table depth across the contiguous**
267 **United States.** Shapley values illustrate the impact of Water-table depth (WTD) on SIF/PAR. In this
268 figure, the calculation of the causal Shapley values is based on a single XGBoost model that includes
269 all vegetation groups combined (see Methods).

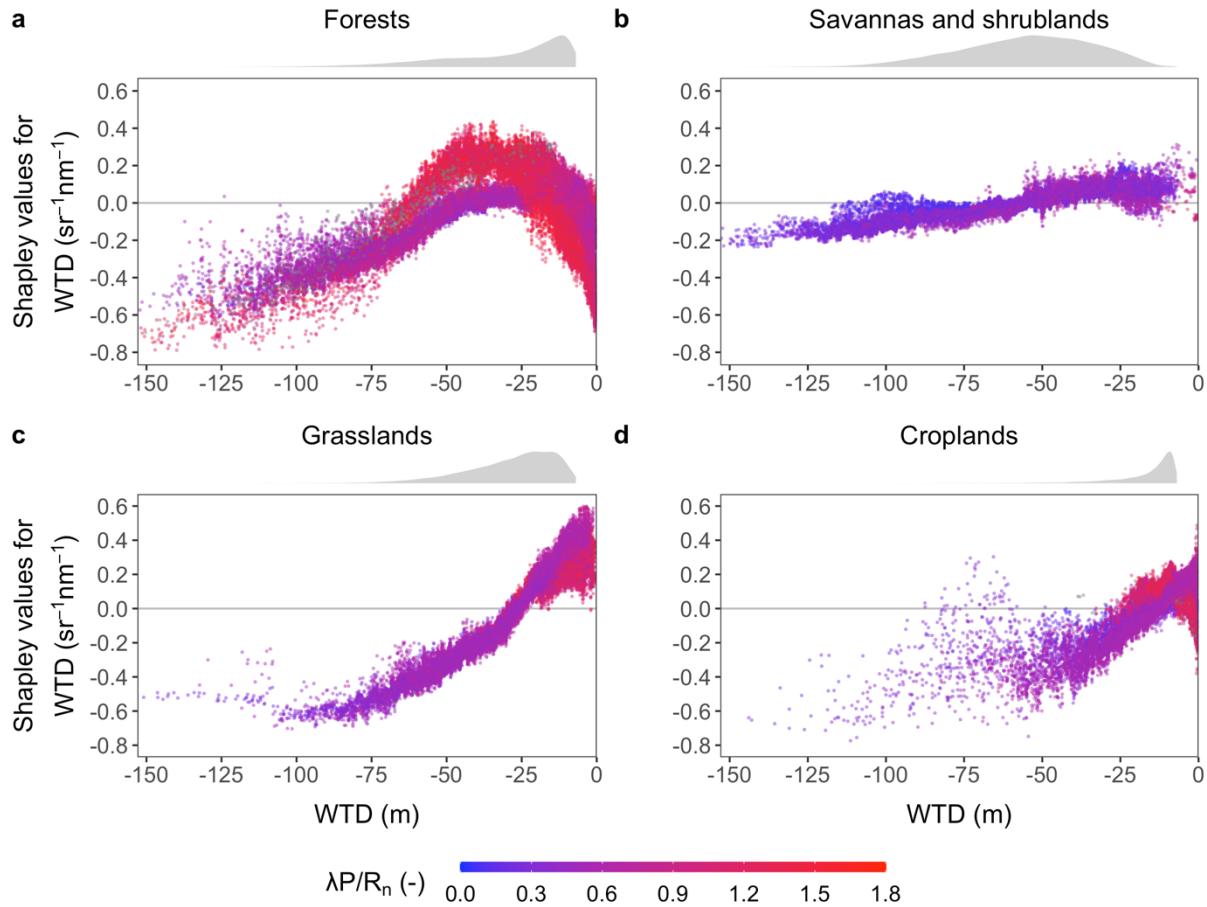
270
271 To further support our findings, we compared the absolute Shapley values of WTD
272 between groundwater-dependent ecosystems (GDEs) and non-GDE pixels, using a recently
273 published map of GDEs⁶⁸ (Supplementary Fig. 4c). We use absolute Shapley values to
274 capture the magnitude of their impact, irrespective of their sign. Our results consistently show
275 that the distribution of absolute Shapley values for WTD at GDE pixels is significantly
276 greater than at non-GDE pixels (Supplementary Fig. 9).

277

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

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

283



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

294 The dependence plot of forests reveals a range of intermediate WTD values associated
295 with positive Shapley values, indicative of stress-free and positive conditions for vegetation
296 photosynthesis (Fig. 5a). Deviating from this range, either towards higher or lower WTD
297 values, respectively corresponds to water-logging or highly arid conditions, resulting in
298 vegetation stress, as evidenced by the corresponding negative Shapley values. These results
299 are consistent with the negative effects of perched water tables on photosynthesis observed
300 under water-logging conditions, which leads to poor oxygenation of the roots^{69,70}. Laboratory
301 experiments have shown that GPP and stomatal conductance exhibit negative responses
302 shortly after the onset of water-logging conditions^{69,70}. Moreover, long-term observations in
303 boreal forests have demonstrated that persistent water-logging conditions reduce surface

conductance, evapotranspiration (ET), and GPP⁷¹. 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^{9,11,32}. 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) reflects their greater adaptation to water-logging conditions⁷². Only the negative effects resulting from very low WTD values are evident (i.e. left tail of Fig. 5c), a behaviour that is also observed in Savannahs and Shrublands (Fig. 5b). 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. 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.

328

329 Conclusion

330 In this study, we focused on disentangling the effects of aridity versus groundwater on
331 photosynthesis. A key point in our analysis is the causal Shapley values approach, which
332 accounts for the causal relationships between predictors, thus avoiding the confounding
333 effects of cross-correlation. The high R^2 of our XGBoost models ($R^2 \geq 0.69$ on the test set for
334 all vegetation groups, Supplementary Fig. 7) contributes to the robustness of our explainable
335 machine-learning approach. In line with a principle of simplicity, we focused on long-term
336 means, reflecting the perspective of the long-term partitioning of water and energy of the
337 Budyko-like framework (Fig. 1). Working with long-term means also has the advantage of

338 avoiding dealing with interannual variability, which can further complicate the relationship of
339 photosynthesis with groundwater. Groundwater and aridity modulate the long-term mean of
340 SIF, but it is challenging to quantify this effect with a physical model. In this context, our
341 data-driven explainable machine-learning approach proves to be particularly useful. We
342 acknowledge that the reliability of our results is fundamentally linked to the quality of the
343 input data, notably the WTD estimates. We thus repeated the analysis with two WTD datasets
344 and report our results as intervals derived from both. Future research could focus on
345 continuing to develop WTD datasets that are as precise and robust as possible.

346 Our results demonstrate that groundwater is of the same order of magnitude as aridity
347 (quantified by the moisture index) in regulating photosynthesis in forests functional types in
348 CONUS (relative importance of 48 to 101%), with similar importance in savannahs and
349 shrublands (30-58%), grasslands (22-42%), and croplands (15-32%) compared to the
350 moisture index. Our findings highlight the crucial role of groundwater in modulating
351 ecosystem productivity across vegetation types, particularly in a world where increasing
352 groundwater extraction adds further pressure on these biomes¹.

353 **Methods**

354 **Data sets**

355 We obtain Level 3 quality-checked SIF from the Sentinel-5 Precursor mission funded by the
356 European Space Agency (ESA)⁴⁹, which is available at a grid resolution of $0.083^\circ \times 0.083^\circ$ in
357 eight-day composites between March 2018 and March 2021. The Level 3 product aggregates
358 the native TROPOMI footprints ($\sim 7 \text{ km} \times 3.5 \text{ km}$, or $\sim 0.063^\circ \times 0.031^\circ$ at nadir). This
359 temporal compositing and spatial averaging reduce cloud gaps and sensor noise while
360 ensuring uniform data density across grid cells⁴⁹. We exclude grid cells that do not have at
361 least one high-quality data point per month over the period of availability. We use WTD
362 estimates from an inverse modelling study³² and repeat the analysis using a WTD dataset
363 from a process-based hydrological model³⁷. PAR is calculated based on the monthly surface
364 solar radiation downwards data from ERA5-Land at a resolution of $\sim 9 \text{ km}$ ⁵⁰, with a scaling
365 factor of 2.04 ($\mu\text{mol J}^{-1}$)⁷³. Net radiation is computed using surface net solar and net thermal
366 radiation from monthly ERA5-Land⁵⁰ and converted to mm/year dividing it by the latent heat
367 of vaporization (J kg^{-1}) as a function of air temperature, also from ERA5-Land⁵¹.
368 Precipitation is from version 2.8 of MSWEP⁵², which has been shown to outperform other
369 products for hydrological applications within the CONUS^{74,75}. Elevation is obtained from the
370 Global Multi-resolution Terrain Elevation Data 2010 provided by the U.S. Geological
371 Survey. For each variable, we calculate the long-term mean over the available temporal
372 record in every cell⁷⁶.

373 We obtain pixel-wise information about the dominant plant functional types (PFTs) from the
374 300-m ESA CCI land cover map³⁶ (Supplementary Fig. 6). We classify the four major
375 vegetation types for the year 2020 (v2.1.1). The analysis is only conducted for the areas
376 which experience no land cover change for the period between 1992 and 2020, based on the
377 annual ESA CCI land cover maps. We only selected rainfed croplands based on the ESA CCI
378 land cover map³⁶ (Fig. S6).

379 To compare the contributions of WTD to SIF/PAR between groundwater-dependent
380 ecosystems (GDEs) and non-GDEs, we use a GDE map of the CONUS from a recent study⁶⁸.
381 GDEs areas are defined as grid cells with a GDE area density greater than 0.6.
382 For consistency, we use bilinear interpolation to convert most datasets to the common 0.083°
383 $\times 0.083^\circ$ grid, matching the SIF resolution. Since WTD is available at a finer resolution
384 ($0.0083^\circ \times 0.0083^\circ$), we average it to $0.083^\circ \times 0.083^\circ$.

385

386 **Selection of datasets and predictors**

387 WTD is from a mechanistic inverse modelling study that integrates soil moisture, dynamic
388 plant root-uptake for transpiration, groundwater flow, river–floodplain interactions and
389 satellite-based leaf area index (LAI) data at hourly time steps, all on a sub-kilometer grid³².
390 Water table observations help constrain a key parameter (the rate at which porosity and
391 permeability decline with depth), so the model aligns well with global-scale hydrologic
392 patterns, including basic water-balance components such as evapotranspiration and river
393 discharge. Monthly water table depths are outputted from the model for 2003 to 2013 and
394 then averaged over that entire period to obtain a mean seasonal cycle³². Although this record
395 does not overlap with the SIF dataset (February 2018 to October 2021), it remains a reliable
396 basis for our analysis because, in most natural and undisturbed ecosystems (i.e. not heavily
397 pumped or drained), the seasonal cycle of water-table levels shows only modest inter-annual
398 variability⁷⁷. This assumption may not fully hold in semi-arid regions, introducing potential
399 uncertainty; nevertheless, calculating a decadal mean (2003–2013) provides a robust
400 benchmark for comparison with longer-term means of other variables. While it is difficult to
401 precisely quantify the uncertainty related to this WTD dataset, we have checked the
402 consistency of our results by repeating the analysis with a recently published WTD product
403 (GLOBGM v1.0)³⁷, which confirms our findings.

404 SIF was chosen as proxy of photosynthesis because it captures the light emitted during the
405 photosynthetic process, providing a direct measure of plant physiological activity and thus a
406 reliable estimate of both photosynthetic rate and physiological stress^{35,78}. In contrast,
407 traditional vegetation indices such as NDVI—which is often used to derive Leaf Area Index
408 (LAI)—depend on reflectance measurements and offer only indirect information about
409 canopy structure and general vegetation health⁷⁹. Even when averaged over multi-year
410 periods, SIF still retains short-term physiological variability that LAI and other structural
411 indices cannot capture, because those indices saturate at high biomass and respond more
412 slowly to environmental stress^{80–82}. SIF has been shown to outperform traditional vegetation
413 indices (like NDVI or LAI) as a proxy for GPP across all temporal scales, owing to its ability
414 to capture high-frequency fluctuations in chlorophyll fluorescence⁸¹. Among current SIF
415 products available contiguously in space, TROPOMI SIF offers the finest spatial
416 resolution^{49,83}, which is sufficient to disentangle the large-scale impacts of aridity and
417 groundwater on photosynthesis. Additionally, reflectance-based vegetation indices, while
418 available at higher spatial resolution, suffer from cloud-induced temporal aliasing⁸⁴. SIF,
419 retrieved in the red or far-red O₂ bands where thin-cloud scattering is weak, passes the cloud

420 screen more frequently^{85,86}. Therefore, each coarser-grid SIF cell (e.g. from TROPOMI)
421 receives far more valid clear-sky observations than an equivalent Landsat composite, greatly
422 reducing aliasing^{83,84}. Although TROPOMI SIF has spatial gaps over mountains and deserts,
423 our analysis is restricted to vegetated areas, so the impact is negligible^{49,86}. In summary,
424 reflectance-based vegetation indices at higher resolution were not considered in our analysis
425 due to their indirect relationship with plant physiology and susceptibility to cloud
426 contamination, which would have introduced additional uncertainty without offsetting
427 benefits.

428 In this study, we restricted our predictors to WTD, $\lambda P/R_n$, and elevation to analyse the
429 relative effects of aridity and groundwater on photosynthesis, in line with our central
430 hypothesis. WTD and $\lambda P/R_n$ are our primary variables of interest (Fig. 3), and results relative
431 to elevation are provided in the supplementary material (Supplementary Fig. 8). Since we
432 focus on long-term photosynthesis, other predictors (e.g. temperature, relative humidity) were
433 excluded due to their strong correlation with the local climate over such timescales⁸⁷. Our
434 objective was not to identify all possible predictors of SIF/PAR for perfect prediction but to
435 focus on the main drivers. The XGB models demonstrate robust predictive capability ($R^2 \geq$
436 0.69 on the test set for all vegetation groups), particularly given the inherent challenges in
437 modelling ecosystem processes.

438

439 **Explainable tree-based models**

440 To test the relative importance of water table depth vs the moisture index in explaining the
441 variance of SIF/PAR, we trained XGBoost models using SIF/PAR as the target variable, with
442 $\lambda P/R_n$, WTD, and elevation as predictors. We trained one model per vegetation group, which
443 were defined in accordance with previous studies^{62,63}: forests (open and closed evergreen and
444 deciduous, needle-leaved and broadleaved, and mixed forests), shrublands and savannahs
445 (savannahs and woody savannahs, open and closed shrublands), grasslands, and croplands.
446 For the Spatial representation of causal Shapley values (Fig. 4 and Supplementary Fig. 3) we
447 trained one model for all vegetation groups combined. This approach allows the direct
448 comparison of feature importance across the entire spatial area, thus avoiding the misleading
449 juxtaposition of Shapley values from different models with different baseline values in the
450 same map.

451 XGBoost models have recently been employed within the tree-based Shapley framework
452 across various scientific fields to quantify feature contribution^{62,65,88,89}. These models tend to
453 converge faster compared to neural networks, particularly in regression problems, where each

454 feature holds individual significance⁶⁵. The XGBoost algorithm uses an iterative decision tree
455 model composed of multiple decision trees^{90,91} and incorporates shrinkage to prevent
456 overfitting, and column subsampling to accelerate the training process⁹¹. We focused on the
457 CONUS, where WTD data are constrained by more observations and are thus more reliable³².
458 The original global data sets were thus cropped between 24°N to 50°N and 125°W to 65°W.
459 A smaller area also avoids training the model with data from very different environmental
460 conditions within the same PFT, possibly reducing the explanatory capacity of the model. We
461 used 80% data for model training and the remaining 20% for model performance testing. The
462 XGBoost model hyperparameters were tuned using grid search and tenfold cross-validation
463 and yield $R^2 \geq 0.69$ on the test set for all vegetation groups (Supplementary Fig. 7).
464 Intermediary data and all code are available in our shared Zenodo repository (see the 'Data
465 Availability' section).
466 Here, we applied the causal Shapley framework⁶⁴ to calculate Shapley values, mean absolute
467 Shapley values and dependence plots to disentangle how our predictors contribute to
468 determining the outcome of SIF/PAR. Compared to traditional Shapley values, causal
469 Shapley values offer a more nuanced approach to understanding feature contributions to
470 SIF/PAR by explicitly incorporating causal relationships between variables (Supplementary
471 Fig. 10). Unlike the conventional SHAP framework, which assumes independence between
472 features, causal Shapley values rely on Pearl's do-calculus to account for both direct and
473 indirect effects of features on the model output. This is particularly useful in scenarios where
474 features are correlated, as causal Shapley values can distinguish between the immediate
475 contribution of a feature to the target variable (direct effect) and its mediated influence
476 through other features (indirect effect). In other words, causal Shapley values allow for
477 clearer interpretations of feature relevance in systems where causal structures are known or
478 hypothesized. This is achieved by conditioning on interventions rather than observations,
479 capturing both causal effects and the dependencies between variables. As a result, causal
480 Shapley values provide a more robust framework for feature attribution in complex models,
481 improving the interpretability of machine learning models in a causally meaningful way.

482 **Data availability**

483 All data used in this study are openly available.

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

485 - Water table depth product from Ref.³²: [http://thredds-](http://thredds-gfnl.usc.es/thredds/catalog/GLOBALWTDFTP/annualmeans/catalog.html)

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

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

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

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

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

491 - Global terrain elevation data:

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

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

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

495 **Code availability**

496 All analyses were performed using R Statistical Software⁹². All intermediary data and code
497 supporting this study are available in the Zenodo Digital Repository:

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

499

500 **Acknowledgements**

501 The authors thank the providers of the data sets used in this study, the reviewers and the
502 Editorial team for their time and constructive feedback, which helped improve this
503 manuscript. We also thank Prof. Y. Fan Reinfelder, Prof. Gonzalo Miguez Macho and Dr.
504 Z.A. Pierrat for comments. We would also like to acknowledge funding from the Swiss
505 National Science Foundation (grant no. PCEFP2_181115), as well as support from the
506 LEMONTREE project (Land Ecosystem Models based on New Theory, Observation, and
507 Experiments), funded through the generosity of Eric and Wendy Schmidt, as recommended
508 by the Schmidt Futures program. Additionally, we acknowledge the USMILE European
509 Research Council Synergy Grant and the Learning the Earth with Artificial Intelligence and
510 Physics (LEAP) Science and Technology Center, funded by the National Science Foundation
511 (Award #2019625-STC).

512

513 **Author Contributions**

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

517

518 **Competing Financial Interests**

519 The authors declare no competing financial interests.

520 **References**

- 521 1. Famiglietti, J. S. The global groundwater crisis. *Nature Climate Change* vol. 4 945–
522 948 Preprint at <https://doi.org/10.1038/nclimate2425> (2014).
- 523 2. Miguez-Macho, G. & Fan, Y. Spatiotemporal origin of soil water taken up by
524 vegetation. *Nature* **598**, 624–628 (2021).
- 525 3. Salvucci, G. D. & Entekhabi, D. Equivalent steady soil moisture profile and the time
526 compression approximation in water balance modeling. *Water Resour Res* **30**, 2737–
527 2749 (1994).
- 528 4. Seneviratne, S. I. *et al.* Weather and Climate Extreme Events in a Changing Climate.
529 in *Climate Change 2021: The Physical Science Basis. Contribution of Working Group
530 I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*
531 1513–1766 (Cambridge University Press, Cambridge, United Kingdom and New York,
532 NY, USA, 2021). doi:<https://doi.org/10.1017/9781009157896.013>.
- 533 5. Famiglietti, J. S. *et al.* Satellites measure recent rates of groundwater depletion in
534 California’s Central Valley. *Geophys Res Lett* **38**, 2010GL046442 (2011).
- 535 6. Giardina, F., Gentile, P., Konings, A. G., Seneviratne, S. I. & Stocker, B. D.
536 Diagnosing evapotranspiration responses to water deficit across biomes using deep
537 learning. *New Phytologist* **240**, 968–983 (2023).
- 538 7. Oshun, J., Dietrich, W. E., Dawson, T. E. & Fung, I. Dynamic, structured
539 heterogeneity of water isotopes inside hillslopes. *Water Resour Res* **52**, 164–189
540 (2016).
- 541 8. Schwinnning, S. The ecohydrology of roots in rocks. *Ecohydrology* **3**, 238–245 (2010).
- 542 9. Fan, Y., Li, H. & Miguez-Macho, G. Global patterns of groundwater table depth.
543 *Science (1979)* **339**, 940–943 (2013).
- 544 10. Lee, J. E., Oliveira, R. S., Dawson, T. E. & Fung, I. Root functioning modifies
545 seasonal climate. *Proc Natl Acad Sci U S A* **102**, 17576–17581 (2005).
- 546 11. Seneviratne, S. I. *et al.* Investigating soil moisture-climate interactions in a changing
547 climate: A review. *Earth Sci Rev* **99**, 125–161 (2010).
- 548 12. Goedhart, C. M. & Pataki, D. E. Ecosystem effects of groundwater depth in Owens
549 Valley, California. *Ecohydrology* **4**, 458–468 (2011).
- 550 13. Fan, Y. Groundwater in the Earth’s critical zone: Relevance to large-scale patterns and
551 processes. *Water Resour Res* **51**, 3052–3069 (2015).
- 552 14. Clark, M. P. *et al.* Improving the representation of hydrologic processes in Earth
553 System Models. *Water Resour Res* **51**, 5929–5956 (2015).
- 554 15. Maxwell, R. M. & Condon, L. E. Connections between groundwater flow and
555 transpiration partitioning. *Science (1979)* **353**, 377–380 (2016).
- 556 16. Swenson, S. C., Clark, M., Fan, Y., Lawrence, D. M. & Perket, J. Representing
557 Intrahillslope Lateral Subsurface Flow in the Community Land Model. *J Adv Model
558 Earth Syst* **11**, 4044–4065 (2019).
- 559 17. Weil, R. R. & Brady, N. C. *The Nature and Properties of Soils*, Global Edition. 10105
560 (2017).
- 561 18. Dawson, T. E. & Pate, J. S. Seasonal water uptake and movement in root systems of
562 Australian phreatophytic plants of dimorphic root morphology: A stable isotope
563 investigation. *Oecologia* **107**, 13–20 (1996).
- 564 19. Voltas, J., Lucabaugh, D., Chambel, M. R. & Ferrio, J. P. Intraspecific variation in the
565 use of water sources by the circum-Mediterranean conifer *Pinus halepensis*. *New
566 Phytologist* **208**, 1031–1041 (2015).
- 567 20. Grossiord, C. *et al.* Prolonged warming and drought modify belowground interactions
568 for water among coexisting plants. *Tree Physiol* **39**, 55–63 (2018).

- 569 21. Stocker, B. D. *et al.* Global patterns of water storage in the rooting zones of vegetation.
570 *Nat Geosci* **16**, 250–256 (2023).
- 571 22. Rempe, D. M. & Dietrich, W. E. Direct observations of rock moisture, a hidden
572 component of the hydrologic cycle. *Proc Natl Acad Sci U S A* **115**, 2664–2669 (2018).
- 573 23. McCormick, E. L. *et al.* Widespread woody plant use of water stored in bedrock.
574 *Nature* **597**, 225–229 (2021).
- 575 24. Querejeta, J. I., Estrada-Medina, H., Allen, M. F. & Jiménez-Osornio, J. J. Water
576 source partitioning among trees growing on shallow karst soils in a seasonally dry
577 tropical climate. *Oecologia* **152**, 26–36 (2007).
- 578 25. Evaristo, J. & McDonnell, J. J. Prevalence and magnitude of groundwater use by
579 vegetation: A global stable isotope meta-analysis. *Sci Rep* **7**, 1–12 (2017).
- 580 26. Barbeta, A. & Peñuelas, J. Relative contribution of groundwater to plant transpiration
581 estimated with stable isotopes. *Sci Rep* **7**, 1–10 (2017).
- 582 27. Jobbágy, E. G., Noysetto, M. D., Villagra, P. E. & Jackson, R. B. Water subsidies from
583 mountains to deserts: their role in sustaining groundwater-fed oases in a sandy
584 landscape. *Ecological Applications* **21**, 678–694 (2011).
- 585 28. Tumber-Dávila, S. J., Schenk, H. J., Du, E. & Jackson, R. B. Plant sizes and shapes
586 above and belowground and their interactions with climate. *New Phytologist* **235**,
587 1032–1056 (2022).
- 588 29. Elmore, A. J., Manning, S. J., Mustard, J. F. & Craine, J. M. Decline in alkali meadow
589 vegetation cover in California: The effects of groundwater extraction and drought.
590 *Journal of Applied Ecology* **43**, 770–779 (2006).
- 591 30. Howard, J. & Merrifield, M. Mapping Groundwater Dependent Ecosystems in
592 California. *PLoS One* **5**, e11249 (2010).
- 593 31. Sutanto, S. J. *et al.* A perspective on isotope versus non-isotope approaches to
594 determine the contribution of transpiration to total evaporation. *Hydrol Earth Syst Sci*
595 **18**, 2815–2827 (2014).
- 596 32. Fan, Y., Miguez-Macho, G., Jobbágy, E. G., Jackson, R. B. & Otero-Casal, C.
597 Hydrologic regulation of plant rooting depth. *Proc Natl Acad Sci U S A* **114**, 10572–
598 10577 (2017).
- 599 33. Joiner, J. *et al.* Global monitoring of terrestrial chlorophyll fluorescence from
600 moderate-spectral-resolution near-infrared satellite measurements: methodology,
601 simulations, and application to GOME-2. *Atmos Meas Tech* **6**, 2803–2823 (2013).
- 602 34. Frankenberg, C. *et al.* New global observations of the terrestrial carbon cycle from
603 GOSAT: Patterns of plant fluorescence with gross primary productivity. *Geophys Res
604 Lett* **38**, 1–6 (2011).
- 605 35. Green, J. K. *et al.* Regionally strong feedbacks between the atmosphere and terrestrial
606 biosphere. *Nat Geosci* **10**, 410–414 (2017).
- 607 36. Harper, K. L. *et al.* A 29-year time series of annual 300m resolution plant-functional-
608 type maps for climate models. *Earth Syst Sci Data* **15**, 1465–1499 (2023).
- 609 37. Verkaik, J., Sutanudjaja, E. H., Oude Essink, G. H. P., Lin, H. X. & Bierkens, M. F. P.
610 GLOBGM v1.0: A parallel implementation of a 30arcsec PCR-GLOBWB-
611 MODFLOW global-scale groundwater model. *Geosci Model Dev* **17**, 275–300 (2024).
- 612 38. Yang, Y., Donohue, R. J., McVicar, T. R. & Roderick, M. L. An analytical model for
613 relating global terrestrial carbon assimilation with climate and surface conditions using
614 a rate limitation framework. *Geophys Res Lett* **42**, 9825–9835 (2015).
- 615 39. Budyko, M. I. *Climate and Life*. (Academic Press, 1974).
- 616 40. Budyko, M. I. *Тепловой Баланс Земной Поверхности (Heat Balance of the Earth's
617 Surface)*. (Gidrometeoizdat, Leningrad, 1956).

- 618 41. Condon, L. E. & Maxwell, R. M. Systematic shifts in Budyko relationships caused by
619 groundwater storage changes. *Hydrol Earth Syst Sci* **21**, 1117–1135 (2017).
- 620 42. Istanbulluoglu, E., Wang, T., Wright, O. M. & Linters, J. D. Interpretation of
621 hydrologic trends from a water balance perspective: The role of groundwater storage in
622 the Budyko hypothesis. *Water Resour Res* **48**, (2012).
- 623 43. Greve, P., Gudmundsson, L., Orlowsky, B. & Seneviratne, S. I. Introducing a
624 probabilistic Budyko framework. *Geophys Res Lett* **42**, 2261–2269 (2015).
- 625 44. Crosbie, R. S., Holland, K. L. & McVicar, T. R. Regional-scale partitioning of
626 transmission losses and groundwater recharge using satellite estimates of actual
627 evapotranspiration in an arid environment. *Ecohydrology* **16**, e2490 (2023).
- 628 45. O'Grady, A. P., Carter, J. L. & Bruce, J. Can we predict groundwater discharge from
629 terrestrial ecosystems using existing eco-hydrological concepts? *Hydrol Earth Syst Sci*
630 **15**, 3731–3739 (2011).
- 631 46. Maes, W. H. *et al.* Sun-induced fluorescence closely linked to ecosystem transpiration
632 as evidenced by satellite data and radiative transfer models. *Remote Sens Environ* **249**,
633 112030 (2020).
- 634 47. Pagán, B., Maes, W., Gentine, P., Martens, B. & Miralles, D. Exploring the Potential
635 of Satellite Solar-Induced Fluorescence to Constrain Global Transpiration Estimates.
636 *Remote Sens (Basel)* **11**, 413 (2019).
- 637 48. Jonard, F. *et al.* Value of sun-induced chlorophyll fluorescence for quantifying
638 hydrological states and fluxes: Current status and challenges. *Agric For Meteorol* **291**,
639 108088 (2020).
- 640 49. Köhler, P. *et al.* Global Retrievals of Solar-Induced Chlorophyll Fluorescence With
641 TROPOMI: First Results and Intersensor Comparison to OCO-2. *Geophys Res Lett* **45**,
642 10,456–10,463 (2018).
- 643 50. Hersbach, H. *et al.* The ERA5 global reanalysis. *Quarterly Journal of the Royal
644 Meteorological Society* **146**, 1999–2049 (2020).
- 645 51. Pagán, B., Maes, W., Gentine, P., Martens, B. & Miralles, D. Exploring the Potential
646 of Satellite Solar-Induced Fluorescence to Constrain Global Transpiration Estimates.
647 *Remote Sens (Basel)* **11**, 413 (2019).
- 648 52. Beck, H. E. *et al.* MSWEP V2 Global 3-hourly 0.1° Precipitation: Methodology and
649 Quantitative Assessment. *Bull Am Meteorol Soc* **100**, 473–500 (2019).
- 650 53. Gentine, P., D'Odorico, P., Lintner, B. R., Sivandran, G. & Salvucci, G.
651 Interdependence of climate, soil, and vegetation as constrained by the Budyko curve.
652 *Geophys Res Lett* **39**, 2–7 (2012).
- 653 54. Gudmundsson, L., Greve, P. & Seneviratne, S. I. The sensitivity of water availability
654 to changes in the aridity index and other factors-A probabilistic analysis in the Budyko
655 space. *Geophys Res Lett* **43**, 6985–6994 (2016).
- 656 55. Lintner, B. R., Gentine, P., Findell, K. L. & Salvucci, G. D. The Budyko and
657 complementary relationships in an idealized model of large-scale land-atmosphere
658 coupling. *Hydrol Earth Syst Sci* **19**, 2119–2131 (2015).
- 659 56. Prentice, I. C. *et al.* Evidence of a universal scaling relationship for leaf CO₂
660 drawdown along an aridity gradient. *New Phytologist* **190**, 169–180 (2011).
- 661 57. Greve, P., Roderick, M. L. & Seneviratne, S. I. Simulated changes in aridity from the
662 last glacial maximum to 4xCO₂. *Environmental Research Letters* **12**, 114021 (2017).
- 663 58. Manabe, S. *Climate and Ocean Circulation. Part I: The Atmospheric Circulation and
664 the Hydrology of the Earth's Surface*. vol. 97 (Mon. Weather Rev., 1969).
- 665 59. Cutler, D. R. *et al.* Random forests for classification in ecology. *Ecology* **88**, 2783–
666 2792 (2007).

- 667 60. Hutengs, C. & Vohland, M. Downscaling land surface temperatures at regional scales
668 with random forest regression. *Remote Sens Environ* **178**, 127–141 (2016).
- 669 61. Green, J. K. *et al.* Surface temperatures reveal the patterns of vegetation water stress
670 and their environmental drivers across the tropical Americas. *Glob Chang Biol* **28**,
671 2940–2955 (2022).
- 672 62. Wang, H. *et al.* Exploring complex water stress–gross primary production
673 relationships: Impact of climatic drivers, main effects, and interactive effects. *Glob
674 Chang Biol* **28**, 4110–4123 (2022).
- 675 63. Fernández-Martínez, M. *et al.* The role of climate, foliar stoichiometry and plant
676 diversity on ecosystem carbon balance. *Glob Chang Biol* **26**, 7067–7078 (2020).
- 677 64. Heskes, T., Sijben, E., Bucur, I. G. & Claassen, T. Causal Shapley Values: Exploiting
678 Causal Knowledge to Explain Individual Predictions of Complex Models. *Adv Neural
679 Inf Process Syst* **2020-December**, (2020).
- 680 65. Lundberg, S. M. *et al.* From local explanations to global understanding with
681 explainable AI for trees. *Nat Mach Intell* **2**, 56–67 (2020).
- 682 66. Wei, D. *et al.* Elevation-dependent pattern of net CO₂ uptake across China. *Nature
683 Communications* **2024** *15:1* **15**, 1–10 (2024).
- 684 67. Lundberg, S. M. & Lee, S.-I. A Unified Approach to Interpreting Model Predictions.
685 *Adv Neural Inf Process Syst* **30**, 4768–4777 (2017).
- 686 68. Rohde, M. M. *et al.* Groundwater-dependent ecosystem map exposes global dryland
687 protection needs. *Nature* **2024** *632:8023* **632**, 101–107 (2024).
- 688 69. Terazawa, K., Maruyama, Y. & Morikawa, Y. Photosynthetic and stomatal responses
689 of Larix kaempferi seedlings to short-term waterlogging. *Ecol Res* **7**, 193–197 (1992).
- 690 70. Terazawa, K. & Kikuzawa, K. Effects of flooding on leaf dynamics and other seedling
691 responses in flood-tolerant *Alnus japonica* and flood-intolerant *Betula platyphylla* var.
692 *japonica*. *Tree Physiol* **14**, 251–261 (1994).
- 693 71. Ohta, T. *et al.* Effects of waterlogging on water and carbon dioxide fluxes and
694 environmental variables in a Siberian larch forest, 1998–2011. *Agric For Meteorol* **188**,
695 64–75 (2014).
- 696 72. Fan, Y. *et al.* Hillslope Hydrology in Global Change Research and Earth System
697 Modeling. *Water Resour Res* **55**, 1737–1772 (2019).
- 698 73. Meek, D. W., Hatfield, J. L., Howell, T. A., Idso, S. B. & Reginato, R. J. A
699 Generalized Relationship between Photosynthetically Active Radiation and Solar
700 Radiation1. *Agron J* **76**, 939–945 (1984).
- 701 74. Tarek, M., Brissette, F. P. & Arsenault, R. Large-scale analysis of global gridded
702 precipitation and temperature datasets for climate change impact studies. *J
703 Hydrometeorol* **21**, 2623–2640 (2020).
- 704 75. Beck, H. E. *et al.* Daily evaluation of 26 precipitation datasets using Stage-IV gauge-
705 radar data for the CONUS. *Hydrol Earth Syst Sci* **23**, 207–224 (2019).
- 706 76. GMTED2010 Global Grids.
707 https://topotools.cr.usgs.gov/gmted_viewer/gmted2010_global_grids.php.
- 708 77. Duan, L., Liu, T., Wang, X. & Luo, Y. Spatio-Temporal Patterns of Water Table and
709 Vegetation Status of a Deserted Area. *Water* **2015**, Vol. 7, Pages 5788–5805 **7**, 5788–
710 5805 (2015).
- 711 78. Giardina, F. *et al.* Tall Amazonian forests are less sensitive to precipitation variability.
712 *Nat Geosci* **11**, 405–409 (2018).
- 713 79. Zeng, Y. *et al.* Optical vegetation indices for monitoring terrestrial ecosystems
714 globally. *Nature Reviews Earth & Environment* **2022** *3:7* **3**, 477–493 (2022).

- 715 80. Dechant, B. *et al.* Canopy structure explains the relationship between photosynthesis
716 and sun-induced chlorophyll fluorescence in crops. *Remote Sens Environ* **241**, 111733
717 (2020).
- 718 81. Pierrat, Z. *et al.* Diurnal and Seasonal Dynamics of Solar-Induced Chlorophyll
719 Fluorescence, Vegetation Indices, and Gross Primary Productivity in the Boreal Forest.
720 *J Geophys Res Biogeosci* **127**, e2021JG006588 (2022).
- 721 82. Sun, Y. *et al.* Overview of Solar-Induced chlorophyll Fluorescence (SIF) from the
722 Orbiting Carbon Observatory-2: Retrieval, cross-mission comparison, and global
723 monitoring for GPP. *Remote Sens Environ* **209**, 808–823 (2018).
- 724 83. Guanter, L. *et al.* The TROPOSIF global sun-induced fluorescence dataset from the
725 Sentinel-5P TROPOMI mission. *Earth Syst Sci Data* **13**, 5423–5440 (2021).
- 726 84. Ju, J. & Roy, D. P. The availability of cloud-free Landsat ETM+ data over the
727 conterminous United States and globally. *Remote Sens Environ* **112**, 1196–1211
728 (2008).
- 729 85. Joiner, J. *et al.* Filling-in of near-infrared solar lines by terrestrial fluorescence and
730 other geophysical effects: Simulations and space-based observations from
731 SCIAMACHY and GOSAT. *Atmos Meas Tech* **5**, 809–829 (2012).
- 732 86. Frankenberg, C. *et al.* New global observations of the terrestrial carbon cycle from
733 GOSAT: Patterns of plant fluorescence with gross primary productivity. *Geophys Res
734 Lett* **38**, 1–6 (2011).
- 735 87. McColl, K. A., Salvucci, G. D. & Gentine, P. Surface Flux Equilibrium Theory
736 Explains an Empirical Estimate of Water-Limited Daily Evapotranspiration. *J Adv
737 Model Earth Syst* **11**, 2036–2049 (2019).
- 738 88. Yang, C., Chen, M. & Yuan, Q. The application of XGBoost and SHAP to examining
739 the factors in freight truck-related crashes: An exploratory analysis. *Accid Anal Prev*
740 **158**, 106153 (2021).
- 741 89. Guo, M. *et al.* Older Pedestrian Traffic Crashes Severity Analysis Based on an
742 Emerging Machine Learning XGBoost. *Sustainability* **13**, 926 (2021).
- 743 90. Yan, F., Song, K., Liu, Y., Chen, S. & Chen, J. Predictions and mechanism analyses of
744 the fatigue strength of steel based on machine learning. *J Mater Sci* **55**, 15334–15349
745 (2020).
- 746 91. Chen, T. & Guestrin, C. XGBoost: A scalable tree boosting system. in *Proceedings of
747 the ACM SIGKDD International Conference on Knowledge Discovery and Data
748 Mining* vols 13–17–August–2016 785–794 (Association for Computing Machinery,
749 New York, NY, USA, 2016).
- 750 92. R Core Team. R: A language and environment for statistical computing. R Foundation
751 for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>. (2023).
- 752