Groundwater rivals aridity in determining global photosynthesis

Francesco Giardina[1,2], Sonia I. Seneviratne[1], Jiangong Liu[2], Benjamin D. Stocker[3,4], Pierre Gentine[2,5]

[1]Institute for Atmospheric and Climate Science, Department of Environmental Systems Science, ETH Zurich, CH-8092 Zürich, Switzerland
[2]Department of Earth and Environmental Engineering, Columbia University, New York, New York 10027, USA
[3]Institute of Geography, University of Bern, Hallerstrasse 12, 3012 Bern, Switzerland
[4]Oeschger Centre for Climate Change Research, University of Bern, Falkenplatz 16, 3012 Bern, Switzerland
[5]Center for Learning the Earth with Artificial intelligence and Physics (LEAP), Columbia University, New York, New York 10027, USA

Author for correspondence:
Francesco Giardina
Email: fgiardina@ethz.ch

Abstract

Understanding the role of groundwater in regulating photosynthesis is key in land-climate interactions. However, the impact of groundwater on terrestrial ecosystem productivity remains poorly understood. Here, we use satellite observations of solar-induced fluorescence as a proxy for photosynthesis, together with estimates of water table depth (WTD) and aridity as quantified by the annual moisture index with reanalysis data, to investigate the relationships between groundwater and photosynthesis. Using explainable machine learning (Shapley additive explanations or SHAP), we demonstrate that groundwater plays a crucial role in determining the spatial patterns of global photosynthesis, with varying importance across different ecosystem types, and that this effect is comparable to aridity. We show that in forests across the contiguous USA (CONUS), the relative importance of groundwater represents 89% of the effect attributed to the moisture index in modulating photosynthesis, and this impact is even more pronounced globally (105% relative importance). The relative importance of groundwater compared to the moisture index remains substantial in grasslands (37%), savannahs and shrublands (25%), and croplands (15%) in CONUS. Global analysis confirms these trends for grasslands (41%) and croplands (15%) but indicates a 60% importance in savannahs and shrublands. Our findings highlight the key global role of groundwater in driving ecosystem productivity.

Main text

Groundwater is a critical component of terrestrial water storage[1], yet its impact on photosynthesis remains poorly constrained. Fluctuations in water table levels exhibit distinct spatiotemporal patterns compared to surface soil moisture[2,3], as they can be related to long-range landscape topography or soil components in addition to local climate. Understanding how groundwater modulates photosynthesis and its impact on the global water and energy fluxes is essential in the context of the rising occurrence of dry extremes in several land regions[4]. Belowground availability of water becomes indispensable during dry spells[5,6], which are projected to become more frequent and intense under climate change[4]. Deep-rooted plants can tap into underground water pools stored in a saturated zone[7–9] and then release water into the atmosphere through leaf stomata, regulating atmospheric humidity and influencing regional climate[10,11]. Determining the control of groundwater on photosynthesis is thus key to accurately predicting land-surface processes and their effects on climate[12,13]. There has been a growing effort to incorporate groundwater into Earth System Models in recent years[13–16], motivated by the evidence that water tables can influence vegetation over notable portions of the land area[9,13]. However, to improve the representation of groundwater-surface interactions in models, it is necessary to develop a solid empirical basis for detecting, quantifying, and mapping the control of groundwater on plant physiological processes and ecosystem carbon and water fluxes.

Systematically assessing the contribution of groundwater to photosynthesis poses significant challenges due to data sampling limitations and the representativity of the in-situ data used, given the complex nature of groundwater across soil type and topographic gradients. Recent research based on inverse modelling has shown that, at the global and annual scales, the primary source of plant-available water is recent precipitation stored in shallow soil layers[2]. Parallel to this, further studies have documented extensive plant water uptake from deeper soil[17–20], rock moisture[21–23] or groundwater[24–27]. Global syntheses of stable isotope measurements have revealed that groundwater use is widespread across various biomes globally, with an increased contribution in arid or seasonally dry regions[24,25]. Yet, the presence of deep-rooted plants worldwide suggests that groundwater usage may extend beyond those sparse sampled sites[28].

Investigations into the uptake of groundwater by plants have thus either focused regionally on syntheses of measurements from stable isotope techniques[2,24,25], and at larger scales, they have relied on indirect methods such as geospatial data-based techniques[29,30], land-surface model[14,31] and inverse modelling approaches[2,9,32]. However, there is a lack of

research at the multi-year scale that incorporates consistent large-scale data while at the same time accounting for plant physiologies and aridity gradients. Using a global remotely sensed dataset of solar-induced fluorescence (SIF) as a proxy for photosynthesis[33–35], coupled with global maps of plant functional types (PFTs)[36] and water table depth (WTD) estimates[32], we evaluate the contribution of groundwater to photosynthesis globally and with a specific focus on the United States. SIF is nearly proportional with gross primary production (GPP) over large temporal and spatial scales[34,35]. Applying explainable machine learning, we show that groundwater plays a role equivalent to aridity in regulating photosynthesis across plant functional types around the globe.

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

A pivotal contribution to the conceptual understanding of regional dryness patterns is the Budyko relationship, which can also illuminate the role of long-term water stress on evapotranspiration and photosynthesis. We present it here in one of its formulations most aligned with our study[37,38]:

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

Where ET is the actual evapotranspiration, R_n is net radiation and P is precipitation multiplied by λ, the latent heat of evaporation[37,38]. The relationship outlines the partitioning of water across an aridity gradient, from water-limited to energy-limited regimes. It conceptualizes the fact that in drier climates, a larger fraction of the available energy goes into evapotranspiration relative to the amount of precipitation[37,38] (Fig. 1). However, a major limitation of the traditional Budyko framework is that it assumes that the main regulator of ET on long-term equilibrium is local climate[39,40] and it does not explicitly account for the effect of long-term water storage changes by groundwater, despite its significant role in the water balance. Recent studies, however, have suggested that groundwater can alter this relationship[39–41]. The inclusion of groundwater in the Budyko framework influences its shape[39,40]. In particular, a positive groundwater contribution leads to higher evapotranspiration for a given aridity index (or "radiative index of dryness", as referred to by Budyko), thus resulting in points above the Budyko curve in its water-limited region[39,40] (i.e. $\lambda P/R_n < 1$ in Fig. 1a). This indicates that groundwater access is supporting additional water consumption by plants through transpiration, enhancing ET beyond the constraints of the local water balance represented by the Budyko curve[39,40]. We here use a similar conceptual approach but for photosynthesis, as photosynthesis and ET are tightly coupled through the plant water use efficiency[42–44].

In our study, we want first to estimate the extent of groundwater use across an aridity gradient. Given the near-linear relationship between SIF and GPP[34,35] at long time scales, we use SIF based on the high-resolution global TROPOMI sensor[45] as a proxy for ecosystem-level photosynthesis. To assess variations in SIF beyond climate forcings, we normalize SIF by dividing it by photosynthetically active radiation (PAR) from MODIS[46] (SIF/PAR, see Methods). SIF has been shown to accurately quantify the impact of environmental stress on

ecosystem transpiration (T)[43,44] and shows a closer relation to T than any other space-based measurement[42]. We thus use SIF/PAR as an observational proxy of the evaporative fraction on the y-axis in the Budyko framework (Fig. 1b). We define aridity as the ratio of MSWEP precipitation[47] (P) to ERA5-Land net radiation[48] (R_n), denoted as $\lambda P/R_n$, where λ is the latent heat of vaporization (J kg^{-1}) for units consistency. This metric provides a simple yet effective way to characterize the water availability of a region in relation to the available energy and, since P and R_n are based on long-term averages, it is an index of the mean climatic conditions of the area[39,40,49–51]. Given that higher $\lambda P/R_n$ values indicate lower aridity (the inverse of Budyko's formulation) and more water availability, we adopt the term moisture index rather than aridity index to avoid confusion, in line with many previous studies[5,52]. It's important to note that, while commonly used in hydroclimatological studies[39,40,49–51], the index is just one of the possible formulations of aridity[53].

As large-scale remote sensing datasets do not provide the exact variables used by Budyko (e.g. ET), we describe our approach as 'Budyko-like', indicating an adaptation of the original framework rather than a direct application (Fig. 1). This is similar to a moisture limitation model previously formulated by Manabe based on Budyko ("bucket model")[11,54], but on long-time scales. In this framework, we expect that a positive contribution of groundwater results in points higher than the Budyko curve in the water-limited regime (Fig. 1a, red versus blue lines). We identified pixels deviating from the moisture index as those potentially located in zones where plants might access groundwater. These points account for 15% of the points within CONUS (Fig. 1b, points within the red triangle). Note that this figure is not a quantitative assessment of groundwater access; rather, it offers an overview of the issue to contextualize our study and motivates the subsequent sections.

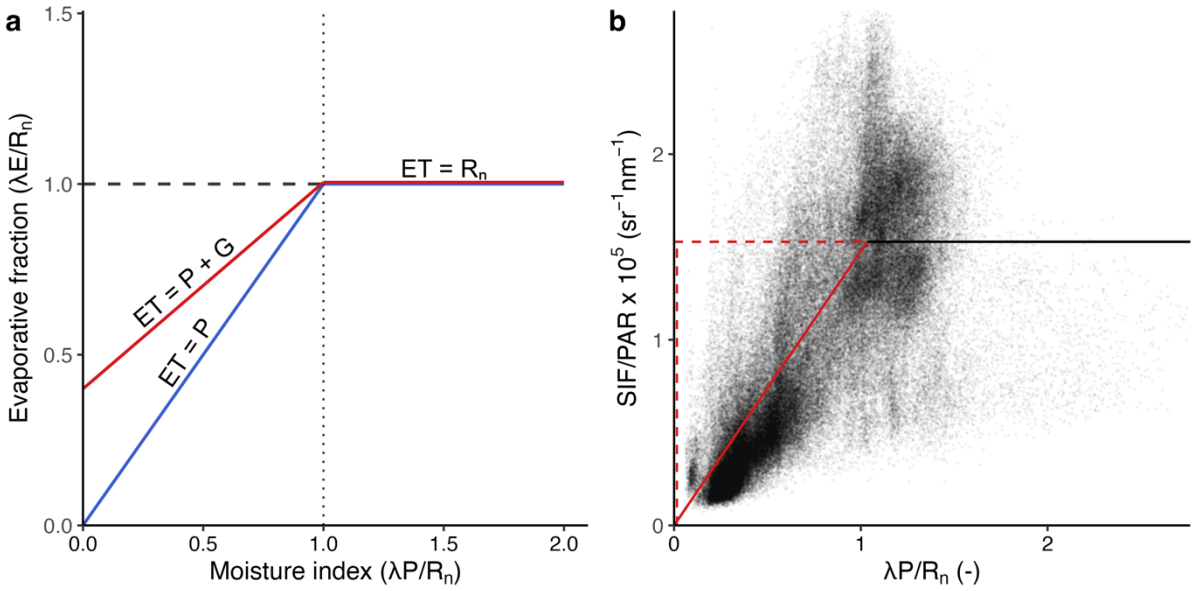

Fig. 1 | Groundwater in the Budyko framework. a, Conceptual Budyko-like framework. The area between the blue and red lines in the water-limited regime ($\lambda P/R_n < 1$) indicates groundwater access due to ET values exceeding expected limits from the moisture index[39]. ET: evaporation, P: precipitation, G: groundwater contribution. R_n: net radiation. λ: latent heat of vaporization (J kg^{-1}). **b**, SIF/PAR distribution in the Budyko-like space across the USA for savannahs. Continuous red and black segments depict a segmented regression model between SIF/PAR and $\lambda P/R_n$. SIF/PAR is used as an observational proxy of the evaporative fraction. Points within the red triangle correspond to zones where plants potentially access groundwater as per framework in **a**.

Local contributions of groundwater table and moisture index

To quantitively study the contribution of groundwater versus aridity on photosynthesis, we trained Extreme gradient boosting (XGBoost) models with SIF/PAR (Fig. 2a) as target variable and WTD and $\lambda P/R_n$ as predictors (Fig. 2b and c, respectively; see Methods). Thanks to their ability to encode non-linear relationships, tree-based machine learning models have gained increasing popularity in ecological applications[55–57]. Consistently with previous studies[58,59], we aggregated PFTs into four groups: forests, shrublands and savannahs, grasslands, and croplands, and trained one XGBoost model per PFT group (Fig. 2d, see Methods). As TROPOMI was launched in 2018, a four-year record of SIF data was used and we calculate the mean over this period for each variable to yield one value for each 0.083° pixel. The predictors $\lambda P/R_n$ and WTD show relatively low correlation coefficients across PFT groups with no pronounced linear relationship, indicating that cross-correlation does not impede the ability of our machine-learning models to capture complex patterns and nonlinearities in the data (Supplementary Fig. 1; see Methods). Our models show good predictive performance for the United States, with $R^2 \geq 0.49$ on the test set for all PFT groups (R^2=0.49 for forests, R^2=0.82 for savannahs and shrublands, R^2=0.64 for

grasslands and R^2=0.66 for croplands, Supplementary Table 1). This demonstrates that despite the simplicity of our models (i.e. only two predictors), they are effective in predicting long-term values. When performing the same analysis using global data, the model performance is lower (R^2=0.20 for forests, R^2=0.31 for savannahs and shrublands, R^2=0.27 for grasslands and R^2=0.36 for croplands, see Supplementary Table 1), likely due to the higher quality of groundwater data in CONUS, due to extensive data sampling in the region[32]. Therefore, while retaining our global scope, we placed an additional focus on the United States for the rest of our analysis. To assess the individual contributions of WTD and $\lambda P/R_n$ to the sensitivity of SIF/PAR, we calculated Shapley additive explanations (SHAP)[60,61] for each XGBoost model, i.e. for each PFT group (Fig. 2d).

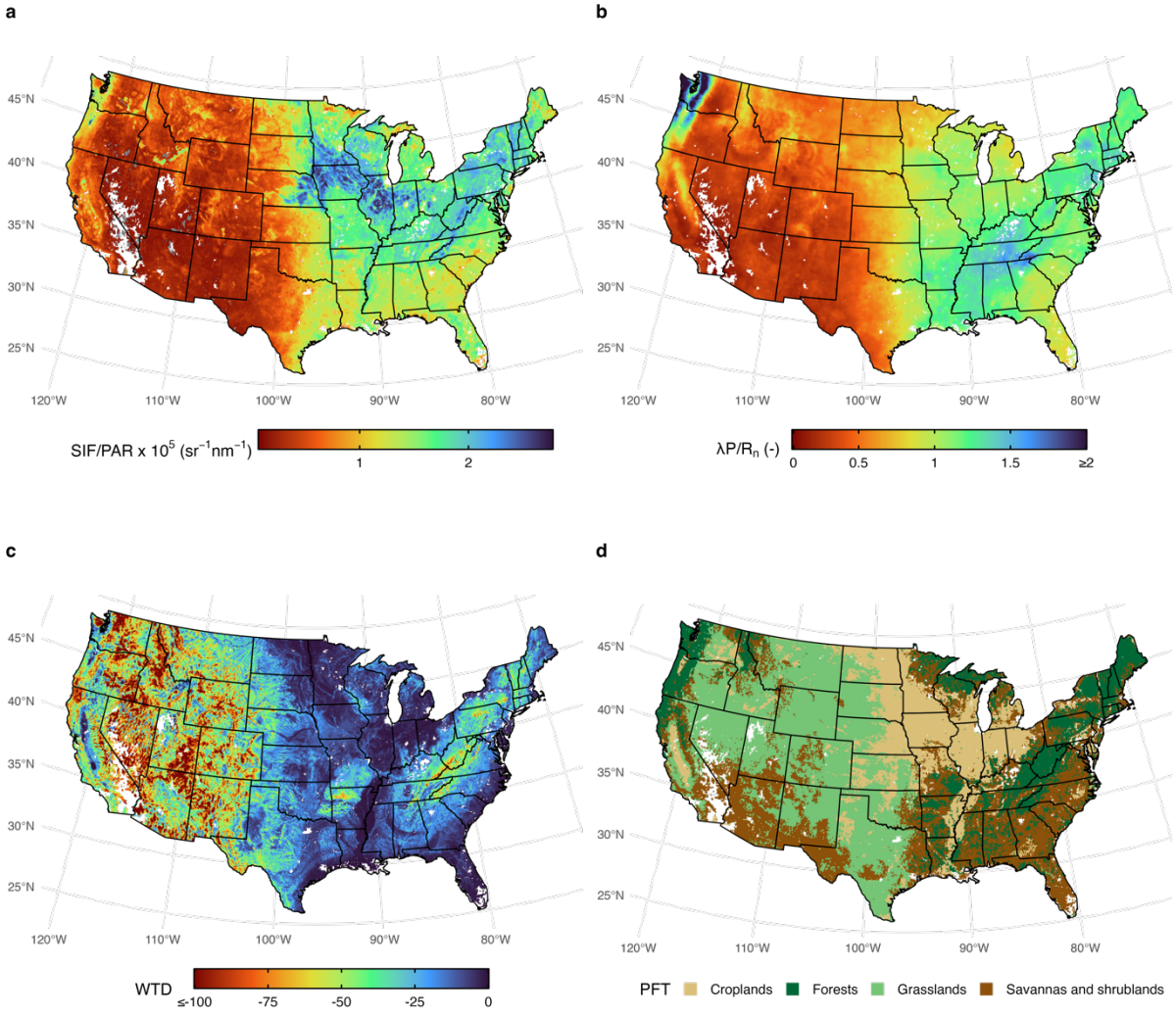

Fig. 2 | Spatial patterns of plant-functional types and long-term means of climatic and environmental variables in the United States. a, Solar-induced fluorescence to photosynthetically active radiation ratio (SIF/PAR). **b,** Moisture index represented as the ratio of precipitation to net radiation (P/Rn). **c,** Water table depth (WTD). **d,** Classification of plant functional types (PFTs).

Machine learning models are based on nonlinear algorithms that typically yield higher accuracy than linear models, especially on large datasets, yet they often lack interpretability ('Black box' model predictions[60]). SHAP values have been used in ecology for interpreting model predictions[58]. They are based on an algorithm rooted in game theory and describe the influence of each input feature (or predictor, i.e. WTD and $\lambda P/R_n$) on a specific prediction (i.e. SIF/PAR value predicted by the XGBoost models). SHAP values directly quantify local (i.e., for a single data point) effects of individual predictors, while at the same time allowing to account for the global structure of the model, thus unveiling meaningful patterns that might otherwise remain undetected[60,61]. A positive SHAP value indicates that a particular predictor leads to an increase in the target variable (i.e. SIF/PAR) above the average outcome predicted by the XGBoost model. A negative SHAP value denotes the opposite. SHAP values that approach zero indicate that SIF/PAR is nearing its average value across the dataset. This, however, should not be interpret as lack of sensitivity with the corresponding predictor.

We compute the average SHAP values (i.e. the mean SHAP values among all SHAP values of a specific predictor) for WTD and $\lambda P/R_n$ across all pixels in CONUS (Fig. 3, left of each plot). Our analysis indicates that in forested areas, the relative importance of groundwater (average SHAP value of 0.164, Fig. 3a) is 89% of the effect linked to the moisture index (average SHAP value of 0.184, Fig. 3a) in determining ecosystem productivity. In grasslands, the relative importance of groundwater compared to the moisture index is 37% (average SHAP values of 0.06 and 0.161, respectively, Fig. 3d), decreasing to 25% in savannahs and shrublands (average SHAP values of 0.109 and 0.441, respectively, Fig. 3b), and to 15% in croplands (average SHAP values of 0.048 and 0.326, Fig. 3c). This outcome is consistent with the fact that forests typically have deeper roots compared to other PFTs, allowing them to access water sources at greater depths[32,62]. As a result, the productivity of forests is more affected by changes in WTD. Note that many croplands may be irrigated, so the relative importance of groundwater could be affected.

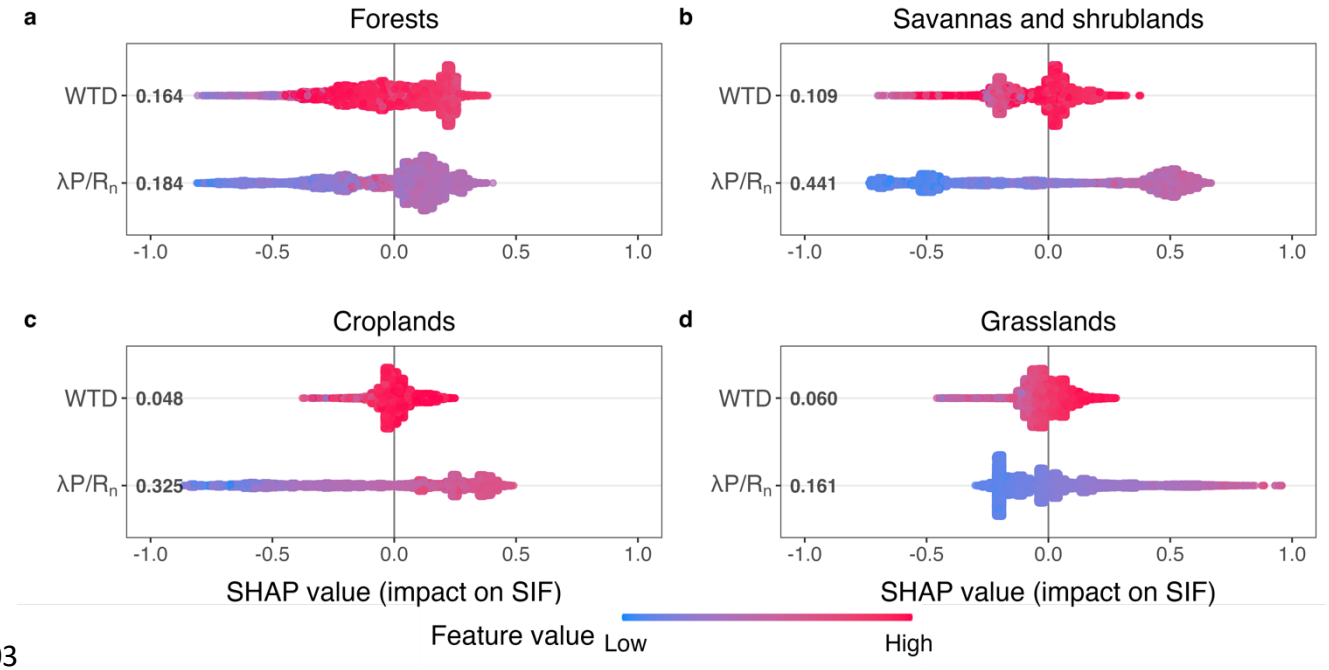

Fig. 3 | Local explanations (SHAP values) of ecosystem productivity drivers, based on a gradient boosting decision tree model trained on a dataset of solar-induced fluorescence. a-d, SHAP summary plots, which show the effect of different predictors on each model outcome. Each dot corresponds to the long-term mean of an individual pixel from global datasets. The average SHAP 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 SHAP value of the predictor, i.e. how a predictor affects the outcome of the model (i.e. SIF/PAR). A positive SHAP value suggests an increase in SIF/PAR while a negative one indicates the opposite. Overlapping dots at an *x*-coordinate denote higher density, suggesting similar effects across multiple points. WTD: water table depth, $\lambda P/R_n$: moisture index. **a**, Forests (evergreen and deciduous, needle-leaved and broadleaved, and mixed forests). **b**, Savannahs and shrublands (savannahs and woody savannahs, open and closed shrublands). **c**, Croplands. **d**, Grasslands.

We obtain consistent results when performing the analysis with data points from the entire globe (Supplementary Fig. 2). At the global level, the influence of groundwater relative to aridity is even more pronounced in forests, showing a 105% relative importance, while it is analogous for grasslands (41%) and identical for croplands (15%), but shows a relative importance of 60% in savannahs and shrublands. The stronger control of groundwater on photosynthesis in savannahs and shrublands at the global scale is consistent with the fact that access to groundwater is typically observed in arid climates, such as most savannahs and shrublands[24,25]. The diminished significance of groundwater observed in CONUS may be due to large climate gradients in savannahs and shrublands in this region (e.g. Fig. 2b, d), which are reconciled when considering the global perspective.

By disentangling the magnitude, prevalence, and direction of an effect from a single metric to a plot, SHAP 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 Fig. 3) tend to yield higher local output values of SIF/PAR, as demonstrated by the corresponding positive SHAP 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 average SHAP values for WTD (Fig. 4) aligns well with an observational map recently published comparing water levels in 4.2 millions wells and streams across the United States (Fig. 2 in Ref.[63]; the data is not publicly available and cannot be reproduced here). Our approach successfully highlights similar key patterns. For instance, regions like New England emerge as zones where WTD has a positive effect on photosynthesis (i.e. positive SHAP value, blue areas in Fig. 4), a characteristic similarly observed in the Interior Highlands (i.e. Oklahoma, Arkansas, and Missouri), as well as the north-western Pacific and Central Valley of California (Fig. 4). In contrast, areas such as the coastal plains of the Southern US, the arid Southwestern US, and the central lowland consistently present as zones where WTD has a negative effect on photosynthesis (red areas in Fig. 4). These zones coincide with locations where the WTD is lower than stream surfaces and thus local rivers flow into the aquifers[63]. On the other end, the spatial distribution of the average SHAP values for $\lambda P/R_n$ (Supplementary Fig. 4) aligns with the long-term mean values of $\lambda P/R_n$ (Fig. 2b).

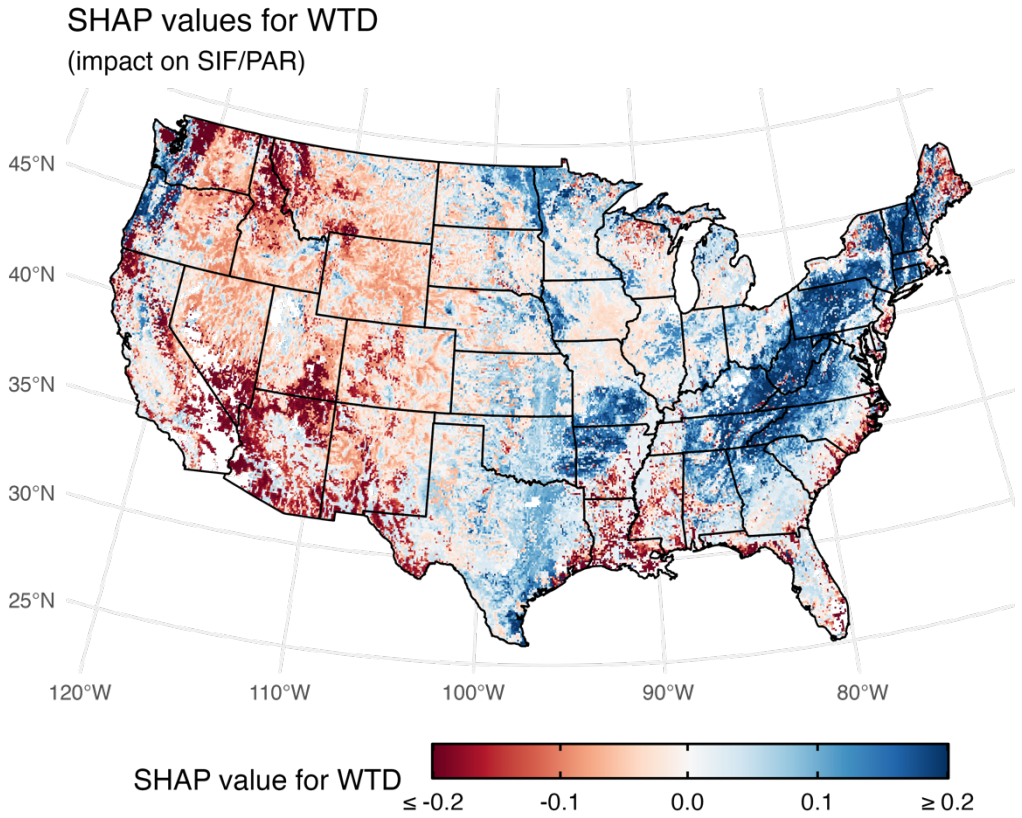

Fig. 4 | Spatial representation of the average SHAP values of WTD across the continental United States. SHAP values illustrate the effect of groundwater (WTD) on SIF/PAR. Positive SHAP values suggests that WTD has a positive impact on SIF/PAR (blue points), thus meaning that groundwater is supporting additional photosynthetic activity. Conversely, negative SHAP values indicate that the WTD value has a negative effect on SIF/PAR in that specific pixel (red points). The map is based on the results of the four PFT-specific XGBoosting models (see Methods).

Coupling effect between water table depth and moisture index

SHAP dependence plots reveal the impact on the prediction (*y*-axis) of a particular variable (*x*-axis) for each sample in our dataset, as well as the coupling or dependence effect between two factors (Fig. 5). A coupling effect between WTD and the moisture index is discernible in Fig. 5, characterized by a vertical dispersion in the SHAP dependence plots.

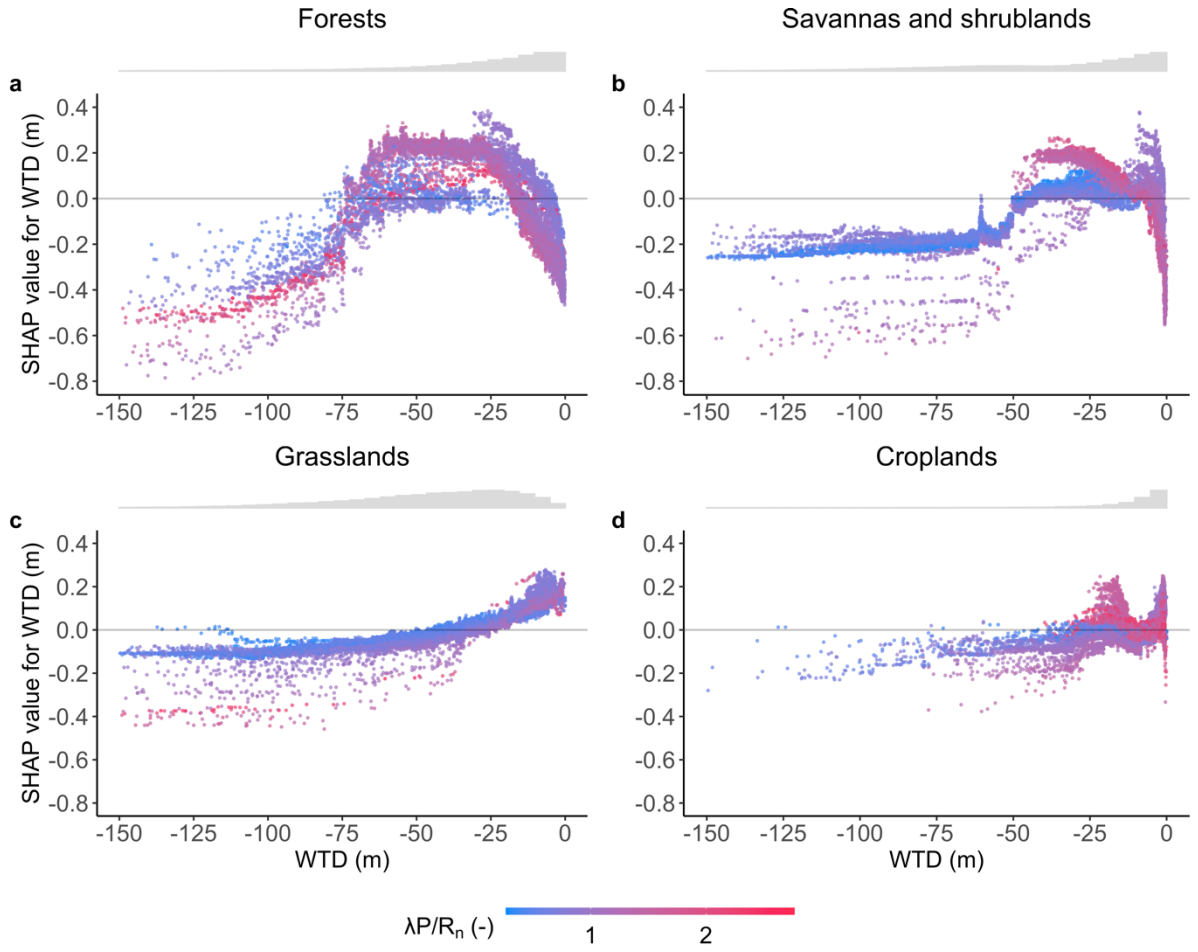

Fig. 5 | SHAP dependence plots of water table depth versus its SHAP value along aridity gradients across plant functional types. SHAP dependence plot show how a specific predictor affects model outcomes while accounting for interaction effects between predictors. Each dot represents the long-term mean of an individual pixel from global datasets. The colour of a dot indicates the magnitude of the moisture index at that location. WTD: water table depth, $\lambda P/R_n$: moisture index. **a**, Forests (evergreen and deciduous, needle-leaved and broadleaved, and mixed forests). **b**, Savannahs and shrublands (savannahs and woody savannahs, open and closed shrublands). **c**, Grasslands. **d**, Croplands. Grey histograms on top of each panel depict marginal density plots for WTD.

The dependence plot of forests and savannahs and shrublands on WTD reveals a range of intermediate WTD values associated with positive SHAP values, indicative of stress-free and positive conditions for vegetation photosynthesis (Fig. 5a,b). Deviating from this range, either towards higher or lower WTD values, respectively corresponds to water-logging or highly arid conditions, resulting in vegetation stress, as evidenced by the corresponding negative SHAP values. These results align with the negative effects of perched water tables on photosynthesis observed under water-logging conditions, which leads to poor oxygenation of the roots[64,65]. Laboratory experiments have shown that GPP and stomatal conductance exhibit negative responses shortly after the onset of water-logging conditions[64,65]. Moreover,

long-term observations in boreal forests have demonstrated that persistent water-logging conditions reduce surface conductance, evapotranspiration (ET), and GPP[66]. Conversely, low WTD values produce a negative effect on SIF sensitivity, as indicated by the corresponding negative SHAP 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]. In both forests and savannahs and shrublands, distinct WTD dependency 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 PFT along the prevailing climatic gradient. In grasslands, the absence of negative effects for shallow WTD values (i.e. positive SHAP even at very shallow WTD in Fig. 5c) aligns with their greater adaptation to water-logging conditions[67]. 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 dependence plot for shallow WTD values is characterized by a noisier signal; this may be caused by the interference of anthropogenic disturbances, such as irrigation, which adds another water source and can at times be coupled to groundwater. 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. 3).

A lower threshold of WTD stress on vegetation activity can be identified across PFTs, below which the relationship between WTD and its SHAP value reaches a plateau (Fig. 5). Below these thresholds, further increases in WTD result in negligible loss in SIF/PAR, as this water is virtually inaccessible even by the deepest rooting systems[28,68,69]. This finding aligns with observations of maximum rooting depths and rooting profiles, indicating a consistent pattern across functional groups, from trees to shrubs and herbaceous plants[28,68,69]. This is also consistent with the maximum observed rooting depth, that for some tree species can go down to 100 meters[32]. Note that the majority of the points is well above these lower thresholds, as indicated by the marginal density plots (i.e. grey histograms in Fig. 5). It is also 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 regulating soil moisture in the unsaturated zone[3].

Conclusion

In this study, we focused on disentangling the effects of aridity versus groundwater on photosynthesis. We thus opted for a simpler approach and selected fewer predictors. Key points in our analysis are the low correlation between WTD and aridity and the high R^2 of our XGBoost models ($R^2 \geq 0.49$ on the test set for all PFT groups, Supplementary Table 1), which contribute to the robustness of our explainable machine-learning approach. In line with this principle of simplicity, we also focused on long-term means, connecting with Budyko's perspective of the long-term partitioning of water and energy. This approach ensures that our analysis remains straightforward yet effective, mirroring the simplicity inherent in Budyko's work. Working with long-term means also has the advantage of avoiding dealing with interannual variability, which can further complicate the relationship with groundwater. Groundwater and aridity modulate the long-term mean of SIF, but it is challenging to quantify this effect with a physical model. In this context, our explainable machine-learning approach proves to be particularly useful. Future research could investigate the role of other drivers (e.g. elevation, atmospheric aridity, irrigation) in this process.

Our results demonstrate that groundwater is nearly as relevant as aridity (quantified by the moisture index) in regulating photosynthesis in forests functional types in CONUS (relative importance of 89%), while its importance diminishes in grasslands (37%), savannahs and shrublands (25%), and croplands (15%) compared to the moisture index. When expanding our focus to a global scale, the influence of groundwater on forest productivity is even more pronounced, exceeding the effect of the moisture index by 5%. Our global analysis also confirms the influence of groundwater in other biomes: in savannahs and shrublands (60%), grasslands (41%) and croplands (15%). Our findings highlight the crucial role of groundwater in modulating ecosystem productivity across biomes, particularly in a world where escalating human groundwater extraction adds further pressure on these ecosystems[1].

Methods

Data sets

SIF, PAR, net radiation, air temperature, precipitation, soil moisture and land cover data are all from global data sets. SIF is from the Sentinel-5 Precursor mission funded by the European Space Agency (ESA)[45]. PAR is from version 6.1 of MCD18C2[46]. Net radiation is calculated using surface net solar and net thermal radiation from ERA5-Land[48] and converted to mm/year multiplying it by the latent heat of vaporization (J kg^{-1}) as a function of air temperature, which is also from ERA5-Land[48]. Precipitation is from version 2.8 of MSWEP[47], which has been shown to outperform other products for hydrological applications[70]. This is consistent with the large positive biases of ERA5-Land precipitation in tropical regions[48]. Soil moisture is from version 5 of SMAP SPL3SMP_E[71]. We obtained pixel-wise information about the dominant plant functional types (PFTs) from a global land cover dataset from MODIS, based on the International Geosphere-Biosphere Programme (IGBP) classification[36]. The water table depth (WTD) was derived from an inverse modelling study that accounted for the coupling between soil water balance and groundwater recharge and discharge, and was benchmarked against satellite observations of Leaf-Area Index (LAI)[32]. A limitation of using these estimates is that the model was run over the 2004-2014 decade, which differs from the record availability of SIF (February 2018 to October 2021). A longer and matching data record would yield more accurate results.

Compared to other products, TROPOMI SIF offers higher spatial resolution, while it suffers from data gaps, especially in mountainous areas and deserts. These missing values can potentially bias our analyses (although we only focus on vegetated land, so this effect should be minor). SIF data was available at a grid resolution of 0.083°x0.083° every 8 days. Since some data sets are available at a lower spatial resolution, for consistency they were converted to a common grid resolution of 0.083°x0.083° using bilinear interpolation. WTD was averaged from 0.0083°x0.0083° to 0.083°x0.083°. We used quality-checked SIF values[45]. We excluded cells that had less than one high-quality data point per month over the period of availability. To focus on the growing season, in each cell we retained SIF values greater than its long-term mean. We applied a similar processing to PAR to match the data gaps present in the SIF time series.

Explainable tree-based models

For each variable, we calculated the mean over the available temporal record in every cell. Net radiation and precipitation were converted to mm/year. We removed points with an annual mean precipitation value greater than 5000 mm/year as considered outliers.

After calculating the ratios of SIF over PAR and precipitation over net radiation (P/Rn), we removed outliers that laid outside of the interval: $\overline{\text{ratio}} \pm 3 * \text{std}(\text{ratio})$.

To test the relative importance of water table depth vs the moisture index in explaining the variance of SIF/PAR, we trained XGBoost models using SIF/PAR as target variable, with $\lambda P/R_n$ and WTD as predictors. We trained one model per each PFT group, defined in accordance with previous studies[58,59]: forests (evergreen and deciduous, needle-leaved and broadleaved, and mixed forests), shrublands and savannahs (savannahs and woody savannahs, open and closed shrublands), grasslands, and croplands.

XGBoost models have recently been employed within the SHAP framework across different scientific fields to quantify feature contribution[58,60,72,73]. These models tend to converge faster compared to neural networks, particularly in regression problems, where each feature holds individual significance[60]. The XGBoost algorithm uses an iterative decision tree model composed of multiple decision trees[74,75] and incorporates shrinkage to prevent overfitting, and column subsampling to accelerate the training process[75]. Here, we applied the SHAP framework to calculate the SHAP value, mean SHAP value and dependence plot between the two features (WTD and $\lambda P/R_n$), to disentangle how the two variables contribute to determining the outcome of SIF/PAR.

We focused on the area of the United States, where WTD data are constrained by more observations and are thus more reliable[32]. The original global data sets were thus cropped to a latitude range of 24°N to 50°N and a longitude range of 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. The XGBoost model hyperparameters were tuned using grid search and tenfold cross-validation and yielded $R^2 \geq 0.49$ on the test set for all PFT groups (Supplementary Table 1). The minimal difference in R^2 values between the training and test sets confirms that our models did not overfit (Supplementary Table 1). We repeated the analysis using all available data at the global scale (Supplementary Table 1 and Supplementary Fig. 2), which showed consistent SHAP values, although the XGBoost models are less accurate (lower R2), as expected.

In this study, we restricted our predictors to WTD and $\lambda P/R_n$ to analyse the relative effects of aridity and groundwater on photosynthesis, in line with our central hypothesis. Since we

focus on long-term photosynthesis, other predictors (e.g. temperature) are not included, because they are very correlated to the local climate over such time scales. Therefore, our objective was not to identify every potential predictor of SIF/PAR for perfect prediction, but only the primary ones. Our XGB models demonstrate robust predictive capability ($R^2 \geq 0.49$ on the test set for all PFT groups), especially considering the inherent challenges in modelling ecosystem processes. Including more predictors could introduce correlated variables (Supplementary Fig. 1c,f,i,l), potentially introducing a confounding effect and undermining the predictive ability of the model[61].

The Pearson correlation coefficient (r) between our predictors, $\lambda P/R_n$ and WTD, is low for Grasslands and Forests (respectively $r = 0.38$ and $r = -0.11$, Supplementary Fig. 1c,i). In contrast, for Croplands, as well as Savannahs and Shrublands, $r \geq 0.52$ (Supplementary Fig. 1f,l). Yet, a closer examination of the scatter plots reveals an absence of a pronounced linear relationship between $\lambda P/R_n$ and WTD, as evidenced by two distinct data clusters in both cases. Comparing the Pearson correlation coefficients between predictors and target variable, SIF/PAR shows a stronger correlation with $\lambda P/R_n$ than with WTD (Supplementary Fig. 1a,b,d,e,g,h,j,k), suggesting that $\lambda P/R_n$ potentially has a greater predictive power. Nonetheless, the average SHAP values of WTD are comparable to those of $\lambda P/R_n$ (Fig. 3), further attesting to the robustness of our machine-learning model in capturing intricate patterns and nonlinearities in our data. It is also important to note that correlation coefficients mainly capture linear relationships; a low correlation does not necessarily indicate a lack of predictiveness, as relationships might be nonlinear.

Data availability

All analyses were performed using R Statistical Software[76]. All data used in this study are openly available.

- TROPOMI SIF: ftp://fluo.gps.caltech.edu/data/tropomi/gridded/SIF740/
- MCD18C2 PAR: https://lpdaac.usgs.gov/products/mcd18c2v061/
- ERA5-Land net radiation and temperature: https://cds.climate.copernicus.eu/cdsapp#!/home
- MSWEP precipitation: http://www.gloh2o.org/mswep/
- SMAP soil moisture: https://nsidc.org/data/spl3smp_e/versions/5
- MODIS land cover: https://lpdaac.usgs.gov/products/mcd12c1v006/
- Modelled water table depth: http://thredds-gfnl.usc.es/thredds/catalog/GLOBALWTDFTP/annualmeans/catalog.html

Acknowledgements

The authors also thank the providers of the other data sets used in this study. The authors would like to acknowledge funding from the LEMONTREE (Land Ecosystem Models based On New Theory, obseRvation and ExpEriments) project, funded through the generosity of Eric and Wendy Schmidt by recommendation of the Schmidt Futures programme as well as the USMILE European Research Council synergy grant, as well as the Learning the Earth with artificial intelligence and Physics (LEAP) Science and Technology Center from the National Science Foundation Award #2019625-STC.

Author Contributions

F.G. wrote the main manuscript in collaboration with P.G. F.G. and P.G. designed the study. F.G. performed the analysis and prepared the figures. All authors reviewed and edited the manuscript.

Competing Financial Interests

The authors declare no competing financial interests.

References

1. Famiglietti, J. S. The global groundwater crisis. *Nature Climate Change* vol. 4 945–948 Preprint at https://doi.org/10.1038/nclimate2425 (2014).
2. Miguez-Macho, G. & Fan, Y. Spatiotemporal origin of soil water taken up by vegetation. *Nature* **598**, 624–628 (2021).
3. Salvucci, G. D. & Entekhabi, D. Equivalent steady soil moisture profile and the time compression approximation in water balance modeling. *Water Resour Res* **30**, 2737–2749 (1994).
4. Seneviratne, S. I. *et al.* Weather and Climate Extreme Events in a Changing Climate. *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* **Cambridge**, (2021).
5. Giardina, F., Gentine, P., Konings, A. G., Seneviratne, S. I. & Stocker, B. D. Diagnosing evapotranspiration responses to water deficit across biomes using deep learning. *New Phytologist* (2023) doi:10.1111/nph.19197.
6. Famiglietti, J. S. *et al.* Satellites measure recent rates of groundwater depletion in California's Central Valley. *Geophys Res Lett* **38**, 2010GL046442 (2011).
7. Oshun, J., Dietrich, W. E., Dawson, T. E. & Fung, I. Dynamic, structured heterogeneity of water isotopes inside hillslopes. *Water Resour Res* **52**, 164–189 (2016).
8. Schwinning, S. The ecohydrology of roots in rocks. *Ecohydrology* **3**, n/a-n/a (2010).
9. Fan, Y., Li, H. & Miguez-Macho, G. Global patterns of groundwater table depth. *Science (1979)* **339**, 940–943 (2013).
10. Lee, J. E., Oliveira, R. S., Dawson, T. E. & Fung, I. Root functioning modifies seasonal climate. *Proc Natl Acad Sci U S A* **102**, 17576–17581 (2005).
11. Seneviratne, S. I. *et al.* Investigating soil moisture-climate interactions in a changing climate: A review. *Earth Sci Rev* **99**, 125–161 (2010).
12. Goedhart, C. M. & Pataki, D. E. Ecosystem effects of groundwater depth in Owens Valley, California. *Ecohydrology* **4**, 458–468 (2011).
13. Fan, Y. Groundwater in the Earth's critical zone: Relevance to large-scale patterns and processes. *Water Resour Res* **51**, 3052–3069 (2015).
14. Clark, M. P. *et al.* Improving the representation of hydrologic processes in Earth System Models. *Water Resour Res* **51**, 5929–5956 (2015).
15. Maxwell, R. M. & Condon, L. E. Connections between groundwater flow and transpiration partitioning. *Science (1979)* **353**, 377–380 (2016).
16. Swenson, S. C., Clark, M., Fan, Y., Lawrence, D. M. & Perket, J. Representing Intrahillslope Lateral Subsurface Flow in the Community Land Model. *J Adv Model Earth Syst* **11**, 4044–4065 (2019).
17. Dawson, T. E. & Pate, J. S. Seasonal water uptake and movement in root systems of Australian phraeatophytic plants of dimorphic root morphology: A stable isotope investigation. *Oecologia* **107**, 13–20 (1996).
18. Voltas, J., Lucabaugh, D., Chambel, M. R. & Ferrio, J. P. Intraspecific variation in the use of water sources by the circum-Mediterranean conifer *Pinus halepensis*. *New Phytologist* **208**, 1031–1041 (2015).
19. Grossiord, C. *et al.* Prolonged warming and drought modify belowground interactions for water among coexisting plants. *Tree Physiol* **39**, 55–63 (2018).
20. Stocker, B. D. *et al.* Global patterns of water storage in the rooting zones of vegetation. *Nat Geosci* **16**, 250–256 (2023).
21. Rempe, D. M. & Dietrich, W. E. Direct observations of rock moisture, a hidden component of the hydrologic cycle. *Proc Natl Acad Sci U S A* **115**, 2664–2669 (2018).
22. McCormick, E. L. *et al.* Widespread woody plant use of water stored in bedrock. *Nature* **597**, 225–229 (2021).
23. Querejeta, J. I., Estrada-Medina, H., Allen, M. F. & Jiménez-Osornio, J. J. Water source partitioning among trees growing on shallow karst soils in a seasonally dry tropical climate. *Oecologia* **152**, 26–36 (2007).

24. Evaristo, J. & McDonnell, J. J. Prevalence and magnitude of groundwater use by vegetation: A global stable isotope meta-analysis. *Sci Rep* **7**, 1–12 (2017).
25. Barbeta, A. & Peñuelas, J. Relative contribution of groundwater to plant transpiration estimated with stable isotopes. *Sci Rep* **7**, 1–10 (2017).
26. Jobbágy, E. G., Nosetto, M. D., Villagra, P. E. & Jackson, R. B. Water subsidies from mountains to deserts: their role in sustaining groundwater-fed oases in a sandy landscape. *Ecological Applications* **21**, 678–694 (2011).
27. Stocker, B. , *et al.* Towards a better understanding of deep belowground water stores and their influence on land-atmosphere exchange and drought impacts. *EGU General Assembly 2023* **EGU23-9503**, (2023).
28. Tumber-Dávila, S. J., Schenk, H. J., Du, E. & Jackson, R. B. Plant sizes and shapes above and belowground and their interactions with climate. *New Phytologist* **235**, 1032–1056 (2022).
29. Elmore, A. J., Manning, S. J., Mustard, J. F. & Craine, J. M. Decline in alkali meadow vegetation cover in California: The effects of groundwater extraction and drought. *Journal of Applied Ecology* **43**, 770–779 (2006).
30. Howard, J. & Merrifield, M. Mapping Groundwater Dependent Ecosystems in California. *PLoS One* **5**, e11249 (2010).
31. Sutanto, S. J. *et al.* A perspective on isotope versus non-isotope approaches to determine the contribution of transpiration to total evaporation. *Hydrol Earth Syst Sci* **18**, 2815–2827 (2014).
32. Fan, Y., Miguez-Macho, G., Jobbágy, E. G., Jackson, R. B. & Otero-Casal, C. Hydrologic regulation of plant rooting depth. *Proc Natl Acad Sci U S A* **114**, 10572–10577 (2017).
33. Joiner, J. *et al.* Global monitoring of terrestrial chlorophyll fluorescence from moderate-spectral-resolution near-infrared satellite measurements: methodology, simulations, and application to GOME-2. *Atmos Meas Tech* **6**, 2803–2823 (2013).
34. Frankenberg, C. *et al.* New global observations of the terrestrial carbon cycle from GOSAT: Patterns of plant fluorescence with gross primary productivity. *Geophys Res Lett* **38**, 1–6 (2011).
35. Green, J. K. *et al.* Regionally strong feedbacks between the atmosphere and terrestrial biosphere. *Nature Geosci* **advance on**, (2017).
36. Friedl, M. A. *et al.* MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sens Environ* **114**, 168–182 (2010).
37. Budyko, M. I. *Climate and Life*. (Academic Press, 1974).
38. Budyko, M. I. *Тепловой баланс земной поверхности (Heat Balance of the Earth's Surface)*. (Gidrometeoizdat, 1956).
39. Condon, L. E. & Maxwell, R. M. Systematic shifts in Budyko relationships caused by groundwater storage changes. *Hydrol Earth Syst Sci* **21**, 1117–1135 (2017).
40. Istanbulluoglu, E., Wang, T., Wright, O. M. & Lenters, J. D. Interpretation of hydrologic trends from a water balance perspective: The role of groundwater storage in the Budyko hypothesis. *Water Resour Res* **48**, (2012).
41. Greve, P., Gudmundsson, L., Orlowsky, B. & Seneviratne, S. I. Introducing a probabilistic Budyko framework. *Geophys Res Lett* **42**, 2261–2269 (2015).
42. Maes, W. H. *et al.* Sun-induced fluorescence closely linked to ecosystem transpiration as evidenced by satellite data and radiative transfer models. *Remote Sens Environ* **249**, 112030 (2020).
43. Pagán, B., Maes, W., Gentine, P., Martens, B. & Miralles, D. Exploring the Potential of Satellite Solar-Induced Fluorescence to Constrain Global Transpiration Estimates. *Remote Sens (Basel)* **11**, 413 (2019).
44. Jonard, F. *et al.* Value of sun-induced chlorophyll fluorescence for quantifying hydrological states and fluxes: Current status and challenges. *Agricultural and Forest Meteorology* vol. 291 108088 Preprint at https://doi.org/10.1016/j.agrformet.2020.108088 (2020).
45. Köhler, P. *et al.* Global Retrievals of Solar-Induced Chlorophyll Fluorescence With TROPOMI: First Results and Intersensor Comparison to OCO-2. *Geophys Res Lett* **45**, 10,456-10,463 (2018).
46. Wang, D. MODIS/Terra+Aqua Photosynthetically Active Radiation Daily/3-Hour L3 Global 0.05Deg CMG V061 [Data set]. *NASA EOSDIS Land Processes DAAC* (2021).

47. Beck, H. E. *et al.* MSWEP V2 Global 3-hourly 0.1° Precipitation: Methodology and Quantitative Assessment. *Bull Am Meteorol Soc* **100**, 473–500 (2019).
48. Hersbach, H. *et al.* The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society* **146**, 1999–2049 (2020).
49. Gentine, P., D'Odorico, P., Lintner, B. R., Sivandran, G. & Salvucci, G. Interdependence of climate, soil, and vegetation as constrained by the Budyko curve. *Geophys Res Lett* **39**, 2–7 (2012).
50. Gudmundsson, L., Greve, P. & Seneviratne, S. I. The sensitivity of water availability to changes in the aridity index and other factors-A probabilistic analysis in the Budyko space. *Geophys Res Lett* **43**, 6985–6994 (2016).
51. Lintner, B. R., Gentine, P., Findell, K. L. & Salvucci, G. D. The Budyko and complementary relationships in an idealized model of large-scale land-atmosphere coupling. *Hydrol Earth Syst Sci* **19**, 2119–2131 (2015).
52. Prentice, I. C. *et al.* Evidence of a universal scaling relationship for leaf CO2 drawdown along an aridity gradient. *New Phytologist* **190**, 169–180 (2011).
53. Greve, P., Roderick, M. L. & Seneviratne, S. I. Simulated changes in aridity from the last glacial maximum to 4xCO2. *Environmental Research Letters* **12**, 114021 (2017).
54. Manabe, S. *Climate and ocean circulation. Part I: the atmospheric circulation and the hydrology of the earth's surface.* vol. 97 (Mon. Weather Rev., 1969).
55. Cutler, D. R. *et al.* Random forests for classification in ecology. *Ecology* **88**, 2783–2792 (2007).
56. Hutengs, C. & Vohland, M. Downscaling land surface temperatures at regional scales with random forest regression. *Remote Sens Environ* **178**, 127–141 (2016).
57. Green, J. K. *et al.* Surface temperatures reveal the patterns of vegetation water stress and their environmental drivers across the tropical Americas. *Glob Chang Biol* **28**, 2940–2955 (2022).
58. Wang, H. *et al.* Exploring complex water stress–gross primary production relationships: Impact of climatic drivers, main effects, and interactive effects. *Glob Chang Biol* **28**, 4110–4123 (2022).
59. Fernández-Martínez, M. *et al.* The role of climate, foliar stoichiometry and plant diversity on ecosystem carbon balance. *Glob Chang Biol* **26**, 7067–7078 (2020).
60. Lundberg, S. M. *et al.* From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell* **2**, 56–67 (2020).
61. Lundberg, S. M. & Lee, S.-I. A Unified Approach to Interpreting Model Predictions. *Adv Neural Inf Process Syst* **30**, 4768–4777 (2017).
62. Tumber-Dávila, S. J., Schenk, H. J., Du, E. & Jackson, R. B. Plant sizes and shapes above and belowground and their interactions with climate. *New Phytologist* **235**, 1032–1056 (2022).
63. Jasechko, S., Seybold, H., Perrone, D., Fan, Y. & Kirchner, J. W. Widespread potential loss of streamflow into underlying aquifers across the USA. *Nature* **591**, 391–395 (2021).
64. Terazawa, K., Maruyama, Y. & Morikawa, Y. Photosynthetic and stomatal responses of Larix kaempferi seedlings to short-term waterlogging. *Ecol Res* **7**, 193–197 (1992).
65. Terazawa, K. & Kikuzawa, K. Effects of flooding on leaf dynamics and other seedling responses in flood-tolerant Alnus japonica and flood-intolerant Betula platyphylla var. japonica. *Tree Physiol* **14**, 251–261 (1994).
66. Ohta, T. *et al.* Effects of waterlogging on water and carbon dioxide fluxes and environmental variables in a Siberian larch forest, 1998-2011. *Agric For Meteorol* **188**, 64–75 (2014).
67. Fan, Y. *et al.* Hillslope Hydrology in Global Change Research and Earth System Modeling. *Water Resour Res* **55**, 1737–1772 (2019).
68. Canadell, J. *et al.* Maximum rooting depth of vegetation types at the global scale. *Oecologia* **108**, 583–595 (1996).
69. Jackson, R. B. *et al.* A global analysis of root distributions for terrestrial biomes. *Oecologia* vol. 108 389–411 Preprint at https://doi.org/10.1007/BF00333714 (1996).
70. Tarek, M., Brissette, F. P. & Arsenault, R. Large-scale analysis of global gridded precipitation and temperature datasets for climate change impact studies. *J Hydrometeorol* **21**, 2623–2640 (2020).

71. O'Neill, P. E. , *et al.* SMAP Enhanced L3 Radiometer Global and Polar Grid Daily 9 km EASE-Grid Soil Moisture, Version 5 [Data Set]. *Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center.* https://doi.org/10.5067/4DQ54OUIJ9DL (2021).
72. Yang, C., Chen, M. & Yuan, Q. The application of XGBoost and SHAP to examining the factors in freight truck-related crashes: An exploratory analysis. *Accid Anal Prev* **158**, 106153 (2021).
73. Guo, M. *et al.* Older Pedestrian Traffic Crashes Severity Analysis Based on an Emerging Machine Learning XGBoost. *Sustainability* **13**, 926 (2021).
74. Yan, F., Song, K., Liu, Y., Chen, S. & Chen, J. Predictions and mechanism analyses of the fatigue strength of steel based on machine learning. *J Mater Sci* **55**, 15334–15349 (2020).
75. Chen, T. & Guestrin, C. XGBoost: A scalable tree boosting system. in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* vols 13-17-August-2016 785–794 (Association for Computing Machinery, 2016).
76. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/. (2023).