Geophysical Research Letters

RESEARCH LETTER
10.1029/2021GL092856

Revisiting Global Vegetation Controls Using Multi-Layer Soil Moisture

Wantong Li[1], Mirco Migliavacca[1], Matthias Forkel[2], Sophia Walther[1], Markus Reichstein[1], and René Orth[1]

[1]Department of Biogeochemical Integration, Max Planck Institute for Biogeochemistry, Jena, Germany, [2]Technische Universität Dresden, Institute of Photogrammetry and Remote Sensing, Dresden, Germany

Key Points:

- Vertically resolved soil moisture (SM) improves the understanding of large-scale vegetation productivity and yields extended water-related controls
- Data-driven evidence for the meaningfulness of the long-standing modeling paradigm of vertical soil layer discretization
- Comparatively deep SM is most relevant in semi-arid areas and for grasses and shrubs

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

W. Li,
wantong@bgc-jena.mpg.de

Citation:

Li, W., Migliavacca, M., Forkel, M., Walther, S., Reichstein, M., & Orth, R. (2021). Revisiting global vegetation controls using multi-layer soil moisture. *Geophysical Research Letters*, *48*, e2021GL092856. https://doi.org/10.1029/2021GL092856

Received 12 FEB 2021
Accepted 18 MAY 2021

Abstract The productivity of terrestrial vegetation is determined by multiple land surface and atmospheric drivers. Water availability is critical for vegetation productivity, but the role of vertical variability of soil moisture (SM) is largely unknown. Here, we analyze dominant controls of global vegetation productivity represented by sun-induced fluorescence and spectral vegetation indices at the half-monthly time scale. We apply random forests to predict vegetation productivity from several hydrometeorological variables including multi-layer SM and quantify the variable importance. Dominant hydrometeorological controls generally vary with latitudes: temperature in higher latitudes, solar radiation in lower latitudes, and root-zone SM in between. We find that including vertically resolved SM allows a better understanding of vegetation productivity and reveals extended water-related control. The deep(er) SM control for semi-arid grasses and shrubs illustrates the potential of deep(er) rooting systems to adapt to water limitation. This study highlights the potential to infer sub-surface processes from remote sensing observations.

1. Introduction

Terrestrial vegetation couples the global water and carbon cycles between the atmosphere and the land surface. Vegetation productivity is determined by a multitude of hydrometeorological variables (Monteith & Unsworth, 1990; Nemani et al., 2003; Piao et al., 2020). The underlying relationships are complex in time and space because of interactions among variables (Cox et al., 2013; Garonna et al., 2018; Pearson et al., 2013) and short-term variations of hydrometeorology potentially influencing ecosystems through nonlinear effects (De Keersmaecker et al., 2015; Fatichi & Ivanov, 2014; Paschalis et al., 2015; Reichstein et al., 2013). The hydrometeorological controls of anomalies in global vegetation productivity are still not fully understood at the half-month time scale, because underlying variables were not investigated comprehensively. For example, previous research often used hand-designed approaches to represent the plant-available water (e.g., lagging precipitation in specific months). The knowledge gap contributes to uncertainties in assessing the sensitivity and resilience of ecosystems to different climate drivers (Sakschewski et al., 2016), and in future climate projections (Duveiller et al., 2018; Feng et al., 2014; Novick et al., 2016).

Previous studies investigated dominant hydrometeorological controls of vegetation productivity at a global scale and across different ecosystems (Beer et al., 2010; Jung et al. 2011, 2017; Li & Xiao, 2020; Madani et al., 2017; Seddon et al., 2016; Walther et al., 2019; D. Wu et al., 2015). While these studies and recent gross primary production estimates agree that vegetation in (semi-)arid area is significantly impacted by soil moisture (SM) (Stocker et al., 2018, 2020), a corresponding global analysis including the impact of SM from multiple depths is lacking. Several studies have already highlighted the local relevance of multi-layer SM to ecosystems: root water uptake from deeper soil layers can help mitigate water stress and maintain plant transpiration (Migliavacca et al., 2009; Schulze et al., 1996); Yinglan et al. (2019) demonstrated divergent relative importance of surface SM versus deeper SM depending on land cover types; and Schlaepfer et al. (2017) simulated lowered sub-surface SM compared to surface SM which largely impact vegetation dynamics in temperate drylands. This way, distinguishing shallow and deep SM is expected to allow for a more accurate identification of global vegetation controls as the accessibility and availability of water for plants varies in space and time. For this purpose, the state-of-the-art ERA5 reanalysis provides multi-layer SM (Hersbach et al., 2020) that has been successfully applied in hydrometeorological studies (Kolluru et al., 2020; Li, Wu, & Ma, 2020; Tarek et al., 2020).

Reliable observation-based global photosynthesis proxies are available for recent years through satellite-derived sun-induced fluorescence (SIF, Frankenberg et al., 2011; Joiner et al., 2013). SIF data are increasingly used to study the relationships between global vegetation productivity and hydrometeorological drivers (Chen et al., 2020; Jiao et al., 2019; Li, Xiao, et al., 2020; Wagle et al., 2016; Walther et al., 2019; Yang et al., 2015; Zuromski et al., 2018). Besides, spectral vegetation indices and biophysical parameters from multi-spectral satellite instruments are widely used to study drivers of vegetation phenology and productivity (Buermann et al., 2018; Forkel et al. 2015). In this study, we consider SIF alongside two spectral indices (the normalized difference vegetation index, NDVI, Tucker, 1979; and near-infrared reflectance of terrestrial vegetation, NIRv, Badgley et al., 2017), and a comprehensive set of explanatory variables representing energy (temperature; radiation; vapor pressure deficit, VPD) and water availability (precipitation; multi-layer SM) to revisit global photosynthesis and greenness controls.

2. Data and Methods

2.1. Remote Sensing Proxies of Vegetation Productivity and Greenness

2.1.1. SIF

SIF is used as a proxy for variations in photosynthesis because it captures radiation emitted by chlorophyll molecules and is related to photosynthetic activity. We use one of the longest available satellite-derived SIF records which is based on the Global Ozone Monitoring Experiment–2 (GOME-2) instrument and ranges from 2007 to 2018 (Köhler et al., 2015). The raw global SIF observations at daily time scale are filtered to remove data based on (i) high solar zenith angles (>70°), (ii) large differences to the normal local overpass time (2 p.m.–8 a.m. in the next day), and (iii) large cloud cover (>50%), as done by Köhler et al. (2015). We note that different levels of cloud filtering of the satellite retrievals yield similar SIF anomalies, such that a cloud cover threshold of 50% is a reasonable compromise between SIF data quality and quantity (Köhler et al., 2015).

2.1.2. Vegetation Indices

To complement the photosynthesis analysis we use NDVI and NIRv as spectral vegetation indices (Badgley et al., 2017; Huete et al., 2002). NIRv is defined as NDVI multiplied by the near-infrared reflectance (Badgley et al., 2017). We obtain red and near-infrared reflectances from MOD13C1 v006 product (https://lpdaac.usgs.gov/products/mod13c1v006/) in an original 16-day and 0.05° resolution. NDVI and NIRv are computed from data with quality flags 0 and 1 representing good and marginal data, thereby ignoring low-quality data.

2.2. Hydrometeorological Data

We consider a comprehensive selection of energy and water-related variables from the ERA5 reanalysis. This dataset accounts for spatial variations in soil types and assimilates spaceborne microwave instruments of surface SM (Balsamo et al., 2009; Hersbach et al., 2020). Even though SM estimates from deeper layers in ERA5 are less constrained by observations, previous studies illustrated the agreement of ERA5 SM data in each layer with independent observations (Albergel et al., 2012, 2013, 2018; Li, Wu, & Ma, 2020; Liu et al., 2013; Hersbach et al., 2020; Jing et al., 2018; see Text S1 for details). Energy-related variables include temperature at 2-m height (temperature), surface downward solar radiation (solar radiation) and VPD, and the water-related variables are total precipitation (precipitation), SM layer 1 (0–7 cm), layer 2 (7–28 cm), layer 3 (28–100 cm) and layer 4 (100–289 cm). For comparison, we compute total SM by averaging values across the individual layers weighted by their thickness. It is to note that VPD is related to the relative humidity and temperature. Hence, it is an energy-related variable but representing the demand of the water in the atmosphere.

2.3. Climate and Vegetation Regimes

To evaluate the results of our analyses, we compute the aridity index for each grid cell as the ratio between the long-term averages of net radiation (expressed in mm) and precipitation using ERA5 data (Budyko, 1974).

We distinguish climate regimes using long-term mean temperatures and aridity index. We use fractional vegetation cover data from the AVHRR vegetation continuous fields products (VCF5KYR, https://lpdaac.usgs.gov/products/vcf5kyrv001/) from 2007 to 2016 to classify the fraction of total vegetation cover, and the fraction of tree cover in total vegetation cover (Song et al., 2018).

2.4. Methods

2.4.1. Data Pre-processing

The data pre-processing is illustrated in Figure S1. All vegetation data and hydrometeorological data are aggregated to 0.5° spatial and half-monthly temporal resolution where SIF is available, and 16-day original NDVI and NIRv are linearly interpolated to half-monthly resolution. The first half-month consists of the first 15 days of the month, and the second half-month consists of the remaining days in the respective month. The study period is limited to 2007–2018 because of the SIF availability. In all SIF-based analyses we focus on data with SIF > 0.5 $mW/m^2/sr/nm$ to filter out sparse or dormant vegetation. This filtering is also applied in the NDVI and NIRv analyses, where additionally negative NDVI and NIRv values are filtered out. Grid cells are only considered in the analysis if more than 15 data points are left after filtering, and if the vegetation cover fraction exceeds 5%. For all target and predictor variables, we obtain half-monthly anomalies by subtracting the mean seasonal cycles determined by averaging values from all years for each of the 24 half-monthly periods between the first half of January and the second half of December. As we focus on relationships at short time scales, we remove long-term trends for each grid cell determined by a locally weighted smoothing filter (Cleveland et al., 1979) with a smoothing span of 0.4. Moreover, this helps to filter out any signal introduced by potential satellite sensor degradation.

2.4.2. Identification of Relative Importance of Hydrometeorological Variables Using Random Forests (RF)

RF is a nonparametric regression-based methods requiring no statistical assumptions on predictor and target variables, and is designed to process large amounts of diverse input data (Breiman, 2001). With this flexibility it is better placed than traditional statistical methods to detect nonlinear relationships among short-term hydrometeorological variations and vegetation productivity. We also note that our RF analysis can indicate plausible governing processes from emergent relationships, but by construction it does not suggest causality. In this study, all hydrometeorological anomalies are used as predictor variables, and anomalies of SIF and vegetation indices are employed as target variables per each grid cell, respectively (Figure S1). RF training is done using information from each grid cell and the surrounding grid cells (forming 3×3 grid cell matrices) to obtain sufficient data. The performance of the RF model is then evaluated by the fraction of explained variance in regression analysis carried out with the linear least squares using cross-validation (hereafter referred to as R^2; see details in Text S2.1). Grid cells with R^2 lower than or equal to 0 are filtered out. Finally, two experiments are performed with RF models differing in the SM data used: total versus multi-layer SM alongside precipitation, VPD, solar radiation, and temperature.

Permutation importance is one of the most common methods for RF to measure the relative importance of each predictor variable (Lunetta et al., 2004; Nicodemus, 2011; Zhang & Yang, 2020). It is inferred from the difference of error before and after a temporal permutation applied to the particular variable (Cutler et al., 2012; Gómez-Ramírez et al., 2020). To validate results of permutation importance we employ two more methods and to infer confounding effects we also quantify the sensitivity of SIF response to each predictor variable (see Text S2.1 for variable importance identification methods, Text S2.2 for sensitivity algorithm).

When considering multiple hydro-meteorological variables, the identification of global vegetation controls is challenged by potential high collinearity (Dormann et al., 2013) between some of the variables. Most previous studies did not consider more than three variables, thereby somewhat circumventing this problem while ignoring potentially important variables (Claessen et al., 2019; Garonna et al., 2018; Li & Xiao, 2020; Seddon et al., 2016). Though RF are also challenged by the collinearity in the input data, they yield more robust results with noisy training data and have high interpretability (Zhang & Yang, 2020), for example by inferring the sign of the sensitivity of vegetation productivity to particular predictor variables. As the

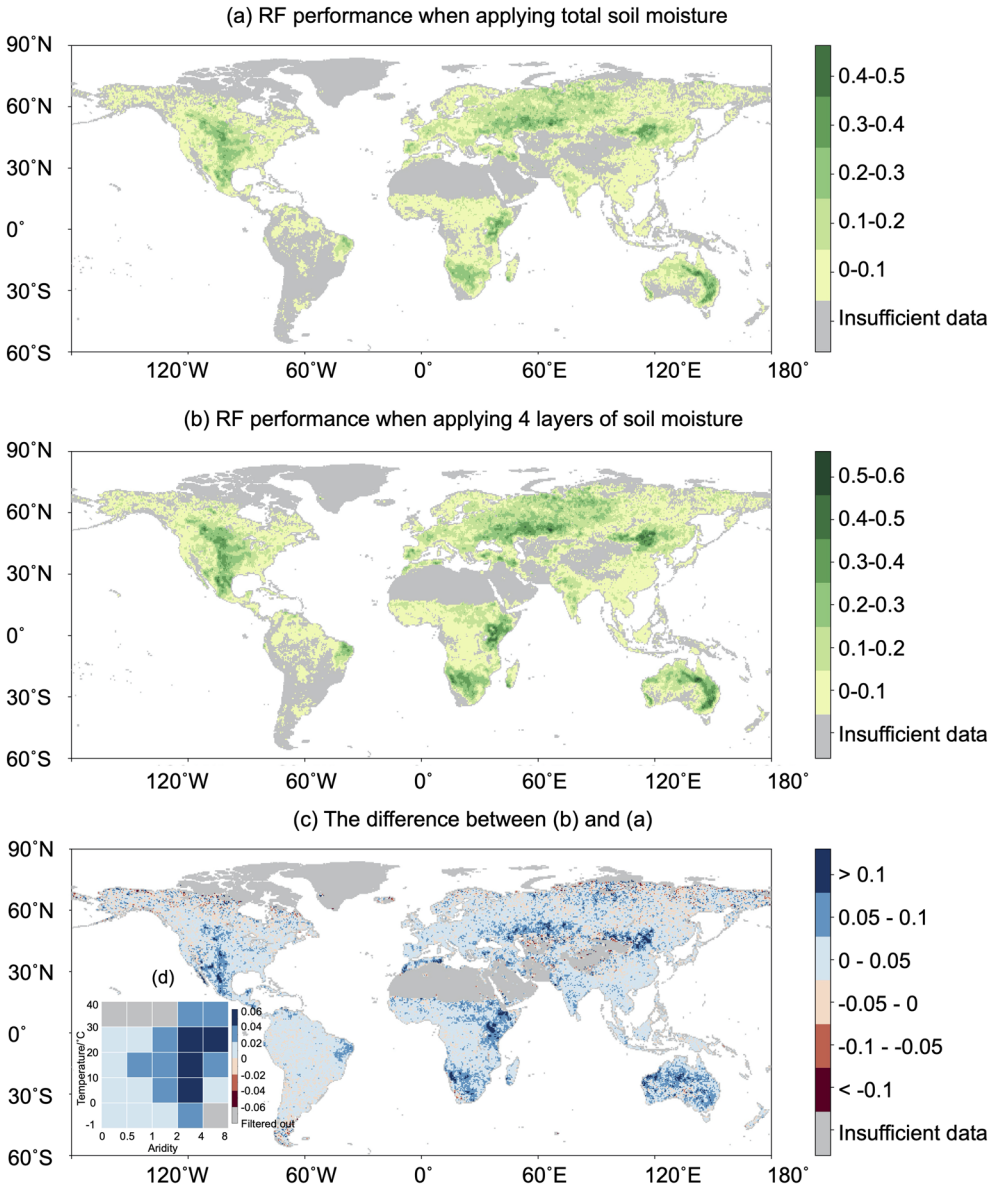

Figure 1. RF model performance (R^2) in predicting sun-induced fluorescence (SIF) when using (a) total and (b) multi-layer soil moisture (SM). (c) is the difference between (b) and (a); and (d) summarizes (c) across climate regimes (i.e., temperature and aridity).

computed relationships and their signs are consistent with previous literature and physical expectations (see Section 3), we note that RF are applicable in our multi-variate context. Further, collinearity can sometimes be mitigated through a pre-processing of the data by removing long-term trends or the mean seasonal cycle (see Section 2.4.1), as long as the collinearity is mainly driven by them, that is trends or seasonal cycles would be similar between variables while shorter-term dynamics are not.

3. Results and Discussion

3.1. Model Performance

The spatial patterns of RF model performance are similar between the two experiments using total and multi-layer SM with comparatively high R^2 (>0.3) in the central North America, central Eurasia, southern and eastern Africa, central Asia, and eastern Australia (Figure 1). The predictive performance is improved

in most regions across the globe when using multi-layer instead of total SM. Vertical SM information improves model performance particularly in semi-arid regions such as Australia, central North America and central Asia (Figures 1c and 1d). In these regions plant rooting systems apparently adapt to compensate for local water deficits which arise from divergent dynamics of surface and root-zone SM across time and space (Berg et al., 2017; Lian et al., 2020; Schlaepfer et al., 2017; Zhang et al., 2016).

Although the performance of SIF prediction is improved with multi-layer SM, the R^2 values are still relatively low, especially in large regions in South America. This relates to a typically significantly decreased model performance in predicting global vegetation productivity for anomalies compared to absolute data and mean seasonal cycles (Kraft et al., 2019). Further, it is related to input data quality where the satellite-based SIF retrievals are strongly impacted by noises in large regions in South America (Joiner et al., 2013; Köhler et al., 2015). Furthermore, using relatively coarser GOME2 pixels to derive the SIF data includes more residual cloud contamination than for example the finer-scale MODIS footprints (Joiner et al., 2013). Therefore, our RF models show much better performance in the case of NDVI and NIRv (Figure S2) compared with SIF. In this context the performance of RF is generally lower in humid regions compared to drier regions, related to increased cloud cover which degrades the quality of the satellite retrievals as previously showed (Kraft et al., 2019; Linscheid et al., 2020). We include regions with weak model performance in the case of SIF for the subsequent analyses, and we believe that our methodology is robust, because (i) our main goal is to rank the relative importance of predictors instead of accurately capturing their dynamics; (ii) we find strongly similar global patterns of main controlling variables across SIF, NIRv and NDVI; and (iii) controlling patterns are largely in line with previous studies (Madani et al., 2017; Nemani et al., 2003; Seddon et al., 2016).

Additionally, we verify that the efficiency of using multi-layer SM in RF is not an artifact of over-fitting (Figures S3 and S4), and we find similar results when using the root fractions in each soil layer as weights in the total SM vertical average computation rather than the layer depths (Figure S5; root fractions are derived from ERA5 data by following ECMWF, 2020). See Text S3 for uncertainties in vegetation data and model tests in details.

3.2. Main Hydrometeorological Controls on Global Vegetation Productivity

The global partitioning of water- vs energy-related controls of vegetation productivity significantly varies between the two experiments involving total and multi-layer SM (Figure 2). Apparently, total SM does not provide sufficient information to the RF model to detect water-controlled regions (Figure 2a). Overall, in the multi-layer SM analysis (Figure 2b) temperature is identified as the main driver of SIF in the higher northern latitudes, solar radiation dominantly controls SIF in tropical regions, and VPD emerges as a main control in parts of the western Amazon forests (Green et al., 2020), eastern North America, northern Eurasia and eastern Asia. In between the tropics and the higher latitudes, where mostly (semi-)arid climate regimes are prevailing, water-related variables play the dominant role. Precipitation and surface SM (0–7 cm in ERA5) control SIF in central India, western Sahel and in central-southern Africa. Root-zone SM (7–28 cm in ERA5) mainly controls SIF in southern North America, southern Europe, and many parts of Eurasia, India and Australia. In general, root-zone SM (7–28 cm in ERA5) emerges as the most relevant water reservoir for vegetation productivity, while deeper SM (28–100 cm in ERA5) is particularly important in the transitional zones and temperate dry regions, such as central North America and southern Europe. To further test the assumption of energy variables misleadingly being detected as the main control where surface SM is the actual main driver due to the insufficiency of the total SM experiment (energy variables negatively contribute to the variation of SIF), we repeat the analysis from Figure 2 while only considering variables with positive contributions to SIF prediction. The result confirms our hypothesis and illustrates that confounding effects can be minimized using multi-layer SM (Figure S6; see Text S4.1).

Confirming these SIF-based results, we find similar global patterns of the main controlling variables in the case of NIRv and NDVI (Figure S7), even though they show extended SM-controlled regions. Furthermore, indicating physical meaningfulness of the obtained global patterns, we find that the sensitivities of SIF to the respective diagnosed hydrometeorological main controls are typically strongly positive (Figure S8). This is also true for precipitation when it is identified as the most important predictor.

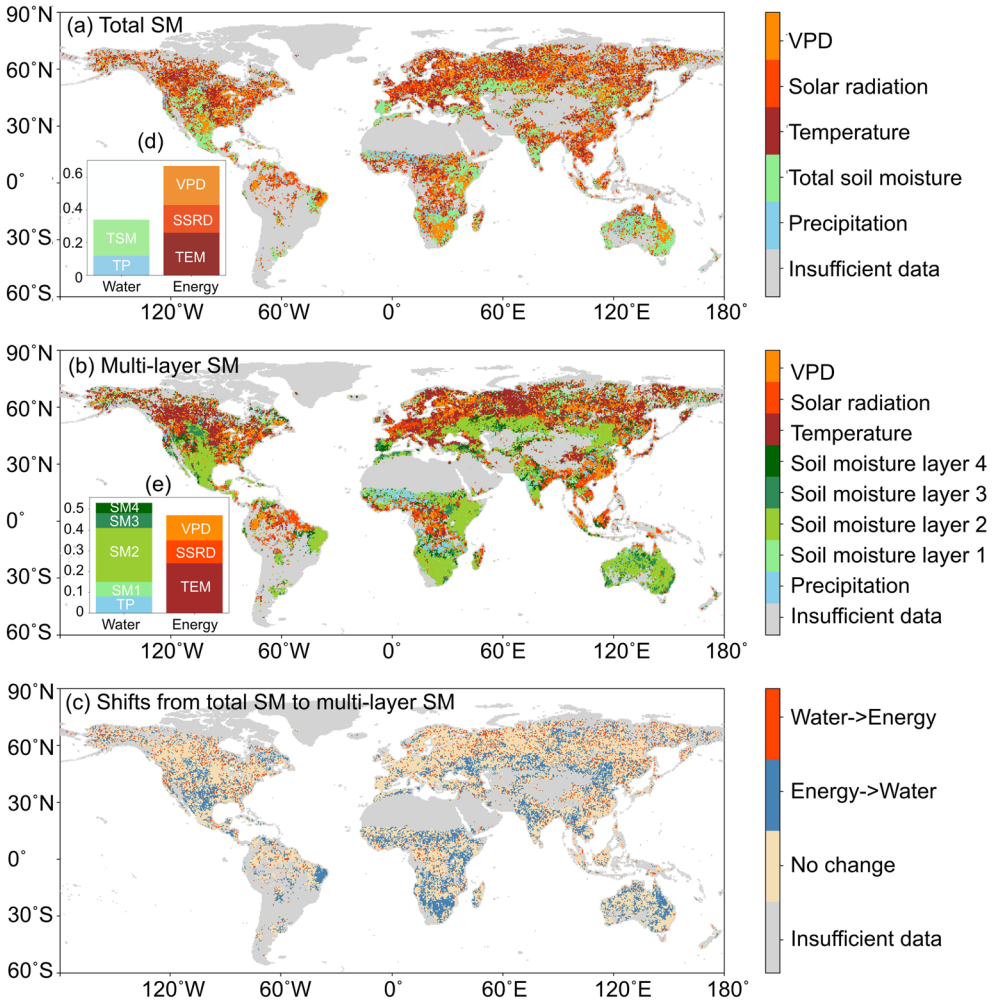

Figure 2. Main hydrometeorological controls on sun-induced fluorescence (SIF) using (a) total and (b) multi-layer soil moisture (SM). (c) Shifts between the energy and water controls from (a) to (b). Proportions of study area where each variable is the most important factor are shown in (d and e). TP denotes precipitation; TSM denotes total SM; SM1, 2, 3, 4 denote SM in layers 1, 2, 3, 4 respectively; TEM denotes temperature; SSRD denotes solar radiation; VPD denotes vapor pressure deficit.

Next, we analyze the prevailing main controls across climate and vegetation regimes. Overall, controlling patterns across climate regimes in Figure 3a are in line with first-order constraints for evapotranspiration from Seneviratne et al. (2010), and from Denissen et al. (2020) in a European study. Root-zone SM (7–28 cm in EAR5) is identified as the main control in (semi-)arid regions. In humid regions energy variables are the most relevant, and overall temperature is the most important while solar radiation also plays a role. Solar radiation best explains SIF variability for forests in humid regions, because SIF is mechanistically linked through absorbed photosynthetically active radiation which depends on the variability of the irradiance (Li & Xiao, 2020). In extreme warm and dry conditions, precipitation is identified as the dominant control. This could be explained as (usually weak) rainfall would not significantly wet the soil surface because of quick evaporation from warm surfaces (Seneviratne et al., 2010) and soil water repellency after long dry periods (Song & Wang, 2019). The rain-induced evaporation in turn mitigates atmospheric water stress (high VPD), and thereby contributes to the precipitation control on SIF. In Figures 3d–3f we show that grasses and shrubs with a low fraction of tree cover are most water-controlled, regions with intermediate tree cover are typically temperature-controlled, and regions with the highest tree cover are mostly radiation-controlled. Energy controls involve a relatively lower vulnerability of tree ecosystems to droughts than other ecosystems (Huang & Xia, 2019), as droughts are typically associated with above-average solar radiation

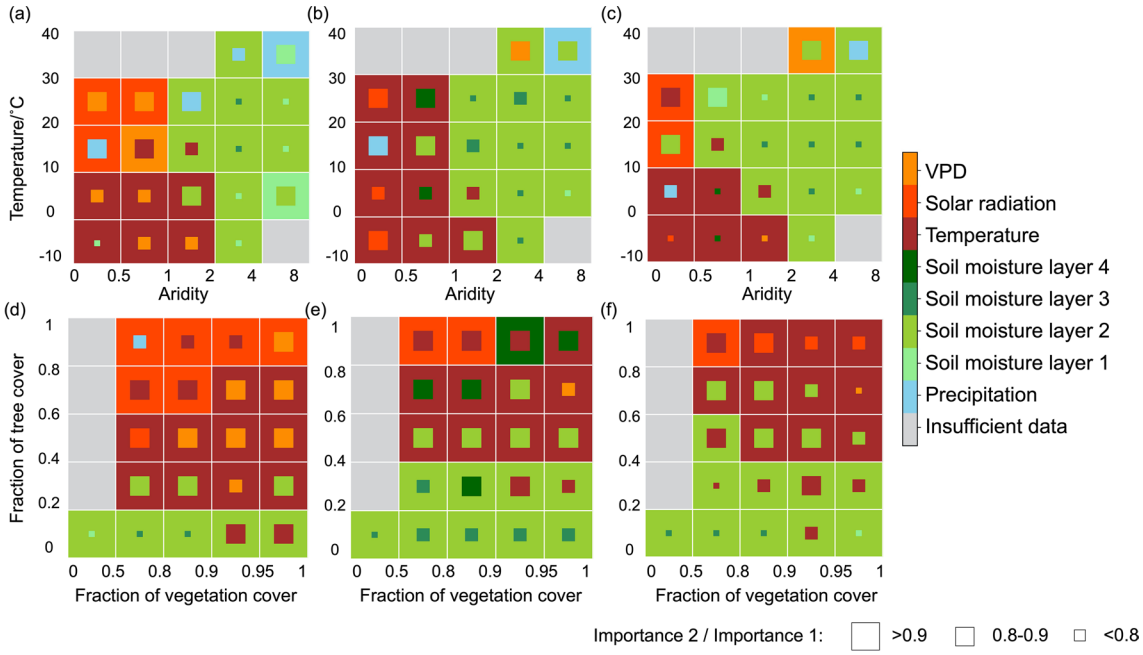

Figure 3. Main hydrometeorological controls on (a), (d) sun-induced fluorescence (SIF) (b), (e) NIRv and (c), (f) normalized difference vegetation index (NDVI) across climate and vegetation regimes. The most important variables are indicated by the color of the temperature-aridity and tree-vegetation boxes, respective second most important variables are denoted by the inner square color, where the size indicates the relative importance compared to the most important variable. Temperature-aridity and tree-vegetation boxes containing less than 10 available data points (grid cells) are shown in gray. The aridity index and the fraction of vegetation cover are visualized by nonlinear sequences following their skewed distributions.

and newly developing leaves that can support photosynthesis (Hutyra et al., 2007; X. Li et al., 2018; Orth & Destouni, 2018; J. Wu et al., 2016; Yan et al., 2019). While the NDVI and NIRv analyses overall confirm the SIF results, they show extended water-related controls in arid regions and in tree-grass mixed regimes, consistent with previous findings (Walther et al., 2019). This is more pronounced for NDVI, potentially due to larger confounding effects of background brightness in NDVI as a response to SM changes, while NIRv contains more information about vegetation canopy structure and partly overcomes this issue (Badgley et al., 2017, 2019). Figure S9 finally investigates jointly the role of fraction of tree cover and aridity and shows that the main hydrometeorological controls change in response to both of them.

We furthermore analyze variations of hydrometeorological controls between early and late growing seasons. Temperature control can be found in larger regions in the early growing season compared to the later growing season, while the control of SM expanded in the late growing season (Figures S10 and S11) is in line with previous studies (Buermann et al., 2018; Lian et al., 2020; Zhang et al., 2020). Overall, patterns for these two sub-periods do not differ much from the results for the entire growing season (See Text S4.2 for details).

3.3. Main Water-Related Controls on Global Vegetation Productivity

Focusing exclusively on water-related first- and second-order controls reveals that the most important soil layer varies across climate-vegetation regimes (Figures 4a–4c). Shallow (7–28 cm in ERA5) and deep root-zone SM (28–100 cm in ERA5) are the most relevant layers for semi-arid grasses or shrubs, indicating that plants can adapt to water-scarce conditions at the surface with establishing deeper-reaching rooting systems (Fan et al., 2017). This is in line with previous but smaller-scale study from Yinglan et al. (2019); further studies confirm that in dry surface soils in (semi-)arid regions, root plasticity and morphology support water uptake from deeper soil layers (Schulze et al., 1996), for instance in local Mediterranean grass (Barkaoui et al., 2016), savannas ecosystems (Hoekstra et al., 2014; Nippert & Holdo, 2014) or central Brazilian savannas (Oliveira et al., 2005). For even drier climate conditions, shallower soil layers become more relevant (Figures 4a–4c), probably because intermittent vegetation productivity mostly benefits from rainfed surface

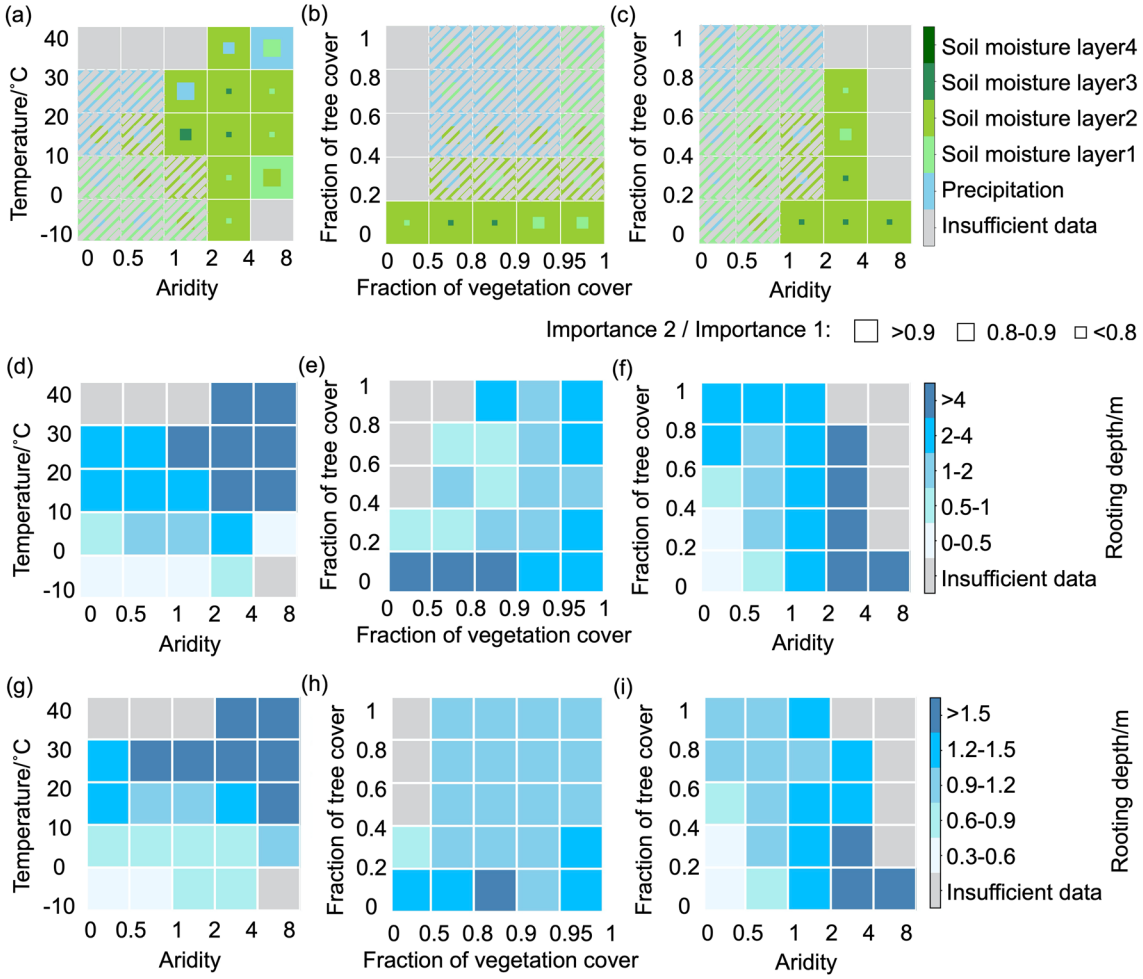

Figure 4. (a–c) Main water-related controls on sun-induced fluorescence (SIF), and distributions of rooting depths from (d–f) Fan et al. (2017) and (g–i) Schenk and Jackson. (2009) across climate regimes and vegetation characteristics. (a–c) Similar to Figure 3 but focusing on SIF and water-related controls only. Dark-gray hatching indicates that an energy variable is identified as the main control on SIF in these boxes in Figure 3.

SM, or deeper fine roots down to the water table capillary fringe play a role (Fan et al., 2017). Interestingly, toward humid conditions our analysis shows a dominant role of surface SM (0–7 cm in ERA5) and precipitation, while these regions are characterized by high tree cover with expected deep rooting systems. This could be due to frequent precipitation keeping surface soil layers wet such that trees absorb significant fractions of water through the near-surface roots (S. G. Li et al., 2007), while the dependence on deeper layers for trees during short droughts is not reflected here. Furthermore, we note that these regions are characterized by first-order energy controls (Figure 3) such that the results could also partly be an artifact as precipitation and surface SM are expected to co-vary with the dominant energy variables.

We validate our findings on the relative importance of the different soil layers by comparing them with multiple global rooting-depth products from two perspectives, effective root-zone water uptake and physical maximum rooting depths (Figures 4d–4i). In general, rooting depths from Fan et al. (2017) and Schenk and Jackson (2009) although from the physical maximum perspective show similar patterns as our results with deepest roots in semi-arid areas and for nontree vegetation such as grasses and shrubs. Canadell et al. (1996) states that maximum rooting depths of trees, shrubs and grasses can exceed 2 meters, respectively. Our results indicate that grasses and shrubs use their deep roots more often such that we can detect a respective relevance of deep SM reservoirs while trees predominantly use shallow roots as water is often easily accessible in wet surface layers. See Text S5 for data details.

To illustrate the robust importance-based analyses we use: (i) Spearman correlation (Figure S13) and (ii) SHAP feature importance (Figure S14) and find similar results. Further, we employ alternative SM products, namely GLEAM, MERRA-2 and SoMo.ml (see data details in Text S5; Figure S15), all of which lead to similar results as found with the ERA5 SM. We note that using a suite of SM products derived independently with physical models or machine learning approaches highlights that our results are not an artifact of one single SM model. In particular we find across these datasets that generally shallow soil layers are more relevant in humid regimes and deep(er) soil layers are more important in semi-arid areas and for nontree vegetation such as grasses and shrubs, confirming our previous results, while the layer depths and amounts are different across products and cannot be readily compared. In addition, we convert multi-layer SM to water potential. Thereby, the water potential is computed as the difference between actual SM and the permanent wilting point which is inferred from the soil texture in each grid cell (ECMWF, 2020). We find similar results indicating no major influence of the considered water availability metric on our results (Figure S16, as the role of different soil types might be diluted at large spatial scales where usually different soil types coincide in individual grid cells. Furthermore, we acknowledge, however, that our analyses do not consider water storage strategies related to hydraulic traits (Matheny et al., 2015) and irrigation effects.

4. Conclusions

This study illustrates that vegetation productivity relies on water from different soil depths while these characteristic depths vary with climate and vegetation types. In particular, we elucidate at the global scale that semi-arid areas and vegetation types such as grasses and shrubs are controlled by comparatively deep layers as they have deep rooting systems. This complexity is not yet sufficiently acknowledged by dynamic global vegetation models (Guimberteau et al., 2018; Schaphoff et al., 2018; Warren et al., 2015) which apply globally constant soil depths, and do not account for deep rooting strategies or potential physical barriers for vertical soil water transport (Sakschewski et al., 2020). This highlights the relevance of our results, and of our approach illustrating that sub-surface soil processes can be inferred with the help of surface-based Earth observations.

Further, we compare the hydrometeorological controls of vegetation productivity obtained with different respective proxy metrics. SIF is more strongly related to photosynthesis, compared with NDVI and NIRv, but SIF data are only available for recent years. Our results show that NDVI and NIRv yield similar spatial patterns and largely confirm the SIF-based results. However, we find extended water-related controls in the cases of NDVI and NIRv, especially in temperate wet regimes, probably induced by changes of soil background reflectance as a response to SM changes.

Overall, our study contributes to an advanced understanding of the global-scale partitioning of hydrometeorological controls of vegetation productivity by benefitting from the ever-growing suite of global eco-hydrological data streams.

Acknowledgments
The authors thank Ulrich Weber for processing the hydrometeorological data and the MODIS data, Tina Trautmann for processing the global rooting-depth data, and Gianpaolo Balsamo for guidance on the ERA5 vertical root distribution data. W. Li acknowledges funding from a PhD scholarship from the China Scholarship Council. W. Li acknowledges support from the International Max Planck Research School for Global Biogeochemical Cycles. R. Orth is funded by the German Research Foundation (Emmy Noether grant number 391059971), and S. Walther acknowledges funding by an ESA Living Planet Fellowship "Vad3e mecum". Open access funding enabled and organized by Projekt DEAL.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

SIF GFZ data have been retrieved from ftp://fluo.gps.caltech.edu/data/Philipp/GOME-2/ungridded/, ERA5 data from https://cds.climate.copernicus.eu/, GLEAM SM from https://www.gleam.eu/, MERRA-2 SM data from https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/FAQ/, and SoMo.ml from https://www.bgc-jena.mpg.de/geodb/projects/Home.php. All links are valid as of May 27, 2021.

References

Albergel, C., De Rosnay, P., Balsamo, G., Isaksen, L., & Muñoz-Sabater, J. (2012). Soil moisture analyses at ECMWF: Evaluation using global ground-based in situ observations. *Journal of Hydrometeorology*, *13*(5), 1442–1460. https://doi.org/10.1175/JHM-D-11-0107.1

Albergel, C., Dorigo, W., Reichle, R. H., Balsamo, G., De Rosnay, P., Muñoz-Sabater, J., et al. (2013). Skill and global trend analysis of soil moisture from reanalyses and microwave remote sensing. *Journal of Hydrometeorology*, *14*(4), 1259–1277. https://doi.org/10.1175/JHM-D-12-0161.1

Albergel, C., Dutra, E., Munier, S., Calvet, J.-C., Munoz-Sabater, J., de Rosnay, P., & Balsamo, G. (2018). ERA-5 and ERA-Interim driven ISBA land surface model simulations: Which one performs better?. *Hydrology and Earth System Sciences*, *22*(6), 3515–3532. https://doi.org/10.5194/hess-22-3515-2018

Badgley, G., Anderegg, L. D. L., Berry, J. A., & Field, C. B. (2019). Terrestrial gross primary production: Using NIR V to scale from site to globe. *Global Change Biology*, *25*(11), 3731–3740. https://doi.org/10.1111/gcb.14729

Badgley, G., Field, C. B., & Berry, J. A. (2017). Canopy near-infrared reflectance and terrestrial photosynthesis. *Science Advances*, *3*(3), e1602244. https://doi.org/10.1126/sciadv.1602244

Balsamo, G., Beljaars, A., Scipal, K., Viterbo, P., van den Hurk, B., Hirschi, M., & Betts, A. K. (2009). A revised hydrology for the ECMWF model: Verification from field site to terrestrial water storage and impact in the integrated forecast system. *Journal of Hydrometeorology*, *10*(3), 623–643. https://doi.org/10.1175/2008JHM1068.1

Barkaoui, K., Roumet, C., & Volaire, F. (2016). Mean root trait more than root trait diversity determines drought resilience in native and cultivated Mediterranean grass mixtures. *Agriculture, Ecosystems & Environment*, *231*, 122–132. https://doi.org/10.1016/j.agee.2016.06.035

Beer, C., Reichstein, M., Tomelleri, E., Ciais, P., Jung, M., Carvalhais, N., et al. (2010). Terrestrial gross carbon dioxide uptake: Global distribution and covariation with climate. *Science*, *329*(5993), 834–838. https://doi.org/10.1126/science.1184984

Berg, A., Sheffield, J., & Milly, P. C. D. (2017). Divergent surface and total soil moisture projections under global warming. *Geophysical Research Letters*, *44*, 236–244. https://doi.org/10.1002/2016GL071921

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32. https://doi.org/10.1023/a:1010933404324

Budyko, M. I., Miller, D. H., & Miller, D. H. (1974). *Climate and life*. New York. Academic press. https://doi.org/10.2307/3897455

Buermann, W., Forkel, M., O'Sullivan, M., Sitch, S., Friedlingstein, P., Haverd, V., et al. (2018). Widespread seasonal compensation effects of spring warming on northern plant productivity. *Nature*, *562*(7725), 110–114. https://doi.org/10.1038/s41586-018-0555-7

Canadell, J., Jackson, R. B., Ehleringer, J. B., Mooney, H. A., Sala, O. E., & Schulze, E.&D., (1996). Maximum rooting depth of vegetation types at the global scale. *Oecologia*, *108*(4), 583–595. http://www.jstor.org/stable/4221458

Chen, A., Mao, J., Ricciuto, D., Xiao, J., Frankenberg, C., Li, X., et al. (2020). Moisture availability mediates the relationship between terrestrial gross primary production and solar-induced chlorophyll fluorescence: Insights from global-scale variations. *Global Change Biology*, *27*(6), 1144–1156. https://doi.org/10.1111/gcb.15373

Claessen, J., Molini, A., Martens, B., Detto, M., Demuzere, M., & Miralles, D. G. (2019). Global biosphere-climate interaction: A causal appraisal of observations and models over multiple temporal scales. *Biogeosciences*, *16*(24), 4851–4874. https://doi.org/10.5194/bg-16-4851-2019

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, *74*(368), 829–836. https://doi.org/10.1080/01621459.1979.10481038

Cox, P. M., Pearson, D., Booth, B. B., Friedlingstein, P., Huntingford, C., Jones, C. D., & Luke, C. M. (2013). Sensitivity of tropical carbon to climate change constrained by carbon dioxide variability. *Nature*, *494*(7437), 341–344. https://doi.org/10.1038/nature11882

Cutler, A., Cutler, D. R., & Stevens, J. R. (2012). Random forests. In C. Zhang, & Y. Ma (Eds.), *Ensemble machine learning* (pp. 157–175). Boston, MA. Springer. https://doi.org/10.1007/978-1-4419-9326-7_5

De Keersmaecker, W., Lhermitte, S., Tits, L., Honnay, O., Somers, B., & Coppin, P. (2015). A model quantifying global vegetation resistance and resilience to short-term climate anomalies and their relationship with vegetation cover. *Global Ecology and Biogeography*, *24*, 539–548. https://doi.org/10.1111/geb.12279

Denissen, J. M. C., Teuling, A. J., Reichstein, M., & Orth, R. (2020). Critical soil moisture derived from satellite observations over Europe. *Journal of Geophysical Research: Atmospheres*, *125*, e2019JD031672. https://doi.org/10.1029/2019JD031672

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., et al. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, *36*(1), 27–46. https://doi.org/10.1111/j.1600-0587.2012.07348.x

Duveiller, G., Hooker, J., & Cescatti, A. (2018). The mark of vegetation change on Earth's surface energy balance. *Nature Communications*, *9*, 679. https://doi.org/10.1038/s41467-017-02810-8

ECMWF. (2020). *Surface parametrization. IFS Documentation CY47R1 - Part I: Observations* (pp. 127–170). England. ECMWF. https://doi.org/10.21957/cpmkqvhja

Fan, Y., Miguez-Macho, G., Jobbágy, E. G., Jackson, R. B., & Otero-Casal, C. (2017). Hydrologic regulation of plant rooting depth. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(40), 10572–10577. https://doi.org/10.1073/pnas.1712381114

Fatichi, S., & Ivanov, V. Y. (2014). Interannual variability of evapotranspiration and vegetation productivity. *Water Resources Research*, *50*, 3275–3294. https://doi.org/10.1002/2013WR015044

Feng, S., Hu, Q., Huang, W., Ho, C.-H., Li, R., & Tang, Z. (2014). Projected climate regime shift under future global warming from multi-model, multi-scenario CMIP5 simulations. *Global and Planetary Change*, *112*, 41–52. https://doi.org/10.1016/j.gloplacha.2013.11.002

Forkel, M., Andela, N., Harrison, S. P., Lasslop, G., van Marle, M., Chuvieco, E., et al. (2019). Emergent relationships with respect to burned area in global satellite observations and fire-enabled vegetation models. *Biogeosciences*, *16*(1), 57–76. https://doi.org/10.5194/bg-16-57-2019

Frankenberg, C., Fisher, J. B., Worden, J., Badgley, G., Saatchi, S. S., Lee, J.-E., et al. (2011). New global observations of the terrestrial carbon cycle from GOSAT: Patterns of plant fluorescence with gross primary productivity. *Geophysical Research Letters*, *38*, L17706. https://doi.org/10.1029/2011gl048738

Garonna, I., de Jong, R., Stöckli, R., Schmid, B., Schenkel, D., Schimel, D., & Schaepman, M. E. (2018). Shifting relative importance of climatic constraints on land surface phenology. *Environmental Research Letters*, *13*(2), 024025. https://doi.org/10.1088/1748-9326/aaa17b

Gómez-Ramírez, J., Ávila-Villanueva, M., & Fernández-Blázquez, M. Á. (2020). Selecting the most important self-assessed features for predicting conversion to Mild cognitive impairment with Random Forest and Permutation-based methods. *Scientific Reports*, *10*(1), 1–15. https://doi.org/10.1038/s41598-020-77296-4

Green, J. K., Berry, J., Ciais, P., Zhang, Y., & Gentine, P. (2020). Amazon rainforest photosynthesis increases in response to atmospheric dryness. *Science Advances*, *6*(47), eabb7232. https://doi.org/10.1126/sciadv.abb7232

Guimberteau, M., Zhu, D., Maignan, F., Huang, Y., Yue, C., Dantec-Nédélec, S., et al. (2018). ORCHIDEE-MICT (v8.4.1), a land surface model for the high latitudes: Model description and validation. *Geoscientific Model Development*, *11*, 121–163. https://doi.org/10.5194/gmd-11-121-2018

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Hoekstra, N. J., Finn, J. A., Hofer, D., & Lüscher, A. (2014). The effect of drought and interspecific interactions on depth of water uptake in deep- and shallow-rooting grassland species as determined by $\delta^{18}O$ natural abundance. *Biogeosciences*, *11*(16), 4493–4506. https://doi.org/10.5194/bg-11-4493-2014

Huang, K., & Xia, J. (2019). High ecosystem stability of evergreen broadleaf forests under severe droughts. *Global Change Biology*, *25*(10), 3494–3503. https://doi.org/10.1111/gcb.14748

Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., & Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, *83*(1–2), 195–213. https://doi.org/10.1016/S0034-4257(02)00096-2

Hutyra, L. R., Munger, J. W., Saleska, S. R., Gottlieb, E., Daube, B. C., Dunn, A. L., et al. (2007). Seasonal controls on the exchange of carbon and water in an Amazonian rain forest. *Journal of Geophysical Research*, *112*, G03008. https://doi.org/10.1029/2006JG000365

Jiao, W., Chang, Q., & Wang, L. (2019). The sensitivity of satellite solar-induced chlorophyll fluorescence to meteorological drought. *Earth's Future*, *7*(5), 558–573. https://doi.org/10.1029/2018ef001087

Jing, W., Song, J., & Zhao, X. (2018). Validation of ECMWF multi-layer reanalysis soil moisture based on the OzNet hydrology network. *Water*, *10*(9), 1123. https://doi.org/10.3390/w10091123

Joiner, J., Guanter, L., Lindstrot, R., Voigt, M., Vasilkov, A. P., Middleton, E. M., et al. (2013). Global monitoring of terrestrial chlorophyll fluorescence from moderate-spectral-resolution near-infrared satellite measurements: Methodology, simulations, and application to GOME-2. *Atmospheric Measurement Techniques*, *6*(10), 2803–2823. https://doi.org/10.5194/amt-6-2803-2013

Jung, M., Reichstein, M., Margolis, H. A., Cescatti, A., Richardson, A. D., Arain, M. A., et al. (2011). Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations. *Journal of Geophysical Research*, *116*, G00J07. https://doi.org/10.1029/2010JG001566

Jung, M., Reichstein, M., Schwalm, C. R., Huntingford, C., Sitch, S., Ahlström, A., et al. (2017). Compensatory water effects link yearly global land CO2 sink changes to temperature. *Nature*, *541*(7638), 516–520. https://doi.org/10.1038/nature20780

Köhler, P., Guanter, L., & Joiner, J. (2015). A linear method for the retrieval of sun-induced chlorophyll fluorescence from GOME-2 and SCIAMACHY data. *Atmospheric Measurement Techniques*, *8*(6), 2589–2608. https://doi.org/10.5194/amt-8-2589-2015

Kolluru, V., Kolluru, S., & Konkathi, P. (2020). Evaluation and integration of reanalysis rainfall products under contrasting climatic conditions in India. *Atmospheric Research*, *246*, 105121. https://doi.org/10.1016/j.atmosres.2020.105121

Kraft, B., Jung, M., Körner, M., Requena Mesa, C., Cortés, J., & Reichstein, M. (2019). Identifying dynamic memory effects on vegetation state using recurrent neural networks. *Frontiers in Big Data*, *2*(31). https://doi.org/10.3389/fdata.2019.00031

Li, M., Wu, P., & Ma, Z. (2020). A comprehensive evaluation of soil moisture and soil temperature from third-generation atmospheric and land reanalysis data sets. *International Journal of Climatology*, *40*, 5744–5766. https://doi.org/10.1002/joc.6549

Li, S. G., Romero-Saltos, H., Tsujimura, M., Sugimoto, A., Sasaki, L., Davaa, G., & Oyunbaatar, D. (2007). Plant water sources in the cold semiarid ecosystem of the upper Kherlen River catchment in Mongolia: A stable isotope approach. *Journal of Hydrology*, *333*(1), 109–117. https://doi.org/10.1016/j.jhydrol.2006.07.020

Li, X., & Xiao, J. (2020). Global climatic controls on interannual variability of ecosystem productivity: Similarities and differences inferred from solar-induced chlorophyll fluorescence and enhanced vegetation index. *Agricultural and Forest Meteorology*, *288–289*, 108018. https://doi.org/10.1016/j.agrformet.2020.108018

Li, X., Xiao, J., & He, B. (2018). Higher absorbed solar radiation partly offset the negative effects of water stress on the photosynthesis of Amazon forests during the 2015 drought. *Environmental Research Letters*, *13*, 044005. https://doi.org/10.1088/1748-9326/aab0b1

Li, X., Xiao, J., Kimball, J. S., Reichle, R. H., Scott, R. L., Litvak, M. E., et al. (2020). Synergistic use of SMAP and OCO-2 data in assessing the responses of ecosystem productivity to the 2018 U.S. drought. *Remote Sensing of Environment*, *251*, 112062. https://doi.org/10.1016/j.rse.2020.112062

Lian, X., Piao, S., Li, L. Z. X., Li, Y., Huntingford, C., Ciais, P., et al. (2020). Summer soil drying exacerbated by earlier spring greening of northern vegetation. *Science Advances*, *6*(1), eaax0255. https://doi.org/10.1126/sciadv.aax0255

Linscheid, N., Estupinan-Suarez, L. M., Brenning, A., Carvalhais, N., Cremer, F., Gans, F., et al. (2020). Towards a global understanding of vegetation-climate dynamics at multiple timescales. *Biogeosciences*, *17*(4), 945–962. https://doi.org/10.5194/bg-17-945-2020

Liu, L., Zhang, R., & Zuo, Z. (2014). Intercomparison of spring soil moisture among multiple reanalysis data sets over eastern China. *Journal of Geophysical Research: Atmospheres*, *119*(1), 54–64. https://doi.org/10.1002/2013JD020940

Lunetta, K. L., Hayward, L. B., Segal, J., & Van Eerdewegh, P. (2004). Screening large-scale association study data: Exploiting interactions using random forests. *BMC Genetics*, *5*, 32. https://doi.org/10.1186/1471-2156-5-32

Madani, N., Kimball, J., Jones, L., Parazoo, N., & Guan, K. (2017). Global analysis of bioclimatic controls on ecosystem productivity using satellite observations of solar-induced chlorophyll fluorescence. *Remote Sensing*, *9*(6), 530. https://doi.org/10.3390/rs9060530

Matheny, A. M., Bohrer, G., Garrity, S. R., Morin, T. H., Howard, C. J., & Vogel, C. S. (2015). Observations of stem water storage in trees of opposing hydraulic strategies. *Ecosphere*, *6*(9), 1–13. https://doi.org/10.1890/ES15-00170.1

Migliavacca, M., Meroni, M., Manca, G., Matteucci, G., Montagnani, L., Grassi, G., et al. (2009). Seasonal and interannual patterns of carbon and water fluxes of a poplar plantation under peculiar eco-climatic conditions. *Agricultural and Forest Meteorology*, *149*(9), 1460–1476. https://doi.org/10.1016/j.agrformet.2009.04.003

Monteith, J. L., & Unsworth, M. H. (1990). *Principles of environmental physics*. London. Edward Arnold.

Nemani, R. R., Keeling, C. D., Hashimoto, H., Jolly, W. M., Piper, S. C., Tucker, C. J., et al. (2003). Climate-driven increases in global terrestrial net primary production from 1982 to 1999. *Science*, *300*(5625), 1560–1563. https://doi.org/10.1126/science.1082750

Nicodemus, K. K. (2011). Letter to the editor: On the stability and ranking of predictors from random forest variable importance measures. *Briefings in Bioinformatics*, *12*, 369–373. https://doi.org/10.1093/bib/bbr016

Nippert, J. B., & Holdo, R. M. (2015). Challenging the maximum rooting depth paradigm in grasslands and savannas. *Functional Ecology*, *29*(6), 739–745. https://doi.org/10.1111/1365-2435.12390

Novick, K. A., Ficklin, D. L., Stoy, P. C., Williams, C. A., Bohrer, G., Oishi, A. C., et al. (2016). The increasing importance of atmospheric demand for ecosystem water and carbon fluxes. *Nature Climate Change*, *6*(11), 1023–1027. https://doi.org/10.1038/nclimate3114

Oliveira, R. S., Bezerra, L., Davidson, E. A., Pinto, F., Klink, C. A., Nepstad, D. C., & Moreira, A. (2005). Deep root function in soil water dynamics in cerrado savannas of central Brazil. *Functional Ecology*, *19*, 574–581. https://doi.org/10.1111/j.1365-2435.2005.01003.x

Orth, R., & Destouni, G. (2018). Drought reduces blue-water fluxes more strongly than green-water fluxes in Europe. *Nature Communications*, *9*, 3602. https://doi.org/10.1038/s41467-018-06013-7

Paschalis, A., Fatichi, S., Katul, G. G., & Ivanov, V. Y. (2015). Cross-scale impact of climate temporal variability on ecosystem water and carbon fluxes. *Journal of Geophysical Research: Biogeosciences*, *120*, 1716–1740. https://doi.org/10.1002/2015JG003002

Pearson, R. G., Phillips, S. J., Loranty, M. M., Beck, P. S. A., Damoulas, T., Knight, S. J., & Goetz, S. J. (2013). Shifts in Arctic vegetation and associated feedbacks under climate change. *Nature Climate Change*, *3*(7), 673–677. https://doi.org/10.1038/nclimate1858

Piao, S., Wang, X., Wang, K., Li, X., Bastos, A., Canadell, J. G., et al. (2020). Interannual variation of terrestrial carbon cycle: Issues and perspectives. *Global Change Biology*, *26*(1), 300–318. https://doi.org/10.1111/gcb.14884
Reichstein, M., Bahn, M., Ciais, P., Frank, D., Mahecha, M. D., Seneviratne, S. I., et al. (2013). Climate extremes and the carbon cycle. *Nature*, *500*(7462), 287–295. https://doi.org/10.1038/nature12350
Sakschewski, B., Von Bloh, W., Boit, A., Poorter, L., Peña-Claros, M., Heinke, J., et al. (2016). Resilience of Amazon forests emerges from plant trait diversity. *Nature Climate Change*, *6*(11), 1032–1036. https://doi.org/10.1038/nclimate3109
Sakschewski, B., von Bloh, W., Drüke, M., Sörensson, A. A., Ruscica, R., Langerwisch, F. et al. (2020). Variable tree rooting strategies improve tropical productivity and evapotranspiration in a dynamic global vegetation model. *Biogeosciences Discuss. [preprint]*. https://doi.org/10.5194/bg-2020-97
Schaphoff, S., von Bloh, W., Rammig, A., Thonicke, K., Biemans, H., Forkel, M., et al. (2018). LPJmL4 - A dynamic global vegetation model with managed land - Part 1: Model description. *Geoscientific Model Development*, *11*(4), 1343–1375. https://doi.org/10.5194/gmd-11-1343-2018
Schenk, H. J., & Jackson, R. B. (2009). ISLSCP II Ecosystem Rooting Depths. In F. G. Hall, G. Collatz, B. Meeson, S. Los, E. Brown de Colstoun, & D. Landis (Eds.), *ISLSCP Initiative II Collection*. Data set. Oak Ridge National Laboratory Distributed Active Archive Center. U.S.A. https://doi.org/10.3334/ORNLDAAC/929
Schlaepfer, D. R., Bradford, J. B., Lauenroth, W. K., Munson, S. M., Tietjen, B., Hall, S. A., et al. (2017). Climate change reduces extent of temperate drylands and intensifies drought in deep soils. *Nature Communications*, *8*. 14196. https://doi.org/10.1038/ncomms14196
Schulze, E.-D., Mooney, H. A., Sala, O. E., Jobbagy, E., Buchmann, N., Bauer, G., et al. (1996). Rooting depth, water availability, and vegetation cover along an aridity gradient in Patagonia. *Oecologia*, *108*(3), 503–511. https://doi.org/10.1007/BF00333727
Seddon, A. W. R., Macias-Fauria, M., Long, P. R., Benz, D., & Willis, K. J. (2016). Sensitivity of global terrestrial ecosystems to climate variability. *Nature*, *531*(7593), 229–232. https://doi.org/10.1038/nature16986
Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., et al. (2010). Investigating soil moisture-climate interactions in a changing climate: A review. *Earth-Science Reviews*, *99*(3–4), 125–161. https://doi.org/10.1016/j.earscirev.2010.02.004
Song, S., & Wang, W. (2019). Impacts of antecedent soil moisture on the rainfall-runoff transformation process based on high-resolution observations in soil tank experiments. *Water*, *11*(2), 296. https://doi.org/10.3390/w11020296
Song, X.-P., Hansen, M. C., Stehman, S. V., Potapov, P. V., Tyukavina, A., Vermote, E. F., & Townshend, J. R. (2018). Global land change from 1982 to 2016. *Nature*, *560*(7720), 639–643. https://doi.org/10.1038/s41586-018-0411-9
Stocker, B. D., Wang, H., Smith, N. G., Harrison, S. P., Keenan, T. F., Sandoval, D., et al. (2020). P-model v1.0: An optimality-based light use efficiency model for simulating ecosystem gross primary production. *Geoscientific Model Development*, *13*(3), 1545–1581. https://doi.org/10.5194/gmd-13-1545-2020
Stocker, B. D., Zscheischler, J., Keenan, T. F., Prentice, I. C., Peñuelas, J., & Seneviratne, S. I. (2018). Quantifying soil moisture impacts on light use efficiency across biomes. *New Phytologist*, *218*(4), 1430–1449. https://doi.org/10.1111/nph.15123
Tarek, M., Brissette, F. P., & Arsenault, R. (2020). Evaluation of the ERA5 reanalysis as a potential reference dataset for hydrological modelling over North America. *Hydrology and Earth System Sciences*, *24*(5), 2527–2544. https://doi.org/10.5194/hess-24-2527-2020
Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, *8*(2), 127–150. https://doi.org/10.1016/0034-4257(79)90013-0
Wagle, P., Zhang, Y., Jin, C., & Xiao, X. (2016). Comparison of solar-induced chlorophyll fluorescence, light-use efficiency, and process-based GPP models in maize. *Ecological Applications*, *26*(4), 1211–1222. https://doi.org/10.1890/15-1434
Walther, S., Duveiller, G., Jung, M., Guanter, L., Cescatti, A., & Camps-Valls, G. (2019). Satellite observations of the contrasting response of trees and grasses to variations in water availability. *Geophysical Research Letters*, *46*, 1429–1440. https://doi.org/10.1029/2018gl080535
Warren, J. M., Hanson, P. J., Iversen, C. M., Kumar, J., Walker, A. P., & Wullschleger, S. D. (2015). Root structural and functional dynamics in terrestrial biosphere models - Evaluation and recommendations. *New Phytologist*, *205*(1), 59–78. https://doi.org/10.1111/nph.13034
Wu, D., Zhao, X., Liang, S., Zhou, T., Huang, K., Tang, B., & Zhao, W. (2015). Time-lag effects of global vegetation responses to climate change. *Global Change Biology*, *21*, 3520–3531. https://doi.org/10.1111/gcb.12945
Wu, J., Albert, L. P., Lopes, A. P., Restrepo-Coupe, N., Hayek, M., Wiedemann, K. T., et al. (2016). Leaf development and demography explain photosynthetic seasonality in Amazon evergreen forests. *Science*, *351*(6276), 972–976. https://doi.org/10.1126/science.aad5068
Yan, H., Wang, S. Q., Huete, A., & Shugart, H. H. (2019). Effects of light component and water stress on photosynthesis of amazon rainforests during the 2015/2016 El Niño drought. *Journal of Geophysical Research: Biogeosciences*, *124*, 1574–1590. https://doi.org/10.1029/2018JG004988
Yang, X., Tang, J., Mustard, J. F., Lee, J.-E., Rossini, M., Joiner, J., et al. (2015). Solar-induced chlorophyll fluorescence that correlates with canopy photosynthesis on diurnal and seasonal scales in a temperate deciduous forest. *Geophysical Research Letters*, *42*, 2977–2987. https://doi.org/10.1002/2015GL063201
Yinglan, A., Wang, G., Liu, T., Xue, B., & Kuczera, G. (2019). Spatial variation of correlations between vertical soil water and evapotranspiration and their controlling factors in a semi-arid region. *Journal of Hydrology*, *574*, 53–63. https://doi.org/10.1016/j.jhydrol.2019.04.023
Zhang, F., & Yang, X. (2020). Improving land cover classification in an urbanized coastal area by random forests: The role of variable selection. *Remote Sensing of Environment*, *251*, 112105. https://doi.org/10.1016/j.rse.2020.112105
Zhang, Y., Parazoo, N. C., Williams, A. P., Zhou, S., & Gentine, P. (2020). Large and projected strengthening moisture limitation on end-of-season photosynthesis. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(17), 9216–9222. https://doi.org/10.1073/pnas.1914436117
Zhang, Y., Peña-Arancibia, J. L., McVicar, T. R., Chiew, F. H. S., Vaze, J., Liu, C., et al. (2016). Multi-decadal trends in global terrestrial evapotranspiration and its components. *Scientific Reports*, *6*, 19124. https://doi.org/10.1038/srep19124
Zuromski, L. M., Bowling, D. R., Köhler, P., Frankenberg, C., Goulden, M. L., Blanken, P. D., & Lin, J. C. (2018). Solar-induced fluorescence detects interannual variation in gross primary production of coniferous forests in the Western United States. *Geophysical Research Letters*, *45*, 7184–7193. https://doi.org/10.1029/2018GL077906

References From the Supporting Information

Balsamo, G., Albergel, C., Beljaars, A., Boussetta, S., Brun, E., Cloke, H., et al. (2015). ERA-Interim/Land: A global land surface reanalysis data set. *Hydrology and Earth System Sciences*, *19*(1), 389–407. https://doi.org/10.5194/hess-19-389-2015
Breiman, L. (1996). *Out-of-bag estimation*.

Drusch, M., Scipal, K., De Rosnay, P., Balsamo, G., Andersson, E., Bougeault, P., & Viterbo, P. (2009). Towards a Kalman Filter based soil moisture analysis system for the operational ECMWF Integrated Forecast System. *Geophysical Research Letters*, *36*, L10401. https://doi.org/10.1029/2009GL037716

Forkel, M., Migliavacca, M., Thonicke, K., Reichstein, M., Schaphoff, S., Weber, U., & Carvalhais, N. (2015). Codominant water control on global interannual variability and trends in land surface phenology and greenness. *Global Change Biology*, *21*(9), 3414–3435. https://doi.org/10.1111/gcb.12950

Fournier, A., Daumard, F., Champagne, S., Ounis, A., Goulas, Y., & Moya, I. (2012). Effect of canopy structure on sun-induced chlorophyll fluorescence. *ISPRS Journal of Photogrammetry and Remote Sensing*, *68*, 112–120. https://doi.org/10.1016/j.rse.2020.112195

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., et al. (2017). The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *Journal of Climate*, *30*(14), 5419–5454. https://doi.org/10.1175/JCLI-D-16-0758.1

Guanter, L., Frankenberg, C., Dudhia, A., Lewis, P. E., Gómez-Dans, J., Kuze, A., et al. (2012). Retrieval and global assessment of terrestrial chlorophyll fluorescence from GOSAT space measurements. *Remote Sensing of Environment*, *121*, 236–251. https://doi.org/10.1016/j.rse.2012.02.006

He, L., Chen, J. M., Liu, J., Mo, G., & Joiner, J. (2017). Angular normalization of GOME-2 Sun-induced chlorophyll fluorescence observation as a better proxy of vegetation productivity. *Geophysical Research Letters*, *44*, 5691–5699. https://doi.org/10.1002/2017GL073708

Helm, L. T., Shi, H., Lerdau, M. T., & Yang, X. (2020). Solar-induced chlorophyll fluorescence and short-term photosynthetic response to drought. *Ecological Applications*, *30*(5), e02101. https://doi.org/10.1002/eap.2101

Huete, A., Liu, H. Q., Batchily, K., & Van Leeuwen, W. (1997). A comparison of vegetation indices over a global set of TM images for EOS-MODIS. *Remote Sensing of Environment*, *59*(3), 440–451. https://doi.org/10.1016/S0034-4257(96)00112-5

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, *9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55

Köhler, P., Guanter, L., Kobayashi, H., Walther, S., & Yang, W. (2018). Assessing the potential of sun-induced fluorescence and the canopy scattering coefficient to track large-scale vegetation dynamics in Amazon forests. *Remote Sensing of Environment*, *204*, 769–785. https://doi.org/10.1016/j.rse.2017.09.025

Lundberg, S. M., & Lee, S. I. (2017). *A unified approach to interpreting model predictions*. arXiv. 1705.07874.

Magney, T. S., Bowling, D. R., Logan, B. A., Grossmann, K., Stutz, J., Blanken, P. D., et al. (2019). Mechanistic evidence for tracking the seasonality of photosynthesis with solar-induced fluorescence. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(24), 11640–11645 https://doi.org/10.1073/pnas.1900278116

Martens, B., Miralles, D. G., Lievens, H., Van Der Schalie, R., De Jeu, R. A. M., Fernández-Prieto, D., et al. (2017). GLEAM v3: Satellite-based land evaporation and root-zone soil moisture. *Geoscientific Model Development*, *10*(5), 1903–1925. https://doi.org/10.5194/gmd-10-1903-2017

Migliavacca, M., Perez-Priego, O., Rossini, M., El-Madany, T. S., Moreno, G., van der Tol, C., et al. 2017). Plant functional traits and canopy structure control the relationship between photosynthetic CO 2 uptake and far-red sun-induced fluorescence in a Mediterranean grassland under different nutrient availability. *New Phytologist*, *214*(3), 1078–1091. https://doi.org/10.1111/nph.14437

O and Orth. (2021). Global soil moisture data derived through machine learning trained with in-situ measurements. *Scientific Data. [accepted]*.

Oliphant, T. E. (2006). *A guide to NumPy*. Trelgol Publishing USA.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Skipper, S., & Perktold, J. (2010). Sstatsmodels: Econometric and statistical modeling with python. *Proceedings of the 9th Python in Science Conference*.

Sundararajan, M., & Najmi, A. (2019). *The many Shapley values for model explanation*. arXiv:1908.08474.

Wohlfahrt, G., Gerdel, K., Migliavacca, M., Rotenberg, E., Tatarinov, F., Müller, J., et al. (2018). Sun-induced fluorescence and gross primary productivity during a heat wave. *Scientific Reports*, *8*, 14169. https://doi.org/10.1038/s41598-018-32602-z

Yang, Y., Donohue, R. J., & McVicar, T. R. (2016). Global estimation of effective plant rooting depth: Implications for hydrological modeling. *Water Resources Research*, *52*, 8260–8276. https://doi.org/10.1002/2016WR019392

Zeng, Y., Badgley, G., Dechant, B., Ryu, Y., Chen, M., & Berry, J. A. (2019). A practical approach for estimating the escape ratio of near-infrared solar-induced chlorophyll fluorescence. *Remote Sensing of Environment*, *232*, 111209. https://doi.org/10.1016/j.rse.2019.05.028

Zwillinger, D., & Kokoska, S. (2000). *Standard probability and statistics tables and formulae*. Florida. Chapman & Hall/CRC.