https://doi.org/10.5194/egusphere-2023-770
Preprint. Discussion started: 8 May 2023

Relevance of near-surface soil moisture vs. terrestrial water storage for global vegetation functioning

Prajwal Khanal[1,2], Anne J. Hoek Van Dijke[1], Timo Schaffhauser[2], Wantong Li[1], Sinikka J. Paulus[1,3], Chunhui Zhan[1,4], René Orth[1]

[1]Department of Biogeochemical Integration, Max Planck Institute for Biogeochemistry, Hans-Knöll-Straße 10, 07745 Jena, -Germany
[2]Chair of Hydrology and River Basin Management, Technical University of Munich, Arcisstraße 21, 80333 Munich, Germany
[3]Chair of Terrestrial Ecohydrology, University of Jena, Burgweg 11, 07749, Jena, Germany
[4]Land Surface-Atmosphere Interactions, Technical University of Munich, TUM School of Life Sciences Weihenstephan, 85354 Freising,Germany

Correspondence to: Prajwal Khanal (ktm.prajwalkhanal@gmail.com)

Abstract. Soil water availability is an essential prerequisite for vegetation functioning. Vegetation takes up water from varying soil depths depending on the characteristics of their rooting system and soil moisture availability across depth. The depth of vegetation water uptake is largely unknown across large spatial scales as a consequence of sparse ground measurements. At the same time, emerging satellite-derived observations of vegetation functioning, surface soil moisture and terrestrial water storage, present an opportunity to assess the depth of vegetation water uptake globally. In this study, we characterise vegetation functioning through the Near-Infrared Reflectance of Vegetation (NIRv), and compare its relation to (i) near-surface soil moisture from ESA-CCI and (ii) total water storage from GRACE at the monthly time scale during the growing season. The relationships are quantified through partial correlations to mitigate the influence of confounding factors such as energy-related variables. We find that vegetation functioning is generally more strongly related to near-surface soil moisture, particularly in semi-arid regions and areas with low tree cover. In contrast, in regions with high tree cover and in arid regions, the correlation with terrestrial water storage is comparable to or even higher than with near-surface soil moisture, indicating that trees can and do make use of their deeper rooting systems to access deeper soil moisture, similar to vegetation in arid regions. In line with this, an attribution analysis that examines the relative importance of these soil water storages for vegetation reveals that they are controlled by (i) water availability influenced by the climate and (ii) vegetation type reflecting adaptation of ecosystems to local water resources. Next to variations in space, the vegetation water uptake depth also varies in time. During dry periods, the relative importance of terrestrial water storage increases, highlighting the relevance of deeper water resources during rain-scarce periods. Overall, the synergistic exploitation of state-of-the-art satellite data products to disentangle the relevance of near-surface vs. terrestrial water storage for vegetation functioning can inform the representation of vegetation-water interactions in land surface models to support more accurate climate change projections.

https://doi.org/10.5194/egusphere-2023-770
Preprint. Discussion started: 8 May 2023

1. Introduction

The regulation of water, energy, and biogeochemical cycling between land and atmosphere is primarily dependent on vegetation. In addition, global vegetation provides essential ecosystem services such as food production and uptake of some of the anthropogenic carbon dioxide emissions (Keenan & Williams, 2018). Vegetation growth depends on nutrient, water and energy availability. As a result, on a global scale, there are regions with energy or water limited vegetation functioning (Orth, 2021). In energy-limited regions, the functioning of vegetation is controlled by radiation and temperature, as they often lack sunny and warm conditions but have ample soil moisture. In contrast, soil moisture becomes critical for vegetation growth in water-limited regions. Plant photosynthesis involves opening the stomata for the uptake of CO_2, while at the same time water is lost through transpiration. However, in water-limited conditions, plants can reduce the stomatal opening to avoid water loss, leading to a decrease in photosynthesis. Hence, variations in soil moisture are likely to affect vegetation functioning in water-limited conditions. Moreover, climate change has led to an expanded water limitation on vegetation (Denissen et al., 2022) and increased vegetation sensitivity to soil moisture (Li et al., 2022). For these reasons, it is essential to better understand the dependence of vegetation functioning on soil moisture to comprehend their coping mechanisms during drought to predict the future of global water, energy, and carbon cycles.

Plants extract water from varying soil depths based on the positioning of their roots and the availability of soil moisture. In general, the plant water uptake depth further differs spatially across different climate regimes and vegetation types, and temporally between seasons. Vegetation in arid regions is more susceptible to fluctuations in near-surface soil moisture compared to vegetation in humid regions (Xie et al., 2019). Grasses, which generally have shorter roots than trees and shrubs, are more reliant on near-surface moisture than deeper moisture (Schenk & Jackson, 2002). Further, root water uptake profiles vary within individual plant types according to above-ground biomass and age, with larger and older trees having deeper roots capable of extracting water from deeper soil layers (Schenk & Jackson, 2002; Tao et al., 2021). Additionally, within similar climate regimes, plant water uptake varies across topographic positions. Upland and lowland roots tend to be shallower, making vegetation more reliant on near-surface soil moisture, while roots go deeper in steep terrain between these landscapes to access both surface and deep moisture (Fan et al., 2017).

Though spatial variations of plant water uptake depths across vegetation types and climate regimes, and temporal shift during dry-months, are widely studied at point scale, inadequate deep soil moisture records pose a major obstacle to study vegetation root water uptake at a global scale. Microwave remote sensing allows to infer near-surface soil moisture dynamics globally. However, such data may not fully represent root-zone soil moisture as microwaves can only propagate through the top few centimetres in soil (Capehart & Carlson, 1997). Land surface models provide an alternative source of global soil moisture data across depths, but they are subject to uncertainties arising from meteorological data, inaccurate knowledge of soil and vegetation characteristics, and the representation of complex processes such as photosynthesis, infiltration, and evaporation

https://doi.org/10.5194/egusphere-2023-770
Preprint. Discussion started: 8 May 2023

(Koster et al., 2009; Seneviratne et al., 2010). Hence, some studies have employed reanalysis-based soil moisture estimates, to investigate the relationship between vegetation and soil moisture at the globa l(Li et al., 2021; Miguez-Macho & Fan, 2021); but those are likely to be impacted by model assumptions affecting soil moisture dynamics, particularly for deeper layers where less observational constraints are available. Thus, studying vegetation interactions with the entire water column, including near-surface and deep soil moisture, at a global scale using exclusively observation-based dataset is imperative to enhance the understanding of relevance of near-surface and deep soil moisture for vegetation functioning.

The Gravity Recovery and Climate Experiment (GRACE) satellite mission, launched in 2002, provides total water storage (TWS) anomalies observations at the global scale and offers a unique opportunity to investigate the relationship between vegetation and the total water column. Furthermore, the inter-annual carbon dioxide growth rate in the atmosphere has been found to be well correlated with the total water storage anomalies on a global scale, underlining the relevance of total water column for vegetation functioning (Humphrey et al., 2018). The TWS captures not only soil water but also snow and ice, canopy water, surface water, and groundwater. TWS anomalies can be used to estimate the variation of overall water availability (near-surface + deep soil moisture) for vegetation under (i) snow-free conditions, and assuming that (ii) water storage variations in lakes or groundwater are negligible at the monthly time scale, (iii) and canopy water storage is much smaller than soil water storage and hence also negligible.

This study focuses on understanding the relevance of near-surface vs. total water storage for vegetation functioning on a global scale using observation-based datasets, thereby inferring vegetation's large-scale water uptake depth from observation-based datasets. For this purpose, we utilise TWS and near-surface soil moisture and correlate them with vegetation functioning, represented by Near-Infrared Reflectance of Vegetation (NIRv). In particular, we analyse (1) what is the relevance of near-surface soil moisture vs. the terrestrial water storage for vegetation functioning?, (2) how does the importance of near-surface soil moisture vs. terrestrial water storage change during dry months? and (3) how do climatic, vegetation, and topographic characteristics explain the variability in the relevance of near-surface vs. terrestrial water storage for vegetation functioning?

2. Data and Methodology

Table 1: Table summarising all the datasets.

Datasets	Variables	Source	Spatial Resolution	References
Vegetation Functioning	Near Infrared Reflectance of Vegetation (NIRv)	MODIS/MOD13C1 v061	0.05 degree	(Badgley et al., 2017)
	Solar Induced Chlorophyll Fluorescence (SIF)	GOME-2	0.5 degree	(Köhler et al., 2015)

https://doi.org/10.5194/egusphere-2023-770
Preprint. Discussion started: 8 May 2023

Soil Water Storage	Near-surface soil moisture (SSM)	ESA-CCI v04.4	0.25 degree	(Dorigo et al., 2017)
	Total Water Storage (TWS) Anomalies	GRACE	0.5 degree	(Landerer & Swenson, 2012)
Meteorological	Air Temperature (T_a)	ERA-5	0.25 degree	(Hersbach et al., 2020)
	Precipitation (P)			
	Net Radiation (R_n)			
Vegetation and Land cover class	Tree cover fraction	VFC5KYR	0.05 degree	(Hansen, Matthew & Song, Xiao-Peng, 2018)
	Land cover data	ESA-CCI	0.05 degree	ESA. Land Cover CCI Product User Guide Version 2. Tech. Rep. (2017)
Topographical data	Elevation	Earthenv	1 km	(Amatulli et al., 2018)
	Slope			
Soil data	Fraction of sand	FAO	0.05 degree	(Reynolds et al., 2000)
	Fraction of clay			
Irrigation	Percentage of Irrigated area	HID	5 arcmin	(Siebert et al., 2015)

2.1 Data

2.1.1 Vegetation Functioning:

In our study, vegetation functioning is characterised by satellite measurements of Near-Infrared Reflectance of vegetation (NIRv) and Solar Induced Fluorescence (SIF) (**Table 1**). NIRv is the product of near-infrared reflectance and the normalised difference vegetation index (NDVI) and represents the vegetation structure and vegetation greenness (Badgley et al., 2017). The NIRv data is available at a high spatial resolution of 0.05°, and the original 16-day data was aggregated to the monthly NIRv data. SIF is directly related to the photosynthetic activity of plants because the excess energy from sunlight, that triggers the light reaction during photosynthesis, is dissipated by leaf as chlorophyll fluorescence (Mohammed et al., 2019). SIF data is derived from the Global Ozone Monitoring Experiment (GOME-2), because GOME-2 provides relatively reliable data over a long period (2007-2018). The 0.5° spatial and 16-day temporal resolution SIF data is processed into monthly data as described by (Köhler et al., 2015).

The high spatial resolution of NIRv allows for a detailed study of the correlation of vegetation functioning with soil water availability. Therefore, we performed the main analyses using NIRv data. However, SIF is more sensitive to drought stress than NIRv (Qiu et al., 2022). Therefore, we perform additional analyses with SIF to show that the relationships hold for a different and more direct indicator of vegetation functioning.

https://doi.org/10.5194/egusphere-2023-770
Preprint. Discussion started: 8 May 2023

2.1.2 Soil Water Storage

This study includes two different measures of soil water availability. The near-surface soil moisture (SSM) provides an estimate of water availability in the top layer of the soil, while the Terrestrial Water Storage (TWS) Anomaly provides an estimate of the overall water column of the soil. The SSM data is derived from the European Space Agency (ESA) Climate Change Initiative Program (CCI), which combines active and passive satellite microwave measurements to provide reliable estimates of SSM (Dorigo et al., 2017). The ESA CCI soil moisture data, at a daily temporal resolution, was aggregated to monthly temporal resolution. The TWS Anomaly data is derived from the GRACE mission, which measures changes in the Earth's gravity field (Landerer & Swenson, 2012). Here, we use the JPL-Mascons product of TWS Anomalies which is available at a 0.5° spatial and monthly temporal resolution.

2.1.3 Meteorological Data

Employed climate variables include monthly air temperature (T_a), precipitation (P), and net radiation (R_n) from the ERA5 reanalysis products at a 0.25° spatial resolution. The aridity index is calculated from the ratio between the long-term mean R_n (mm y^{-1}) (1 MJ/sq.m/day = 0.408 mm/day) and P (mm y^{-1}) for each grid cell (Budyko, 1974). In addition, the mean and standard deviation of the climate variables are calculated and incorporated in the attribution analysis (**Section 2.2.3**).

2.1.4 Vegetation, soil, and topography data

To evaluate the resulting correlation of vegetation functioning and water storages with respect to vegetation characteristics, we employ the tree cover fraction data from the AVHRR vegetation continuous fields products (VCF5KYR, https://lpdaac.usgs.gov/products/vcf5kyrv001/) (Hansen, Matthew & Song, Xiao-Peng, 2018). For this purpose, the mean of tree cover fraction for the years between 2007 and 2016 is calculated.

Topographical variables such as elevation and slope are incorporated along with other climatological variables to determine the relative contribution of different variables to the correlation between vegetation functioning and water storage. Topographic data at a 5 km resolution were downloaded from the EarthEnv. These data are calculated based on the 250 m GMTED dataset, and compared against the 90 m SRTM 4.1 dev dataset. The data were resampled to a coarser resolution of 5 km using various aggregation techniques, details of which are in (Amatulli et al., 2018). Furthermore for each grid cell, the fraction of sand and clay in soil (Reynolds et al., 2000) along with the percentage of irrigated area (Siebert et al., 2015) were considered in attribution analysis.

https://doi.org/10.5194/egusphere-2023-770
Preprint. Discussion started: 8 May 2023

2.2 Methodology

2.2.1 Data pre-processing

A flowchart of the data pre-processing and analyses is presented in **Figure S1**. The time period of analysis is from 2007 to 2018 constrained by the concurrent availability of all involved datasets. All the analyses were performed in monthly temporal resolution and at 0.05° spatial resolution (for NIRv) and 0.5° spatial resolution (for SIF). The SSM and TWS data were initially available at 0.25° and 0.5° resolution, but were disaggregated or aggregated to 0.05° or 0.5° degrees , depending on the spatial resolution of the analysis performed, based on the assumption that the soil water storage anomalies are representative over larger areas. Also the meteorological data and vegetation, soil, and topographic data were resampled into the same resolution. After aggregating all the datasets to 0.05° resolution, the monthly anomalies were calculated by subtracting the long term mean monthly cycle and by removing linear trends. A SIF threshold was applied in each grid cell to filter out non-growing season data. For this purpose, we filtered out all the months from 2007-2018 when the mean-monthly SIF value was below the threshold of 0.2 mW/m2/sr/nm. We apply an additional temperature threshold ($T_a > 5$°C) to remove the months with frozen soil and snow cover, similar to (Li et al., 2021). Last, all months with missing soil water storage or vegetation functioning records were excluded.

2.2.2 Calculate the relevance of near-surface (SSM) soil moisture and terrestrial water storage (TWS) for vegetation functioning

We calculated the spearman correlation between vegetation functioning (NIRv) and soil water storages (SSM and TWS) for each grid cell during growing season months when observations for at least 40 months were available. In addition to soil moisture, also air temperature (T_a) and net radiation (R_n) affect the vegetation functioning. To focus exclusively on the effects of water availability on vegetation functioning, we corrected for the confounding effects of T_a and R_n, by computing the partial correlation between NIRv and water storages while controlling for T_a and R_n. Since we focus on understanding the role of soil moisture on vegetation functioning, which is primarily critical in water-limited conditions, we removed the grids cells with insignificant ($p <0.05$) and negative partial correlations from our analysis. Such negative partial correlations may hint at vegetation's converse effect on soil moisture (when increasing vegetation activity depletes the soil moisture) and a negative correlation could occur in the grid cells where water limits vegetation productivity through oxygen limitation (Ohta et al., 2014).

To analyse how the importance of SSM and TWS changes during dry months, we selected the months with the lowest 10% SSM for each grid cell. The partial correlations between NIRv and water storages, r(NIRv~SSM) and r(NIRv~TWS) were calculated separately for dry months. To focus on vegetation response to similar extent of dryness spatially, only grid cells with greater than 100 monthly observations were considered for the dry months analysis. In addition, only the grid cells which had significant partial correlation in growing season months were included for the dry months analysis.

https://doi.org/10.5194/egusphere-2023-770
Preprint. Discussion started: 8 May 2023

After computing the partial correlations, we grouped the grid cells by aridity and tree cover or land cover classes and calculated the mean correlation, for each aridity-tree cover class with sufficient number of observations for both growing season and dry months. This allowed us to analyse the evolution of correlations and the difference between the partial correlation across aridity and vegetation classes.

Moreover, to test the robustness of the results, we did additional correlation analyses, for which we correlated the SIF (instead of NIRv) with SSM and TWS. The analyses with SIF were performed at a spatial resolution of 0.5°, at which SIF data was available.

2.2.3 Attribution Analysis

We used a random forest model to understand the spatial variability in the relevance of SSM versus TWS for NIRv. Random forest is a nonparametric based regression algorithm which does not require any statistical assumptions on the predictor and target variables which makes it particularly useful for detecting the nonlinear relationship (Breiman, 2001). Given potential nonlinear impacts of various factors (climate, soil types, vegetation) on the relationship between moisture storages and vegetation functioning, this study employed the random forest method to assess the relative contributions of these variables.

In our study, 13 predictors were included in the random forest model based on their potential physical relevance to the target variable, which is the difference in correlation between SSM and TWS with NIRv in growing season months. These predictors included mean and standard deviation of climate variables (T_a, R_n, and P), aridity index, topographical variables (elevation and slope), vegetation variable (tree cover), soil-related variables (fraction of clay and sand), and percentage of irrigated areas for each grid cell. We calculated the mean and standard deviation of the climate variables only during the growing months (the months with $SIF_{mean\text{-}monthly} < 0.2$ mW/m2/sr/nm were excluded). Furthermore, only the grid cells having significant and positive partial correlation between NIRv and SSM as well as NIRv and TWS during growing season-months were included in the random forest analysis. For training a random forest model, we used the "xgboost" package in R (Chen & Guestrin, 2016).

We further incorporate SHAP (SHapley Additive exPlanations) values for interpreting the predictions of the random forest model (Lundberg et al., 2020). The SHAP value for a feature is the average difference in prediction of the model when that feature is included compared to when it is excluded, over all possible combinations of features. By calculating SHAP values for each feature in the model, we identified which features were most important in explaining the spatial variability in the relevance of SSM versus TWS. For calculating the SHAP values, we employed "SHAPforxgboost" package in R.

https://doi.org/10.5194/egusphere-2023-770
Preprint. Discussion started: 8 May 2023

3. Results and Discussion

3.1 Coupling of vegetation functioning with surface soil moisture and total water storage in the growing season

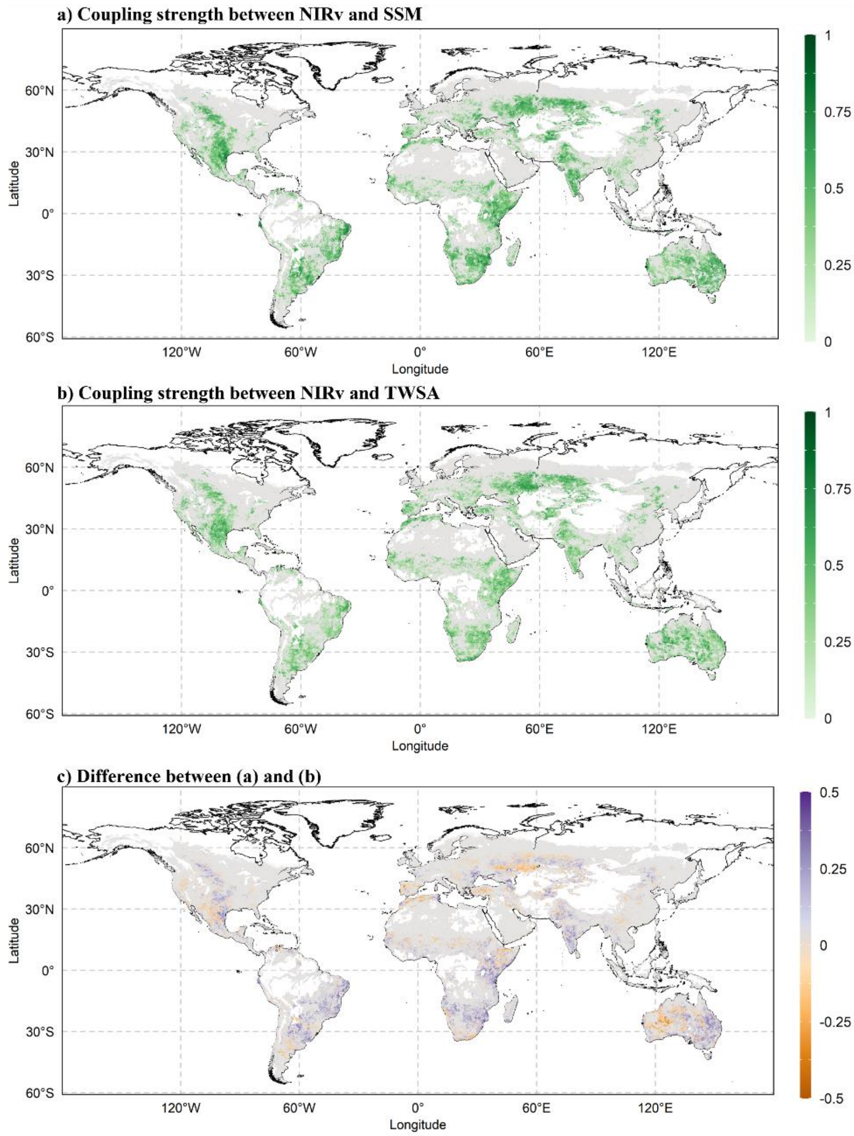

Figure 1: Coupling strength between vegetation functioning (NIRv) and (a) near-surface soil moisture (SSM), and (b) total water storage (TWS) during the growing season months. Monthly anomalies of all variables are used to calculate the partial correlation. (c) Difference between (a) and (b). The purple colour in (c) indicates the greater partial correlation of NIRv with SSM compared to the partial correlation of NIRv with TWS while orange colour indicates the opposite. Grid cells with significant ($p < 0.05$) and positive relationships for both correlations (a) and (b) are shown in (c) with blueish and orange colours. Light grey colour indicates insignificant and/or negative partial correlations between NIRv and water storage.

The partial correlation of NIRv with near-surface soil moisture varies globally during growing-season months (**Figure 1a**). NIRv demonstrates stronger correlation with near-surface soil moisture within semi-arid climates, Central North America,

https://doi.org/10.5194/egusphere-2023-770
Preprint. Discussion started: 8 May 2023

South America, regions in South Africa and Australia. The correlation is stronger in Southern Europe and the Mediterranean region compared to central and Northern Europe. The correlation gradient from the hot and dry Mediterranean region to wet and cold Northern Europe corresponds to the gradient of water-limited ecosystems to energy-limited ecosystems obtained in other studies (Denissen et al., 2022; Teuling et al., 2009).

The global correlation of NIRv with TWS follows a similar pattern as with SSM (**Figure 1b**) in growing-season months. The correlation of NIRv with TWS is higher in drier central northern America and Australia compared to other regions. The similarities in the correlation of NIRv with SSM and TWS are expected because the monthly anomalies of SSM and TWS are highly correlated during growing season months in most of our study area **(Figure S2).**

The difference between the partial correlation of NIRv with SSM and TWS (**Figure 1c**) indicates that the NIRv correlates stronger with TWS in Western America, Southern Europe, and arid regions of Australia compared to other regions globally during growing-season months. In South America and Southern Africa, however, the NIRv shows a stronger correlation with SSM. It is difficult to determine which soil water storage (SSM or TWS) is more critical for vegetation, because the near-surface soil moisture is included in the measurements of TWS, and both datasets have very different noise levels.

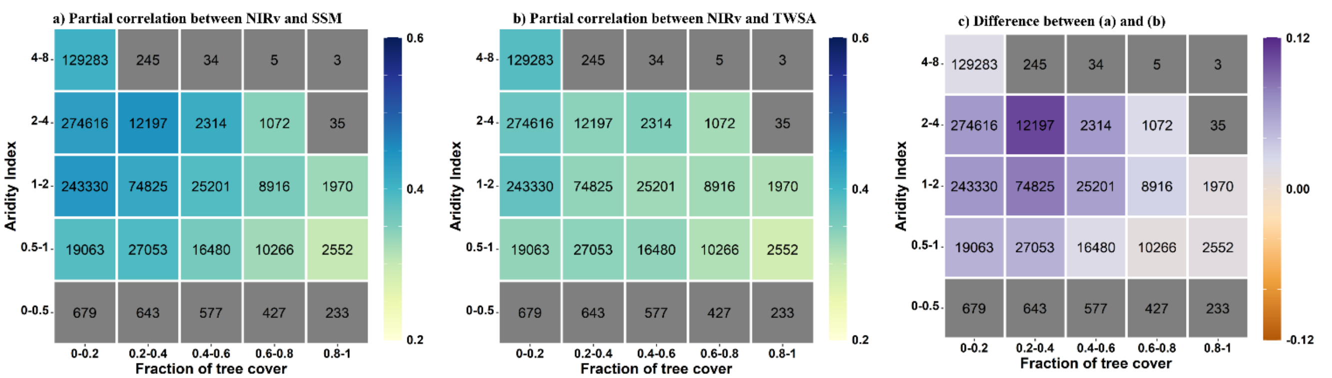

Figure 2: Summarising the coupling strengths of vegetation functioning (NIRv) with (a) near-surface soil moisture (SSM) and (b) terrestrial water storage (TWS) in the growing season-months across climate (aridity index) and vegetation regimes (fraction of tree cover). (c) shows the difference between (a) and (b). Numbers within the boxes denote the number of grid cells for each aridity-tree cover class. Aridity-tree cover classes containing less than 1000 grid cells are shown in grey. The colour bar indicates a mean partial correlation for each class. Only grid cells with significant ($p < 0.05$) and positive relationships are considered.

Next, we analyse the partial correlation between NIRv and soil water storages across different aridity and tree cover fraction classes during growing season months. For this, we group the grid cells into different aridity and tree cover fraction classes and then compute mean partial correlation for each class with more than 1000 grid cells. We find that the partial correlation of NIRv with SSM (**Figure 2a**) increases with increasing aridity for aridity index 0.5 – 4. This can be attributed to the

https://doi.org/10.5194/egusphere-2023-770
Preprint. Discussion started: 8 May 2023

intensification of water stress on vegetation under increasingly arid conditions, resulting in a stronger correlation between NIRv and SSM. However, for a further increase in aridity (4-8), the strength of the correlation of NIRv with SSM declines. This is due to a low soil moisture availability and low temporal variability under extremely arid conditions (**Figure S7**). The pattern of increasing correlation along aridity index is also observed in the partial correlation between NIRv and TWS. (**Figure 2b**).

Furthermore, the correlation of NIRv with SSM decreases for higher tree cover fractions **(Figure 2a**). However, such a gradient along tree cover fraction is less pronounced in the partial correlation of the NIRv with TWS (**Figure 2b**). This overall depicts that the coupling of vegetation functioning with SSM is generally higher for non-forested areas compared to forested areas while this gradient is less pronounced in the case of TWS.

It is difficult to conclude which soil moisture storage (SSM or TWS) is more important for a certain aridity-vegetation class because it inherently includes the difference in noise levels associated with SSM and TWS. However, we can compare the evolution of the gradient along tree cover or aridity index and assert how the relevance of SSM and TWS changes with varying tree cover or aridity index. Taking this into account, we find that NIRv correlates more strongly with near-surface soil moisture compared to terrestrial water storage in semi-arid regions with low tree cover (**Figure 2c**), suggesting that the vegetation preferentially takes up water from SSM whenever available to meet its transpiration demand. This might be due to lower energy expenditure on root water uptake, abundant nutrients and reduced chance of root water logging in the near-surface soil moisture (Andrew Feldman et al., 2022; Schenk & Jackson, 2002; Tao et al., 2021). Conversely, the correlation between the NIRv and TWS in arid areas (AI 4-8) and regions with a high fraction of tree cover is equivalent to or greater than that of SSM, suggesting that trees can utilise their extensive root systems to access deeper soil moisture, as observed in arid vegetation. This is consistent with previous studies reporting that the vegetation dependence on sub-surface soil moisture is higher in arid and seasonal-arid climates (Miguez-Macho & Fan, 2021).

Note that while our analysis focuses on regions with water-controlled vegetation as denoted by significantly positive correlations between NIRv and the considered soil water storages, some of these grid cells are located in comparatively wet climate regimes with aridity index values between 0.5 and 1 (**Figure 2**). This highlights the relevance of non-climatic factors such as soil and vegetation types or topography in determining vegetation-water relationships in addition to the climate regime. Next to this, in **Figure 2c** it seems that the relevance of terrestrial water storage is comparatively higher in wet climate (aridity 0.5-1) than in transitional climate regimes (aridity 1-2) as shown with the smaller correlation differences. This, however, is probably not the case and simply a reflection of reduced variability in surface soil moisture (**Figure S7**).

To ascertain that our results are not impacted by outliers, we analysed the heatmaps with 10^{th} and 90^{th} percentile correlation values for each aridity-vegetation class, instead of the mean correlation value (**Figure S3**). This shows consistent patterns of the partial correlation of NIRv with soil water storages as **in Figure 2** and indicates that the gradients with tree cover and aridity are valid throughout the entire dataset.

3.2 Coupling of vegetation functioning with surface soil moisture and total water storage in dry months

The correlation between NIRv and soil water storage increases during dry months (**Figure 3a,b**) compared to growing season months (**Figure 2a,b**). This increase is consistent for both SSM and TWS and across all tree cover fractions and aridity classes. This is because the water limitation on vegetation increases in dry months and so does the vegetation's sensitivity to the moisture. During the dry months, the correlation with near-surface soil moisture tends to rise, but the correlation with terrestrial water storage increases even more significantly **(Figure 3c**). This hints the relevance of deeper water resources during periods of scarce rainfall. The partial correlation maps **(Figure S4**) also reveal that NIRv's correlation with TWS increases more than its correlation with SSM for most grid cells.

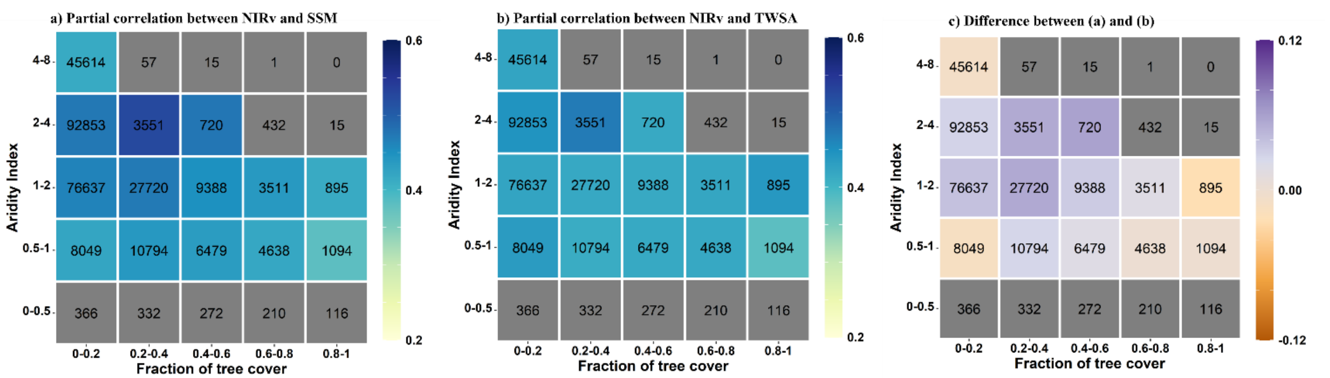

Figure 3: Similar to Figure 2, but only considering the 10% driest months in each grid cell.

During dry months, the number of analysed grid cells (**Figure 3**) is lower compared to all growing season months (**Figure 2**). We performed a reanalysis of the correlation patterns within aridity-tree cover classes by selecting only those grid cells that displayed significant and positive partial correlation between NIRv and soil water storages during both the dry months and the growing season months. The results demonstrate that the previously observed patterns remain valid, thereby eliminating the impact of the differing numbers of grid cells analysed. (**Figure S5)**.

https://doi.org/10.5194/egusphere-2023-770
Preprint. Discussion started: 8 May 2023

3.3 Climate, vegetation, and topographic controls on the relevance of surface soil moisture vs. total water storage on vegetation

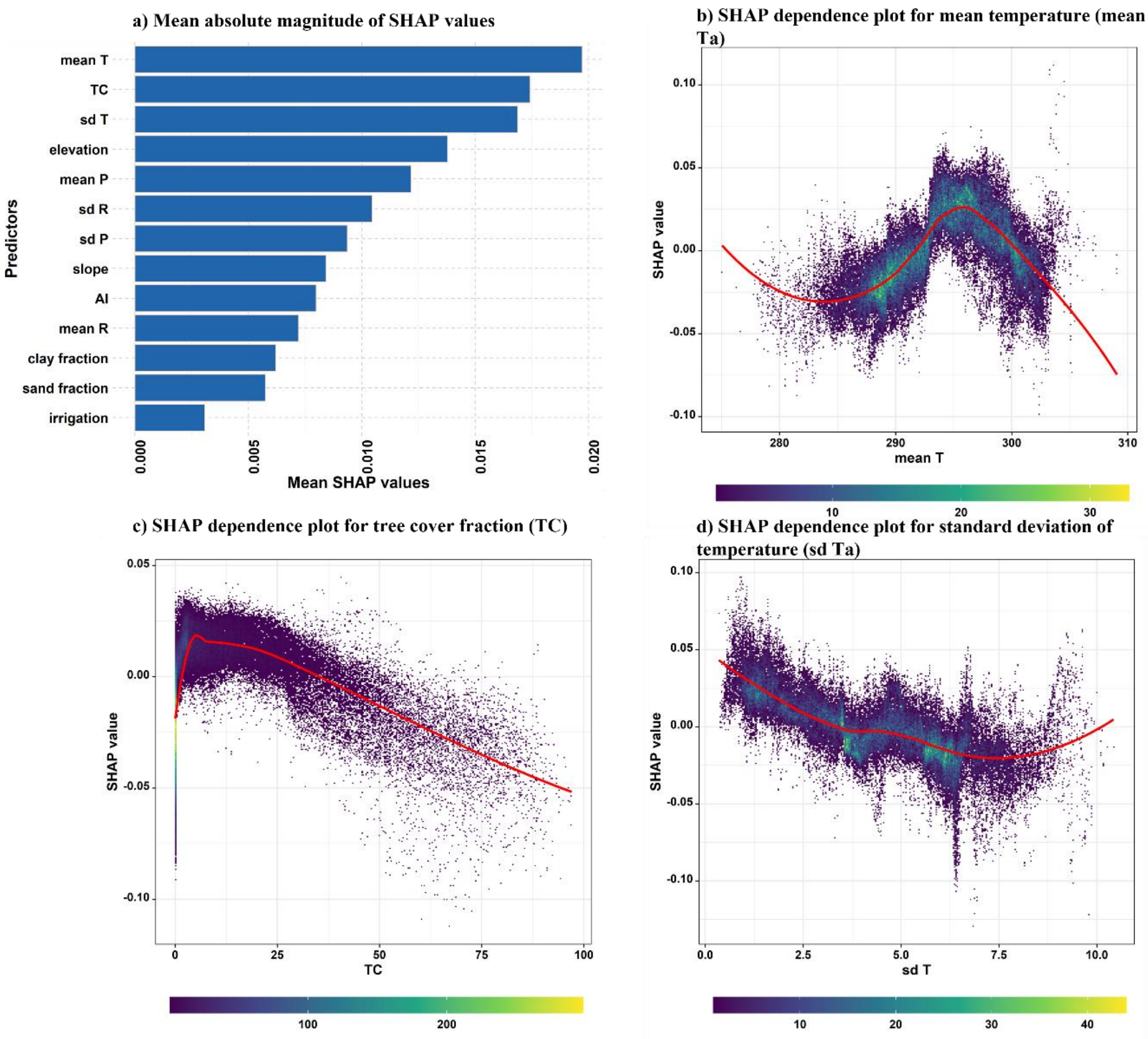

Figure 4: (a) Global feature importance based on the mean absolute magnitude of the SHAP values. The higher the mean SHAP values, the greater the predictor's relevance. (b-d) Evaluation of SHAP values (=contributions to the correlation difference illustrated in Figure 1c) against predictor values for the 3 most relevant predictors mean temperature during the growing season months (mean T_a), tree cover fraction (TC), and variability of temperature during the growing season months (sd T_a). The colour indicates the density of data points. For plotting (b), (c) and (d), only 10 percent random samples of the whole dataset are utilised.

We use a random forest model to understand the spatial variability in the relevance of SSM versus TWS for NIRv. The model was trained with 13 climatic, vegetation, and topographic predictors against the target variable which is the difference of the partial correlations of NIRv with SSM and TWS during growing season-months (R^2 = 0.64, see **methods section 2.2.3**). The mean absolute SHAP value plot shows that the climate variables (mean and standard deviation of T_a) and tree cover are most important variables for explaining the spatial variability in the relative importance of SSM vs. TWS for vegetation functioning

https://doi.org/10.5194/egusphere-2023-770
Preprint. Discussion started: 8 May 2023

(**Figure 4a**). This overall highlights that the relative importance of SSM vs. TWS for the vegetation is broadly controlled by climate, influencing water availability and vegetation type, reflecting the local adaptation of ecosystem (Stocker et al., 2023).

The relative importance of SSM and TWS varies non-linearly with the mean growing season temperature (**Figure 4b)**. TWS tends to be more crucial for vegetation functioning in areas with low (approximately below 20°C) or high (above 27°C) growing season temperatures, while SSM has greater importance in regions with moderate growing season air temperatures. One possible explanation for this trend is that high temperatures induce a strong atmospheric water demand that dries near-surface soil layers, which leads vegetation to increase water extraction from deep soils. In contrast, SSM is more available during growing seasons characterised by moderate temperatures. Regions that experience relatively cold growing season temperatures exhibit stronger temperature and weather variability that may contribute to longer dry periods and, thus, emphasises the importance of deeper soil moisture for vegetation functioning. However, it should be noted that our findings regarding the relevance of TWS at high temperatures must be interpreted with caution due to the exclusion of most tropical forest regions from our analysis (**Figure S6**). As a result, most warm regions are dry, and there are only a few hot and wet regions included in our training data.

In addition to mean growing season T_a, tree cover fraction is an important factor in determining the relevance of SSM and TWS for vegetation functioning (**Figure 4c**). Regions with a high tree cover are more dependent on TWS, as trees generally have deeper root systems that allow them to adjust water uptake between different depths (Tao et al., 2021). Grasslands on the other hand have shallow roots that are more susceptible to surface soil moisture variations (Yang et al., 2014).

Not only the mean of the growing season temperature, but also its variability is crucial for explaining the significance of SSM and TWS for vegetation functioning (**Figure 4d**). A higher temporal variability in temperature increases the importance of TWS for vegetation. This is because atmospheric water demand scales with temperature. Hence, higher variability in temperature implies more peaks in related atmospheric water demand which is a stronger incentive for plants to access deeper water storages which are more often available to meet the vegetation´s transpiration demand.

Figure S8 illustrates the effect of the other six important predictors on the model output. Apart from climatological parameters (mean P, variability in R_n, and P, and aridity index), elevation and slope explain part of the variability in the relevance of SSM vs. TWS for NIRv. Although the reasons for increasing relevance of TWS for vegetation functioning at higher elevation remain unclear, it may be due to elevation's strong correlation with other climatic variables such as T_a and P.

3.4 Robustness Tests

Although NIRv can largely reflect vegetation functioning (Badgley et al., 2017), we repeat our analysis with SIF, which is an alternative and independent indicator for vegetation functioning and shows a near-linear relationship with gross primary productivity at the ecosystem level (Guanter et al., 2012). However, SIF is only available at a coarse resolution of 0.5 degree. The partial correlations, r(SIF~SSM) and r(SIF~TWS) largely agree with the pattern of r(NIRv~SSM) and r(NIRv~TWS) across varying aridity index and tree cover classes (**Figure S9**). This suggests that our overall conclusion on the relevance of SSM or TWS for vegetation functioning is robust across different indicators of vegetation productivity.

The partial correlation of NIRv with TWS is confounded by the presence of SSM within TWS, which makes it challenging to determine the relative importance of SSM and TWS for vegetation functioning. To address this issue, we re-calculated the partial correlation of NIRv with TWS while additionally controlling for SSM (next to T_a and R_n) during growing season months. With this additional control variable, we observed fewer grid cells with positive and significant correlations compared to the analysis without controlling for SSM. Additionally, the magnitude of the partial correlation of NIRv with TWS slightly decreased in most grid cells when controlling for SSM (**Figure S10**). Nevertheless, we still observed the decreasing relevance of SSM and increasing relevance of TWS along an increasing tree cover fraction. Similar gradient across the aridity index is also observed in this analysis controlling for SSM. Thus, we conclude that our findings hold even after controlling for the effect of SSM in TWS .

4. Summary and Conclusions

In this study we compare the relevance of near-surface soil moisture and of terrestrial water storage for vegetation functioning across the globe. We find that in semi-arid regions and regions with low tree cover, vegetation preferentially utilises the water from shallow soil, which is related to continuous availability of near-surface water availability and lack of deep rooting systems respectively. By contrast, in mostly forested regions and in relatively dry climate regimes, the correlation with terrestrial water storage is comparable or higher than with near-surface soil moisture, indicating that trees and vegetation in arid regions use their deep root systems to access deeper soil moisture.

We also find that vegetation's preferential water uptake depth changes over time. During particularly dry months, the relative importance of terrestrial water storage is higher, highlighting the importance of deep water resources during periods of low soil water availability. This is in line with previous studies showing changes in vegetation's water uptake depth during drought periods at small spatial scales where accessing water in deeper soil layers helps plants to alleviate water stress and maintain transpiration (Migliavacca et al., 2009; Tao et al., 2021).

Furthermore, we show that the spatial variability of the importance of near-surface soil moisture vs. terrestrial water storage for vegetation functioning is influenced by temperature and the fraction of tree cover. This emphasises the role of climate in determining shallow vs. deep soil water resources, and the role of vegetation in adapting to different soil water availability patterns.

https://doi.org/10.5194/egusphere-2023-770
Preprint. Discussion started: 8 May 2023

Vegetation functioning and soil water storages are generally coupled in both directions, i.e. while soil moisture availability affects vegetation functioning (positive coupling), this in turn also affects soil moisture through transpiration (negative coupling). As our study focuses on water-controlled vegetation we only consider positive couplings and filter out grid cells with negative correlations. Future research may consider the relevance of soil moisture across depths for the positive coupling regions.

Overall, our analysis illustrates that satellite-based data can be used for belowground analysis at large spatial scales thanks to the fact that satellite retrievals can assess soil water storage dynamics across depths and because vegetation in water-controlled areas can be used as an indicator of soil water dynamics. Such novel ways to improve our understanding of belowground water dynamics is necessary and valuable as respective in-situ observations are scarce and of limited representativeness for larger areas, particularly given the typical spatial heterogeneity of soils and vegetation. Our results can further inform a better representation of belowground processes in global models in order to support more accurate projections of future changes in climate, water resources, and ecosystem services.

Data availability

The monthly SIF data is available from https://www.gfz-potsdam.de/sektion/fernerkundungund-geoinformatik/projekte/global-monitoring-of-vegetation-fluorescence-globfluo/daten.The NIRv was calculated from the red and near-infrared reflectance obtained from the MOD13C1 v006 product (https://lpdaac.usgs.gov/products/mod13c1v061/). The ESA-CCI soil moisture can be accessed through https://esa-soilmoisture-cci.org/ and Terrestrial Water Storage Anomaly data can be accessed through https://podaac.jpl.nasa.gov/dataset/TELLUS_GRACGRFO_MASCON_CRI_GRID_RL06_V2. The ERA5 climate variables are available from https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5 . Tree cover fraction data is available from the AVHRR vegetation continuous fields products https://lpdaac.usgs.gov/products/vcf5kyrv001/, land cover data is available from https://www.esa-landcover-cci.org/, and topographic data is available via https://www.earthenv.org/topography. Similarly, the irrigation fraction data could be accessed from https://mygeohub.org/publications/8 .

Competing Interests

The contact author has declared that none of the authors has any competing interests.

Acknowledgements

The authors thank Ulrich Weber for help with obtaining and processing the data, and the Hydrology–Biosphere–Climate Interactions group at the Max Planck Institute for Biogeochemistry for fruitful discussions. Prajwal Khanal, Anne Hoek van Dijke and Rene Orth acknowledge funding by the German Research Foundation (Emmy Noether grant no. 391059971).

References

Amatulli, G., Domisch, S., Tuanmu, M.-N., Parmentier, B., Ranipeta, A., Malczyk, J., & Jetz, W. (2018). A suite of global, cross-scale topographic variables for environmental and biodiversity modeling. *Scientific Data*, *5*(1), 180040. https://doi.org/10.1038/sdata.2018.40

Andrew Feldman, Gianotti, D., Dong, J., Akbar, R., Crow, W., McColl, K., Nippert, J., Tumber-Dávila, S. J., Holbrook, N. M., & al., F. R. et. (2022). Satellites capture soil moisture dynamics deeper than a few centimeters and are relevant to plant water uptake. *Earth and Space Science Open Archive*. https://doi.org/10.1002/essoar.10511280.1 DA - 2022

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

Breiman, L. (2001). [No title found]. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Budyko, M. I. (1974). *Climate and life*. Academic press.

Capehart, W. J., & Carlson, T. N. (1997). Decoupling of surface and near-surface soil water content: A remote sensing perspective. *Water Resources Research*, *33*(6), 1383–1395. https://doi.org/10.1029/97WR00617

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

Denissen, J. M. C., Teuling, A. J., Pitman, A. J., Koirala, S., Migliavacca, M., Li, W., Reichstein, M., Winkler, A. J., Zhan, C., & Orth, R. (2022). Widespread shift from ecosystem energy to water limitation with climate change. *Nature Climate Change*, *12*(7), 677–684. https://doi.org/10.1038/s41558-022-01403-8

Dorigo, W., Wagner, W., Albergel, C., Albrecht, F., Balsamo, G., Brocca, L., Chung, D., Ertl, M., Forkel, M., Gruber, A., Haas, E., Hamer, P. D., Hirschi, M., Ikonen, J., de Jeu, R., Kidd, R., Lahoz, W., Liu, Y. Y., Miralles, D., … Lecomte,

P. (2017). ESA CCI Soil Moisture for improved Earth system understanding: State-of-the art and future directions. *Remote Sensing of Environment*, *203*, 185–215.

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

Guanter, L., Frankenberg, C., Dudhia, A., Lewis, P. E., Gómez-Dans, J., Kuze, A., Suto, H., & Grainger, R. G. (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

Hansen, Matthew, & Song, Xiao-Peng. (2018). *Vegetation Continuous Fields (VCF) Yearly Global 0.05 Deg* [Data set]. NASA EOSDIS Land Processes DAAC. https://doi.org/10.5067/MEASURES/VCF/VCF5KYR.001

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., … Thépaut, J.-N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Humphrey, V., Zscheischler, J., Ciais, P., Gudmundsson, L., Sitch, S., & Seneviratne, S. I. (2018). Sensitivity of atmospheric CO2 growth rate to observed changes in terrestrial water storage. *Nature*, *560*(7720), 628–631. https://doi.org/10.1038/s41586-018-0424-4

Keenan, T. F., & Williams, C. A. (2018). The Terrestrial Carbon Sink. *Annual Review of Environment and Resources*, *43*(1), 219–243. https://doi.org/10.1146/annurev-environ-102017-030204

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

Koster, R. D., Guo, Z., Yang, R., Dirmeyer, P. A., Mitchell, K., & Puma, M. J. (2009). On the nature of soil moisture in land surface models. *Journal of Climate*, *22*(16), 4322–4335. https://doi.org/10.1175/2009JCLI2832.1

Landerer, F. W., & Swenson, S. C. (2012). Accuracy of scaled GRACE terrestrial water storage estimates. *Water Resources Research*, *48*(4). https://doi.org/10.1029/2011WR011453

Li, W., Migliavacca, M., Forkel, M., Denissen, J. M. C., Reichstein, M., Yang, H., Duveiller, G., Weber, U., & Orth, R. (2022). Widespread increasing vegetation sensitivity to soil moisture. *Nature Communications*, *13*(1), 3959. https://doi.org/10.1038/s41467-022-31667-9

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*(11). https://doi.org/10.1029/2021GL092856

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, *2*(1), 56–67. https://doi.org/10.1038/s42256-019-0138-9

Migliavacca, M., Meroni, M., Manca, G., Matteucci, G., Montagnani, L., Grassi, G., Zenone, T., Teobaldelli, M., Goded, I., Colombo, R., & Seufert, G. (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

Miguez-Macho, G., & Fan, Y. (2021). Spatiotemporal origin of soil water taken up by vegetation. *Nature*, *598*(7882), 624–628. https://doi.org/10.1038/s41586-021-03958-6

Mohammed, G. H., Colombo, R., Middleton, E. M., Rascher, U., van der Tol, C., Nedbal, L., Goulas, Y., Pérez-Priego, O., Damm, A., Meroni, M., Joiner, J., Cogliati, S., Verhoef, W., Malenovský, Z., Gastellu-Etchegorry, J.-P., Miller, J. R., Guanter, L., Moreno, J., Moya, I., … Zarco-Tejada, P. J. (2019). Remote sensing of solar-induced chlorophyll fluorescence (SIF) in vegetation: 50 years of progress. *Remote Sensing of Environment*, *231*, 111177.

Ohta, T., Kotani, A., Iijima, Y., Maximov, T. C., Ito, S., Hanamura, M., Kononov, A. V., & Maximov, A. P. (2014). Effects of waterlogging on water and carbon dioxide fluxes and environmental variables in a Siberian larch forest, 1998–2011. *Agricultural and Forest Meteorology*, *188*, 64–75. https://doi.org/10.1016/j.agrformet.2013.12.012

Orth, R. (2021). When the Land Surface Shifts Gears. *AGU Advances*, *2*(2). https://doi.org/10.1029/2021AV000414

Qiu, R., Li, X., Han, G., Xiao, J., Ma, X., & Gong, W. (2022). Monitoring drought impacts on crop productivity of the U.S. Midwest with solar-induced fluorescence: GOSIF outperforms GOME-2 SIF and MODIS NDVI, EVI, and NIRv. *Agricultural and Forest Meteorology*, *323*, 109038. https://doi.org/10.1016/j.agrformet.2022.109038

Reynolds, C. A., Jackson, T. J., & Rawls, W. J. (2000). Estimating soil water-holding capacities by linking the Food and Agriculture Organization Soil map of the world with global pedon databases and continuous pedotransfer functions. *Water Resources Research*, *36*(12), 3653–3662. https://doi.org/10.1029/2000WR900130

Schenk, H. J., & Jackson, R. B. (2002). Rooting depths, lateral root spreads and below-ground/above-ground allometries of plants in water-limited ecosystems. *Journal of Ecology*, *90*(3), 480–494. https://doi.org/10.1046/j.1365-2745.2002.00682.x

Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., Orlowsky, B., & Teuling, A. J. (2010). *Investigating soil moisture-climate interactions in a changing climate: A review* (Vol. 99, Issues 3–4, pp. 125–161). https://doi.org/10.1016/j.earscirev.2010.02.004

Siebert, S., Kummu, M., Porkka, M., Döll, P., Ramankutty, N., & Scanlon, B. (2015). *Historical Irrigation Dataset (HID)* [Data set]. MyGeoHUB. https://doi.org/10.13019/M20599

Stocker, B. D., Tumber-Dávila, S. J., Konings, A. G., Anderson, M. C., Hain, C., & Jackson, R. B. (2023). Global patterns of water storage in the rooting zones of vegetation. *Nature Geoscience*, *16*(3), 250–256. https://doi.org/10.1038/s41561-023-01125-2

Tao, Z., Neil, E., & Si, B. (2021). Determining deep root water uptake patterns with tree age in the Chinese loess area. *Agricultural Water Management*, *249*, 106810.

Teuling, A. J., Hirschi, M., Ohmura, A., Wild, M., Reichstein, M., Ciais, P., Buchmann, N., Ammann, C., Montagnani, L., Richardson, A. D., Wohlfahrt, G., & Seneviratne, S. I. (2009). A regional perspective on trends in continental evaporation. *Geophysical Research Letters*, *36*(2). https://doi.org/10.1029/2008GL036584

Xie, X., He, B., Guo, L., Miao, C., & Zhang, Y. (2019). Detecting hotspots of interactions between vegetation greenness and terrestrial water storage using satellite observations. *Remote Sensing of Environment*, *231*, 111259.

Yang, Y., Long, D., Guan, H., Scanlon, B. R., Simmons, C. T., Jiang, L., & Xu, X. (2014). GRACE satellite observed hydrological controls on interannual and seasonal variability in surface greenness over mainland Australia. *Journal of Geophysical Research: Biogeosciences*, *119*(12), 2245–2260. https://doi.org/10.1002/2014JG002670