remote sensing

MDPI

Article

The Value of L-Band Soil Moisture and Vegetation Optical Depth Estimates in the Prediction of Vegetation Phenology

Bonan Li *, Stephen P. Good and Dawn R. URycki

Department of Biological & Ecological Engineering, Oregon State University, Corvallis, OR 97331, USA; stephen.good@oregonstate.edu (S.P.G.); uryckid@oregonstate.edu (D.R.U.)
* Correspondence: libon@oregonstate.edu

Abstract: Vegetation phenology is a key ecosystem characteristic that is sensitive to environmental conditions. Here, we examined the utility of soil moisture (SM) and vegetation optical depth (VOD) observations from NASA's L-band Soil Moisture Active Passive (SMAP) mission for the prediction of leaf area index (LAI), a common metric of canopy phenology. We leveraged mutual information theory to determine whether SM and VOD contain information about the temporal dynamics of LAI that is not contained in traditional LAI predictors (i.e., precipitation, temperature, and radiation) and known LAI climatology. We found that adding SMAP SM and VOD to multivariate non-linear empirical models to predict daily LAI anomalies improved model fit and reduced error by 5.2% compared with models including only traditional LAI predictors and LAI climatology (average R^2 = 0.22 vs. 0.15 and unbiased root mean square error [ubRMSE] = 0.130 vs. 0.137 for cross-validated models with and without SM and VOD, respectively). SMAP SM and VOD made the more improvement in model fit in grasslands (R^2 = 0.24 vs. 0.16 and ubRMSE = 0.118 vs. 0.126 [5.7% reduction] for models with and without SM and VOD, respectively); model predictions were least improved in shrublands. Analysis of feature importance indicates that LAI climatology and temperature were overall the two most informative variables for LAI anomaly prediction. SM was more important in drier regions, whereas VOD was consistently the second least important factor. Variations in total LAI were mostly explained by local daily LAI climatology. On average, the R^2s and ubRMSE of total LAI predictions by the traditional drivers and its climatology are 0.81 and 0.137, respectively. Adding SMAP SM and VOD to these existing predictors improved the R^2s to 0.83 (0.02 improvement in R^2s) and reduced the ubRMSE to 0.13 (5.2% reduction). Though these improvements were modest on average, in locations where LAI climatology is not reflective of LAI dynamics and anomalies are larger, we find SM and VOD to be considerably more useful for LAI prediction. Overall, we find that L-band SM and VOD observations can be useful for prediction of LAI, though the informational contribution varies with land cover and environmental conditions.

Keywords: SMAP; soil moisture; vegetation optical depth; regression

check for updates

Citation: Li, B.; Good, S.P.; URycki, D.R. The Value of L-Band Soil Moisture and Vegetation Optical Depth Estimates in the Prediction of Vegetation Phenology. *Remote Sens.* **2021**, *13*, 1343. https://doi.org/10.3390/rs13071343

Academic Editors: Arnaud Mialon, Nemesio Rodriguez-Fernandez, Mariette Vreugdenhil, Tianjie Zhao and Roberto Fernández Morán

Received: 18 February 2021
Accepted: 29 March 2021
Published: 1 April 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Vegetation phenology is the study of plant life cycles such as budburst, flowering and leaf senescence, and the impact of seasonal and inter-annual climate variability on the timing and magnitude of these events [1]. Recent climate warming has impacted vegetation phenology, which may consequently have profound implications for agriculture and forest productivity [2]. Accurately tracking land surface vegetation phenology is therefore critical to enhance our understanding of the water–food–energy nexus and carbon exchanges between terrestrial ecosystems and the atmosphere [3]. Satellite derived vegetation indices are frequently used to track and map the spatial and temporal dynamics of vegetation phenology at large scales and can overcome the limitations of traditional phenology studies that heavily rely on in situ observations [4]. Various vegetation indices have been applied to assess and track vegetation phenology. The Normalized Difference Vegetation Index

(NDVI) is considered a proxy of plant phenology and was first applied to the study of the spring vegetation growing season in the 1970s [5]. Since then, numerous phenology studies using NDVI have been conducted across different biomes and climate regimes. Shifts in the start of growing season were identified across the northern hemisphere using NDVI during the 1980s and 1990s [6]. The Enhanced Vegetation Index (EVI) has been applied to map global vegetation phenology and investigate the peak growing season of dense evergreen forests [7]. Normalized Difference Water Index (NDWI) has the potential to determine the phenological dates over boreal regions [8]. Although various vegetation indices have been applied to track and map vegetation phenology, most of these vegetation indices represent vegetation "greenness," which corresponds to how lush the vegetation is [9], and are not precise indicators of vegetation leaf area [10]. In addition, the retrieval accuracy of satellite vegetation indices is often affected by satellite viewing geometry, atmospheric condition, and background soil properties [11].

One key aspect of vegetation phenology that is needed to be predicted for a number of environmental applications are the dynamics of leaf area index (LAI), a commonly used metric characterizing the surface area available for photosynthesis in an ecosystem. LAI is the one-sided green leaf area per unit vegetated ground surface in a vegetation canopy and can better reflect plant growth and leaf density than do optical indices based on greenness [12]. A large set of physically and empirically based algorithms coupled with remotely sensed datasets have been proposed to estimate LAI at large scales [13–15]. Various of regression techniques such as simple linear regression, multivariate linear regression, and partial least squares regression have been used to empirically link LAI with vegetation indices by combing different spectral observations [16,17]. Machine learning algorithms and geostatistical prediction methods were attempted in previous studies to explore the non-linearity between LAI and a set of remotely sensed bio-geophysical variables [18–20]. Process-based model was also used to estimate LAI by optimizing the parameters that are related to LAI in a terrestrial ecosystem model [21].

LAI has been shown to be related to plant water content, a canopy characteristic that can be inferred from vegetation optical depth (VOD) at different microwave frequencies [22]. Soil Moisture and Ocean Salinity (SMOS) L-band VOD was found to be highly correlated with LAI, with correlation strength varying with landcover type [23]. In particular, LAI was found to vary synchronously with VOD in boreal and temperate forests, while this relationship was found to be reversed for the tropical woodlands [24]. A comparative study has also shown that SMOS L-band VOD contains information about LAI that varies for different crop zones in the USA [25]. A conceptual framework showed that VOD sensitivity to leaf water potential and LAI varies with climate conditions [26]. These studies suggest that VOD may provide additional information for LAI prediction, yet the accuracy of LAI predictions based on microwave observations is not known.

Vegetation phenology and soil moisture (SM) are linked in the land surface system [27]. Microwave remote sensing of SM has shown to be useful for characterizing vegetation soil water use behavior by inferring key plant functional parameters such as soil water potential [28]. A global study on vegetation dynamics has shown that the synthesized microwave SM dataset can model and estimate vegetation phenology in drylands [29]. In addition, microwave retrieved SM often requires the description of the overlaying canopy water status, which is related to the vegetation phenology [30]. These studies all suggest that satellite retrieved SM may possess useful information about plant development and vegetation phenology (i.e., LAI).

Launched in 2015, NASA's Soil Moisture Passive Active (SMAP) mission is the most recent space-borne SM dedicated satellite. SMAP uses L-band measurements to map surface/root zone SM and its freeze–thaw state [31]. SMAP SM products have been extensively validated against in situ SM in a sequence of core and spare validation campaigns [32,33]. Results from these validations have shown that SMAP SM estimates meet the expected mission performance of 0.04 m^3/m^3 (unbiased root mean square error, ubRMSE) volumetric

SM [34,35]. Global scale inter-comparisons among SMAP and other satellite SM suggested that SMAP is more robust and exhibits smaller ubRMSE [36].

The SMAP Level-2 dataset, provides SM and vegetation opacity data that can be converted to VOD by accounting for the satellite viewing angle, with different active, passive, and active/passive derived products at ascending (afternoon retrieval) and descending (morning retrieval) orbits. The SMAP SM product has been used to understand plant water uptake activities in response to soil water availability [28,37]. However, limited research has focused on the application of SMAP Level-2 VOD product. This product measures how microwaves are attenuated through vegetation canopy, which is directly related to vegetation water content and above-ground canopy biomass [36]. Furthermore, no effort has been made to quantify the added value of SM Wildfire Predictions AP Level-2 SM and VOD in the prediction of LAI. Given the known response of vegetation to SM status [37], we hypothesize that combining SMAP observations with traditionally phonological drivers (temperature, precipitation, and radiation) will improve the predictions of LAI.

The objective of this study is to evaluate how much information SMAP Level-2 SM and VOD products can provide in the prediction of LAI at daily scales. We leverage mutual information theory to determine the additional information that SMAP Level-2 SM and VOD provided for LAI prediction. We then use a multivariate non-linear empirical model (random forest regression) to quantify the additional information from SMAP Level-2 SM and VOD. Finally, we assess the conditions under which SMAP SM and VOD provide more predictive information.

2. Materials

2.1. Study Sites

This study was conducted at 500 sites that were initially selected at random from within the contiguous United States between 48.57° N and 28.33° N, and between −65.07° E and −124.16° E. The selected sites were excluded if they were identified as water bodies or bare ground as indicated by a mean annual LAI of zero. The selected sites were further excluded if no data were available in their correspondent quality-controlled SMAP datasets. Our final set was comprised of 216 study sites, including 120 grasslands (55% of the total sites), 47 croplands (22%), 22 savannas (10%) and 27 shrublands (13%; Figure 1). Mean annual LAI and mean annual precipitation of these study sites range from 0.14 to 2.4, and from 133 mm to 1895 mm, respectively.

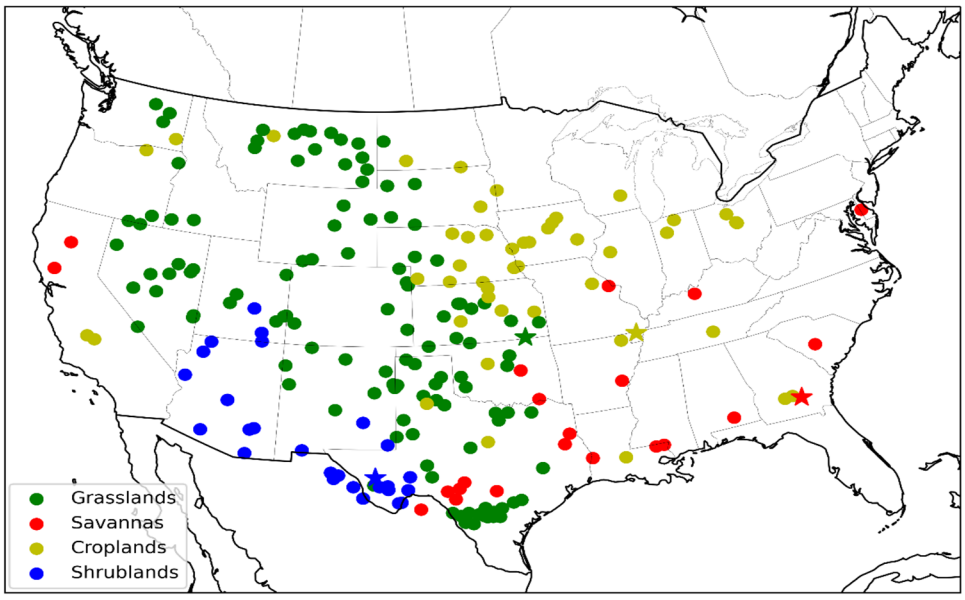


Figure 1. Spatial distribution of selected terrestrial study sites and their correspondent dominant landcovers (stared locations are representative sites of each landcover as shown in Figures S1–S4).

2.2. LAI Datasets

Satellite derived LAI was used as an estimate of the canopy leaf area status because of its temporal consistency and calibration to ground based datasets. A time series of MODIS LAI (MCD15A3H Version 6) from April 2004 to October 2019 was obtained for each site using MODIS Fixed Sites Subsetting and Visualization Tool [38,39], which produced a product with a 4-day temporal resolution and user-specified spatial extent. We specified an 8.5 km by 8.5 km MODIS LAI extent that was centered at each study site to match the spatial extent of SMAP Level−2 9 km product. The specified MODIS spatial extent contains 289 original 500-m by 500-m MODIS LAI pixels. The MODIS LAI dataset contains an intrinsic quality control metric characterizing if the estimated LAI meet the desired quality level. The obtained MODIS LAI was first filtered by the total number of pixels ($\geq$232) that passed quality control to ensure the accuracy of the dataset. All quality-controlled MODIS LAI estimates within a SMAP pixel were then average to represent the phenology status at each location. Previous studies have shown that MODIS LAI captures the seasonal dynamics of in situ LAI and performs better than other satellite derived LAI products [40,41]. To minimize large LAI data gaps, we excluded sites where over 30% of the data during our study period April 2015 to October 2019 was missing. After this step, the average number of data points per site ranged from 292 to a maximum of 412 data points.

As SMAP SM and VOD do not align with MODIS data in time, and because LAI observations can contain noise induced by environmental conditions such as clouds, a continuous and smoothed LAI time series was created. We first applied linear interpolation at each site to the quality-controlled LAI to generate continuous LAI at the daily scale. Next, the interpolated LAI was smoothed using Savitzky–Golay filter that consists of a local polynomial fitting with two parameters: order of the polynomial and temporal window. It has been previously suggested that a polynomial order of 2–4 and the window length of 72–112 days generally performs well [42]; these parameter bounds result in a total of 123 different possible combinations. At each site, we chose the parameter combination that minimized the squared differences between the smoothed LAI and interpolated LAI. We calculated the daily LAI anomaly and the daily mean LAI climatology (LAI_C) (both terms as explicitly defined later in Section 3.2) based on the smoothed LAI. See Figures S1–S4 for plotted time series of the smoothed LAI, LAI climatology, and LAI anomalies at a representative site for each landcover (the started locations in Figure 1).

2.3. Landcover Information

The yearly MODIS Landcover dataset (MCD12Q1) was obtained with the same user-specified spatial extent as LAI (see Section 2.2). The original spatial resolution of MCD12Q1 is 500 m and contains five types of landcover classification schemes [43]. We used the landcover scheme classified by International Geosphere–Biosphere Programme (IGBP). The IGBP classifies the land surface into 17 categories: evergreen needleleaf/broadleaf forest, deciduous needleleaf/broadleaf forest, Mixed forest, closed/open shrublands, woody savanna/savanna, grasslands, croplands, permanent wetlands, urban and build-up lands, croplands/natural vegetation mosaics, snow and ice, barren, and water bodies [44]. The landcover of each study site was the mode of the landcovers in the user-specified extent (8.5 km by 8.5 km, 289 original 500 m by 500 m MODIS pixels). Study sites dominated by woody savanna were further classified as savanna; closed/open shrubland were classified as shrubland; and cropland/natural vegetation mosaic were classified as cropland.

2.4. Grid-Scale Geophysical Data

We used SMAP Level-4 Global 3-hourly 9 km EASE-Grid Surface and Root Zone Soil Moisture Geophysical Data, Version 4 (SMAP-L4) from April 2015 to October 2019 to incorporate meteorological variables [45,46] that have been shown to be related plant phenology: temperature, radiation, and, precipitation [47,48]. SMAP-L4 geophysical datasets were generated using an ensemble Kalman filter data assimilation system that assimilates SMAP L-band brightness temperature measurements into the Catchment Land

Surface Model driven by Goddard Earth Observing System Model, Version 5 surface metrological forcing data with gauge-corrected precipitation [45]. The assimilation system interpolates and extrapolates model estimates and SMAP measurements in time and space producing geophysical variables at a 3-hourly temporal resolution on a 9 km modeling grid. We obtained a 3-hourly time series of total surface precipitation flux, surface temperature, and downward shortwave radiation from the SMAP-L4 dataset at each study site. The downward shortwave radiation and surface temperature were averaged to obtain the daily downward shortwave radiation (R [W/m^2]) and daily surface temperature (T [K]). The 3-hourly total surface precipitation flux was averaged and then converted to daily precipitation (P [mm/day]). The R, T, and P time series at a representative site for each landcover (marked as stars in Figure 1) are shown in Figures S1–S4.

2.5. L-Band Microwave Data

SMAP SM and VOD time series from April 2015 to October 2019 were acquired from SMAP Enhanced Level-2 Radiometer Half-Orbit 9 km EASE-Grid Soil Moisture, Version 3 (SMAP-L2) [46,49]. SMAP-L2 was generated using SMAP Level-1B interpolated antenna temperatures as primary input and other datasets such as surface temperature, soil texture as ancillary inputs. Three different retrieval algorithms were implemented in the SMAP-L2 operational processing software that includes Single Channel Algorithm Horizontal and Vertical polarization (SCA-H and SCA-V) and Modified Dual Channel Algorithm (MDCA). SMAP retrieves SM based on the '*tau-omega*' model, a well-known radiative transfer-based SM retrieval framework in the passive microwave SM community [31]. SCAs requires the horizontally or vertically polarized brightness temperature as the main input and parameterized by overlaying vegetation and soil surface information. In general, the SCAs link one unknow (SM) with one equation. In contrast, MDCA feed the '*tau-omega*' model with initial guesses of surface SM and vegetation optical depth [50]. The guesses of SM and vegetation optical depth are adjusted iteratively until they minimize the squared distance difference between satellite observed brightness temperatures. The guesses of SM and VOD are adjusted iteratively until they minimize the squared distance difference between satellite observed brightness temperatures and the estimated brightness temperature. Detailed discussions of these algorithms have been previously published in [50,51].

We used the MDCA retrieved SM and VOD time series in this study. The SM and VOD datasets were filtered by the SM quality flag to minimize bias induced by poor data quality. SMAP retrieves SM and VOD in mornings and afternoons, so we averaged the quality-controlled morning and afternoon SM and VOD estimates to produce time series of mean daily SM and VOD. See Figures S1–S4 for plotted time series of SM and VOD at a representative site for each landcover (the stared locations in Figure 1).

3. Methodology

3.1. Shannon's Entropy and Mutual Information

Shannon's entropy represents the amount of information that is needed to fully describe a random variable [52]. Shannon's entropy of a single random variable is defined as

$$H(\mathrm{X}) = -\sum_x p(x) log_2 p(x) \quad (1)$$

where p(x) is the probability mass function associated with random variable X. We computed p(x) by discretizing X to a fixed number of equal bins. The bin number of X is determined by Scott's rule, which bins the random variable X based on the number of data points and standard deviation of X [53]. We then applied the Miller-Madow correction and a normalization method [54] to reduced error that may be induced by data length of X. The corrected and normalized entropy ($\widetilde{H}$, here after entropy) is expressed as

$$\widetilde{H}(\mathrm{X}) = \frac{H(X) + \frac{K-1}{2n}}{log_2 n} \quad (2)$$

where $H(X)$ is the Shannon's entropy, n is the number of data points of X, K is the number of non-zero bins while estimating $p(x)$.

For multiple variables X_i or a set of random variables $\{X_1, \ldots, X_N\}$, the joint entropy is

$$H(X_1, \ldots, X_N) = -\sum_{x_1} \ldots \sum_{x_N} p(x_1, \ldots, x_N) log_2 p(x_1, \ldots, x_N) \quad (3)$$

where $p(x_1, \ldots, x_N)$ is the joint probability mass function of a set of random variable $\{X_1, \ldots, X_N\}$ that is estimated using the Scott's rule data binning method [53]. We applied the same correction and normalization method as in Equation (2). The corrected and normalized joint entropy ($\widetilde{H}$, here after joint entropy) is

$$\widetilde{H}(X_1, \ldots, X_N) = \frac{H(X_1, \ldots, X_N) + \frac{K-1}{2n}}{log_2 n} \quad (4)$$

where $H(X_1, \ldots, X_N)$ is the original joint entropy from Equation (3), n is the number of data points of $\{X_1, \ldots, X_N\}$, K is the number of non-zero bins in estimating the joint probability mass function in Equation (3).

Entropy and joint entropy are the basic building blocks of mutual Information (I). Mutual information represents the amount of information known about one or a set of random variables, given the knowledge of other random variables. Mathematically, mutual information can be written as

$$I\Big(Y; X_1, \ldots, X_N\Big) = \widetilde{H}(Y) + \widetilde{H}(X_1, \ldots, X_N) - \widetilde{H}(X_1, \ldots, X_N, Y) \quad (5)$$

where $\widetilde{H}(Y)$ is computed by Equation (2), $\widetilde{H}(X_1, \ldots, X_N)$ and $\widetilde{H}(X_1, \ldots, X_N, Y)$ can be calculated by Equation (4).

Leverage Equations (1)–(5), we denoted the mutual information between LAI and $\{LAI_C, P, T, R\}$ as I(LAI; LAI_C, P, T, R), {SM, LAI_C, P, T, R} as I(LAI; SM, LAI_C,P, T, R), {VOD, LAI_C, P, T, R} as I(LAI; VOD, LAI_C, P, T, R), {SM, VOD, LAI_C, P, T, R} as I(LAI; SM, VOD, LAI_C,P, T, R). Therefore, the information gap between I(LAI; SM, LAI_C,P, T, R) and I(LAI; LAI_C, P, T, R) represents the additional information of SM for LAI prediction. The information gap between I(LAI; VOD, LAI_C,P, T, R) and I(LAI; LAI_C, P, T, R) represents the additional information of VOD for LAI prediction. Finally, the additive information of SM and VOD can be express as the information gap between I(LAI; SM, VOD, LAI_C, P, T, R) and I(LAI; LAI_C, P, T, R). These informational gaps represent the upped bound of additive information that SM, VOD, and SM with VOD can provide to LAI prediction. In practice, we often quantify the additive information of SM and VOD in a modeling scheme, though this approach is model dependent and may not produce the same additive information as shown by mutual information approach outlined above.

3.2. Random Forest Regression

Vegetation passes through known cycles on a seasonal basis and a significant portion of the variance of LAI can be explained by its day of year climatology (LAI_C). We define LAI_C on a certain day as the average smoothed LAI on that day between 2004 to 2019. As the LAI_C is static year-to-year, and therefore assumed known once it has been determined, new remote sensing observations are most useful if they aid in the prediction of deviations LAI from the known climatology. These deviations (hereafter termed LAI anomalies) may be caused, in-part, by unusual local environmental conditions such as droughts, cool snaps. We thus focus first on estimation of the LAI day-of-year anomaly. At each study site, we developed a "*null*" regression model that included P, T, R and LAI_C (base predictors). A "*full*" model was then built that included all base predictors as well as SM and VOD. The "*null*" and "*full*" models were established using the random forest (RF) regressor.

Random Forests (RF) is a supervised learning algorithm that uses ensemble learning method for regression tasks using multiple decision trees and a bagging statistical technique based on discrete or continuous datasets [55]. RF builds multiple decision trees and merges

the predictions from individual tree together result in more superior predictions compared with predictions solely rely on individual trees. Each individual tree learns from a random sample of training observations that are drawn with replacement. Despite higher variance in each individual tree on a particular set of training dataset, the idea of training each tree on different samples can lead to a lower variance and a lower bias for the entire forest [20].

A RF model contains multiple hyperparameters, and its application often involves complex hyperparameter tuning. We randomly partitioned the datasets used in the "*null*" and "*full*" models into training and testing datasets of 80% and 20% at each study site, respectively. We used training datasets to tune the hyperparameters that were shown to significantly impact RF performance [56] including the number of trees in the forest(100, 250, 500, 750, 1000), the number of features to consider when looking for the best split ('auto', 'sqrt', 'log2'), the minimum number of samples required to split an internal node (2, 30, 50), the minimum number of samples required to be at a leaf node (1, 10, 20, 30), and whether bootstrap samples are used when building trees ("True", "False"), where values in parentheses denote hyperparameter values we evaluated. We tuned the "*null*" and "*full*" models by specifying a finite range of the selected parameters and explore the potential parameter combinations iteratively using a grid search with a 5-fold cross validation.

The best "*null*" and "*full*" models were selected based on the model performance on the training dataset that was evaluated by coefficient of determination (R^2). R^2 and unbiased root mean square error (ubRMSE) were calculated from the testing dataset as

$$\mathrm{R2} = 1 - \frac{\sum_i (y_i - \hat{y})^2}{\sum_i (\overline{y} - \hat{y})^2} \quad (6)$$

and

$$\mathrm{ubRMSE} = \sqrt{E\left[((\hat{y} - E[\hat{y}]) - (y_i - E[\overline{y}]))^2\right]} \quad (7)$$

where y_i is the daily LAI anomaly, $\hat{y}$ is the estimated daily LAI anomaly by the RF models, $E[\hat{y}]$ and $E[\overline{y}]$ are the mean of model estimated LAI anomalies and daily LAI anomaly, respectively.

Predicted LAI anomalies from the "*full*" and "*null*" models were added back to the LAI climatology to get the final respective total LAI predictions. We then computed R^2 of the reconstructed LAI and the daily LAI using Equation (6) and replacing y_i with LAI and $\hat{y}$ with the respective "*full*" and "*null*" model reconstructed LAI. The ubRMSE of the reconstructed LAI and daily LAI were computed using Equation (7), replacing y_i with LAI, $\hat{y}$ and $E[\hat{y}]$ with the "full" and "null" model reconstructed LAI and their expectations, and $E[\overline{y}]$ with the mean of LAI. Note that the ubRMSE of the LAI anomaly prediction and LAI prediction are the same for both "*full*" and "*null*" models due to the ubRMSE formulation and the way we reconstruct total LAI as the sum of the climatology and the anomalies.

Feature importance is a score of input features based on how useful they are at predicting a target variable [55]. Feature importance was extracted from the "*full*" RF models to evaluate which variables impacted the LAI anomaly prediction. Negative R^2 values for "*full*" models indicate that the models are worse than simply using the mean value to predict LAI anomalies. Therefore, only study sites with worthwhile "*full*" model ($R^2 > 0$) were retained for feature importance analysis (115 grasslands, 44 croplands, 22 Savannas, 26 shrublands). The feature scores were ranked from least important (smaller values) to most important (larger values). A moving average with a window of 32 sites (~15% of the sample sites) of each feature importance was calculated and evaluated as a function of mean site SM.

4. Results

4.1. Input Dataset Characteristics

During the study period (April 2015 to October 2019), the mean site smoothed LAI in our study sites ranges from 0.17 to 2.42 and the standard deviation ranges from 0.02 to 1.46.

The ranges of skewness and kurtosis are −1.66 to 1.82 and −1.63 to 3.74, respectively. The mean site precipitation and standard deviation ranges are 0.14-mm/day to 3.79-mm/day and 1.01 mm/day to 21.95 mm/day, respectively. The precipitation skewness and kurtosis ranges are 2.99 to 25.45 and 10.72 to 675.21, respectively. The temperature mean ranges from 285.77 K to 298.2 K and the standard deviation ranges from 5.82 K to 10.59 K. The temperature skewness and kurtosis ranges from −0.95 to 0.06 and from −1.16 to 0.55, respectively. The mean site radiation ranges from 187.9 W/m^2 to 279.36 W/m^2 with the standard deviation ranges from 68.5 W/m^2 to 100.77 W/m^2. The radiation skewness and kurtosis ranges from −0.73 to 0.16 and from −1.23 to −0.14, respectively. Mean site SM during the study period ranges from 0.05 to 0.38 and the standard deviation ranges from 0.01 to 0.09. The SM skewness and kurtosis ranges are −0.38 to 2.41 and −1.13 to 8.38, respectively. The mean site VOD ranges from 0.05 to 0.51 and the standard deviation of VOD ranges from 0.03 to 0.11. The VOD skewness and kurtosis ranges are 0.06 to 1.9 and −0.79 to 12.3, respectively.

4.2. Mutual Information Analysis

Figure 2 shows how much information the base predictors contain about LAI as plotted against the mutual information of the base predictors combined with SMAP SM, SMAP VOD, or SM and VOD jointly. It is observable that the base predictors with SM and VOD individually, or with SM and VOD jointly contains more information about LAI than just the base predictor themselves. The base predictors with VOD contain slightly more information about LAI than the base predictors with SM. The averaged mutual information between LAI and base predictors is 0.24, while the mean mutual information between LAI and base predictors with SM is 0.29, and the mean mutual information between LAI and base predictors with VOD is 0.30 (Figure 2). The mean value of mutual information between LAI and base predictors with SM and VOD jointly is 0.32 (Figure 2), roughly a 33% increase in the information content.

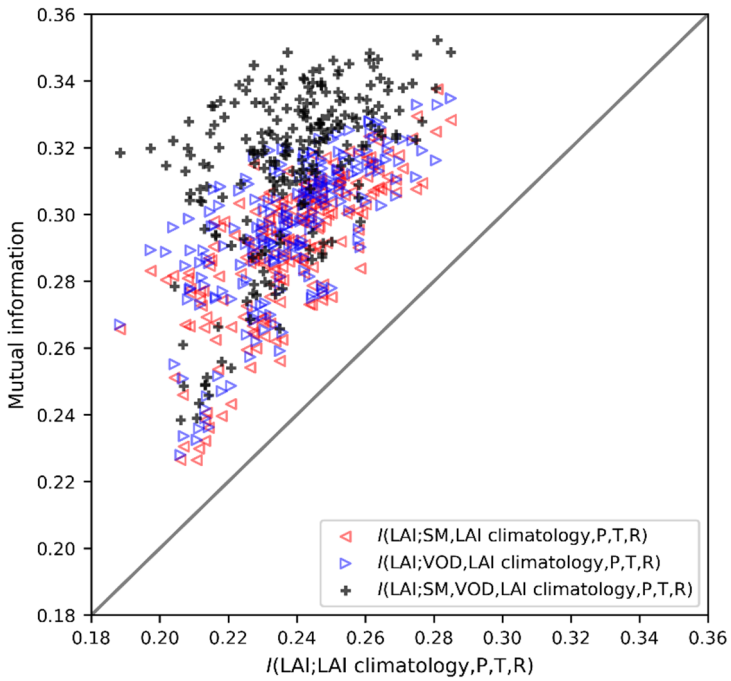


Figure 2. Mutual information between LAI and LAI climatology (LAI$_C$), precipitation (P), temperature (T) and radiation (R) (based predictors) against the mutual information between LAI and soil moisture (SM) and base predictors (red triangles), the mutual information between LAI and VOD and base predictors (blue triangles), and mutual information between LAI and base predictors and VOD and SM (black crosses).

4.3. LAI Anomaly Estimations

The performance of the random forest regressor on the prediction of LAI anomalies is shown in Figure 3. In general, the "*full*" models that were trained with the base predictors and SM and VOD jointly show a better prediction accuracy in higher R^2s and lower ubRMSE than the "*null*" models. Most of the estimates fall above the one-to-one line, with more than 80% of the "*full*" model outperform the "*null*" model in terms of their R^2s. The mean R^2s and ubRMSE of the "*null*" models are 0.15 (Table 1) and 0.137 (Table 2), respectively. The "*full*" models perform better in predicting LAI anomalies in grasslands and croplands, while compared with the model performance in other landcovers. However, the "*full*" model in shrublands on average has smaller ubRMSE (0.053), which possibly due to smaller LAI anomaly magnitude. The mean "*full*" model R^2s and ubRMSE in grasslands are 0.24 (Table 1) and 0.118 (Table 2), respectively. The "*null*" model results consistently slightly lower than the "*full*" model results across all landcovers (Table 1 and Figure 3). Mean improvements across sites for all landcover anomaly predictions were found to be greater than zero using a 1-sided *t*-test at a significance level of 0.05. While these R^2 values are not large, it is important to note that we are estimating anomalies, and much of the known deterministic behavior of these systems accounted for in the seasonal cycle.

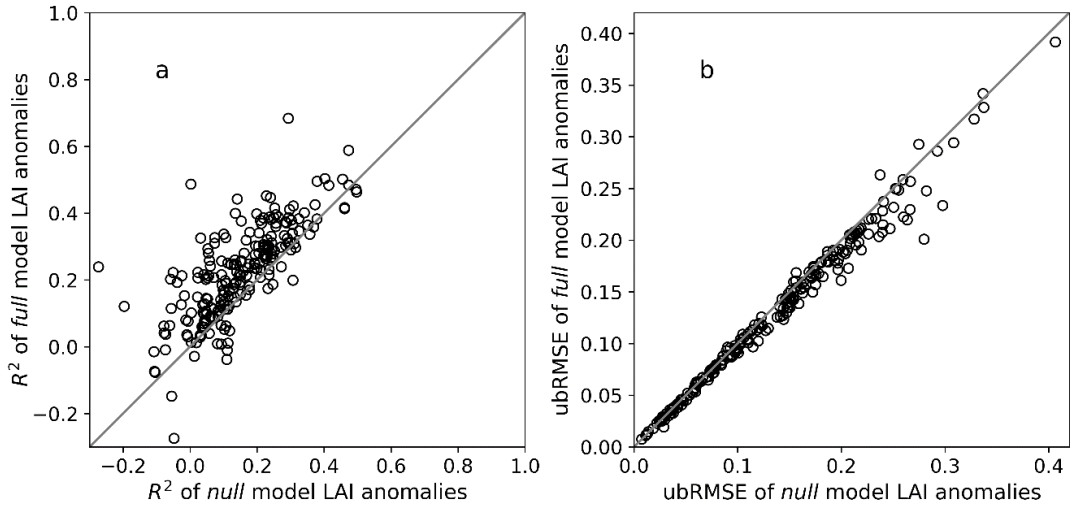


Figure 3. The R^2s of "*full*" model predicted LAI anomalies against the R^2s of the "*null*" model predicted LAI anomalies (**a**), and ubRMSE of "*full*" model predicted LAI anomalies against the of the "*null*" model predicted LAI anomalies (**b**).

Table 1. Statistics of "*null*" model R^2s and "*full*" model R^2s for LAI anomaly predictions and the statistics of the difference in R^2s between the "*full*" models and "*null*" models and for the LAI anomaly prediction. (asterisks indicate the R^2s of the models are significantly greater than 0 with different significant levels * 0.05; ** 0.01).

	***full* Model R^2s**	***null* Model R^2s**	***full* Model R^2s–*null* Model R^2s**	**Percentage Improvements**	**Number of Sites**
	Mean	**Mean**	**Mean**	**(% of *null*)**	***n***
Grasslands	0.24 **	0.16 **	0.08 **	50%	120
Shrublands	0.17 **	0.14 **	0.03 **	19%	27
Croplands	0.23 **	0.15 **	0.08 **	56%	47
Savannas	0.20 **	0.12 **	0.08 **	57%	22
All	0.22 **	0.15 **	0.07 **	50%	216

Table 2. Statistics of "*null*" model ubRMSE and "*full*" model ubRMSE for LAI anomaly predictions and the statistics of the difference in ubRMSE between the "*full*" models and "*null*" models and for the LAI anomaly predictions or LAI predictions (asterisks indicate the ubRMSE of the models are significantly greater than 0 with different significant levels * 0.05; ** 0.01. Note: This table is the same for LAI predictions due to how we reconstruct LAI using the "*full*" model and "*null*" model predicted LAI anomalies combined with static LAI climatology).

	***full* Model ubRMSE**	***null* Model ubRMSE**	***null* Model ubRMSE–*full* Model ubRMSE**	**Percentage Improvements**	**Number of Sites**
	Mean	**Mean**	**Mean**	**(% of *null*)**	***n***
Grasslands	0.118 **	0.126 **	0.008 **	5.7%	120
Shrublands	0.053 **	0.054 **	0.001 *	1.7%	27
Croplands	0.193 **	0.203 **	0.01 **	5.1%	47
Savannas	0.157 **	0.165 **	0.008 **	4.4%	22
All	0.130 **	0.137 **	0.007 **	5.2%	216

On average, the improvement of the "*full*" model over "*null*" model R^2s and ubRMSE are 0.07 (Table 1) and 0.007 (Table 2), respectively. These results show that SM and VOD are more informative in improving LAI anomaly prediction in grasslands, croplands, and savanna with the mean value of change of R^2 from "*null*" models to "*full*" models being 0.08. SMAP SM and VOD provide the least amount of additional information in shrublands for LAI anomaly prediction (Table 1). In terms of ubRMSE, shrublands has the smallest reduction (0.001) in ubRMSE from the "*null*" models to the "*full*" model (Table 2), while the largest reduction in ubRMSE is found in croplands (0.01). Though the reductions in ubRMSE from "*null*" models to the "*full*" models are small, the percentage improvements are non-trivial as indicated by statistical significance levels below 0.05 for all landcover classes. The grasslands and croplands have 5.7% and 5.1% reduction in ubRMSE relative to the mean ubRMSE of the "*null*" models, respectively.

4.4. LAI Estimations

Total modeled LAI was reconstructed by adding the LAI climatology to the "*null*" and "*full*" LAI anomaly models. In Figure 4a,b, the R^2s and ubRMSE of LAI climatology themselves (when compared with the true LAI) are plotted against and R^2s and ubRMSE of reconstructed total LAI. It is shown that majority of R^2s from reconstructed "*full*" model and "*null*" model LAI prediction outperform the prediction of LAI using only LAI climatology (Figures 4a and 5), while this pattern is reversed for ubRMSE (Figures 4b and 6). The reconstructed LAI by the "*full*" model often outperform the "*null*" reconstructed LAI in terms of both R^2s and ubRMSE (Figure 4a−b). On average, the "*full*" model R^2 and the "*null*" model R^2 are 0.83 and 0.81, respectively. The "*full*" model ubRMSE and the "*null*" model ubRMSE are 0.13 and 0.137, respectively. The R^2 and ubRMSE of LAI climatology are 0.75 and 0.15, respectively. As is summarized in Table 3, the mean R^2 of the "*full*" model is higher in grasslands (0.82), croplands (0.92), and savannas (0.91), while the "*full*" models are less effective for predicting LAI in shrublands (0.64). Overall, we note that the R^2 improvements (Table 3) are modest, but statically significant, after adding SM and VOD for the prediction of LAI. This is because the LAI climatology can explain a large amount of variance in LAI (Table 3).

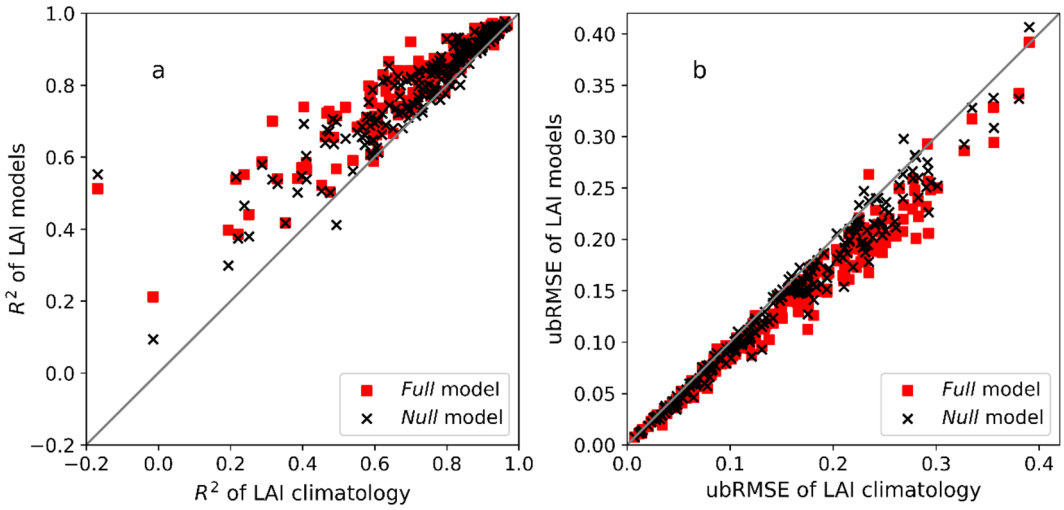


Figure 4. The R^2s of reconstructed LAI by adding "*full*" model predicted and "*null*" model predicted LAI anomalies to LAI climatology against the R^2s of LAI climatology (**a**), and the ubRMSE of reconstructed LAI by adding "*full*" model predicted and "*null*" model predicted LAI anomalies to LAI climatology against the ubRMSE of LAI climatology (**b**).

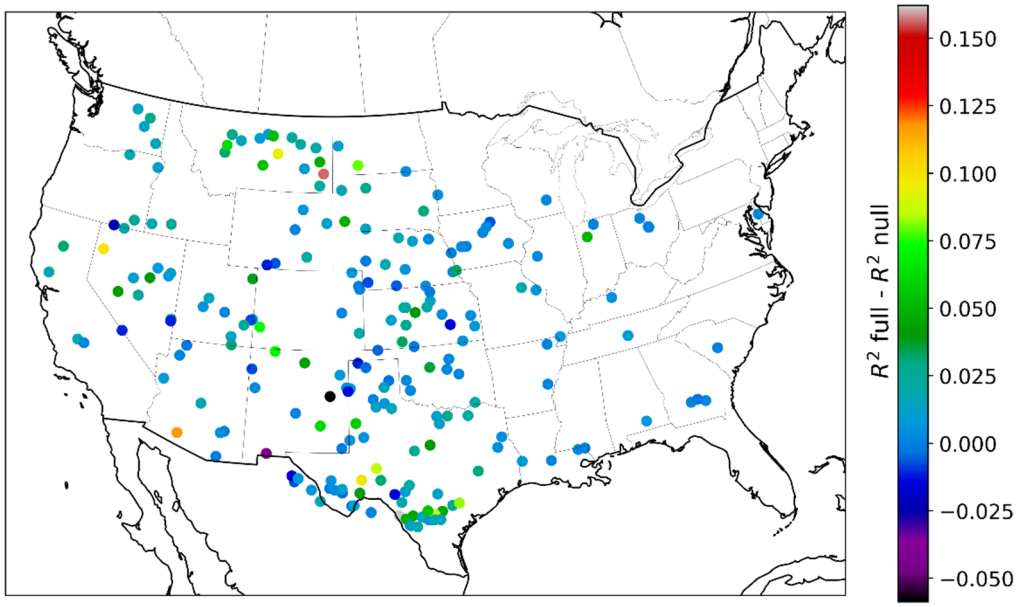


Figure 5. The spatial mapping of difference in R^2s between the "*full*" models and "*null*" models for the LAI prediction.

Table 3. Statistics of "*null*" model R^2s, "*full*" model R^2s, R^2s of LAI climatology for LAI predictions and the statistics of the difference in R^2s between the "*full*" model and "*null*" model LAI predictions. (asterisks indicate the R^2s of the models are significantly greater than 0 with different significant levels * 0.05; ** 0.01).

	full Model R^2s	*null* Model R^2s	Climatology R^2s	*full* Model R^2s–*null* Model R^2s	Percentage Improvements	Number of Sites
	Mean	Mean	Mean	Mean	(% of *null*)	*n*
Grasslands	0.82 **	0.80 **	0.73 **	0.02 **	2.7%	120
Shrublands	0.64 **	0.63 **	0.50 **	0.01 *	2.3%	27
Croplands	0.92 **	0.91 **	0.89 **	0.01 **	1.1%	47
Savannas	0.91 **	0.90 **	0.86 **	0.01 **	1.0%	22
All	0.83 **	0.81 **	0.75 **	0.02 **	2.1%	216

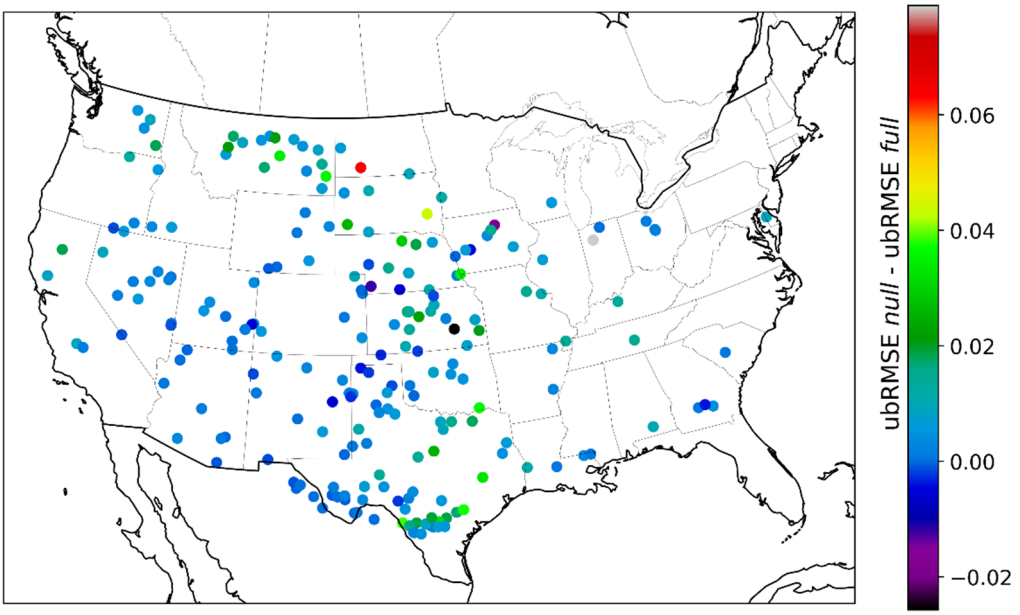


Figure 6. The spatial mapping of difference in ubRMSE between the "*null*" models and "*full*" models for the LAI prediction.

Overall, the improvements of the "*full*" model R^2s over "*null*" model in R^2 is 0.02 (Table 3), which is about 2.1% relative to the "*null*" model R^2s. In terms of percentage improvement relative to the "*null*" model, SM and VOD are more informative in improving LAI estimations in grasslands with the mean value of change of R^2 and ubRMSE being 5.7% and 2.7%, respectively. In croplands, shrublands and savannas, the SM and VOD provide less information for the LAI prediction with the mean value of change of R^2 from "*null*" models to "*full*" models being 0.01 (Table 3). In terms of ubRMSE, the SM and VOD provide the least amount of additional information for LAI prediction with the mean reduction in ubRMSE from the "*null*" models to the "*full* "models being 0.001 (1.7% of "*null*" model ubRMSE).

5. Discussion

5.1. Theoretical Additive Information of L-Band VOD and SM

The mutual information analysis in this study serves as the theoretical basis of exploring the potential usage of L-band SM and VOD to predict LAI given other base environmental variables. We consistently showed that the information shared between the base predictor and LAI can be improve after adding either L-band SM or VOD individually into the system. Incorporating L-band VOD, SM and base predictors in conjunction can provide more predictive skill. The improved skill indicates that microwave remotely sensed SM and VOD capture distinct phenological information that is not reflected in these base predictors. It is important to acknowledge that the mechanism of mutual information is purely data driven and is based on the marginal or joint probabilistic relationships between variables in this mutual information function. Therefore, the mutual information between LAI and these predictors reflects the theorical "*true*" relationships. However, this "*true*" relationship may change given the interference of other predictors.

It is important to note that any newly added predictors, such as SM or VOD, may be inter-correlated with some of the selected base predictors evaluated here. In fact, numerous previous studies have highlighted the spatial and temporal patterns of SM depend on the variability of precipitation forming a positive or negative SM-precipitation feedbacks that varies across different soil conditions [56–59]. Similarly, L-band VOD represents how microwave is attenuated through the canopy and has found to be tightly related to the vegetation canopy water storage [60]. Vegetation canopy water storage is not only a function of vegetation traits such as leaf water potential but also a function of external driving force such as radiation [61]. Therefore, radiation may contain information about

VOD. Given that SM and VOD not only share information about LAI, but also contains information about other base predictors, the theoretical explanatory power after adding SM or VOD into the system may not be as strong as expected since part of variability in LAI may be redundantly explained by the base predictors and L-band SM and VOD.

Overall, we found that the improvement in predictive skill after adding VOD and SM together shows a better performance than individually. This indicates that SM and VOD may share unique information with LAI and this uniquely shared information may reflect that the theoretical true vegetation phenology dynamics are driven by the mixture of the biotic (VOD) and abiotic (SM) factors.

5.2. Additive Information of L-Band SM and VOD for LAI Anomaly

We observed that there exist theoretical additional explanatory power of VOD and SM for LAI prediction from the mutual information analysis. However, in practice, it is extremely hard to approach the theoretical bound due to the algorithm uncertainties. In this study, we first evaluated how much additional information can SM and VOD provide for the prediction of LAI anomalies. Our results demonstrated that SMAP L-band VOD and SM have more skill for the prediction of LAI anomalies in grasslands. The stronger predictive power of SM and VOD in grasslands relative to shrublands could be physiological or could be due to the measurement accuracy of SMAP SM and VOD products. Previous SMAP SM validation studies have demonstrated that the SMAP SM estimates in grassland are generally more accurate than other landcovers. The poor performance of SMAP SM in shrubland has also been confirmed in other SMAP validation studies where the correlation between SMAP SM is significantly lower than those of other landcovers [62,63]. In addition, the nominal sensing depth of L-band SM is 5 cm, whereas other studies have demonstrated that the penetration depth of L-band was found to be much shallower (~1 cm) and can be sensitive to surface meteorological conditions [64,65]. Therefore, L-band SM may capture more information about plant-available water in grasslands where rooting depths are shallower than shrublands on average [64].

The SM and VOD sensed by SMAP is more representative of the average SM and VOD within a single pixel when landcover is homogeneous (as assumed with SMAP algorithms). However, the vegetation patterns in shrublands are tend to very patchy and inhomogeneous and shrubs contain more woody branches, which can cause the SMAP SM and VOD to be less representative, and hence may provide less predictive skill. From biophysical perspective, shrublands are more resilient to soil variations and disturbance than grasslands and croplands [66]. Changes in SM or vegetation water status (as reflected by VOD) in shrublands may be less influential on LAI anomalies, which may partially explain why SM and VOD produce possess less skill for the prediction of LAI anomalies.

The feature importance from "*full*" models was extracted and evaluated by their mean SM. As shown in Figure 7, it is not surprising that the LAI climatology is consistently the most important feature for the prediction of daily LAI anomalies. We also find vegetation tends to be sensitive to water availability in relatively dry SM conditions. Our study sites have relatively small LAI values and hence change in absolute LAI values may not synchronize with LAI climatology, which may cause a decrease in LAI climatology importance on LAI anomaly predictions in low SM regimes. Temperature was found to be the second most important factor for LAI anomaly prediction (Figure 7). High temperatures often lead to an increase of evapotranspiration therefore affect the leaf development. SM, the third highest rated factor, influences LAI anomalies via the amount of water that is available for plants. Previous study suggested that savannas and grasslands have the most resistant water uptake strategies because they are more effective at extracting water from the soils at drier soil conditions [28]. Our study sites are mainly dominated by grasslands and therefore might be able to explain why SM is relatively important for LAI anomalies in drier soil conditions.

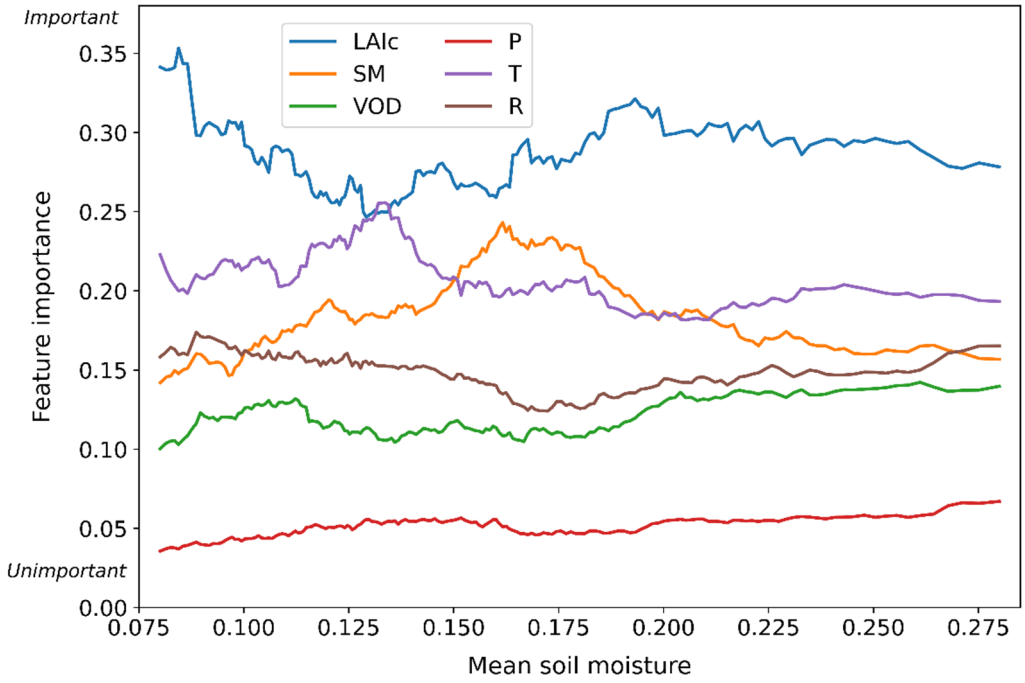


Figure 7. The level of feature importance of LAI climatology (LAI_C), soil moisture (SM), vegetation optical depth (VOD), precipitation (P), temperature (T) and radiation (R) from the "*full*" models against different soil moisture conditions.

In higher SM regimes, SM is no longer the key limiting factor since competition for water by vegetation is not as strong, this likely results a reduction in the importance of SM for driving LAI anomalies. Overall, VOD is found to be less important than the aforementioned factors (Figure 7). Information lost in the SMAP algorithms may contribute to this low performance [67], and much of the VOD information may be correlated with LAI climatology data. VOD is the representation of vegetation water status and it is possible that there exist some lag relationships between LAI anomalies and VOD that this study did not consider.

Precipitation was consistently found to be the least important factor that influences LAI anomaly predictions. Precipitation is an indirect water resource when compared with SM, which plants directly access. Under drier conditions, vegetation may be more sensitive to precipitation since the amount of moisture that can be provided by soil may not be sufficient for the plant to grow, while in wetter conditions plants mainly uptake the moisture from soil therefore mitigate the influence of precipitation on LAI anomalies. However, given the availability of both precipitation and SM, SM and VOD are much more relevant for LAI anomalies.

5.3. Additive Information of SM and VOD for LAI

Figure 8 shows the additive information of SM and VOD for the prediction of total LAI for different SM conditions. There is more opportunity for SMAP VOD and SM to be useful under intermediate (around ~0.16) SM regimes. Under drier SM conditions, plants often experience water stress, and therefore external water availability and internal water status, as indicated by SM and VOD, may be more relevant to plant water strategy. Hence, SM and VOD have more opportunity to capture unique information that is not expressed by traditional phenology predictors. A past study has found that VOD and water stress inferred from remotely sensed datasets is correlated with LAI. The correlation strength has shown to vary with different SM conditions and canopy characteristics [26]. Under wetter conditions, water availability is no more the key limiting factor since the vegetation leaf development is a collective effect of energy, water, and nutrients [68]. Therefore, the SM and VOD may provide less unique information for as vegetation is well-watered.

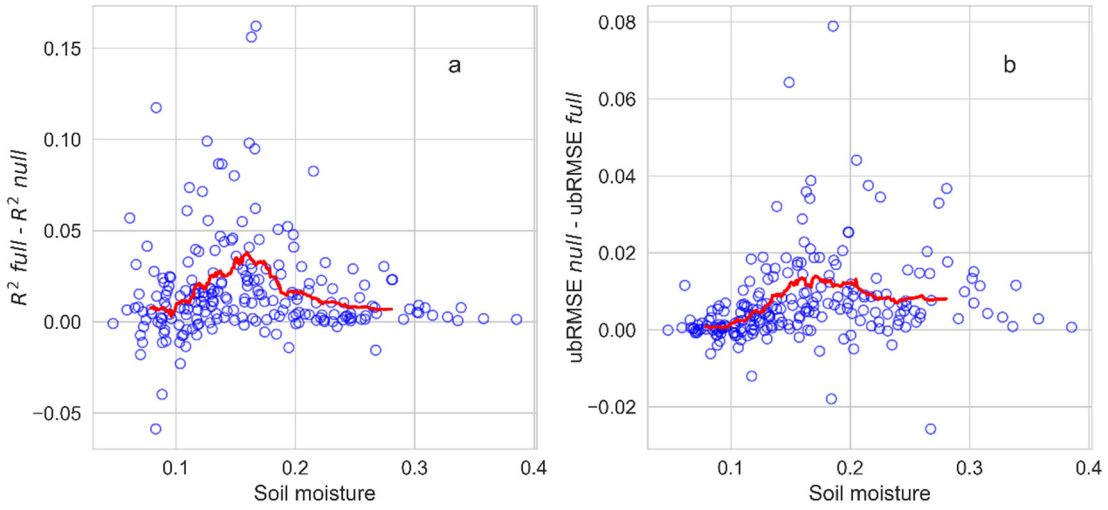


Figure 8. The difference in R^2s between "*full*" model and "*null*" model reconstructed LAI predictions as a function of soil moisture (**a**), and the reduction in ubRMSE from the "*null*" model to "*full*" model as a function of soil moisture (**b**). The red lines in (**a**,**b**) are the moving average using a window length of 15% of the data.

It is not surprising that LAI climatology explains most of the variability in LAI. In the locations where there are less inter-seasonal variabilities in vegetation dynamics/patterns such as shrublands, majority of the variation in LAI can be captured by its climatology. In these locations, there is less opportunity for SM and VOD to provided additional information. However, in locations where LAI climatology is not reflective of LAI dynamics, SM and VOD can be more informative, with the improvements up to 0.05 in R^2s, on average (Figure 9a). This is more than double the average value reported in Table 3. We found a strong correlation between the LAI variance that is not explained by the climatology and the improvement of the "*full*" model over the "*null*" model (Pearson correlation of 0.40, Figure 9a; Pearson correlation of 0.37, Figure 9b). This demonstrated that SM and VOD can be considered as much more useful potential predictors in locations that often experience inter-seasonal vegetation variability that cannot be fully represented by its daily LAI climatology. The additive information of SMAP-L2 SM and VOD exhibit a large variance (Figure 9a, b), which might be an indication of complexity of ecosystems.

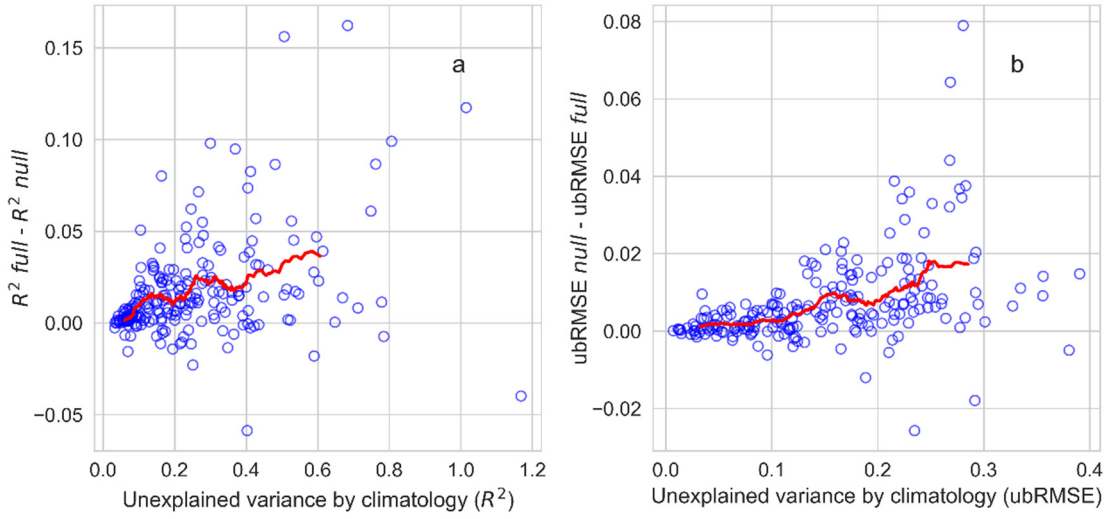


Figure 9. The difference in R^2s between "*full*" model and "*null*" model reconstructed LAI predictions as a function unexplained variance by LAI climatology (1-R^2 between LAI and its climatology) (**a**), and the reduction in ubRMSE from the "*null*" model to "*full*" model as a function of unexplained variance by LAI climatology (ubRMSE of LAI climatology). The red lines in (**a**,**b**) are the moving average using a window length of 15% of the data.

It is worth noting that there are numerous of studies that have focused on improving LAI prediction with various methods and data involved. Previous studies have shown that LAI can be better predicted with the help of additional remotely sensed datasets (improvements of ~0.05 in RMSE) [69,70]. LAI predictive accuracy can be improved with the combination of color and textures indices from unmanned aerial vehicle-based remote sensing using a random forest modeling approach (best R^2 values of 0.84 reported) [71]. Moreover, LAI estimations can be improved by leveraging a data assimilation technique in conjunction with terrestrial ecosystem models fed by satellite observations (R^2 values of 0.83) [72]. A study focusing on estimating LAI using Landsat datasets over tropical savanna and rangelands demonstrated that LAI can generally be estimated with R^2 values ranging 0.62–0.72 and 0.62–0.63 for from dry and wet seasons, respectively [73]. It was shown that LAI estimation accuracy can be improved by incorporating background, topography, and foliage clumping information (R^2 values of 0.42 & 0.69 when compared with in situ TRAC + LAI 2000 and TRAC measurements, respectively) [74]. While direct comparison is challenging due to differences in scale, extent, and other factors, the modest yet significant improvements from SMAP VOD and SM (Table 3) reported in this study are of similar magnitude as other studies, particularly in locations where interannual variability is large. In the end, a combination of many approaches is likely the optimum, however the global extent, increasing availability of microwave data, and results of this study suggest VOD and SM data should be considered moving forward.

5.4. Uncertainties, Limitations and Potential Applications

Our study demonstrated that the L-band SM and VOD can potentially provide additional information for improving the LAI predictions, but it is necessary to acknowledge that uncertainties and limitations still exist in this analysis. Firstly, the data uncertainty from SMAP SM and VOD can never be neglected. Although we filtered the SM and VOD using the intrinsic quality flags, there are still chances where these measurements can be noisy. We interpolated 4-day LAI to daily scale LAI. It is possible that the interpolated and smoothed LAI may not be reflective to the actual vegetation dynamics, which can cause the SM and VOD being less informative to predict LAI. There are many ways of interpolating and smoothing LAI that can lead to distinct LAI time series that may influence the results here.

The spatial scale of this analysis is relatively coarse (9 km by 9 km). Therefore, the results from this study may not be applicable to smaller scale studies or point scale studies. Our study sites mainly focused on low density landcovers since most of the heavily vegetated locations such as evergreen forests were excluded from this analysis due to the poor data quality from SMAP. We evaluated the additive information of SM and VOD with the interference of base predictors. It is important to note that LAI can be controlled by different environmental factors such as light, water, nutrients, temperature, and ambient carbon concentration collectively. Therefore, the additive information for SM and VOD may be less than what has shown in this study if more predictors are considered in the system.

This study can provide guidance for improving vegetation phenology predictions in locations where the vegetation phenology cannot be accurately captured by the daily climatology. In addition, the results can used as a reference for large scale LAI predictions and estimations. Other LAI prediction studies may consider coupling or fusing L-band SM and VOD in their modelling framework to get better accuracy of LAI predictions.

6. Conclusions

This study evaluated and quantified how informative L-band VOD and SM products are for the prediction of LAI using a machine learning approach. We first predicted the LAI daily anomaly by the random forest models. We showed that adding SMAP SM and VOD product can improve 0.07 in R^2s and reduce 0.007 (5.2% reduction of the model without SM and VOD) in ubRMSE for the LAI anomaly predictions. SMAP SM and VOD contain additive information that results in more skillful LAI anomaly predictions

in grasslands relative to shrublands. LAI climatology and temperature were overall the two most important variables for LAI anomalies prediction. SM is more important under drier conditions than wetter conditions. VOD is consistently the second least important factor. On average, R^2 of LAI prediction can be improved by 0.02 (2.1% improvement of the model without SM and VOD) after incorporating SMAP SM and VOD product the prediction of LAI can be improved to additional variance beyond what can be explained by the traditional drivers and LAI climatology.

Based on the results of our study, SM and VOD tend to be more useful for LAI prediction when LAI cannot be predicted well by its daily climatology. These locations, where inter-annual variability is high, are challenging to predict with the traditional drivers explored here. These large deviations from climatological expectations are also likely to be some of the most important periods to be predicted, as they are expected to correspond to unusual conditions with high societal relevance (e.g., droughts). Overall, the results of this study provide additional information for LAI prediction using L-band microwave derived SM and VOD products. It also provides insights about the relative importance of environmental drivers of daily LAI anomalies under different surface soil conditions.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/rs13071343/s1, Figure S1: Daily smoothed LAI and LAI climatology (a), daily LAI anomalies (b), soil moisture (c), vegetation optical depth (VOD) (c), precipitation (d), temperature (e), and radiation (f) time series at the representative grassland site (green star in Figure 1), Figure S2: Daily smoothed LAI and LAI climatology (a), daily LAI anomalies (b), soil moisture (c), vegetation optical depth (VOD) (c), precipitation (d), temperature (e), and radiation (f) time series at the representative savanna site (red start in Figure 1), Figure S3: Daily smoothed LAI and LAI climatology (a), daily LAI anomalies (b), soil moisture (c), vegetation optical depth (VOD) (c), precipitation (d), temperature (e), and radiation (f) time series at the representative cropland site (yellow star in Figure 1), Figure S4: Daily smoothed LAI and LAI climatology (a), daily LAI anomalies (b), soil moisture (c), vegetation optical depth (VOD) (c), precipitation (d), temperature (e), and radiation (f) time series at the representative shrubland site (blue star in Figure 1).

Author Contributions: Conceptualization, B.L. and S.P.G.; Methodology, B.L. and S.P.G.; Writing—original draft, B.L. and S.P.G.; Writing—review and editing, B.L., S.P.G. and D.R.U.; Funding Acquisition, S.P.G.; Supervision, S.P.G. All authors have read and agreed to the published version of the manuscript.

Funding: B.L. and S.P.G. acknowledge the support of National Aeronautics and Space Administration under grant NNX16AN13G.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: MCD15A3H, Version 6 and MCD12Q1, Version 6 dataset were acquired from https://modis.ornl.gov/globalsubset/ accessed on 24 February 2020. SMAP Enhanced Level-2 Radiometer Half-Orbit 9 km EASE-Grid Soil Moisture, Version 3 (SMAP-L2) and SMAP Level -4 Global 3-hourly 9 km EASE-Grid Surface and Root Zone Soil Moisture Geophysical Data, Version 4 datasets were obtained from National Snow & Ice Data Center (https://nsidc.org/data/smap/smap-data.html accessed on 2 October 2019).

Acknowledgments: The authors acknowledge Science Utilization of SMAP (SUSMAP) team for processing and providing the SMAP-L2 and SMAP-L4 datasets. The authors also would like to acknowledge the administrative support of Oregon State University department of Biological and Ecological Engineering and the Water Resources Graduate program.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Way, D.A.; Montgomery, R.A. Photoperiod constraints on tree phenology, performance and migration in a warming world. *Plant Cell Environ.* **2014**, *38*, 1725–1736. [CrossRef]
2. Gornall, J.; Betts, R.; Burke, E.; Clark, R.; Camp, J.; Willett, K.; Wiltshire, A. Implications of climate change for agricultural productivity in the early twenty-first century. *Philos. Trans. R. Soc. B Biol. Sci.* **2010**, *365*, 2973–2989. [CrossRef]
3. Piao, S.; Fang, J.; Zhou, L.; Ciais, P.; Zhu, B. Variations in satellite-derived phenology in China's temperate vegetation. *Glob. Chang. Biol.* **2006**, *12*, 672–685. [CrossRef]
4. Justice, C.; Townshend, J.R.G.; Holben, B.N.; Tucker, C.J. Analysis of the phenology of global vegetation using meteorological satellite data. *Int. J. Remote Sens.* **1985**, *6*, 1271–1318. [CrossRef]
5. Piao, S.; Liu, Q.; Chen, A.; Janssens, I.A.; Fu, Y.; Dai, J.; Liu, L.; Lian, X.; Shen, M.; Zhu, X. Plant phenology and global climate change: Current progresses and challenges. *Glob. Chang. Biol.* **2019**, *25*, 1922–1940. [CrossRef]
6. Liu, Q.; Piao, S.; Janssens, I.A.; Fu, Y.; Peng, S.; Lian, X.; Ciais, P.; Myneni, R.B.; Peñuelas, J.; Wang, T. Extension of the growing season increases vegetation exposure to frost. *Nat. Commun.* **2018**, *9*, 426. [CrossRef] [PubMed]
7. Gerard, F.F.; George, C.T.; Hayman, G.; Chavana-Bryant, C.; Weedon, G.P. Leaf phenology amplitude derived from MODIS NDVI and EVI: Maps of leaf phenology synchrony for Meso- and South America. *Geosci. Data J.* **2020**, *7*, 13–26. [CrossRef]
8. Delbart, N.; Kergoat, L.; Le Toan, T.; Lhermitte, J.; Picard, G. Determination of phenological dates in boreal regions using normalized difference water index. *Remote Sens. Environ.* **2005**, *97*, 26–38. [CrossRef]
9. Gao, L.; Wang, X.; Johnson, B.A.; Tian, Q.; Wang, Y.; Verrelst, J.; Mu, X.; Gu, X. Remote sensing algorithms for estimation of fractional vegetation cover using pure vegetation index values: A review. *ISPRS J. Photogramm. Remote Sens.* **2020**, *159*, 364–377. [CrossRef]
10. Ju, J.; Masek, J.G. The vegetation greenness trend in Canada and US Alaska from 1984–2012 Landsat data. *Remote Sens. Environ.* **2016**, *176*, 1–16. [CrossRef]
11. Xue, J.; Su, B. Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications. *J. Sens.* **2017**, *2017*, 1–17. [CrossRef]
12. Chen, M.; Willgoose, G.R.; Saco, P.M. Investigating the impact of leaf area index temporal variability on soil moisture predictions using remote sensing vegetation data. *J. Hydrol.* **2015**, *522*, 274–284. [CrossRef]
13. Martinez, B.; Cassiraga, E.; Camacho, F.; Garcia-Haro, J. Geostatistics for Mapping Leaf Area Index over a Cropland Landscape: Efficiency Sampling Assessment. *Remote Sens.* **2010**, *2*, 2584–2606. [CrossRef]
14. Sun, L.; Gao, F.; Anderson, M.C.; Kustas, W.P.; Alsina, M.M.; Sanchez, L.; Sams, B.; McKee, L.; Dulaney, W.; White, W.A.; et al. Daily Mapping of 30 m LAI and NDVI for Grape Yield Prediction in California Vineyards. *Remote Sens.* **2017**, *9*, 317. [CrossRef]
15. Li, Z.; Wang, J.; Tang, H.; Huang, C.; Yang, F.; Chen, B.; Wang, X.; Xin, X.; Ge, Y. Predicting Grassland Leaf Area Index in the Meadow Steppes of Northern China: A Comparative Study of Regression Approaches and Hybrid Geostatistical Methods. *Remote Sens.* **2016**, *8*, 632. [CrossRef]
16. Darvishzadeh, R.; Skidmore, A.; Schlerf, M.; Atzberger, C.; Corsi, F.; Cho, M. LAI and chlorophyll estimation for a heterogeneous grassland using hyperspectral measurements. *ISPRS J. Photogramm. Remote Sens.* **2008**, *63*, 409–426. [CrossRef]
17. Xie, Q.; Huang, W.; Liang, D.; Chen, P.; Wu, C.; Yang, G.; Zhang, J.; Huang, L.; Zhang, D. Leaf Area Index Estimation Using Vegetation Indices Derived from Airborne Hyperspectral Images in Winter Wheat. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 3586–3594. [CrossRef]
18. Houborg, R.; Boegh, E. Mapping leaf chlorophyll and leaf area index using inverse and forward canopy reflectance modeling and SPOT reflectance data. *Remote Sens. Environ.* **2008**, *112*, 186–202. [CrossRef]
19. Omer, G.; Mutanga, O.; Abdel-Rahman, E.M.; Adam, E. Empirical Prediction of Leaf Area Index (LAI) of Endangered Tree Species in Intact and Fragmented Indigenous Forests Ecosystems Using WorldView-2 Data and Two Robust Machine Learning Algorithms. *Remote Sens.* **2016**, *8*, 324. [CrossRef]
20. Houborg, R.; McCabe, M.F. A hybrid training approach for leaf area index estimation via Cubist and random forests machine-learning. *ISPRS J. Photogramm. Remote Sens.* **2018**, *135*, 173–188. [CrossRef]
21. Qu, Y.; Zhuang, Q. Modeling leaf area index in North America using a process-based terrestrial ecosystem model. *Ecosphere* **2018**, *9*, e02046. [CrossRef]
22. Wigneron, J.-P.; Li, X.; Frappart, F.; Fan, L.; Al-Yaari, A.; De Lannoy, G.; Liu, X.; Wang, M.; Le Masson, E.; Moisy, C. SMOS-IC data record of soil moisture and L-VOD: Historical development, applications and perspectives. *Remote Sens. Environ.* **2021**, *254*, 112238. [CrossRef]
23. Grant, J.; Wigneron, J.-P.; De Jeu, R.; Lawrence, H.; Mialon, A.; Richaume, P.; Al Bitar, A.; Drusch, M.; van Marle, M.; Kerr, Y. Comparison of SMOS and AMSR-E vegetation optical depth to four MODIS-based vegetation indices. *Remote Sens. Environ.* **2016**, *172*, 87–100. [CrossRef]
24. Tian, F.; Wigneron, J.-P.; Ciais, P.; Chave, J.; Ogée, J.; Peñuelas, J.; Ræbild, A.; Domec, J.-C.; Tong, X.; Brandt, M.; et al. Coupling of ecosystem-scale plant water storage and leaf phenology observed by satellite. *Nat. Ecol. Evol.* **2018**, *2*, 1428–1435. [CrossRef]
25. Lawrence, H.; Wigneron, J.-P.; Richaume, P.; Novello, N.; Grant, J.; Mialon, A.; Al Bitar, A.; Merlin, O.; Guyon, D.; Leroux, D.; et al. Comparison between SMOS Vegetation Optical Depth products and MODIS vegetation indices over crop zones of the USA. *Remote Sens. Environ.* **2014**, *140*, 396–406. [CrossRef]

26. Momen, M.; Wood, J.D.; Novick, K.A.; Pangle, R.; Pockman, W.T.; McDowell, N.G.; Konings, A.G. Interacting Effects of Leaf Water Potential and Biomass on Vegetation Optical Depth. *J. Geophys. Res. Biogeosci.* **2017**, *122*, 3031–3046. [CrossRef]
27. Wang, C.; Fu, B.; Zhang, L.; Xu, Z. Soil moisture–plant interactions: An ecohydrological review. *J. Soils Sediments* **2018**, *19*, 1–9. [CrossRef]
28. Bassiouni, M.; Good, S.P.; Still, C.J.; Higgins, C.W. Plant Water Uptake Thresholds Inferred from Satellite Soil Moisture. *Geophys. Res. Lett.* **2020**, *47*, 47. [CrossRef]
29. Boke-Olén, N.; Ardö, J.; Eklundh, L.; Holst, T.; Lehsten, V. Remotely sensed soil moisture to estimate savannah NDVI. *PLoS ONE* **2018**, *13*, e0200328. [CrossRef]
30. Tong, C.; Wang, H.; Magagi, R.; Goïta, K.; Zhu, L.; Yang, M.; Deng, J. Soil Moisture Retrievals by Combining Passive Microwave and Optical Data. *Remote Sens.* **2020**, *12*, 3173. [CrossRef]
31. Entekhabi, D.; Njoku, E.G.; O'Neill, P.E.; Kellogg, K.H.; Crow, W.T.; Edelstein, W.N.; Entin, J.K.; Goodman, S.D.; Jackson, T.J.; Johnson, J.; et al. The Soil Moisture Active Passive (SMAP) Mission. *Proc. IEEE* **2010**, *98*, 704–716. [CrossRef]
32. Colliander, A.; Cosh, M.H.; Misra, S.; Jackson, T.J.; Crow, W.T.; Chan, S.; Bindlish, R.; Chae, C.; Collins, C.H.; Yueh, S.H. Validation and scaling of soil moisture in a semi-arid environment: SMAP validation experiment 2015 (SMAPVEX15). *Remote Sens. Environ.* **2017**, *196*, 101–112. [CrossRef]
33. Chan, S.K.; Bindlish, R.; O'Neill, P.E.; Njoku, E.; Jackson, T.; Colliander, A.; Chen, F.; Burgin, M.; Dunbar, S.; Piepmeier, J.; et al. Assessment of the SMAP Passive Soil Moisture Product. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 4994–5007. [CrossRef]
34. Burgin, M.S.; Colliander, A.; Njoku, E.G.; Chan, S.K.; Cabot, F.; Kerr, Y.H.; Bindlish, R.; Jackson, T.J.; Entekhabi, D.; Yueh, S.H. A Comparative Study of the SMAP Passive Soil Moisture Product with Existing Satellite-Based Soil Moisture Products. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 2959–2971. [CrossRef] [PubMed]
35. Colliander, A.; Jackson, T.J.; Bindlish, R.; Chan, S.; Das, N.; Kim, S.B.; Cosh, M.H.; Dunbar, R.S.; Dang, L.; Pashaian, L.; et al. Validation of SMAP surface soil moisture products with core validation sites. *Remote Sens. Environ.* **2017**, *191*, 215–231. [CrossRef]
36. Tian, F.; Brandt, M.; Liu, Y.Y.; Verger, A.; Tagesson, T.; Diouf, A.A.; Rasmussen, K.; Mbow, C.; Wang, Y.; Fensholt, R. Remote sensing of vegetation dynamics in drylands: Evaluating vegetation optical depth (VOD) using AVHRR NDVI and in situ green biomass data over West African Sahel. *Remote Sens. Environ.* **2016**, *177*, 265–276. [CrossRef]
37. Feldman, A.F.; Gianotti, D.J.S.; Konings, A.G.; McColl, K.A.; Akbar, R.; Salvucci, G.D.; Entekhabi, D. Moisture pulse-reserve in the soil-plant continuum observed across biomes. *Nat. Plants* **2018**, *4*, 1026–1033. [CrossRef]
38. Global Subsets Tool: MODIS/VIIRS Land Products. Available online: https://modis.ornl.gov/globalsubset/ (accessed on 24 February 2020).
39. Vannan, S.K.S.; Cook, R.B.; Holladay, S.K.; Olsen, L.M.; Dadi, U.; Wilson, B.E. A Web-Based Subsetting Service for Regional Scale MODIS Land Products. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2009**, *2*, 319–328. [CrossRef]
40. Fang, H.; Wei, S.; Liang, S. Validation of MODIS and CYCLOPES LAI products using global field measurement data. *Remote Sens. Environ.* **2012**, *119*, 43–54. [CrossRef]
41. Yan, K.; Park, T.; Yan, G.; Liu, Z.; Yang, B.; Chen, C.; Nemani, R.R.; Knyazikhin, Y.; Myneni, R.B. Evaluation of MODIS LAI/FPAR Product Collection 6. Part 2: Validation and Intercomparison. *Remote Sens.* **2016**, *8*, 460. [CrossRef]
42. Kandasamy, S.; Baret, F.; Verger, A.; Neveux, P.; Weiss, M. A comparison of methods for smoothing and gap filling time series of remote sensing observations—Application to MODIS LAI products. *Biogeosciences* **2013**, *10*, 4055–4071. [CrossRef]
43. Sulla-Menashe, D.; Gray, J.M.; Abercrombie, S.P.; Friedl, M.A. Hierarchical mapping of annual global land cover 2001 to present: The MODIS Collection 6 Land Cover product. *Remote Sens. Environ.* **2019**, *222*, 183–194. [CrossRef]
44. Kwa, C. Local Ecologies and Global Science. *Soc. Stud. Sci.* **2005**, *35*, 923–950. [CrossRef]
45. Reichle, R.; De Lannoy, G.; Koster, R.D.; Crow, W.T.; Kimball, J.S.; Liu, Q. *SMAP L4 Global 3-Hourly 9 km EASE-Grid Surface and Root Zone Soil Moisture Geophysical Data, Version 4*; NASA National Snow and Ice Data Center Distributed Active Archive Center: Boulder, CO, USA, 2018. [CrossRef]
46. National Snow & Ice Data Center NASA Distributed Active Archive Center (DACC) at NSIDC. Available online: https://nsidc.org/data/smap/smap-data.html (accessed on 2 October 2019).
47. Scranton, K.; Amarasekare, P. Predicting phenological shifts in a changing climate. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 13212–13217. [CrossRef]
48. Bradley, A.V.; Gerard, F.F.; Barbier, N.; Weedon, G.P.; Anderson, L.O.; Huntingford, C.; Aragão, L.E.O.C.; Zelazowski, P.; Arai, E. Relationships between phenology, radiation and precipitation in the Amazon region. *Glob. Chang. Biol.* **2011**, *17*, 2245–2260. [CrossRef]
49. O'Neill, P.E.; Chan, S.; Njoku, E.G.; Jackson, T.; Bindlish, R. *SMAP Enhanced L2 Radiometer Global Daily 9 km EASE-Grid Soil Moisture, Version 3*; NASA National Snow and Ice Data Center Distributed Active Archive Center: Boulder, CO, USA, 2019. [CrossRef]
50. Chaubell, M.J.; Yueh, S.H.; Dunbar, R.S.; Colliander, A.; Chen, F.; Chan, S.K.; Entekhabi, D.; Bindlish, R.; O'Neill, P.E.; Asanuma, J.; et al. Improved SMAP Dual-Channel Algorithm for the Retrieval of Soil Moisture. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 3894–3905. [CrossRef]
51. Wigneron, J.-P.; Jackson, T.; O'Neill, P.; De Lannoy, G.; de Rosnay, P.; Walker, J.; Ferrazzoli, P.; Mironov, V.; Bircher, S.; Grant, J.; et al. Modelling the passive microwave signature from land surfaces: A review of recent results and application to the L-band SMOS & SMAP soil moisture retrieval algorithms. *Remote Sens. Environ.* **2017**, *192*, 238–262. [CrossRef]
52. Kraskov, A.; Stögbauer, H.; Grassberger, P. Estimating mutual information. *Remote Sens. Environ.* **2004**, *69*, 066138. [CrossRef]

53. Scott, D.W. Scott's rule. *Wiley Interdiscip. Rev. Comput. Stat.* **2010**, *2*, 497–502. [CrossRef]
54. Zhang, Z.; Grabchak, M. Bias Adjustment for a Nonparametric Entropy Estimator. *Entropy* **2013**, *15*, 1999–2011. [CrossRef]
55. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
56. Probst, P.; Wright, M.N.; Boulesteix, A. Hyperparameters and tuning strategies for random forest. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2019**, *9*, 9. [CrossRef]
57. Hohenegger, C.; Brockhaus, P.; Bretherton, C.S.; Schär, C. The Soil Moisture–Precipitation Feedback in Simulations with Explicit and Parameterized Convection. *J. Clim.* **2009**, *22*, 5003–5020. [CrossRef]
58. Asharaf, S.; Dobler, A.; Ahrens, B. Soil Moisture–Precipitation Feedback Processes in the Indian Summer Monsoon Season. *J. Hydrometeorol.* **2012**, *13*, 1461–1474. [CrossRef]
59. Hsu, H.; Lo, M.-H.; Guillod, B.P.; Miralles, D.G.; Kumar, S. Relation between precipitation location and antecedent/subsequent soil moisture spatial patterns. *J. Geophys. Res. Atmos.* **2017**, *122*, 6319–6328. [CrossRef]
60. Vittucci, C.; Laurin, G.V.; Tramontana, G.; Ferrazzoli, P.; Guerriero, L.; Papale, D. Vegetation optical depth at L-band and above ground biomass in the tropical range: Evaluating their relationships at continental and regional scales. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *77*, 151–161. [CrossRef]
61. Liu, R.; Wen, J.; Wang, X.; Wang, Z.; Li, Z.; Xie, Y.; Zhu, L.; Li, D. Derivation of Vegetation Optical Depth and Water Content in the Source Region of the Yellow River using the FY-3B Microwave Data. *Remote Sens.* **2019**, *11*, 1536. [CrossRef]
62. Zhang, R.; Kim, S.; Sharma, A. A comprehensive validation of the SMAP Enhanced Level-3 Soil Moisture product using ground measurements over varied climates and landscapes. *Remote Sens. Environ.* **2019**, *223*, 82–94. [CrossRef]
63. Chan, S.; Bindlish, R.; O'Neill, P.; Jackson, T.; Njoku, E.; Dunbar, S.; Chaubell, J.; Piepmeier, J.; Yueh, S.; Entekhabi, D.; et al. Development and assessment of the SMAP enhanced passive soil moisture product. *Remote Sens. Environ.* **2018**, *204*, 931–941. [CrossRef] [PubMed]
64. El Hajj, M.; Baghdadi, N.; Bazzi, H.; Zribi, M. Penetration Analysis of SAR Signals in the C and L Bands for Wheat, Maize, and Grasslands. *Remote Sens.* **2018**, *11*, 31. [CrossRef]
65. Escorihuela, M.; Chanzy, A.; Wigneron, J.; Kerr, Y. Effective soil moisture sampling depth of L-band radiometry: A case study. *Remote Sens. Environ.* **2010**, *114*, 995–1001. [CrossRef]
66. Stavi, I. Wildfires in Grasslands and Shrublands: A Review of Impacts on Vegetation, Soil, Hydrology, and Geomorphology. *Water* **2019**, *11*, 1042. [CrossRef]
67. Bonan, L.; Good, S.P. Information–based uncertainty decomposition in dual channel microwave remote sensing of soil moisture. *Hydrol. Earth Syst. Sci. Discuss.* **2020**, *534*, 1–16. [CrossRef]
68. Cowling, S.A.; Field, C.B. Environmental control of leaf area production: Implications for vegetation and land-surface modeling. *Glob. Biogeochem. Cycles* **2003**, *17*, 7-1. [CrossRef]
69. Roosjen, P.P.; Brede, B.; Suomalainen, J.M.; Bartholomeus, H.M.; Kooistra, L.; Clevers, J.G. Improved estimation of leaf area index and leaf chlorophyll content of a potato crop using multi-angle spectral data—Potential of unmanned aerial vehicle imagery. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *66*, 14–26. [CrossRef]
70. Korhonen, L.; Hadi; Packalen, P.; Rautiainen, M. Comparison of Sentinel-2 and Landsat 8 in the estimation of boreal forest canopy cover and leaf area index. *Remote Sens. Environ.* **2017**, *195*, 259–274. [CrossRef]
71. Li, S.; Yuan, F.; Ata-Ui-Karim, S.T.; Zheng, H.; Cheng, T.; Liu, X.; Tian, Y.; Zhu, Y.; Cao, W.; Cao, Q. Combining Color Indices and Textures of UAV-Based Digital Imagery for Rice LAI Estimation. *Remote Sens.* **2019**, *11*, 1763. [CrossRef]
72. Li, R.; Li, C.-J.; Dong, Y.-Y.; Liu, F.; Wang, J.-H.; Yang, X.-D.; Pan, Y.-C. Assimilation of Remote Sensing and Crop Model for LAI Estimation Based on Ensemble Kaiman Filter. *Agric. Sci. China* **2011**, *10*, 1595–1602. [CrossRef]
73. Dube, T.; Pandit, S.; Shoko, C.; Ramoelo, A.; Mazvimavi, D.; Dalu, T. Numerical Assessments of Leaf Area Index in Tropical Savanna Rangelands, South Africa Using Landsat 8 OLI Derived Metrics and In-Situ Measurements. *Remote Sens.* **2019**, *11*, 829. [CrossRef]
74. Gonsamo, A.; Chen, J.M. Improved LAI Algorithm Implementation to MODIS Data by Incorporating Background, Topography, and Foliage Clumping Information. *IEEE Trans. Geosci. Remote Sens.* **2013**, *52*, 1076–1088. [CrossRef]