AN ABSTRACT OF THE DISSERTATION OF

Bonan Li for the degree of Doctor of Philosophy in Water Resources Science presented on August 19, 2021.

Title: The Evaluation of the Additive Information Contained in New Ecohydrological Measurements

Abstract approved: _____

Stephen P. Good

The development of ecohydrological frameworks and theories under the ongoing global climate crisis depends on the development of new and advanced ecohydrological measurements. Currently, numerous of datasets have been collected at plot and ecosystem levels to understand the complex interactions of along the soil, plant, and atmosphere continuum. The development of new measurements costs considerable effort, time, and funding. Therefore, it is important to quantify if new measurements contain useful information for ecohydrological studies, and how the information encoded in the new measurements are transferred to understand and predict key environmental variables.

In this dissertation, we show that predictions in vegetation dynamics can be improved by adding observations from advanced satellite and in situ sensor systems. First, we evaluated new satellite soil moisture and vegetation optical depth measurements by Soil Moisture Passive Active (SMAP). We find that there is more opportunity for SMAP soil moisture and vegetation observations to be useful in locations where the daily vegetation climatology cannot adequately reflect observed vegetation dynamics. We also find that the uncertainties SMAP dual channel algorithm (DCA) are largely contributed by uncertainty of the algorithm's inputs (horizontal and vertical polarized brightness temperature), while the informational uncertainty of the SMAP DCA model itself is more related to the retrieval quality of soil moisture. The informational redundancy and synergy of the two brightness temperature measurements from SMAP are tightly related to the SMAP soil moisture retrieval quality. Finally, we apply a similar informational analysis to new

isotopic measurements at the National Ecological Observation Network (NEON). We find that majority of the information from the isotope measurements collected by NEON is unique, which cannot be obtained by other meteorological variables. Carbon isotope ($\delta^{13}C$) provides more additional information about *LH* in arid locations, while the water isotope ($\delta^{2}H$) provides more additional information about *LH* at locations with higher aridity, lower mean annual temperature, and lower mean site elevation. These studies show that informational analysis is useful to evaluated how additional information is encoded in new ecohydrological measurements.

The Evaluation of the Additive Information Contained in New Ecohydrological Measurements

by
Bonan Li

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented August 19, 2021
Commencement June 2022

Doctor of Philosophy dissertation of Bonan Li presented on August 19, 2021

APPROVED:

Major Professor, representing Water Resources Science

Director of the Water Resources Graduates Program

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

Bonan Li, Author

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the support of National Aeronautics and Space Administration and National Ecological Observatory Network.

I would like to express my sincerest appreciation of my major advisor Dr. Stephen Good for his instruction and dedication during my Ph.D. I also appreciate and would like to thank all my colleagues in the Water Resources Graduate Program for providing me guidance from professional and non-professional aspects. I acknowledge all the co-authors who have been helpful and patient on writing up the dissertation chapters.

I would also like to acknowledge my family for being my strongest support over the course of my Ph.D. I am grateful for all the people who I met over the last four years. Your encouragement makes my Ph.D a wonderful adventure.

CONTRIBUTION OF AUTHORS

Dawn R. URycki was involved in manuscript review of chapter 2. Catherine Finkenbiner, Gabriel Bowen, David Noone, Christopher Still, and William Anderegg were involved in the manuscript review of chapter 4. Richard Fiorella provided instructions on data acquisition and data calibration in chapter 4.

Stephen P. Good was involved in the conceptualizations, methodology, data analysis, and manuscript review from chapter 2 to chapter 4.

TABLE OF CONTENTS

Page

TABLE OF CONTENTS (Continued)

Page

TABLE OF CONTENTS (Continued)

Page

LIST OF FIGURES

Figure Page

LIST OF FIGURES (Continued)

LIST OF FIGURES (Continued)

Figure Page

LIST OF TABLES

Table Page

Chapter 1. General Introduction

Ecohydrological studies have been providing guidance and methodological approaches direct revealing at the underlying mechanisms of the ecosystem's function in response to climate, social, and water resources variations (Vose et al., 2011). Understanding plant-water interactions are central of the ecohydrological investigations (Asbjornsen et al., 2011). The availability, quality, and spatial distribution of water resources such as soil moisture and precipitation affect the plant development and shape ecosystem structures (Pugnaire et al., 2019). Conversely, the variation in plant community composition with diversified plant traits influence the variations in water flux exchange (Fischer et al., 2019).

The availability of long-term datasets from plot scale to landscape level under contrasting climate regimes are often pivotal in ecohydrological investigations to effectively address and understand the governing mechanisms in the soil-plant-water continuum (Vose et al., 2011). Traditional ecohydrological investigations heavily rely on using manually collected field samples of water and vegetation. For instance, the study of plant-water interactions in response to various of climatic conditions requires the vegetation canopy and water flux exchange information that can be represented leaf area index (LAI) (Kumar et al., 2019), which is often measured by direct destructive sampling that involves the in removal of several plants, the separation and measurement of the individual leaf areas (Fang et al., 2019). Ground measurement of evapotranspiration (ET) often rely on the installation lysimeters in the field plots (Gebler et al., 2015), which requires significant amount of human, materials, and financial resources. These datasets and data collection methods are effective for the study of local ecohydrological process but cannot be used to address macroscale issues due to the spatial and temporal heterogeneity of soil properties, plant community compositions, and weather conditions at the ecosystem level (Wang & Dickinson, 2012). Scaling local scale datasets to study landscape scale problems, which remain one of the under addressed questions in ecohydrological investigations (Asbjornsen et al., 2011) requires new observations of these complex systems.

Resolving large scale ecohydrology processes requires long-term and high frequency datasets that can adequately reflect the holistic dynamics of the ecosystem (Krause et al., 2015).

Recently, considerable efforts have been made to measure and estimate the key ecohydrological variables at ecosystem level. The long heritage of soil moisture observation from space has enabled researchers to resolve various of ecohydrological problems in the past (Mohanty et al., 2017). The newly launched Soil Moisture Active Passive (SMAP) can provide soil moisture and vegetation observations with higher temporal and spatial resolution (Entekhabi et al., 2010a), which may hold useful information about ecosystem functions. However, little attention has been given to how much new information can the SMAP observations provided for the understanding of vegetation dynamics beyond what can be informed from other metalogical variables that are known to affect the vegetation. In addition, the soil moisture observations from have been validated against in situ soil moisture measurements at a few core and sparse validation sites (Chen et al., 2017; Colliander et al., 2017), yet little is known about why and how information transferred from the raw SMAP algorithm inputs to the end-user product. Knowing where and when does the SMAP algorithm tend to be more uncertain will provide insights to algorithm designers to refine the algorithms thereby better algorithm performance.

Therefore, the chapter 2 of this dissertation we evaluated how much information SMAP soil moisture and vegetation observations can provide in the prediction of LAI at daily scales. Mutual information theory (Cover & Thomas, 2005) was used to broadly determine the additional information that captures both linear and non-linear information transferred from SMAP soil moisture and vegetation observations to the prediction of LAI. A multivariate non-linear empirical model was created to quantify the additional information of SMAP observations and the conditions under which SMAP observations are more informative.

In chapter 3, we assessed how information flows through satellite algorithms from SMAP Dual Channel Algorithm (DCA) input data is translated to end user products. The informational random uncertainty and informational model uncertainty were quantified by leveraging the framework built by (Gong et al., 2013). The relationships between these informational uncertainties and the retrieval quality of the SMAP algorithm were analyzed. Finally, the multivariate information from SMAP brightness temperature observations (the main measurements from SMAP) to SMAP soil moisture retrieval were decomposition using partial

information decomposition. The relationships between these decomposed components were assessed against the SMAP soil moisture retrieval quality.

Another important component in ecohydrological studies are the carbon and water fluxes. Carbon and water fluxes, which circulate between the terrestrial ecosystems and the atmosphere, are equally important to as they are jointly influence the earth's surface hydrology, ecology, and climate (Gentine et al., 2019). The advent of the eddy covariance (EC) technique has been providing valuable water and carbon fluxes observations at the ecosystem level that have been leveraged to address a wide array of ecohydrological issues (Baldocchi, 2020; Yaseef et al., 2009). The National Ecological Observatory Network (NEON), for the first time, provides a standardized measurements long record of the stable isotope ratios of water vapor and carbon dioxide measurements of the atmosphere that can be used to understand the ecosystem's water and carbon cycles (Fiorella et al., 2021). Mounting and calibrating such isotope equipment on the EC towers requires considerable effort, yet little is known how much additional constraints on carbon and water fluxes these isotopes can provide beyond traditional meteorological observations (e.g., vapor pressure deficit, air temperature, net radiation, windspeed) that are less costly and easier to be obtained. Therefore, it is crucial to know if these measurements contain information about the water and carbon fluxes of the ecosystem. In fact, past literature has highlighted that stable isotope holds the potential to investigate carbon and water cycles at different locations (Bowen et al., 2019; Esposito et al., 2019; Li et al., 2019).

Therefore, chapter 4 of this dissertation focused on answering three simple questions: (1) do isotope observations from NEON contain useful information about the bulk water and carbon fluxes across North America, (2) can any information provided about the bulk fluxes by isotope measurements be gleaned from other meteorological variables, and (3) where the does the isotope contain more information about these bulk fluxes. We approach the answers to these questions by leveraging information theory and multivariate mutual information decomposition technique.

Finally, the general findings of this dissertation is summarized in chapter 5 of this dissertation.

Chapter 2. 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

Status: Published in MDPI *Remote Sensing*,

Volume 13, Issue 7, 2021

2. 1 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.

2.2 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 (Way & Montgomery, 2015). Recent climate warming has impacted vegetation phenology, which may consequently have profound implications for agriculture and forest productivity (Gornall et al., 2010). 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 (Piao et al., 2006). 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 (Justice et al., 1985). 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 (Piao et al., 2019). 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 (Liu et al., 2018). The Enhanced Vegetation Index (EVI) has been applied to map global vegetation phenology and investigate the peak growing season of dense evergreen forests (Gerard et al., 2020). Normalized Difference Water Index (NDWI) has the potential to determine the phenological dates over boreal regions (Delbart et al., 2005). 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 (Gao et al., 2020), and are not precise indicators of vegetation leaf area (Ju & Masek, 2016). In addition, the retrieval accuracy of satellite vegetation indices is often affected by satellite viewing geometry, atmospheric condition, and background soil properties (Xue & Su, 2017).

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 (Chen et al., 2015). A large set of physically and empirically based algorithms coupled with remotely sensed datasets have been proposed to estimate LAI at large scales (Li et al., 2016; Martinez et al., 2010; Sun et al., 2017). 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 (Darvishzadeh et al., 2008; Xie et al., 2014). 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 (Houborg & Boegh, 2008; Houborg & McCabe, 2018; Omer et al., 2016). Process-based model was also used to estimate LAI by optimizing the parameters that are related to LAI in a terrestrial ecosystem model (Qu & Zhuang, 2018).

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 (Wigneron et al., 2021). Soil Moisture and Ocean Salinity (SMOS) L-band VOD was found to be highly correlated with LAI, with correlation strength varying with landcover type (Grant et al., 2016). 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 (Tian et al., 2018). A comparative study has also shown that SMOS L-band VOD contains information about LAI that varies for different crop zones in the USA (Lawrence et al., 2014). A conceptual framework showed that VOD sensitivity to leaf water potential and LAI varies with climate conditions (Momen et al., 2017). 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 (Wang et al., 2019). 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 (Bassiouni et al., 2020). A global study on vegetation dynamics has shown that the

synthesized microwave SM dataset can model and estimate vegetation phenology in drylands (Boke-Olén et al., 2018). In addition, microwave retrieved SM often requires the description of the overlaying canopy water status, which is related to the vegetation phenology (Tong et al., 2020). 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 (Entekhabi et al., 2010b). SMAP SM products have been extensively validated against in situ SM in a sequence of core and spare validation campaigns (S. K. Chan et al., 2016; Andreas Colliander et al., 2017). 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 (Burgin et al., 2017; Colliander et al., 2017). Global scale inter-comparisons among SMAP and other satellite SM suggested that SMAP is more robust and exhibits smaller ubRMSE (Tian et al., 2016).

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 (Bassiouni et al., 2020; Feldman et al., 2018). 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 (Tian et al., 2016). 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 (Feldman et al., 2018), 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.

2.3 Materials

2.3.1 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 2.1). 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.

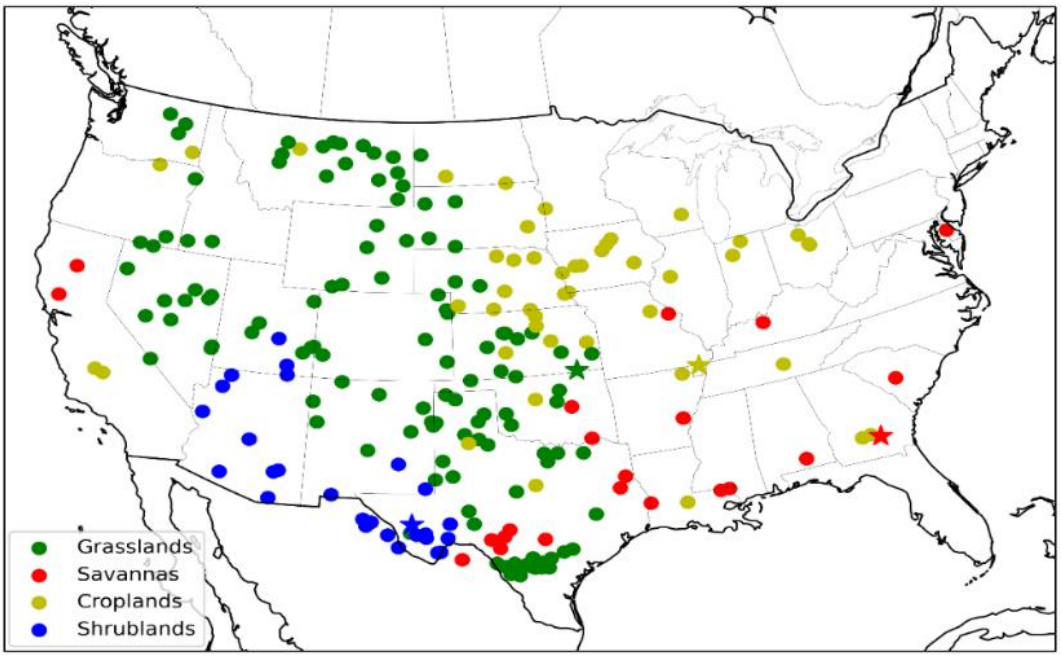

Figure 2.1 Spatial distribution of selected terrestrial study sites and their correspondent dominant landcovers (stared locations are representative sites of each landcover as shown in Figures 2. S1–2. S4).

2.3.2 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 (ORNL DAAC, 2018; Santhana Vannan et al., 2009), 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 ($\geqslant$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 (Fang et al., 2012; Yan et al., 2016). 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 (Kandasamy et al., 2013); 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 2.4.2) based on the smoothed LAI. See Figures 2.S1–2.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 2.1).

2.3.3 Landcover Information

The yearly MODIS Landcover dataset (MCD12Q1) was obtained with the same user-specified spatial extent as LAI (see Section 2.3.2). The original spatial resolution of MCD12Q1 is 500 m and contains five types of landcover classification schemes (Sulla-Menashe et al., 2019). 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 (Kwa, 2005). 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.

2.3.4 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 (Reichle et al., 2018) that have been shown to be related plant phenology: temperature, radiation, and precipitation (Bradley et al., 2011; Scranton & Amarasekare, 2017). 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 (Reichle et al., 2018). 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 2.1) are shown in Figures 2.S1–2.S4.

2.3.5 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) (O'Neill et al., 2019). 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 (Entekhabi et al., 2010b). 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 (Chaubell et al., 2020). 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 (Wigneron et al., 2017) and (Chaubell et al., 2020).

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 2.S1–2.S4 for plotted time series of SM and VOD at a representative site for each landcover (the stared locations in Figure 2.1).

2.4 Methodology

2.4.1. Shannon's Entropy and Mutual Information

Shannon's entropy represents the amount of information that is needed to fully describe a random variable (Kraskov et al., 2004). 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 (D. W. Scott, 2010). We then applied the Miller-Madow correction and a normalization method (Z. Zhang & Grabchak, 2013) 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 $H(\mathrm{X})$ 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 $\mathrm{X_i}$ or a set of random variables $\{\mathrm{X_1}, \ldots, \mathrm{X_N}\}$, the joint entropy is

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

where $p(x_1,\ldots, x_N)$ is the joint probability mass function of a set of random variable $\{\mathrm{X_1},\ldots, \mathrm{X_N}\}$ that is estimated using the Scott's rule data binning method (Scott, 2010). 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

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

where $H(\mathrm{X_1}, \ldots, \mathrm{X_N})$ is the original joint entropy from Equation (3), n is the number of data points of $\{\mathrm{X_1}, \ldots, \mathrm{X_N}\}$, 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

$$I(Y; X_1, \ldots, X_N) = \widetilde{H}(Y) + \widetilde{H}(X_1, \ldots, X_N) - \widetilde{H}(X_1, \ldots, X_N, Y), \quad (3)$$

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.

2.4.2 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 (Breiman, 2001). 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 (Houborg & McCabe, 2018).

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 (Probst et al., 2019) 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

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y})^2}{\sum_i (\bar{y} - y_i)^2}, \quad (6)$$

and

$$\text{ubRMSE} = \sqrt{E[((\hat{y} - E[\hat{y}]) - (y_i - E[\bar{y}]))^2]}, \quad (4)$$

where y_i is the daily LAI anomaly, $\hat{y}$ is the estimated daily LAI anomaly by the RF models, $E[\hat{y}]$ and $E[\bar{y}]$ 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 $E[\bar{y}]$ 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 (Breiman, 2001). 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.

2.5 Results

2.5.1. 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.

2.5.2. Mutual Information Analysis

Figure 2.2 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 2.2). The mean value of mutual information between LAI and base predictors with SM and VOD jointly is 0.32 (Figure 2.2), roughly a 33% increase in the information content.

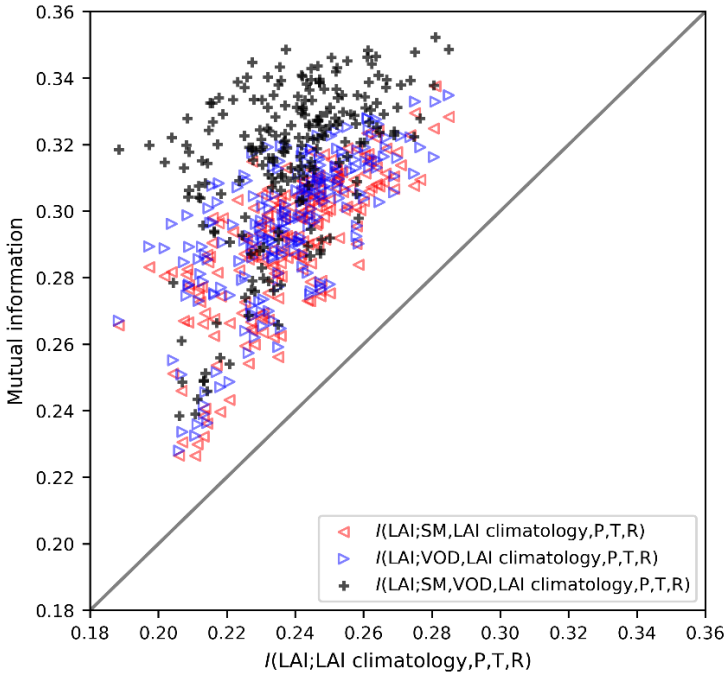

Figure 2.2 Mutual information between LAI and LAI climatology (LAIC), precipitation (P), temperature (T) and radiation (R) (based predictors) against the mutual information between LAI and soil moisture (SM) and base predictors (red triangles), the mutual information between LAI and VOD and based predictors (blue triangles), and mutual information between LAI and base predictors and VOD and SM (black crosses).

2.5.3 LAI Anomaly Estimations

The performance of the random forest regressor on the prediction of LAI anomalies is shown in Figure 2.3. 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 2.1) and 0.137 (Table 2.2), 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 2.1) and 0.118 (Table 2.2), respectively. The "null" model results consistently slightly lower than the "full" model results

across all landcovers (Table 2.1 and Figure 2.3). 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.

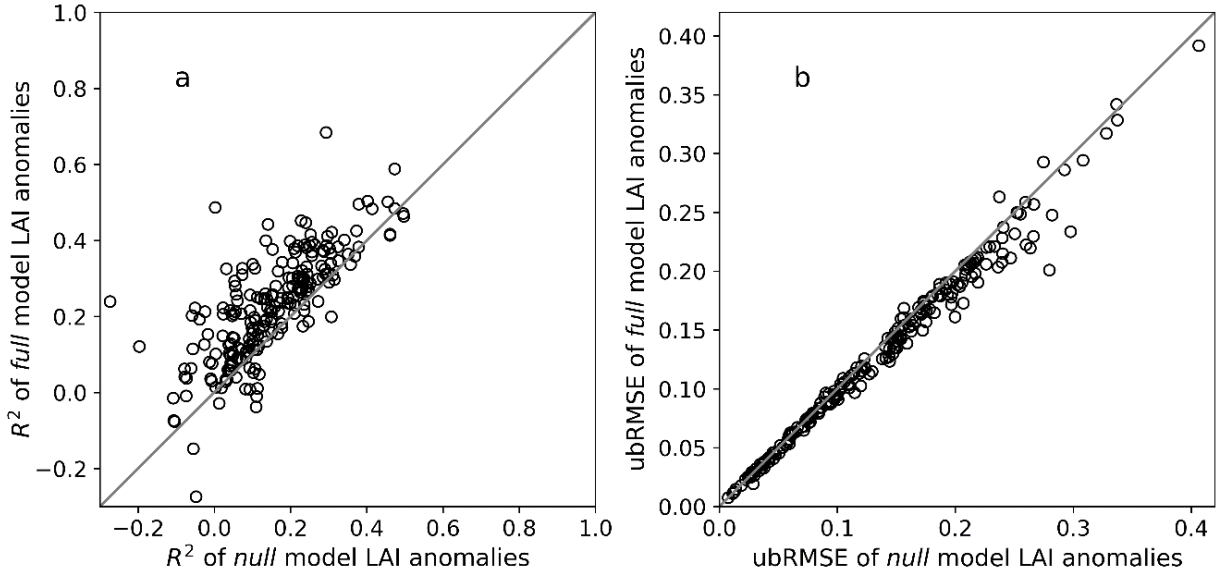

Figure 2.3 The R^2s of "full" model predicted LAI anomalies against the R^2s of the "null" model predicted LAI anomalies (a), and ubRMSE of "full" model predicted LAI anomalies against the of the "null" model predicted LAI anomalies (b).

Table 2.1 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 2.2 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 2.1) and 0.007 (Table 2.2), 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 shrub-lands for LAI anomaly prediction (Table 2.1). In terms of ubRMSE, shrublands has the smallest reduction (0.001) in ubRMSE from the "null" models to the "full" model (Table 2.2), 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.

2.5.4. LAI Estimations

Total modeled LAI was reconstructed by adding the LAI climatology to the "null" and "full" LAI anomaly models. In Figure 2.4 a, b, the R^2s and ubRMSE of LAI climatology themselves (when compared with the true LAI) are plotted against and R2s 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 2.4a and 2.5), while this pattern is reversed for ubRMSE (Figures 2.4b and 2.6). The reconstructed LAI by the "full" model often outperform the "null" reconstructed LAI in terms of both R^2s and ubRMSE (Figure 2.4a−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 2.3) 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).

Table 2.3 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

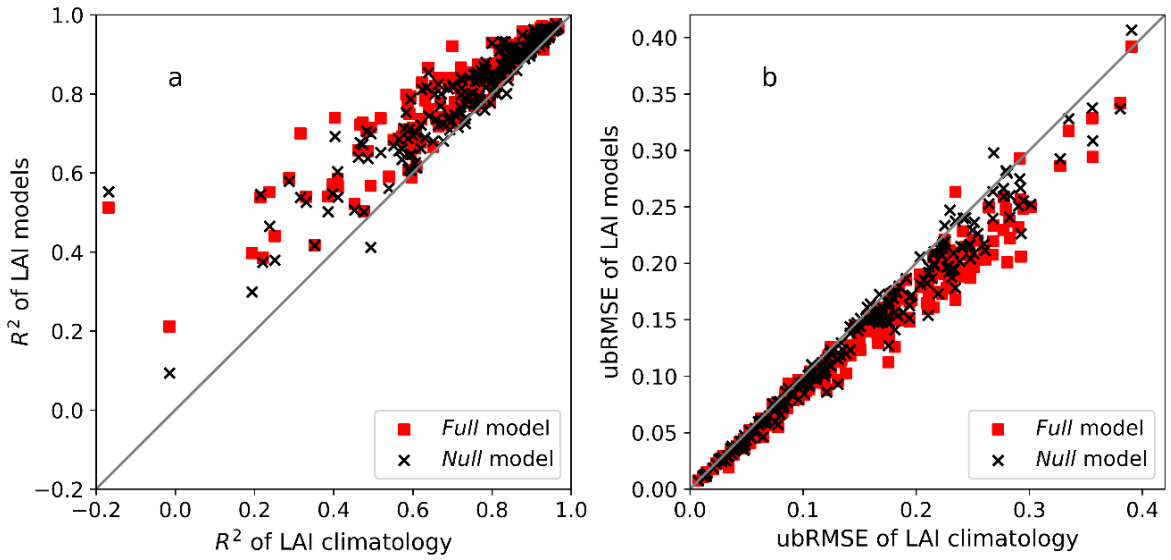

Figure 2.4 The R^2s of reconstructed LAI by adding "full" model predicted and "null" model predicted LAI anomalies to LAI climatology against the R^2s of LAI climatology (a), and the ubRMSE of reconstructed LAI by adding "full" model predicted and "null" model predicted LAI anomalies to LAI climatology against the ubRMSE of LAI climatology (b).

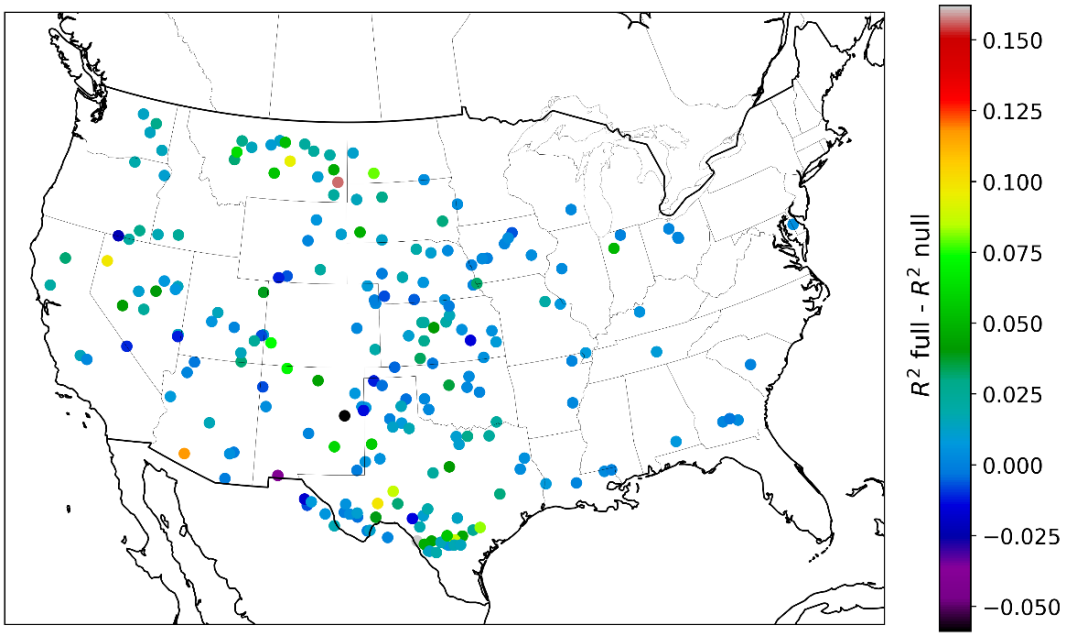

Figure 2.5 The spatial mapping of difference in R^2s between the "full" models and "null" models for the LAI prediction.

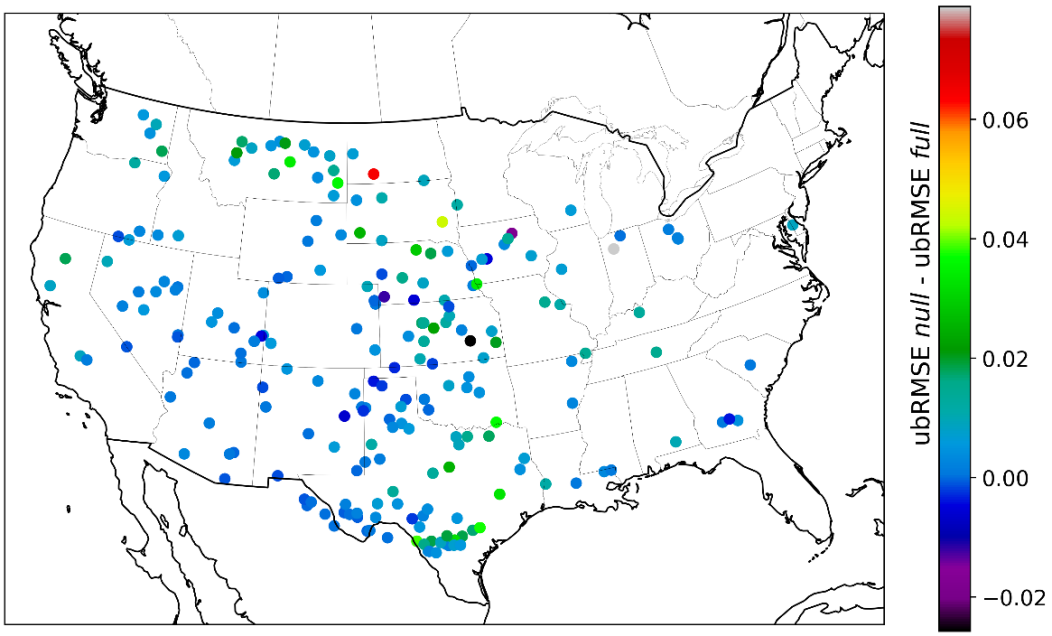

Figure 2.6 The spatial mapping of difference in ubRMSE between the "null" models and "full" models for the LAI prediction.

Overall, the improvements of the "full" model R^2s over "null" model in R^2 is 0.02 (Table 2.3), which is about 2.1% relative to the "null" model R^2s. In terms of percentage im-provident 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 2.3). In terms of ubRMSE, the SM and VOD pro-vide 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).

2.6 Discussion

2.6.1 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 in-formation 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 (Asharaf et al., 2012; Hohenegger et al., 2009; Hsu et al., 2017). 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 (Vittucci et al., 2019). 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 (Liu et al., 2019). 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.

2.6.2 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 physio-logical 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 (Liu et al., 2019; Zhang et al., 2019). 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 (El Hajj et al., 2018; Escorihuela et al., 2010). Therefore, L-band SM may capture more information about plant-available water in grasslands where rooting depths are shallower than shrublands on average (El Hajj et al., 2018).

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 in-homogeneous 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 bio-physical perspective, shrublands are more resilient to soil variations and disturbance than grasslands and croplands (Stavi, 2019). Changes in SM or vegetation water status (as reflected by VOD) in shrublands may be less influential on LAI anomalies, which may partially ex-plain 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 2.7, 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 2.7). 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 (Bassiouni et al., 2020). 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.

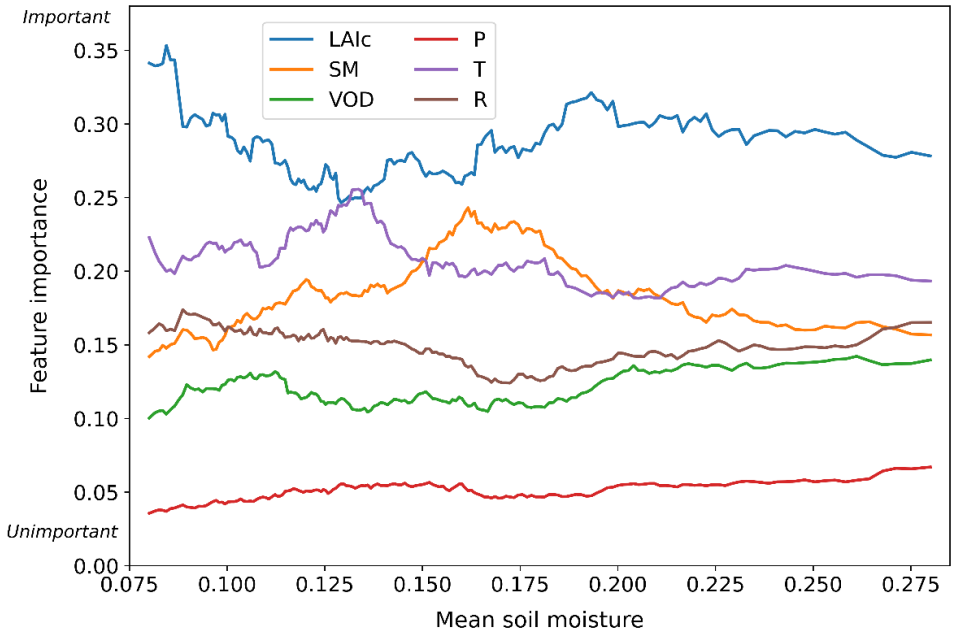

Figure 2.7 The level of feature importance of LAI climatology (LAI_C), soil moisture (SM), vegetation optical depth (VOD), precipitation (P), temperature (T) and radiation (R) from the "full" models against different soil moisture conditions.

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 2.7). Information lost in the SMAP algorithms may contribute to this low performance (Li & Good, 2020), 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.

2.6.3 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 (Momen et al., 2017). 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 (Cowling & Field, 2003). Therefore, the SM and VOD may provide less unique information for as vegetation is well-watered.

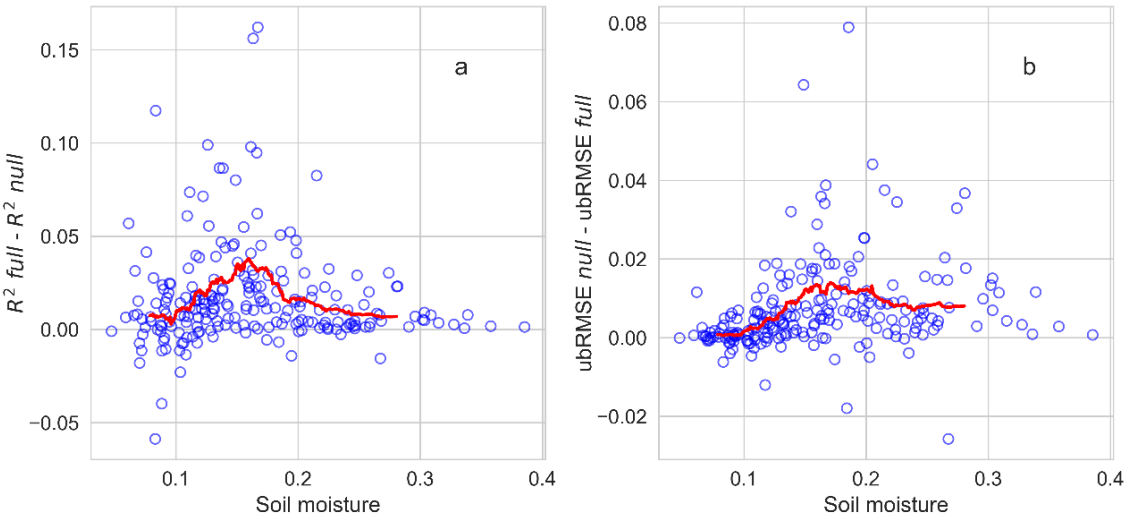

Figure 2.8 The difference in R^2s between "full" model and "null" model reconstructed LAI predictions as a function of soil moisture (a), and the reduction in ubRMSE from the "null" model to "full" model as a function of soil moisture (b). The red lines in (a,b) are the moving average using a window 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 2.9a). 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 2.9b). 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 2.9a, b), which might be an indication of complexity of ecosystems.

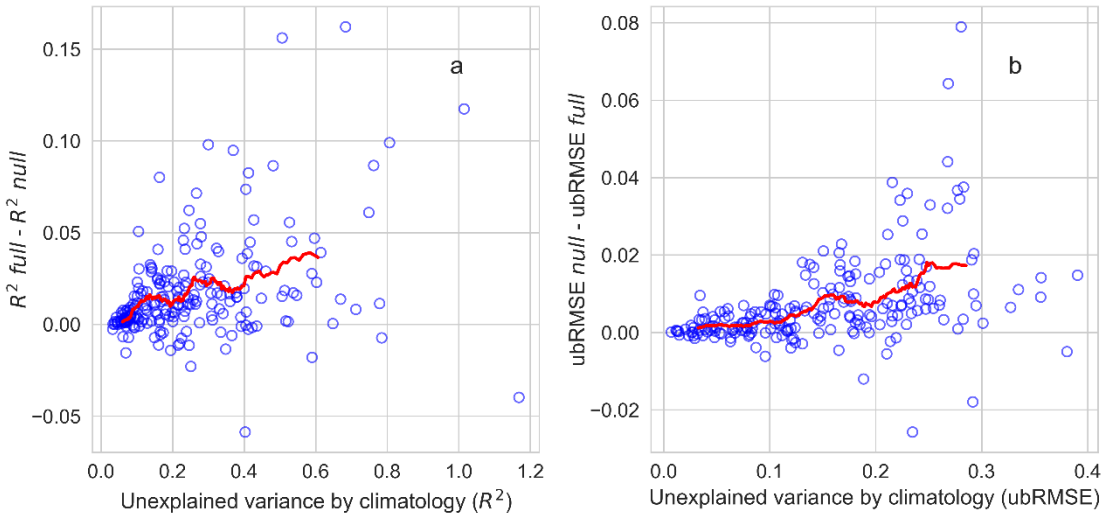

Figure 2.9 The difference in R^2s between "full" model and "null" model reconstructed LAI predictions as a function unexplained variance by LAI climatology (1-R^2 between LAI and its climatology) (a), and the reduction in ubRMSE from the "null" model to "full" model as a function of unexplained variance 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) (Korhonen et al., 2017; Roosjen et al., 2018). 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) (Li et al., 2019). Moreover, LAI estimations can be improved by leveraging a data assimilation technique in conjunction with terrestrial ecosystem models fed by satellite observations (R^2 value of 0.83) (Li et al., 2011). 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 (Dube et al., 2019). 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) (Gonsamo & Chen, 2014). 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 2.3) 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.

2.6.4 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.

2.7 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 environ-mental drivers of daily LAI anomalies under different surface soil conditions.

2.8 References

Asharaf, S., Dobler, A., & Ahrens, B. (2012). Soil Moisture–Precipitation Feedback Processes in the Indian Summer Monsoon Season. *Journal of Hydrometeorology*, *13*(5), 1461–1474. https://doi.org/10.1175/JHM-D-12-06.1

Bassiouni, M., Good, S. P., Still, C. J., & Higgins, C. W. (2020). Plant Water Uptake Thresholds Inferred From Satellite Soil Moisture. *Geophysical Research Letters*, *47*(7). https://doi.org/10.1029/2020GL087077

Boke-Olén, N., Ardö, J., Eklundh, L., Holst, T., & Lehsten, V. (2018). Remotely sensed soil moisture to estimate savannah NDVI. *PLOS ONE*, *13*(7), e0200328. https://doi.org/10.1371/journal.pone.0200328

Bonan Li. (2020). Information–based uncertainty decomposition in dual channel microwave remote sensing of soil moisture. *Hydrology and Earth System Sciences Discussion*. https://doi.org/https://doi.org/10.5194/hess-2020-534

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. (2011). Relationships between phenology, radiation and precipitation in the Amazon region. *Global Change Biology*, *17*(6), 2245–2260. https://doi.org/10.1111/j.1365-2486.2011.02405.x

Breiman, L. (2001). Rndom Forests. *Macchine Learning*, *45*, 5–32. https://doi.org/https://doi.org/10.1023/A:1010933404324

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. (2017). A Comparative Study of the SMAP Passive Soil Moisture Product With Existing Satellite-Based Soil Moisture Products. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(5), 2959–2971. https://doi.org/10.1109/TGRS.2017.2656859

Chan, S. K., Bindlish, R., O'Neill, P. E., Njoku, E., Jackson, T., Colliander, A., Chen, F., Burgin, M., Dunbar, S., Piepmeier, J., Yueh, S., Entekhabi, D., Cosh, M. H., Caldwell, T., Walker, J., Wu, X., Berg, A., Rowlandson, T., Pacheco, A., … Kerr, Y. (2016). Assessment of the SMAP Passive Soil Moisture Product. *IEEE Transactions on Geoscience and Remote Sensing*, *54*(8), 4994–5007. https://doi.org/10.1109/TGRS.2016.2561938

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., Berg, A. A., Bosch, D. D., Caldwell, T., Cosh, M. H., Holifield Collins, C., Martinez-Fernandez, J., Seyfried, M., Starks, P. J., Su, Z., … Walker, J. (2020). Improved SMAP Dual-Channel Algorithm for the Retrieval of Soil Moisture. *IEEE Transactions on Geoscience and Remote Sensing*, *58*(6), 3894–3905. https://doi.org/10.1109/TGRS.2019.2959239

Chen, M., Willgoose, G. R., & Saco, P. M. (2015). Investigating the impact of leaf area index temporal variability on soil moisture predictions using remote sensing vegetation data. *Journal of Hydrology*, *522*, 274–284. https://doi.org/10.1016/j.jhydrol.2014.12.027

Colliander, A., Jackson, T. J., Bindlish, R., Chan, S., Das, N., Kim, S. B., Cosh, M. H., Dunbar, R. S., Dang, L., Pashaian, L., Asanuma, J., Aida, K., Berg, A., Rowlandson, T., Bosch, D., Caldwell, T., Caylor, K., Goodrich, D., al Jassar, H., … Yueh, S. (2017). Validation of

SMAP surface soil moisture products with core validation sites. *Remote Sensing of Environment*, *191*, 215–231. https://doi.org/10.1016/j.rse.2017.01.021

Colliander, Andreas, Cosh, M. H., Misra, S., Jackson, T. J., Crow, W. T., Chan, S., Bindlish, R., Chae, C., Holifield Collins, C., & Yueh, S. H. (2017). Validation and scaling of soil moisture in a semi-arid environment: SMAP validation experiment 2015 (SMAPVEX15). *Remote Sensing of Environment*, *196*, 101–112. https://doi.org/10.1016/j.rse.2017.04.022

Cowling, S. A., & Field, C. B. (2003). Environmental control of leaf area production: Implications for vegetation and land-surface modeling. *Global Biogeochemical Cycles*, *17*(1), 7-1-7–14. https://doi.org/10.1029/2002GB001915

Darvishzadeh, R., Skidmore, A., Schlerf, M., Atzberger, C., Corsi, F., & Cho, M. (2008). LAI and chlorophyll estimation for a heterogeneous grassland using hyperspectral measurements. *ISPRS Journal of Photogrammetry and Remote Sensing*, *63*(4), 409–426. https://doi.org/10.1016/j.isprsjprs.2008.01.001

Delbart, N., Kergoat, L., Le Toan, T., Lhermitte, J., & Picard, G. (2005). Determination of phenological dates in boreal regions using normalized difference water index. *Remote Sensing of Environment*, *97*(1), 26–38. https://doi.org/10.1016/j.rse.2005.03.011

Dube, T., Pandit, S., Shoko, C., Ramoelo, A., Mazvimavi, D., & Dalu, T. (2019). Numerical Assessments of Leaf Area Index in Tropical Savanna Rangelands, South Africa Using Landsat 8 OLI Derived Metrics and In-Situ Measurements. *Remote Sensing*, *11*(7), 829. https://doi.org/10.3390/rs11070829

El Hajj, M., Baghdadi, N., Bazzi, H., & Zribi, M. (2018). Penetration Analysis of SAR Signals in the C and L Bands for Wheat, Maize, and Grasslands. *Remote Sensing*, *11*(1), 31. https://doi.org/10.3390/rs11010031

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., Kimball, J., Piepmeier, J. R., Koster, R. D., Martin, N., McDonald, K. C., Moghaddam, M., Moran, S., Reichle, R., Shi, J. C., … Van Zyl, J. (2010). The Soil Moisture Active Passive (SMAP) Mission. *Proceedings of the IEEE*, *98*(5), 704–716. https://doi.org/10.1109/JPROC.2010.2043918

Escorihuela, M. J., Chanzy, A., Wigneron, J. P., & Kerr, Y. H. (2010). Effective soil moisture sampling depth of L-band radiometry: A case study. *Remote Sensing of Environment*, *114*(5), 995–1001. https://doi.org/10.1016/j.rse.2009.12.011

Fang, H., Wei, S., & Liang, S. (2012). Validation of MODIS and CYCLOPES LAI products using global field measurement data. *Remote Sensing of Environment*, *119*, 43–54. https://doi.org/10.1016/j.rse.2011.12.006

Feldman, A. F., Short Gianotti, D. J., Konings, A. G., McColl, K. A., Akbar, R., Salvucci, G. D., & Entekhabi, D. (2018). Moisture pulse-reserve in the soil-plant continuum observed across biomes. *Nature Plants*, *4*(12), 1026–1033. https://doi.org/10.1038/s41477-018-0304-9

Gao, L., Wang, X., Johnson, B. A., Tian, Q., Wang, Y., Verrelst, J., Mu, X., & Gu, X. (2020). Remote sensing algorithms for estimation of fractional vegetation cover using pure vegetation index values: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, *159*, 364–377. https://doi.org/10.1016/j.isprsjprs.2019.11.018

Gerard, F. F., George, C. T., Hayman, G., Chavana-Bryant, C., & Weedon, G. P. (2020). Leaf phenology amplitude derived from MODIS NDVI and EVI: Maps of leaf phenology

synchrony for Meso- and South America. *Geoscience Data Journal*, *7*(1), 13–26. https://doi.org/10.1002/gdj3.87

Gonsamo, A., & Chen, J. M. (2014). Improved LAI Algorithm Implementation to MODIS Data by Incorporating Background, Topography, and Foliage Clumping Information. *IEEE Transactions on Geoscience and Remote Sensing*, *52*(2), 1076–1088. https://doi.org/10.1109/TGRS.2013.2247405

Gornall, J., Betts, R., Burke, E., Clark, R., Camp, J., Willett, K., & Wiltshire, A. (2010). Implications of climate change for agricultural productivity in the early twenty-first century. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1554), 2973–2989. https://doi.org/10.1098/rstb.2010.0158

Grant, J. P., Wigneron, J.-P., De Jeu, R. A. M., Lawrence, H., Mialon, A., Richaume, P., Al Bitar, A., Drusch, M., van Marle, M. J. E., & Kerr, Y. (2016). Comparison of SMOS and AMSR-E vegetation optical depth to four MODIS-based vegetation indices. *Remote Sensing of Environment*, *172*, 87–100. https://doi.org/10.1016/j.rse.2015.10.021

Hohenegger, C., Brockhaus, P., Bretherton, C. S., & Schär, C. (2009). The Soil Moisture–Precipitation Feedback in Simulations with Explicit and Parameterized Convection. *Journal of Climate*, *22*(19), 5003–5020. https://doi.org/10.1175/2009JCLI2604.1

HOUBORG, R., & BOEGH, E. (2008). Mapping leaf chlorophyll and leaf area index using inverse and forward canopy reflectance modeling and SPOT reflectance data. *Remote Sensing of Environment*, *112*(1), 186–202. https://doi.org/10.1016/j.rse.2007.04.012

Houborg, R., & McCabe, M. F. (2018). A hybrid training approach for leaf area index estimation via Cubist and random forests machine-learning. *ISPRS Journal of Photogrammetry and Remote Sensing*, *135*, 173–188. https://doi.org/10.1016/j.isprsjprs.2017.10.004

Hsu, H., Lo, M.-H., Guillod, B. P., Miralles, D. G., & Kumar, S. (2017). Relation between precipitation location and antecedent/subsequent soil moisture spatial patterns. *Journal of Geophysical Research: Atmospheres*, *122*(12), 6319–6328. https://doi.org/10.1002/2016JD026042

Ju, J., & Masek, J. G. (2016). The vegetation greenness trend in Canada and US Alaska from 1984–2012 Landsat data. *Remote Sensing of Environment*, *176*, 1–16. https://doi.org/10.1016/j.rse.2016.01.001

JUSTICE, C. O., TOWNSHEND, J. R. G., HOLBEN, B. N., & TUCKER, C. J. (1985). Analysis of the phenology of global vegetation using meteorological satellite data. *International Journal of Remote Sensing*, *6*(8), 1271–1318. https://doi.org/10.1080/01431168508948281

Kandasamy, S., Baret, F., Verger, A., Neveux, P., & Weiss, M. (2013). A comparison of methods for smoothing and gap filling time series of remote sensing observations – application to MODIS LAI products. *Biogeosciences*, *10*(6), 4055–4071. https://doi.org/10.5194/bg-10-4055-2013

Korhonen, L., Hadi, Packalen, P., & Rautiainen, M. (2017). Comparison of Sentinel-2 and Landsat 8 in the estimation of boreal forest canopy cover and leaf area index. *Remote Sensing of Environment*, *195*, 259–274. https://doi.org/10.1016/j.rse.2017.03.021

Kraskov, A., Stögbauer, H., & Grassberger, P. (2004). Estimating mutual information. *Physical Review E*, *69*(6), 066138. https://doi.org/10.1103/PhysRevE.69.066138

Kwa, C. (2005). Local Ecologies and Global Science. *Social Studies of Science*, *35*(6), 923–950. https://doi.org/10.1177/0306312705052100

Lawrence, H., Wigneron, J.-P., Richaume, P., Novello, N., Grant, J., Mialon, A., Al Bitar, A., Merlin, O., Guyon, D., Leroux, D., Bircher, S., & Kerr, Y. (2014). Comparison between SMOS Vegetation Optical Depth products and MODIS vegetation indices over crop zones of the USA. *Remote Sensing of Environment*, *140*, 396–406. https://doi.org/10.1016/j.rse.2013.07.021

LI, R., LI, C., DONG, Y., LIU, F., WANG, J., YANG, X., & PAN, Y. (2011). Assimilation of Remote Sensing and Crop Model for LAI Estimation Based on Ensemble Kaiman Filter. *Agricultural Sciences in China*, *10*(10), 1595–1602. https://doi.org/10.1016/S1671-2927(11)60156-9

Li, S., Yuan, F., Ata-UI-Karim, S. T., Zheng, H., Cheng, T., Liu, X., Tian, Y., Zhu, Y., Cao, W., & Cao, Q. (2019). Combining Color Indices and Textures of UAV-Based Digital Imagery for Rice LAI Estimation. *Remote Sensing*, *11*(15), 1763. https://doi.org/10.3390/rs11151763

Li, Z., Wang, J., Tang, H., Huang, C., Yang, F., Chen, B., Wang, X., Xin, X., & Ge, Y. (2016). Predicting Grassland Leaf Area Index in the Meadow Steppes of Northern China: A Comparative Study of Regression Approaches and Hybrid Geostatistical Methods. *Remote Sensing*, *8*(8), 632. https://doi.org/10.3390/rs8080632

Liu, Q., Piao, S., Janssens, I. A., Fu, Y., Peng, S., Lian, X., Ciais, P., Myneni, R. B., Peñuelas, J., & Wang, T. (2018). Extension of the growing season increases vegetation exposure to frost. *Nature Communications*, *9*(1), 426. https://doi.org/10.1038/s41467-017-02690-y

Liu, R., Wen, J., Wang, X., Wang, Z., Li, Z., Xie, Y., Zhu, L., & Li, D. (2019). Derivation of Vegetation Optical Depth and Water Content in the Source Region of the Yellow River using the FY-3B Microwave Data. *Remote Sensing*, *11*(13), 1536. https://doi.org/10.3390/rs11131536

Martinez, B., Cassiraga, E., Camacho, F., & Garcia-Haro, J. (2010). Geostatistics for Mapping Leaf Area Index over a Cropland Landscape: Efficiency Sampling Assessment. *Remote Sensing*, *2*(11), 2584–2606. https://doi.org/10.3390/rs2112584

Momen, M., Wood, J. D., Novick, K. A., Pangle, R., Pockman, W. T., McDowell, N. G., & Konings, A. G. (2017). Interacting Effects of Leaf Water Potential and Biomass on Vegetation Optical Depth. *Journal of Geophysical Research: Biogeosciences*, *122*(11), 3031–3046. https://doi.org/10.1002/2017JG004145

O'Neill, P. E., Chan, S., Njoku, E. G., Jackson, T., Bindlish, R., & Chaubell, J. (2019). 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.* https://doi.org/https://doi.org/10.5067/017XZSKMLTT2

Omer, G., Mutanga, O., Abdel-Rahman, E., & Adam, E. (2016). 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 Sensing*, *8*(4), 324. https://doi.org/10.3390/rs8040324

ORNL DAAC. (2018). MODIS and VIIRS Land Products Global Subsetting and Visualization Tool. *Available Online: Https://Modis.Ornl.Gov/Globalsubset/ (Accessed on 24 February 2020)*. https://doi.org/https://doi.org/10.3334/ORNLDAAC/1379

PIAO, S., FANG, J., ZHOU, L., CIAIS, P., & ZHU, B. (2006). Variations in satellite-derived phenology in China's temperate vegetation. *Global Change Biology*, *12*(4), 672–685. https://doi.org/10.1111/j.1365-2486.2006.01123.x

Piao, S., Liu, Q., Chen, A., Janssens, I. A., Fu, Y., Dai, J., Liu, L., Lian, X., Shen, M., & Zhu, X. (2019). Plant phenology and global climate change: Current progresses and challenges. *Global Change Biology*, *25*(6), 1922–1940. https://doi.org/10.1111/gcb.14619

Probst, P., Wright, M. N., & Boulesteix, A. (2019). Hyperparameters and tuning strategies for random forest. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *9*(3). https://doi.org/10.1002/widm.1301

Qu, Y., & Zhuang, Q. (2018). Modeling leaf area index in North America using a process-based terrestrial ecosystem model. *Ecosphere*, *9*(1). https://doi.org/10.1002/ecs2.2046

Reichle, R., De Lannoy, G., Koster, R. D., Crow, W. T., Kimball, J. S., & Liu, Q. (2018). 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*.

Roosjen, P. P. J., Brede, B., Suomalainen, J. M., Bartholomeus, H. M., Kooistra, L., & Clevers, J. G. P. W. (2018). 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. *International Journal of Applied Earth Observation and Geoinformation*, *66*, 14–26. https://doi.org/10.1016/j.jag.2017.10.012

Santhana Vannan, S. K., Cook, R. B., Holladay, S. K., Olsen, L. M., Dadi, U., & Wilson, B. E. (2009). A Web-Based Subsetting Service for Regional Scale MODIS Land Products. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *2*(4), 319–328. https://doi.org/10.1109/JSTARS.2009.2036585

Scott, D. W. (2010). Scott's rule. *Wiley Interdisciplinary Reviews: Computational Statistics*, *2*(4), 497–502. https://doi.org/10.1002/wics.103

Scranton, K., & Amarasekare, P. (2017). Predicting phenological shifts in a changing climate. *Proceedings of the National Academy of Sciences*, *114*(50), 13212–13217. https://doi.org/10.1073/pnas.1711221114

Stavi, I. (2019). Wildfires in Grasslands and Shrublands: A Review of Impacts on Vegetation, Soil, Hydrology, and Geomorphology. *Water*, *11*(5), 1042. https://doi.org/10.3390/w11051042

Sulla-Menashe, D., Gray, J. M., Abercrombie, S. P., & Friedl, M. A. (2019). Hierarchical mapping of annual global land cover 2001 to present: The MODIS Collection 6 Land Cover product. *Remote Sensing of Environment*, *222*, 183–194. https://doi.org/10.1016/j.rse.2018.12.013

Sun, L., Gao, F., Anderson, M., Kustas, W., Alsina, M., Sanchez, L., Sams, B., McKee, L., Dulaney, W., White, W., Alfieri, J., Prueger, J., Melton, F., & Post, K. (2017). Daily Mapping of 30 m LAI and NDVI for Grape Yield Prediction in California Vineyards. *Remote Sensing*, *9*(4), 317. https://doi.org/10.3390/rs9040317

Tian, F., Brandt, M., Liu, Y. Y., Verger, A., Tagesson, T., Diouf, A. A., Rasmussen, K., Mbow, C., Wang, Y., & Fensholt, R. (2016). 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 Sensing of Environment*, *177*, 265–276. https://doi.org/10.1016/j.rse.2016.02.056

Tian, F., Wigneron, J.-P., Ciais, P., Chave, J., Ogée, J., Peñuelas, J., Ræbild, A., Domec, J.-C., Tong, X., Brandt, M., Mialon, A., Rodriguez-Fernandez, N., Tagesson, T., Al-Yaari, A., Kerr, Y., Chen, C., Myneni, R. B., Zhang, W., Ardö, J., & Fensholt, R. (2018). Coupling of ecosystem-scale plant water storage and leaf phenology observed by satellite. *Nature Ecology & Evolution*, *2*(9), 1428–1435. https://doi.org/10.1038/s41559-018-0630-3

Tong, C., Wang, H., Magagi, R., Goïta, K., Zhu, L., Yang, M., & Deng, J. (2020). Soil Moisture Retrievals by Combining Passive Microwave and Optical Data. *Remote Sensing*, *12*(19), 3173. https://doi.org/10.3390/rs12193173

Vittucci, C., Vaglio Laurin, G., Tramontana, G., Ferrazzoli, P., Guerriero, L., & Papale, D. (2019). Vegetation optical depth at L-band and above ground biomass in the tropical range: Evaluating their relationships at continental and regional scales. *International Journal of Applied Earth Observation and Geoinformation*, *77*, 151–161. https://doi.org/10.1016/j.jag.2019.01.006

Wang, C., Fu, B., Zhang, L., & Xu, Z. (2019). Soil moisture–plant interactions: an ecohydrological review. *Journal of Soils and Sediments*, *19*(1), 1–9. https://doi.org/10.1007/s11368-018-2167-0

WAY, D. A., & MONTGOMERY, R. A. (2015). Photoperiod constraints on tree phenology, performance and migration in a warming world. *Plant, Cell & Environment*, *38*(9), 1725–1736. https://doi.org/10.1111/pce.12431

Wigneron, J.-P., Jackson, T. J., O'Neill, P., De Lannoy, G., de Rosnay, P., Walker, J. P., Ferrazzoli, P., Mironov, V., Bircher, S., Grant, J. P., Kurum, M., Schwank, M., Munoz-Sabater, J., Das, N., Royer, A., Al-Yaari, A., Al Bitar, A., Fernandez-Moran, R., Lawrence, H., … Kerr, Y. (2017). 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 Sensing of Environment*, *192*, 238–262. https://doi.org/10.1016/j.rse.2017.01.024

Wigneron, Jean-Pierre, Li, X., Frappart, F., Fan, L., Al-Yaari, A., De Lannoy, G., Liu, X., Wang, M., Le Masson, E., & Moisy, C. (2021). SMOS-IC data record of soil moisture and L-VOD: Historical development, applications and perspectives. *Remote Sensing of Environment*, *254*, 112238. https://doi.org/10.1016/j.rse.2020.112238

Xie, Q., Huang, W., Liang, D., Chen, P., Wu, C., Yang, G., Zhang, J., Huang, L., & Zhang, D. (2014). Leaf Area Index Estimation Using Vegetation Indices Derived From Airborne Hyperspectral Images in Winter Wheat. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *7*(8), 3586–3594. https://doi.org/10.1109/JSTARS.2014.2342291

Xue, J., & Su, B. (2017). Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications. *Journal of Sensors*, *2017*, 1–17. https://doi.org/10.1155/2017/1353691

Yan, K., Park, T., Yan, G., Liu, Z., Yang, B., Chen, C., Nemani, R., Knyazikhin, Y., & Myneni, R. (2016). Evaluation of MODIS LAI/FPAR Product Collection 6. Part 2: Validation and Intercomparison. *Remote Sensing*, *8*(6), 460. https://doi.org/10.3390/rs8060460

Zhang, R., Kim, S., & Sharma, A. (2019). A comprehensive validation of the SMAP Enhanced Level-3 Soil Moisture product using ground measurements over varied climates and landscapes. *Remote Sensing of Environment*, *223*, 82–94. https://doi.org/10.1016/j.rse.2019.01.015

Zhang, Z., & Grabchak, M. (2013). Bias Adjustment for a Nonparametric Entropy Estimator. *Entropy*, *15*(12), 1999–2011. https://doi.org/10.3390/e15061999

2.9 Supplementary Materials

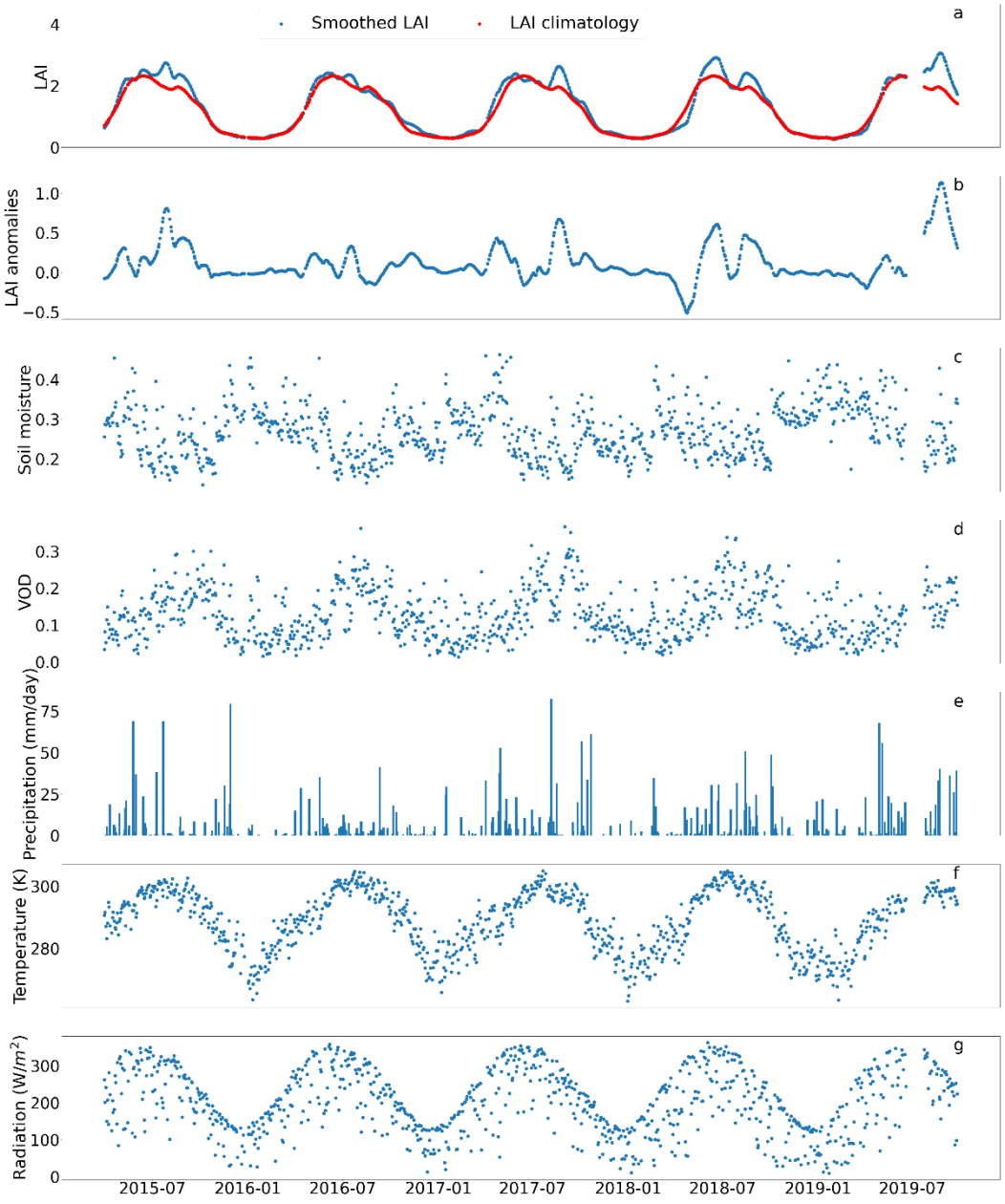

Figure 2.S1 Daily smoothed LAI and LAI climatology (a), daily LAI anomalies (b), soil moisture (c), vegetation optical depth (VOD) (d), precipitation (e), temperature (f), and radiation (g) time series at the representative grassland site (green star in figure 2.1).

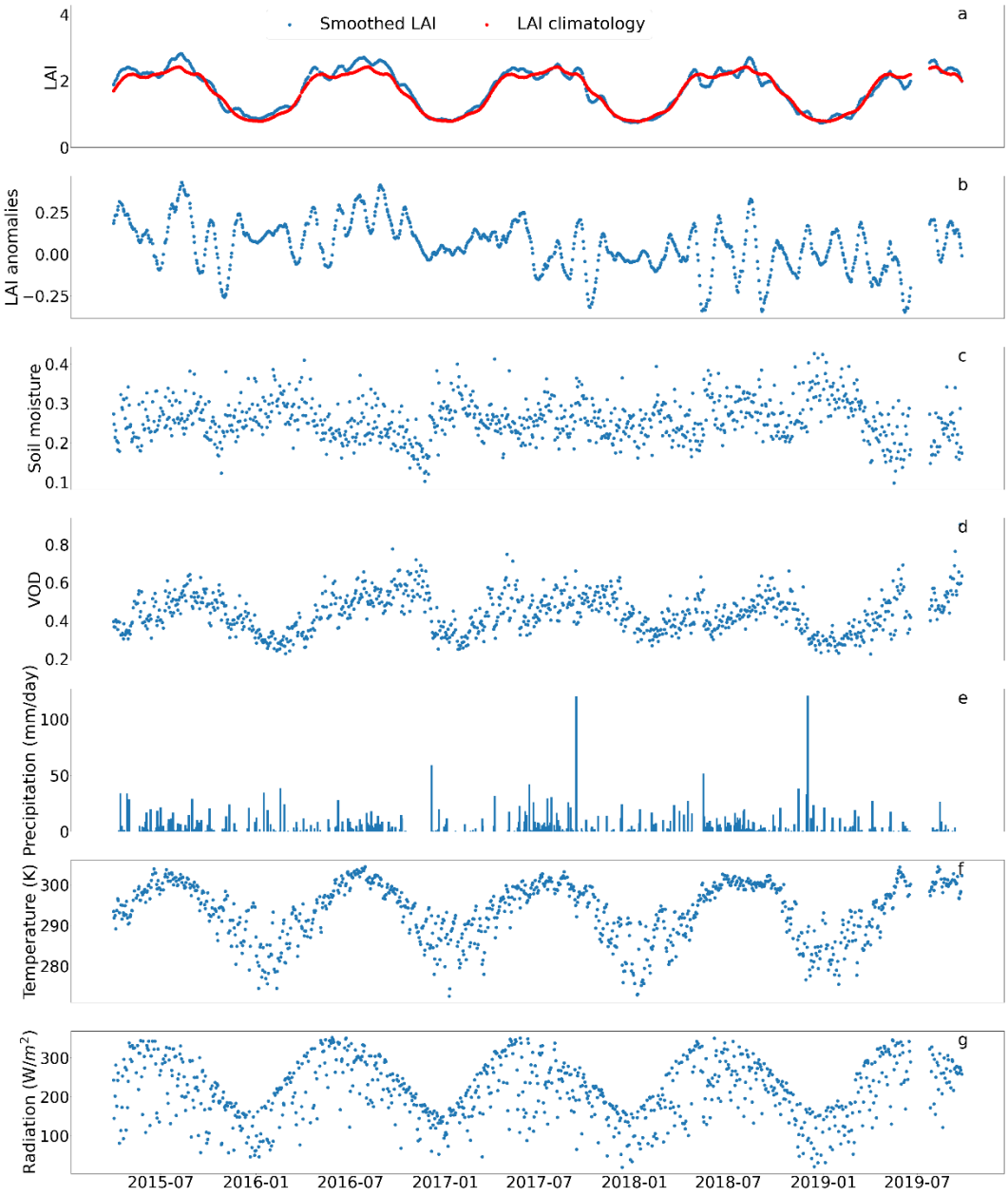

Figure 2.S2 Daily smoothed LAI and LAI climatology (a), daily LAI anomalies (b), soil moisture (c), vegetation optical depth (VOD) (d), precipitation (e), temperature (f), and radiation (g) time series at the representative savanna site (red start in figure 2.1).

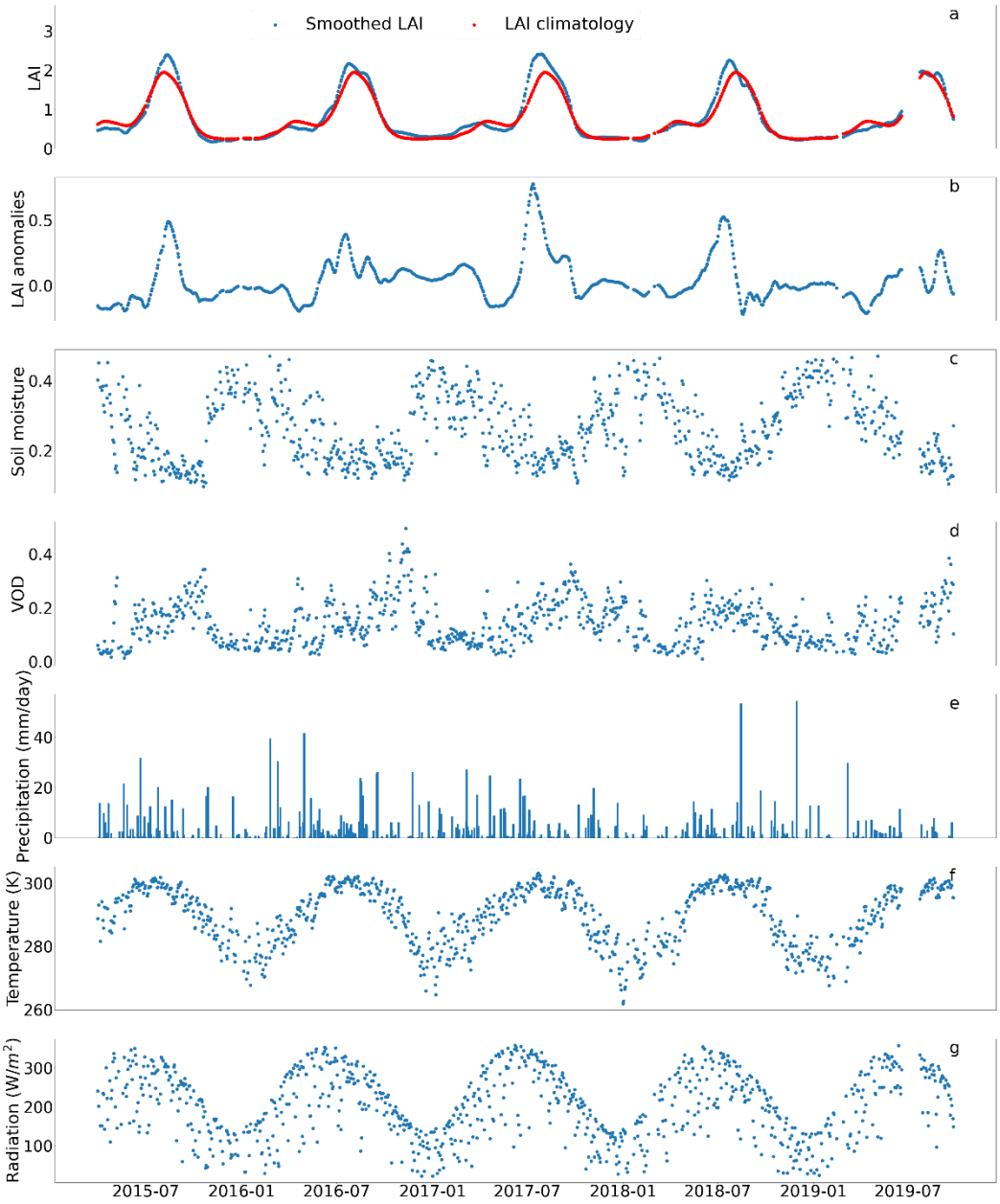

Figure 2.S3 Daily smoothed LAI and LAI climatology (a), daily LAI anomalies (b), soil moisture (c), vegetation optical depth (VOD) (d), precipitation (e), temperature (f), and radiation (g) time series at the representative cropland site (yellow star in figure 2.1).

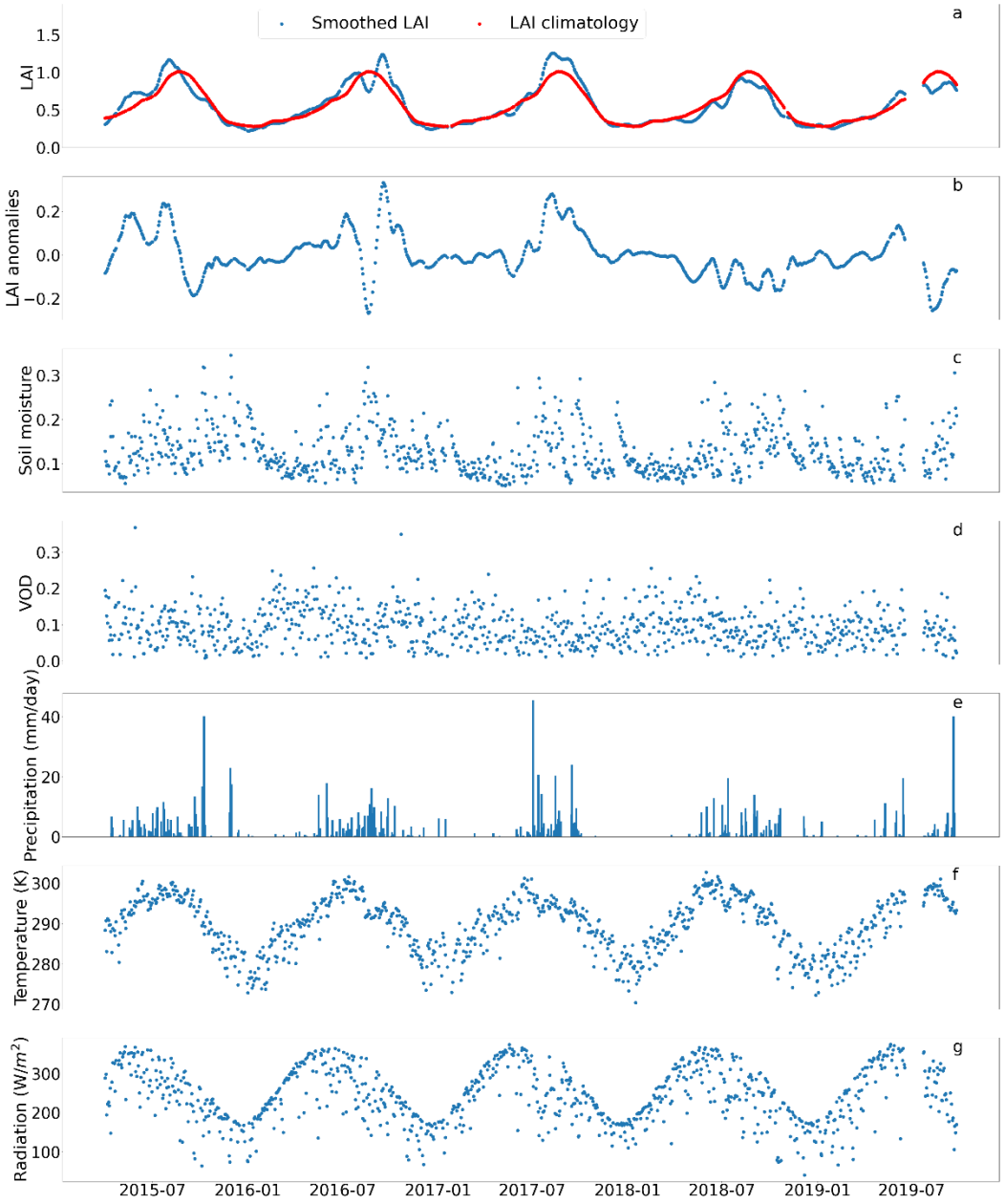

Figure 2.S4 Daily smoothed LAI and LAI climatology (a), daily LAI anomalies (b), soil moisture (c), vegetation optical depth (VOD) (d), precipitation (e), temperature (f), and radiation (g) time series at the representative shrubland site (blue star in figure 2.1).

Chapter 3. Information - Based Uncertainty Decomposition in Dual Channel Microwave Remote Sensing of Soil Moisture

Bonan Li[1,2], Stephen P. Good[1,2]

[1]Department of Biological & Ecological Engineering, Oregon State University

[2]Water Resources Graduate Program, Oregon State University

Status: Accepted by *Hydrology and Earth System Sciences*

3.1 Abstract

NASA's Soil Moisture Active-Passive (SMAP) mission characterizes global spatiotemporal patterns in surface soil moisture using dual L-band microwave retrievals of horizontal (T_{Bh}) and vertical (T_{Bv}) polarized microwave brightness temperatures through a modeled mechanistic relationship between vegetation opacity, surface scattering albedo, and soil effective temperature (T_{eff}). Although this model has been validated against in situ soil moisture, there is a lack of systematic characterization of where and why SMAP estimates deviate from the *in situ* observations. Here, we assess how the information content of in situ soil moisture observations from the US Climate Reference Network contrasts with (1) the information contained within raw SMAP observations (i.e., 'informational random uncertainty') derived from T_{Bh}, T_{Bv} and T_{eff} themselves, and (2) with the information contained in SMAP's Dual Channel Algorithm (DCA) soil moisture estimates (i.e., 'informational model uncertainty') derived from the model's inherent structure and parameterizations. The results show that, on average, 80% of the information in the *in situ* soil moisture is unexplained by SMAP DCA soil moisture estimates. 35% of the unexplained information is caused by the loss of information in the DCA modeling process while the remainder is induced by a lack of additional explanatory power within T_{Bh}, T_{Bv} and T_{eff}. Overall, retrieval quality of SMAP DCA soil moisture, denoted as the Pearson correlation coefficient between SMAP DCA soil moisture and in situ soil moisture, is negatively correlated with the informational uncertainties, with slight differences across different landcovers. The informational model uncertainty (Pearson correlation of -0.59) was found to be more influential than the informational random uncertainty (Pearson correlation of -0.34), suggesting that the poor performance of SMAP DCA at some locations is driven by model parameterization and/or structure and not underlying satellite measurements of T_{Bh} and T_{Bv}. A decomposition of mutual information between T_{Bh}, T_{Bv} and DCA soil moisture shows that on average 58% of information provided by T_{Bh} and T_{Bv} to DCA estimates is redundancy. The amount of information redundantly and synergistically provided by T_{Bh} and T_{Bv} was found to be tightly related (Pearson correlation of 0.79 and -0.82, respectively) to the retrieval quality of SMAP DCA. T_{Bh} and T_{Bv} tend to contribute large redundant information to DCA estimates under

surfaces or conditions where DCA makes better retrievals. This suggests that the informational redundancy and synergy between these remotely sensed observations can be indicative about soil moisture retrieval quality. This study provides a baseline approach that can also be applied to evaluate other remote sensing models and understand informational loss as satellite retrievals are translated to end user products.

3.2 Introduction

Accurate information on soil moisture is of great importance for understanding various biophysical processes in hydrology, agronomy, and ecosystem sciences (Bassiouni et al., 2020; Uber et al., 2018). The poor spatial representativeness of *in-situ* soil moisture sensors, combined with their labor-intensive installation and maintenance, impedes the application these sensors to understand large scale ecosystem phenomena (Babaeian et al., 2019; Petropoulos et al., 2015). Spaceborne passive microwave remote sensing has been developed as a reliable method to estimate surface soil moisture at large scales (Wigneron et al., 2017). It leverages the large discrepancies in dielectric properties between liquid water and dry soil that result in a high dependency of soil dielectric constants on soil moisture (Njoku & Entekhabi, 1996). Various microwave frequencies have been available to date, amongst which the L-band microwave frequencies were found to be desirable for soil moisture estimations because they can sense soil moisture at a relatively deeper layer (~5cm) and can provide greater vegetation penetration power (Mohanty et al., 2017). Though microwave remote sensing has been investigated for decades, significant uncertainties still exist in both microwave radiometry and in the algorithms used to translate microwave observations to soil moisture estimates (Gruber et al., 2020).

Passive L-band remote sensing soil moisture estimation uses a radiometer to measure surface emission intensity, which is proportional to the brightness temperature (Wang & Qu, 2009). The brightness temperature is linked with soil moisture and vegetation opacity through the '*tau-omega'* emission model and parameterized by soil and vegetation functions (Jackson et al., 1982; Mo et al., 1982). The '*tau-omega'* model rationale has been adopted by NASA's Soil Moisture Active-Passive (SMAP) mission, which is one of the earth observation missions dedicated to estimate soil moisture at L-band microwave frequency (Entekhabi et al., 2010a). SMAP implemented two primary algorithms: (1) the single channel algorithm (SCA) that uses one polarized brightness temperature as primary input to retrieve soil moisture and (2) the dual channel algorithm (DCA) that retrieves soil moisture and vegetation opacity simultaneously by taking the polarized brightness temperature information in both horizontal and vertical directions (O'Neill et al., 2020a) . There is strong interest in the DCA approach because of its independent

estimation of vegetation opacity in lieu of the specified vegetation climatology employed by the SCA (O'Neill et al., 2020a). Other L-band focused satellite mission such as Soil Moisture and Ocean Salinity (SMOS) retrieves both soil moisture and vegetation optical depth by using numerous brightness measurements for different incidence angles (Kerr et al., 2012). Additionally, it has been suggested that using a time-integrated vegetation opacity, as is employed in the multi-temporal dual channel algorithm (MT-DCA) for instance (Konings et al., 2016), improves the estimates of soil and vegetation state. These contrasting approaches, as well as other studies on SMAP's temporal polarized ratio algorithm (TPRA) (Gao et al., 2020) and regularized dual channel algorithm (RDCA) (Chaubell et al., 2020), suggested there is still uncertainty about how SMAP observations of horizontal and vertical brightness temperature can be best translated into estimates of surface properties. Although SMAP can provide spatially explicit soil moisture estimates that have been shown to be useful to understand a set of ecohydrological problems (Dadap et al., 2019; Feldman et al., 2018), the soil moisture retrievals are still subject to significant amount of uncertainty due to the imperfection of the model and the forcing datasets. It is also important to consider the how the amount of duplicate information carried within a set of observations limits the number of independent parameters to be inferred (Konings et al., 2015). Therefore, it is critical to diagnosis and quantify the causality of the uncertainty caused by the SMAP algorithm to improve the soil moisture and vegetation opacity retrieval quality.

SMAP soil moisture products have been extensively validated against well-calibrated *in situ* soil moisture using unbiased root mean square error (ubRMSE), bias, RMSE Pearson correlation coefficients and triple collocation method at 'core' and 'sparse' validation sites (Chan et al., 2016; Chen et al., 2017; Colliander et al., 2017; Zhang et al., 2019). These validation investigations found that SMAP met the required accuracy target (ubRMSE, 0.04 m^3/m^3) on average, while there exist some locations where the performance of SMAP did not meet the expected performance. All these validation studies were focused on finding the general uncertainty of SMAP (which is the deviation of SMAP soil moisture from the *in situ* or reference soil moisture) and cannot diagnose and differentiate where the uncertainty arises. Indeed, the causality of uncertainty of SMAP soil moisture may arise from two aspects: (1) the uncertainty

due to the inaccuracies from forcing the datasets and (2) the uncertainty due to poor model structure and parameterizations. In addition, the assessment metrics used in these evaluation studies are either heavily dependent on *in situ* soil moisture or additional reference datasets, which does not allow for SMAP to be validated in some remote and inaccessible areas.

The challenges faced by previous SMAP evaluation investigations can be resolved by leveraging two information quantities (Shannon, 1948): (1) Shannon's entropy, which is the amount of information required to fully describe a random variable and (2) mutual information (Cover & Thomas, 2005), which represents the amount information of knowing one variable given the knowledge of another or a set of random variables. (Gong et al., 2013) first leveraged these information quantities to partition overall uncertainty in the hydrological modeling process into two categories: (1) random uncertainty that arises by incompleteness of exploratory variable and/or inherent stochasticity of forcing datasets, and (2) model uncertainty that is contributed by poor model parameterization or formulation. The random uncertainty is not resolvable for the given system as it is only related to the probability distributions of the forcing data itself, while the model uncertainty is reduceable by a better model parameterization.

Given that both horizontal and vertical polarized brightness temperatures are measured by SMAP, it is unclear how each polarization contributes information to the overall performance of the DCA. Recent research on partial information decomposition has provided tremendous opportunities for understanding the nuanced interactions among different variables and model structure. Initially proposed by (Williams & Beer, 2010) and further advanced by (Goodwell & Kumar, 2017), this approach has been used to understand environmental processes that link two source variables with a target variable by partitioning multivariate mutual information into unique, redundant and synergistic components. The unique information represents the amount of information shared with the target variable from each individual source variable separately (Finn & Lizier, 2018). Synergistic information is the information provided to the target while both source variables act jointly (Kunert-Graf et al., 2020). Redundant information is the overlapping information that both source variables redundantly provide to a target (Wibral et al., 2017). Information partitioning brings new insight by unambiguously characterizing the

interdependencies between source variables and a target variable without any underlying assumption (Goodwell et al., 2018).

The overall objective of this study is to demonstrate that by assessing how information flows through satellite algorithms from raw retrievals to end user products, we can illuminate areas where improvements can be made and diagnose instances where algorithm estimates are expected to be uncertain. In this study, we focus on (1) quantifying the random uncertainty and model uncertainty in SMAP's Dual Channel Algorithm (DCA) and understand how these uncertainties are related to DCA retrieval quality; (2) exploring how the partial information components between SMAP DCA soil moisture and horizontally polarized and vertically polarized brightness temperature can be used to indicate overall DCA soil moisture retrieval performance.

3.3 Material and methods

3.3.1 In situ Soil Moisture

The US Climate Reference Network (USCRN) is a systematic and sustained network that is operated and maintained by National Oceanic and Atmospheric Administration (NOAA) to support climate-impact research with continuous high-quality field observed soil moisture, soil temperature and windspeed at different temporal scales (Diamond et al., 2013). The USCRN provides soil moisture observations at five different standard depth (5 cm, 10 cm 20 cm, 50 cm, and 100 cm) in 114 locations of Contiguous U.S. (CONUS) (Bell et al., 2013). These *in situ* datasets have been used for a wide variety of research such as drought evaluation and satellite soil moisture validation (Bell et al., 2015; Leeper et al., 2017). The hourly soil moisture (beta version product) datasets at the depth of 5 cm were collected from 58 (15 croplands, 32 grasslands, 5 shrublands, 2 savannas, 4 mixed) selected USCRN stations (Fig. 3.1 and Table 3.S1) based on the availability of *in situ* soil moisture dataset and the data quality of SMAP pixels in the study period of March 31, 2015 to December 10, 2020.

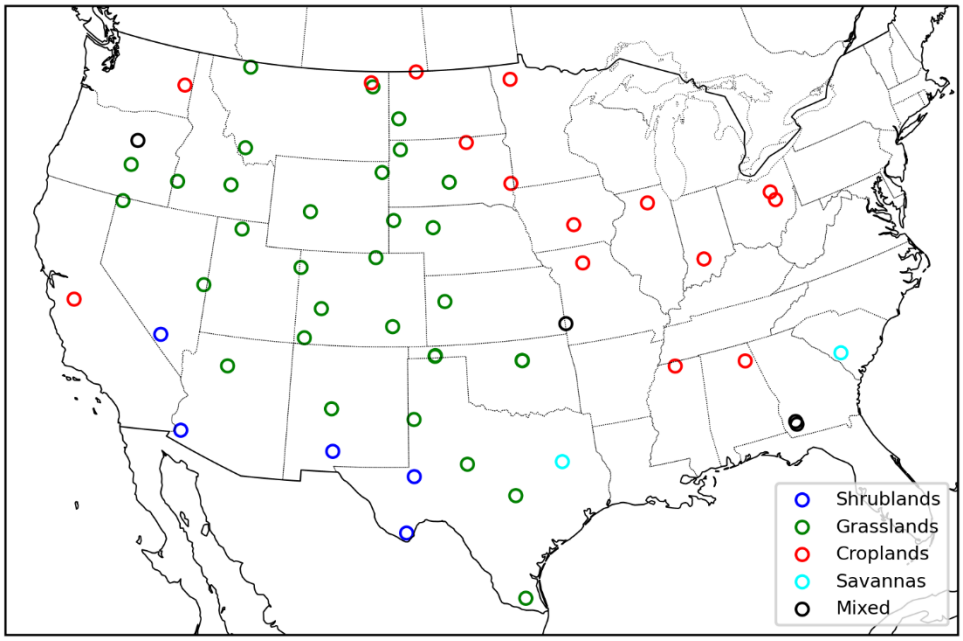

Figure 3.1 Spatial distribution of selected USCRN stations classified by landcovers.

3.3.2 SMAP Level-2 Datasets

In this study, we acquired the water body corrected horizontally polarized brightness temperature (T_{Bh}), vertically polarized brightness temperature (T_{Bv}), soil effective temperature (T_{eff}), DCA soil moisture and the fraction of landcover at each selected USCRN station from SMAP Level-2 Radiometer Half-Orbit 36 km EASE-Grid Soil moisture, Version 7 data product (O'Neill et al., 2020b) in the same period as the USCRN soil moisture at every station. The extracted data series were filtered by the internal quality flags of T_{Bh} ("tb_qual_flag_h"), T_{Bv} ("tb_qual_flag_v") and DCA ("retrieval_qual_flag_option3") as provided in SMAP data files. We retain data points at a particular SMAP observation time when they all simultaneous pass quality control and fall within their correspondent valid ranges (e.g., 0 ~ 330K for T_{Bh} and T_{Bv}, 253.15K ~ 313.15K for T_{eff}, > 0.02m^3/m^3 for DCA soil moisture) as specified in the SMAP documentation (Chan, 2020). On average, the number of datapoints across all the sites is 1090 with a minimum of 225 and a maximum of 1651. DCA retrieves soil moisture based on the '*tau-omega'* model (Jackson et al., 1982; Mo et al., 1982), which is a well-known radiative transfer based soil moisture retrieval algorithm in the passive microwave soil moisture community. It

requires the brightness temperatures as the main inputs, soil effective temperature as an ancillary input, and is parameterized based on overlaying vegetation and soil surface information (Njoku & Entekhabi, 1996). The DCA iteratively feeds the '*tau-omega'* model with initial guesses of soil moisture and vegetation optical depth. The retrieved soil moisture is assumed to be close to the real value when the estimated brightness temperatures are similar to the satellite observed brightness temperature (Konings et al., 2017; O'Neill et al., 2020a). Compared to the SCAs, the DCA uses a different polarization mixing factor function and different values of vegetation single scattering albedo (O'Neill et al., 2020a).

The SMAP fraction of landcover data field provides the fraction of top three dominate landcovers that were classified by International Geosphere – Biosphere Programme (IGBP) ecosystem surface classification scheme at each pixel (Chan, 2020). The IGBP classified land surface into water, evergreen needleleaf forest, evergreen broadleaf forest, deciduous needleleaf forest, deciduous broadleaf forest, mixed forest, closed shrublands, open shrublands, woody savannas, savannas, grasslands, permanent wetlands, croplands, urban and built-up, croplands/natural vegetation mosaics, snow and ice, barren (Seitzinger et al., 2015). In this study, the landcover of the study site was classified as the most dominate landcover if the fraction of the most dominate landcover was greater than 50%. Otherwise, the landcover of the study site is classified as the "mixed" landcover. Furthermore, the study sites that are dominated by woody savanna were classified as savannas, by closed/open shrublands were classified as shrublands, by cropland/natural vegetation mosaics were classified as croplands. Sites meeting specified data requirements and their associated landcover classification are shown in Figure 2.1. Additionally, the 500m leaf area index (LAI) of each site was obtained from NASAs Moderate Resolution (Myneni et al., 2015; ORNL DAAC, 2018) Imaging Spectrometer (MODIS) mission and averaged in time. Within each site the mean and standard deviation of LAI of all pixels within each SMAP pixel was calculated as a measure of vegetation biomass and variability.

3.3.3 Information – Based Uncertainty Partitioning

The fundamental quantity of information theory is Shannon's entropy (Shannon, 1948), which represents the amount of information required to fully describe a random variable (Cover

& Thomas, 2005). Shannon's entropy is the basic building block of computing mutual information and the informational uncertainties. The entropy of a single random variable is defined as

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

where $p(x)$ is the probability mass function of random variable X. The estimation of $p(x)$ often involves discretizing the values of X into a set of bins and then the $p(x)$ of a specific bin is computed by dividing the total number of datapoints within a specific bin by the total of number of data points of X. The number of bins in this study is estimated by Freedman-Diaconis binning method (Freedman & Diaconis, 1981). The entropy calculated by eq. (1) is in unit of bits.

Previous study has indicated that this method (eq. (1)) may underestimate the true entropy (Paninski, 2003). Therefore, we leveraged the simple Miller-Madow corrected entropy estimator (Zhang & Grabchak, 2013) and we also normalization the entropy to remove the bias that may cause by the heterogeneity in length of available datasets across all stations. We acknowledge that there exist several entropies estimation methods. However, we select the Miller-Madow correction based on its simplicity and effectiveness. The corrected and normalized entropy is then expressed as

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

where $H_{\mathrm{CN}}(\mathrm{X})$ is the Miller-Madow corrected and normalized entropy of random variable X (hereafter entropy), $H(\mathrm{X})$ is the uncorrected entropy from eq. (1), n is the number of data points of X , K is the number of non-zero probabilities (bins contains more than one data point) based on the fixed binned method (Freedman & Diaconis, 1981). In this study, all entropies of single random variables in the later equations (i.e., $H_{CN}(\mathrm{T_{Bh}})$, $H_{CN}(\mathrm{T_{Bv}})$, $H_{CN}(\textit{in situ})$ etc.) are computed using the combination of eq. (1) and eq. (2) with the replacement of $p(\bullet)$ by their individual probability mass functions.

The joint entropy (Cover & Thomas, 2005) is a critical intermediate information quantity to calculate these informational uncertainties. It represents the amount of information required to describe a set of random variables. The joint entropy for two random variables is defined as

$$H(\mathrm{X}, \mathrm{Y}) = -\sum_{x \in X} \sum_{y \epsilon Y} p(x, y) log_2\, p(x, y), \tag{3}$$

where $p(x, y)$ is the joint probability mass function associated with X and Y that is estimated by the same method mentioned above. The same normalization and correction method of eq. (2) is applied to joint entropy of eq. (3). The entropy after the correction and normalization is formulated as

$$H_{CN}(\mathrm{X}, \mathrm{Y}) = \frac{H(\mathrm{X},\mathrm{Y}) + \frac{K-1}{2n}}{log_2\, n}, \tag{4}$$

where H_{CN}(X, Y) is the corrected and normalized joint entropy of random variable associated with $\{X, Y\}$, H(X, Y) is the uncorrected and unnormalized entropy from eq. (3), n is the number of data points that were used to calculate the normalized joint entropy (hereafter joint entropy), K is the number of non-zero joint probabilities based on the Freeman and Diaconis method (Freedman & Diaconis, 1981). All the joint entropies that are associated with two or more random variables in the later equations (i.e., H_{CN}(*in situ*, DCA), H_{CN}(T_{Bh}, T_{Bv}, DCA), H_{CN}(T_{Bh}, T_{Bv}, T_{eff}, *in situ*) etc.) are computed using the combination of eq. (3) and eq. (4) with the replacement of $p(\bullet)$ by their joint probability mass functions, respectively.

Generally, modeling efforts are focused on capturing the information of a random variable of interest via other explanatory variables through some physically- or empirically- based models. However, most of models being constructed of natural processes are not perfect, and the model outputs are often not capable of capturing the exact relationship between the available input variables and the variable of interest (Gong et al., 2013). There exists a maximum achievable performance of a model that describes the variable of interest the best for a particular system given the available datasets (Gong et al., 2013); yet the detailed structure of this model is often unknown. Mutual information (Cover & Thomas, 2005), for instance I(A; B), is a measure of the amount information due to the knowledge of knowing either random variable A or B in the

function $I(\bullet;\bullet)$. Mutual information between model inputs and *in situ* observations of model output (hereafter *in situ* observations) can be used as a useful and effective measure of best achievable performance model because it links the model inputs and *in situ* observations only through the joint and marginal probability mass functions that do not involve any priori model assumptions (Gong et al., 2013).

The mutual information can be defined based on entropy and joint entropy (Cover & Thomas, 2005). The mutual information between T_{Bh} and DCA, and the mutual information between T_{Bv} and DCA, are computed as

$$I(T_{Bh}; \text{DCA}) = H_{CN}(T_{Bh}) + H_{CN}(\text{DCA}) - H_{CN}(T_{Bh}, \text{DCA}), \quad (5)$$

and

$$I(T_{Bv}; \text{DCA}) = H_{CN}(T_{Bv}) + H_{CN}(\text{DCA}) - H_{CN}(T_{Bv}, \text{DCA}), \quad (6)$$

The mutual information between *in situ* and DCA soil moisture is computed as

$$I(\text{DCA}; \textit{in situ}) = H_{CN}(\text{DCA}) + H_{CN}(\textit{ in situ }) - H_{CN}(\text{DCA}, \textit{in situ }), \quad (7)$$

The mutual information between DCA and *in situ* soil moisture is calculated as

$$I(T_{Bh}, T_{Bv}; \text{DCA}) = H_{CN}(T_{Bh}, T_{Bv}) + H_{CN}(\text{DCA}) - H_{CN}(T_{Bh}, T_{Bv}, \text{DCA}), \quad (8)$$

The mutual information between T_{Bh}, T_{Bv}, T_{eff} and *in situ* soil moisture is computed as

$$I(T_{Bh}, T_{Bv}, T_{eff}; \textit{in situ}) = H_{CN}(T_{Bh}, T_{Bv}, T_{eff}) + H_{CN}(\textit{in situ}) - H_{CN}(T_{Bh}, T_{Bv}, T_{eff}, \textit{in situ}), \quad (9)$$

We adopted the information uncertainty analysis by (Gong et al., 2013) and applied it to SMAP DCA. For a given system in which the inputs and output are linked via mathematical

functions, the mutual information between model output and *in situ* observation can never exceed the entropy of the *in situ* observations. Conceptually, the entropies of model output and *in situ* observations can be considered as two circles (of equal or unequal sizes) and the mutual information between model output and *in situ* observation can be viewed as the overlapping area of these two circles (Uda, 2020). Therefore, the maximum mutual information shared between model output and *in situ* is the minimum of the entropy of model output and *in situ* observations, i.e: $I(\text{DCA}, \textit{in situ}) \leq \min[H_{CN}(\text{DCA}), H_{CN}(\textit{in situ})]$. Intuitively, the overlapping area of two circles cannot be larger that of the smaller circle. Because we are focused on representing the observed soil condition, the information gap between *in situ* observations, $H_{CN}(\textit{in situ})$, and the mutual information shared between *in situ* observations and model output, $I(\text{DCA}, \textit{in situ})$, is defined as informational total uncertainty (I_{Tot}). This quantity describes how much of the information within *in situ* observations, as measured by $H_{CN}(\textit{in situ})$, is not captured by the estimator, as measured by $I(\text{DCA}, \textit{in situ})$. The mutual information between the *in situ* observations and the available explanatory variables is also always smaller than the entropy of *in situ* observations. This information gap, defined as informational random uncertainty (I_{Rnd}), is caused by the existence of inherent data uncertainty of the explanatory variables and a lack of complete explanatory variables to fully capture the information in the *in situ* observations (Gong et al., 2013). Furthermore, the mutual information between model inputs and *in situ* observations should equal to the mutual information between *in situ* observations and model output if the model hypothesis completely captures the true relationship between model inputs and *in situ* observations. However, it's commonly stated that "*All models are wrong*" and model assumptions typically cannot fully express the true relationship between the explanatory (Box, 1976) variables and *in situ* observations. Hence, the mutual information between model output and *in situ* observation is expected to be smaller than the mutual information between model inputs and *in situ* observations (Gong et al., 2013). This information gap, defined as informational model uncertainty (I_{Mod}) is induced by poor model assumption, formulations, and/or inappropriate model parameterizations. Therefore, the informational total uncertainty (I_{Tot}) is the sum of the informational random uncertainty and informational model uncertainty come naturally given the explicitly definition of these informational uncertainties.

In this study, the explanatory variables of DCA are T_{Bh}, T_{Bv} and the T_{eff}. The *in situ* observation and model output are *in situ* USCRN soil moisture and DCA soil moisture, respectively. Leveraging eq. (7) and eq. (9), the DCA informational random uncertainty (I_{Rnd}), DCA informational model uncertainty (I_{Mod}), and DCA total informational uncertainty (I_{Tot}) calculated are calculated as

$$I_{Rnd} = H_{CN}(in\ situ) - I(T_{Bh}, T_{Bv}, T_{eff}; in\ situ), \tag{10}$$

$$I_{Mod} = I(T_{Bh}, T_{Bv}, T_{eff}; in\ situ) - I(\mathrm{DCA}; in\ situ), \tag{11}$$

and

$$I_{Tot} = H_{CN}(in\ situ) - I(\mathrm{DCA}; in\ situ) = I_{Rnd} + I_{Mod}, \tag{12}$$

3.3.4 Partial Information Decomposition

The distinct informational contributions of T_{Bh} and T_{Bv} to the DCA soil moisture are assessed through a decomposition of the mutual information. This method partitions multivariate mutual information to unique, redundant and synergistic components (P. L. Williams & Beer, 2010). The decomposed information components on the DCA model inputs and outputs are expected to indicative of informational flow as model inputs are translated to end user products, and these components may have potential for evaluating model performance. The partial information decomposition of $I(T_{Bh}, T_{Bv}; \mathrm{DCA})$ can be expressed as

$$\begin{aligned} I(T_{Bh}, T_{Bv}; \mathrm{DCA}) = {} & U_h(T_{Bh}; \mathrm{DCA}) + U_v(T_{Bv}; \mathrm{DCA}) + \\ & R(T_{Bh}, T_{Bv}; \mathrm{DCA}) + S(T_{Bh}, T_{Bv}; \mathrm{DCA}), \end{aligned} \tag{13}$$

where U_h and U_v are unique information of T_{Bh} and T_{Bv} shared with DCA, respectively. S and R are the synergistic information and redundant information that T_{Bh} and T_{Bv} shared with DCA estimates, respectively. All the decomposed components are non-negative real values (P. L. Williams & Beer, 2010).

The mutual information between T_{Bh} and DCA and mutual information between T_{Bv} and DCA are formulated as

$$I(T_{Bh}; DCA) = U_h(T_{Bh};DCA) + R(T_{Bh},T_{Bv}; DCA), \quad (14)$$

and

$$I(T_{Bv}; DCA) = U_v(T_{Bv}; DCA) + R(T_{Bh},T_{Bv}; DCA), \quad (15)$$

In this approach, U_h, U_v, S and R are unknowns in the systems of equations (13) - (15). (Goodwell & Kumar, 2017) showed that the R can be formulated as

$$R = R_{min} + I_s*(R_{MMI} - R_{min}), \quad (16)$$

where

$$I_s = \frac{I(T_{Bh};T_{Bv})}{\min\{H_{CN}(T_{Bh}); H_{CN}(T_{Bv})\}}, \quad (17)$$

$$R_{MMI} = \min[I(T_{Bh}; DCA), I(T_{Bv}; DCA)], \quad (18)$$

and

$$R_{min} = \max(0, -II), \quad (19)$$

The II is the interaction information of T_{Bh}, T_{Bv}, DCA and can be computed as:

$$II = I(T_{Bh}; DCA| T_{Bv}) - I(T_{Bh}; DCA) = H_{CN}(T_{Bh}, DCA) + H_{CN}(T_{Bv}, DCA) + H_{CN}(T_{Bh}, T_{Bv}) - H_{CN}(T_{Bh}) - H_{CN}(T_{Bv}) - H_{CN}(DCA) - H_{CN}(T_{Bh}, T_{Bv}, DCA), \quad (20)$$

It is important to note that we used the point based *in situ* soil moisture as the ground truth in this analysis. Due to coarse spatial resolution of SMAP products, we acknowledge that *in situ* soil moisture may not be able to represent the spatial averaged soil moisture well. Although the nominal sensing depth of L-band SMAP soil moisture is 5 cm, the penetration depth was found to be even shallower in wetter regions (Shellito et al., 2016). In fact, the L-band sensing depth was found to as little as ~1cm (Jackson et al., 2012) and was found to vary with surface soil moisture conditions (Escorihuela et al., 2010; Raju et al., 1995). The heterogeneity in each pixel relative to the *in situ* observations together with the sensing depth disparity may influence the results of this study and can bias the estimation of informational uncertainties. We also acknowledge the existence of upscaling methods for matching the *in situ* soil moisture to satellite footprint (Crow et al., 2012). However, most of upscaling methods are achieved under the assistance of additional reference soil moisture datasets. This process introduces additional pieces of information in the DCA system making the separation of the uncertainty induced by the upscaling algorithm or additional dataset from other informational uncertainties much harder. Additionally, we used the hourly *in situ* data to best match the SMAP DCA soil moisture retrievals in time (within an hour). Based on current technologies, it is difficult to find a reference dataset with high frequency in time domain and good spatial coverage. Here we consider the informational uncertainty caused by the spatial mismatch and sensing depth mismatch between *in situ* and DCA soil moisture as part of the informational random uncertainty (I_{Rnd}) because the DCA is essentially a mathematical function and does not inherently require the inputs to be at a specific resolution. The spatial resolution is often the inherent attribute of the data. The reader should also keep these in mind while interpreting and adopting the results in this study.

3.4 Results

3.4.1 Information Quantities and System Informational Uncertainties

The estimated entropies across all the study sites are shown in figure 3.2 while the mutual information quantities are shown in figure 3.3. The brightness temperature entropies, $H_{CN}(\mathrm{T_{Bh}})$

and $H_{CN}(T_{Bv})$, generally follow the same pattern across sites with both having an average value of 0.37. Although the patterns of $H_{CN}(T_{Bh})$ and $H_{CN}(T_{Bv})$ are similar, the $H_{CN}(T_{Bh})$ is slightly more variable than $H_{CN}(T_{Bv})$ with the coefficients of variation (CV) being 0.053 and 0.046, respectively. $H_{CN}(T_{eff})$ shares the same average with $H_{CN}(T_{Bh})$ and $H_{CN}(T_{Bv})$, whereas the pattern of $H_{CN}(T_{eff})$ is quite different (Figure 3.2). On average, the H_{CN}(*in situ*) is 0.35, while H_{CN}(DCA) is 0.38. In general, H_{CN}(DCA) follows the pattern of H_{CN}(in situ) with the CV of H_{CN}(DCA) (0.064) being smaller than the CV of H_{CN}(*in situ*) (0.081).

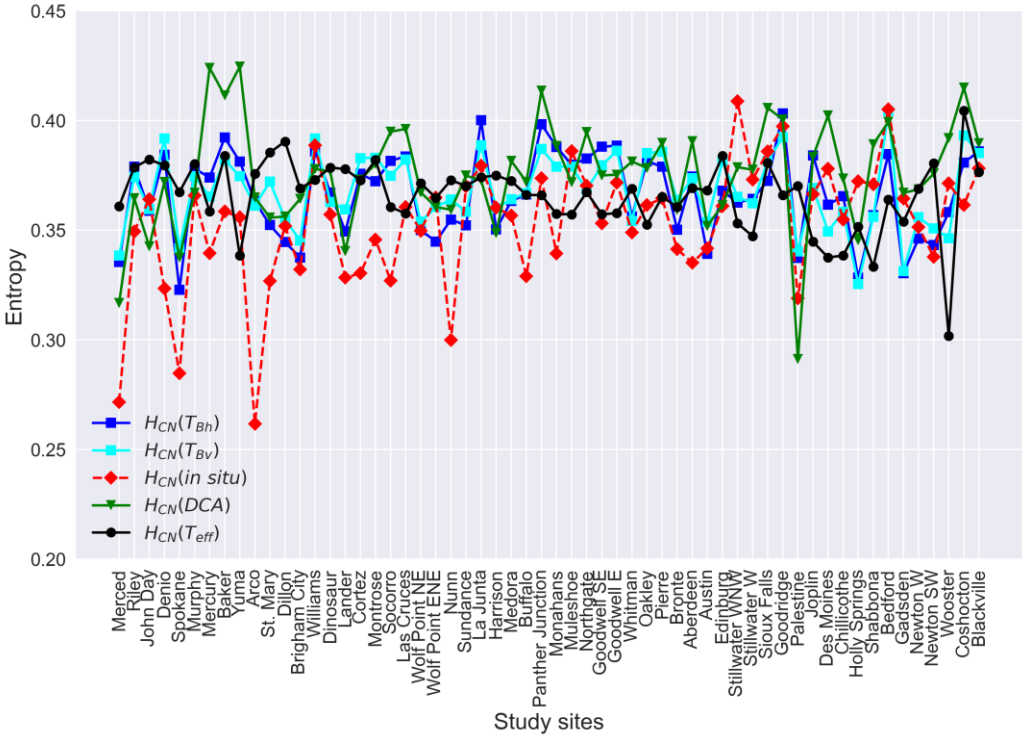

Figure 3.2 Entropies of horizontally polarized brightness temperature (T_{Bh}), vertically polarized brightness temperature (T_{Bv}), *in situ* soil moisture, DCA soil moisture, and soil effective temperature (T_{eff}) across the study sites. The sites are ordered by longitude (West to East).

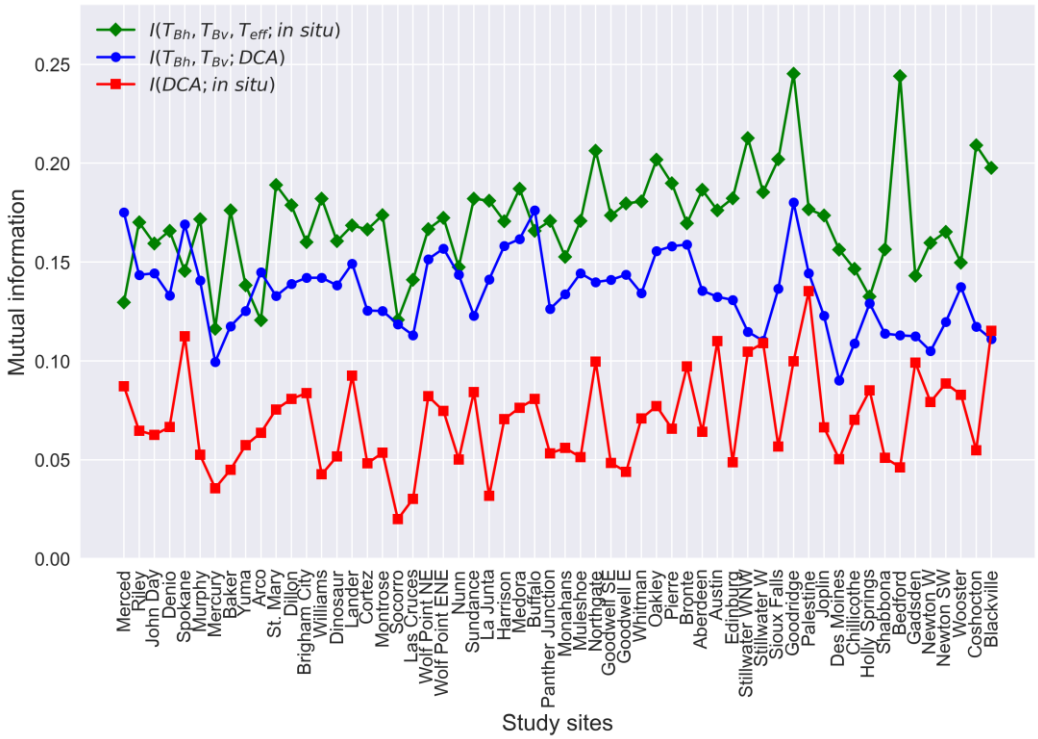

Figure 3.3 Mutual information between horizontally polarized brightness temperature (T_{Bh}), vertically polarized brightness temperature (T_{Bv}), soil effective temperature (T_{eff}) and *in situ* soil moisture; mutual information between horizontally polarized brightness temperature (T_{Bh}), vertically polarized brightness temperature (T_{Bv}) and DCA soil moisture; mutual information between DCA soil moisture and *in situ* soil moisture. See figure 3.2 caption for site ordering.

As shown in figure 3.4a, the entropies of the retrieved brightness temperatures and DCA model output, $H_{CN}(T_{Bh})$, $H_{CN}(T_{Bv})$ and H_{CN}(DCA), are significantly correlated with the entropy of *in situ* observations, H_{CN}(*in situ*), while no significant correlation is found between H_{CN}(*in situ*) and $H_{CN}(T_{eff})$. The H_{CN}(DCA) has the strongest correlation strength with H_{CN}(*in situ*) compared with other entropy quantities (Figure 3.4a). As expected, the mutual information quantities (Figure 3.3) are shown to be generally smaller than the entropy quantities (Figure 3.2). On average, $I(T_{Bh},T_{Bv}; DCA)$ is 0.14, while the I(DCA; *in situ*) and $I(T_{Bh},T_{Bv}, T_{eff}; in\ situ)$ are 0.07 and 0.17 (Figure 3.3), respectively. $I(T_{Bh},T_{Bv}, T_{eff}; in\ situ)$ and $I(T_{Bh},T_{Bv}; DCA)$ are significantly correlated (0.58 and -0.30) with H_{CN}(*in situ*), while no significant correlation is found for I(DCA; i*n situ*) and H_{CN}(*in situ*) (Figure 3.4b).

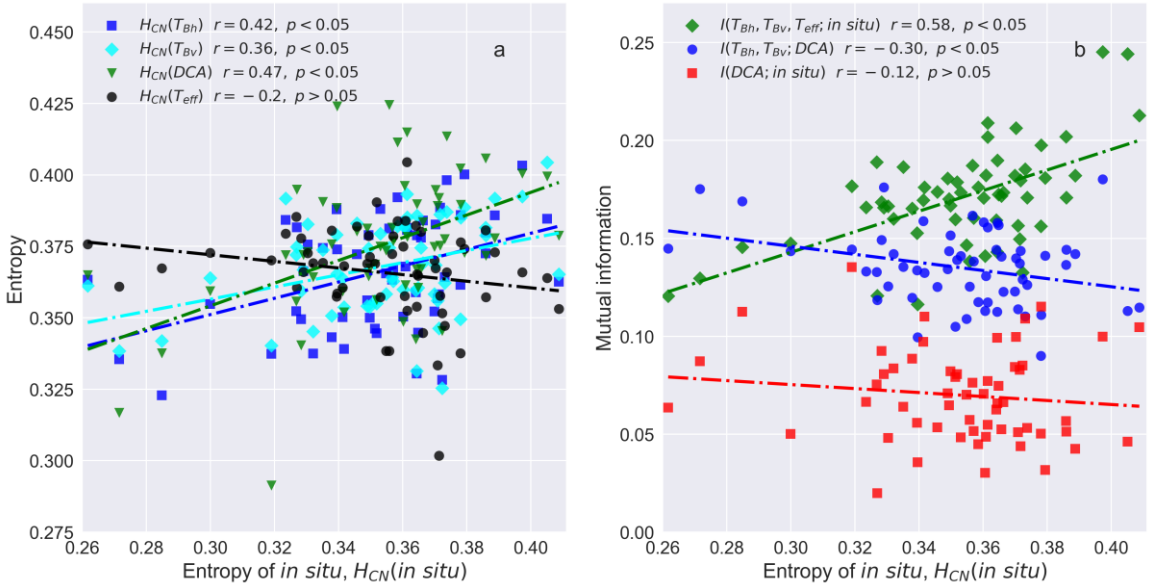

Figure 3.4 Entropy of *in situ* soil moisture against the entropies of DCA soil moisture, horizontally polarized brightness temperature (T_{Bh}), vertically polarized brightness temperature (T_{Bv}) and soil effective temperature (T_{eff}) (a) and mutual information quantities (b).

It is noticeable that there exists a large information gap between H_{CN}(*in situ*) in figure 3.2 and I(T_{Bh},T_{Bv}, T_{eff}; *in situ*) and I(T_{Bh},T_{Bv}, T_{eff}; *in situ*) and I(DCA; *in situ*) in figure 3.3. These information gaps confirm the existence of informational random uncertainty (I_{Rnd}) and informational model uncertainty (I_{Mod}) in the SMAP DCA system. When calculating informational quantities on a site-by-site basis and then averaging, the SMAP DCA explains 20% of the H_{CN}(*in situ*) leaving 80% of the H_{CN}(*in situ*) that is unexplained (Table 1) as informational total uncertainty (I_{Tot}). 35% of the I_{Tot} is caused by I_{Mod}, while the rest is induced by I_{Rnd}. The information uncertainties vary slightly across different landcovers. On average across sites, the SMAP DCA system is capable of capturing more information of H_{CN}(*in situ*) at croplands and savannas (Table 3.1). Shrublands have largest absolute I_{Rnd} (0.21) than other landcovers, while savannas have the largest proportion of I_{Rnd} to I_{Tot} (Table 1). I_{Mod} in absolute value is greater in shrublands, grasslands, and croplands with grasslands have the largest proportion of I_{Mod} to I_{Tot} (Table 3.1). When lumping all the datasets together and recalculating informational quantities, we observe that SMAP DCA captures 10% of the information in the *in situ* soil moisture and the proportion of I_{Mod} to I_{Tot} is larger.

Table 3.1 The amount of informational uncertainties in percentage. The values in the table are the average of each landcover. The values in "Overall" are the average of all the sites. The "Lumped" field is computed using all available dataset.

Landcover	Informational random uncertainty, I_{Rnd} (and its % of I_{Tot})	Informational model uncertainty, I_{Mod} (and its % of I_{Tot})	Informational total uncertainty, I_{Tot} (and its % of H_{CN}(*in situ*))	Number of Sites
Shrublands	0.21 (68%)	0.10 (32%)	0.31 (87%)	5
Grasslands	0.18 (63%)	0.11 (37%)	0.28 (81%)	32
Croplands	0.18 (66%)	0.10 (34%)	0.28 (78%)	15
Savannas	0.16 (73%)	0.06 (27%)	0.22 (64%)	2
Mixed	0.19 (68%)	0.09 (32%)	0.28 (79%)	4
Lumped	0.14 (46%)	0.17 (54%)	0.32 (90%)	58
Overall	0.18 (65%)	0.10 (35%)	0.28 (80%)	58

The relationship between different informational uncertainties and the Pearson correlation coefficients between *in situ* soil moisture and SMAP DCA soil moisture, a commonly adopted relative model evaluation metric in SMAP studies (Chan et al., 2016; Colliander et al., 2017), was evaluated. The I_{Tot}, I_{Mod} and I_{Rnd} are shown to be related how well the SMAP DCA soil moisture is correlated with *in situ* soil moisture (Figure 3.5). I_{Tot} is found to be negatively correlated (r = -0.69, Figure 3.5a) with the Pearson correlation between *in situ* soil moisture and SMAP DCA soil moisture. Similarly, I_{Mod} and I_{Rnd} are also shown to be negatively (-0.59 and -0.34 respectively) related to the Pearson correlation between *in situ* soil moisture and SMAP DCA soil moisture with I_{Mod} being more influential than I_{Rnd} (Figure 3.5b and 3.5c). These negative relationships are consistent with general expectations since SMAP tends to capture more information about the *in situ* soil moisture (i.e. lower I_{Tot}, I_{Mod} and I_{Rnd}) when it retrieves high quality datasets (higher correlation between *in situ* soil moisture and SMAP DCA soil moisture).

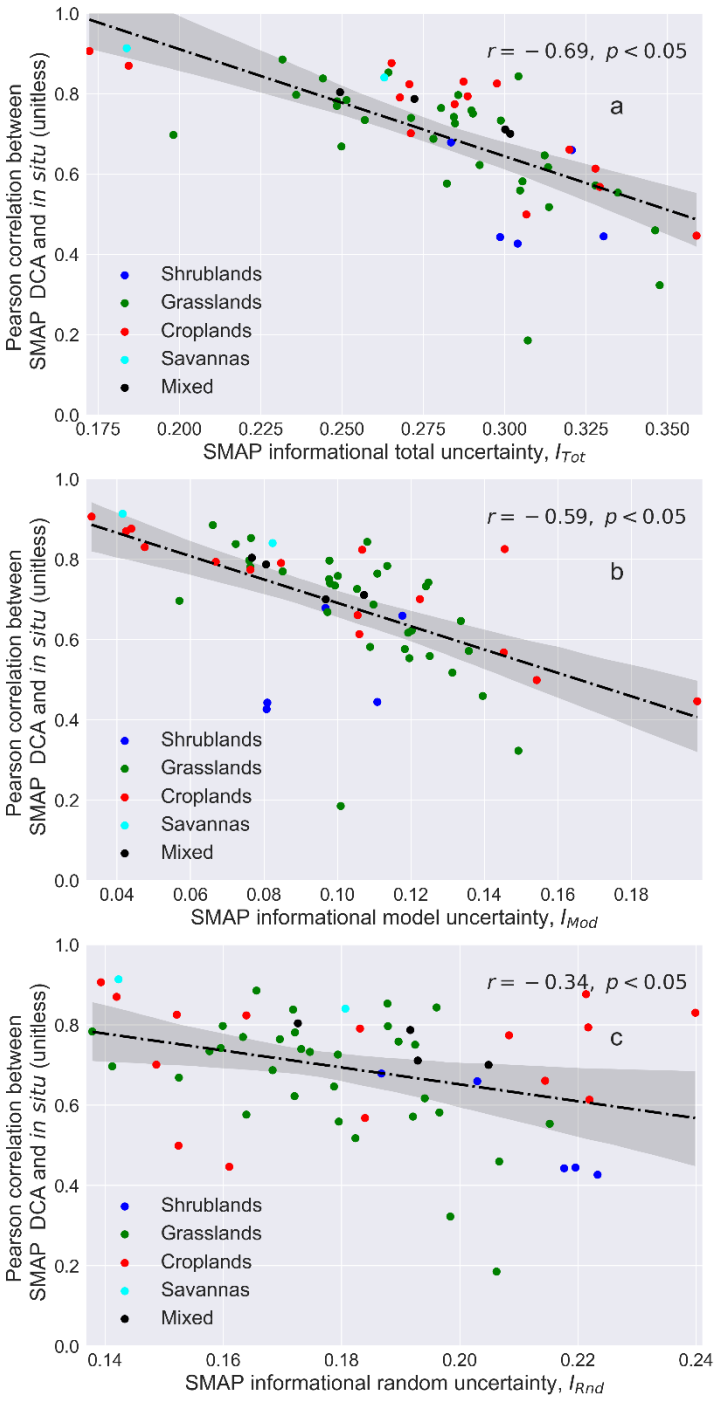

Figure 3.5 SMAP informational total uncertainty (a), SMAP informational model uncertainty (b) and SMAP informational random uncertainty (c) against Pearson correlation between *in situ* soil moisture and DCA soil moisture

3.4.2 Partial Information Decomposition of DCA

The partial information decompositions were assessed on a site basis and are shown in figure 3.6. The fractional contribution of each component to that site's mutual information between brightness temperatures and DCA estimates, $I(T_{Bh},T_{Bv}; DCA)$, was also calculated and are given in Table 3.2. Generally, the majority of $I(T_{Bh},T_{Bv}; DCA)$ is redundantly (R) shared by T_{Bh} and T_{Bv}, which is about 0.08 (58% of $I(T_{Bh},T_{Bv}; DCA)$) on average (Table 3.2). The mean values of unique information of T_{Bh} (U_h) and synergistic information (S) of T_{Bh} and T_{Bv} are 0.024 (18% of $I(T_{Bh},T_{Bv}; DCA)$) and 0.018 (14% of $I(T_{Bh},T_{Bv}; DCA)$), respectively (Table 3.2). Compared to other decomposed information components, U_v is the smallest with its mean being 0.013 (10% of $I(T_{Bh},T_{Bv}; DCA)$). Savannas have the highest absolute and fraction of R (0.101 ,78% of $I(T_{Bh},T_{Bv}; DCA)$) (Table 3.2). In general, the DCA system is mainly dominated by R as indicated by both site wise decomposition and when lumping all datasets together (45% of $I(T_{Bh},T_{Bv}; DCA)$) and S is consistently the lowest (Table 3.2).

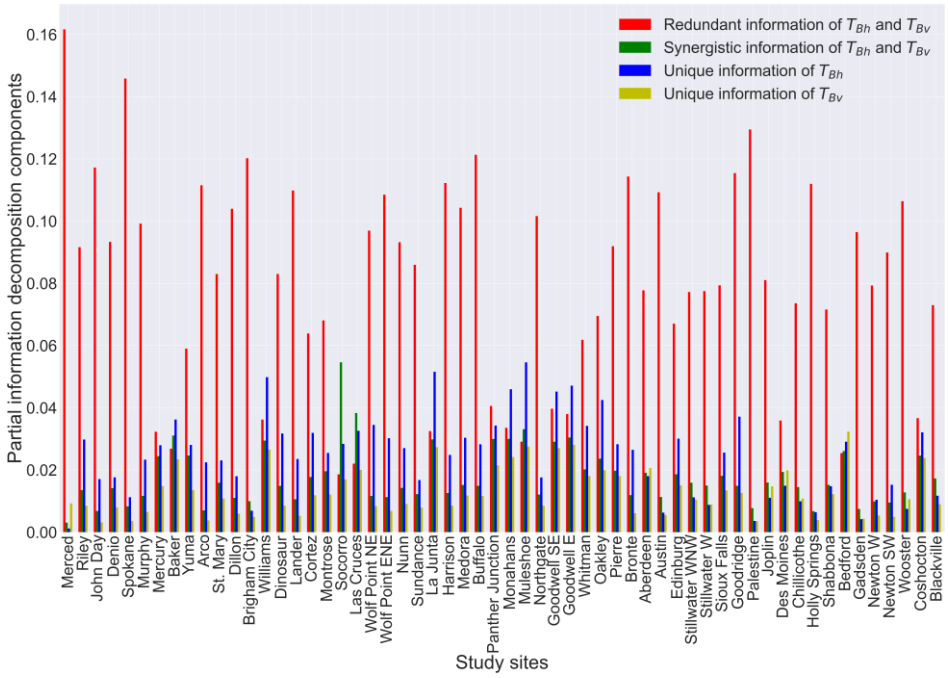

Figure 3.6 Partial information decomposition components between horizontally (T_{Bh}) and vertically (T_{Bv}) polarized brightness temperature and DCA soil moisture. See figure 3.2 caption for site ordering.

Table 3.2 The partial information decomposition components. The values in the table are the average of each landcover. The values in "Overall" are the average of all the sites. The "Lumped" field is computed using all available dataset.

Landcover	Unique information of T_{Bh} (U_h) (and its % I(T_{Bh}, T_{Bv}; DCA))	Unique information of T_{Bv} (U_v) (and its % I(T_{Bh}, T_{Bv}; DCA))	Synergistic information of T_{Bh} and T_{Bv} (S) (and its % I(T_{Bh}, T_{Bv}; DCA))	Redundant information of T_{Bh} and T_{Bv} (R) (and its % I(T_{Bh}, T_{Bv}; DCA))	Mutual information (I(T_{Bh}, T_{Bv}; DCA))	Number of sites
Shrublands	0.034 (28%)	0.019(16%)	0.029 (25%)	0.037 (31%)	0.120	5
Grasslands	0.028 (20%)	0.013 (10%)	0.019 (14%)	0.079 (56%)	0.140	32
Croplands	0.018 (13%)	0.013 (10%)	0.014 (11%)	0.089 (65%)	0.134	15
Savannas	0.008 (7%)	0.006 (5%)	0.012 (10%)	0.101 (78%)	0.128	2
Mixed	0.013(11%)	0.007 (6%)	0.011 (9%)	0.092 (74%)	0.123	4
Lumped	0.014 (19%)	0.019 (25%)	0.008 (11%)	0.034 (45%)	0.076	58
Overall	0.024 (18%)	0.013 (10%)	0.018 (14%)	0.080 (58%)	0.135	58

Through this analysis, it is shown (Figure 3.7) that there are strong relationships between SMAP DCA retrieval quality and decomposed information components. In general, the correlation strength between DCA and *in situ* soil moisture is higher when U_h, U_v and S are low and R is high. This is demonstrated by a significant correlation of these components with the Pearson correlation between *in situ* and DCA soil moisture. The negative relationship between increasing S and decreasing DCA quantity is strongest of the decomposed components, though the positive relationship between increasing R and decreasing DCA is of similar correlation strength. This indicates that R or S contains useful information about DCA soil moisture quality.

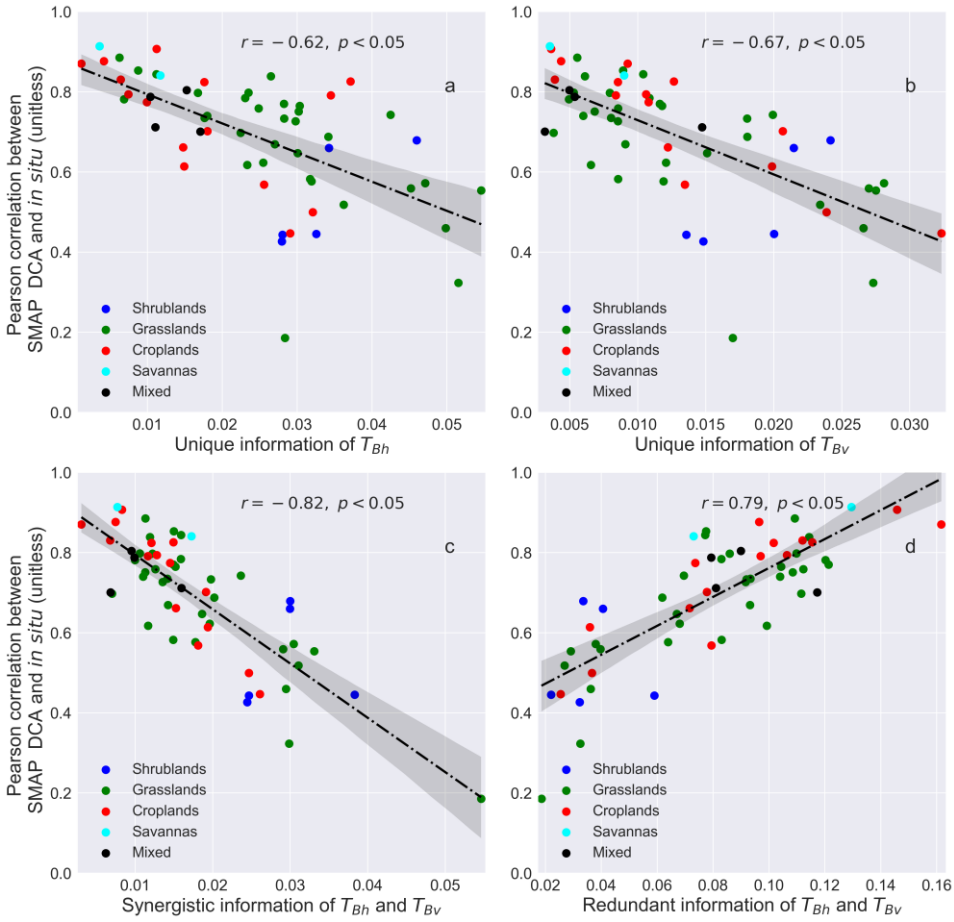

Figure 3.7 Partial information decomposition components between horizontally (T_{Bh}) and vertically (T_{Bv}) polarized brightness temperature against Pearson correlation coefficient between in situ and DCA soil moisture.

3.5 Discussion

3.5.1 DCA Informational Uncertainties

The first objective of this study is to leverage information theory to quantitatively decompose the informational total uncertainty into informational random uncertainty and informational model uncertainty in the DCA as an approach to understand where retrieval uncertainties arise. This information theory approach can provide new insight to SMAP modeling diagnosis. It offers an opportunity of partitioning the total informational uncertainty in the DCA into the uncertainty due to the input datasets and the uncertainty due to model structure

and model parameterizations. This partition process cannot be achieved by leveraging the common DCA assessment metrics (Chan et al., 2016) (e.g., Pearson correlation, ubRMSE) that only involve the DCA soil moisture and *in situ* soil moisture.

The DCA model structure is inherently a hypothesis that relates the input datasets to soil moisture based on prior physical knowledge. The DCA is thus a procedure of processing the input dataset into estimates soil moisture. Thus, models, even those that perform the best, can only reduce the available information in its inputs and are not capable of adding new information about the "true" soil moisture. Hence, there is no possibility of building a model that is better than the one with the best achievable performance of the input data themselves (yet even achieving this theoretically limit is nearly impossible) (Gong et al., 2013). If, however, more freedom on available datasets to incorporate is given, it is possible to build models that outperform the best achievable model performance by adding new explanatory variables which may lead to a family of models that have completely different model structure. Based on Table 3.1, we find that the DCA has more informational uncertainty in shrublands than grasslands and croplands. This might be due to stronger variability in vegetation in for shrublands while grasslands and croplands tend to be more uniform and homogeneous. It is worth noting that these finding are based on averaging our studied sites within different landcover categories, and results may be different while comparing two specific sites from different landcovers. In addition, we find the proportion of informational uncertainty increases as the data is lumped together relative to averaging these statics calculated on a site-by-site basis (Table 3.1). Treating all the surfaces together as a whole does not reduce the informational total uncertainty because the lumping process contains both "high quality" and "low quality" (as assessed by the Pearson correlation between *in situ* and DCA soil moisture) datasets. The uncertainties in these datasets may accumulate while lumping them together and result in an increase in total informational uncertainty.

The fraction that informational random uncertainty contributes to the informational total uncertainty is quite significant (65% on average) in this study. The informational random uncertainty in the system may arises from the inherent error due to calibration of T_{Bh} and T_{Bv} (Al-Yaari et al., 2017), the mismatch in the scale of observations, and the presence water bodies

(Ye et al., 2015). If poorly calibrated, the soil moisture estimations can be exacerbated due to the error propagation that hinders the correct information being expressed. Furthermore, SMAP attempts to T_{eff} to capture both soil and canopy temperature because the differences between canopy and soil temperature are minimized in the morning and afternoon orbits. The T_{eff} is computed based on a model that uses the information from average soil temperature of first layer and second layer and interpolated in time in order to match SMAP observations (O'Neill et al., 2020a). These interpolation and modeling processes may produce erroneous T_{eff} dataset and hence contribute the informational random uncertainty of DCA. Therefore, a better and robust calibration strategy of T_{Bh} and T_{Bv} to the presence of water bodies and a comprehensive assessment of T_{eff} may be needed to reduce some of the information random uncertainty.

Informational model uncertainty contributes an unneglectable portion to the informational total uncertainty (35% on average). This model uncertainty may arise from poor model parameterizations, which may vary with site soil moisture dynamics (H_{CN}(*in situ*)). As shown in figure 3.4b, the I(T_{Bh},T_{Bv}, T_{eff}; *in situ*) increases as the *in situ* soil moisture is more dynamic as reflected by high values of H_{CN}(T_{Bh}) and H_{CN}(T_{Bv}). The raw observations (T_{Bv}, T_{Bh}, and T_{eff}) provide more available information to the system, whereas such information is not properly captured by the algorithm as reflected by low correlation strength between H_{CN}(*in situ*) and I(DCA; *in situ*). Therefore, it is more likely to observe large information model uncertainty where the soil moisture is more dynamic, which may cause a low efficiency of DCA to correctly transmit the available information. It is known that DCA is parameterized with a set of surface and vegetation parameters such as vegetation single scattering albedo (ω), surface height standard deviation (s), etc. These parameter values are landcover dependent and are derived from past studies as well as prior experience and some information discussions with experts, all of which could be biased and inaccurate (O'Neill et al., 2020a). These parameter values also are not differentiated by landcover microwave polarization directions and were assumed to be constant in time. It is possible that these parameters (such as ω) vary in time (Konings et al., 2017) and shift during senescence or harvesting seasons. It is observed that the proportion of the informational model uncertainty is slightly smaller in shrublands (Table 3.1) (here we do not include savannas in the discussion since this landcover only have 2 sites), while these

proportions are larger in croplands and grasslands (Table 3.1). This might because the model parameterizations are more reasonable in shrublands than other landcovers. In addition, croplands and grasslands may have seasonal harvesting and therefore may more subject to changes in these values, while shrublands may not. Additionally, when averaging informational values site-by-site, the informational random uncertainty is a larger fraction of the total uncertainty, whereas when all data are lumped together, the informational model uncertainty is a larger fraction (Table 3.1). DCA parameters are different with respect to each landcover, and the biases induced by these parameters at each site may accumulate through the system resulting a dominance in informational model uncertainty over informational random uncertainty when all sites are lumped together.

To summarize, this is the first attempt of leveraging mutual information approach to analyze the uncertainty components in microwave remote sensing models. The results of this study can be further used as guidance in assessing of SMAP algorithm and can quantitively identify where information lost in the process of SMAP soil moisture modeling. More broadly, this study, though focused on SMAP, can be transferred and extended to analyze other remote sensing algorithms. Over many decades, a lot of effort, resources, and time have been devoted to the launch numerous of satellite missions to retrieve the key environmental variables such as evapotranspiration and vegetation biomass (Dubayah et al., 2020; Hulley et al., 2017). Performing such analysis on these retrieval algorithms is expected to be beneficial to understanding the informational flow in these algorithms and may provide insights to further improve the data retrieval accuracy as well as making maximum use of data collected at greater expense.

3.5.2 Model Evaluation from Another Perspective

The second objective of this study was to demonstrate that the partitioned information components contain useful information about DCA model performance that does not depend on *in situ* soil moisture and other ancillary datasets. We find a strong linear relationship between redundant (R) and synergistic (S) information of the polarized brightness temperatures and Pearson correlation between DCA and *in situ* soil moisture. In general, it is more likely to

observe higher R and lower S (and U_h and U_v) in the less woody landcovers such as croplands and grasslands, where the range of brightness temperature may possibly be greater. These information components were found to be marginally correlated with factors such as vegetation density (the Pearson correlation of average LAI with R, S, U_h, U_v are 0.23, -0.38, -0.54, and -0.19 respectively) and vegetation heterogeneity (the Pearson correlation of LAI standard deviation with R, S, U_h, U_v are 0.22, -0.39, -0.53, and -0.22 respectively). Additionally, these informational components were also found to be correlated with the mutual information shared between brightness temperatures and DCA estimates (the Pearson correlation of $I(T_{Bh},T_{Bv}; \text{DCA})$ with R, S, U_h, U_v are 0.6, -0.27, 0.22, and -0.16 respectively), the informational total uncertainty (the Pearson correlation of I_{Tot} with R, S, U_h, U_v are -0.75, 0.62, 0.55, and 0.68 respectively), informational random uncertainty (the Pearson correlation of I_{Rnd} with R, S, U_h, U_v are -0.41, 0.30, 0.05, and 0.15 respectively), and informational model uncertainty (the Pearson correlation of I_{Mod} with R, S, U_h, U_v are -0.62, 0.56, 0.66, and 0.74 respectively). This indicates that these informational components in the DCA system are not only physically driven by both vegetation density and heterogeneity but also other factors such as how algorithm processes the information from T_{Bh} and T_{Bv} to produce the DCA outputs. It is more likely to observe higher R and lower S in locations where vegetation is denser and more heterogeneous, yet the correlation of these variables with model quality (0.47 for mean LAI and 0.42 for the standard deviation of LAI) are weaker than the correlations found between R and S and model quality shown in Figure 7. The R and S metric in this study can thus not only integrate information about how the surface vegetation density and heterogeneity influence the algorithm performance but provided insight into how effectively DCA algorithm uses the information from T_{Bh} and T_{Bv}.

Compared with other ancillary and *in situ* independent metrics such as correlation strength between Pearson correlation of T_{Bh} with T_{Bv} and the Pearson correlation between *in situ* and DCA soil moisture (0.67), the correlation strength of S and R with Pearson correlation of *in situ* and DCA soil moisture are tighter (0.79 and -0.82 for R and S). This suggests the complex non-linear relationship between of T_{Bh}, T_{Bv} with DCA soil moisture is better captured by R and S as compared to the direct correlation between the two brightness temperatures themselves. Given the strength of this relationship, the R and S holds the potential to be used as a DCA evaluation

metric that does not depend on *in situ* measurement and ancillary dataset. It is also useful for SMAP DCA soil moisture users to have a rough estimation of how high the quality (as characterized as the correlation strength between DCA and *in situ*) of the obtained DCA soil moisture without actually knowing the *in situ* soil moisture. However, this depends on specific user requirements for data quality. In general, the DCA soil moisture tends to be in high end in term retrieval quality (~ 0.75 in Pearson correlation) when the R is greater 0.1 or S is smaller than 0.015. It is important to note that the decomposed information components are dependent on the DCA parameterizations (e.g., ω, h. etc.) that may influence how the T_{Bh} and T_{Bv} are probabilistically linked with the DCA and hence may alter the partitioned information components.

3.5.3 Approach Limitations

While we expect that this approach can be generalized to analyze other remote sensing models, it may be difficult to compute the joint probability density functions for models with high-dimensional inputs. Difficulty in determining the joint probability density functions hinders the estimation of high dimensional joint entropy and mutual information components, and these are still open questions in the field of information theory. Although there exist serval data dimension reduction techniques, these dimension reduction techniques are mostly based some assumptions (Xu et al., 2019). In practice, most of the systems with high dimension inputs tend to be complex. Therefore, there is a strong risk of introducing additional uncertainty if one chooses an inappropriate technique.

It is important to understand that SMAP DCA system retrieves soil moisture with the help of vegetation water content climatology derived from the MODIS NDVI data stream. This is specified as a set value for each location and day of year combination and is used to estimate the initial guess for the unknown vegetation optical depth (O'Neill et. al., 2020a). The reader should keep in mind that this study considers such data as a dynamic time-varying parameter and it is not treated as a data input in this study. Adding NDVI as a data input would result in $I(T_{Bh}, T_{Bv,} T_{eff}, \text{NDVI}; \textit{in situ})$ being larger than or equal to $I(T_{Bh}, T_{Bv,} T_{eff}; \textit{in situ})$ in the calculation of I_{Rnd}, and therefore I_{Rnd} would decrease. Since, I_{Tot} only considers DCA output and *in situ* data it is not

altered by adding dynamic parameters and I_{Mod} would therefore increase. Thus, consideration of additional dynamic parameters in this informational assessment would serve to shift uncertainties from those attributed to the input data themselves to uncertainties attributed to the model structure and parameterizations.

This study was conducted only at locations where *in situ* soil moisture is readily available. It could be an interesting topic to explore if, and how, information-based uncertainty analysis can be applied in the locations without *in situ* soil moisture measurements. We would expect the informational uncertainty analysis to provide the estimates of random and model uncertainties. The best performance we can expect from this current uncertainty analysis is to use all the available datasets we have; yet we believe that uncertainty estimations of this approach should be stabilized given adequate representative locations and data records.

3. 6 Conclusions

This study differentiates and quantifies the uncertainty sources in the SMAP DCA using information theory. We found that on average DCA soil moisture explains 20% of the information in the *in situ* soil moisture leaving 80% unexplained. Among the unexplained information, 65% is informational random uncertainty that is caused by the inherent stochasticity of the explanatory variables of SMAP DCA and a lack of additional explanatory variables in the system, while the rest of the informational uncertainty is caused by inappropriateness of the assumption of DCA model structure and parameterizations. We show that informational random uncertainty contributes a larger proportion of the informational total uncertainty across different landcovers. However, the informational model uncertainty contributes more to total uncertainty when lumping all the datasets together. The performance of SMAP DCA is negatively correlated to all the information uncertainties, with the informational model uncertainty being more reflective of overall SMAP DCA retrieval quality than the informational random uncertainty.

The decomposition of the mutual information has shown that all decomposed components are significantly related to the Pearson correlation between *in situ* and DCA soil moisture, with the redundant and synergistic information being the strongest. Good DCA model performance (as measured by Pearson correlation between *in situ* and DCA soil moisture) is more likely to be

found in locations where the redundant information of brightness temperatures shared with DCA soil moisture is high and is more dominant relative to other components. The informational uncertainty decomposition analysis opens a new window for SMAP algorithm uncertainty diagnosis. SMAP DCA users may examine to the *R* and *S* components to have an approximate estimation of the soil moisture data quality obtained when no *in situ* soil moisture is readily available.

3.7 References

Al-Yaari, A., Wigneron, J.-P., Kerr, Y., Rodriguez-Fernandez, N., O'Neill, P. E., Jackson, T. J., De Lannoy, G. J. M., Al Bitar, A., Mialon, A., Richaume, P., Walker, J. P., Mahmoodi, A., & Yueh, S. (2017). Evaluating soil moisture retrievals from ESA's SMOS and NASA's SMAP brightness temperature datasets. *Remote Sensing of Environment*, *193*, 257–273. https://doi.org/10.1016/j.rse.2017.03.010

Babaeian, E., Sadeghi, M., Jones, S. B., Montzka, C., Vereecken, H., & Tuller, M. (2019). Ground, Proximal, and Satellite Remote Sensing of Soil Moisture. *Reviews of Geophysics*, *57*(2), 530–616. https://doi.org/10.1029/2018RG000618

Bassiouni, M., Good, S. P., Still, C. J., & Higgins, C. W. (2020). Plant Water Uptake Thresholds Inferred From Satellite Soil Moisture. *Geophysical Research Letters*, *47*(7). https://doi.org/10.1029/2020GL087077

Bell, J. E., Leeper, R. D., Palecki, M. A., Coopersmith, E., Wilson, T., Bilotta, R., & Embler, S. (2015). Evaluation of the 2012 Drought with a Newly Established National Soil Monitoring Network. *Vadose Zone Journal*, *14*(11), vzj2015.02.0023. https://doi.org/10.2136/vzj2015.02.0023

Bell, J. E., Palecki, M. A., Baker, C. B., Collins, W. G., Lawrimore, J. H., Leeper, R. D., Hall, M. E., Kochendorfer, J., Meyers, T. P., Wilson, T., & Diamond, H. J. (2013). U.S. Climate Reference Network Soil Moisture and Temperature Observations. *Journal of Hydrometeorology*, *14*(3), 977–988. https://doi.org/10.1175/JHM-D-12-0146.1

Box, G. E. P. (1976). Science and Statistics. *Journal of the American Statistical Association*, *71*(356), 791–799. https://doi.org/10.1080/01621459.1976.10480949

Chan, S. (2020). Soil Moisture Active Passive (SMAP) Level 2 Passive Soil Moisture Product Specification Documen. *Jet Propulsion Laboratory California Institute of Technology, Pasadena, USA, JPL D-72547 (Version 7.0)*.

Chan, S. K., Bindlish, R., O'Neill, P. E., Njoku, E., Jackson, T., Colliander, A., Chen, F., Burgin, M., Dunbar, S., Piepmeier, J., Yueh, S., Entekhabi, D., Cosh, M. H., Caldwell, T., Walker, J., Wu, X., Berg, A., Rowlandson, T., Pacheco, A., … Kerr, Y. (2016). Assessment of the SMAP Passive Soil Moisture Product. *IEEE Transactions on Geoscience and Remote Sensing*, *54*(8), 4994–5007. https://doi.org/10.1109/TGRS.2016.2561938

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., Berg, A. A., Bosch, D. D., Caldwell, T., Cosh, M. H., Holifield Collins, C., Martinez-Fernandez, J., Seyfried, M., Starks, P. J., Su, Z., … Walker, J. (2020). Improved SMAP Dual-Channel Algorithm for the Retrieval of Soil Moisture. *IEEE Transactions on Geoscience and Remote Sensing*, *58*(6), 3894–3905. https://doi.org/10.1109/TGRS.2019.2959239

Chen, F., Crow, W. T., Colliander, A., Cosh, M. H., Jackson, T. J., Bindlish, R., Reichle, R. H., Chan, S. K., Bosch, D. D., Starks, P. J., Goodrich, D. C., & Seyfried, M. S. (2017). Application of Triple Collocation in Ground-Based Validation of Soil Moisture Active/Passive (SMAP) Level 2 Data Products. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *10*(2), 489–502. https://doi.org/10.1109/JSTARS.2016.2569998

Colliander, A., Jackson, T. J., Bindlish, R., Chan, S., Das, N., Kim, S. B., Cosh, M. H., Dunbar, R. S., Dang, L., Pashaian, L., Asanuma, J., Aida, K., Berg, A., Rowlandson, T., Bosch, D., Caldwell, T., Caylor, K., Goodrich, D., al Jassar, H., … Yueh, S. (2017a). Validation of SMAP surface soil moisture products with core validation sites. *Remote Sensing of Environment*, *191*, 215–231. https://doi.org/10.1016/j.rse.2017.01.021

Colliander, A., Jackson, T. J., Bindlish, R., Chan, S., Das, N., Kim, S. B., Cosh, M. H., Dunbar, R. S., Dang, L., Pashaian, L., Asanuma, J., Aida, K., Berg, A., Rowlandson, T., Bosch, D., Caldwell, T., Caylor, K., Goodrich, D., al Jassar, H., … Yueh, S. (2017b). Validation of SMAP surface soil moisture products with core validation sites. *Remote Sensing of Environment*, *191*, 215–231. https://doi.org/10.1016/j.rse.2017.01.021

Cover, T. M., & Thomas, J. A. (2005). *Elements of Information Theory*. Wiley. https://doi.org/10.1002/047174882X

Crow, W. T., Berg, A. A., Cosh, M. H., Loew, A., Mohanty, B. P., Panciera, R., de Rosnay, P., Ryu, D., & Walker, J. P. (2012). Upscaling sparse ground-based soil moisture observations for the validation of coarse-resolution satellite soil moisture products. *Reviews of Geophysics*, *50*(2). https://doi.org/10.1029/2011RG000372

Dadap, N. C., Cobb, A. R., Hoyt, A. M., Harvey, C. F., & Konings, A. G. (2019). Satellite soil moisture observations predict burned area in Southeast Asian peatlands. *Environmental Research Letters*, *14*(9), 094014. https://doi.org/10.1088/1748-9326/ab3891

Diamond, H. J., Karl, T. R., Palecki, M. A., Baker, C. B., Bell, J. E., Leeper, R. D., Easterling, D. R., Lawrimore, J. H., Meyers, T. P., Helfert, M. R., Goodge, G., & Thorne, P. W. (2013). U.S. Climate Reference Network after One Decade of Operations: Status and Assessment. *Bulletin of the American Meteorological Society*, *94*(4), 485–498. https://doi.org/10.1175/BAMS-D-12-00170.1

Dubayah, R., Blair, J. B., Goetz, S., Fatoyinbo, L., Hansen, M., Healey, S., Hofton, M., Hurtt, G., Kellner, J., Luthcke, S., Armston, J., Tang, H., Duncanson, L., Hancock, S., Jantz, P., Marselis, S., Patterson, P. L., Qi, W., & Silva, C. (2020). The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth's forests and topography. *Science of Remote Sensing*, *1*, 100002. https://doi.org/10.1016/j.srs.2020.100002

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., Kimball, J., Piepmeier, J. R., Koster, R. D., Martin, N., McDonald, K. C., Moghaddam, M., Moran, S., Reichle, R., Shi, J. C., … Van Zyl, J. (2010). The Soil Moisture Active Passive (SMAP) Mission. *Proceedings of the IEEE*, *98*(5), 704–716. https://doi.org/10.1109/JPROC.2010.2043918

Escorihuela, M. J., Chanzy, A., Wigneron, J. P., & Kerr, Y. H. (2010). Effective soil moisture sampling depth of L-band radiometry: A case study. *Remote Sensing of Environment*, *114*(5), 995–1001. https://doi.org/10.1016/j.rse.2009.12.011

Feldman, A. F., Short Gianotti, D. J., Konings, A. G., McColl, K. A., Akbar, R., Salvucci, G. D., & Entekhabi, D. (2018). Moisture pulse-reserve in the soil-plant continuum observed across biomes. *Nature Plants*, *4*(12), 1026–1033. https://doi.org/10.1038/s41477-018-0304-9

Finn, C., & Lizier, J. (2018). Pointwise Partial Information Decomposition Using the Specificity and Ambiguity Lattices. *Entropy*, *20*(4), 297. https://doi.org/10.3390/e20040297

Freedman, D., & Diaconis, P. (1981). On the histogram as a density estimator:L 2 theory. *Zeitschrift Für Wahrscheinlichkeitstheorie Und Verwandte Gebieter*, *57*(4), 453–476. https://doi.org/10.1007/BF01025868

Gao, L., Sadeghi, M., Ebtehaj, A., & Wigneron, J.-P. (2020). A temporal polarization ratio algorithm for calibration-free retrieval of soil moisture at L-band. *Remote Sensing of Environment*, *249*, 112019. https://doi.org/10.1016/j.rse.2020.112019

Gong, W., Gupta, H. V., Yang, D., Sricharan, K., & Hero, A. O. (2013). Estimating epistemic and aleatory uncertainties during hydrologic modeling: An information theoretic approach. *Water Resources Research*, *49*(4), 2253–2273. https://doi.org/10.1002/wrcr.20161

Goodwell, A. E., & Kumar, P. (2017). Temporal information partitioning: Characterizing synergy, uniqueness, and redundancy in interacting environmental variables. *Water Resources Research*, *53*(7), 5920–5942. https://doi.org/10.1002/2016WR020216

Goodwell, A. E., Kumar, P., Fellows, A. W., & Flerchinger, G. N. (2018). Dynamic process connectivity explains ecohydrologic responses to rainfall pulses and drought. *Proceedings of the National Academy of Sciences*, *115*(37), E8604–E8613. https://doi.org/10.1073/pnas.1800236115

Gruber, A., De Lannoy, G., Albergel, C., Al-Yaari, A., Brocca, L., Calvet, J.-C., Colliander, A., Cosh, M., Crow, W., Dorigo, W., Draper, C., Hirschi, M., Kerr, Y., Konings, A., Lahoz, W., McColl, K., Montzka, C., Muñoz-Sabater, J., Peng, J., … Wagner, W. (2020). Validation practices for satellite soil moisture retrievals: What are (the) errors? *Remote Sensing of Environment*, *244*, 111806. https://doi.org/10.1016/j.rse.2020.111806

Hulley, G., Hook, S., Fisher, J., & Lee, C. (2017). ECOSTRESS, A NASA Earth-Ventures Instrument for studying links between the water cycle and plant health over the diurnal cycle. *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 5494–5496. https://doi.org/10.1109/IGARSS.2017.8128248

Jackson, T. J., Bindlish, R., Cosh, M. H., Zhao, T., Starks, P. J., Bosch, D. D., Seyfried, M., Moran, M. S., Goodrich, D. C., Kerr, Y. H., & Leroux, D. (2012). Validation of Soil Moisture and Ocean Salinity (SMOS) Soil Moisture Over Watershed Networks in the U.S. *IEEE Transactions on Geoscience and Remote Sensing*, *50*(5), 1530–1543. https://doi.org/10.1109/TGRS.2011.2168533

Jackson, T. J., Schmugge, T. J., & Wang, J. R. (1982). Passive microwave sensing of soil moisture under vegetation canopies. *Water Resources Research*, *18*(4), 1137–1142. https://doi.org/10.1029/WR018i004p01137

Kerr, Y. H., Waldteufel, P., Richaume, P., Wigneron, J. P., Ferrazzoli, P., Mahmoodi, A., Al Bitar, A., Cabot, F., Gruhier, C., Juglea, S. E., Leroux, D., Mialon, A., & Delwart, S. (2012). The SMOS Soil Moisture Retrieval Algorithm. *IEEE Transactions on Geoscience and Remote Sensing*, *50*(5), 1384–1403. https://doi.org/10.1109/TGRS.2012.2184548

Konings, A. G., McColl, K. A., Piles, M., & Entekhabi, D. (2015). How Many Parameters Can Be Maximally Estimated From a Set of Measurements? *IEEE Geoscience and Remote Sensing Letters*, *12*(5), 1081–1085. https://doi.org/10.1109/LGRS.2014.2381641

Konings, A. G., Piles, M., Das, N., & Entekhabi, D. (2017). L-band vegetation optical depth and effective scattering albedo estimation from SMAP. *Remote Sensing of Environment*, *198*, 460–470. https://doi.org/10.1016/j.rse.2017.06.037

Konings, A. G., Piles, M., Rötzer, K., McColl, K. A., Chan, S. K., & Entekhabi, D. (2016). Vegetation optical depth and scattering albedo retrieval using time series of dual-polarized L-band radiometer observations. *Remote Sensing of Environment*, *172*, 178–189. https://doi.org/10.1016/j.rse.2015.11.009

Kunert-Graf, J., Sakhanenko, N., & Galas, D. (2020). Partial Information Decomposition and the Information Delta: A Geometric Unification Disentangling Non-Pairwise Information. *Entropy*, *22*(12), 1333. https://doi.org/10.3390/e22121333

Leeper, R. D., Bell, J. E., Vines, C., & Palecki, M. (2017). An Evaluation of the North American Regional Reanalysis Simulated Soil Moisture Conditions during the 2011–13 Drought Period. *Journal of Hydrometeorology*, *18*(2), 515–527. https://doi.org/10.1175/JHM-D-16-0132.1

Mo, T., Choudhury, B. J., Schmugge, T. J., Wang, J. R., & Jackson, T. J. (1982). A model for microwave emission from vegetation-covered fields. *Journal of Geophysical Research*, *87*(C13), 11229. https://doi.org/10.1029/JC087iC13p11229

Mohanty, B. P., Cosh, M. H., Lakshmi, V., & Montzka, C. (2017). Soil Moisture Remote Sensing: State-of-the-Science. *Vadose Zone Journal*, *16*(1), vzj2016.10.0105. https://doi.org/10.2136/vzj2016.10.0105

Myneni, R., Knyazikhin, Y., & Park, T. (2015). MCD15A3H MODIS/Terra+Aqua Leaf Area Index/FPAR 4-day L4 Global 500m SIN Grid V006. *NASA EOSDIS Land Processes DAAC*. https://doi.org/https://doi.org/10.5067/MODIS/MCD15A3H.006

Njoku, E. G., & Entekhabi, D. (1996). Passive microwave remote sensing of soil moisture. *Journal of Hydrology*, *184*(1–2), 101–129. https://doi.org/10.1016/0022-1694(95)02970-2

O'Neill, P., Bindlish, R., Chan, S., Njoku, E., and Jackson, T. (2020a). Algorithm theoretical basis document: Level 2 & 3 soil moisture (passive) data products. *Level 2 & 3 Soil Moisture (Passive) Data Products, Jet Propulsion Laboratory/California Institute of Technology, Pasadena, USA, JPL D-66480 (Revision F)*, 100.

O'Neill, P. E., Chan, S. K., Njoku, E. G., Tom, J., Bindlish, R., & Chaubell, M. J. (2020b). SMAP L2 Radiometer Half-Orbit 36 km EASE-Grid Soil Moisture, Version 7. *[March 31 2015 to December 2020],Boulder, Colorado. USA. NASA National. Snow Ice Data Cent. Distrib. Act. Arch. Center [Date Accessed December 10 2020]*. https://doi.org/https://doi.org/10.5067/F1TZ0CBN1F5N

ORNL DAAC. (2018). MODIS and VIIRS Land Products Global Subsetting and Visualization Tool. *Available Online: Https://Modis.Ornl.Gov/Globalsubset/ (Accessed on 24 February 2020)*. https://doi.org/https://doi.org/10.3334/ORNLDAAC/1379

Paninski, L. (2003). Estimation of Entropy and Mutual Information. *Neural Computation*, *15*(6), 1191–1253. https://doi.org/10.1162/089976603321780272

Petropoulos, G. P., Ireland, G., & Barrett, B. (2015). Surface soil moisture retrievals from remote sensing: Current status, products & future trends. *Physics and Chemistry of the Earth, Parts A/B/C*, *83–84*, 36–56. https://doi.org/10.1016/j.pce.2015.02.009

Raju, S., Chanzy, A., Wigneron, J.-P., Calvet, J.-C., Kerr, Y., & Laguerre, L. (1995). Soil moisture and temperature profile effects on microwave emission at low frequencies. *Remote Sensing of Environment*, *54*(2), 85–97. https://doi.org/10.1016/0034-4257(95)00133-L

Seitzinger, S. P., Gaffney, O., Brasseur, G., Broadgate, W., Ciais, P., Claussen, M., Erisman, J. W., Kiefer, T., Lancelot, C., Monks, P. S., Smyth, K., Syvitski, J., & Uematsu, M. (2015). International Geosphere–Biosphere Programme and Earth system science: Three decades of co-evolution. *Anthropocene*, *12*, 3–16. https://doi.org/10.1016/j.ancene.2016.01.001

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Shellito, P. J., Small, E. E., Colliander, A., Bindlish, R., Cosh, M. H., Berg, A. A., Bosch, D. D., Caldwell, T. G., Goodrich, D. C., McNairn, H., Prueger, J. H., Starks, P. J., van der Velde, R., & Walker, J. P. (2016). SMAP soil moisture drying more rapid than observed in situ following rainfall events. *Geophysical Research Letters*, *43*(15), 8068–8075. https://doi.org/10.1002/2016GL069946

Uber, M., Vandervaere, J.-P., Zin, I., Braud, I., Heistermann, M., Legoût, C., Molinié, G., & Nord, G. (2018). How does initial soil moisture influence the hydrological response? A case study from southern France. *Hydrology and Earth System Sciences*, *22*(12), 6127–6146. https://doi.org/10.5194/hess-22-6127-2018

Uda, S. (2020). Application of information theory in systems biology. *Biophysical Reviews*, *12*(2), 377–384. https://doi.org/10.1007/s12551-020-00665-w

Wang, L., & Qu, J. J. (2009). Satellite remote sensing applications for surface soil moisture monitoring: A review. *Frontiers of Earth Science in China*, *3*(2), 237–247. https://doi.org/10.1007/s11707-009-0023-7

Wibral, M., Priesemann, V., Kay, J. W., Lizier, J. T., & Phillips, W. A. (2017). Partial information decomposition as a unified approach to the specification of neural goal functions. *Brain and Cognition*, *112*, 25–38. https://doi.org/10.1016/j.bandc.2015.09.004

Wigneron, J.-P., Jackson, T. J., O'Neill, P., De Lannoy, G., de Rosnay, P., Walker, J. P., Ferrazzoli, P., Mironov, V., Bircher, S., Grant, J. P., Kurum, M., Schwank, M., Munoz-Sabater, J., Das, N., Royer, A., Al-Yaari, A., Al Bitar, A., Fernandez-Moran, R., Lawrence, H., … Kerr, Y. (2017). 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 Sensing of Environment*, *192*, 238–262. https://doi.org/10.1016/j.rse.2017.01.024

Williams, P. L., & Beer, R. D. (2010). *Nonnegative Decomposition of Multivariate Information*. http://arxiv.org/abs/1004.2515

Xu, X., Liang, T., Zhu, J., Zheng, D., & Sun, T. (2019). Review of classical dimensionality reduction and sample selection methods for large-scale data processing. *Neurocomputing*, *328*, 5–15. https://doi.org/10.1016/j.neucom.2018.02.100

Ye, N., Walker, J. P., Guerschman, J., Ryu, D., & Gurney, R. J. (2015). Standing water effect on soil moisture retrieval from L-band passive microwave observations. *Remote Sensing of Environment*, *169*, 232–242. https://doi.org/10.1016/j.rse.2015.08.013

Zhang, R., Kim, S., & Sharma, A. (2019). A comprehensive validation of the SMAP Enhanced Level-3 Soil Moisture product using ground measurements over varied climates and landscapes. *Remote Sensing of Environment*, *223*, 82–94. https://doi.org/10.1016/j.rse.2019.01.015

Zhang, Z., & Grabchak, M. (2013). Bias Adjustment for a Nonparametric Entropy Estimator. *Entropy*, *15*(12), 1999–2011. https://doi.org/10.3390/e15061999

3.8 Supplementary Materials

Table 3.S1 The USCRN study site information

USCRN sites	Lat.	Lon.	USCRN sites	Lat.	Lon.
AL_Gadsden_19_N	34.29	-85.96	CO_Montrose_11_ENE	38.54	-107.69
GA_Newton_11_SW	31.19	-84.45	CO_Nunn_7_NNE	40.81	-104.76
GA_Newton_8_W	31.31	-84.47	KS_Oakley_19_SSW	38.87	-100.96
IA_Des_Moines_17_E	41.56	-93.29	MT_Wolf_Point_29_ENE	48.31	-105.1
IL_Shabbona_5_NNE	41.84	-88.85	MT_Wolf_Point_34_NE	48.49	-105.21
IN_Bedford_5_WNW	38.89	-86.57	NE_Harrison_20_SSE	42.42	-103.74
MN_Goodridge_12_NNW	48.31	-95.87	NE_Whitman_5_ENE	42.07	-101.44
MO_Chillicothe_22_ENE	39.87	-93.15	NM_Las_Cruces_20_N	32.61	-106.74
MO_Joplin_24_N	37.43	-94.58	NM_Socorro_20_N	34.36	-106.89
MS_Holly_Springs_4_N	34.82	-89.43	OK_Goodwell_2_E	36.6	-101.59
ND_Northgate_5_ESE	48.97	-102.17	OK_Goodwell_2_SE	36.57	-101.61
SD_Aberdeen_35_WNW	45.71	-99.13	OK_Stillwater_2_W	36.12	-97.09
SD_Sioux_Falls_14_NNE	43.73	-96.62	OK_Stillwater_5_WNW	36.13	-97.11
AZ_Williams_35_NNW	35.76	-112.34	SD_Buffalo_13_ESE	45.52	-103.3
AZ_Yuma_27_ENE	32.84	-114.19	SD_Pierre_24_S	44.02	-100.35
CA_Merced_23_WSW	37.24	-120.88	TX_Bronte_11_NNE	32.04	-100.25
ID_Arco_17_SW	43.46	-113.56	TX_Monahans_6_ENE	31.62	-102.81
ID_Murphy_10_W	43.2	-116.75	TX_Muleshoe_19_S	33.96	-102.77
NV_Baker_5_W	39.01	-114.21	TX_Panther_Junction_2_N	29.35	-103.21
NV_Denio_52_WSW	41.85	-119.64	WY_Sundance_8_NNW	44.52	-104.44
NV_Mercury_3_SSW	36.62	-116.02	TX_Austin_33_NW	30.62	-98.08
OR_John_Day_35_WNW	44.56	-119.65	TX_Edinburg_17_NNE	26.53	-98.06
OR_Riley_10_WSW	43.47	-119.69	TX_Palestine_6_WNW	31.78	-95.72
UT_Brigham_City_28_WNW	41.62	-112.54	CO_Cortez_8_SE	37.26	-108.5
WA_Spokane_17_SSW	47.42	-117.53	CO_Dinosaur_2_E	40.24	-108.97
OH_Coshocton_8_NNE	40.37	-81.78	WY_Lander_11_SSE	42.68	-108.67
OH_Wooster_3_SSE	40.76	-81.91	ND_Medora_7_E	46.89	-103.38
SC_Blackville_3_W	33.36	-81.33	MT_Dillon_18_WSW	45.16	-113.01
CO_La_Junta_17_WSW	37.86	-103.82	MT_St._Mary_1_SSW	48.74	-113.43

Chapter 4. Stable Isotopes Contain Substantial Additive Information about Carbon and Water Cycles in Arid Environment

Bonan Li[1,2] Stephen P. Good[1,2], Catherine Finkenbiner[1,2], Richard Fiorella[3,4], Gabriel Bowen[3,4], David Noone[5], Christopher Still[6], William Anderegg[4,7]

[1]Departement of Biological & Ecological Engineering, Oregon State University,

[2]Water Resources Graduate Program, Oregon State University

[3]Department of Geology and Geophysics, University of Utah

[4]Global Change and Sustainability Center, University of Utah

[5]Department of Physics, University of Auckland

[6]Department of Forest Ecosystems and Society, Oregon State University

[7]School of Biological Sciences, University of Utah

Status: In preparation to submit to *Proceedings of the National Academy of Sciences* of the United States of America

4.1 Abstract

Water and carbon exchanges between the land and atmosphere reflect key ecosystem processes, from global climate change to local biogeochemistry. Environmental stable isotope ratios of H_2O and CO_2 fluxes have been used to study these processes, yet measurement constraints have limited macroscale surface-atmosphere isotope flux evaluations. Across North American biomes within the US National Ecological Observation Network, we assessed how much information stable isotope measurements ($\delta^{13}C$, $\delta^{2}H$, and deuterium excess [*d-ex*]) contain about latent heat (*LH*) and net ecosystem exchange (*NEE*) fluxes. Overall, $\delta^{13}C$, $\delta^{2}H$, and *d-ex* contain non-trivial information about the *LH* and *NEE* fluxes, with isotope measurements carrying about the same amount of information as can be obtained by wind speed measurements, but less than carried by vapor pressure deficit, air temperature, and net radiation measurements. The decomposition of multivariate mutual information about bulk fluxes from the isotopes and the individual meteorological variables suggested that $\delta^{13}C$, $\delta^{2}H$, and *d-ex* contain significant unique information about the carbon and water surface-atmosphere exchange. The sum of the unique and synergistic information, which represents the total additional information from isotopes, was shown to be significantly driven by site-specific factors for *LH*, but not *NEE*. Generally, $\delta^{13}C$ provides more additional information about *LH* in arid locations, while the $\delta^{2}H$ provides more additional information about *LH* at locations with higher aridity, lower mean annual temperature, and lower mean site elevation. This study demonstrates that stable isotope measurements, particularly in arid environments, can provide new information to help understand carbon and water cycles.

4.2 Significance statement

Under global climate change, it is crucial to know how water and carbon are exchanged between the terrestrial and atmospheric environments. Stable isotopes are useful as they can be adopted as tracers to understand water and carbon flues. Here, we show that isotopes contain useful information about these fluxers, roughly as much as wind speed. We show that this information is largely new information that cannot be provided by other variables. It is more

likely for δ^2H to provide additional information under drier conditions. Integrate isotope datasets into ecohydrological models may help to improve the model performance. Researchers should expect a better improvement in model performance under drier conditions.

4.3 Introduction

The terrestrial carbon and water cycles are strongly coupled and are jointly the most important components to support life on earth (Gentine et al., 2019). The partitioning of water and carbon fluxes into their respective constituents enhance our ability to understand, model and differentiate abiotic and biotic ecosystem processes that are driven by distinct environmental forcings (Scanlon et al., 2019; Scott & Biederman, 2017). Understanding the interaction and feedbacks of exchanging carbon and water between the terrestrial ecosystems and the atmosphere is crucial to understand earth's paleoclimate and forecast the future climate change (Ferguson & Veizer, 2007; Keenan & Williams, 2018). Accordingly, over the past decades, significant efforts have been made to measure and monitor carbon and water fluxes and the eddy covariance (EC) technique has become the most widely adopted approach in the earth science community to measure water and carbon fluxes at ecosystem levels (Baldocchi, 2014; Foken et al., 2012). EC flux towers can provide continuous net ecosystem exchange (*NEE*) of CO_2 between land surface and the atmosphere at various frequency domains as an indicator of terrestrial carbon balance. Besides *NEE*, latent heat flux (*LH*), representing evaporation and transpiration from soils, water bodies, and plant canopies, is also measured by EC flux towers and is of great value for understanding of regional and global water budgets as well as agricultural applications (Trenberth et al., 2007; Zhou et al., 2018). These EC flux measurements have been used for a variety of environmental applications such as calibrating and validating of remotely sensed flux estimations (Jia et al., 2012), parameterizing land surface models (Williams et al., 2009), modeling seasonal crop coefficients (Li et al., 2008), and investigating post-fire carbon balance (Lupascu et al., 2020).While measurements of *LH* and *NEE* can quantify flux amounts, better knowledge of the processes that drive carbon and water cycles themselves is needed to improve models and overall understanding of the earth's ecosystem as a whole.

The naturally occurring water and carbon isotopes are powerful tools for understanding a wide array of ecohydrological and biophysical problems. Regional weather characteristics can be gleaned from the water isotopes in precipitation (Scholl & Murphy, 2014). Past climate patterns can be reconstructed based on the relationship between stable isotopes in precipitation and temperature (ROZANSKI et al., 1997). Stable isotope signatures of groundwater are mixtures of different water sources, and these sources can be disentangled based on isotope composition (Kuang et al., 2019). More importantly, water isotopes (δ^2H, $\delta^{18}O$) of *LH* are widely used approaches to partition *ET* to evaporation and transpiration because heavier isotope are more enriched in evaporated water vapor due to the significant fractionations of soil water in contrast to the isotope composition of transpiration which remains similar to plant water (Xiao et al., 2018). Carbon isotope CO_2 ($\delta^{13}C$) is useful to infer the mechanism of changes in atmosphere CO_2 during the last deglaciation (Bauska et al., 2016). It has also been applied to separate *NEE* into its constituent fluxes, as photosynthesis creates a unique pathway to preferentially fix CO_2 as plant biomass and alters the isotope composition of atmosphere CO_2 (Lee et al., 2020; ZHANG et al., 2006). Because of measurement difficulties and costs, previous studies of δ^2H, $\delta^{18}O$ and $\delta^{13}C$ examined patterns across ecosystems using cryogenic baths and flask samples. However, these methods constrain in their ability to understand ecosystem scale processes that generally require finer temporal and spatial sampling coverage (Gemery et al., 1996; Orlowski et al., 2018).

The development of automated laser spectroscopy systems mounted on the EC towers provided new opportunities of obtaining higher spatially and temporally resolved atmosphere profiles of these isotopes (Fiorella et al., 2021). The National Ecological Observatory Network (NEON), for the first time, provides standardized measurements of the stable isotope ratios of water vapor and carbon dioxide that can be converted to δ^2H, $\delta^{18}O$ of *LH* and $\delta^{13}C$ of *NEE*. These measurements were made at considerable cost and effort, yet it is not known how much additional constraints on carbon and water fluxes these isotopes are able to provide beyond traditional meteorological observations (e.g., vapor pressure deficit [*VPD*], air temperature [*T*], net radiation [R_n], windspeed [*u*]) (Wong, 2016). Recent advances in information theory allow for quantification of linear and nonlinear interactions between variables (termed mutual

information) (Cover & Thomas, 2005) as well as approaches to diagnose how unique the information provided by new data sources is relative to others (Goodwell & Kumar, 2017). Given that stable isotopes have been used to investigate carbon and water cycles at numerous sites (Brady et al., 2019; Vargas et al., 2017; Wang et al., 2019), this study addresses three simple questions: (1) do δ^2H, $\delta^{18}O$ of *LH* and $\delta^{13}C$ contain useful information about the bulk CO_2 and H_2O fluxes across North America, (2) can any information provided about the bulk fluxes by isotope measurements be obtained from other meteorological variables, and (3) under what circumstances do the isotopes provide more additional information?

4.4 Results

4.4.1 Individual Mutual Information

Across the NEON sites of North America, we evaluated the mutual information that different individual meteorological variables and isotopes share with *NE*E and *LH* (Figure 4.1A). Consistently, R_n and *T* are found to contain the most information about the *NEE* across all isotope datasets, with R_n being the most informative single variable of the *NEE*. *VPD* and *u* also provides information about *LH* but less than the information that can be gleaned from R_n and *T*. Broadly, the isotopes provided at least the same amount of information as *u* to *NEE*, with $\delta^{13}C$ generally containing more information about *NEE* than the other two isotopes. Overall, there is more inter-site variability in mutual information for *VPD*, *T* and R_n than the isotopes and *u*. While the information provided by R_n is almost an order of magnitude larger than the information from *u* and isotopes, *u* is a well-established driver of surface-atmosphere water and carbon exchange and a driver to turbulence in the planetary boundary layer (Yusup & Liu, 2020). The amount of information that can be inferred from isotopes (and other variables) about *NEE* are significantly highly unlikely to be obtained by random process ($p < 0.01$).

The mutual information between each individual variable and isotopes with *LH* demonstrates similar patterns to those of *NEE*, though overall, both meteorological variables and isotope values consistently share more information with LH than *NEE*. R_n contains the most information about *LH*, then *T* and *VPD*, and finally with isotopes and *u*. Again, the information

of isotopes shared with *LH* is at least as much as that can be obtained by *u*. Measurements of δ^2H and $\delta^{13}C$ generally contain more information about the *LH* than *NEE* with $\delta^{13}C$ being the most informative isotope to *LH* and *NEE*, respectively. In addition, there is more inter-site variability in terms of mutual information between each isotope and *LH* than the mutual information between each isotope and *NEE*.

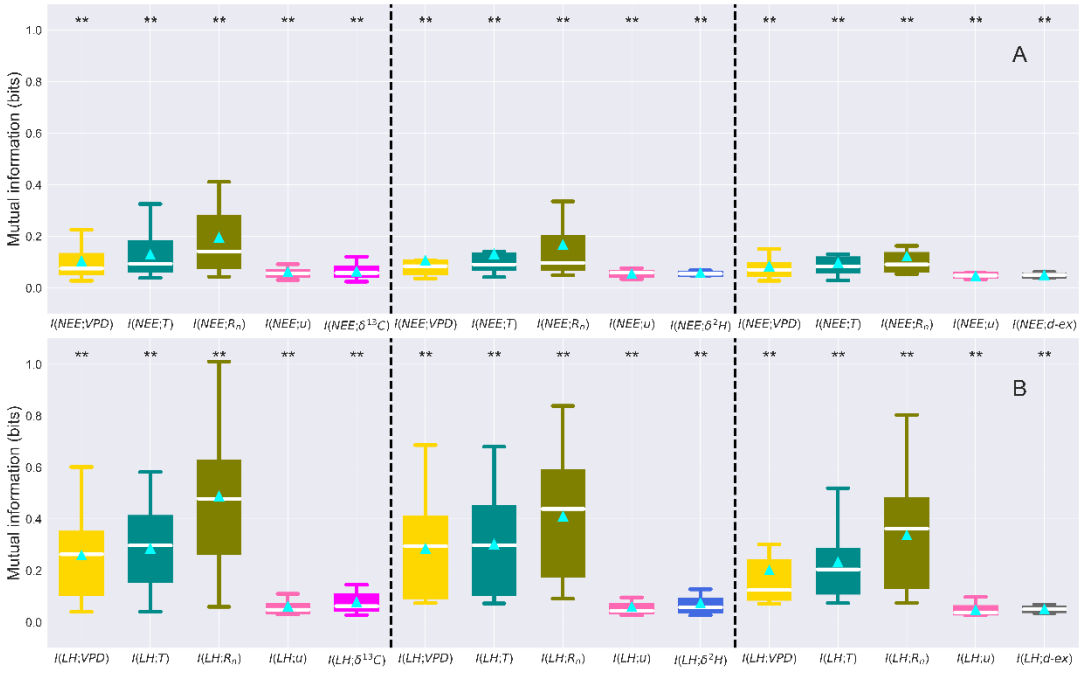

Figure 4.1 Mutual information that meteorological variables (vapor pressure deficit [*VPD*], air temperature [*T*], net radiation [R_n], windspeed [*u*]) and stable isotopes (δ^2H, $\delta^{13}C$, *d-ex*) contain about net ecosystem exchange (*NEE*) (A) and latent heat fluxes (*LH*) (B). Note: ** indicates a significant p-value (<0.01) of the paired t-test between the observed mutual information and the mutual information from 50 shuffle iterations.

4.4.2 Decomposition of Multivariate Mutual Information

We decomposed the multivariate mutual information between *NEE*/*LH*, isotopes and other variables and evaluated them in figure 4.2. We find that most of the information provided by isotopes to *NEE* is unique (Figure 4.S1). All the unique information provided isotopes are non-trivial and highly unlikely to be obtained by random ($p < 0.01$). The synergistic information of *NEE* is similar across all isotope datasets with the synergistic information from $\delta^{13}C$ being

slightly higher (Figure 4.S2). The redundant information provided by these isotopes are small compared to the unique and synergistic information (Figure 4.S3). The unique information provided by $\delta^{13}C$ to *NEE* has larger inter-site variability than the unique information of $\delta^{2}H$ and *d-ex* provided to *NEE*. All the partial information components of the isotopes about *NEE* are significant and non-trivial ($p < 0.01$).

Like the information components of *NEE*, the unique information and synergistic information of these isotopes to *LH* is higher compared to redundant information. The synergistic information provided by the isotopes to *LH* are similar with only synergistic information of $\delta^{13}C$ being non-trivial ($p < 0.01$). The $\delta^{13}C$ tends to provide more unique information to *LH* than the rest of isotope datasets. The unique information provided by $\delta^{13}C$ to *LH* is slightly higher compared to $\delta^{2}H$ and *d-ex*. The unique information provided by $\delta^{13}C$ and $\delta^{2}H$ to *LH* are higher than the unique information provided by $\delta^{13}C$ and $\delta^{2}H$ to *NEE*. The synergistic components of these isotopes about *LH* are quite similar to that found for *NEE* but are more variable. The redundant components of isotopes about the *LH* are the consistently smallest informational component when compared to the unique and synergistic information of $\delta^{13}C$ and $\delta^{2}H$ about *LH* and are larger than the redundant information of $\delta^{13}C$ and $\delta^{2}H$ about *NEE*.

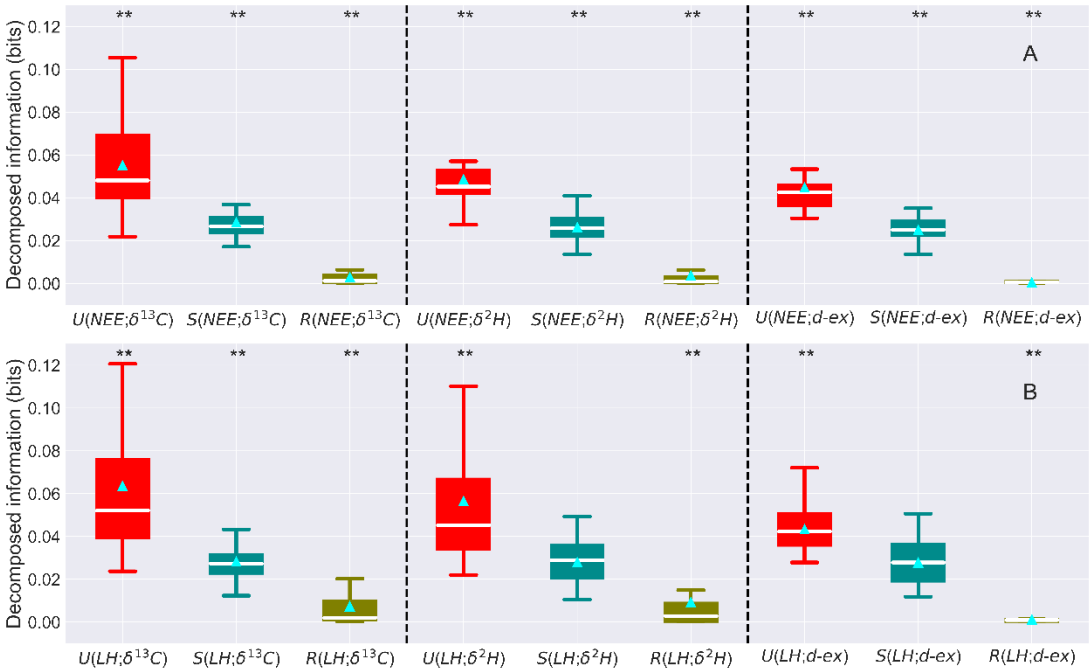

Figure 4.2 The averaged unique information (*U*), synergistic information (*S*) and redundant information (*R*) of the stable isotopes provided to *NEE* (A) and *LH* (B). Note: the ** indicates a significant p-value (<0.01) of the paired t-test between the observed decomposed components and decomposed components from 50 shuffle iterations.

4.4.3 The Additive Information of Isotopes

The additional information, represented by the sum of the synergistic information and the unique information, provided by the isotopes to *LH*/*NEE* are evaluated across NEON sites of North America (Figure 4.3). This additional information is the information provided by the isotopes beyond what can be obtained by other variables. In general, $\delta^{13}C$ tend to provide more additional information to *NEE*, while the *d-ex* provides the least amount of additive information to *NEE*. The fraction of additive information of isotopes about *NEE* to the sum of unique information, synergistic information, and redundant information are 0.97 for $\delta^{13}C$, 0.96 for $\delta^{2}H$, and 0.99 for *d-ex*, respectively. $\delta^{13}C$ also provide more additive information about *LH*, then $\delta^{2}H$ and *d-ex* (Figure 4.3A). The fraction of information provided by the isotopes that is (i.e. (*U* +*S*)/(*U*+*S*+*R*)) 0.94 for $\delta^{13}C$, 0.92 for $\delta^{2}H$, and 0.98 for *d-ex*, respectively. The additive information of $\delta^{13}C$ and $\delta^{2}H$ about *LH* have larger variability than these isotopes carry for *NEE*

(Figure 4.3A-4.3B and Figure 4.3D-4.3E), while there is less variability in the additive information of *d-ex* about *LH* (Figure 4.3C and Figure 4.3F) than *NEE*. Overall, the majority of the additive information of these isotopes about *NEE* and *LH* are non-trivial ($p < 0.01$) except the additive information of *d-ex* about *NEE*.

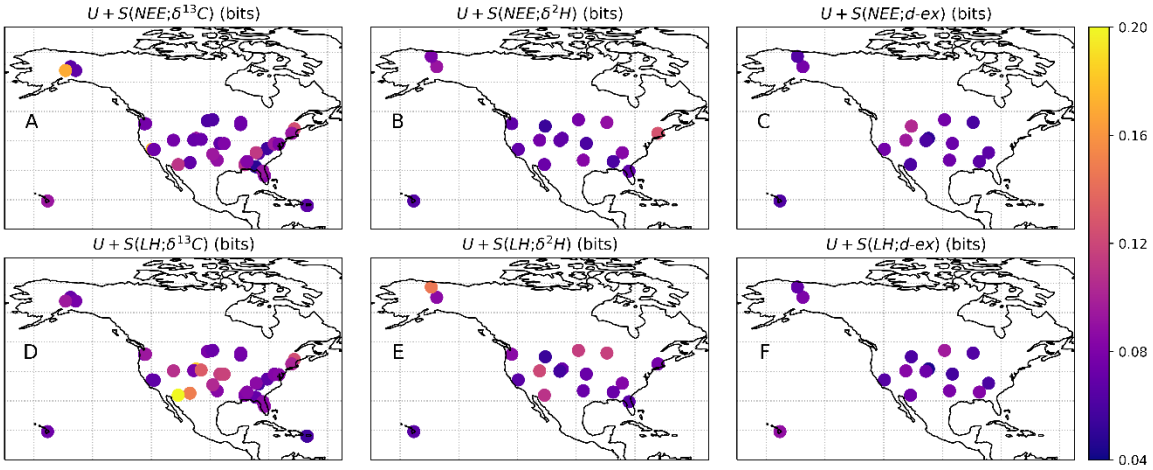

Figure 4.3 The spatial distribution of added information, calculated as the sum of unique information (*U*) and synergistic information (*S*), of different stable isotopes about *NEE* (A-C) and *LH* (D-F).

4.5 Discussion

This is the first evaluation of the capacity of isotopes to provide useful information about the bulk carbon and water fluxes across continental scale gradients. One of key motivations for measuring stable isotopes of water and carbon is that they provide a unique way of partitioning the bulk fluxes into their respective constituents (Conrad et al., 2012; Good et al., 2014; Wang et al., 2010) that represents distinct biophysical processes, and it is possible that when used to study these flux sub-components the isotopes will be even more informative (Wong, 2016). We showed that the information individually provided by these isotopes is not as much as *VPD*, *T*, and R_n but similar to *u*. These meteorological variables are traditionally known to be related to bulk fluxes and are standard measurements at weather stations. In addition, some of the meteorological variables evaluated here are known to be inter-related to some extent. For instance, the *VPD* is strongly dependent on *T* due to the Classuis-Claperyon relationship, and *T* is

tightly related to the amount of R_n. Therefore, it is expected that these variables contain more information about *NEE* and *LH*. Many studies have highlighted how *NEE* and *LH* respond to changes in *VPD*, *T* and R_n across various scales, seasons, and ecosystems (Chen et al., 2020; Gu et al., 2006; Niu et al., 2012). *VPD* was found to have a direct effect on surface energy partitioning as high *VPD* represents high atmosphere demand and hence high *LH* (Gu et al., 2006). Yet, high *VPD* reduces the stomata conductance and thereby negatively affects *LH* (Grossiord et al., 2020). The *u* also influences the *LH* non-linearly by the atmospheric conductivity (Davarzani et al., 2014). The divergent effect of *VPD* on *LH* may be underrepresented by other metrics but can be captured if evaluated using information theory-based metrics explored here.

Although individual isotopes are not as informative as other variables, the information carried by these isotopes was found to be unique and cannot be obtained from the other variables alone. It is crucial to know how different variables interactively provide information to a target of interest. Knowing the interactive dependencies between the inputs and outputs of a studied system is critical for model development and model simplification because it is more desired that most of the inputs of the model provide unique or synergistic piece of information rather than the redundant information (Wibral et al., 2017). The decomposition of the multivariate mutual information between isotopes, other meteorological variables, and the bulk fluxes offers an opportunity to elucidate in what form the information from isotopes is transferred to these bulk fluxes. We showed that isotopes carry unique signals about the bulk fluxes that cannot be implied by other variables. The portion of unique information from isotopes measurements is these isotopes is non-trivial and larger than the redundant information. Larger unique information often indicates less inter-dependency of the isotopes on other meteorological variables, suggesting that the isotopes of the fluxes may be collectively interacting with these meteorological variables by different pathways. However, we observed that there are inter-sites variations in the unique information provided by the isotopes, indicating that the unique information from these isotopes may be site condition dependent. Similarly, isotopes can provide synergistic information about the bulk fluxes suggesting that the patterns of bulk fluxes might be better characterized with the isotopes included as an additional constraint.

The additional information provided by isotopes to these bulk fluxes are described by the sum of unique information and synergistic information. The sum of these two components, on average, represents how remaining uncertainty can be reduced with isotopes observations given other meteorological variables in the systems. In other words, how much information is transferred from isotopes to the bulk fluxes. On average, the uncertainty of the fluxes can be reduced given the isotope information. Numerous approaches have been developed to estimate and forecast *NEE* and *LH* using empirically or physically based models (Fisher et al., 2008; Su, 2002; Veroustraete et al., 1996; Wood, 2021). However, most of these methods are not perfect and often either must make some assumptions or simplifications and can be subject to significant amounts of uncertainty (Papale et al., 2006; Zhao et al., 2020).

Results here demonstrate that fusing isotope data products can be useful to better monitor and predictions of *NEE* and *LH*, as these isotope datasets provide additional information beyond what can be obtained in traditional meteorological variables. But it is not necessary to conclude that the isotopes datasets are informative under all circumstances. We evaluated the additive information of isotopes based on NEON site conditions via a multivariate linear regression (Table 4.S1). We show that the additive information of $\delta^{13}C$ provided to *LH* is significantly influenced by how arid the site is as indicated by a significant positive coefficient of the linear regression with the R^2 value of 0.41. This suggests that $\delta^{13}C$ likely provides more useful information about *LH* in locations with higher atmospheric evaporative demand relative to supplied precipitation. The additive information of $\delta^{2}H$ about *LH* is shown to be driven by multiple factors which can reliably (R^2 of 0.55) use predict when these observations provide additional constraints. Based on our analysis, $\delta^{2}H$ likely provides more provide additional information about *LH* at lower altitude or in cooler places. However, $\delta^{2}H$ tends to provide more information or at drier locations (Figure 4.4). The water isotopes are generally used to partition evapotranspiration to evaporation and transpiration. However, the results were found to be different across various ecosystems with different accuracy (Lu et al., 2017; Lixin Wang et al., 2014; Wen et al., 2016). Given our analysis, researchers should expect to see those approaches that involve $\delta^{2}H$ show better results in arid locations than in humid locations for the estimation or partition of evapotranspiration.

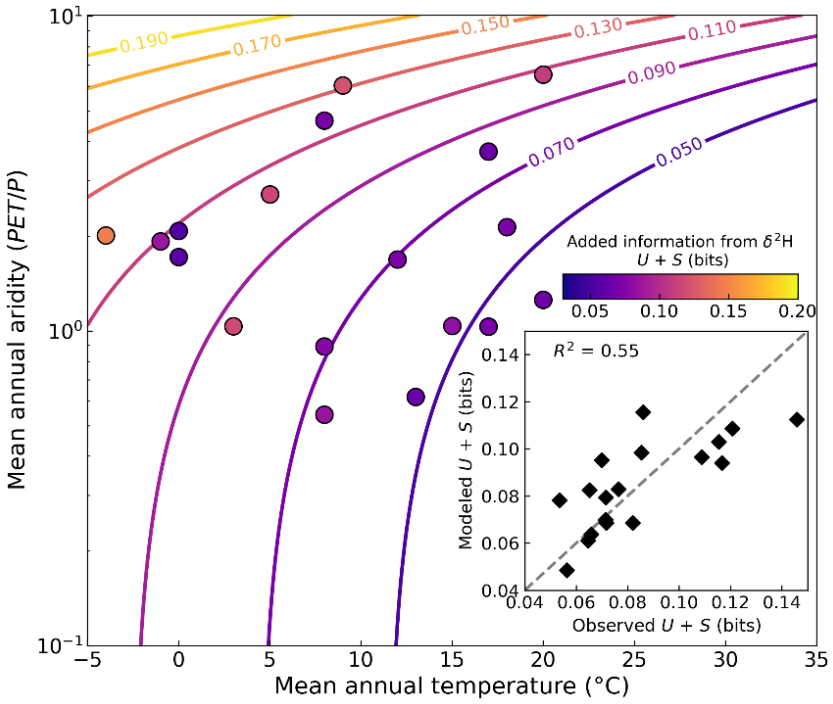

Figure 4.4 The mean annual aridity against mean annual temperature colored by the added information ($U + S$) of δ^2H about LH. The multivariate linear regression modeled added information of δ^2H against observed added information (inset).

4.6 Material and Methods

4.6.1 NEON Sites

This study was conducted at terrestrial sites part of the National Ecological Observatory Network (NEON), which is a continental scale research platform for understanding the fundamental principles that govern ecological responses to climate change, land use change and species invasion. NEON statistically divided US territory into 20 ecoclimatic domains across different biomes and landforms to effectively capture key biological aspects of US ecology (Elmendorf et al., 2016). NEON collects and delivers coordinated and standardized data about plants, animals, soil, nutrients, freshwater and the atmosphere alongside field measurements and airborne remote sensing that can provide the capacity to forecast future states of ecological systems and enable the scientific community to effectively address critical ecological issues (Barnett et al., 2019).

4.6.2 Data Preparation

The NEON's eddy covariance bundled product provides data about eddy covariance storage exchange fluxes, basic meteorological data and isotopes ratio product of water vapor and carbon isotope of atmospheric CO_2. These data were collected at multiple measurement heights along the tower profile and were then bounded in NEON data product DP4.00200.001 (NEON, 2021a). In this study, we extracted the 30 min aggregated *NEE*, *LH*, total net radiation (R_n), tower top air temperature (T), and the two-dimensional wind speed (u) dataset. The vapor pressure deficit data were computed from the relative humidity product DP1.00098.001 (NEON, 2021b)and air temperature. There exist some unusual spikes and data gaps in the *NEE* and *LH* datasets due instrumentation failure and calibration issues. We filtered the *NEE* by friction velocity and gap filled the *NEE* and *LH* using "ReddyProc" R package (Wutzler et al., 2018). We then filtered *NEE*, *LH*, and other meteorological datasets by an interquartile filter that only keeps the data points within 1.5 interquartile range (IQR) below the 25th percentile (p25) and 1.5*IQR above the 75th (p75).

4.6.3 Stable Isotope Data

The isotope ratios from the eddy covariance bundle represent the isotope composition of the gases at the specific measurement layer and hence cannot fully reflect the isotope composition of *NEE* and *LH* at ecosystem scale. In addition, the isotope datasets offered by NEON are frequently experiencing instrumental drift which will deviate from the Vienna Standard Mean Ocean Water scale (VSMOW) scale. Efforts have been made to calibrate the carbon isotope ratio of the atmosphere CO_2 in the "NEONiso" R package (Fiorella et al., 2021). This package has been extended to calibrate the water vapor isotopes. Therefore, the estimation of the isotope composition of fluxes in this study requires two procedures: calibrating the NEON's isotope dataset to VPDB scale and estimating the daily isotope composition of the fluxes using the calibrated isotope products at ecosystem scale. The most adopted way of estimating the isotope composition is using the Keeling plots, with the basic principle of the Keeling plot method is the conservation of mass that estimates the isotope ratio of *LH* and *NEE* at ecosystem scale by

leveraging the isotope composition and their mole fractions (Pataki et al., 2003). A common variant of Keeling plot is Miller-Tans mixing model that estimates the isotope ratio of ecosystem scale fluxes by fitting a simple linear regression line between the product of isotopes of the measurements at all measurement heights and their mole fraction and the mole fraction (Miller & Tans, 2003). The slopes of such regression lines are estimates of the isotope composition of the *NEE* and *LH* at ecosystem level. We filtered the isotope ratio data that is generated by fitting less than 5 data points into the linear regression and only keep the slopes that have R^2 greater 0.9. Finally, we keep the isotope data points within p_{25} – 3*IQR and p_{75} + 3*IQR. The deuterium excess (*d-ex*) is computed using the daily $\delta^2 H$ and $\delta^{18} O$ of *LH*.

4.6.4 Mutual Information Measures

Mutual information is a measure of how two random variables are probabilistically dependent on each other (Cover & Thomas, 2005). It quantifies how much uncertainty of one random variable can be reduced given the knowledge of another random variable. Probabilistically, the mutual information can be expressed as

$$I(\mathrm{X};\mathrm{Y}) = \sum p(x,y)\log_2\left(\frac{p(x,y)}{p(x)\mathrm{p}(y)}\right) \quad (1)$$

where $p(x)$, $p(y)$, and $p(x,y)$ are probability density functions of random variables X, Y, and {X,Y} respectively.

The multivariate mutual information of a single random variable (Z) and a set of random variables {X, Y} characterize the uncertainty of Z can be reduced by the knowledge of {X, Y} and can be expressed as

$$I(\mathrm{X},\mathrm{Y};\mathrm{Z}) = \sum p(x,y,z)\log 2\left(\frac{\mathrm{p}(x,y,z)}{\mathrm{p}(x,y)\mathrm{p}(z)}\right) \quad (2)$$

where $p(z)$, $p(x,y)$, and $p(x,y,z)$ are the probability density functions of variable Z, {X,Y}, and {X,Y,Z}, respectively. All the probability density functions were estimated using kernel density estimation (KDE) method with a gaussian kernel. We normalized the data point to the common

range of [0, 1] before using KDE. We then evaluate the probability density functions from 0, 1 with a step size of 0.05.

We computed the mutual information shared from *VPD*, *T*, R_n, *u*, $\delta^{13}C$, $\delta^{2}H$, and *d-ex* about *NEE* and *LH* iteratively. Due to the limitation of isotope datasets, we computed the mutual information of each variable with the *NEE* and *LH* by subsample 40 data points without replacement 500 times. The average of the mutual information from the resampling datasets are as the mutual information between the variable of interest and the flux.

4.6.5 Partial Information Decomposition

It was shown that the multivariate mutual information can be decomposed into different informational components via a partial information decomposition framework (PID) (Goodwell & Kumar, 2017). The PID can decompose I(X,Y;Z) into (1) unique information (U) that is only provided by X or Y solely to the Z, respectively; (2) synergistic information (S) that is the information provide to the Z when X and Y act jointly; (3) redundant information (R) that is the overlapping information provided both by X and Y to the Z. The PID framework can be formulated as

$$I(\mathrm{X,Y;\ Z}) = U_{\mathrm{X}} + U_{\mathrm{Y}} + R + S \tag{3}$$

$$I(\mathrm{X;\ Z}) = U_{\mathrm{X}} + R \tag{4}$$

$$I(\mathrm{Y;\ Z}) = U_{\mathrm{Y}} + R \tag{5}$$

Where U_{X} and U_{Y} are the unique information of X and Y to the Z, respectively. R and S are the redundant and synergistic information of X and Y to the Z, respectively. All PID components are non-negative real numbers.

In this study, we quantified the between the fluxes and isotopes by leveraging the PID framework. We first compute the unique information of the isotope that contributes to the bulk fluxes. The unique information can only be solved when other variables (besides the bulk flux and isotope) are available in the PID system. Therefore, we define the decomposed information

components of isotope fluxes provided to the bulk fluxes as the averaged unique information across all other meteorological variables (*VPD*, *T*, R_n, *u*). Like the computation of the individual mutual information, we also subsample 40 data points from each dataset without replacement 500 times. The partial information components of the isotopes are then computed as the averaged information components from 500 iterations.

4.7 References

Baldocchi, D. (2014). Measuring fluxes of trace gases and energy between ecosystems and the atmosphere - the state and future of the eddy covariance method. *Global Change Biology*, *20*(12), 3600–3609. https://doi.org/10.1111/gcb.12649

Barnett, D. T., Adler, P. B., Chemel, B. R., Duffy, P. A., Enquist, B. J., Grace, J. B., Harrison, S., Peet, R. K., Schimel, D. S., Stohlgren, T. J., & Vellend, M. (2019). The plant diversity sampling design for The National Ecological Observatory Network. *Ecosphere*, *10*(2). https://doi.org/10.1002/ecs2.2603

Bauska, T. K., Baggenstos, D., Brook, E. J., Mix, A. C., Marcott, S. A., Petrenko, V. V., Schaefer, H., Severinghaus, J. P., & Lee, J. E. (2016). Carbon isotopes characterize rapid changes in atmospheric carbon dioxide during the last deglaciation. *Proceedings of the National Academy of Sciences*, *113*(13), 3465–3470. https://doi.org/10.1073/pnas.1513868113

Brady, E., Stevenson, S., Bailey, D., Liu, Z., Noone, D., Nusbaumer, J., Otto-Bliesner, B. L., Tabor, C., Tomas, R., Wong, T., Zhang, J., & Zhu, J. (2019). The Connected Isotopic Water Cycle in the Community Earth System Model Version 1. *Journal of Advances in Modeling Earth Systems*, *11*(8), 2547–2566. https://doi.org/10.1029/2019MS001663

Chen, J., Wen, J., Kang, S., Meng, X., Tian, H., Ma, X., & Yuan, Y. (2020). Assessments of the factors controlling latent heat flux and the coupling degree between an alpine wetland and the atmosphere on the Qinghai-Tibetan Plateau in summer. *Atmospheric Research, 240*, 104937. https://doi.org/10.1016/j.atmosres.2020.104937

Conrad, R., Klose, M., Yuan, Q., Lu, Y., & Chidthaisong, A. (2012). Stable carbon isotope fractionation, carbon flux partitioning and priming effects in anoxic soils during methanogenic degradation of straw and soil organic matter. *Soil Biology and Biochemistry*, *49*, 193–199. https://doi.org/10.1016/j.soilbio.2012.02.030

Cover, T. M., & Thomas, J. A. (2005). *Elements of Information Theory*. Wiley. https://doi.org/10.1002/047174882X

Davarzani, H., Smits, K., Tolene, R. M., & Illangasekare, T. (2014). Study of the effect of wind speed on evaporation from soil through integrated modeling of the atmospheric boundary layer and shallow subsurface. *Water Resources Research*, *50*(1), 661–680. https://doi.org/10.1002/2013WR013952

Elmendorf, S. C., Jones, K. D., Cook, B. I., Diez, J. M., Enquist, C. A. F., Hufft, R. A., Jones, M. O., Mazer, S. J., Miller-Rushing, A. J., Moore, D. J. P., Schwartz, M. D., & Weltzin, J. F. (2016). The plant phenology monitoring design for The National Ecological Observatory Network. *Ecosphere*, *7*(4), e01303. https://doi.org/10.1002/ecs2.1303

Ferguson, P. R., & Veizer, J. (2007). Coupling of water and carbon fluxes via the terrestrial biosphere and its significance to the Earth's climate system. *Journal of Geophysical Research*, *112*(D24), D24S06. https://doi.org/10.1029/2007JD008431

Fiorella, R. P., Good, S. P., Allen, S. T., Guo, J. S., Still, C. J., Noone, D. C., Anderegg, W. R. L., Florian, C. R., Luo, H., Pingintha-Durden, N., & Bowen, G. J. (2021). Calibration Strategies for Detecting Macroscale Patterns in NEON Atmospheric Carbon Isotope

Observations. *Journal of Geophysical Research: Biogeosciences*, *126*(3). https://doi.org/10.1029/2020JG005862

Fisher, J. B., Tu, K. P., & Baldocchi, D. D. (2008). Global estimates of the land–atmosphere water flux based on monthly AVHRR and ISLSCP-II data, validated at 16 FLUXNET sites. *Remote Sensing of Environment*, *112*(3), 901–919. https://doi.org/10.1016/j.rse.2007.06.025

Foken, T., Aubinet, M., & Leuning, R. (2012). The Eddy Covariance Method. In *Eddy Covariance* (pp. 1–19). Springer Netherlands. https://doi.org/10.1007/978-94-007-2351-1_1

Gemery, P. A., Trolier, M., & White, J. W. C. (1996). Oxygen isotope exchange between carbon dioxide and water following atmospheric sampling using glass flasks. *Journal of Geophysical Research: Atmospheres*, *101*(D9), 14415–14420. https://doi.org/10.1029/96JD00053

Gentine, P., Green, J. K., Guérin, M., Humphrey, V., Seneviratne, S. I., Zhang, Y., & Zhou, S. (2019). Coupling between the terrestrial carbon and water cycles—a review. *Environmental Research Letters*, *14*(8), 083003. https://doi.org/10.1088/1748-9326/ab22d6

Good, S. P., Soderberg, K., Guan, K., King, E. G., Scanlon, T. M., & Caylor, K. K. (2014). δ 2 H isotopic flux partitioning of evapotranspiration over a grass field following a water pulse and subsequent dry down. *Water Resources Research*, *50*(2), 1410–1432. https://doi.org/10.1002/2013WR014333

Goodwell, A. E., & Kumar, P. (2017). Temporal information partitioning: Characterizing synergy, uniqueness, and redundancy in interacting environmental variables. *Water Resources Research*, *53*(7), 5920–5942. https://doi.org/10.1002/2016WR020216

Grossiord, C., Buckley, T. N., Cernusak, L. A., Novick, K. A., Poulter, B., Siegwolf, R. T. W., Sperry, J. S., & McDowell, N. G. (2020). Plant responses to rising vapor pressure deficit. *New Phytologist*, *226*(6), 1550–1566. https://doi.org/10.1111/nph.16485

Gu, L., Meyers, T., Pallardy, S. G., Hanson, P. J., Yang, B., Heuer, M., Hosman, K. P., Riggs, J. S., Sluss, D., & Wullschleger, S. D. (2006). Direct and indirect effects of atmospheric conditions and soil moisture on surface energy partitioning revealed by a prolonged drought at a temperate forest site. *Journal of Geophysical Research*, *111*(D16), D16102. https://doi.org/10.1029/2006JD007161

Jia, Z., Liu, S., Xu, Z., Chen, Y., & Zhu, M. (2012). Validation of remotely sensed evapotranspiration over the Hai River Basin, China. *Journal of Geophysical Research: Atmospheres*, *117*(D13), n/a-n/a. https://doi.org/10.1029/2011JD017037

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

Kuang, X., Luo, X., Jiao, J. J., Liang, S., Zhang, X., Li, H., & Liu, J. (2019). Using stable isotopes of surface water and groundwater to quantify moisture sources across the Yellow River source region. *Hydrological Processes*, hyp.13441. https://doi.org/10.1002/hyp.13441

Lee, S.-C., Christen, A., Black, T. A., Jassal, R. S., Ketler, R., & Nesic, Z. (2020). Partitioning of net ecosystem exchange into photosynthesis and respiration using continuous stable isotope measurements in a Pacific Northwest Douglas-fir forest ecosystem. *Agricultural and Forest Meteorology*, *292–293*, 108109. https://doi.org/10.1016/j.agrformet.2020.108109

Li, S., Kang, S., Li, F., & Zhang, L. (2008). Evapotranspiration and crop coefficient of spring maize with plastic mulch using eddy covariance in northwest China. *Agricultural Water Management*, *95*(11), 1214–1222. https://doi.org/10.1016/j.agwat.2008.04.014

Lu, X., Liang, L. L., Wang, L., Jenerette, G. D., McCabe, M. F., & Grantz, D. A. (2017). Partitioning of evapotranspiration using a stable isotope technique in an arid and high temperature agricultural production system. *Agricultural Water Management*, *179*, 103–109. https://doi.org/10.1016/j.agwat.2016.08.012

Lupascu, M., Akhtar, H., Smith, T. E. L., & Sukri, R. S. (2020). Post-fire carbon dynamics in the tropical peat swamp forests of Brunei reveal long-term elevated CH 4 flux. *Global Change Biology*, *26*(9), 5125–5145. https://doi.org/10.1111/gcb.15195

MILLER, J. B., & TANS, P. P. (2003). Calculating isotopic fractionation from atmospheric measurements at various scales. *Tellus B*, *55*(2), 207–214. https://doi.org/10.1034/j.1600-0889.2003.00020.x

National Ecological Observatory Network (NEON). (2021a). *Bundled data products - eddy covariance (DP4.00200.001)*. National Ecological Observatory Network (NEON). https://doi.org/10.48443/BWAY-HC74

National Ecological Observatory Network (NEON). (2021b). *Relative humidity (DP1.00098.001)*. National Ecological Observatory Network (NEON). https://doi.org/10.48443/S9WY-T644

Niu, S., Luo, Y., Fei, S., Yuan, W., Schimel, D., Law, B. E., Ammann, C., Altaf Arain, M., Arneth, A., Aubinet, M., Barr, A., Beringer, J., Bernhofer, C., Andrew Black, T., Buchmann, N., Cescatti, A., Chen, J., Davis, K. J., Dellwik, E., … Zhou, X. (2012). Thermal optimality of net ecosystem exchange of carbon dioxide and underlying mechanisms. *New Phytologist*, *194*(3), 775–783. https://doi.org/10.1111/j.1469-8137.2012.04095.x

Orlowski, N., Breuer, L., Angeli, N., Boeckx, P., Brumbt, C., Cook, C. S., Dubbert, M., Dyckmans, J., Gallagher, B., Gralher, B., Herbstritt, B., Hervé-Fernández, P., Hissler, C., Koeniger, P., Legout, A., Macdonald, C. J., Oyarzún, C., Redelstein, R., Seidler, C., … McDonnell, J. J. (2018). Inter-laboratory comparison of cryogenic water extraction systems for stable isotope analysis of soil water. *Hydrology and Earth System Sciences*, *22*(7), 3619–3637. https://doi.org/10.5194/hess-22-3619-2018

Papale, D., Reichstein, M., Aubinet, M., Canfora, E., Bernhofer, C., Kutsch, W., Longdoz, B., Rambal, S., Valentini, R., Vesala, T., & Yakir, D. (2006). Towards a standardized processing of Net Ecosystem Exchange measured with eddy covariance technique: algorithms and uncertainty estimation. *Biogeosciences*, *3*(4), 571–583. https://doi.org/10.5194/bg-3-571-2006

Pataki, D. E., Ehleringer, J. R., Flanagan, L. B., Yakir, D., Bowling, D. R., Still, C. J., Buchmann, N., Kaplan, J. O., & Berry, J. A. (2003). The application and interpretation of Keeling plots in terrestrial carbon cycle research. *Global Biogeochemical Cycles*, *17*(1). https://doi.org/10.1029/2001GB001850

ROZANSKI, K., JOHNSEN, S. J., SCHOTTERER, U., & THOMPSON, L. G. (1997). Reconstruction of past climates from stable isotope records of palaeo-precipitation preserved in continental archives. *Hydrological Sciences Journal*, *42*(5), 725–745. https://doi.org/10.1080/02626669709492069

Scanlon, T. M., Schmidt, D. F., & Skaggs, T. H. (2019). Correlation-based flux partitioning of water vapor and carbon dioxide fluxes: Method simplification and estimation of canopy water use efficiency. *Agricultural and Forest Meteorology*, *279*, 107732. https://doi.org/10.1016/j.agrformet.2019.107732

Scholl, M. A., & Murphy, S. F. (2014). Precipitation isotopes link regional climate patterns to water supply in a tropical mountain forest, eastern Puerto Rico. *Water Resources Research*, *50*(5), 4305–4322. https://doi.org/10.1002/2013WR014413

Scott, R. L., & Biederman, J. A. (2017). Partitioning evapotranspiration using long-term carbon dioxide and water vapor fluxes. *Geophysical Research Letters*, *44*(13), 6833–6840. https://doi.org/10.1002/2017GL074324

Su, Z. (2002). The Surface Energy Balance System (SEBS) for estimation of turbulent heat fluxes. *Hydrology and Earth System Sciences*, *6*(1), 85–100. https://doi.org/10.5194/hess-6-85-2002

Trenberth, K. E., Smith, L., Qian, T., Dai, A., & Fasullo, J. (2007). Estimates of the Global Water Budget and Its Annual Cycle Using Observational and Model Data. *Journal of Hydrometeorology*, *8*(4), 758–769. https://doi.org/10.1175/JHM600.1

Vargas, A. I., Schaffer, B., Yuhong, L., & Sternberg, L. da S. L. (2017). Testing plant use of mobile vs immobile soil water sources using stable isotope experiments. *New Phytologist*, *215*(2), 582–594. https://doi.org/10.1111/nph.14616

Veroustraete, F., Patyn, J., & Myneni, R. B. (1996). Estimating net ecosystem exchange of carbon using the normalized difference vegetation index and an ecosystem model. *Remote Sensing of Environment*, *58*(1), 115–130. https://doi.org/10.1016/0034-4257(95)00258-8

Wang, B., Zhang, H., Liang, X., Li, X., & Wang, F. (2019). Cumulative effects of cascade dams on river water cycle: Evidence from hydrogen and oxygen isotopes. *Journal of Hydrology*, *568*, 604–610. https://doi.org/10.1016/j.jhydrol.2018.11.016

Wang, L., Caylor, K. K., Villegas, J. C., Barron-Gafford, G. A., Breshears, D. D., & Huxman, T. E. (2010). Partitioning evapotranspiration across gradients of woody plant cover: Assessment of a stable isotope technique. *Geophysical Research Letters*, *37*(9), n/a-n/a. https://doi.org/10.1029/2010GL043228

Wang, L., Good, S. P., & Caylor, K. K. (2014). Global synthesis of vegetation control on evapotranspiration partitioning. *Geophysical Research Letters*, *41*(19), 6753–6757. https://doi.org/10.1002/2014GL061439

Wen, X., Yang, B., Sun, X., & Lee, X. (2016). Evapotranspiration partitioning through in-situ oxygen isotope measurements in an oasis cropland. *Agricultural and Forest Meteorology*, *230–231*, 89–96. https://doi.org/10.1016/j.agrformet.2015.12.003

Wibral, M., Priesemann, V., Kay, J. W., Lizier, J. T., & Phillips, W. A. (2017). Partial information decomposition as a unified approach to the specification of neural goal functions. *Brain and Cognition*, 112, 25–38. https://doi.org/10.1016/j.bandc.2015.09.004

Williams, M., Richardson, A. D., Reichstein, M., Stoy, P. C., Peylin, P., Verbeeck, H., Carvalhais, N., Jung, M., Hollinger, D. Y., Kattge, J., Leuning, R., Luo, Y., Tomelleri, E., Trudinger, C. M., & Wang, Y.-P. (2009). Improving land surface models with FLUXNET data. *Biogeosciences*, *6*(7), 1341–1359. https://doi.org/10.5194/bg-6-1341-2009

Wong, T. E. (2016). *The Impact of Stable Water Isotopic Information on Parameter Calibration in a Land Surface Model*.

Wood, D. A. (2021). Net ecosystem carbon exchange prediction and insightful data mining with an optimized data-matching algorithm. *Ecological Indicators*, *124*, 107426. https://doi.org/10.1016/j.ecolind.2021.107426

Wutzler, T., Lucas-Moffat, A., Migliavacca, M., Knauer, J., Sickel, K., Šigut, L., Menzer, O., & Reichstein, M. (2018). Basic and extensible post-processing of eddy covariance flux data with REddyProc. *Biogeosciences*, *15*(16), 5015–5030. https://doi.org/10.5194/bg-15-5015-2018

Xiao, W., Wei, Z., & Wen, X. (2018). Evapotranspiration partitioning at the ecosystem scale using the stable isotope method—A review. *Agricultural and Forest Meteorology*, *263*, 346–361. https://doi.org/10.1016/j.agrformet.2018.09.005

Yusup, Y., & Liu, H. (2020). Effects of persistent wind speeds on turbulent fluxes in the water-atmosphere interface. *Theoretical and Applied Climatology*, 140(1–2), 313–325. https://doi.org/10.1007/s00704-019-03084-4

ZHANG, J., GRIFFIS, T. J., & BAKER, J. M. (2006). Using continuous stable isotope measurements to partition net ecosystem CO2 exchange. *Plant, Cell and Environment*, *29*(4), 483–496. https://doi.org/10.1111/j.1365-3040.2005.01425.x

Zhao, W. L., Qiu, G. Y., Xiong, Y. J., Paw U, K. T., Gentine, P., & Chen, B. Y. (2020). Uncertainties Caused by Resistances in Evapotranspiration Estimation Using High-Density Eddy Covariance Measurements. *Journal of Hydrometeorology*, *21*(6), 1349–1365. https://doi.org/10.1175/JHM-D-19-0191.1

Zhou, S., Yu, B., Zhang, Y., Huang, Y., & Wang, G. (2018). Water use efficiency and evapotranspiration partitioning for three typical ecosystems in the Heihe River Basin, northwestern China. *Agricultural and Forest Meteorology*, *253–254*, 261–273. https://doi.org/10.1016/j.agrformet.2018.02.002

4.8 Supplementary Materials

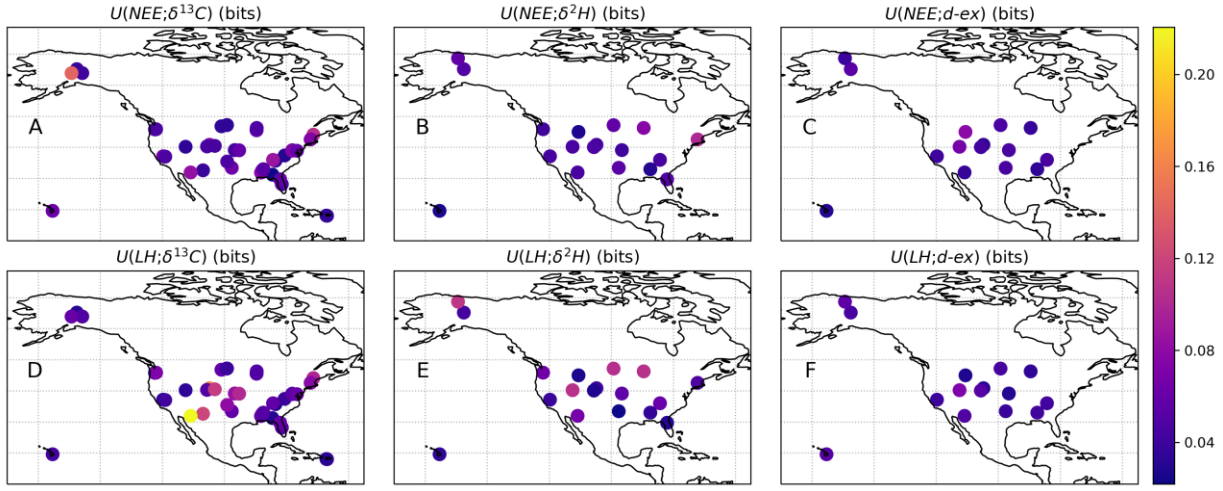

Figure 4.S1 The spatial distribution of unique information (*U*) of different isotopes about *NEE* (A-C) and *LH* (D-F).

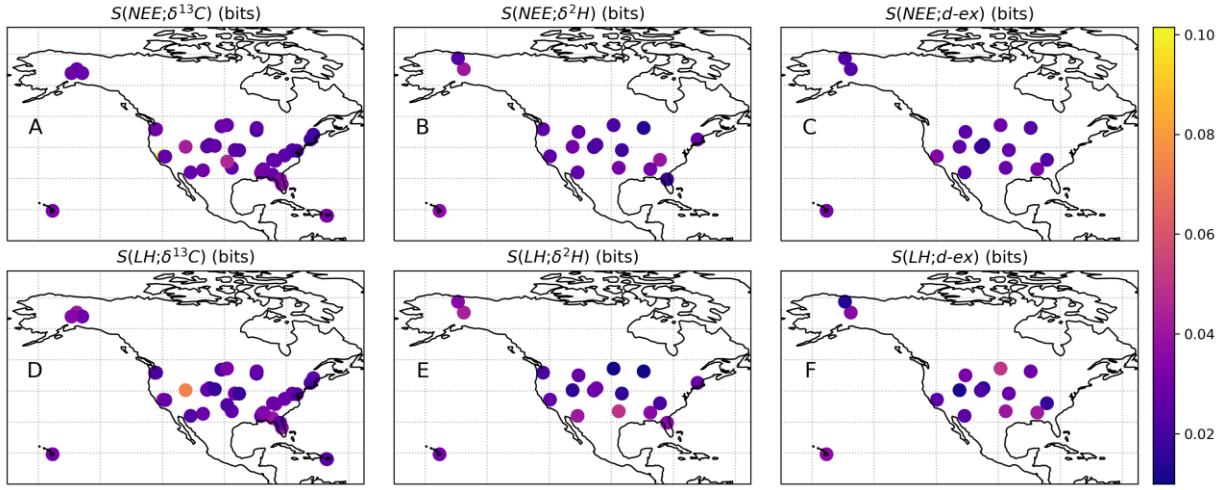

Figure 4.S2 The spatial distribution of synergistic information (*S*) of different isotopes (δ^2H, $\delta^{13}C$, *d-ex*) *NEE* (A-C) and *LH* (D-F).

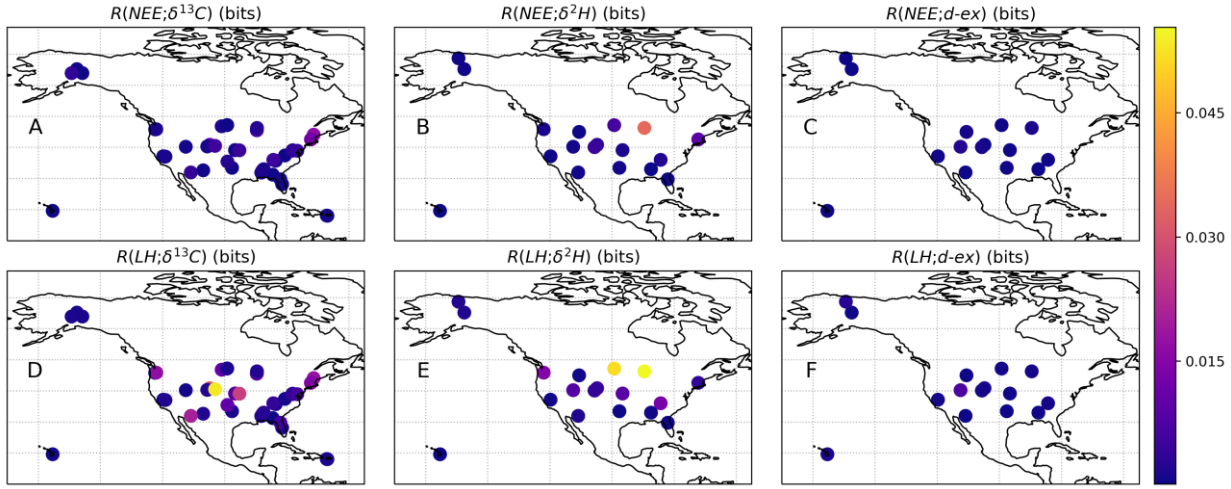

Figure 4.S3 The spatial distribution of redundant information (*R*) of different isotopes (δ^2H, $\delta^{13}C$, *d-ex*) about *NEE* (A-C) and *LH* (D-F).

Table 4.S1 The coefficients and p-values of the multivariate linear regression of the additive information (*U* + *S*) of isotopes and different site condition variables. Note: the asterisk indicates a p-value is significant at the level of 0.05.

	Aridity index	Mean annual temperature	Mean annual precipitation	Elevation	R^2
U + *S*(*NEE*;$\delta^{13}C$)	-1.0e-4	1.0e-4	-6.4e-6	-4.1e-6	0.02
U + *S*(*LH*;$\delta^{13}C$)	0.015*	3.6e-5	6.8e-6	-6.0e-6	0.41
U + *S*(*NEE*;δ^2H)	-1.0e-3	-6.0e-4	-4.9e-6	-8.0e-6	0.23
U + *S*(*LH*;δ^2H)	0.012*	-2.9e-3*	1.3e-5	-2.0e-5*	0.55
U + *S*(*NEE*;*d-ex*)	2.0e-4	-2e-4	-4.3e-6	-2.6e-8	0.06
U + *S*(*LH*;*d-ex*)	4.0e-3	-7.0e-4	1.6e-5	-4.9e-6	0.19

Chapter 5. General Conclusions

The challenges of understanding the interactions between land surface and atmosphere require advances in the environment sensing, and data analysis. Such challenge has been gradually resolved by launching new ecohydrological measurements in space and field. These new measurements in environmental sensing and data may possess the potential of improving understanding by further disentangling the complex nature of the Earth's ecosystems. This dissertation used cutting-edge methods including machine learning and information theory to the quantify how much additional information from SMAP and NEON's new measurements can be transferred to understand vegetation dynamics, carbon, and water cycles.

In chapter 2, I found that new soil moisture and vegetation optical depth observations from NASA SMAP is useful for the prediction of how vegetation leaf changes. The prediction power of these SMAP observations varies across different landcovers. The variations in vegetation leaf dynamics can be mostly explained by their daily climatology. In general, adding SMAP soil moisture and VOD observations can improve the predictive power of vegetation dynamics models and these observations added more predictive power in locations where the vegetation dynamics cannot be captured by their daily vegetation climatology.

In chapter 3, I found that majority of the uncertainty from SMAP DCA algorithm comes from informational random uncertainty induced by a lack of explanatory power of DCA model inputs. The informational model uncertainties are tightly related to the DCA soil moisture retrieval quality. The decomposition of multivariate mutual information between DCA soil moisture and DCA's major model inputs shown that the synergistic information and redundant information provided by the major inputs are strongly correlated to DCA soil moisture retrieval quality. SMAP DCA users may examine to the redundant or synergistic components to have an approximate estimation of the soil moisture data quality obtained when no in situ soil moisture is readily available.

In chapter 4, I found that the NEON's isotope observations contain non-trivial information about the bulk water and carbon fluxes. The individual information provided by the isotope

dataset to the bulk fluxes are not as much as the information provided by air temperature, net radiation and VPD, but they contain as least as much as the information provided by windspeed. The multivariate decomposition of mutual information between bulk fluxes, isotopes and other meteorological variables shown that isotopes contain unique information that cannot be simple obtained by other more common meteorological variables alone. They also provided synergistic information along with other meteorological observations and most of synergistic information is non-trivial. The sum of unique information and synergistic information, characterize the additive information of these isotopes, was found to be non-trivial. Therefore, synthesizing the isotope datasets a long with other environmental datasets may enhance the understanding of environmental fluxes. The multivariate linear regression analysis shown that there is more opportunity for δ^2H to be useful under arid conditions. Therefore, researchers may expect to see betters results in locations with high atmospheric demand.

Overall, the methods/frameworks used in this dissertation can be further applied to evaluate the added values of other new ecohydrological datasets as new sensors and observations are available in the future.

Bibliography

Al-Yaari, A., Wigneron, J.-P., Kerr, Y., Rodriguez-Fernandez, N., O'Neill, P. E., Jackson, T. J., De Lannoy, G. J. M., Al Bitar, A., Mialon, A., Richaume, P., Walker, J. P., Mahmoodi, A., & Yueh, S. (2017). Evaluating soil moisture retrievals from ESA's SMOS and NASA's SMAP brightness temperature datasets. *Remote Sensing of Environment*, *193*, 257–273. https://doi.org/10.1016/j.rse.2017.03.010

Asbjornsen, H., Goldsmith, G. R., Alvarado-Barrientos, M. S., Rebel, K., Van Osch, F. P., Rietkerk, M., Chen, J., Gotsch, S., Tobon, C., Geissert, D. R., Gomez-Tagle, A., Vache, K., & Dawson, T. E. (2011). Ecohydrological advances and applications in plant-water relations research: a review. *Journal of Plant Ecology*, *4*(1–2), 3–22. https://doi.org/10.1093/jpe/rtr005

Asharaf, S., Dobler, A., & Ahrens, B. (2012). Soil Moisture–Precipitation Feedback Processes in the Indian Summer Monsoon Season. *Journal of Hydrometeorology*, *13*(5), 1461–1474. https://doi.org/10.1175/JHM-D-12-06.1

Babaeian, E., Sadeghi, M., Jones, S. B., Montzka, C., Vereecken, H., & Tuller, M. (2019). Ground, Proximal, and Satellite Remote Sensing of Soil Moisture. *Reviews of Geophysics*, *57*(2), 530–616. https://doi.org/10.1029/2018RG000618

Baldocchi, D. (2014). Measuring fluxes of trace gases and energy between ecosystems and the atmosphere - the state and future of the eddy covariance method. *Global Change Biology*, *20*(12), 3600–3609. https://doi.org/10.1111/gcb.12649

Baldocchi, D. D. (2020). How eddy covariance flux measurements have contributed to our understanding of Global Change Biology. *Global Change Biology*, *26*(1), 242–260. https://doi.org/10.1111/gcb.14807

Barnett, D. T., Adler, P. B., Chemel, B. R., Duffy, P. A., Enquist, B. J., Grace, J. B., Harrison, S., Peet, R. K., Schimel, D. S., Stohlgren, T. J., & Vellend, M. (2019). The plant diversity sampling design for The National Ecological Observatory Network. *Ecosphere*, *10*(2). https://doi.org/10.1002/ecs2.2603

Bassiouni, M., Good, S. P., Still, C. J., & Higgins, C. W. (2020). Plant Water Uptake Thresholds Inferred From Satellite Soil Moisture. *Geophysical Research Letters*, *47*(7). https://doi.org/10.1029/2020GL087077

Bauska, T. K., Baggenstos, D., Brook, E. J., Mix, A. C., Marcott, S. A., Petrenko, V. V., Schaefer, H., Severinghaus, J. P., & Lee, J. E. (2016). Carbon isotopes characterize rapid changes in atmospheric carbon dioxide during the last deglaciation. *Proceedings of the National Academy of Sciences*, *113*(13), 3465–3470. https://doi.org/10.1073/pnas.1513868113

Bell, J. E., Leeper, R. D., Palecki, M. A., Coopersmith, E., Wilson, T., Bilotta, R., & Embler, S. (2015). Evaluation of the 2012 Drought with a Newly Established National Soil Monitoring Network. *Vadose Zone Journal*, *14*(11), vzj2015.02.0023. https://doi.org/10.2136/vzj2015.02.0023

Bell, J. E., Palecki, M. A., Baker, C. B., Collins, W. G., Lawrimore, J. H., Leeper, R. D., Hall, M. E., Kochendorfer, J., Meyers, T. P., Wilson, T., & Diamond, H. J. (2013). U.S. Climate Reference Network Soil Moisture and Temperature Observations. *Journal of Hydrometeorology*, *14*(3), 977–988. https://doi.org/10.1175/JHM-D-12-0146.1

Boke-Olén, N., Ardö, J., Eklundh, L., Holst, T., & Lehsten, V. (2018). Remotely sensed soil moisture to estimate savannah NDVI. *PLOS ONE*, *13*(7), e0200328. https://doi.org/10.1371/journal.pone.0200328

Bonan Li. (2020). Information–based uncertainty decomposition in dual channel microwave remote sensing of soil moisture. *Hydrology and Earth System Sciences Discussion*. https://doi.org/https://doi.org/10.5194/hess-2020-534

Bowen, G. J., Cai, Z., Fiorella, R. P., & Putman, A. L. (2019). Isotopes in the Water Cycle: Regional- to Global-Scale Patterns and Applications. *Annual Review of Earth and Planetary Sciences*, *47*(1), 453–479. https://doi.org/10.1146/annurev-earth-053018-060220

Box, G. E. P. (1976). Science and Statistics. *Journal of the American Statistical Association*, *71*(356), 791–799. https://doi.org/10.1080/01621459.1976.10480949

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. (2011). Relationships between phenology, radiation and precipitation in the Amazon region. *Global Change Biology*, *17*(6), 2245–2260. https://doi.org/10.1111/j.1365-2486.2011.02405.x

Brady, E., Stevenson, S., Bailey, D., Liu, Z., Noone, D., Nusbaumer, J., Otto-Bliesner, B. L., Tabor, C., Tomas, R., Wong, T., Zhang, J., & Zhu, J. (2019). The Connected Isotopic Water Cycle in the Community Earth System Model Version 1. *Journal of Advances in Modeling Earth Systems*, *11*(8), 2547–2566. https://doi.org/10.1029/2019MS001663

Breiman, L. (2001). Rndom Forests. *Macchine Learning*, *45*, 5–32. https://doi.org/https://doi.org/10.1023/A:1010933404324

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. (2017). A Comparative Study of the SMAP Passive Soil Moisture Product With Existing Satellite-Based Soil Moisture Products. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(5), 2959–2971. https://doi.org/10.1109/TGRS.2017.2656859

Chan, S. (2020). Soil Moisture Active Passive (SMAP) Level 2 Passive Soil Moisture Product Specification Documen. *Jet Propulsion Laboratory California Institute of Technology, Pasadena, USA, JPL D-72547 (Version 7.0)*.

Chan, S. K., Bindlish, R., O'Neill, P. E., Njoku, E., Jackson, T., Colliander, A., Chen, F., Burgin, M., Dunbar, S., Piepmeier, J., Yueh, S., Entekhabi, D., Cosh, M. H., Caldwell, T., Walker, J., Wu, X., Berg, A., Rowlandson, T., Pacheco, A., … Kerr, Y. (2016). Assessment of the SMAP Passive Soil Moisture Product. *IEEE Transactions on Geoscience and Remote Sensing*, *54*(8), 4994–5007. https://doi.org/10.1109/TGRS.2016.2561938

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., Berg, A. A., Bosch, D. D., Caldwell, T., Cosh, M. H., Holifield Collins, C., Martinez-Fernandez, J., Seyfried, M., Starks, P. J., Su, Z., … Walker, J. (2020). Improved SMAP Dual-Channel Algorithm for the Retrieval of Soil

Moisture. *IEEE Transactions on Geoscience and Remote Sensing*, *58*(6), 3894–3905. https://doi.org/10.1109/TGRS.2019.2959239

Chen, F., Crow, W. T., Colliander, A., Cosh, M. H., Jackson, T. J., Bindlish, R., Reichle, R. H., Chan, S. K., Bosch, D. D., Starks, P. J., Goodrich, D. C., & Seyfried, M. S. (2017). Application of Triple Collocation in Ground-Based Validation of Soil Moisture Active/Passive (SMAP) Level 2 Data Products. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *10*(2), 489–502. https://doi.org/10.1109/JSTARS.2016.2569998

Chen, J., Wen, J., Kang, S., Meng, X., Tian, H., Ma, X., & Yuan, Y. (2020). Assessments of the factors controlling latent heat flux and the coupling degree between an alpine wetland and the atmosphere on the Qinghai-Tibetan Plateau in summer. *Atmospheric Research*, *240*, 104937. https://doi.org/10.1016/j.atmosres.2020.104937

Chen, M., Willgoose, G. R., & Saco, P. M. (2015). Investigating the impact of leaf area index temporal variability on soil moisture predictions using remote sensing vegetation data. *Journal of Hydrology*, *522*, 274–284. https://doi.org/10.1016/j.jhydrol.2014.12.027

Colliander, A., Jackson, T. J., Bindlish, R., Chan, S., Das, N., Kim, S. B., Cosh, M. H., Dunbar, R. S., Dang, L., Pashaian, L., Asanuma, J., Aida, K., Berg, A., Rowlandson, T., Bosch, D., Caldwell, T., Caylor, K., Goodrich, D., al Jassar, H., … Yueh, S. (2017a). Validation of SMAP surface soil moisture products with core validation sites. *Remote Sensing of Environment*, *191*, 215–231. https://doi.org/10.1016/j.rse.2017.01.021

Colliander, A., Jackson, T. J., Bindlish, R., Chan, S., Das, N., Kim, S. B., Cosh, M. H., Dunbar, R. S., Dang, L., Pashaian, L., Asanuma, J., Aida, K., Berg, A., Rowlandson, T., Bosch, D., Caldwell, T., Caylor, K., Goodrich, D., al Jassar, H., … Yueh, S. (2017b). Validation of SMAP surface soil moisture products with core validation sites. *Remote Sensing of Environment*, *191*, 215–231. https://doi.org/10.1016/j.rse.2017.01.021

Colliander, Andreas, Cosh, M. H., Misra, S., Jackson, T. J., Crow, W. T., Chan, S., Bindlish, R., Chae, C., Holifield Collins, C., & Yueh, S. H. (2017). Validation and scaling of soil moisture in a semi-arid environment: SMAP validation experiment 2015 (SMAPVEX15). *Remote Sensing of Environment*, *196*, 101–112. https://doi.org/10.1016/j.rse.2017.04.022

Conrad, R., Klose, M., Yuan, Q., Lu, Y., & Chidthaisong, A. (2012). Stable carbon isotope fractionation, carbon flux partitioning and priming effects in anoxic soils during methanogenic degradation of straw and soil organic matter. *Soil Biology and Biochemistry*, *49*, 193–199. https://doi.org/10.1016/j.soilbio.2012.02.030

Cover, T. M., & Thomas, J. A. (2005). *Elements of Information Theory*. Wiley. https://doi.org/10.1002/047174882X

Cowling, S. A., & Field, C. B. (2003). Environmental control of leaf area production: Implications for vegetation and land-surface modeling. *Global Biogeochemical Cycles*, *17*(1), 7-1-7–14. https://doi.org/10.1029/2002GB001915

Crow, W. T., Berg, A. A., Cosh, M. H., Loew, A., Mohanty, B. P., Panciera, R., de Rosnay, P., Ryu, D., & Walker, J. P. (2012). Upscaling sparse ground-based soil moisture observations for the validation of coarse-resolution satellite soil moisture products. *Reviews of Geophysics*, *50*(2). https://doi.org/10.1029/2011RG000372

Dadap, N. C., Cobb, A. R., Hoyt, A. M., Harvey, C. F., & Konings, A. G. (2019). Satellite soil moisture observations predict burned area in Southeast Asian peatlands. *Environmental Research Letters*, *14*(9), 094014. https://doi.org/10.1088/1748-9326/ab3891

Darvishzadeh, R., Skidmore, A., Schlerf, M., Atzberger, C., Corsi, F., & Cho, M. (2008). LAI and chlorophyll estimation for a heterogeneous grassland using hyperspectral measurements. *ISPRS Journal of Photogrammetry and Remote Sensing*, *63*(4), 409–426. https://doi.org/10.1016/j.isprsjprs.2008.01.001

Davarzani, H., Smits, K., Tolene, R. M., & Illangasekare, T. (2014). Study of the effect of wind speed on evaporation from soil through integrated modeling of the atmospheric boundary layer and shallow subsurface. *Water Resources Research*, *50*(1), 661–680. https://doi.org/10.1002/2013WR013952

Delbart, N., Kergoat, L., Le Toan, T., Lhermitte, J., & Picard, G. (2005). Determination of phenological dates in boreal regions using normalized difference water index. *Remote Sensing of Environment*, *97*(1), 26–38. https://doi.org/10.1016/j.rse.2005.03.011

Diamond, H. J., Karl, T. R., Palecki, M. A., Baker, C. B., Bell, J. E., Leeper, R. D., Easterling, D. R., Lawrimore, J. H., Meyers, T. P., Helfert, M. R., Goodge, G., & Thorne, P. W. (2013). U.S. Climate Reference Network after One Decade of Operations: Status and Assessment. *Bulletin of the American Meteorological Society*, *94*(4), 485–498. https://doi.org/10.1175/BAMS-D-12-00170.1

Dubayah, R., Blair, J. B., Goetz, S., Fatoyinbo, L., Hansen, M., Healey, S., Hofton, M., Hurtt, G., Kellner, J., Luthcke, S., Armston, J., Tang, H., Duncanson, L., Hancock, S., Jantz, P., Marselis, S., Patterson, P. L., Qi, W., & Silva, C. (2020). The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth's forests and topography. *Science of Remote Sensing*, *1*, 100002. https://doi.org/10.1016/j.srs.2020.100002

Dube, T., Pandit, S., Shoko, C., Ramoelo, A., Mazvimavi, D., & Dalu, T. (2019). Numerical Assessments of Leaf Area Index in Tropical Savanna Rangelands, South Africa Using Landsat 8 OLI Derived Metrics and In-Situ Measurements. *Remote Sensing*, *11*(7), 829. https://doi.org/10.3390/rs11070829

El Hajj, M., Baghdadi, N., Bazzi, H., & Zribi, M. (2018). Penetration Analysis of SAR Signals in the C and L Bands for Wheat, Maize, and Grasslands. *Remote Sensing*, *11*(1), 31. https://doi.org/10.3390/rs11010031

Elmendorf, S. C., Jones, K. D., Cook, B. I., Diez, J. M., Enquist, C. A. F., Hufft, R. A., Jones, M. O., Mazer, S. J., Miller-Rushing, A. J., Moore, D. J. P., Schwartz, M. D., & Weltzin, J. F. (2016). The plant phenology monitoring design for The National Ecological Observatory Network. *Ecosphere*, *7*(4), e01303. https://doi.org/10.1002/ecs2.1303

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., Kimball, J., Piepmeier, J. R., Koster, R. D., Martin, N., McDonald, K. C., Moghaddam, M., Moran, S., Reichle, R., Shi, J. C., … Van Zyl, J. (2010a). The Soil Moisture Active Passive (SMAP) Mission. *Proceedings of the IEEE*, *98*(5), 704–716. https://doi.org/10.1109/JPROC.2010.2043918

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., Kimball, J., Piepmeier, J. R., Koster, R. D., Martin, N., McDonald, K. C., Moghaddam, M., Moran, S., Reichle, R., Shi, J. C., …

Van Zyl, J. (2010b). The Soil Moisture Active Passive (SMAP) Mission. *Proceedings of the IEEE*, *98*(5), 704–716. https://doi.org/10.1109/JPROC.2010.2043918

Escorihuela, M. J., Chanzy, A., Wigneron, J. P., & Kerr, Y. H. (2010). Effective soil moisture sampling depth of L-band radiometry: A case study. *Remote Sensing of Environment*, *114*(5), 995–1001. https://doi.org/10.1016/j.rse.2009.12.011

Esposito, M., Achterberg, E. P., Bach, L. T., Connelly, D. P., Riebesell, U., & Taucher, J. (2019). Application of Stable Carbon Isotopes in a Subtropical North Atlantic MesocosmStudy: A New Approach to Assess CO2 Effects on the Marine Carbon Cycle. *Frontiers in Marine Science*, *6*. https://doi.org/10.3389/fmars.2019.00616

Fang, H., Baret, F., Plummer, S., & Schaepman-Strub, G. (2019). An Overview of Global Leaf Area Index (LAI): Methods, Products, Validation, and Applications. *Reviews of Geophysics*, *57*(3), 739–799. https://doi.org/10.1029/2018RG000608

Fang, H., Wei, S., & Liang, S. (2012). Validation of MODIS and CYCLOPES LAI products using global field measurement data. *Remote Sensing of Environment*, *119*, 43–54. https://doi.org/10.1016/j.rse.2011.12.006

Feldman, A. F., Short Gianotti, D. J., Konings, A. G., McColl, K. A., Akbar, R., Salvucci, G. D., & Entekhabi, D. (2018). Moisture pulse-reserve in the soil-plant continuum observed across biomes. *Nature Plants*, *4*(12), 1026–1033. https://doi.org/10.1038/s41477-018-0304-9

Ferguson, P. R., & Veizer, J. (2007). Coupling of water and carbon fluxes via the terrestrial biosphere and its significance to the Earth's climate system. *Journal of Geophysical Research*, *112*(D24), D24S06. https://doi.org/10.1029/2007JD008431

Finn, C., & Lizier, J. (2018). Pointwise Partial Information Decomposition Using the Specificity and Ambiguity Lattices. *Entropy*, *20*(4), 297. https://doi.org/10.3390/e20040297

Fiorella, R. P., Good, S. P., Allen, S. T., Guo, J. S., Still, C. J., Noone, D. C., Anderegg, W. R. L., Florian, C. R., Luo, H., Pingintha-Durden, N., & Bowen, G. J. (2021). Calibration Strategies for Detecting Macroscale Patterns in NEON Atmospheric Carbon Isotope Observations. *Journal of Geophysical Research: Biogeosciences*, *126*(3). https://doi.org/10.1029/2020JG005862

Fischer, C., Leimer, S., Roscher, C., Ravenek, J., de Kroon, H., Kreutziger, Y., Baade, J., Beßler, H., Eisenhauer, N., Weigelt, A., Mommer, L., Lange, M., Gleixner, G., Wilcke, W., Schröder, B., & Hildebrandt, A. (2019). Plant species richness and functional groups have different effects on soil water content in a decade-long grassland experiment. *Journal of Ecology*, *107*(1), 127–141. https://doi.org/10.1111/1365-2745.13046

Fisher, J. B., Tu, K. P., & Baldocchi, D. D. (2008). Global estimates of the land–atmosphere water flux based on monthly AVHRR and ISLSCP-II data, validated at 16 FLUXNET sites. *Remote Sensing of Environment*, *112*(3), 901–919. https://doi.org/10.1016/j.rse.2007.06.025

Foken, T., Aubinet, M., & Leuning, R. (2012). The Eddy Covariance Method. In *Eddy Covariance* (pp. 1–19). Springer Netherlands. https://doi.org/10.1007/978-94-007-2351-1_1

Freedman, D., & Diaconis, P. (1981). On the histogram as a density estimator:L 2 theory. *Zeitschrift Für Wahrscheinlichkeitstheorie Und Verwandte Gebieter*, *57*(4), 453–476. https://doi.org/10.1007/BF01025868

Gao, Lin, Wang, X., Johnson, B. A., Tian, Q., Wang, Y., Verrelst, J., Mu, X., & Gu, X. (2020). Remote sensing algorithms for estimation of fractional vegetation cover using pure

vegetation index values: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, *159*, 364–377. https://doi.org/10.1016/j.isprsjprs.2019.11.018

Gao, Lun, Sadeghi, M., Ebtehaj, A., & Wigneron, J.-P. (2020). A temporal polarization ratio algorithm for calibration-free retrieval of soil moisture at L-band. *Remote Sensing of Environment*, *249*, 112019. https://doi.org/10.1016/j.rse.2020.112019

Gebler, S., Hendricks Franssen, H.-J., Pütz, T., Post, H., Schmidt, M., & Vereecken, H. (2015). Actual evapotranspiration and precipitation measured by lysimeters: a comparison with eddy covariance and tipping bucket. *Hydrology and Earth System Sciences*, *19*(5), 2145–2161. https://doi.org/10.5194/hess-19-2145-2015

Gemery, P. A., Trolier, M., & White, J. W. C. (1996). Oxygen isotope exchange between carbon dioxide and water following atmospheric sampling using glass flasks. *Journal of Geophysical Research: Atmospheres*, *101*(D9), 14415–14420. https://doi.org/10.1029/96JD00053

Gentine, P., Green, J. K., Guérin, M., Humphrey, V., Seneviratne, S. I., Zhang, Y., & Zhou, S. (2019). Coupling between the terrestrial carbon and water cycles—a review. *Environmental Research Letters*, *14*(8), 083003. https://doi.org/10.1088/1748-9326/ab22d6

Gerard, F. F., George, C. T., Hayman, G., Chavana-Bryant, C., & Weedon, G. P. (2020). Leaf phenology amplitude derived from MODIS NDVI and EVI: Maps of leaf phenology synchrony for Meso- and South America. *Geoscience Data Journal*, *7*(1), 13–26. https://doi.org/10.1002/gdj3.87

Gong, W., Gupta, H. V., Yang, D., Sricharan, K., & Hero, A. O. (2013). Estimating epistemic and aleatory uncertainties during hydrologic modeling: An information theoretic approach. *Water Resources Research*, *49*(4), 2253–2273. https://doi.org/10.1002/wrcr.20161

Gonsamo, A., & Chen, J. M. (2014). Improved LAI Algorithm Implementation to MODIS Data by Incorporating Background, Topography, and Foliage Clumping Information. *IEEE Transactions on Geoscience and Remote Sensing*, *52*(2), 1076–1088. https://doi.org/10.1109/TGRS.2013.2247405

Good, S. P., Soderberg, K., Guan, K., King, E. G., Scanlon, T. M., & Caylor, K. K. (2014). δ 2 H isotopic flux partitioning of evapotranspiration over a grass field following a water pulse and subsequent dry down. *Water Resources Research*, *50*(2), 1410–1432. https://doi.org/10.1002/2013WR014333

Goodwell, A. E., & Kumar, P. (2017). Temporal information partitioning: Characterizing synergy, uniqueness, and redundancy in interacting environmental variables. *Water Resources Research*, *53*(7), 5920–5942. https://doi.org/10.1002/2016WR020216

Goodwell, A. E., Kumar, P., Fellows, A. W., & Flerchinger, G. N. (2018). Dynamic process connectivity explains ecohydrologic responses to rainfall pulses and drought. *Proceedings of the National Academy of Sciences*, *115*(37), E8604–E8613. https://doi.org/10.1073/pnas.1800236115

Gornall, J., Betts, R., Burke, E., Clark, R., Camp, J., Willett, K., & Wiltshire, A. (2010). Implications of climate change for agricultural productivity in the early twenty-first century. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1554), 2973–2989. https://doi.org/10.1098/rstb.2010.0158

Grant, J. P., Wigneron, J.-P., De Jeu, R. A. M., Lawrence, H., Mialon, A., Richaume, P., Al Bitar, A., Drusch, M., van Marle, M. J. E., & Kerr, Y. (2016). Comparison of SMOS and AMSR-E vegetation optical depth to four MODIS-based vegetation indices. *Remote Sensing of Environment*, *172*, 87–100. https://doi.org/10.1016/j.rse.2015.10.021

Grossiord, C., Buckley, T. N., Cernusak, L. A., Novick, K. A., Poulter, B., Siegwolf, R. T. W., Sperry, J. S., & McDowell, N. G. (2020). Plant responses to rising vapor pressure deficit. *New Phytologist*, *226*(6), 1550–1566. https://doi.org/10.1111/nph.16485

Gruber, A., De Lannoy, G., Albergel, C., Al-Yaari, A., Brocca, L., Calvet, J.-C., Colliander, A., Cosh, M., Crow, W., Dorigo, W., Draper, C., Hirschi, M., Kerr, Y., Konings, A., Lahoz, W., McColl, K., Montzka, C., Muñoz-Sabater, J., Peng, J., … Wagner, W. (2020). Validation practices for satellite soil moisture retrievals: What are (the) errors? *Remote Sensing of Environment*, *244*, 111806. https://doi.org/10.1016/j.rse.2020.111806

Gu, L., Meyers, T., Pallardy, S. G., Hanson, P. J., Yang, B., Heuer, M., Hosman, K. P., Riggs, J. S., Sluss, D., & Wullschleger, S. D. (2006). Direct and indirect effects of atmospheric conditions and soil moisture on surface energy partitioning revealed by a prolonged drought at a temperate forest site. *Journal of Geophysical Research*, *111*(D16), D16102. https://doi.org/10.1029/2006JD007161

Hohenegger, C., Brockhaus, P., Bretherton, C. S., & Schär, C. (2009). The Soil Moisture–Precipitation Feedback in Simulations with Explicit and Parameterized Convection. *Journal of Climate*, *22*(19), 5003–5020. https://doi.org/10.1175/2009JCLI2604.1

HOUBORG, R., & BOEGH, E. (2008). Mapping leaf chlorophyll and leaf area index using inverse and forward canopy reflectance modeling and SPOT reflectance data. *Remote Sensing of Environment*, *112*(1), 186–202. https://doi.org/10.1016/j.rse.2007.04.012

Houborg, R., & McCabe, M. F. (2018). A hybrid training approach for leaf area index estimation via Cubist and random forests machine-learning. *ISPRS Journal of Photogrammetry and Remote Sensing*, *135*, 173–188. https://doi.org/10.1016/j.isprsjprs.2017.10.004

Hsu, H., Lo, M.-H., Guillod, B. P., Miralles, D. G., & Kumar, S. (2017). Relation between precipitation location and antecedent/subsequent soil moisture spatial patterns. *Journal of Geophysical Research: Atmospheres*, *122*(12), 6319–6328. https://doi.org/10.1002/2016JD026042

Hulley, G., Hook, S., Fisher, J., & Lee, C. (2017). ECOSTRESS, A NASA Earth-Ventures Instrument for studying links between the water cycle and plant health over the diurnal cycle. *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 5494–5496. https://doi.org/10.1109/IGARSS.2017.8128248

Jackson, T. J., Bindlish, R., Cosh, M. H., Zhao, T., Starks, P. J., Bosch, D. D., Seyfried, M., Moran, M. S., Goodrich, D. C., Kerr, Y. H., & Leroux, D. (2012). Validation of Soil Moisture and Ocean Salinity (SMOS) Soil Moisture Over Watershed Networks in the U.S. *IEEE Transactions on Geoscience and Remote Sensing*, *50*(5), 1530–1543. https://doi.org/10.1109/TGRS.2011.2168533

Jackson, T. J., Schmugge, T. J., & Wang, J. R. (1982). Passive microwave sensing of soil moisture under vegetation canopies. *Water Resources Research*, *18*(4), 1137–1142. https://doi.org/10.1029/WR018i004p01137

Jia, Z., Liu, S., Xu, Z., Chen, Y., & Zhu, M. (2012). Validation of remotely sensed evapotranspiration over the Hai River Basin, China. *Journal of Geophysical Research: Atmospheres*, *117*(D13), n/a-n/a. https://doi.org/10.1029/2011JD017037

Ju, J., & Masek, J. G. (2016). The vegetation greenness trend in Canada and US Alaska from 1984–2012 Landsat data. *Remote Sensing of Environment*, *176*, 1–16. https://doi.org/10.1016/j.rse.2016.01.001

JUSTICE, C. O., TOWNSHEND, J. R. G., HOLBEN, B. N., & TUCKER, C. J. (1985). Analysis of the phenology of global vegetation using meteorological satellite data. *International Journal of Remote Sensing*, *6*(8), 1271–1318. https://doi.org/10.1080/01431168508948281

Kandasamy, S., Baret, F., Verger, A., Neveux, P., & Weiss, M. (2013). A comparison of methods for smoothing and gap filling time series of remote sensing observations – application to MODIS LAI products. *Biogeosciences*, *10*(6), 4055–4071. https://doi.org/10.5194/bg-10-4055-2013

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

Kerr, Y. H., Waldteufel, P., Richaume, P., Wigneron, J. P., Ferrazzoli, P., Mahmoodi, A., Al Bitar, A., Cabot, F., Gruhier, C., Juglea, S. E., Leroux, D., Mialon, A., & Delwart, S. (2012). The SMOS Soil Moisture Retrieval Algorithm. *IEEE Transactions on Geoscience and Remote Sensing*, *50*(5), 1384–1403. https://doi.org/10.1109/TGRS.2012.2184548

Konings, A. G., McColl, K. A., Piles, M., & Entekhabi, D. (2015). How Many Parameters Can Be Maximally Estimated From a Set of Measurements? *IEEE Geoscience and Remote Sensing Letters*, *12*(5), 1081–1085. https://doi.org/10.1109/LGRS.2014.2381641

Konings, A. G., Piles, M., Das, N., & Entekhabi, D. (2017). L-band vegetation optical depth and effective scattering albedo estimation from SMAP. *Remote Sensing of Environment*, *198*, 460–470. https://doi.org/10.1016/j.rse.2017.06.037

Konings, A. G., Piles, M., Rötzer, K., McColl, K. A., Chan, S. K., & Entekhabi, D. (2016). Vegetation optical depth and scattering albedo retrieval using time series of dual-polarized L-band radiometer observations. *Remote Sensing of Environment*, *172*, 178–189. https://doi.org/10.1016/j.rse.2015.11.009

Korhonen, L., Hadi, Packalen, P., & Rautiainen, M. (2017). Comparison of Sentinel-2 and Landsat 8 in the estimation of boreal forest canopy cover and leaf area index. *Remote Sensing of Environment*, *195*, 259–274. https://doi.org/10.1016/j.rse.2017.03.021

Kraskov, A., Stögbauer, H., & Grassberger, P. (2004). Estimating mutual information. *Physical Review E*, *69*(6), 066138. https://doi.org/10.1103/PhysRevE.69.066138

Krause, S., Lewandowski, J., Dahm, C. N., & Tockner, K. (2015). Frontiers in real-time ecohydrology - a paradigm shift in understanding complex environmental systems. *Ecohydrology*, *8*(4), 529–537. https://doi.org/10.1002/eco.1646

Kuang, X., Luo, X., Jiao, J. J., Liang, S., Zhang, X., Li, H., & Liu, J. (2019). Using stable isotopes of surface water and groundwater to quantify moisture sources across the Yellow River source region. *Hydrological Processes*, hyp.13441. https://doi.org/10.1002/hyp.13441

Kumar, S. V., M. Mocko, D., Wang, S., Peters-Lidard, C. D., & Borak, J. (2019). Assimilation of Remotely Sensed Leaf Area Index into the Noah-MP Land Surface Model: Impacts on

Water and Carbon Fluxes and States over the Continental United States. *Journal of Hydrometeorology*, *20*(7), 1359–1377. https://doi.org/10.1175/JHM-D-18-0237.1

Kunert-Graf, J., Sakhanenko, N., & Galas, D. (2020). Partial Information Decomposition and the Information Delta: A Geometric Unification Disentangling Non-Pairwise Information. *Entropy*, *22*(12), 1333. https://doi.org/10.3390/e22121333

Kwa, C. (2005). Local Ecologies and Global Science. *Social Studies of Science*, *35*(6), 923–950. https://doi.org/10.1177/0306312705052100

Lawrence, H., Wigneron, J.-P., Richaume, P., Novello, N., Grant, J., Mialon, A., Al Bitar, A., Merlin, O., Guyon, D., Leroux, D., Bircher, S., & Kerr, Y. (2014). Comparison between SMOS Vegetation Optical Depth products and MODIS vegetation indices over crop zones of the USA. *Remote Sensing of Environment*, *140*, 396–406. https://doi.org/10.1016/j.rse.2013.07.021

Lee, S.-C., Christen, A., Black, T. A., Jassal, R. S., Ketler, R., & Nesic, Z. (2020). Partitioning of net ecosystem exchange into photosynthesis and respiration using continuous stable isotope measurements in a Pacific Northwest Douglas-fir forest ecosystem. *Agricultural and Forest Meteorology*, *292–293*, 108109. https://doi.org/10.1016/j.agrformet.2020.108109

Leeper, R. D., Bell, J. E., Vines, C., & Palecki, M. (2017). An Evaluation of the North American Regional Reanalysis Simulated Soil Moisture Conditions during the 2011–13 Drought Period. *Journal of Hydrometeorology*, *18*(2), 515–527. https://doi.org/10.1175/JHM-D-16-0132.1

LI, R., LI, C., DONG, Y., LIU, F., WANG, J., YANG, X., & PAN, Y. (2011). Assimilation of Remote Sensing and Crop Model for LAI Estimation Based on Ensemble Kaiman Filter. *Agricultural Sciences in China*, *10*(10), 1595–1602. https://doi.org/10.1016/S1671-2927(11)60156-9

Li, Sien, Kang, S., Li, F., & Zhang, L. (2008). Evapotranspiration and crop coefficient of spring maize with plastic mulch using eddy covariance in northwest China. *Agricultural Water Management*, *95*(11), 1214–1222. https://doi.org/10.1016/j.agwat.2008.04.014

Li, Songyang, Yuan, F., Ata-UI-Karim, S. T., Zheng, H., Cheng, T., Liu, X., Tian, Y., Zhu, Y., Cao, W., & Cao, Q. (2019). Combining Color Indices and Textures of UAV-Based Digital Imagery for Rice LAI Estimation. *Remote Sensing*, *11*(15), 1763. https://doi.org/10.3390/rs11151763

Li, Zhenwang, Wang, J., Tang, H., Huang, C., Yang, F., Chen, B., Wang, X., Xin, X., & Ge, Y. (2016). Predicting Grassland Leaf Area Index in the Meadow Steppes of Northern China: A Comparative Study of Regression Approaches and Hybrid Geostatistical Methods. *Remote Sensing*, *8*(8), 632. https://doi.org/10.3390/rs8080632

Li, Zongxing, Gui, J., Wang, X., Feng, Q., Zhao, T., Ouyang, C., Guo, X., Zhang, B., & Shi, Y. (2019). Water resources in inland regions of central Asia: Evidence from stable isotope tracing. *Journal of Hydrology*, *570*, 1–16. https://doi.org/10.1016/j.jhydrol.2019.01.003

Liu, Q., Piao, S., Janssens, I. A., Fu, Y., Peng, S., Lian, X., Ciais, P., Myneni, R. B., Peñuelas, J., & Wang, T. (2018). Extension of the growing season increases vegetation exposure to frost. *Nature Communications*, *9*(1), 426. https://doi.org/10.1038/s41467-017-02690-y

Liu, R., Wen, J., Wang, X., Wang, Z., Li, Z., Xie, Y., Zhu, L., & Li, D. (2019). Derivation of Vegetation Optical Depth and Water Content in the Source Region of the Yellow River

using the FY-3B Microwave Data. *Remote Sensing*, *11*(13), 1536. https://doi.org/10.3390/rs11131536

Lu, X., Liang, L. L., Wang, L., Jenerette, G. D., McCabe, M. F., & Grantz, D. A. (2017). Partitioning of evapotranspiration using a stable isotope technique in an arid and high temperature agricultural production system. *Agricultural Water Management*, *179*, 103–109. https://doi.org/10.1016/j.agwat.2016.08.012

Lupascu, M., Akhtar, H., Smith, T. E. L., & Sukri, R. S. (2020). Post-fire carbon dynamics in the tropical peat swamp forests of Brunei reveal long-term elevated CH 4 flux. *Global Change Biology*, *26*(9), 5125–5145. https://doi.org/10.1111/gcb.15195

Martinez, B., Cassiraga, E., Camacho, F., & Garcia-Haro, J. (2010). Geostatistics for Mapping Leaf Area Index over a Cropland Landscape: Efficiency Sampling Assessment. *Remote Sensing*, *2*(11), 2584–2606. https://doi.org/10.3390/rs2112584

MILLER, J. B., & TANS, P. P. (2003). Calculating isotopic fractionation from atmospheric measurements at various scales. *Tellus B*, *55*(2), 207–214. https://doi.org/10.1034/j.1600-0889.2003.00020.x

Mo, T., Choudhury, B. J., Schmugge, T. J., Wang, J. R., & Jackson, T. J. (1982). A model for microwave emission from vegetation-covered fields. *Journal of Geophysical Research*, *87*(C13), 11229. https://doi.org/10.1029/JC087iC13p11229

Mohanty, B. P., Cosh, M. H., Lakshmi, V., & Montzka, C. (2017). Soil Moisture Remote Sensing: State-of-the-Science. *Vadose Zone Journal*, *16*(1), vzj2016.10.0105. https://doi.org/10.2136/vzj2016.10.0105

Momen, M., Wood, J. D., Novick, K. A., Pangle, R., Pockman, W. T., McDowell, N. G., & Konings, A. G. (2017). Interacting Effects of Leaf Water Potential and Biomass on Vegetation Optical Depth. *Journal of Geophysical Research: Biogeosciences*, *122*(11), 3031–3046. https://doi.org/10.1002/2017JG004145

Myneni, R., Knyazikhin, Y., & Park, T. (2015). MCD15A3H MODIS/Terra+Aqua Leaf Area Index/FPAR 4-day L4 Global 500m SIN Grid V006. *NASA EOSDIS Land Processes DAAC*. https://doi.org/https://doi.org/10.5067/MODIS/MCD15A3H.006

National Ecological Observatory Network (NEON). (2021a). *Bundled data products - eddy covariance (DP4.00200.001)*. National Ecological Observatory Network (NEON). https://doi.org/10.48443/BWAY-HC74

National Ecological Observatory Network (NEON). (2021b). *Relative humidity (DP1.00098.001)*. National Ecological Observatory Network (NEON). https://doi.org/10.48443/S9WY-T644

Niu, S., Luo, Y., Fei, S., Yuan, W., Schimel, D., Law, B. E., Ammann, C., Altaf Arain, M., Arneth, A., Aubinet, M., Barr, A., Beringer, J., Bernhofer, C., Andrew Black, T., Buchmann, N., Cescatti, A., Chen, J., Davis, K. J., Dellwik, E., … Zhou, X. (2012). Thermal optimality of net ecosystem exchange of carbon dioxide and underlying mechanisms. *New Phytologist*, *194*(3), 775–783. https://doi.org/10.1111/j.1469-8137.2012.04095.x

Njoku, E. G., & Entekhabi, D. (1996). Passive microwave remote sensing of soil moisture. *Journal of Hydrology*, *184*(1–2), 101–129. https://doi.org/10.1016/0022-1694(95)02970-2

O'Neill, P., Bindlish, R., Chan, S., Njoku, E., and Jackson, T. (2020a). Algorithm theoretical basis document: Level 2 & 3 soil moisture (passive) data products. *Level 2 & 3 Soil*

Moisture (Passive) Data Products, Jet Propulsion Laboratory/California Institute of Technology, Pasadena, USA, JPL D-66480 (Revision F), 100.

O'Neill, P.E., Chan, S., Njoku, E. G., Jackson, T., Bindlish, R., & Chaubell, J. (2019). 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.* https://doi.org/https://doi.org/10.5067/017XZSKMLTT2

O'Neill, Peggy E., Chan, S. K., Njoku, E. G., Tom, J., Bindlish, R., & Chaubell, M. J. (2020b). SMAP L2 Radiometer Half-Orbit 36 km EASE-Grid Soil Moisture, Version 7. *[March 31 2015 to December 2020],Boulder, Colorado. USA. NASA National. Snow Ice Data Cent. Distrib. Act. Arch. Center [Date Accessed December 10 2020]*. https://doi.org/https://doi.org/10.5067/F1TZ0CBN1F5N

Omer, G., Mutanga, O., Abdel-Rahman, E., & Adam, E. (2016). 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 Sensing*, *8*(4), 324. https://doi.org/10.3390/rs8040324

Orlowski, N., Breuer, L., Angeli, N., Boeckx, P., Brumbt, C., Cook, C. S., Dubbert, M., Dyckmans, J., Gallagher, B., Gralher, B., Herbstritt, B., Hervé-Fernández, P., Hissler, C., Koeniger, P., Legout, A., Macdonald, C. J., Oyarzún, C., Redelstein, R., Seidler, C., … McDonnell, J. J. (2018). Inter-laboratory comparison of cryogenic water extraction systems for stable isotope analysis of soil water. *Hydrology and Earth System Sciences*, *22*(7), 3619–3637. https://doi.org/10.5194/hess-22-3619-2018

ORNL DAAC. (2018). MODIS and VIIRS Land Products Global Subsetting and Visualization Tool. *Available Online: Https://Modis.Ornl.Gov/Globalsubset/ (Accessed on 24 February 2020)*. https://doi.org/https://doi.org/10.3334/ORNLDAAC/1379

Paninski, L. (2003). Estimation of Entropy and Mutual Information. *Neural Computation*, *15*(6), 1191–1253. https://doi.org/10.1162/089976603321780272

Papale, D., Reichstein, M., Aubinet, M., Canfora, E., Bernhofer, C., Kutsch, W., Longdoz, B., Rambal, S., Valentini, R., Vesala, T., & Yakir, D. (2006). Towards a standardized processing of Net Ecosystem Exchange measured with eddy covariance technique: algorithms and uncertainty estimation. *Biogeosciences*, *3*(4), 571–583. https://doi.org/10.5194/bg-3-571-2006

Pataki, D. E., Ehleringer, J. R., Flanagan, L. B., Yakir, D., Bowling, D. R., Still, C. J., Buchmann, N., Kaplan, J. O., & Berry, J. A. (2003). The application and interpretation of Keeling plots in terrestrial carbon cycle research. *Global Biogeochemical Cycles*, *17*(1). https://doi.org/10.1029/2001GB001850

Petropoulos, G. P., Ireland, G., & Barrett, B. (2015). Surface soil moisture retrievals from remote sensing: Current status, products & future trends. *Physics and Chemistry of the Earth, Parts A/B/C*, *83–84*, 36–56. https://doi.org/10.1016/j.pce.2015.02.009

PIAO, S., FANG, J., ZHOU, L., CIAIS, P., & ZHU, B. (2006). Variations in satellite-derived phenology in China's temperate vegetation. *Global Change Biology*, *12*(4), 672–685. https://doi.org/10.1111/j.1365-2486.2006.01123.x

Piao, S., Liu, Q., Chen, A., Janssens, I. A., Fu, Y., Dai, J., Liu, L., Lian, X., Shen, M., & Zhu, X. (2019). Plant phenology and global climate change: Current progresses and challenges. *Global Change Biology*, *25*(6), 1922–1940. https://doi.org/10.1111/gcb.14619

Probst, P., Wright, M. N., & Boulesteix, A. (2019). Hyperparameters and tuning strategies for random forest. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *9*(3). https://doi.org/10.1002/widm.1301

Pugnaire, F. I., Morillo, J. A., Peñuelas, J., Reich, P. B., Bardgett, R. D., Gaxiola, A., Wardle, D. A., & van der Putten, W. H. (2019). Climate change effects on plant-soil feedbacks and consequences for biodiversity and functioning of terrestrial ecosystems. *Science Advances*, *5*(11), eaaz1834. https://doi.org/10.1126/sciadv.aaz1834

Qu, Y., & Zhuang, Q. (2018). Modeling leaf area index in North America using a process-based terrestrial ecosystem model. *Ecosphere*, *9*(1). https://doi.org/10.1002/ecs2.2046

Raju, S., Chanzy, A., Wigneron, J.-P., Calvet, J.-C., Kerr, Y., & Laguerre, L. (1995). Soil moisture and temperature profile effects on microwave emission at low frequencies. *Remote Sensing of Environment*, *54*(2), 85–97. https://doi.org/10.1016/0034-4257(95)00133-L

Reichle, R., De Lannoy, G., Koster, R. D., Crow, W. T., Kimball, J. S., & Liu, Q. (2018). 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*.

Roosjen, P. P. J., Brede, B., Suomalainen, J. M., Bartholomeus, H. M., Kooistra, L., & Clevers, J. G. P. W. (2018). 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. *International Journal of Applied Earth Observation and Geoinformation*, *66*, 14–26. https://doi.org/10.1016/j.jag.2017.10.012

ROZANSKI, K., JOHNSEN, S. J., SCHOTTERER, U., & THOMPSON, L. G. (1997). Reconstruction of past climates from stable isotope records of palaeo-precipitation preserved in continental archives. *Hydrological Sciences Journal*, *42*(5), 725–745. https://doi.org/10.1080/02626669709492069

Santhana Vannan, S. K., Cook, R. B., Holladay, S. K., Olsen, L. M., Dadi, U., & Wilson, B. E. (2009). A Web-Based Subsetting Service for Regional Scale MODIS Land Products. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *2*(4), 319–328. https://doi.org/10.1109/JSTARS.2009.2036585

Scanlon, T. M., Schmidt, D. F., & Skaggs, T. H. (2019). Correlation-based flux partitioning of water vapor and carbon dioxide fluxes: Method simplification and estimation of canopy water use efficiency. *Agricultural and Forest Meteorology*, *279*, 107732. https://doi.org/10.1016/j.agrformet.2019.107732

Scholl, M. A., & Murphy, S. F. (2014). Precipitation isotopes link regional climate patterns to water supply in a tropical mountain forest, eastern Puerto Rico. *Water Resources Research*, *50*(5), 4305–4322. https://doi.org/10.1002/2013WR014413

Scott, D. W. (2010). Scott's rule. *Wiley Interdisciplinary Reviews: Computational Statistics*, *2*(4), 497–502. https://doi.org/10.1002/wics.103

Scott, R. L., & Biederman, J. A. (2017). Partitioning evapotranspiration using long-term carbon dioxide and water vapor fluxes. *Geophysical Research Letters*, *44*(13), 6833–6840. https://doi.org/10.1002/2017GL074324

Scranton, K., & Amarasekare, P. (2017). Predicting phenological shifts in a changing climate. *Proceedings of the National Academy of Sciences*, *114*(50), 13212–13217. https://doi.org/10.1073/pnas.1711221114

Seitzinger, S. P., Gaffney, O., Brasseur, G., Broadgate, W., Ciais, P., Claussen, M., Erisman, J. W., Kiefer, T., Lancelot, C., Monks, P. S., Smyth, K., Syvitski, J., & Uematsu, M. (2015). International Geosphere–Biosphere Programme and Earth system science: Three decades of co-evolution. *Anthropocene*, *12*, 3–16. https://doi.org/10.1016/j.ancene.2016.01.001

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Shellito, P. J., Small, E. E., Colliander, A., Bindlish, R., Cosh, M. H., Berg, A. A., Bosch, D. D., Caldwell, T. G., Goodrich, D. C., McNairn, H., Prueger, J. H., Starks, P. J., van der Velde, R., & Walker, J. P. (2016). SMAP soil moisture drying more rapid than observed in situ following rainfall events. *Geophysical Research Letters*, *43*(15), 8068–8075. https://doi.org/10.1002/2016GL069946

Stavi, I. (2019). Wildfires in Grasslands and Shrublands: A Review of Impacts on Vegetation, Soil, Hydrology, and Geomorphology. *Water*, *11*(5), 1042. https://doi.org/10.3390/w11051042

Su, Z. (2002). The Surface Energy Balance System (SEBS) for estimation of turbulent heat fluxes. *Hydrology and Earth System Sciences*, *6*(1), 85–100. https://doi.org/10.5194/hess-6-85-2002

Sulla-Menashe, D., Gray, J. M., Abercrombie, S. P., & Friedl, M. A. (2019). Hierarchical mapping of annual global land cover 2001 to present: The MODIS Collection 6 Land Cover product. *Remote Sensing of Environment*, *222*, 183–194. https://doi.org/10.1016/j.rse.2018.12.013

Sun, L., Gao, F., Anderson, M., Kustas, W., Alsina, M., Sanchez, L., Sams, B., McKee, L., Dulaney, W., White, W., Alfieri, J., Prueger, J., Melton, F., & Post, K. (2017). Daily Mapping of 30 m LAI and NDVI for Grape Yield Prediction in California Vineyards. *Remote Sensing*, *9*(4), 317. https://doi.org/10.3390/rs9040317

Tian, F., Brandt, M., Liu, Y. Y., Verger, A., Tagesson, T., Diouf, A. A., Rasmussen, K., Mbow, C., Wang, Y., & Fensholt, R. (2016). 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 Sensing of Environment*, *177*, 265–276. https://doi.org/10.1016/j.rse.2016.02.056

Tian, F., Wigneron, J.-P., Ciais, P., Chave, J., Ogée, J., Peñuelas, J., Ræbild, A., Domec, J.-C., Tong, X., Brandt, M., Mialon, A., Rodriguez-Fernandez, N., Tagesson, T., Al-Yaari, A., Kerr, Y., Chen, C., Myneni, R. B., Zhang, W., Ardö, J., & Fensholt, R. (2018). Coupling of ecosystem-scale plant water storage and leaf phenology observed by satellite. *Nature Ecology & Evolution*, *2*(9), 1428–1435. https://doi.org/10.1038/s41559-018-0630-3

Tong, C., Wang, H., Magagi, R., Goïta, K., Zhu, L., Yang, M., & Deng, J. (2020). Soil Moisture Retrievals by Combining Passive Microwave and Optical Data. *Remote Sensing*, *12*(19), 3173. https://doi.org/10.3390/rs12193173

Trenberth, K. E., Smith, L., Qian, T., Dai, A., & Fasullo, J. (2007). Estimates of the Global Water Budget and Its Annual Cycle Using Observational and Model Data. *Journal of Hydrometeorology*, *8*(4), 758–769. https://doi.org/10.1175/JHM600.1

Uber, M., Vandervaere, J.-P., Zin, I., Braud, I., Heistermann, M., Legoût, C., Molinié, G., & Nord, G. (2018). How does initial soil moisture influence the hydrological response? A case study from southern France. *Hydrology and Earth System Sciences*, *22*(12), 6127–6146. https://doi.org/10.5194/hess-22-6127-2018

Uda, S. (2020). Application of information theory in systems biology. *Biophysical Reviews*, *12*(2), 377–384. https://doi.org/10.1007/s12551-020-00665-w

Vargas, A. I., Schaffer, B., Yuhong, L., & Sternberg, L. da S. L. (2017). Testing plant use of mobile vs immobile soil water sources using stable isotope experiments. *New Phytologist*, *215*(2), 582–594. https://doi.org/10.1111/nph.14616

Veroustraete, F., Patyn, J., & Myneni, R. B. (1996). Estimating net ecosystem exchange of carbon using the normalized difference vegetation index and an ecosystem model. *Remote Sensing of Environment*, *58*(1), 115–130. https://doi.org/10.1016/0034-4257(95)00258-8

Vittucci, C., Vaglio Laurin, G., Tramontana, G., Ferrazzoli, P., Guerriero, L., & Papale, D. (2019). Vegetation optical depth at L-band and above ground biomass in the tropical range: Evaluating their relationships at continental and regional scales. *International Journal of Applied Earth Observation and Geoinformation*, *77*, 151–161. https://doi.org/10.1016/j.jag.2019.01.006

Vose, J. M., Sun, G., Ford, C. R., Bredemeier, M., Otsuki, K., Wei, X., Zhang, Z., & Zhang, L. (2011). Forest ecohydrological research in the 21st century: what are the critical needs? *Ecohydrology*, *4*(2), 146–158. https://doi.org/10.1002/eco.193

Wang, B., Zhang, H., Liang, X., Li, X., & Wang, F. (2019). Cumulative effects of cascade dams on river water cycle: Evidence from hydrogen and oxygen isotopes. *Journal of Hydrology*, *568*, 604–610. https://doi.org/10.1016/j.jhydrol.2018.11.016

Wang, C., Fu, B., Zhang, L., & Xu, Z. (2019). Soil moisture–plant interactions: an ecohydrological review. *Journal of Soils and Sediments*, *19*(1), 1–9. https://doi.org/10.1007/s11368-018-2167-0

Wang, K., & Dickinson, R. E. (2012). A review of global terrestrial evapotranspiration: Observation, modeling, climatology, and climatic variability. *Reviews of Geophysics*, *50*(2). https://doi.org/10.1029/2011RG000373

Wang, Lingli, & Qu, J. J. (2009). Satellite remote sensing applications for surface soil moisture monitoring: A review. *Frontiers of Earth Science in China*, *3*(2), 237–247. https://doi.org/10.1007/s11707-009-0023-7

Wang, Lixin, Caylor, K. K., Villegas, J. C., Barron-Gafford, G. A., Breshears, D. D., & Huxman, T. E. (2010). Partitioning evapotranspiration across gradients of woody plant cover: Assessment of a stable isotope technique. *Geophysical Research Letters*, *37*(9), n/a-n/a. https://doi.org/10.1029/2010GL043228

Wang, Lixin, Good, S. P., & Caylor, K. K. (2014). Global synthesis of vegetation control on evapotranspiration partitioning. *Geophysical Research Letters*, *41*(19), 6753–6757. https://doi.org/10.1002/2014GL061439

WAY, D. A., & MONTGOMERY, R. A. (2015). Photoperiod constraints on tree phenology, performance and migration in a warming world. *Plant, Cell & Environment*, *38*(9), 1725–1736. https://doi.org/10.1111/pce.12431

Wen, X., Yang, B., Sun, X., & Lee, X. (2016). Evapotranspiration partitioning through in-situ oxygen isotope measurements in an oasis cropland. *Agricultural and Forest Meteorology*, *230–231*, 89–96. https://doi.org/10.1016/j.agrformet.2015.12.003

Wibral, M., Priesemann, V., Kay, J. W., Lizier, J. T., & Phillips, W. A. (2017). Partial information decomposition as a unified approach to the specification of neural goal functions. *Brain and Cognition*, *112*, 25–38. https://doi.org/10.1016/j.bandc.2015.09.004

Wigneron, J.-P., Jackson, T. J., O'Neill, P., De Lannoy, G., de Rosnay, P., Walker, J. P., Ferrazzoli, P., Mironov, V., Bircher, S., Grant, J. P., Kurum, M., Schwank, M., Munoz-Sabater, J., Das, N., Royer, A., Al-Yaari, A., Al Bitar, A., Fernandez-Moran, R., Lawrence, H., … Kerr, Y. (2017). 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 Sensing of Environment*, *192*, 238–262. https://doi.org/10.1016/j.rse.2017.01.024

Wigneron, Jean-Pierre, Li, X., Frappart, F., Fan, L., Al-Yaari, A., De Lannoy, G., Liu, X., Wang, M., Le Masson, E., & Moisy, C. (2021). SMOS-IC data record of soil moisture and L-VOD: Historical development, applications and perspectives. *Remote Sensing of Environment*, *254*, 112238. https://doi.org/10.1016/j.rse.2020.112238

Williams, M., Richardson, A. D., Reichstein, M., Stoy, P. C., Peylin, P., Verbeeck, H., Carvalhais, N., Jung, M., Hollinger, D. Y., Kattge, J., Leuning, R., Luo, Y., Tomelleri, E., Trudinger, C. M., & Wang, Y.-P. (2009). Improving land surface models with FLUXNET data. *Biogeosciences*, *6*(7), 1341–1359. https://doi.org/10.5194/bg-6-1341-2009

Williams, P. L., & Beer, R. D. (2010). *Nonnegative Decomposition of Multivariate Information*. http://arxiv.org/abs/1004.2515

Wong, T. E. (2016). *The Impact of Stable Water Isotopic Information on Parameter Calibration in a Land Surface Model*.

Wood, D. A. (2021). Net ecosystem carbon exchange prediction and insightful data mining with an optimized data-matching algorithm. *Ecological Indicators*, *124*, 107426. https://doi.org/10.1016/j.ecolind.2021.107426

Wutzler, T., Lucas-Moffat, A., Migliavacca, M., Knauer, J., Sickel, K., Šigut, L., Menzer, O., & Reichstein, M. (2018). Basic and extensible post-processing of eddy covariance flux data with REddyProc. *Biogeosciences*, *15*(16), 5015–5030. https://doi.org/10.5194/bg-15-5015-2018

Xiao, W., Wei, Z., & Wen, X. (2018). Evapotranspiration partitioning at the ecosystem scale using the stable isotope method—A review. *Agricultural and Forest Meteorology*, *263*, 346–361. https://doi.org/10.1016/j.agrformet.2018.09.005

Xie, Q., Huang, W., Liang, D., Chen, P., Wu, C., Yang, G., Zhang, J., Huang, L., & Zhang, D. (2014). Leaf Area Index Estimation Using Vegetation Indices Derived From Airborne

Hyperspectral Images in Winter Wheat. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *7*(8), 3586–3594. https://doi.org/10.1109/JSTARS.2014.2342291

Xu, X., Liang, T., Zhu, J., Zheng, D., & Sun, T. (2019). Review of classical dimensionality reduction and sample selection methods for large-scale data processing. *Neurocomputing*, *328*, 5–15. https://doi.org/10.1016/j.neucom.2018.02.100

Xue, J., & Su, B. (2017). Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications. *Journal of Sensors*, *2017*, 1–17. https://doi.org/10.1155/2017/1353691

Yan, K., Park, T., Yan, G., Liu, Z., Yang, B., Chen, C., Nemani, R., Knyazikhin, Y., & Myneni, R. (2016). Evaluation of MODIS LAI/FPAR Product Collection 6. Part 2: Validation and Intercomparison. *Remote Sensing*, *8*(6), 460. https://doi.org/10.3390/rs8060460

Yaseef, N. R., Yakir, D., Rotenberg, E., Schiller, G., & Cohen, S. (2009). Ecohydrology of a semi-arid forest: partitioning among water balance components and its implications for predicted precipitation changes. *Ecohydrology*, n/a-n/a. https://doi.org/10.1002/eco.65

Ye, N., Walker, J. P., Guerschman, J., Ryu, D., & Gurney, R. J. (2015). Standing water effect on soil moisture retrieval from L-band passive microwave observations. *Remote Sensing of Environment*, *169*, 232–242. https://doi.org/10.1016/j.rse.2015.08.013

Yusup, Y., & Liu, H. (2020). Effects of persistent wind speeds on turbulent fluxes in the water-atmosphere interface. *Theoretical and Applied Climatology*, *140*(1–2), 313–325. https://doi.org/10.1007/s00704-019-03084-4

ZHANG, J., GRIFFIS, T. J., & BAKER, J. M. (2006). Using continuous stable isotope measurements to partition net ecosystem CO2 exchange. *Plant, Cell and Environment*, *29*(4), 483–496. https://doi.org/10.1111/j.1365-3040.2005.01425.x

Zhang, R., Kim, S., & Sharma, A. (2019). A comprehensive validation of the SMAP Enhanced Level-3 Soil Moisture product using ground measurements over varied climates and landscapes. *Remote Sensing of Environment*, *223*, 82–94. https://doi.org/10.1016/j.rse.2019.01.015

Zhang, Z., & Grabchak, M. (2013). Bias Adjustment for a Nonparametric Entropy Estimator. *Entropy*, *15*(12), 1999–2011. https://doi.org/10.3390/e15061999

Zhao, W. L., Qiu, G. Y., Xiong, Y. J., Paw U, K. T., Gentine, P., & Chen, B. Y. (2020). Uncertainties Caused by Resistances in Evapotranspiration Estimation Using High-Density Eddy Covariance Measurements. *Journal of Hydrometeorology*, *21*(6), 1349–1365. https://doi.org/10.1175/JHM-D-19-0191.1

Zhou, S., Yu, B., Zhang, Y., Huang, Y., & Wang, G. (2018). Water use efficiency and evapotranspiration partitioning for three typical ecosystems in the Heihe River Basin, northwestern China. *Agricultural and Forest Meteorology*, *253–254*, 261–273. https://doi.org/10.1016/j.agrformet.2018.02.002