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, Corvallis, OR 97330, USA
[2] Water Resources Graduate Program, Oregon State University, Corvallis, OR 97330, USA
Correspondence to: Bonan Li (libon@oregonstate.edu)

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~~2~~% of the information in the *in situ* soil moisture is unexplained by SMAP DCA soil moisture. ~~36~~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 beyond 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.~~51~~59) was found to be more influential than the informational random uncertainty (Pearson correlation of -0.~~37~~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}.~~.~~ ~~The DCA has a higher informational total uncertainty (88% of unexplained information of *in situ* soil moisture) in shrublands while the informational model uncertainty (31% of the informational total uncertainty) in shrublands is less dominant than other landcovers.~~ A decomposition of mutual information between T_{Bh}, T_{Bv} and DCA soil moisture shows that on average 58~~7~~% of information provided by T_{Bh} and T_{Bv} ~~i~~to DCA estimates is redundancy~~s redundant~~. The amount of information redundantly and synergistically provided by T_{Bh} and T_{Bv} was found to be tightly correlated (Pearson correlation of 0.79 and -0.82, respectively~~ 0.83~~) to the retrieval quality of SMAP DCA~~to how well the DCA correlated to *in situ* soil moisture~~. ~~Higher redundant information provided by T_{Bh} and T_{Bv} tends to be found in landcovers with less woody components.~~ The DCA retrieval quality improves as T_{Bh} and T_{Bv} provide more redundant information for the DCA soil moisture. This suggests that the informational redundancy and synergy between these remotely

sensed observations can be indicative about soil moisture retrieval quality~~used as independent metric to assess the retrieval quality of algorithms using these data streams~~. 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.

1 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)~~(Petropoulos et al., 2015)~~. 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 and Entekhabi, 1996). Various microwave frequencies have been available to date, amongst which the L-band ~~(1.4-1.427 GHz)~~ 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)~~(Njoku and Entekhabi, 1996)~~. 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 ~~a linear function of~~ brightness temperature (Wang and 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., 2010). 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. at., 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. at., 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) (Konings et al., 2016) for instance, 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 ~~The success of retrieving soil moisture and vegetation opacity are interdependent and it~~ is important to consider the how the amount of duplicate information carried within a set of observations limits the number of 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 ~~in order to~~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 and 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). ~~Additionally, the triple collocation method, which combines *in situ* measurements, SMAP observations, and model fields, has been used to characterize systematic biases and error variances.~~ 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 met 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 evaluation 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: (1) Shannon's entropy (Shannon, 1948), which is the amount of information required to fully describe a random variable and (2) mutual information (Cover and 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~~estimated~~ 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 and Beer, 2010 and further advanced by Goodwell and 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 and 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 modeling assumption. The partitioned components hold potential as a new model evaluation metrics that can be used to assess SMAP algorithm performance in remote and inaccessible regions (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 model uncertainty is 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.

developing an *in situ* and ancillary data independent SMAP DCA evaluation metric using partial information decomposition between SMAP DCA soil moisture and horizontally polarized and vertically polarized brightness temperature.

2 Material and Methods

2.1 *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). The *in situ* datasets have been used for a wide variety of research such as drought monitoring and satellite soil moisture evaluationsvalidation (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 51 (12 croplands, 30 grasslands, 5 shrublands, 4 mixed) selected USCRN stations (Fig. 1 and Table S1) based on the availability of *in situ* soil moisture dataset and the data quality of SMAP pixels in the study period of 03/31/2015 – 10/01/2020March 31, 2015 to December 10, 2020.

2.2 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 filter by the internal quality flags of T_{Bh} ("tb_qual_flag_h"), T_{Bv} ("tb_qual_flag_v") and DCA soil moisture ("retrieval_qual_flag_option3") as provided in SMAP data files~~quality flags~~. We ~~only kept the~~retain data points at a particular 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 1652. 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 and 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.,~~P., Bindlish, R., Chan, S., Njoku, E., and Jackson,~~ 2020a) ~~(Konings et al., 2015; O'Neill et al., 2020)~~. 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~~, P., Bindlish,~~. ~~R., Chan, S., Njoku, E., and Jackson~~, 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 1. Additionally, the 500m leaf area index (LAI) of each site was obtained from NASAs Moderate Resolution Imaging Spectrometer (MODIS) mission and averaged in time (Myneni~~,~~ et. al.~~R., Knyazikhin, Y., Park~~, 2015; ORNL DAAC, 2018). 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.

2.3 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 and 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

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

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 and 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 and Grabchak, 2013) and applied a normalization method to remove the bias that may cause by the heterogeneity in length of available datasets across all stations. We acknowledge that there exist several entropy correction and 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 and 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 and 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

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

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

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

where $H_{CN}(\mathrm{X, Y})$ is the corrected and normalized joint entropy of random variable associated with $\{X, Y\}$, $H(\mathrm{X, Y})$ is the uncorrected 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 and Diaconis, 1981). All the joint entropies that are associated with two or more random variables in the later equations (i.e., $H_{CN}(\textit{in situ}, \mathrm{DCA})$, $H_{CN}(\mathrm{T_{Bh}, T_{Bv}, DCA})$, $H_{CN}(\mathrm{T_{Bh}, T_{Bv}, T_{eff}}, \textit{in situ})$ 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.

~~Commonly~~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). ~~In theory,~~ T~~t~~here 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 and 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 is defined based on entropy and joint entropy (Cover and Thomas, 2005). The mutual information between T_{Bh} and DCA, and the mutual information between T_{Bv} and DCA, are computed as

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

and

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

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

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

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

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

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 outputs 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: $I(\mathrm{DCA}, \textit{in situ}) \leq \min[H_{CN}(\mathrm{DCA}), H_{CN}(\textit{in situ})]$. 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}(*in situ*), and the mutual information shared between *in situ* observations and model output, I(DCA, *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}(*in situ*), is not captured by the estimator, as measured by I(DCA, *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 is defined as informational total uncertainty (I_{Tot}). 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 or correctly expresses the true relationship between model inputs and *in situ* observations. However, it's commonly known that "*All models are wrong*, but some are useful" {Formatting Citation}(Peters and Kok, 2016) (Box, 1976) and model assumptions typically cannot fully express the true relationship between the explanatory 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:

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

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

and

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

2.4 Partial information decomposition

The distinct informational contributions of T_{Bh} and T_{Bv} to the DCA outputs are be assessed through a decomposition of the information. This method partitions multivariate mutual information to unique, redundantredundant, and synergistic components (Williams and 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 $I(\mathrm{T_{Bh}}, \mathrm{T_{Bv}}; \mathrm{DCA})$ can be expressed as

$$I(\mathrm{T_{Bh}}, \mathrm{T_{Bv}}; \mathrm{DCA}) = U_{h}(\mathrm{T_{Bh}}; \mathrm{DCA}) + U_{v}(\mathrm{T_{Bv}}; \mathrm{DCA}) + R(\mathrm{T_{Bh}}, \mathrm{T_{Bv}}; \mathrm{DCA}) + S(\mathrm{T_{Bh}}, \mathrm{T_{Bv}}; \mathrm{DCA}), \quad (13)$$

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

The mutual information between $\mathrm{T_{Bh}}$ and DCA and mutual information between $\mathrm{T_{Bv}}$ and DCA were defined as

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

and

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

where ~~U_1~~ U_h, ~~U_2~~ U_v, S and R are unknowns in the systems of equations (13) - (15). Goodwell and Kumar, 2017 showed that the R can be formulated as

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

where

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

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

and

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

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

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

It is important to acknowledge that we used the point based *in situ* soil moisture as the ground truth in this analysis. Due to course 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 and was found to vary with surface soil moisture conditions (Escorihuela et al., 2010; Raju et al., 1995)~~can be more sensitive to surface meteorological conditions and more random than the actual *in situ* soil moisture~~. The heterogeneity in each pixel relative to the *in situ* observations together with the sensing depth disparity can bias~~may negatively influence~~ the estimation of ~~the results of this study and result in an overestimate the actual~~ 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). ~~Therefore~~Based on current technologies, it is difficult~~hard~~ to find ~~a~~ reference dataset at 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~~. Because~~ the DCA essential is 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 sensing depth is more of imperfection L-band sensor. The reader should also keep these in mind while interpreting and adopting the results in this study.

3 Results

3.1 Information quantities and system informational uncertainties

The estimated entropies across all the study sites are shown in Figure 2 while the mutual information quantities are shown in Figure 3. ~~Figure 2 shows the estimated entropies across all the study sites while Figure 3 shows the mutual information quantities.~~ The $H_{CN}(T_{Bh})$ and $H_{CN}(T_{Bv})$ general follow the same pattern 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 patterns of $H_{CN}(T_{eff})$ is quite different (Fig. 2). On average, the H_{CN}(*in situ*) is 0.35, while H_{CN}(DCA) and 0.38. In general, H_{CN}(DCA) follows the pattern of H_{CN}(in situ) with the CV of H_{CN}(DCA) (0.064~~05~~) being smaller than the CV of H_{CN}(*in situ*) (0.081).

As shown in Figure 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 (Fig. 4a). As expected, the mutual information quantities (Fig. 3) are shown to be generally smaller than the entropy quantities (Fig. 2). ~~$H_{CN}(T_{Bh})$, $H_{CN}(T_{Bv})$ and H_{CN}(DCA) are significantly correlated with 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 (Fig. 4a).~~

~~The mutual information quantities (Fig. 3) are shown to be generally smaller than the entropy quantities (Fig. 2).~~ On average, $I(T_{Bh},T_{Bv}$; DCA) is 0.~~14~~13, while the I(DCA; *in situ*) and $I(T_{Bh},T_{Bv}, T_{eff}$; *in situ*) are 0.07~~6~~ and 0.17 (Fig. 3), respectively.

$I(T_{Bh},T_{Bv}, T_{eff}$; *in situ*) is significantly correlated (0.59~~8~~) with H_{CN}(*in situ*), while no significant correlation is found for I(DCA; *in situ*) and H_{CN}(*in situ*) (Fig. 4b)~~other two mutual information quantities (Fig. 4b)~~.

It is noticeable that there exists a large information gap ~~(Fig. 2 and Fig. 3)~~ between H_{CN}(*in situ*) in Figure 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. 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,~~On average,~~ the SMAP DCA explains ~~18~~20% of the H_{CN}(*in situ*) leaving ~~82~~80% of the H_{CN}(*in situ*) that is unexplained (Table 1) as informational total uncertainty (I_{Tot}). ~~36~~35% (Table 1) 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, the SMAP DCA system is capable of capturing more information of H_{CN}(*in situ*) at croplands and savannas (Table 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 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.

~~Grasslands and Mixed landcover have largest absolute I_{Rnd} (0.20) than other landcovers, while shrublands has the largest proportion of I_{Rnd} to I_{Tot} (Table 1). The shrublands have the largest I_{Mod} in absolute value, while grasslands have the largest proportion of I_{Mod} to I_{Tot} (Table 1).~~

~~3.2 Informational uncertainties and retrieval quality~~

The relationship between different informational uncertainties and the Pearson correlation coefficients between *in situ* and SMAP DCA output, 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 (Fig. 5). I_{Tot} is found to be negatively correlated (r = -0.69~~6~~, Fig. 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~~1~~ and -0.34~~7~~ 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} (Fig. 5b and ~~–~~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).

~~The negative relationship between SMAP DCA informational uncertainties are in line with general expectations since SMAP tends to capture more information about the *in situ* soil moisture when it retrieves high quality datasets.~~

3.3 Partial information decomposition of DCA

The partial information decompositions were assessed on a site basis and are shown in Figure 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 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~~7~~% of $I(T_{Bh},T_{Bv}; DCA)$) on average (Table 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~~6~~ (18~~9~~% of $I(T_{Bh},T_{Bv}; DCA)$) and 0.018~~9~~ (14% of $I(T_{Bh},T_{Bv}; DCA)$), respectively (Table 2). Compared to other decomposed information components, U_v is the smallest~~, but is of similar magnitude with S,~~ with its mean being 0.013 (10% of $I(T_{Bh},T_{Bv}; DCA)$). ~~Croplands~~ Savannas have the highest absolute and fraction of R ~~have the highest R in absolute value 0.095~~ (0.101, 74~~68~~% of $I(T_{Bh},T_{Bv}; DCA)$) (Table 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 2).~~while mixed landcover has the highest fraction of R (74% of $I(T_{Bh},T_{Bv}; DCA)$) (Table 2). In general, the DCA system is mainly dominated by R. This indicates that both T_{Bh} and T_{Bv} provide similar information within the DCA.~~

~~**3.4 Partial information decomposition and retrieval quality**~~

Through this analysis, it is shown (Fig. 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~~the DCA tends to retrieve high quality soil moisture when U_h, U_v and S are low (Fig. 7a – Fig. 7c)~~. This is demonstrated by a significant correlation of these components with the Pearson correlation between *in situ* and DCA soil moisture~~This is demonstrated by a negative correlation of these component with the Pearson correlation between *in situ* and DCA soil moisture (Fig. 7a – Fig. 7c)~~. 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. ~~In contrast, R shows the strongest positive correlation (Fig. 7d) with the relative model evaluation metric (r = 0.83). This indicates that R could potentially be a reference metric for DCA evaluation that does not require *in situ* and ancillary datasets.~~

4 Discussion

4.1 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 errors 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.

~~This information theory approach can add considerable power to SMAP modeling diagnosis. Mutual information can provide a way to unambiguously define the best achievement performance of a model that is able to completely transform the available information to the desired target given a set of the input data.~~

Any model based on theThe DCA model structure is a hypothesis that relates the input datasets to soil moisture based on prior physical knowledge. The essence of the modelDCA -iis thuss a procedure of processing the input dataset inin order toto estimate soil moisture. Thus, models, even the one performs 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 chance 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 aforementioned 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 1, we found that the DCA has more informational uncertainty in shrublands than grasslands and croplands which is consistent with previous study (Zhang et al., 2017). This might be due to stronger variability in vegetation types for shrublands while grasslands and croplands tend to be more uniform and homogeneous. Furthermore, shrublands tend to be relatively less sensitive to changes in water availability while grasslands are more sensitive to the soil moisture dynamics in the condition of drought (Geruo et al., 2017). It is worth to noting that these finding are based on lumping our studied sites into 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 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 (654% 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) in the locations, 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 use the T_{eff} to capture both soil and canopy temperature because the differences between canopy and soil temperature are minimized in the morning and dawnafternoon orbits. The T_{eff} is computed based on a model that uses the information from average soil temperature of first layer (5cm–15cm) and second layer (15cm – 35cm) and interpolated in time in order toto match SMAP morning and dawn observations (O'Neill, et.al.P., Bindlish, R., Chan, S., Njoku, E., and Jackson, 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 (356% 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 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's~~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 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.~~, P., Bindlish, R., Chan, S., Njoku, E., and Jackson~~, 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 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 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 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 quantitively analyze the uncertainty components in microwave remote sensing models. The results of this study can be further used as ~~a foundation~~ guidance of SMAP algorithm and ~~assessing approach that~~ 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.

~~This analysis, though focused on DCA soil moisture, can be transferred and extended to analyze other remote sensing algorithms.~~

4.2 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. ~~The second objective of this study was to demonstrate that the partitioned information components can be used as a new DCA model evaluation~~

metric that does not depend on *in situ* soil moisture and other ancillary datasets. We found a strong linear relationship between redundant information (R) 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, -0.52, 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, -0.28, 0.22, and -0.16 respectively), the informational total uncertainty (the Pearson correlation of I_{Tot} with R, S, U_h, U_v are -0.76, 0.62, 0.56, and 0.68 respectively), informational random uncertainty (the Pearson correlation of I_{Rnd} with R, S, U_h, U_v are -0.42 , 0.29, 0.05, and 0.15 respectively), and informational model uncertainty (the Pearson correlation of I_{Mod} with R, S, U_h, U_v are -0.63, 0.56, 0.66, and 0.75 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}.

, which indicated that T_{Bh} and T_{Bv} are highly dependent. R is also the dominant component relative to others quantified here. In general, it is more likely to observe higher R in the less woody landcovers (croplands and grasslands) where the range of brightness temperature may possible be greater. From an information perspective, higher or complete R indicates that one variable is a function of the other, or they share the same source. T_{Bv} and T_{Bh} are known to be highly correlated. It's important to note that the decomposed information component R is dependent on the DCA parameterizations that determines how strong the T_{Bh} and T_{Bv} are linked with the DCA. This stronger linkage is indicated by a higher value of R relative to other components.

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.

We found that DCA model performance, as characterized by the correlation between Person correlation between DCA and *in situ* soil moisture, improves with larger values of *R* that T_{Bh} and T_{Bv} share with DCA estimates. Given the strength of this relationship, the *R* could be potentially used as a DCA evaluation metric that doesn't 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 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.

4.3 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 (Xu et al., 2019), 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 (O'Neill, et.al. P., Bindlish, R., Chan, S., Njoku, E., and Jackson, 2020a). 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. 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 $I(T_{Bh}, T_{Bv}, T_{eff}, NDVI;$ *in situ*) being larger than or equal to $I(T_{Bh}, T_{Bv}, T_{eff};$ *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. The problem of how to leverage information theory to evaluate the error components in the locations without *in situ* soil moisture measurements is challenging and could be an interesting topic for future works. F We inally, we would expect the informational uncertainty analysis to provide theasymptotic estimation of random and model uncertainties. The best performance we can expect from this current uncertainty analysis is to use all ofall the available datasets we have; yet we believe that uncertainty estimations of this approach should be stabilized given adequate representative locations and data records.

5 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 1820% of the information in the *in situ* soil moisture leaving 8280% unexplained. Among the unexplained information, 6465% 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. We showed that the shrublands have smaller informational model uncertainty compared with other landcovers. The performance of SMAP DCA is negatively correlated to all the information uncertainties, with the informational uncertainty being more influential than the informational random uncertainty.

The decomposition of the mutual information has shown that all decomposed components are correlated with the Pearson correlation between in situ and DCA soil moisture, with the redundant information being the tightest. 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. The result of mutual information decomposition analysis can be adopted as a new *in situ* independent SMAP soil moisture evaluation reference metric especially in locations where *in situ* soil moisture is not readily available.

Code availability

The code regarding the SMAP dataset time series, mutual information and partial information decomposition calculation can obtained from https://github.com/libonancaesar/HESS_Information_Uncertaintyhttps://github.com/libonancaesar/HESS_Information_Uncertainty

Data availability

SMAP L2 Radiometer Half-Orbit 36 km EASE-Grid Soil Moisture, Version 7 is was acquired from US National Snow and Ice Data Center (https://nsidc.org/data/smaphttps://nsidc.org/data/smap). The *in situ* soil moisture is accessible through U.S. Climate Reference Network (https://www.ncdc.noaa.gov/crn/https://www.ncdc.noaa.gov/crn/). The leaf area index dataset can be accessed through Oak Ridge National Laboratory Distributed Active Archive Center (https://modis.ornl.gov/globalsubset/)

Author contribution

Bonan Li: conceptualization; data acquisition; formal analysis; methodology; original draft writing and editing; ***Stephen P. Good***: conceptualization; methodology; draft writing, editing and revisions; supervision.

Competing interests

The authors declare no conflicts of interest.

Acknowledgments

This project was supported by The National Aeronautics and Space Administration under grant NNX16AN13G.

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. and Yueh, S.: Evaluating soil moisture retrievals from ESA's SMOS and NASA's SMAP brightness temperature datasets, Remote Sens. Environ., 193, 257–273, doi:10.1016/j.rse.2017.03.010, 2017.

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

Bassiouni, M., Good, S. P., Still, C. J. and Higgins, C. W.: Plant Water Uptake Thresholds Inferred From Satellite Soil Moisture, Geophys. Res. Lett., 47(7), doi:10.1029/2020GL087077, 2020.

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. and Diamond, H. J.: U.S. Climate Reference Network Soil Moisture and Temperature Observations, J. Hydrometeorol., 14(3), 977–988, doi:10.1175/JHM-D-12-0146.1, 2013.

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

Box, G. E. P.: Science and Statistics, J. Am. Stat. Assoc., 71(356), 791–799, doi:10.1080/01621459.1976.10480949, 1976.

Chan, S.: Soil Moisture Active Passive (SMAP) Level 2 Passive Soil Moisture Product Specification Document, Jet Propuls. Lab. Inst. Technol. Pasadena, USA, JPL D-72547 (Version 7.0), 63, 2020.

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., McNairn, H., Thibeault, M., Martinez-Fernandez, J., Gonzalez-Zamora, A., Seyfried, M., Bosch, D., Starks, P., Goodrich, D., Prueger, J., Palecki, M., Small, E. E., Zreda, M., Calvet, J.-C., Crow, W. T. and Kerr, Y.: Assessment of the SMAP Passive Soil Moisture Product, IEEE Trans. Geosci. Remote Sens., 54(8), 4994–5007, doi:10.1109/TGRS.2016.2561938, 2016.

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., Thibeault, M. and Walker, J.: Improved SMAP Dual-Channel Algorithm for the Retrieval of Soil Moisture, IEEE Trans. Geosci. Remote Sens., 58(6), 3894–3905, doi:10.1109/TGRS.2019.2959239, 2020.

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. and Seyfried, M. S.: Application of Triple Collocation in Ground-Based Validation of Soil Moisture Active/Passive (SMAP) Level 2 Data Products, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., 10(2), 489–502, doi:10.1109/JSTARS.2016.2569998, 2017.

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., Lopez-Baeza, E., Martínez-Fernández, J., González-Zamora, A., Livingston, S., McNairn, H., Pacheco, A., Moghaddam, M., Montzka, C., Notarnicola, C., Niedrist, G., Pellarin, T., Prueger, J., Pulliainen, J., Rautiainen, K., Ramos, J., Seyfried, M., Starks, P., Su, Z., Zeng, Y., van der Velde, R., Thibeault, M., Dorigo, W., Vreugdenhil, M., Walker, J. P., Wu, X., Monerris, A., O’Neill, P. E., Entekhabi, D., Njoku, E. G. and Yueh, S.: Validation of SMAP surface soil moisture products with core validation sites, Remote Sens. Environ., 191, 215–231, doi:10.1016/j.rse.2017.01.021, 2017.

Cover, T. M. and Thomas, J. A.: Elements of Information Theory, Wiley., 2005.

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

Dadap, N. C., Cobb, A. R., Hoyt, A. M., Harvey, C. F. and Konings, A. G.: Satellite soil moisture observations predict burned area in Southeast Asian peatlands, Environ. Res. Lett., 14(9), 094014, doi:10.1088/1748-9326/ab3891, 2019.

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. and Thorne, P. W.: U.S. Climate Reference Network after One Decade of Operations: Status and Assessment, Bull. Am. Meteorol. Soc., 94(4), 485–498, doi:10.1175/BAMS-D-12-00170.1, 2013.

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. and Silva, C.: The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth’s forests and topography, Sci. Remote Sens., 1, 100002, doi:10.1016/j.srs.2020.100002, 2020.

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., Spencer, M. W., Thurman, S. W., Tsang, L. and Van Zyl, J.: The Soil Moisture Active Passive (SMAP) Mission, Proc. IEEE, 98(5), 704–716, doi:10.1109/JPROC.2010.2043918, 2010.

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

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

Finn, C. and Lizier, J.: Pointwise Partial Information Decomposition Using the Specificity and Ambiguity Lattices, Entropy,

20(4), 297, doi:10.3390/e20040297, 2018.
Freedman, D. and Diaconis, P.: On the histogram as a density estimator:L2 theory, Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete, 57(4), 453–476, doi:10.1007/BF01025868, 1981.
Gao, L., Sadeghi, M., Ebtehaj, A. and Wigneron, J.-P.: A temporal polarization ratio algorithm for calibration-free retrieval of soil moisture at L-band, Remote Sens. Environ., 249, 112019, doi:10.1016/j.rse.2020.112019, 2020.
Gong, W., Gupta, H. V., Yang, D., Sricharan, K. and Hero, A. O.: Estimating epistemic and aleatory uncertainties during hydrologic modeling: An information theoretic approach, Water Resour. Res., 49(4), 2253–2273, doi:10.1002/wrcr.20161, 2013.
Goodwell, A. E. and Kumar, P.: Temporal information partitioning: Characterizing synergy, uniqueness, and redundancy in interacting environmental variables, Water Resour. Res., 53(7), 5920–5942, doi:10.1002/2016WR020216, 2017.
Goodwell, A. E., Kumar, P., Fellows, A. W. and Flerchinger, G. N.: Dynamic process connectivity explains ecohydrologic responses to rainfall pulses and drought, Proc. Natl. Acad. Sci., 115(37), E8604–E8613, doi:10.1073/pnas.1800236115, 2018.
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., Reichle, R., Richaume, P., Rüdiger, C., Scanlon, T., van der Schalie, R., Wigneron, J.-P. and Wagner, W.: Validation practices for satellite soil moisture retrievals: What are (the) errors?, Remote Sens. Environ., 244, 111806, doi:10.1016/j.rse.2020.111806, 2020.
Hulley, G., Hook, S., Fisher, J. and Lee, C.: ECOSTRESS, A NASA Earth-Ventures Instrument for studying links between the water cycle and plant health over the diurnal cycle, in 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 5494–5496, IEEE., 2017.
Jackson, T. J., Schmugge, T. J. and Wang, J. R.: Passive microwave sensing of soil moisture under vegetation canopies, Water Resour. Res., 18(4), 1137–1142, doi:10.1029/WR018i004p01137, 1982.
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. and Delwart, S.: The SMOS Soil Moisture Retrieval Algorithm, IEEE Trans. Geosci. Remote Sens., 50(5), 1384–1403, doi:10.1109/TGRS.2012.2184548, 2012.
Konings, A. G., McColl, K. A., Piles, M. and Entekhabi, D.: How Many Parameters Can Be Maximally Estimated From a Set of Measurements?, IEEE Geosci. Remote Sens. Lett., 12(5), 1081–1085, doi:10.1109/LGRS.2014.2381641, 2015.
Konings, A. G., Piles, M., Rötzer, K., McColl, K. A., Chan, S. K. and Entekhabi, D.: Vegetation optical depth and scattering albedo retrieval using time series of dual-polarized L-band radiometer observations, Remote Sens. Environ., 172, 178–189, doi:10.1016/j.rse.2015.11.009, 2016.
Konings, A. G., Piles, M., Das, N. and Entekhabi, D.: L-band vegetation optical depth and effective scattering albedo estimation from SMAP, Remote Sens. Environ., 198, 460–470, doi:10.1016/j.rse.2017.06.037, 2017.
Kunert-Graf, J., Sakhanenko, N. and Galas, D.: Partial Information Decomposition and the Information Delta: A Geometric Unification Disentangling Non-Pairwise Information, Entropy, 22(12), 1333, doi:10.3390/e22121333, 2020.
Leeper, R. D., Bell, J. E., Vines, C. and Palecki, M.: An Evaluation of the North American Regional Reanalysis Simulated Soil Moisture Conditions during the 2011–13 Drought Period, J. Hydrometeorol., 18(2), 515–527, doi:10.1175/JHM-D-16-

0132.1, 2017.
Mo, T., Choudhury, B. J., Schmugge, T. J., Wang, J. R. and Jackson, T. J.: A model for microwave emission from vegetation-covered fields, J. Geophys. Res., 87(C13), 11229, doi:10.1029/JC087iC13p11229, 1982.
Mohanty, B. P., Cosh, M. H., Lakshmi, V. and Montzka, C.: Soil Moisture Remote Sensing: State-of-the-Science, Vadose Zo. J., 16(1), vzj2016.10.0105, doi:10.2136/vzj2016.10.0105, 2017.
Myneni, R., Knyazikhin, Y., Park, T.: MCD15A3H MODIS/Terra+Aqua Leaf Area Index/FPAR 4-day L4 Global 500m SIN Grid V006 [Data set.NASA EOSDIS Land Processes DAAC. Accessed 2021-05-20 from, , doi:https://doi.org/10.5067/MODIS/MCD15A3H.006, 2015.
Njoku, E. G. and Entekhabi, D.: Passive microwave remote sensing of soil moisture, J. Hydrol., 184(1–2), 101–129, doi:10.1016/0022-1694(95)02970-2, 1996.
O’Neill, P., Bindlish, R., Chan, S., Njoku, E., and Jackson, T.: Algorithm theoretical basis document: Level 2 & 3 soil moisture (passive) data products, Lev. 2 3 soil moisture data Prod. Jet Propuls. Lab. Inst. Technol. Pasadena, USA, JPL D-66480 (revision F), 100, 2020.a
O’Neill, P. E., S. Chan, E. G. Njoku, T. Jackson, R. Bindlish, and J. C.: SMAP L2 Radiometer Half-Orbit 36 km EASE-Grid Soil Moisture, Version 7, [March 31 2015 to December 2020], Boulder, Color. USA. NASA Natl. Snow Ice Data Cent. Distrib. Act. Arch. Center[Date Accessed December 10 2020], doi:https://doi.org/10.5067/F1TZ0CBN1F5N., 2020.b
ORNL DAAC: MODIS and VIIRS Land Products Global Subsetting and Visualization Tool, , doi:10.3334/ORNLDAAC/1379, 2018.
Paninski, L.: Estimation of Entropy and Mutual Information, Neural Comput., 15(6), 1191–1253, doi:10.1162/089976603321780272, 2003.
Petropoulos, G. P., Ireland, G. and Barrett, B.: Surface soil moisture retrievals from remote sensing: Current status, products & future trends, Phys. Chem. Earth, Parts A/B/C, 83–84, 36–56, doi:10.1016/j.pce.2015.02.009, 2015.
Raju, S., Chanzy, A., Wigneron, J.-P., Calvet, J.-C., Kerr, Y. and Laguerre, L.: Soil moisture and temperature profile effects on microwave emission at low frequencies, Remote Sens. Environ., 54(2), 85–97, doi:10.1016/0034-4257(95)00133-L, 1995.
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. and Uematsu, M.: International Geosphere–Biosphere Programme and Earth system science: Three decades of co-evolution, Anthropocene, 12, 3–16, doi:10.1016/j.ancene.2016.01.001, 2015.
Shannon, C. E.: A Mathematical Theory of Communication, Bell Syst. Tech. J., 27(3), 379–423, doi:10.1002/j.1538-7305.1948.tb01338.x, 1948.
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. and Walker, J. P.: SMAP soil moisture drying more rapid than observed in situ following rainfall events, Geophys. Res. Lett., 43(15), 8068–8075, doi:10.1002/2016GL069946, 2016.
Uber, M., Vandervaere, J.-P., Zin, I., Braud, I., Heistermann, M., Legoût, C., Molinié, G. and Nord, G.: How does initial soil moisture influence the hydrological response? A case study from southern France, Hydrol. Earth Syst. Sci., 22(12), 6127–6146, doi:10.5194/hess-22-6127-2018, 2018.
Uda, S.: Application of information theory in systems biology, Biophys. Rev., 12(2), 377–384, doi:10.1007/s12551-020-

00665-w, 2020.

Wang, L. and Qu, J. J.: Satellite remote sensing applications for surface soil moisture monitoring: A review, Front. Earth Sci. China, 3(2), 237–247, doi:10.1007/s11707-009-0023-7, 2009.

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

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., Mialon, A., Parrens, M., Richaume, P., Delwart, S. and Kerr, Y.: Modelling the passive microwave signature from land surfaces: A review of recent results and application to the L-band SMOS & SMAP soil moisture retrieval algorithms, Remote Sens. Environ., 192, 238–262, doi:10.1016/j.rse.2017.01.024, 2017.

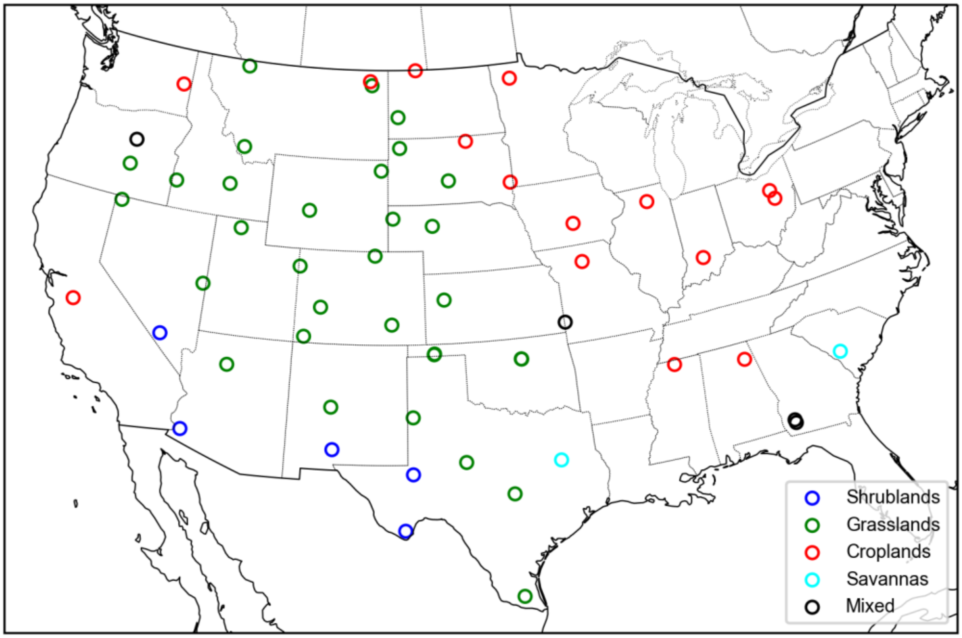
Williams, P. L. and Beer, R. D.: Nonnegative Decomposition of Multivariate Information, [online] Available from: http://arxiv.org/abs/1004.2515, 2010.

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

Ye, N., Walker, J. P., Guerschman, J., Ryu, D. and Gurney, R. J.: Standing water effect on soil moisture retrieval from L-band passive microwave observations, Remote Sens. Environ., 169, 232–242, doi:10.1016/j.rse.2015.08.013, 2015.

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

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



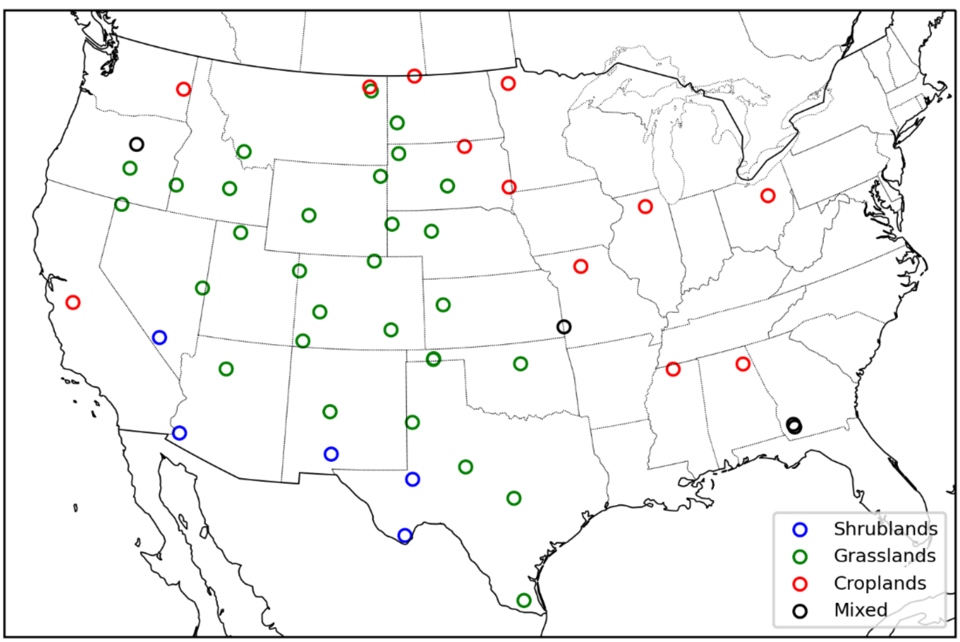
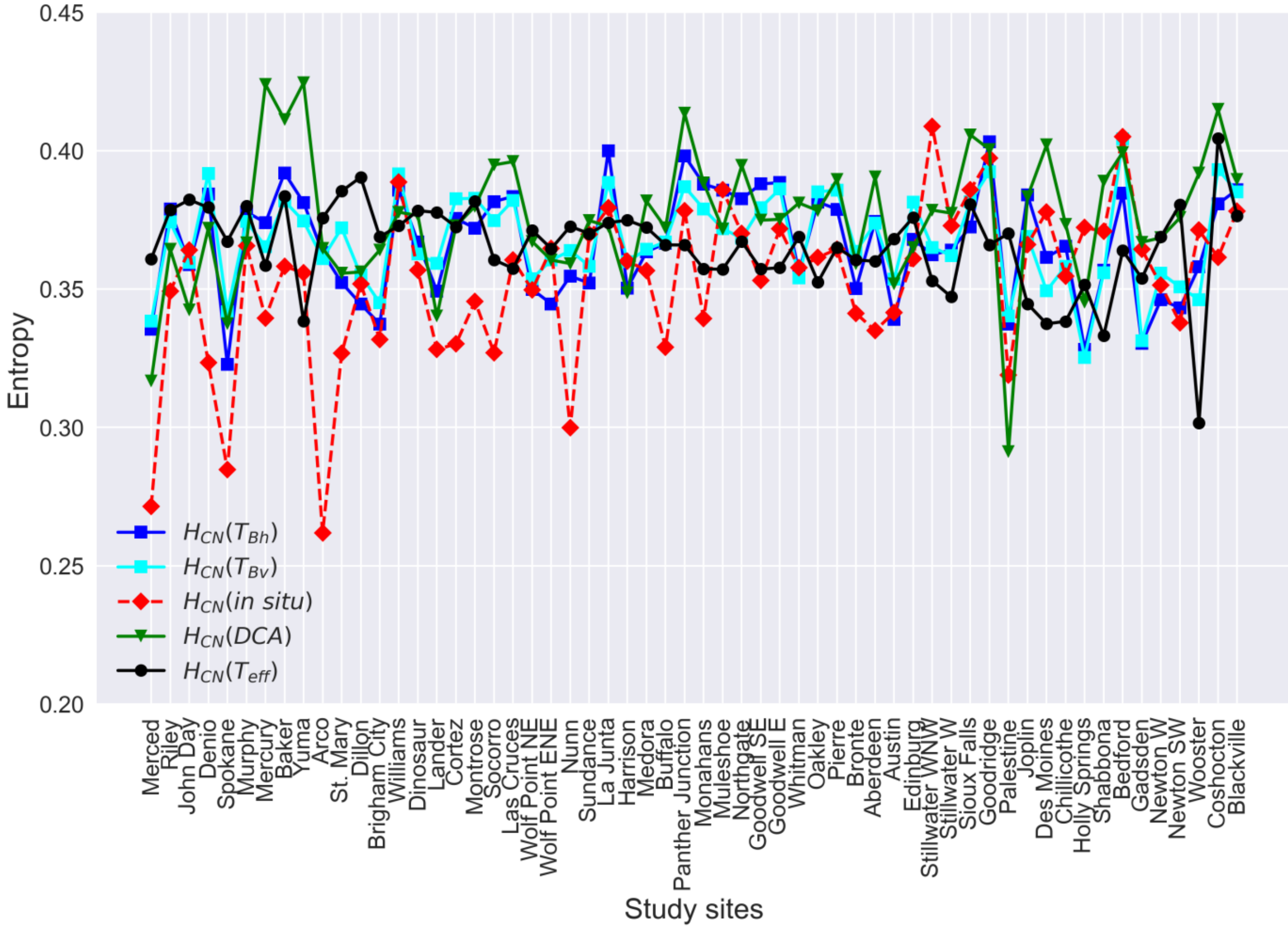


Figure 1 Spatial distribution of selected USCRN stations classified by landcovers.


Entropy
0.45
0.40
0.35
0.30
0.25
0.20
$H_{CN}(T_{Bh})$
$H_{CN}(T_{Bv})$
$H_{CN}(in\ situ)$
$H_{CN}(DCA)$
$H_{CN}(T_{eff})$
Merced
Riley
John Day
Denio
Spokane
Murphy
Mercury
Baker
Yuma
Arco
St. Mary
Dillon
Brigham City
Williams
Dinosaur
Lander
Cortez
Montrose
Socorro
Las Cruces
Wolf Point NE
Wolf Point ENE
Nunn
Sundance
La Junta
Harrison
Medora
Buffalo
Panther Junction
Monahans
Muleshoe
Northgate
Goodwell SE
Goodwell E
Whitman
Oakley
Pierre
Bronte
Aberdeen
Austin
Edinburg
Stillwater WNW
Stillwater W
Sioux Falls
Goodridge
Palestine
Joplin
Des Moines
Chillicothe
Holly Springs
Shabbona
Bedford
Gadsden
Newton W
Newton SW
Wooster
Coshocton
Blackville
Study sites

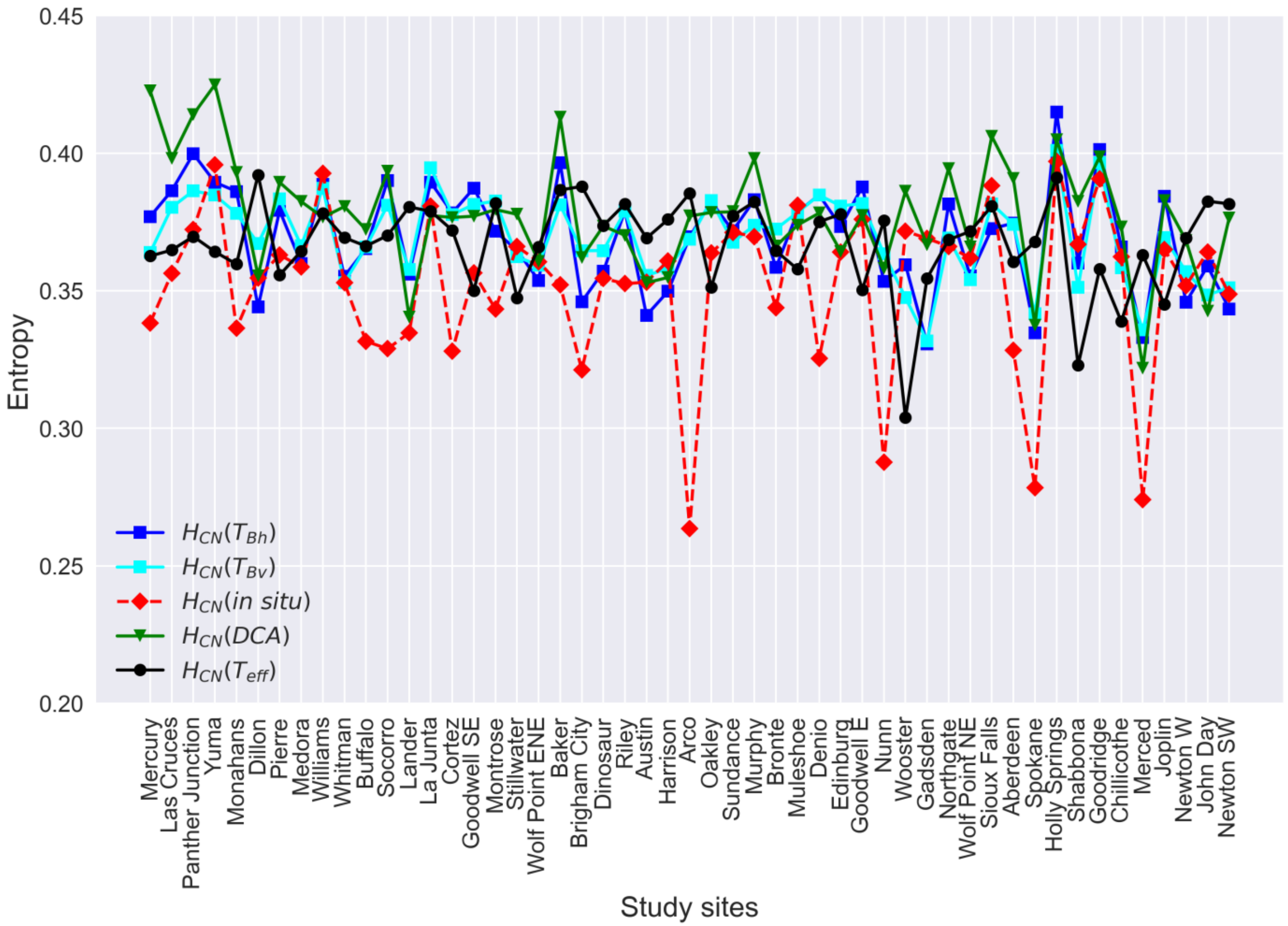
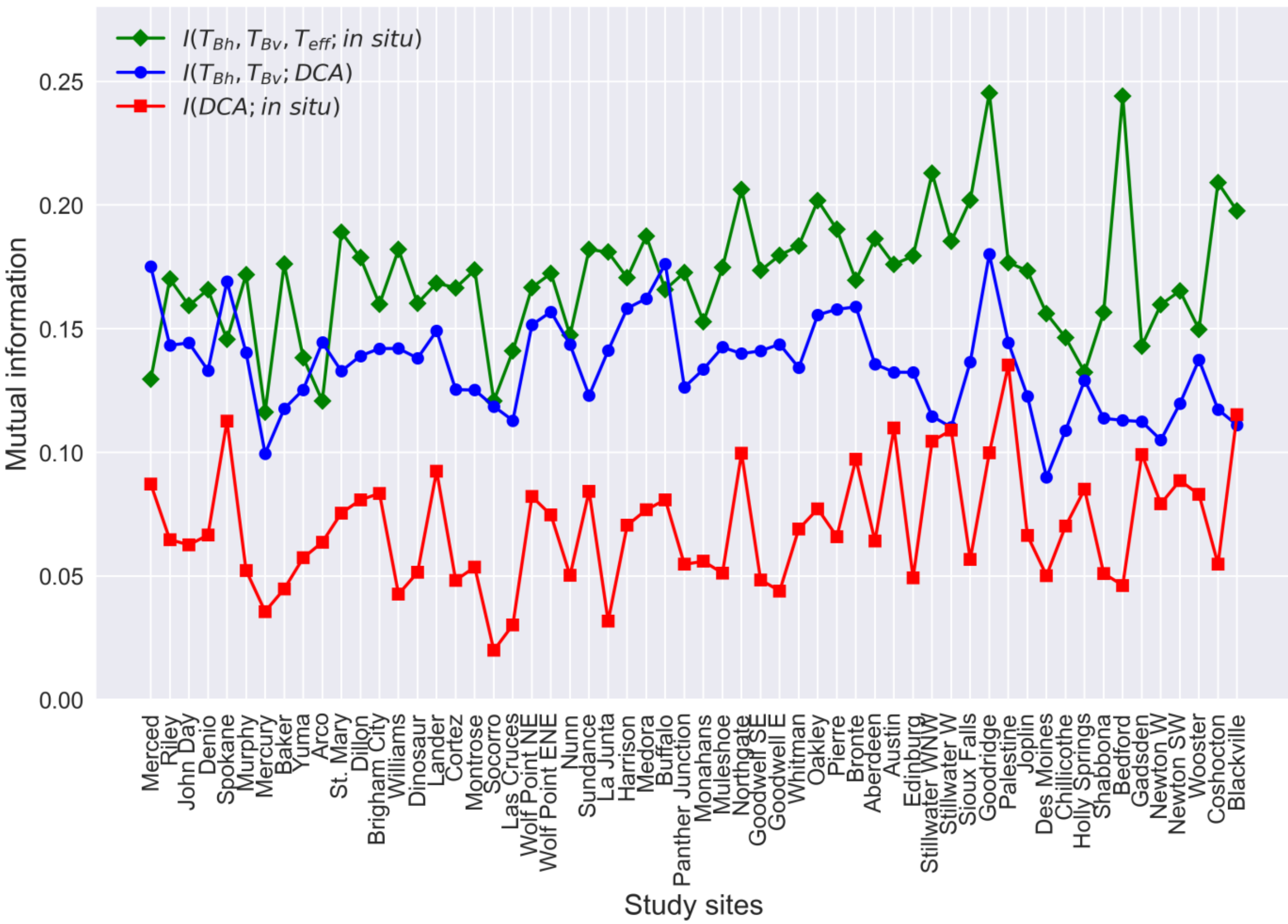


Figure 2 Entropies of *in situ* soil moisture, horizontally polarized brightness temperature (T_{Bh}), vertically polarized brightness temperature (T_{Bv}), soil effective temperature (T_{eff}) and DCA soil moisture across the study sites. The sites are ordered by longitude (West to East).



$I(T_{Bh}, T_{Bv}, T_{eff}; in\ situ)$
$I(T_{Bh}, T_{Bv}; DCA)$
$I(DCA; in\ situ)$
Mutual information
0.00
0.05
0.10
0.15
0.20
0.25
Study sites
Merced
Riley
John Day
Denio
Spokane
Murphy
Mercury
Baker
Yuma
Arco
St. Mary
Dillon
Brigham City
Williams
Dinosaur
Lander
Cortez
Montrose
Socorro
Las Cruces
Wolf Point NE
Wolf Point ENE
Nunn
Sundance
La Junta
Harrison
Medora
Buffalo
Panther Junction
Monahans
Muleshoe
Northgate
Goodwell SE
Goodwell E
Whitman
Oakley
Pierre
Bronte
Aberdeen
Austin
Edinburg
Stillwater WNW
Stillwater W
Sioux Falls
Goodridge
Palestine
Joplin
Des Moines
Chillicothe
Holly Springs
Shabbona
Bedford
Gadsden
Newton W
Newton SW
Wooster
Coshocton
Blackville

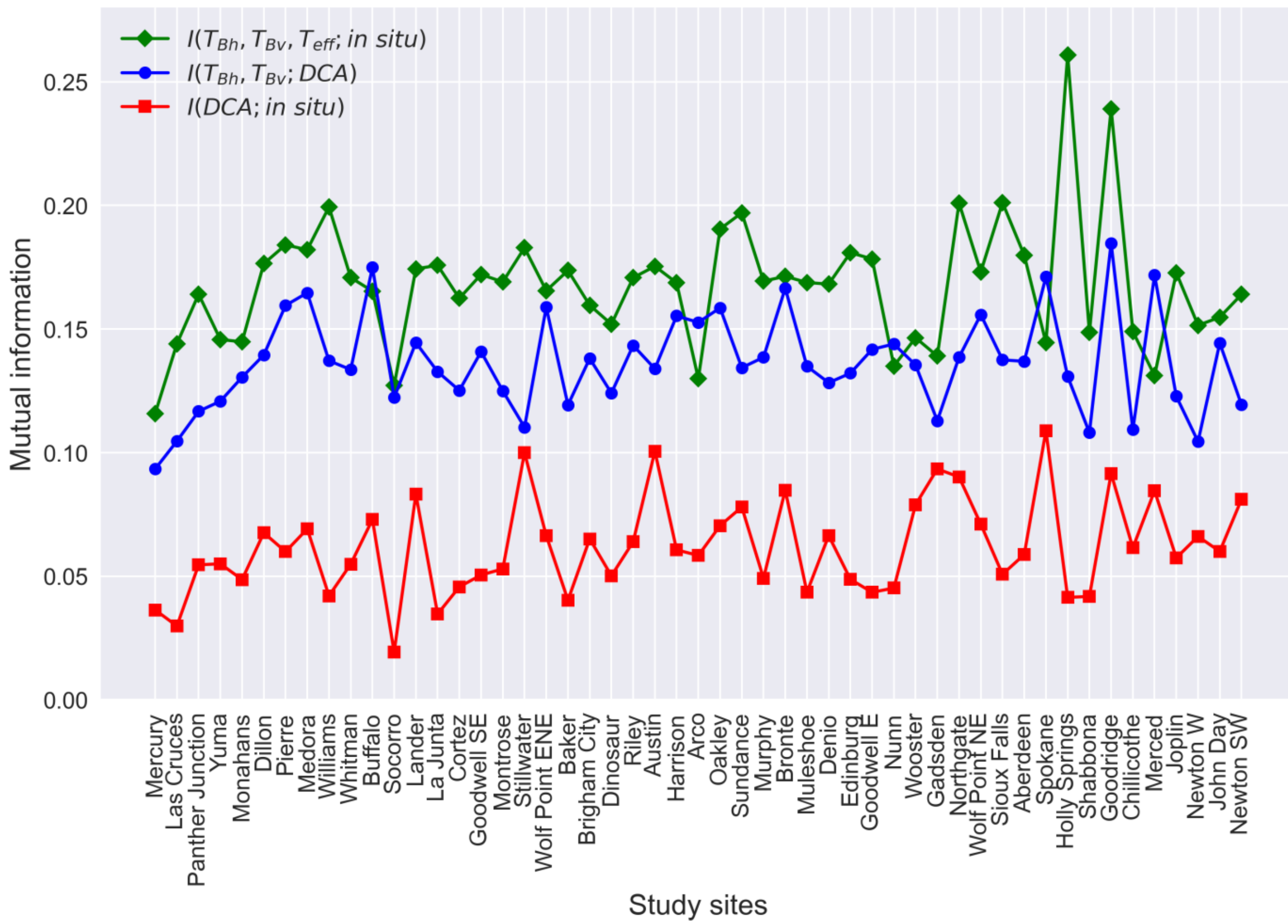


Figure 3 Mutual information between horizontally polarized brightness temperature (T_{Bh}), vertically polarized brightness temperature (T_{Bv}), soil effective temperature (T_{eff}) and *in situ* soil moisture, mutual information between horizontally polarized brightness temperature (T_{Bh}), vertically polarized brightness temperature (T_{Bv}) and DCA soil moisture, mutual information between DCA soil moisture and *in situ* soil moisture. See figure 2 caption for site ordering.

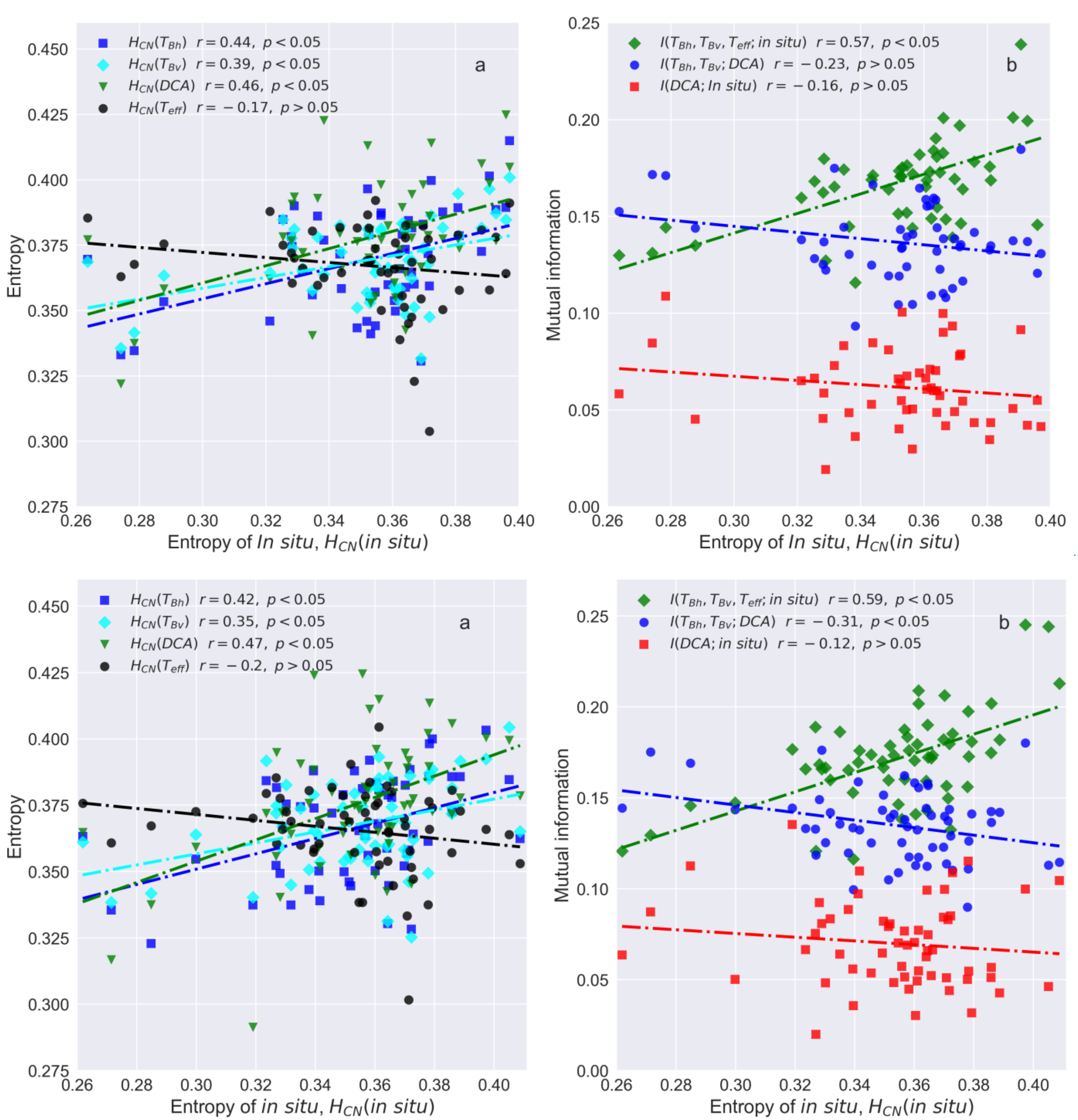
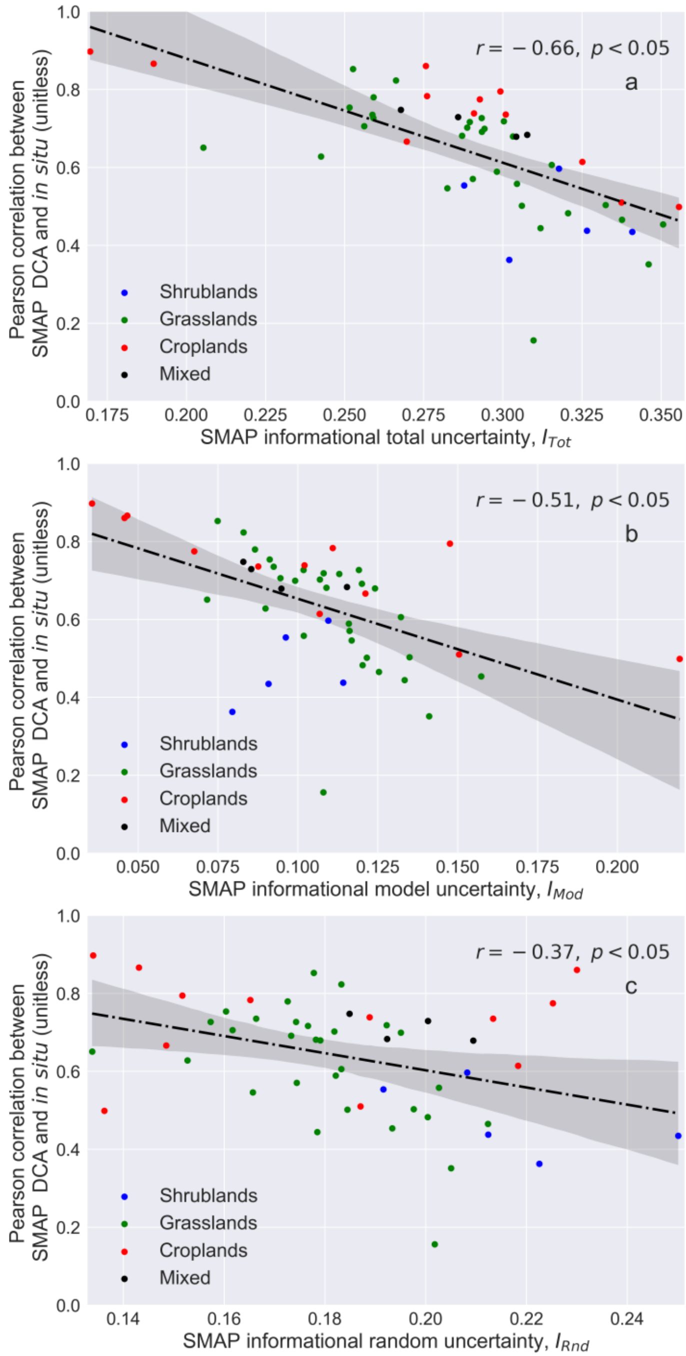
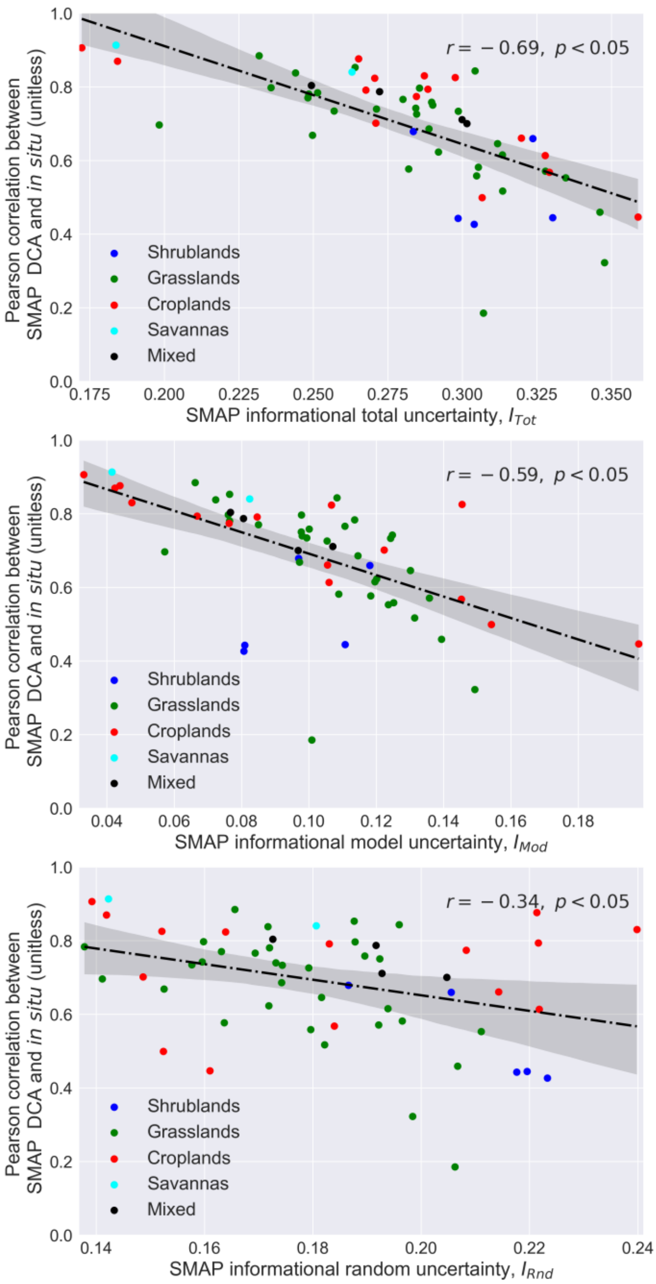


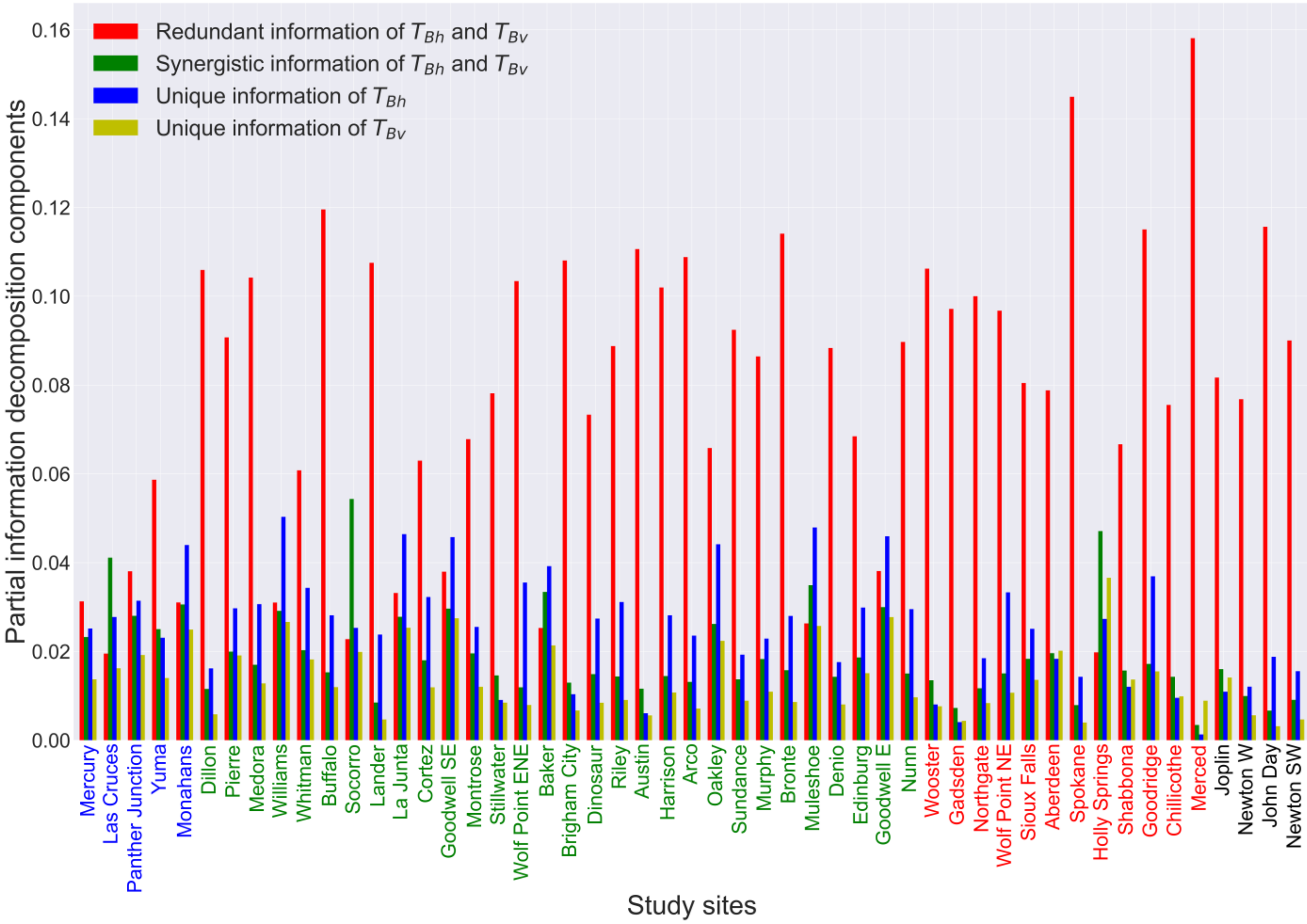
Figure 4 Entropy of *in situ* soil moisture against the entropies of DCA soil moisture, horizontally polarized brightness temperature (T_{Bh}), vertically polarized brightness temperature (T_{Bv}) and soil effective temperature (T_{eff}) (a) and mutual information quantities (b).



$r = -0.66,\ p < 0.05$
a
Shrublands
Grasslands
Croplands
Mixed
Pearson correlation between SMAP DCA and in situ (unitless)
SMAP informational total uncertainty, I_{Tot}
$r = -0.51,\ p < 0.05$
b
SMAP informational model uncertainty, I_{Mod}
$r = -0.37,\ p < 0.05$
c
SMAP informational random uncertainty, I_{Rnd}



825 **Figure 5** SMAP informational total uncertainty (a), SMAP informational model uncertainty (b) and SMAP informational random uncertainty against Pearson correlation between *in situ* soil moisture and DCA soil moisture


Redundant information of T_{Bh} and T_{Bv}
Synergistic information of T_{Bh} and T_{Bv}
Unique information of T_{Bh}
Unique information of T_{Bv}
Partial information decomposition components
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
0.16
Mercury
Las Cruces
Panther Junction
Yuma
Monahans
Dillon
Pierre
Medora
Williams
Whitman
Buffalo
Socorro
Lander
La Junta
Cortez
Goodwell SE
Montrose
Stillwater
Wolf Point ENE
Baker
Brigham City
Dinosaur
Riley
Austin
Harrison
Arco
Oakley
Sundance
Murphy
Bronte
Muleshoe
Denio
Edinburg
Goodwell E
Nunn
Wooster
Gadsden
Northgate
Wolf Point NE
Sioux Falls
Aberdeen
Spokane
Holly Springs
Shabbona
Goodridge
Chillicothe
Merced
Joplin
Newton W
John Day
Newton SW
Study sites

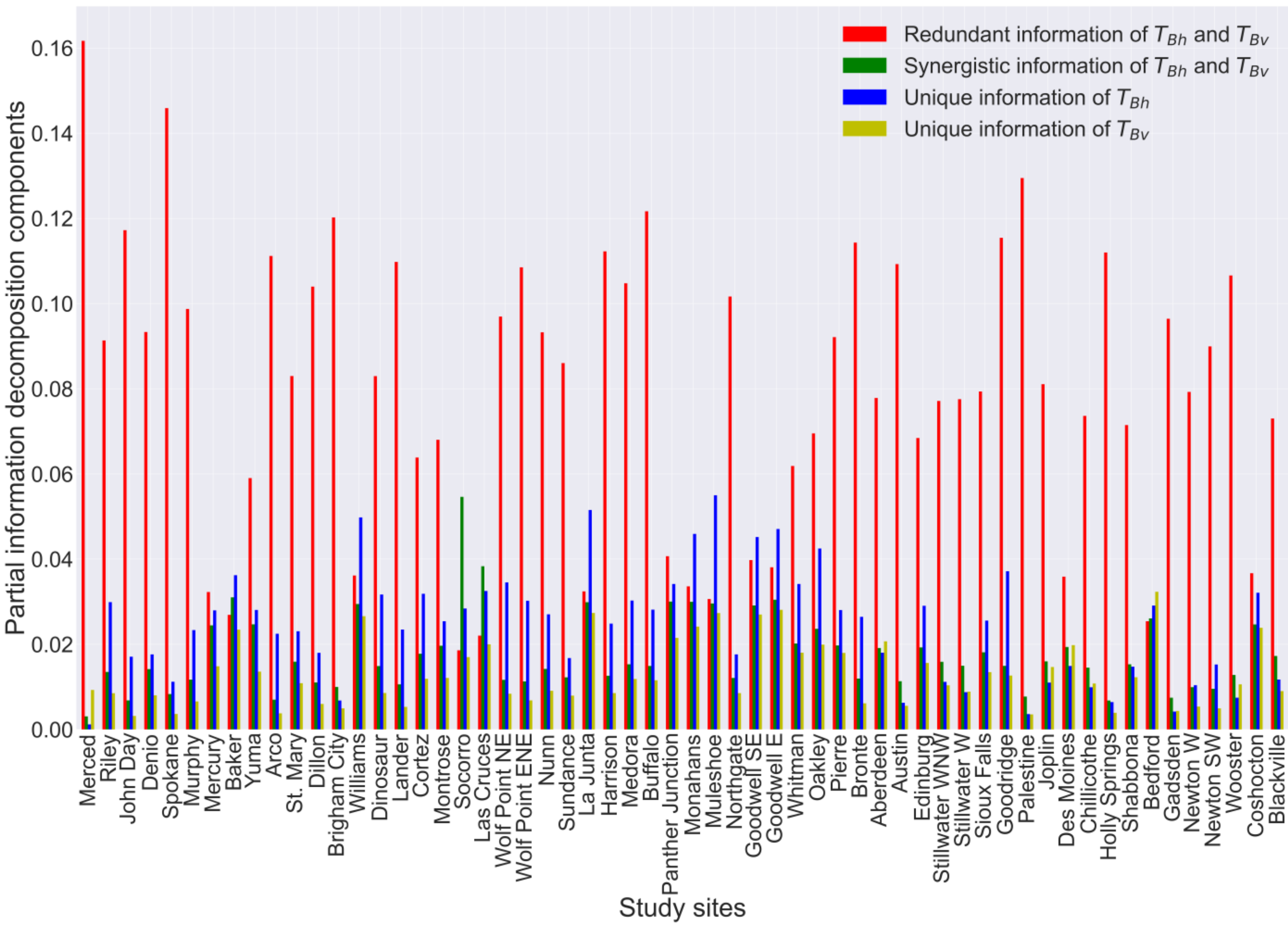
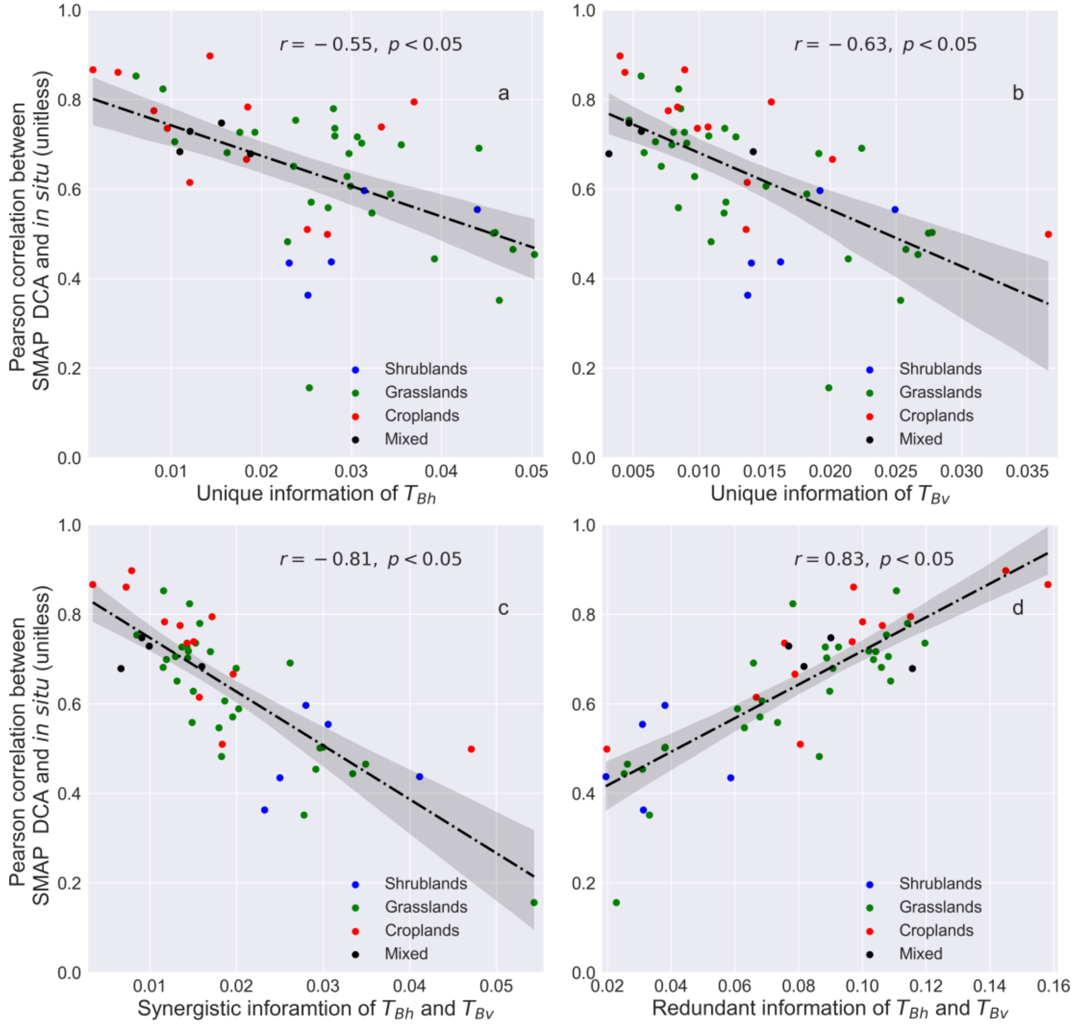


Figure 6 Partial information decomposition components between horizontally (T_{Bh}) and vertically (T_{Bv}) polarized brightness temperature and DCA soil moisture. See figure 2 caption for site ordering.

~~The colored labels of the horizontal axis represent different landcover of the study sites (blue: Shrublands, green: Grasslands, red: Croplands, black: Mixed).~~


$r = -0.55,\ p < 0.05$
a
$r = -0.63,\ p < 0.05$
b
$r = -0.81,\ p < 0.05$
c
$r = 0.83,\ p < 0.05$
d
Shrublands
Grasslands
Croplands
Mixed
Pearson correlation between SMAP DCA and *in situ* (unitless)
Unique information of T_{Bh}
Unique information of T_{Bv}
Synergistic inforamtion of T_{Bh} and T_{Bv}
Redundant information of T_{Bh} and T_{Bv}

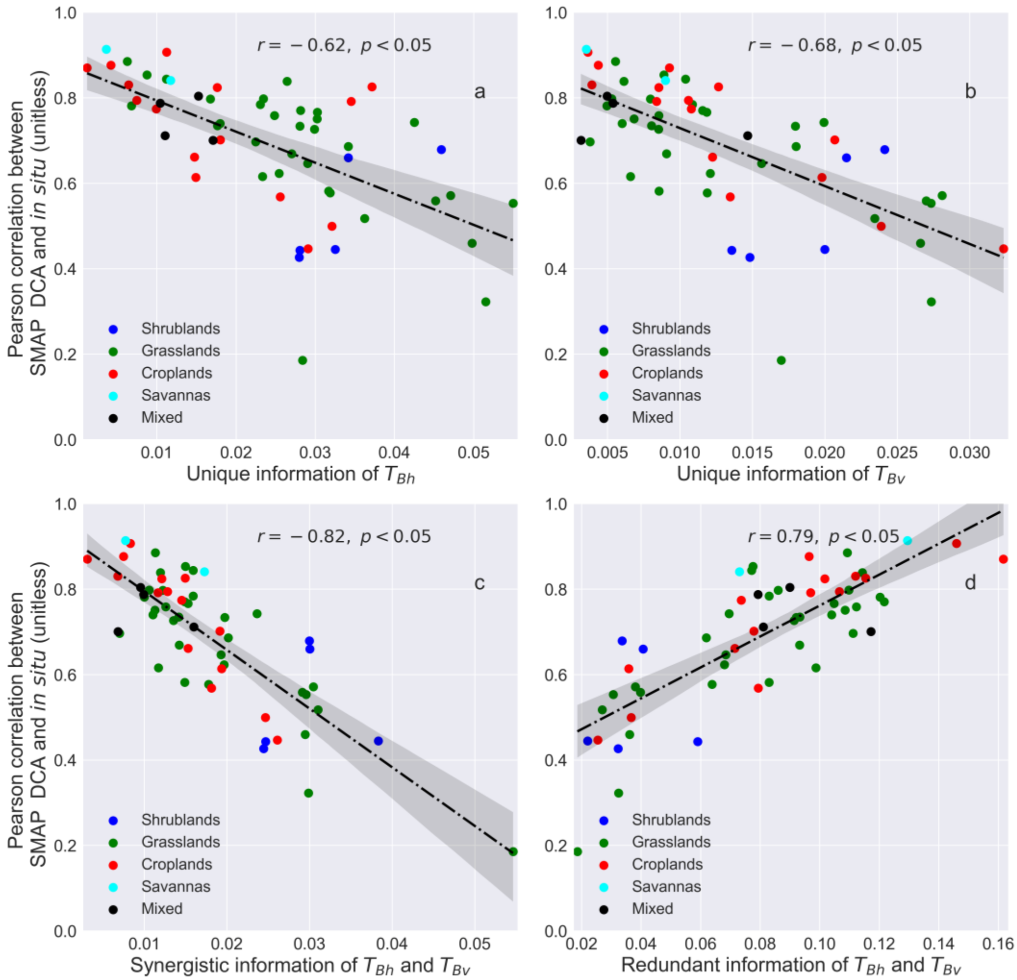


Figure 7 Partial information decomposition components between horizontally (T_{Bh}) and vertically (T_{Bv}) polarized brightness temperature against Pearson correlation coefficient between *in situ* and DCA soil moisture.

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

Table 1 The amount of informational uncertainties in percentage. The values in the table are the average of each landcover. The values in "Overall" is 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*))
Shrublands	0.22 (69%)	0.10 (31%)	0.32 (88%)
Grasslands	0.20 (62%)	0.09 (38%)	0.29 (83%)
Croplands	0.18 (65%)	0.10 (35%)	0.28 (79%)
Mixed	0.20 (68%)	0.09 (32%)	0.29 (81%)
Overall	0.18 (64%)	0.11 (36%)	0.29 (82%)

Table 1 The amount of informational uncertainties in percentage. The values in the table are the average of each landcover. The values in "Overall" is the average of all the sites.

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%)	0.038 (31%)	0.120	5
Grasslands	0.028 (20%)	0.013 (10%)	0.019 (14%)	0.080 (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

Table 2 The partial information decomposition components. The values in the table are the average of each landcover. The values in "Overall" is 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 ($I(T_{Bh}, T_{Bv}; DCA)$)
Shrublands	0.03 (27%)	0.017(15%)	0.03 (26%)	0.036 (32%)	0.113
Grasslands	0.029 (21%)	0.014 (10%)	0.02 (14%)	0.077 (55%)	0.14
Croplands	0.017 (12%)	0.013 (9%)	0.016 (12%)	0.095 (67%)	0.141
Mixed	0.014 (12%)	0.007 (6%)	0.01 (8%)	0.091(74%)	0.122
Overall	0.026 (19%)	0.013 (10%)	0.019 (14%)	0.08 (57%)	0.137

Table 2 The partial information decomposition components. The values in the table are the average of each landcover. The values in "Overall" is the average of all the sites.