https://doi.org/10.5194/hess-2020-534
Preprint. Discussion started: 28 October 2020

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

Information - based uncertainty decomposition in dual channel microwave remote sensing of soil moisture

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

[1]Department of Biological & Ecological Engineering, 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 relationship between vegetation opacity and surface scattering albedo (i.e. '*tau-omega*' model). Although this model has been validated against *in situ* soil moisture measurements across sparse validations sites, there is lack of systematic characterization of where and why SMAP estimates deviate from the *in situ* observations. Here, soil moisture observations from the US Climate Reference Network are used within a mutual information framework to decompose the overall retrieval uncertainty from SMAPs Modified Dual Channel Algorithm (MDCA) into random uncertainty derived from raw data itself and model uncertainty derived from the model's inherent structure. The results shown that, on average, 12% of the uncertainty in SMAP soil moisture estimates is caused by the loss of information in the MDCA model itself while the remainder is induced by inadequacy of T_{Bh} and T_{Bv} observations. We find the fraction of algorithm induced uncertainty is negatively correlated (pearson r of -0.48) with correlations between *in-situ* observations and MDCA estimates. A decomposition of mutual information between T_{Bh}, T_{Bv} and MDCA soil moisture shows that on average 55% of the mutual information is redundantly shared by T_{Bh} and T_{Bv}, while the information provided uniquely from both T_{Bh} and T_{Bv} is 15%. The fraction of information redundantly provided by T_{Bh} and T_{Bv} was found to be tightly correlated (pearson r = -0.7) to how well the MDCA output correlated to *in situ* observations. Thus, MDCA overall quality improves as T_{Bh} and T_{Bv} provide more redundant information for the MDCA. This suggests the informational redundancy between these remotely sensed observations can be used as independent metric to assess the overall 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.

https://doi.org/10.5194/hess-2020-534
Preprint. Discussion started: 28 October 2020

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

1 Introduction

Accurate information on soil moisture is of great importance to understand various of 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 (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 more desirable for soil moisture estimation because they can sense soil moisture at a relatively deeper layer (~5cm) and greater vegetation penetration (Njoku and Entekhabi, 1996). Though microwave remote sensing has been investigated for decades, significant uncertainties still exists in both microwave radiometry and in the algorithms used to translate microwave observations to soil moisture estimates.

L-band remote sensing soil moisture estimation uses a radiometer to measure surface emission intensity, which is a linear function of brightness temperature. The brightness temperature is linked with soil moisture and vegetation opacity through the '*tau-omega'* emission model and parameterized by soil and vegetation functions (Njoku and Entekhabi, 1996). The '*tau-omega'* model rationale has been adopted by SMAP, which is one of the earth observation missions dedicated to soil moisture estimation at L-band microwave frequency. SMAP implemented two primary algorithms: (1) single channel algorithm (SCA) that uses one polarized brightness temperature as primary input to retrieve soil moisture and (2) the modified dual channel algorithm (MDCA) that can retrieve soil moisture and vegetation opacity simultaneously by taking the advantage of polarized brightness temperature in both directions (Peggy O'Neill et al., 2018). There is strong interest in the MDCA approach because of its independent estimation of vegetation water status. Although SMAP can provide spatially explicit soil moisture estimates that have been shown to be useful to understand a set of ecohydrological problems (Jadidoleslam et al., 2019; Williams and Beer, 2010), the soil moisture retrievals are still subject to significant amount of uncertainty due to the imperfection of the model and the forcing datasets. Therefore, it is critical to diagnosis and quantify the causality of the uncertainty caused by the SMAP algorithm in order to improve the soil moisture retrieval accuracy.

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 at 'core' and 'sparse' validation sites (Babaeian et al., 2019; Colliander et al., 2017). 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 (Chen et al., 2017, 2018b). These validation investigations found that SMAP met the required accuracy target (0.04 cm^3/cm^3) on average, while there exist some locations where the performance of SMAP did not met the expected performance. This is because these validation studies all 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 from which the uncertainty arise. 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 form and parameterizations. In addition, the evaluation metrics used in these evaluation studies are either heavily depend on *in situ* soil moisture

https://doi.org/10.5194/hess-2020-534
Preprint. Discussion started: 28 October 2020


Hydrology and Earth System Sciences
Discussions
Open Access
EGU

or additional reference dataset, which challenges 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 describes the inherent uncertainty of a random variable and (2) mutual information, which represents the reduction in uncertainty of one random variable given the knowledge of another random variable. Gong et al. (2013) 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 (2) model uncertainty that is contributed by poor model parameterization. The random uncertainty is not resolvable for the given system as they are only related to the probability densities, while the model uncertainty is reduceable by a better model parameterization.

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 links two source variables with a target variable. It partitions multivariate mutual information into unique, redundant and synergistic components. The unique information represents the amount of information shared with the target variable only from each individual source variable. Synergistic information is the information provided to the target while both source variables act jointly. Redundant information is the overlapping information that both source variables redundantly provide to a target. Information partition brings a new insight into unambiguously characterizing the interdependencies between source variables and a target variable without any underlying modeling assumption. The partitioned components may be used as a new model evaluation metric that can be used to assess SMAP algorithm performance in remote and inaccessible regions.

In this study, we focus on (1) quantifying the random uncertainty and model uncertainty in SMAP Modified Dual Channel Algorithm (MDCA) and understand how model uncertainty is related to MDCA retrieval accuracy; (2) developing an *in situ* and ancillary data independent SMAP MDCA evaluation reference metric using partial information decomposition between SMAP MDCA soil moisture and horizontally polarized (T_{Bh}) and vertically polarized brightness temperature (T_{Bv}).

2 Material and Methods

2.1 *In situ* soil moisture

US Climate Reference Network (USCRN) is a systematic and sustained network 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 (Bell et al., 2013). The USCRN provides soil moisture observations at five different standard depth (5, 10 20, 50 and 100 cm) in 114 locations of Contiguous U.S. (CONUS). The *in situ* datasets have been used for a wide variety of research such as drought monitoring and satellite soil moisture evaluations (Mishra et al., 2017). The hourly soil moisture dataset at the depth of 5 cm was collected from 58 selected USCRN stations (Fig. 1) based on the availability *in situ* soil moisture dataset and the quality of SMAP pixels in the study period of 03/31/2015 – 10/01/2019.

2.2 MDCA soil moisture

In this study, we acquired T_{Bh}, T_{Bv} and MDCA soil moisture from the SMAP Enhanced Level-2 Radiometer Half-Orbit 9 km EASE-Grid Soil Moisture (https://nsidc.org/data/smap), Version 3 in the

https://doi.org/10.5194/hess-2020-534
Preprint. Discussion started: 28 October 2020

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

same period of the USCRN soil moisture at every station (Peggy O'Neill et al., 2018). The extracted data series were filter by their respective quality flags and the T_{Bh}, T_{Bv} and MDCA soil moisture values were kept only when they all simultaneous pass quality control. MDCA retrieves soil moisture based on the '*tau-omega'* model, which is a well- known radiative transfer-based soil moisture retrieval algorithm in the passive microwave soil moisture community. It requires the brightness temperature (T_B) as inputs and parameterized by overlaying vegetation and soil surface information. The MDCA invert the '*tau-omega'* model with initial guesses of surface soil moisture and vegetation optical depth. The guesses of soil moisture and vegetation optical depth are adjusted iteratively until they minimize the difference between satellite observed T_B and inverted T_B from a least square perspective. Compared to the SCAs, the MDCA updates roughness and the polarization mixing parameters (Chaubell et al., 2020).

2.3 Information - based uncertainty decomposition

Shannon's entropy is a quantity that express the inherent uncertainty associated with a random variable. Commonly, modeling efforts are focused on reducing the uncertainty in the variable of interest, which is denoted as $H(Y_{obs})$, using other explanatory variables through some physically- or empirically-based models. Most of models being constructed of natural processes are not perfect, and the model outputs are often not capable of capturing the information of the "truth". In theory, there exists a best achievable model performance that describe the variable of interest the best for a particular system given the available datasets (Gong et al., 2013); yet detailed structure of best achievable model performance is often unknown. Although the detailed structure of best achievable model performance maybe remain unknown, mutual information, denoted as $I(\boldsymbol{X}_{Inputs}; Y_{obs})$ where $\boldsymbol{X}_{Inputs}$ are the available inputs and Y_{obs} is the *in situ* measured variable of interest, can provide a good benchmark measure. The quantity $I(\bullet;\bullet)$ represents the amount of uncertainty reduced due to the knowledge of either variable in this function.

It should be noted that a model is a formal hypothesis that maps input datasets space to output dataset space in the form of a mathematical function. Therefore, the model hypothesis (function), at least, cannot provide new information. This is expressed as the data processing inequality which states that "no clever manipulation of the data can improve the inferences that can be made from the data" (Cover and Thomas, 2005). Formally, if random variables X, Y, Z are said to form a Markov chain (denoted by $X \rightarrow Y \rightarrow Z$), wherein the conditional distribution of Z only depends on Y and is conditionally independent of X, then X can only influence Z via the knowledge of Y and knowing Z can only decrease the amount of X tells about Y. The formula of data processing inequality is defined as:

$$I(X, Y) \geq I(X, Z) \tag{1}$$

Hence, given the measure of best achievable model performance and data processing inequality, the relationship between input, output, and *in situ* measurements in any modeling processes can be expressed as follows:

$$H(Y_{obs}) \geq I(\boldsymbol{X}_{Inputs}; Y_{obs}) \geq I(Y_{model}; Y_{obs}) \tag{2}$$

The relationship equation (2) allow us to differentiate two types of uncertainties, (1) random uncertainty, which unresolvable due to the randomness of the input datasets, that is the difference between $H(Y_{obs})$ and $I(\boldsymbol{X}_{Inputs}; Y_{obs})$; (2) model uncertainty, which is resolvable due to the inadequacy of model, that is the information gap between and $I(\boldsymbol{X}_{Inputs}; Y_{obs})$ and $I(Y_{model}; Y_{obs})$.

https://doi.org/10.5194/hess-2020-534
Preprint. Discussion started: 28 October 2020

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

In our case, $\boldsymbol{X}_{\text{Inputs}}$ are T_{Bh} and T_{Bv}, Y_{obs} is the *in situ* surface soil moisture, Y_{model} is MDCA soil moisture. The $H(Y_{obs})$ can be calculated as:

$$H(Y_{\text{obs}}) = -\sum_{y \in Y_{obs}} p(y) \log_2 p(y) \quad (3)$$

Where $p(y)$ probability mass function of Y_{obs} that is estimated by a fixed bin method (Freedman and Diaconis, 1981). This method calculates $H(Y_{obs})$ in unit of bits. Previous study has indicated that this method may underestimates the true entropy (Paninski, 2003). Therefore, we leveraged the simple Miller-Madow corrected entropy estimator (Chen et al., 2018a) 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 pick this Miller-Madow correction based on its simplicity and effectiveness. The corrected and normalized entropy is then expressed as follows:

$$H_{\text{CN}}(Y_{\text{obs}}) = \frac{H(Y_{\text{obs}}) + \frac{K-1}{2n}}{log_2^n} \quad (4)$$

Where $H_{CN}(Y_{obs})$ is the Miller-Madow corrected and normalized entropy, hereafter entropy, n is the number of data points that were used to calculate the normalized entropy, K is the number of non-zero probabilities associate based on the fixed binned method.

The computation of two types of uncertainties require the estimation of $I(T_{Bv}, T_{Bh}; Y_{obs})$ and $I(Y_{MDCA}; Y_{obs})$, which can be computed via the following equation:

$$I(T_{Bh}, T_{Bv}; Y_{obs}) = H_{CN}(T_{Bh}, TB_v) + H_{CN}(Y_{obs}) - H_{CN}(T_{Bh}, T_{Bv}, Y_{obs}) \quad (5)$$

Where $H_{CN}(T_{Bh}, T_{Bv})$ and $H_{CN}(T_{Bh}, T_{Bv}, Y_{obs})$ the estimated joint entropy that describes the uncertainty associated with a set of variables. $H_{CN}(Y_{MDCA}; Y_{obs})$ can be estimated by replacing the T_{Bh}, T_{Bv} with Y_{MDCA} on both side of the equation. It worth noting that the joint entropies are estimated using equation (3) except they require the estimation of joint probability mass functions that are also estimated using the fixed bin method (Freedman and Diaconis, 1981).

2.4 Partial information decomposition

This method partitions multivariate shared information to unique, redundant and synergistic components. The decomposed information components on the model inputs and outputs maybe indicative on understand informational loss as model inputs are translated to end user products and these components may have the potential for evaluating model performance. The partial information decomposition of MDCA can be expressed as follows:

$$I(T_{Bh}, T_{Bv}; Y_{MDCA}) = U_1(T_{Bh}; Y_{MDCA}) + U_2(T_{Bv}; Y_{MDCA}) + R(T_{Bh}, T_{Bv}; Y_{MDCA}) + S(T_{Bh}, T_{Bv}; Y_{MDCA}) \quad (6)$$

Where U_1 and U_2 are unique information of T_{Bh} and T_{Bv} shared with Y_{MDCA}, respectively. S and R are the synergistic information and redundant information that T_{Bh} and T_{Bv} shared with Y_{MDCA}, respectively. All

https://doi.org/10.5194/hess-2020-534
Preprint. Discussion started: 28 October 2020


Hydrology and Earth System Sciences
Discussions
Open Access
EGU

the decomposed components are non-negative real values.

The individual mutual information between T_{Bh}, T_{Bv} and Y_{MDCA} can be expressed as follows:

$$I(T_{Bh}; Y_{MDCA}) = U_1(T_{Bh}; Y_{MDCA}) + R(T_{Bh},T_{Bv}; Y_{MDCA}) \quad (7)$$

$$I(T_{Bv}; Y_{MDCA}) = U_1(T_{Bv}; Y_{MDCA}) + R(T_{Bh},T_{Bv}; Y_{MDCA}) \quad (8)$$

U_1, U_2, S and R are unknowns in the systems of equations (6) - (8). Therefore, additional information is need to fully estimated one of these unknowns. We used the approach proposed by Goodwell and Kumar (2017) to estimate R as follows:

$$R = R_{min} + I_s*(R_{MMI} - R_{min}) \quad (9)$$

Where R_{min} is represents a lower bound for R that is expressed as:

$$R_{min} = \max(0, -II) \quad (10)$$

The inter-dependency of T_{Bh} and T_{Bv} represented by I_s and computed as:

$$I_s = \frac{I(T_{Bh};T_{Bv})}{\min\{H(T_{Bh}); H(T_{Bv})\}} \quad (11)$$

II is interaction information that can be positive or negative. II is computed as:

$$II = I(T_{Bh}; Y_{MDCA}|T_{Bv}) - I(T_{Bh}; Y_{MDCA}) \quad (12)$$

3 Results

3.1 Information quantities and system uncertainties

Figure 2 shows the estimated entropy and mutual information quantities across the study sites. It is shown that the joint entropy of T_{Bh} and T_{Bv} (H_{CN}(h,v)) are always the largest compared to other information quantities. On average, H_{CN}(h,v) is 0.53 bits, which is greater than the entropies of MDCA soil moisture, H_{CN}(MDCA), and *in situ* soil moisture, H_{CN}(*in situ*), (0.38 and 0.35, respectively). Although the pattern of H_{CN}(MDCA) and H_{CN}(*in situ*) are similar, the H_{CN}(*in situ*) is more variable than H_{CN}(MDCA) with the coefficients of variation (CV) being 0.08 and 0.05, respectively. Mutual information between T_{Bh}, T_{Bv} and *in situ* soil moisture, I(h,v; *In situ*), and mutual information between MDCA soil moisture and *in situ* soil moisture, I(MDCA; *In situ*), are the least information quantities, as they are expected to be. I(h,v; *In situ*) follows the pattern of I(MDCA; *In situ*) with the mean values being 0.09 and 0.06, respectively.

It is noticeable that there exists large information gaps (Fig. 2) between H_{CN}(*in situ*) and I(h,v; *in situ*) and I(h,v; *In situ*) and I(MDCA; *in situ*). H_{CN}(*in situ*) represent the amount of information that is required to fully characterize the "true "soil moisture, while I(h,v; *in situ*) indicates the available information contained in the system input variable about the "true" soil moisture. The information gap between H_{CN}(*in situ*) and I(MDCA; *in situ*) is the overall SMAP uncertainty in which 88% is contributed by the random uncertainty in the systems explanatory variables (Fig. 3). The information gap between

https://doi.org/10.5194/hess-2020-534
Preprint. Discussion started: 28 October 2020


Hydrology and Earth System Sciences
Discussions
Open Access
EGU

I(h,v; *In situ*) and I(MDCA; *in situ*) represents the MDCA model uncertainty, which contributes 12% of the total uncertainty (Fig. 3).

3.2 Model uncertainty and retrieval accuracy

Figure 4 shows the relationship between the fraction of model uncertainty against different commonly adopted absolute (Fig. 4a) and relative model evaluation metrics (Fig. 4b). The model uncertainty is shown to be tightly related to these metrics. It is observed that the fraction of MDCA induced uncertainty is positively correlated (r = 0.28) with RMSE of *in-situ* soil moisture and MDCA soil moisture (Fig. 4a). An obvious negative relationship is found when it comes to the relationship between the fraction of MDCA induced uncertainty and r of MDCA soil moisture and *in situ* soil moisture (r = -0.48). Both the positive and negative relationship are in line with general expectations since model uncertainty should go up when the retrieval accuracy is poor and vice versa.

3.3 Partial information decomposition of MDCA

Figure 5 illustrates that majority of the mutual information between T_{Bh}, T_{Bv} and MDCA (I(T_{Bh},T_{Bv}; MDCA)) is redundantly shared by T_{Bh} and T_{Bv}, which take about 0.55 of I(T_{Bh},T_{Bv}; MDCA) on average (Fig. 5). U_h is comparable to S with a mean value of 0.15, respectively. Compared to other decomposed information components, U_v is the smallest but is of similar magnitude with U_h and S with mean being 0.14. Although the R is the largest information component, it has the smallest CV (0.35) compared to U_h (CV = 0.58), U_v (CV = 0.52) and S (CV = 0.63). In general, the MDCA system is dominated by R. This indicates that both T_{Bh} and T_{Bv} provide information regarding the soil moisture estimations, but these two variables are themselves highly dependent.

3.4 Partial information decomposition and retrieval accuracy

Figure 6 shows the relationship between different decomposed information components and the RMSE of *in situ* and MDCA soil moisture. In general, only U_h is significantly negatively correlated (r = -0.28) with the RMSE of *in situ* and MDCA soil moisture (Fig. 6a), while relationships between RMSE and other components are not statistically significant (Fig. 6b – Fig. 6d). Figure 7 shows the relationship between different information components and the r of in situ and MDCA soil moisture. This demonstrates that all the information components are significantly correlated with the correlation, r, of *in situ* and MDCA soil moisture. U_h, U_v and S are negatively (Fig. 7a – Fig. 7c) correlated with r, while R is positively correlated with r. R shows the strongest correlation (Fig. 7d) with the relative model evaluation metric (r = 0.7). This indicates that R could potentially be a reference metric for MDCA evaluation. It does not require *in situ* soil moisture and shows a better performance than simply using r (Fig. 7d inset).

4 Discussion

4.1 Random uncertainty and model uncertainty

The first objective of this study is to leverage information theory to quantitatively decompose the overall uncertainty to random uncertainty and model uncertainty in the MDCA as an approach to understand where retrieval errors arise. This information theory approach can add considerable power to SMAP modeling diagnosis. Mutual information can provide a way to unambiguously define the best model performance that is able to completely transform the available information to the desired target given a set of the input data.

https://doi.org/10.5194/hess-2020-534
Preprint. Discussion started: 28 October 2020


Hydrology and Earth System Sciences
Discussions
Open Access
EGU

In this study, any model based on MDCA model structure is a hypothesis that relates T_{Bh} and T_{Bv} to soil moisture based on prior physical knowledge. The essence of the model is a procedure of processing the T_{Bh} and T_{Bv} to get soil moisture. The modeled soil moisture is deemed as an estimate of "true" soil moisture and a Markov chain is formed from T_{Bh}, T_{Bv} via MDCA soil moisture to *in situ* soil moisture. Any model, even the one performs the best, can only reduce the available information in its primary inputs (T_{Bh} and T_{Bv}) and is not capable of add new information about the "true" soil moisture. Hence, there is no chance of building a model that is better than the benchmark one (yet even achieving this theoretically limit is nearly impossible) if no freedom is given to the available datasets. If, however, given more freedom on available datasets, it is possible to build models that outperform best achievable model performance by adding new explanatory variables which will lead to a family of models that have completely different model structure. Additionally, the fraction that random uncertainty contributes to the overall uncertainty is quite significant (88% on average) in this study. The random uncertainty in the system may arises from the inherent error due to calibration of T_{Bh} and T_{Bv} in the locations and the presence water body. If poorly calibrated, the soil moisture estimations can be exacerbated due to the error propagation that hinders the correct information being transformed. Therefore, for example, a better and robust calibration strategy of T_{Bh} and T_{Bv} to the presence of water body might need. Furthermore, a better quality-control method or additional data screening metric with respect to water corrected T_{Bh} and T_{Bv} is also required to further reduce the random uncertainty.

Apart from random uncertainty, the model uncertainty contribution is also a significant amount the total (12% on average). This model uncertainty may arise from poor model parameterizations. It's known that the '*tau-omega'* model in MDCA is parameterized by landcover based parameters. The values of these parameters are derived from past studies, past experience and some information discussions with subject matter experts, which could be biased and inaccurate (Peggy O'Neill et al., 2018). In addition, these parameter values are differentiated by landcover and do not vary in time and microwave polarization directions. In fact, these parameters may not vary in short time (days or weeks) but could vary from a long-term perspective (month or years) and the parameter associated with vegetation structure may vary correspondent to different phenology phases.

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 assessing approach that can quantitively identify where information lost in the process of SMAP soil moisture modeling. This analysis, though focused on MDCA soil moisture, can be transferred and extended to analyze any other remote sensing models.

4.2 Model evaluation from another perspective

The second objective was to demonstrate is that partitioned information components can be used as a new MDCA model evaluation metric that does not depend on *in situ* soil moisture and other ancillary datasets. We found a strong linear relationship between *R* and *r* of MDCA and *in situ* soil moisture, which indicated that T_{Bh} and T_{Bv} are highly dependent. *R* is also the dominant component relative to others quantified here. From an information perspective, higher or complete *R* indicates that one source variable is a function of the other, or they share the same source. It can be observed from Figure 7d (inset) that there is a strong linear relationship between TB_v and TB_h ($r \geq 0.94$). Therefore, it is expected that a higher redundancy in the MDCA system. The MDCA takes T_{Bv} and T_{Bh} as primary inputs while T_{Bh} and T_{Bv} share a lot of redundancy. Therefore, it is not surprising that the MDCA soil moisture underperforms

https://doi.org/10.5194/hess-2020-534
Preprint. Discussion started: 28 October 2020


Hydrology and Earth System Sciences
Discussions

SMAP SCA soil moisture due to the error accumulation and error propagation from both channels.

To summarize, the redundant information shown a strong correlation with *r*, which could be potentially used as a MDCA evaluation metric. This metric only involves T_{Bh}, T_{Bv} and MDCA soil moisture and doesn't depend on *in situ* measurement and ancillary dataset. Compared to another *in situ* independent metric, such as pearson *r* of T_{Bh} and T_{Bv}, it shows a better performance (0.70 vs 0.52). This is potentially due to numerous non-linear processes acting within the MDCA, which are not well captured by linear metrics such as the pearson *r* of T_{Bh} and T_{Bv}

5 Conclusion and Limitations

This study attempts to differentiate and quantify the uncertainty sources in MDCA using information theoretic. We found that on average 88% of the uncertainty is contributed by the inadequacy of explanatory variables of SMAP or uncertainties in the estimated brightness temperature, while the rest of the uncertainty is induced by inaccurate MDCA parameterizations. The fraction of the model uncertainty to the overall uncertainty is negatively correlated with the pearson *r* of *in situ* and MDCA soil moisture (r = -0.48) while positively correlated with the error between in situ and MDCA soil moisture (r = 0.28). The decomposition of the mutual information has shown that all decomposed components are correlated with the pearson *r* between in situ and MDCA soil moisture with the redundant information being the tightest (r = 0.7). The uncertainty decomposition analysis opens a new window for SMAP algorithm uncertainty diagnosis. The result of mutual information decomposition analysis can be adopted as a new *in situ* independent SMAP soil moisture evaluation reference metric.

We acknowledge the existence of limitations of this study. First, we expect that this approach can be generalized to analyze other remote sensing models. However, it may be difficult to compute the joint probability density function for models with high-dimensional inputs, and thus also difficult to estimate the joint entropy and mutual information components. Though there exist several approaches for computing joint entropy and mutual information, the caveat here is that it is not guaranteed that the estimated mutual information can be exactly the entropy and joint entropy that fulfils the equality of, for instance, equation (5). Second, this study was conducted at locations where *in situ* soil moisture readily available. 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. Third, we would expect that the information theoretic to provide asymptotic estimation of random and model uncertainties, the best performance we can expect from this current uncertainty analysis is to use all of the available datasets we have; yet we believe that uncertainty estimations of this approach should be stabilized given adequate representative locations and data records.

Code availability

The code regarding the SMAP datasets time series extraction, mutual information and partial information decomposition calculation is upon request.

Data availability

SMAP Enhanced Level-2 Radiometer Half-Orbit 9 km EASE-Grid Soil Moisture, Version 3 is acquired from US National Snow and Ice Data Center (https://nsidc.org/data/smap). The in situ soil moisture is accessible through U.S. Climate Reference Network (https://www.ncdc.noaa.gov/crn/).

Author contribution

https://doi.org/10.5194/hess-2020-534
Preprint. Discussion started: 28 October 2020

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

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.

Chaubell, M. J., Yueh, S. H., Scott Dunbar, R., 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., Collins, C. H., 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, C., Grabchak, M., Stewart, A., Zhang, J. and Zhang, Z.: Normal Laws for Two Entropy Estimators on Infinite Alphabets, Entropy, 20(5), 371, doi:10.3390/e20050371, 2018a.

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.

Chen, F., Crow, W. T., Bindlish, R., Colliander, A., Burgin, M. S., Asanuma, J. and Aida, K.: Global-scale evaluation of SMAP, SMOS and ASCAT soil moisture products using triple collocation, Remote Sens. Environ., 214(March), 1–13, doi:10.1016/j.rse.2018.05.008, 2018b.

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.

https://doi.org/10.5194/hess-2020-534
Preprint. Discussion started: 28 October 2020

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

Freedman, D. and Diaconis, P.: On the histogram as a density estimator:L 2 theory, Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebieter Wahrscheinlichkeitstheorie und Verwandte Gebiete, 57(4), 453–476, doi:10.1007/BF01025868, 1981.

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.

Jadidoleslam, N., Mantilla, R., Krajewski, W. F. and Goska, R.: Investigating the role of antecedent SMAP satellite soil moisture, radar rainfall and MODIS vegetation on runoff production in an agricultural region, J. Hydrol., 579, 124210, doi:10.1016/j.jhydrol.2019.124210, 2019.

Mishra, A., Vu, T., Veettil, A. V. and Entekhabi, D.: Drought monitoring with soil moisture active passive (SMAP) measurements, J. Hydrol., 552(January 2015), 620–632, doi:10.1016/j.jhydrol.2017.07.033, 2017.

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.

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

Peggy O'Neill, Rajat Bindlish, Steven Chan, Eni Njoku and Tom Jackson: Soil Moisture Active Passive (SMAP) Algorithm Theoretical Basis Document SMAP L2 & L3 Radar Soil Moisture (Active) Data Products, Jet Propuls. Lab., Calif. Inst. Technol., Pasadena, CA, USA, JPL D-66480 [online] Available from: http://smap.jpl.nasa.gov/files/smap2/L2&3_SM_AP_InitRel_v11.pdf, 2018.

Petropoulos, G. P., Ireland, G. and Barrett, B.: Surface soil moisture retrievals from remote sensing: Current status, products & future trends, Phys. Chem. Earth, 83–84, 36–56, doi:10.1016/j.pce.2015.02.009, 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.

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.

Williams, P. L. and Beer, R. D.: Nonnegative Decomposition of Multivariate Information, , 1–14 [online] Available from: http://arxiv.org/abs/1004.2515, 2010.

https://doi.org/10.5194/hess-2020-534
Preprint. Discussion started: 28 October 2020

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

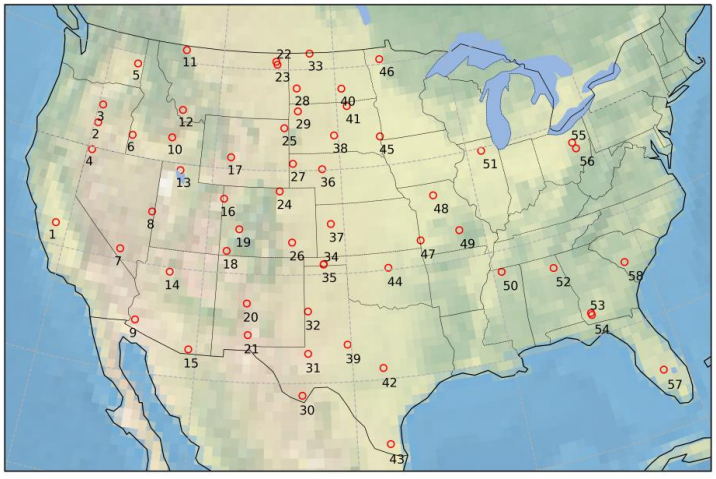


Figure 1 Spatial distribution of selected USCRN stations from west to east. See figure 2 caption for names of individual sites based on numbering.

470

475

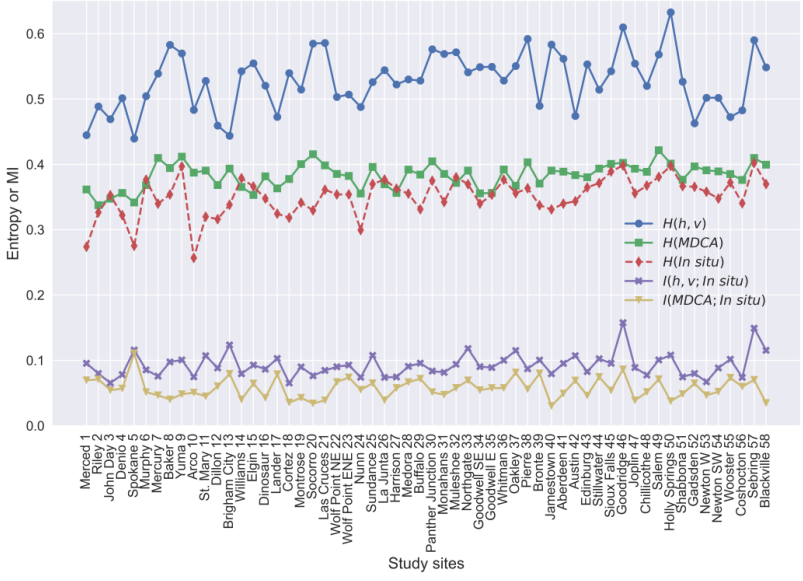
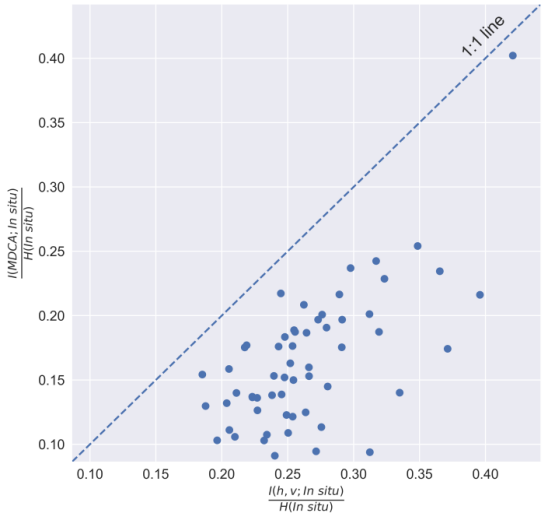


Figure 2 Information quantities of *in situ* soil moisture, T_{Bh}, T_{Bv} and MDCA soil moisture across the study sites.

https://doi.org/10.5194/hess-2020-534
Preprint. Discussion started: 28 October 2020

Hydrology and Earth System Sciences
Discussions
Open Access
EGU



480 **Figure 3 Mutual information between MDCA soil moisture and *in situ* soil moisture against mutual information between T_{Bh}, T_{Bv} and *in situ* soil moisture.**

485

490

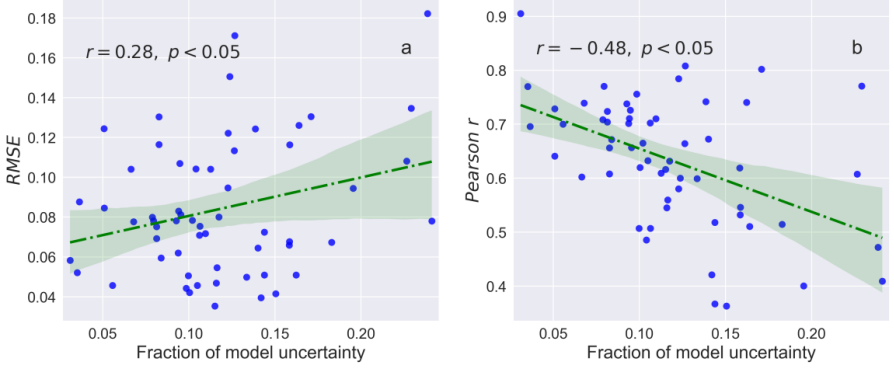


Figure 4 Fraction of MDCA model uncertainty against RMSE of MDCA soil moisture and *in situ* 495 **soil moisture (a) and fraction of MDCA model uncertainty against pearson *r* of MDCA soil moisture and *in situ* soil moisture (b).**

https://doi.org/10.5194/hess-2020-534
Preprint. Discussion started: 28 October 2020

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

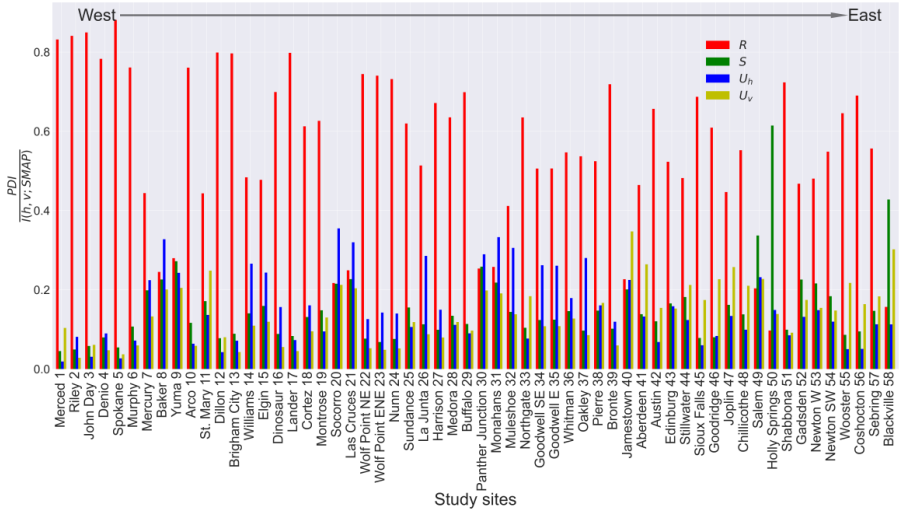
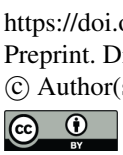


Figure 5 The normalized partial information decomposition components between T_{Bh}, T_{Bv} and MDCA soil moisture.

https://doi.org/10.5194/hess-2020-534
Preprint. Discussion started: 28 October 2020



Hydrology and Earth System Sciences
Discussions
Open Access
EGU

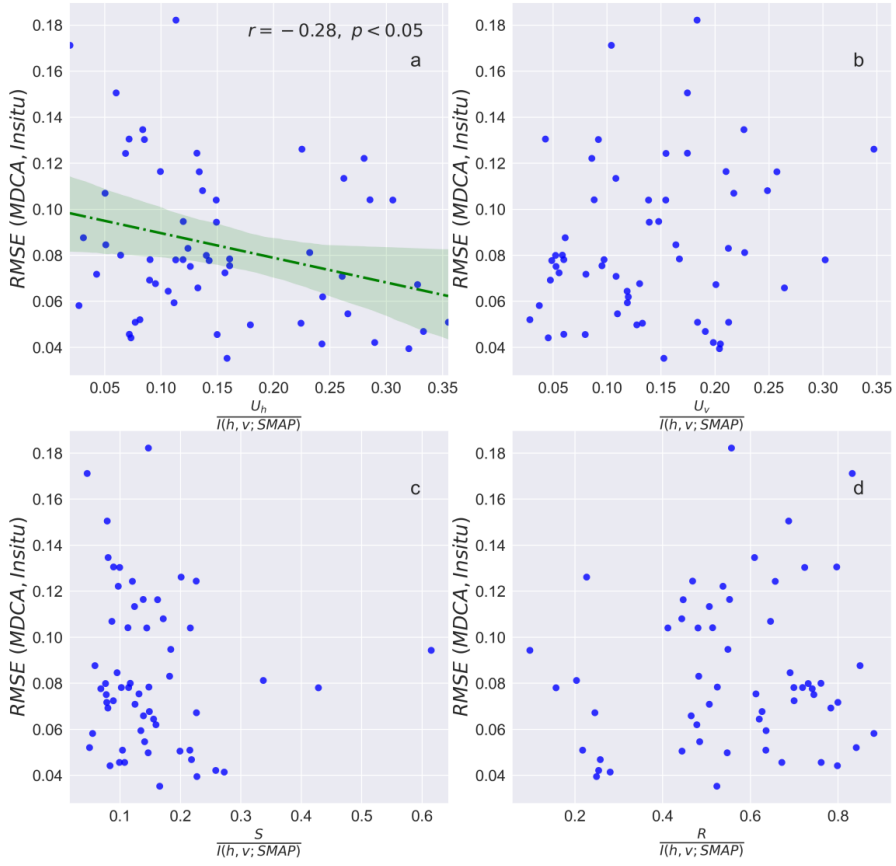
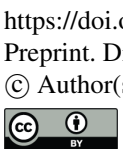


Figure 6 Normalized partial information decomposition components against RMSE of MDCA and *in situ* soil moisture.

https://doi.org/10.5194/hess-2020-534
Preprint. Discussion started: 28 October 2020



Hydrology and Earth System Sciences
Discussions
Open Access
EGU

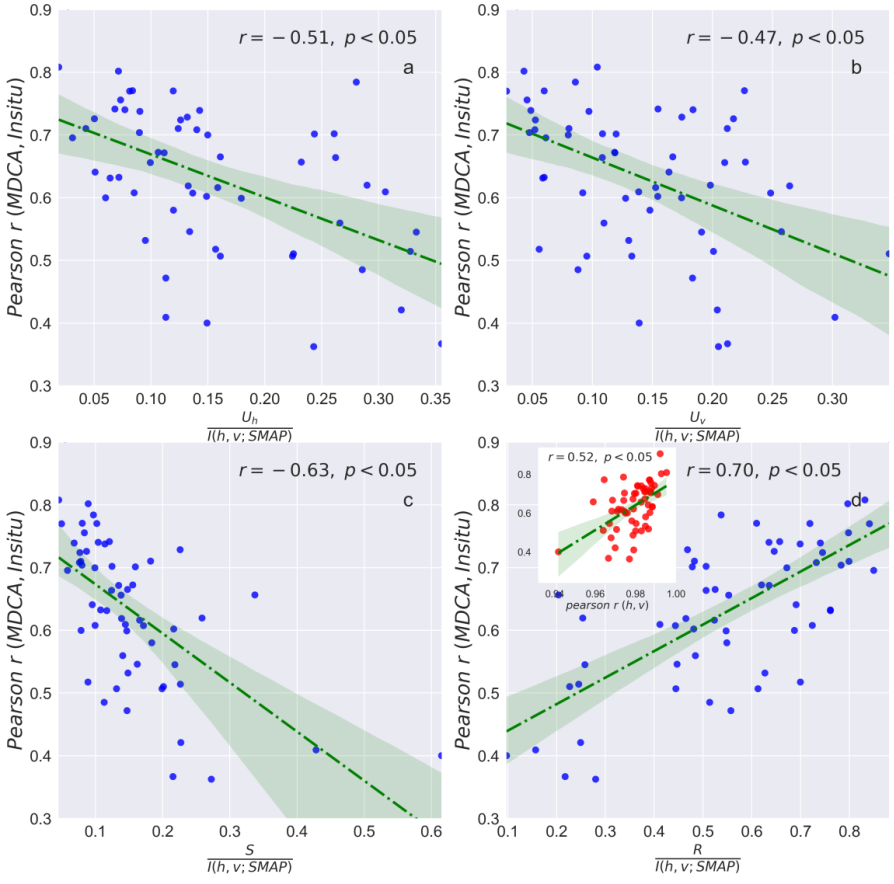


Figure 7 Normalized partial information decomposition components against pearson *r* of MDCA and *in situ* soil moisure.