Nonlinearity and multivariate dependencies in land-atmosphere coupling

Hsin Hsu[1] and Paul A Dirmeyer[1]

[1]George Mason University

August 30, 2023

Abstract

Most studies of land-atmosphere coupling have focused on bivariate linear statistics like correlation. However, more complex dependencies exist, including nonlinear relationships between components of land-atmosphere coupling and the transmutability of relationships between soil moisture and surface heat fluxes under different environmental conditions. In this study, a technique called multivariate mutual information, based on information theory, is used to quantify how surface heat fluxes depend on both surface energy and wetness conditions, i.e. net radiation and soil moisture, across the globe by season using reanalysis data. Such interdependency is then decomposed into linear and nonlinear contributions, which are further decomposed as different components explainable as the unique contribution from individual land surface conditions, redundant contributions shared by both land surface conditions, and the synergistic contribution from the coaction of net radiation and soil moisture. The dependency linearly contributed from soil moisture bears a similar global pattern to previously identified hot spots of coupling. The linear unique contributions of net radiation and soil moisture are mainly nonoverlapping, which suggests two separate regimes are governed by either energy or water limitations. These patterns persist when the nonlinearity is superimposed, thus reinforcing the validity of the land-atmospheric coupling hot spot paradigm and the spatial division of energy-limited as well as water-limited regions. Nevertheless, strong nonlinear relationships are detected, particularly over subtropical regions. Synergistic components are found across the globe, implying widespread multidimensional physical relationships among net radiation, soil moisture, and surface heat fluxes that previously had only been inferred locally.

Nonlinearity and multivariate dependencies in land-atmosphere coupling

Hsin Hsu[1] and Paul A. Dirmeyer[1,2]

[1]George Mason University, Fairfax, VA, USA

[2]Center for Ocean-Land-Atmosphere Studies, George Mason University, Fairfax, VA, USA

Key Points:

- Nonlinear and multivariate dependencies in land-atmospheric coupling are diagnosed using multivariate mutual information (MMI)
- Integrated MMI method decomposes drivers of surface flux variability into linear vs nonlinear; unique, redundant and synergistic components
- Nonlinear and synergistic dependencies of surface heat fluxes on soil moisture and net radiation are widespread across the globe

Abstract

Most studies of land-atmosphere coupling have focused on bivariate linear statistics like correlation. However, more complex dependencies exist, including nonlinear relationships between components of land-atmosphere coupling and the transmutability of relationships between soil moisture and surface heat fluxes under different environmental conditions. In this study, a technique called multivariate mutual information, based on information theory, is used to quantify how surface heat fluxes depend on both surface energy and wetness conditions, i.e. net radiation and soil moisture, across the globe by season using reanalysis data. Such interdependency is then decomposed into linear and nonlinear contributions, which are further decomposed as different components explainable as the unique contribution from individual land surface conditions, redundant contributions shared by both land surface conditions, and the synergistic contribution from the coaction of net radiation and soil moisture. The dependency linearly contributed from soil moisture bears a similar global pattern to previously identified hot spots of coupling. The linear unique contributions of net radiation and soil moisture are mainly nonoverlapping, which suggests two separate regimes are governed by either energy or water limitations. These patterns persist when the nonlinearity is superimposed, thus reinforcing the validity of the land-atmospheric coupling hot spot paradigm and the spatial division of energy-limited as well as water-limited regions. Nevertheless, strong nonlinear relationships are detected, particularly over subtropical regions. Synergistic components are found across the globe, implying widespread multidimensional physical relationships among net radiation, soil moisture, and surface heat fluxes that previously had only been inferred locally.

Plain Language Summary

Most of our knowledge of how land surface states affect weather and climate around the world is based on common statistical methods that assume straight lines can be fitted to determine how evaporation and heating of the air can be controlled by soil moisture. Furthermore, studies have historically looked one-at-a-time at a single cause affecting a single response. But local studies have shown that coupling between land and atmosphere can be more complex, involving multiple factors working simultaneously and often nonlinearly. This study uses a novel combination of techniques to discern how variations in processes like evaporation are driven by multiple factors such as soil moisture and available soil and thermal energy at the land surface that can make separately unique contributions, redundant contributions, and compound contributions due to factor interaction. Each of these contributions is also broken down into simple linear and other nonlinear components, helping show how much land-atmosphere interaction the previous methods were missing, and suggesting which regions, seasons and potential physical processes that need further study.

1. Introduction

Land-atmosphere coupling acting through water and energy cycles is an important component of the Earth system (Seneviratne et al. 2010). In addition to the process of precipitation directly moistening the land surface, the return process of the moistened land influencing the atmosphere is rather intricate (Eltahir 1998). Thus, efforts have been made to understand such soil moisture-precipitation feedbacks. Conventionally, the mechanism behind the feedback can be divided into a terrestrial leg and an atmospheric leg (Santanello et al. 2018). The terrestrial leg of coupling describes the ability of soil moisture to affect the partitioning of surface heat fluxes. The atmospheric leg of coupling characterizes how surface heat fluxes can modify the properties of the lower atmosphere and ultimately cloud formation and precipitation. Generally, with sufficient available energy, anomalously wet soil can moisten the atmosphere by enhancing evapotranspiration, cooling the surface while also cooling the overlying atmosphere by suppressing the release of sensible heat flux. The resulting moistened atmosphere and suppressed boundary layer growth compete to determine the net impact on cloud formation and thus precipitation.

Traditionally, studies attempting to identify regions of strong land-atmosphere coupling have used statistical frameworks with linear dependencies between two factors, of which the influences of temporal variability are hypothesized to be significant in the explored leg of coupling. For instance, coupling simultaneously throughout both legs has been diagnosed by the proportion of precipitation variance explained by soil moisture using output from multiple climate models (Koster et al. 2004) during boreal summer and has shown that regions with strong land-atmosphere coupling are mostly located in the semiarid regions. A similar pattern has been detected by reanalysis data with a sensitivity index involving variances and correlations of soil moisture and surface heat fluxes (Dirmeyer 2011). These studies have described a canonical global pattern of land-atmosphere coupling strength and thus many observational studies have explored the land surface processes over hot spots such as the North American Great Plains (e.g., Santanello et al. 2013, Tao et al. 2019), the Sahel (e.g., Los et al. 2006, Yu et al. 2017), and Australia (e.g., Kala et al. 2015, Herold et al. 2016). Since the soil moisture-surface heat flux-precipitation feedback processes have long been demonstrated as a potential key to improve the skill of subseasonal to seasonal forecasts, the importance of land-atmosphere feedbacks in such regions with strong land-atmosphere coupling is matched only by the eastern equatorial Pacific, a hot spot of ocean-atmosphere coupling associated with the phenomenon known as El Niño-Southern Oscillation (ENSO).

However, dependencies among the environmental factors within land-atmosphere coupling may be more complex than can be depicted by linear frameworks. A classic exemplification is the long-recognized threshold behavior in which the flux of latent heat behaves dramatically differently once soil moisture crosses a certain critical value (Budyko 1963, 1974). Another source is multivariate dependence, which is revealed in a recent in situ observation-based analyses showing the relationship between soil moisture and surface heat flux is non-unique (Haghighi et al. 2018). The embedded complex dependencies that cannot be recognized through the canonical linear coupling framework motivates to this study, re-examining land-atmosphere coupling strength while addressing its nonlinear and multivariate aspects.

Here we assess the terrestrial leg of coupling by quantifying both linear and nonlinear dependencies on surface heat fluxes by energy and moisture availability at the land surface, for which net radiation and soil moisture are used as proxies respectively. The techniques used are based on information theory. Following the commonly used terminology in information theory to describe causes and effects, soil moisture and net radiation are referred to as source variables and surface heat flux is called the target variable. Dependency between the target and one source is called mutual information (MI) and dependency between the target and multiple sources is called multivariate mutual information (MMI). The global pattern of MMI and its seasonal cycle are calculated. Furthermore, we integrate the methods of calculating nonlinearity in mutual information with the method of decomposing the MMI into different contributed components. This enables us also to separate MMI into linear and nonlinear components, interpretable as the unique information contributed by a particular source, the redundant information provided identically by both sources, and the synergistic information created by the interaction of the sources. The diagnosed nonlinearity and synergistic components reveal unexplored aspects of land-atmosphere coupling that can inform process understanding with potential to improve model parameterizations and prediction skill.

2. Methodology and Data

In this section, we first introduce the existing information measurements and their partitioning. We propose a combination of those measurements to decompose the information in land surface states and fluxes into their basic components. Then, the integrated application of the technique is illustrated. Lastly, the data and specifics of significance testing are described.

2.1 Information measurement

2.1.1 Shannon Entropy

Shannon Entropy H (Shannon 1948) quantifies the amount of uncertainty of a single random variable X with the probability distribution function $p(x)$:

$$H(X) = -\sum p(x)\log_2 p(x)$$

Eq 1

Various bases of the logarithm have been used in different applications; here base-2 is used, so the quantity is in units of bits. In this study, only the probability distribution in time is examined. For summation, $p(x)$ must be expressed across a finite number of bins – the procedure for bin selection is discussed in the Appendix.

2.1.2 Mutual information

Conditional entropy $H(X_{tar}|X_s)$ is an expression of Shannon entropy that quantifies the amount of uncertainty of a target variable X_{tar} given knowledge of a single source variable X_s:

$$H(X_{tar} \mid X_s) = -\sum p(x_s, x_{tar})\log_2 \frac{p(x_s, x_{tar})}{p(x_s)}$$

Eq 2

Mutual information (MI) (Cover and Thomas 1991) measures the "similarity" of information between two random variables. In other words, MI quantifies the reduction in uncertainty of one target variable X_{tar} by the knowledge of a source variable X_s. For a pair of random variables $(X_s; X_{tar})$, MI is given by:

$$I(X_s; X_{tar}) = \sum p(x_s, x_{tar}) \log_2\left(\frac{p(x_s, x_{tar})}{p(x_s)\mathrm{p}(x_{\mathrm{tar}})}\right) = H(X_{tar}) - H(X_{tar} \mid X_s) \qquad \text{Eq 3}$$

2.1.3 Multivariate mutual information

Multivariate mutual information (MMI) $I(X_{s1}, \dots X_{sn}; X_{tar})$ measures the reduction in uncertainty of one target variable X_{tar} by knowledge of multiple source variables. In this study, the simplest case involving two source variables X_{s1} and X_{s2} is examined, of which the function is given as:

$$I(X_{s1}; X_{s2}; X_{tar}) = \sum p(x_{s1}, x_{s2}, x_{tar}) \log_2\left(\frac{p(x_{s1}, x_{\mathrm{s}2}, x_{tar})}{p(x_{s1}, x_{s2})\mathrm{p}(x_{\mathrm{tar}})}\right) \qquad \text{Eq 4}$$

2.1.4 Temporal information

Temporal information is a technique to measure the evolution of dependencies among variables by applying MI or MMI with moving time windows (Goodwell and Kumar 2017a, 2017b). Instead of obtaining a single value MI or MMI for complete time series, temporal information technique cuts the whole time series into several time windows and obtains MI or MMI for each time window. In a climate application like this, such a measurement is tailored to detect the dependency among environmental factors considering that the dependency could vary due to factors such as seasonality or weather conditions.

2.2 Existing information partitioning approaches

Two perspectives of the partitioning approach have been proposed in past studies. One is partitioning of information into a linear part and a nonlinear part (Smith 2015), which is applicable to either MI or MMI. The other is only relevant to MMI: partitioning of information into components representing the interactive roles of multiple source variables, namely: unique, redundant and synergistic components (Williams and Beer 2010, Goodwell and Kumar 2017a, 2017b). Here we describe both approaches and show how they may be combined to decompose information in more detail.

2.2.1 Nonlinearity and linearity

Total information can be partitioned into linear information and nonlinear information. The method of calculating the nonlinearity in MI has been proposed by Smith (2015). For a given set of X_s and X_{tar}, the procedure is to: (1) fit a linear regression model in terms of predicting X_{tar} given X_s; (2) obtain $\hat{X}_{tar}$ as the fitted values of X_{tar} and define the nonlinear residual $X'_{tar} = X_{tar} - \hat{X}_{tar}$; (3) normalize X_{tar}' by the quantile normalization based on the value of X_{tar} (quantile normalization makes the distribution of X_{tar}' and X_{tar} identical in statistical properties; see Smith 2015 for more detailed discussion); (4) estimate the MI for both: $I(X_s;X_{tar})$ and $I(X_s;X_{tar}')$. The quantity $I(X_s;X_{tar}')$ is the nonlinear dependency between X_s and X_{tar}. The linear dependency in terms of MI is the difference (total minus nonlinear): $I(X_s;X_{tar}) - I(X_s;X_{tar}')$.

2.2.2 Unique, redundant, and synergistic components

For a system composed of two sources and one target, the total MMI can be decomposed into synergistic, unique, and redundant components. The unique (*U*) component is the peculiar information shared only between an individual source and the target. Redundancy (*R*) is repeated information that both sources share with the target. The synergistic (*S*) component is the extra information arising from the cooperative action among the sources. The expression of partitioning is given as the follows:

$$I(X_{s1}, X_{s2}; X_{tar}) = U_1(X_{tar}; X_{S1}) + U_2(X_{tar}; X_{S2}) + R(X_{tar}; X_{s1}, X_{s2}) + S(X_{tar}; X_{s1}, X_{s2}) \quad \text{Eq 5}$$

where U_1, U_2, R, and S are nonnegative quantities. Note that each source has its own individual unique contribution, whereas the redundant and synergistic contributions involve both sources.

Using this conceptualization, mutual information between each source and the target can be decomposed as the sum of unique and redundant components:

$$I(X_{s1};X_{tar}) = U_1(X_{tar};X_{S1}) + R(X_{tar};X_{s1},X_{s2}) \qquad \text{Eq 6}$$

$$I(X_{s2};X_{tar}) = U_2(X_{tar};X_{S2}) + R(X_{tar};X_{s1},X_{s2}) \qquad \text{Eq 7}$$

The above equation set requires one to seek an additional equation for any of U_1, U_2, R, and S to obtain all components (note that any I can be calculated directly from the probability density function of data). Several ways have been proposed to achieve a well-determined system by estimating U or R, while there is no universal agreement on the appropriate approach. In this study, we use the approach proposed by Goodwell and Kumar (2017a and 2017b), which assumes that the strength of the dependency of the two sources determines the amount of redundant information. To achieve this, a measurement called Rescaled Redundancy (R_s) is introduced as the follows:

$$R_s = R_{min} + I_s(R_{MMI} - R_{min}) \qquad \text{Eq 8}$$

The solution of R_s is obtained by computing the normalized source dependency I_s, the lower bounds of redundancy R_{min}, the upper bounds of redundancy R_{MMI}, and the interaction information □; these are given as follows:

$$I_s = \frac{I(X_{s1};X_{s2})}{\min[H(X_{s1}),H(X_{s2})]} = \frac{I(X_{s1};X_{s2})}{\min[I(X_{s1};X_{s1}),I(X_{s2};X_{s2})]} \qquad \text{Eq 9}$$

$$R_{MMI} = \min[I(X_{s1};X_{tar}),I(X_{s2};X_{tar})] \qquad \text{Eq 10}$$

$$R_{min} = \max(0,-□) \qquad \text{Eq 11}$$

$$□ = I(X_{s1};X_{s2};X_{tar}) \qquad \text{Eq 12}$$

The interaction information □ can be either positive or negative and □ is shown by

Williams and Beer (2010) to be equal to $S - R$. With stronger dependency between the two sources, a larger normalized source dependency I_s results in a larger redundant component R.

2.3. Integrated information measurement

Our attempt to quantify the dependency among multiple interacting variables and to disentangle the information as different explainable components can be achieved by combining the two approaches with an additional step.

We extend the approach of calculating nonlinearity from the bivariate (MI) to the trivariate (MMI) case. For a given set of X_{S1}, X_{S2}, and X_{tar}, a linear regression model is fitted in terms of predicting X_{tar} given both X_{S1} and X_{S2}. The rest of the procedure to obtain the nonlinear multivariate mutual information $I(X_{s1},X_{s2};X_{tar}')$ follows as described above. Subsequently, the two partitioning frameworks can be fitted together perfectly. The full decomposition of MMI by the integrated approach gives eight components relating two source variables to one target variable: four linear components $\overline{U}_1$, $\overline{U}_2$, $\overline{R}$, and $\overline{S}$; and four nonlinear components U_1', U_2', R', and S'. More precisely, total components U_1, U_2, R, and S are calculated by total MMI decomposition and nonlinear components U_1', U_2', R', and S' are calculated by nonlinear MMI decomposition. Then, linear components are calculated by subtracting the nonlinear components from their corresponding total components, e.g., $\overline{U}_1 = U_1 - U_1'$.

2.4 Data and data preprocessing

Daily mean fields are calculated based on UTC dates from MERRA-2 hourly output (GMAO 2015) spanning 1986-2015. We explore the information shared by land surface energy and wetness conditions with both latent heat flux and sensible heat flux. Intuitively, total land energy change (net radiation) and soil moisture are used as the sources and the surface heat fluxes are targets.

The climatological seasonal cycle of the coupling strength quantified by MMI is calculated independently at each ice-free land grid cell of MERRA-2. Variability with frequencies lower than 1/365 days is removed by a high-pass filter. Then, for each variable, time series for each calendar month of the 30-year period are constructed. For instance, a catenated time series for June is produced by connecting each June 30th of a year with June 1st of the next year. Then, we use Tukey's fences to deal with the outliers. The values larger than the upper boundary $Q_{max} = Q_3 + 1.5(Q_3 - Q_1)$ are set as Q_{max}; values smaller than the lower boundary $Q_{min} = Q_1 - 1.5(Q_3 - Q_1)$ are set as Q_{min}. Q_1 and Q_3 are the first and third quartile respectively. Then, each timeseries is normalized into the range [0, 1]; an example is shown in Figure 1a. This time series is used to calculate the total MMI (Fig 1b). A linear model $\hat{X}_{tar} = b + \sum_i a_i X_{s_i}$ is then fitted to the time series to calculate the residual of the target X_{tar}' and quantile normalization is applied on X_{tar}' based on the quantile of the target X_{tar} (Fig 1c) to make the total entropy of X_{tar}' and X_{tar} equivalent, ensuring the comparability between total MMI and nonlinear MMI. Finally, the decomposition of MMI is calculated (Fig 1d). Note that because the MMI is calculated separately for each calendar month, subtracting the monthly climatology from the data does not affect the distribution (i.e., filtering out the seasonal cycle is not necessary).

The unit of MMI is bits, representing the amount of information transmitted from the sources to the target. To make the result more interpretable, obtained MMI values are normalized by the Shannon entropy of the target $H(X_{tar})$. Therefore, the amount of normalized MMI (nMMI) can be interpreted as the fraction of uncertainty of the target that is explained by the source.

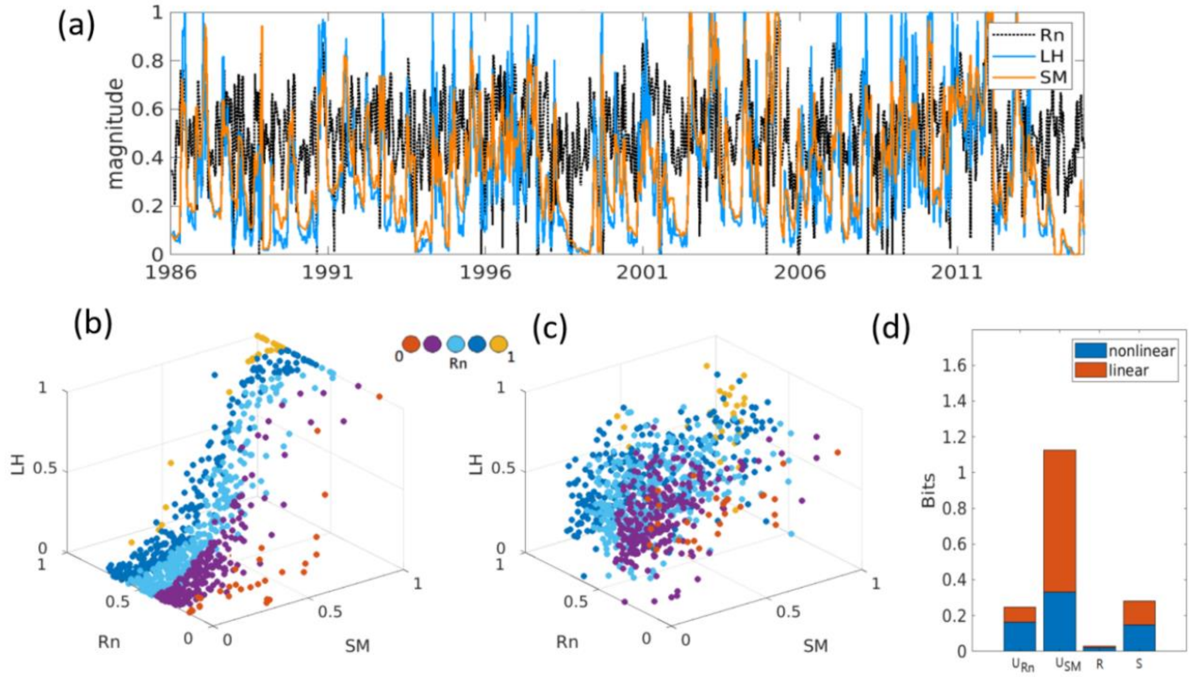

Figure 1. An example of the process of estimating multivariate mutual information (MMI) at a point (10°E,15°N) during June. The source variables are net radiation (Rn) and soil moisture (SM), and latent heat flux (LH) is the target variable. (a) The 30-year June timeseries produced by catenating the 30 years daily data of each June. (b) A 3-d scatter plot of the Rn, SM, and LH used to calculate the total MMI. (c) A 3-d scatter plot of the Rn, SM, and the nonlinear residual of LH determined by subtracting the linear model, used to calculate the nonlinearity in total MMI. (d) The decomposition of total MMI into nonlinear and linear parts after further decomposition into unique, redundant, and synergistic components.

2.5 Significance testing

We test the statistical significance progressively for total MMI, nonlinear MMI, and linear MMI. If a total MMI quantity is found to be significant, full decomposition of both linear and nonlinear unique, redundant as well as synergistic components is computed. The term "preprocessed timeseries" means the timeseries that have undergone high-pass filtering, Tukey's fences, and normalization mentioned in section 2.4.

For the total MMI case, a shuffled surrogates method, with null hypothesis that no total dependency exists, is applied on each grid cell and each calendar month. Once we calculate the MMI from the preprocessed timeseries (observed MMI), we resample the timeseries by randomly permuting the preprocessed timeseries of each of the two sources and calculate

MMI again. By repeating the process 100 times, a probability distribution of MMI as well as its mean μ and standard deviation σ are obtained. We retain the observed MMI that is larger than $\mu+3\sigma$, the level of 99% confidence. A fully nonparametric significance threshold can be directly obtained by repeating the process ~1000 times, but it is much more computationally expensive while yielding very similar results.

An identical procedure and null hypothesis are used to calculate the significance of nonlinear MMI. The observed nonlinear MMI and the MMI computed from the shuffled surrogates method both use the target that has had its linear fit subtracted. An observed nonlinear MMI larger than $\mu+3\sigma$ means the dependence is significant at the 99% confidence level and it can be recognized that only nonlinearity occupies such significant total dependency.

For the linear component, we first find the 99% significance value of the multiple correlation coefficient ρ_c for the given three preprocessed timeseries. Then, the criterion for linear MMI, MMI_c, can be calculated by using the following equivalence between correlation and mutual information:

$$MMI_c = -\frac{1}{2}\log(1-\rho_c{}^2) \tag{Eq 13}$$

3. Results

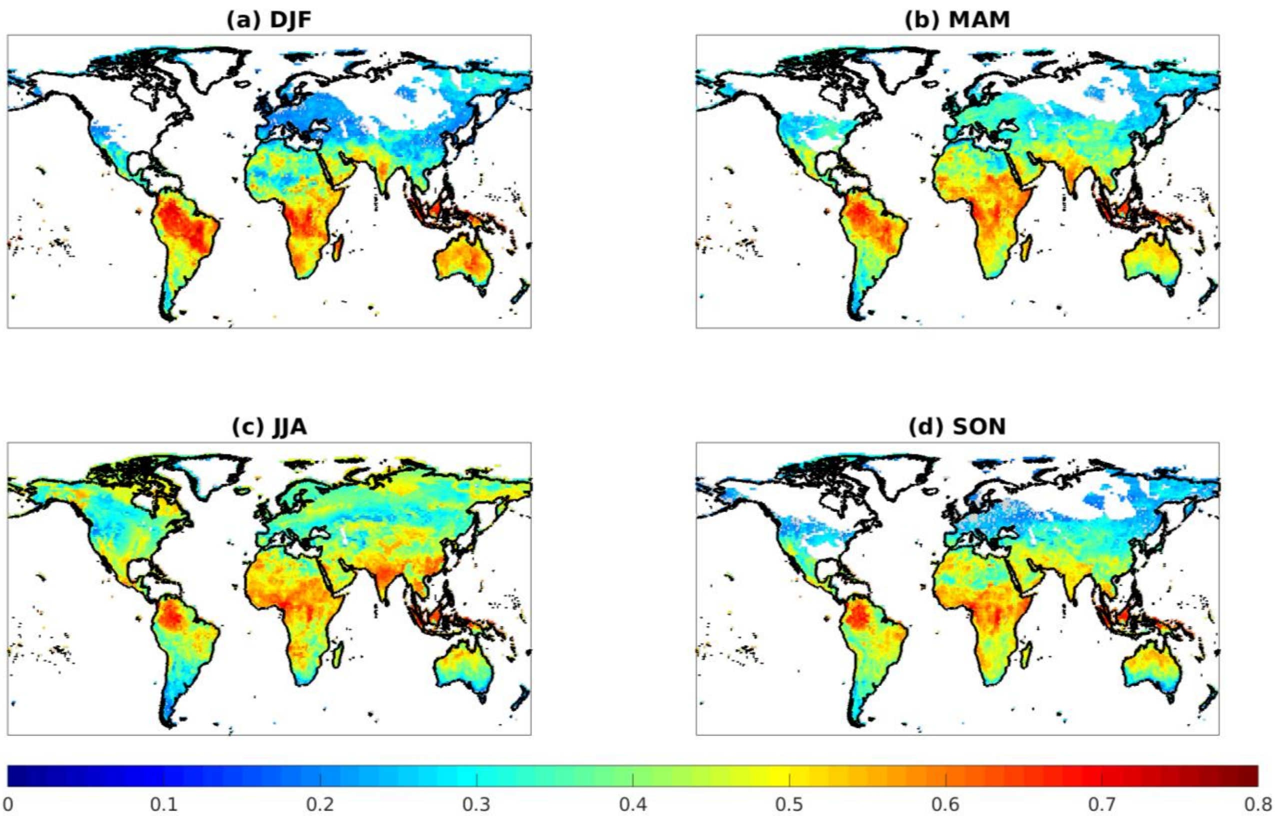

Figure 2. Seasonal total normalized multivariate mutual information (nMMI, unitless) with latent heat flux (LH) as the target, net radiation (Rn) and soil moisture (SM) as the sources. Grey shading means that not all the three analyzed months pass the significance test.

The seasonal cycle of total nMMI (each season is reported as the mean of three analyzed months), quantifying the dependency of latent heat (LH) on land surface net radiation (Rn) and soil moisture (SM), is shown in Figure 2. Regions where the surface temperature is below 0°C for more than half of the days during the analyzed season are masked out and regions where all months do not pass the significance test are shaded grey. With strong seasonality, the total dependency over most of the world is significant. This is not a surprising result since it has long been recognized that latent heat flux is controlled mainly by available moisture and energy. The strongest dependencies are found over rainforests during all seasons. Semiarid regions and the Asian monsoon area show large values during wet seasons (Fig 2c). Dependencies using sensible heat flux (SH) as the target are relatively small over the globe as shown in Figure 3. Rainforest in this case has much lower total nMMI while the large values over semiarid regions during wet seasons remain.

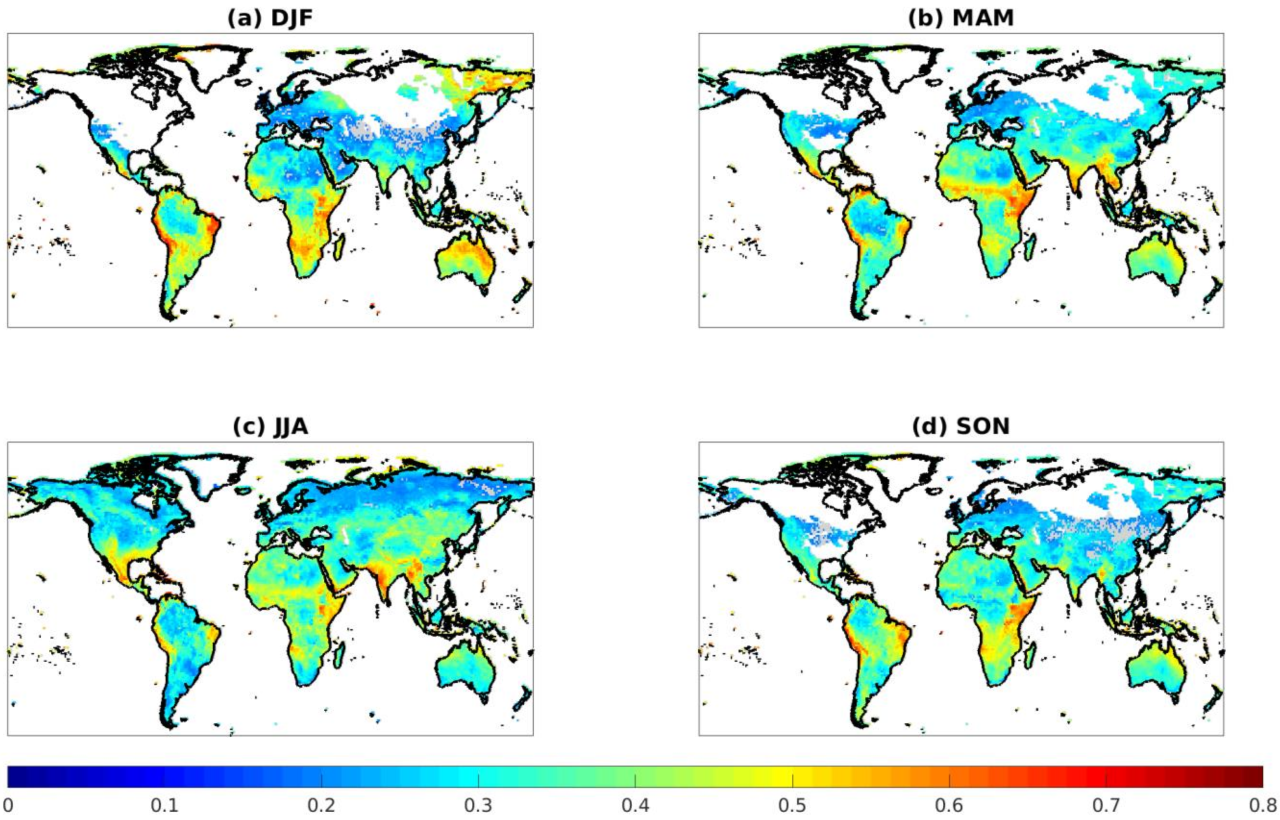

Figure 3. As in Figure 2 but using sensible heat flux (SH) as the target.

The nonlinear and linear components contributing to the total nMMI for LH and SH as the target are shown in Figure 4 and Figure 5 respectively. To make the sum of nonlinear and linear components equal to the total normalized MMI, we only screen out the regions when none of analyzed months are significant. Most areas have moderate to large nonlinearity with some particularly large values over arid regions, for example over the Sahara and the Arabian Peninsula in both seasons and over Western Australia in DJF. In those desert regions, soil moisture content is usually below the critical (wilting) point so that the wetness conditions of the land do not affect the release of latent heat flux. However, once soil moisture content rises high enough, latent heat flux becomes sensitive to the change in soil moisture. Such a transition is infrequent, but it can induce a dramatical change in the relationship between soil moisture and latent heat flux, resulting in large nonlinearity over those areas. Linearity is strong over wet and semi-arid regions (Fig 4c & d) and has a strong north-south gradient during DJF (Fig 4c). The fraction of nonlinearity shows there is generally strong linear dependency over the summer hemisphere, while nonlinear contributions are found to be more important over the winter hemisphere. However, arid and some semi-arid and subtropical

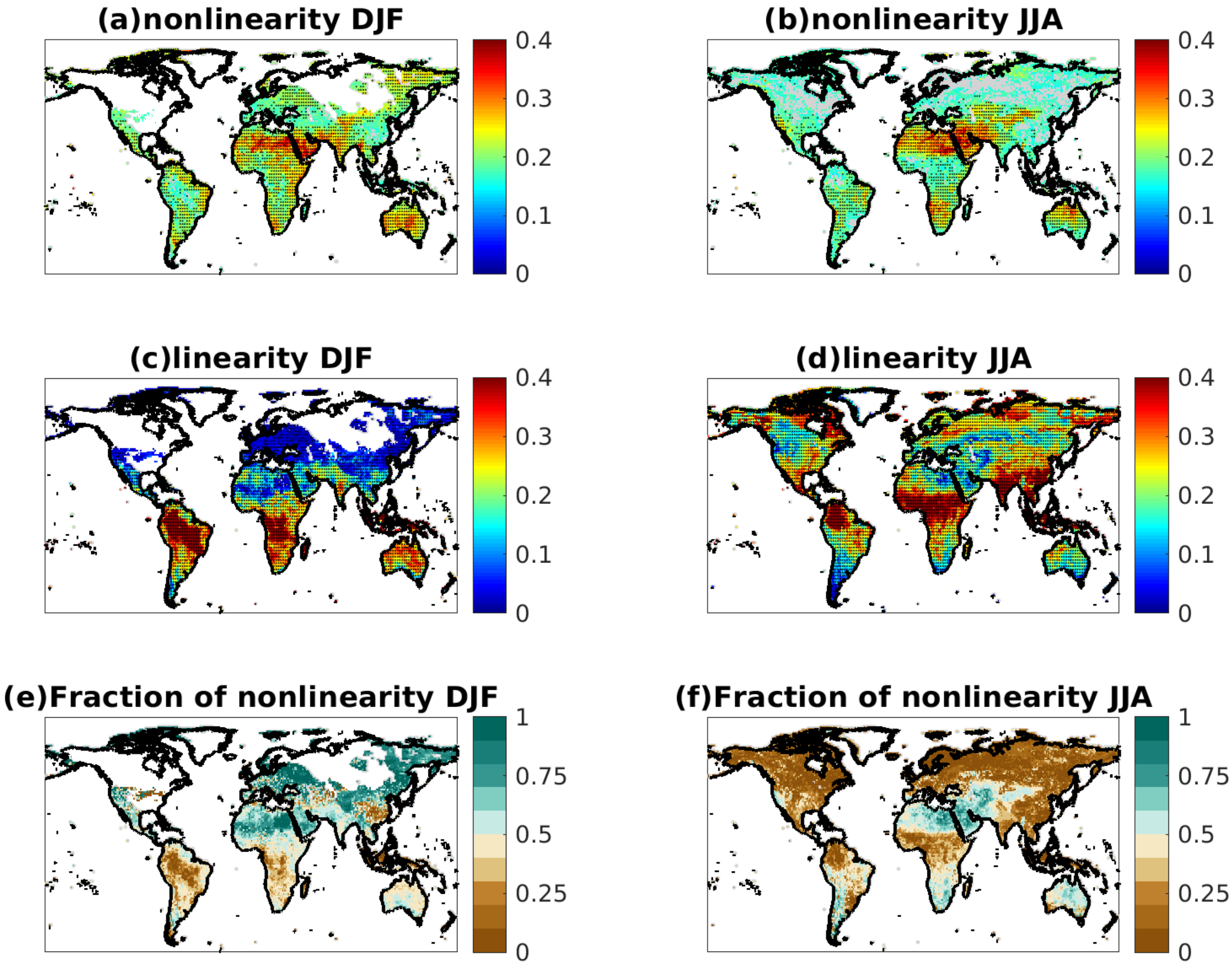

Figure 4. During DJF: (a) Nonlinear nMMI, (c) linear nMMI, and (e) the fraction of nonlinearity to total nMMI with LH as the target and Rn and SM as the sources. (b), (d), and (f) are same as (a), (c), and (e) respectively but for JJA. In panel (a) to (d), areas where none of analyzed months pass the significance test are shaded grey; areas where all analyzed months pass the significance test are stippled. All quantities are unitless.

regions show a substantial fraction of nonlinear contribution throughout the year.

Results using SH as the target (Fig 5) show a slightly different pattern. The nonlinearity is more homogeneous and subdued than in the LH case. Semiarid regions including Mexico, the Sahel, and the Indus Valley during JJA, northern and eastern Australia, the South American lowlands, and southern Africa during DJF are clearly the standouts of linear dependency. The presence of strong nonlinearity in nMMI for SH is largely absent over deserts, and is large only over parts of the tropics, namely in local dry seasons (Fig 5e & f),

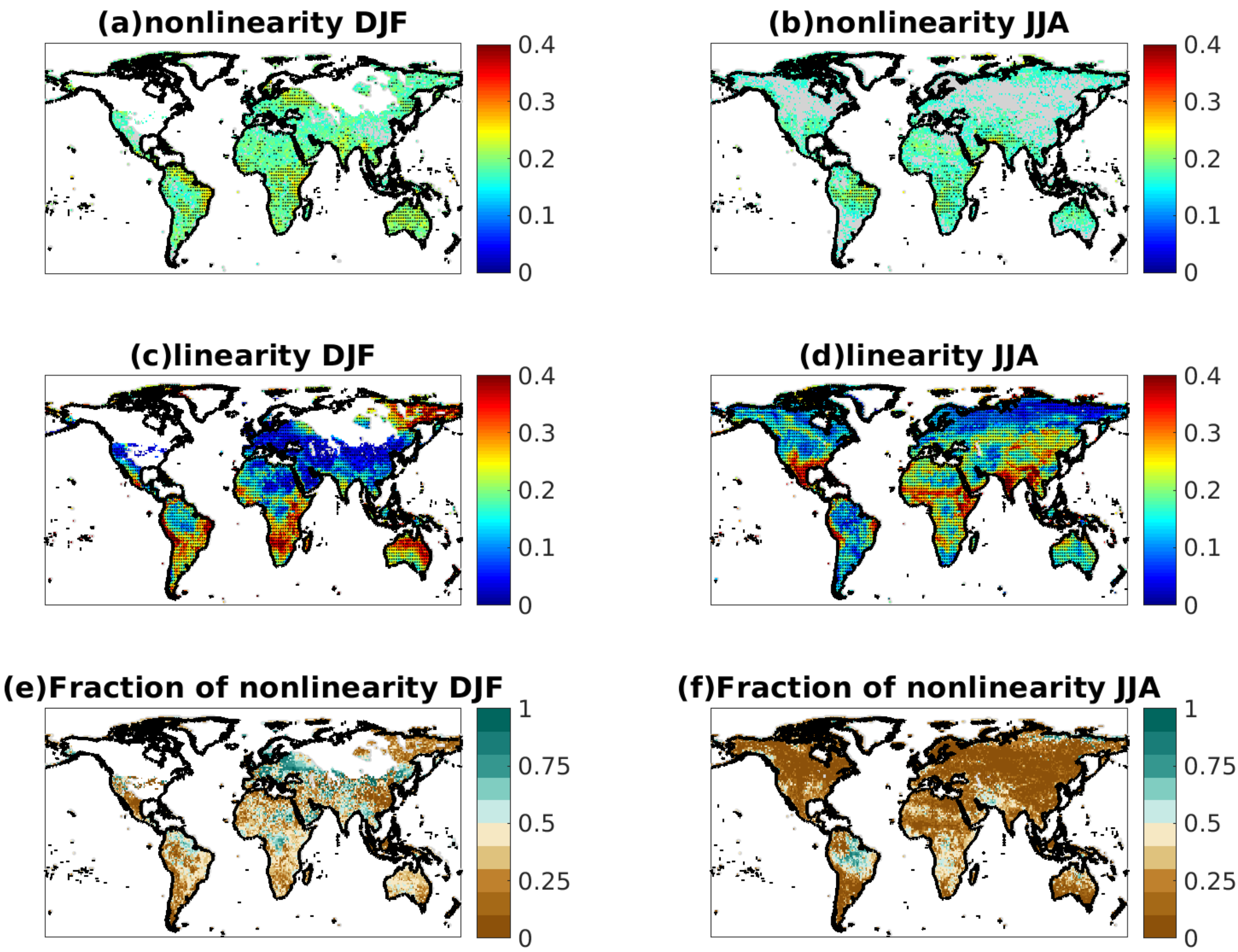

Figure 5. As in Figure 4 but using SH as the target.

and in the Northern Hemisphere mid-latitudes during DJF (Fig 5e).

We next focus on JJA to examine the composition of both the linear and nonlinear dependencies. The decomposition of total, linear, and nonlinear nMMI into unique, redundant and synergistic components with LH as the target is displayed in Figure 6. The linear unique contributions from SM alone (Figure 6b), shows a pattern similar to the canonical land-atmosphere “hot spots”, while the dry regions also depict large dependencies between LH and SM in our analysis. Ignoring the magnitude of change in the variable, which is a considered factor for quantifying coupling strength in past studies (e.g., the standard deviation term in Guo et al. 2006), might lead to this difference in the pattern. By comparing Figure 6f & j, we see many of the areas with strong linear unique contributions from SM also

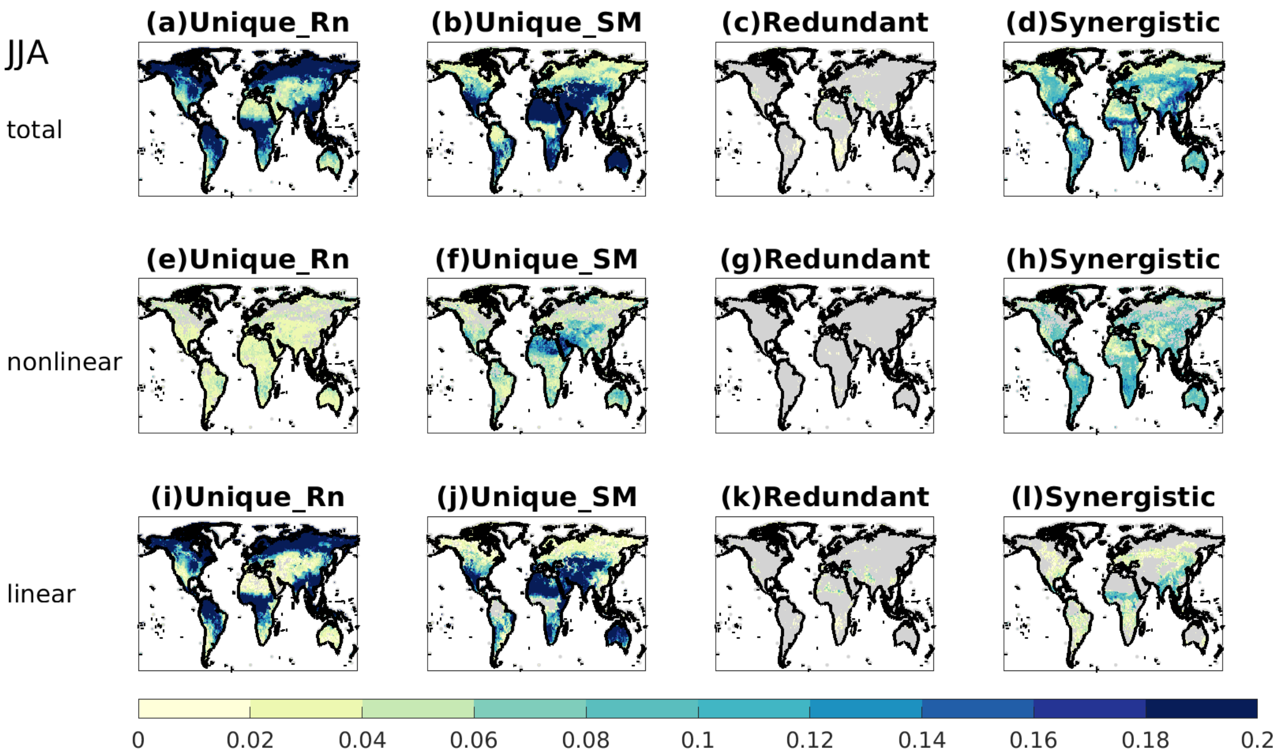

Figure 6. The partitioning of total nMMI with LH as the target, Rn and SM as the sources for boreal summer (JJA); decomposition is calculated over regions of where nMMI is significant for at least one of the three analyzed months. (a) unique information contributed from Rn, (b) unique information contributed from SM, (c) redundant and (d) synergistic components. (e), (f), (g), (h) are same as (a), (b), (c), (d) respectively but for the partitioning of the nonlinear nMMI. (i), (j), (k), (l) are same as (a), (b), (c), (d) respectively but for the partitioning of the linear nMMI. All the quantities are unitless. Regions where decomposition is not calculated or where the value is less than 0.01 are shaded gray.

have little nonlinear SM-LH dependency (Fig 6f). This can be attributed to the threshold behavior, characterized by SM values distributed around the wilting point or critical point, and/or the higher order direct relationships within the transitional zone of SM-LH relationships. We note that nonlinearity is prominent in much of the semiarid area with both of the above-mentioned features (not shown) and their quantification needs further investigation. Such non-negligible contribution of nonlinearity by SM suggests that quantifying the coupling under a linear framework, as in past studies, may underestimate the strength and somewhat misrepresent the character of coupling over the "hot spots". The case using SH as the target (Fig 7f & j) shows similar patterns as that using LH, while the strength of dependencies is weaker overall, and no strong dependency is found over dry regions.

Comparison between the linear unique contributions from Rn and SM (Fig 6i & j) suggests the two patterns are largely out of phase, accompanied by very weak linear redundancy (Fig 6k); they reveal that the two dominant regimes are controlled either solely by energy or moisture. Such a bimodal pattern is evident even when the nonlinearity is included (Fig 6a-d). This strengthens the validity of previous studies that divide the globe into energy-limited regions and soil moisture-limited regions with only considering the linear dependencies. The nonlinear contribution from Rn alone (Fig 6e) is non-zero but much weaker than that from SM (Fig 6f); no bimodal pattern is found as seen in linear part (Fig 6i & j).

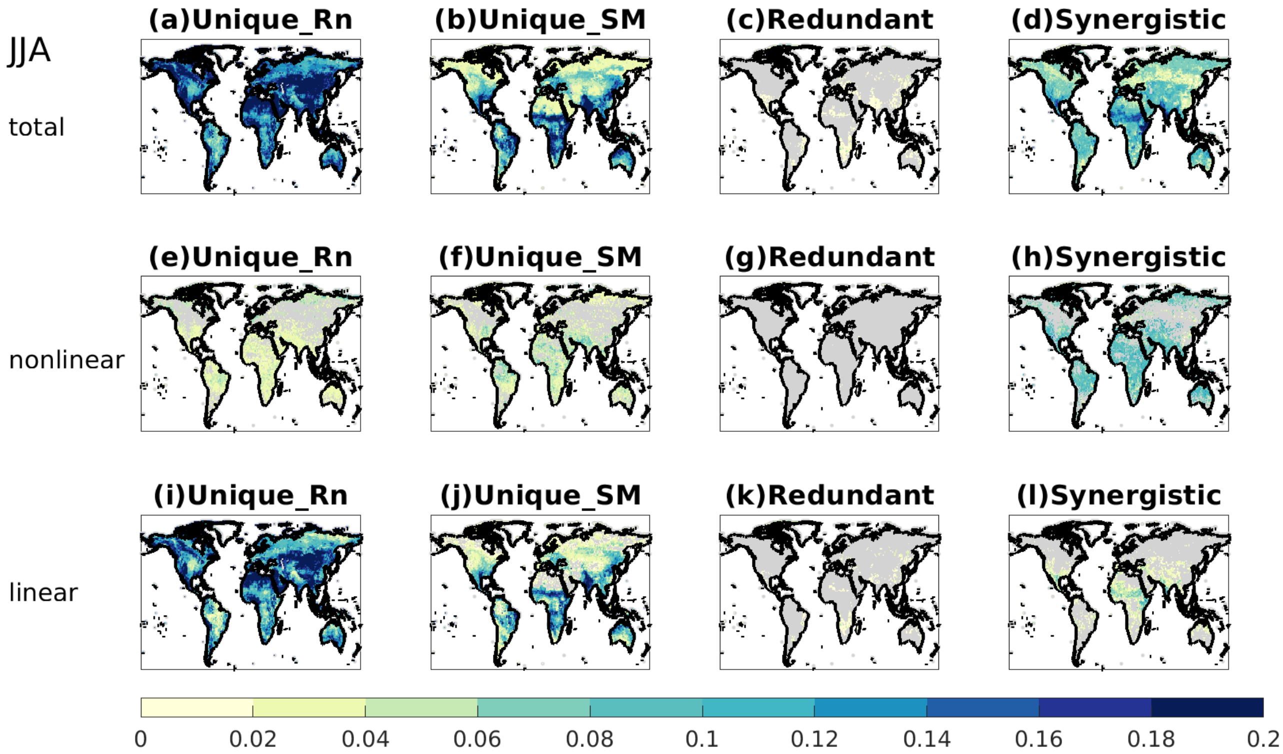

Figure 7. Same as Figure 5 but using SH as the target.

Whereas there is almost no redundancy between SM and Rn (Fig 6c), some degree of synergistic information (Fig 6d) is found over much of the world. The nonlinear synergistic information (Fig 6h) is much larger than the linear part (Fig 6i). This suggests the linear SM-LH relationship is not obviously modulated by Rn; a reason that could lead to a neglect of any multidimensional SM-Rn-LH relationship by statistical frameworks with linear dependencies. Large information for both nonlinear and linear synergistic components is found over many of the semiarid regions, e.g. the Sahel, India, and northern China. In these regions, the soil moisture content typically lies in the transition zone wherein LH is sensitive to fluctuations in SM. Together with the large synergistic information, this result suggests that both the linear and nonlinear relationships between soil moisture and heat fluxes can be modulated by Rn. Our result corroborates the findings in a recent observational station-based analysis that the relationship between soil moisture and heat fluxes is multidimensional (Haghighi et al. 2018). However, here we demonstrate that such a multidimensional concept applies globally.

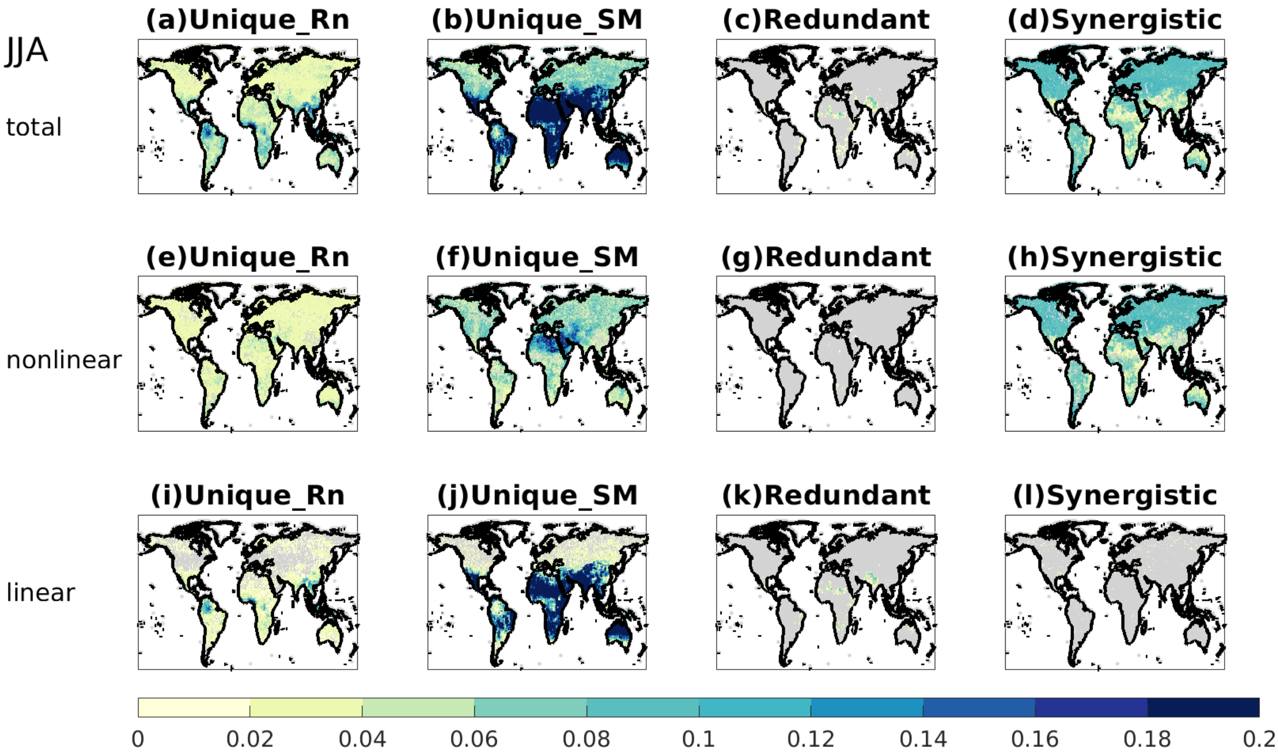

Figure 8. Same as Figure 5 but using evaporative fraction (EF) as the target.

The bimodal pattern seen between Rn and SM sources for LH is less distinct for SH (Fig 7a-b, i-j), in which the linear contribution of Rn dominates the total nMMI and is opposite to that in LH case in many moisture-limited regions, e.g. western North America, the Sahara and the Arabian Peninsula. Intuitively, in those regions, the available energy directly determines the amount of SH most of the time since there is no water to be evaporated; the available moisture determines the amount of LH only when soil moisture content is above a critical value. The nonlinear contribution from SM alone (Fig 7f) is much weaker compared to the case of LH (Fig 6f) and no particularly large value is found. This suggests that when soil moisture passes the critical point, the induced change in the SM-SH relation is not as obvious as that in the SM-LH relation. The nonlinear synergistic contribution and linear contribution of SM are comparable, again suggesting the nonlinear and multidimensional dependencies among Rn, SM and SH.

The appearance of bimodal patterns for LH and to a lesser extent SH gives rise to the question of which source dominates the partitioning of surface heat fluxes. To address this, a decomposition with the target of evaporative fraction (EF), calculated as LH divided by the sum of LH and SH, is shown in Figure 8. The decomposition of MMI reveals that SM plays a critical role in heat flux partitioning; most of the contribution is linear, while the nonlinearity is slightly larger than that in both the LH and SH cases. The total synergistic information is larger than the unique contribution from Rn, suggesting that the SM-EF relationship can be affected by Rn although the Rn-EF relationship is weaker. This highlights the importance of exploring coupling in a multivariate analysis. We note that the nonlinear contribution to the total nMMI in the EF case (not shown) is greater than the separate LH and SH cases, and could arise from the mathematical representation of the EF ratio. Individually, linear SM-LH and SM-SH relationships can still result in a nonlinear SM-EF relationship as the ratio form of EF makes it inherently sensitive to fluctuations in the denominator when it is small. In high latitudes or places with cloud cover, such situations can occur because net radiation is small or even negative, leading to values of EF well outside the nominal range 0-1. After EF is calculated from the original data of MERRA-2, we rely entirely on Tukey's fences so that the pre-processed timeseries of EF is constrained to the range [0, 1].

4. Conclusions

Addressing both nonlinear and multidimensional aspects, a technique based on information theory has been applied to reanalysis data to revisit the global estimation of land-atmosphere coupling. A quantification of multi-dependency using multivariate mutual information (MMI), has been decomposed as different interpretable components by a newly proposed integrated partitioning method. Three combinations of three variables (two sources, one target) have been explored, namely Rn-SM-LH, Rn-SM-SH, and Rn-SM-EF, of which Rn and SM are the sources and their contributions to the change in targets LH, SH, and EF have been quantified respectively.

Our analysis of total multidimensional dependency shows variability in both spatial and temporal aspects. The linear components resemble the canonical land-atmospheric coupling distributions as well as the regions known to be governed by water-limited and energy-limited regimes. The nonlinear components superposed on the linear results do not alter these familiar patterns and thus strengthen the authoritativeness of past findings, while contributing new insights. Nonlinear contributions to LH variability are predominant in arid regions and across midlatitudes and subtropical areas in the winter hemisphere. Most of the nonlinear contribution is from SM, although there are non-negligible contributions from Rn. The existence of water- and energy-dominated regimes seen for LH are also evident for SH. Whereas Rn is a major contributor to variability in LH and SH individually over much of the world, the partitioning of surface heat fluxes is confirmed to be strongly determined by SM by using EF as the target variable. The property of multidimensional dependency among land-atmospheric coupling factors is revealed to exist over the whole globe by the substantial magnitude of the synergistic term, which is greater than the redundancy term in all cases.

We have only applied this analysis to MERRA-2, which is not a perfect representation of reality since values of net radiation and surface heat fluxes are calculated from other assimilated state variables, instead of being measured directly. This leads to an inherent interdependency among the variables analyzed in this study. Further application of our analysis on data from other reanalyses, climate models and satellite data is needed to increase confidence in the global patterns shown here.

We also note that the nonlinearity and synergistic relationships suggested to exist across the globe need further investigation. For instance, it remains to be disentangled how factors including (1) potential critical points that determine the changes in sensitivity of surface heat fluxes to soil moisture, (2) higher order direct relationships between variables, and (3) natural groupings inherent in the data each contribute to the nonlinear relationships found here. Synergism, treated as the ability of a third factor to alter the bivariate relationship between two factors, is worth quantifying to improve our understanding of the interactions in nature and advance predictability in models. For example, the finding of multidimensional relationships among Rn-SM-SH implies that considering Rn as a predictor could improve forecasts of extreme events like heatwaves, since soil moisture-sensible heat flux-surface temperature feedback plays a crucial role in predicting near surface temperature.

Finally, we note that the MMI analysis can be performed with other combinations of source variables such as wind speed and near surface humidity. The comparison among different combinations of source variables may further determine when and where other variables not considered in this study are also important factors for surface heat fluxes. Applying this analysis to outputs from numerical models can help identify shortcomings in the parameterizations of land surface processes and land-atmosphere interactions. In addition, the MMI technique can be extended vertically along the water and energy cycle process chains linking land and atmosphere (Santanello et al. 2018) by using surface heat fluxes as the sources and any property/state of the planetary boundary layer, clouds or precipitation as targets. MMI is a tool that shows great promise for exploring more complex relationships in coupled land-atmosphere processes than have been possible with simple statistics.

Acknowledgements:

We are grateful to Dr. Timothy DelSole for helpful discussions during key phases of the development of this methodology. This research has been supported by grants from the National Science Foundation (AGS-1419445) and the National Oceanographic and Atmospheric Administration (NA16OAR4310095). MERRA-2 hourly output used in this study were downloaded from Global Modeling and Assimilation Office (GMAO 2015, https://disc.gsfc.nasa.gov/datacollection/M2T1NXLND_5.12.4.html).

References

Budyko, M. I. (1963), Evaporation Under Natural Conditions, Isr. Program for Sci. Transl., Off. of Techn. Serv., U.S. Dep. of Commerce, Jerusalem

Budyko, M. I. (1974), Climate and Life, xvii, 508

Dirmeyer P. A. (2011), The terrestrial segment of soil moisture-climate coupling. *Geophysical Research Letters*, *38*, L16702 doi: https://doi.org/10.1029/2011GL048268

Eltahir, E. A. B. (1998), A soil moisture–rainfall feedback mechanism: 1. Theory and observations, *Water Resour. Res.*, *34*, 765–776, doi:10.1029/97WR03499.

Global Modeling and Assimilation Office (GMAO) (2015), MERRA-2 tavg1_2d_lnd_Nx: 2d,1-Hourly,Time-Averaged,Single-Level,Assimilation,Land Surface Diagnostics V5.12.4, Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), Accessed: 27 September 2016, doi: 10.5067/RKPHT8KC1Y1T.

Goodwell, A. E. and P. Kumar (2017), Temporal Information Partitioning Networks (TIPNets): A process network approach to infer ecohydrologic shifts, *Water Resour. Res.*, *53*, 5899–5919, doi:10.1002/2016WR020218.

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

Guo, Z., et al. (2006), GLACE: The Global Land-Atmosphere Coupling Experiment. 2. Analysis. *J. Hydrometeor.*, *7*, 611-625, doi:10.1175/JHM511.1.

Haghighi, E., D. J. Short Gianotti, R. Akbar, G.D. Salvucci, D. Entekhabi (2018) Soil and atmospheric controls on the land surface energy balance: a generalized framework for distinguishing moisture-limited and energy-limited evaporation eegimes, *Water Resour. Res.*, *54*, 1831–1851, doi: https://doi.org/10.1002/2017WR021729

Herold N., J. Kala and L. V. Alexander (2016), The influence of soil moisture deficits on Australian heatwaves, *Environ. Res. Lett*, *11*, doi: http://dx.doi.org/10.1088/1748-9326/11/6/064003

Kala J., J. P. Evans and A. J. Pitman (2015), Influence of antecedent soil moisture conditions on the synoptic meteorology of the Black Saturday bushfire event in southeast Australia, *Q. J. R. Meteorol. Soc. 141*, 3118–29, doi: https://doi.org/10.1002/joc.2206

Koster, R. D. et al. (2004), Regions of strong coupling between soil moisture and precipitation, *Science*, *305*, 1138–1140, doi: 10.1126/science.1100217.

Los, L. O. et al. (2006), An observation-based estimate of the strength of rainfall-vegetation interactions in the Sahel, *Geophys. Res. Lett. 33*, L–16402, doi: 10.1029/2006GL027065.

Paninski, L (2003), Estimation of entropy and mutual information, *Neural Comp, 15(6)*, 1191-1253, doi: https://doi.org/10.1162/089976603321780272

Rienecker, M.M. , M. J. Suarez, R. Gelaro, R. Todling, J. Bacmeister, E. Liu, et al. (2011), MERRA: NASA's modern-era retrospective analysis for research and applications, *J Clim*, *24*, pp. 3624-3648, doi: 10.1175/JCLI-D-11-00015.1

Santanello, J.A., P.A. Dirmeyer, C.R. Ferguson, K.L. Findell, A.B. Tawfik, A. Berg, M. Ek, P. Gentine, B.P. Guillod, C. van Heerwaarden, J. Roundy, and V. Wulfmeyer (2018), Land–atmosphere interactions: the LoCo perspective, *Bull. Amer. Meteor. Soc.*, *99*, 1253–1272, doi: https://doi.org/10.1175/BAM/S-D-17-0001.1

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

Santanello, J.A., C.D. Peters-Lidard, A. Kennedy, and S.V. Kumar (2013), Diagnosing the nature of land–atmosphere coupling: a case study of dry/wet extremes in the U.S. Southern Great Plains, *J. Hydrometeor.*, *14*, 3–24, doi: https://doi.org/10.1175/JHM-D-12-023.1

Santanello, J. A., P. A. Dirmeyer, C. R. Ferguson, K. L. Findell, A. B. Tawfik, A. Berg, M. B. Ek, P. Gentine, B. Guillod, C. van Heerwaarden, J. Roundy, and V. Wulfmeyer (2018), Land-atmosphere interactions: The LoCo perspective. *Bull. Amer. Meteor. Soc., 99*, 1253–1272, https://doi.org/10.1175/BAMS-D-17-0001.1.

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

Smith R. (2015), A mutual information approach to calculating nonlinearity, *The ISI's Journal for the Rapid Dissemination of Statistics Research*, doi: 10.1002/sta4.96

Tao, C., Zhang, Y., Tang, S., Tang, Q.,Ma, H.-Y., Xie, S., and M. Zhang (2019), Regional moisture budget andland-atmosphere coupling over the U.S.Southern Great Plains inferred from the ARM long-term observations, *Journal of Geophysical Research: Atmospheres*, *124*, doi: https://doi.org/10.1029/2019JD030585

Thomas, M. C. and J. A. Thomas (2006), Elements of Information Theory, Wiley-Interscience; 2 edition.

Williams, P.L. and R. D. Beer (2010), Nonnegative decomposition of multivariate information, *arXiv*, arXiv:1004.2515

Yu, Y., M. Notaro, F. Wang et al. (2017) Observed positive vegetation-rainfall feedbacks in the Sahel dominated by a moisture recycling mechanism. *Nat Commun, 8,* 1873, doi: https://doi.org/10.1038/s41467-017-02021-1