Characterizing soil moisture regimes and linear and nonlinear soil moisture-latent heat flux dependency

Hsin Hsu[1] and Paul Dirmeyer[1]

[1]George Mason University

August 30, 2023

Abstract

The control of latent heat flux (LE) by soil moisture (SM) variations is a key process affecting the moisture and energy balance at the land-atmosphere interface. SM-LE coupling is conventionally examined by identifying SM-LE relationships with metrics involving correlation. However, such a traditional approach, which fits a straight line across the full SM-LE space to evaluate the dependency, leaves out certain critical information: nonlinear SM-LE relationships and the long-recognized thresholds that lead to dramatically different behavior in different soil moisture regimes. This study examines three aspects of the SM-LE relationship to diagnose coupling globally: linear dependencies, nonlinear dependencies, and SM-LE threshold behavior. Using data from climate models, reanalyses, and observational-constrained datasets, global patterns of SM-LE regimes are determined by segmented regression. Mutual information analysis is applied only for days when SM is in the transitional regime between critical points defining high sensitivity in the SM-LE dependency. Sensitivity is further decomposed into linear and nonlinear components. Our results show discrepancies in the global pattern of existing SM regimes, but general consistencies among the linear and nonlinear components of SM-LE coupling. This implies that although models simulate different surface hydroclimates, the inherent behavior of how LE interacts with SM is well-described. The pattern of strong SM-LE coupling in the transition regime resembles the conventional distribution of "hot spots" of land-atmosphere interactions. This indicates that only the transitional SM range is necessary to determine the strength of coupling. This framework can be applied to investigate extremes and the shifting surface hydroclimatology in a warming climate.

Characterizing soil moisture regimes and linear and nonlinear soil moisture-latent heat flux dependency

Hsin Hsu[1] (hhsu@gmu.edu) 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

Abstract

The control of latent heat flux (LE) by soil moisture (SM) variations is a key process affecting the moisture and energy balance at the land-atmosphere interface. SM-LE coupling is conventionally examined by identifying SM-LE relationships with metrics involving correlation. However, such a traditional approach, which fits a straight line across the full SM-LE space to evaluate the dependency, leaves out certain critical information: nonlinear SM-LE relationships and the long-recognized thresholds that lead to dramatically different behavior in different soil moisture regimes. This study examines three aspects of the SM-LE relationship to diagnose coupling globally: linear dependencies, nonlinear dependencies, and SM-LE threshold behavior. Using data from climate models, reanalyses, and observational-constrained datasets, global patterns of SM-LE regimes are determined by segmented regression. Mutual information analysis is applied only for days when SM is in the transitional regime between critical points defining high sensitivity in the SM-LE dependency. Sensitivity is further decomposed into linear and nonlinear components. Our results show discrepancies in the global pattern of existing SM regimes, but general consistencies among the linear and nonlinear components of SM-LE coupling. This implies that although models simulate different surface hydroclimates, the inherent behavior of how LE interacts with SM is well-described. The pattern of strong SM-LE coupling in the transition regime resembles the conventional distribution of "hot spots" of land-atmosphere interactions. This indicates that only the transitional SM range is necessary to determine the strength of coupling. This framework can be applied to investigate extremes and the shifting surface hydroclimatology in a warming climate.

Significance Statement

Evaporation is sensitive to soil moisture only within a specific range that is neither too dry nor too wet. This transitional regime is examined to quantify how strongly soil moisture controls local humidification of the air. We identify dry, transitional, and wet regimes across the globe; spatial patterns in climate models and observationally-based datasets often show discrepancies. We determine dependencies between soil moisture and evaporation only for the transitional regime. We find general consistency of locations having simple linear dependencies and more complex nonlinear relationships. We conclude that although surface hydroclimates are different between climate models and observations, how soil moisture controls evaporation is well simulated by models. These results have potential application for improved forecasting and climate change assessment.

1. Introduction

Coupling between soil moisture and evapotranspiration modulates the moisture and energy balances at the land surface (Bonan 2008b; Seneviratne et al. 2010; Santanello et al. 2018). As water escapes from land to atmosphere, it can induce a chain of consequences (Findell and Eltahir 1997; Eltahir 1998). A decline in soil moisture content can cause moisture stress that increases the resistance to evaporation, and secondarily can result in a higher surface albedo that reduces absorbed incoming solar radiation. Surface latent heat flux decreases while surface sensible heat flux increases under conditions with lower moisture and energy availability. This change in the partitioning of surface heat fluxes results in a warmer and drier atmosphere that is further conditioned against the formation of clouds and limits precipitation. Without a source of water going into land surface, it dries further. Such a soil moisture-evapotranspiration-precipitation feedback has a significant impact on the sub-seasonal time scale as it links to extremes such as heat waves and droughts, which can cause huge economic and societal damage (Fischer et al. 2007; Hirschi et al. 2011; Herold et al. 2016; Miralles et al. 2019). Therefore, the role of soil moisture in the current climate and future climate is topic of growing importance.

The existence of such a feedback at any location mainly depends on whether the variations in latent heat flux are controlled by soil moisture content or energy availability (Dirmeyer et al. 2009). As has long been recognized, the relationship between soil moisture and latent heat flux (hereafter abbreviated SM-LE) is not linear. It exhibits threshold behavior; LE behaves dramatically differently when SM crosses critical values (Budyko 1963, 1974). Values typically defined are the permanent wilting point, field capacity, and a critical SM threshold above which the SM-LE relationship weakens or even reverses (Seneviratne et al. 2010). Permanent wilting point is the minimum required SM such that vegetation does not wilt; a criterion of hydraulic pressure such that osmosis, the process that allows roots absorb water, can happen and water can be transmitted up the vascular system of the plants. Field capacity is the amount of water retained in the soil that can be maintained as a balance between gravity and capillary forces.

There is a critical SM threshold that lies below field capacity and depends on available energy. If SM exceeds the critical threshold, further increases in SM do not increase evapotranspiration; rather, available energy determines LE. Thus, the critical

SM threshold depends on multiple climate variables. It has been found that LE increases with increasing SM content when water is between the states of permanent wilting point and the critical threshold. This is called the *transitional regime* for SM. In other words, only when SM lies within, or shifts into, the transitional regime is the land-atmosphere feedback maintained/triggered and SM-LE coupling becomes important. The variations in LE become uncoupled from SM when it is below the wilting point (dry regime) or exceeds the critical SM threshold (wet regime).

Several studies have examined when and where the SM variability-induced feedback exists by using climate models, reanalyses, and observations. Metrics for such identification have been built upon testing the null hypothesis that there is no statistically-significant dependency between SM and other meteorological variables, such as surface heat fluxes (Dirmeyer 2011; Hsu and Dirmeyer 2021), air temperature (Seneviratne et al. 2006; Miralles et al. 2012; Gevaert et al. 2018), and precipitation (Koster et al. 2004; Santanello et al. 2013; Guillod et al. 2015; Hsu et al. 2017; Tao et al. 2019). The feedback is found to be strongest over semi-arid regions, the so-called land-atmosphere coupling "hot spots" (Koster et al. 2004), where SM frequently lies in the transitional regime and varies strongly. Although multiple studies have agreed with the locations of strong SM-LE coupling (e.g. Koster et al. 2006; Zhang et al. 2008; Dirmeyer 2011; Diro et al. 2014; Hirschi et al. 2014; Liu et al. 2014; Lorenz et al. 2015; Hsu and Dirmeyer 2021), the cause-and-effect link of energy and water availability to evapotranspiration has not been fully explored.

Although the existence of a limited sensitive range for SM-LE coupling has been recognized for a long time (e.g., Koster and Milly 1997), the coupling metrics in most studies customarily consider the full range of SM without attempting to quantify how the distribution of the three SM regimes relates to the coupling. This may dilute the aim of detecting locations where SM dominates the LE in two ways. First, during extremely wet and energy-deficient conditions, variations in available energy govern the release of LE. Consequently, LE can decrease while SM increases as a result of reduced net radiation in very wet, cloudy conditions. Such a situation may also be identified via correlations as SM-LE coupled (albeit a negative relationship) even though energy availability is the cause and SM changes are a consequence of evaporation rates.

Second, LH is positively correlated with SM only when SM is in the transitional regime, but is one location found to be more strongly coupled than another simply

because it spends more days within the transition regime, or rather because the sensitivity of LE to variations in SM is actually stronger? Could some overlooked regions be strongly, but infrequently, coupled due to high SM-LE sensitivities in a transitional SM regime that is not routinely experienced? In other words, will locations of strong land-atmosphere coupling corresponding to the canonical "hot spots" remain if the analysis only considers those days when SM values lie within transitional regime?

Taking into account the changeable dependency between SM and surface heat fluxes, Schwingshackl et al. (2017) statistically estimated the critical SM values that separate SM into dry, transitional, and wet regimes with a piecewise-linear regression analysis. The sensitivity of near-surface air temperature to SM variations has also been estimated for transitional regimes to obtain a clearer picture of how SM affects the lower troposphere and episodes of extreme heat. In regional studies, Dirmeyer et al. (2021) and Benson and Dirmeyer (2021) have found that the sensitivity of extremes in near surface air temperature to declining SM is amplified when SM declines below a quantifiable threshold corresponding to the local wilting point. Such a hypersensitive regime arises because a strengthened positive feedback is triggered by a chain of processes linking drier SM, depleted LE, increased sensible heat flux, and increased atmospheric temperature. Denissen et al. (2020) used satellite data to estimate the distribution of the critical SM threshold as a crossover point that signifies whether evapotranspiration is controlled by the availability of water or energy.

This study aims to identify SM-LE coupling across the globe while explicitly considering the points of SM regime shifts. Coupling strength is quantified by mutual information that measures total dependency between variables, which can be further decomposed as linear and nonlinear components. Using observationally-based data sets, reanalyses, and climate models, we first determine the spatial distribution of the various SM regimes including the determination of areas that routinely cross thresholds and inhabit multiple regimes. We then quantify the total, linear, and nonlinear dependencies of LE to SM within the transitional regime using daily fields of SM and LE. This two-step method focuses on the dependency within the transitional regime where the bulk of sensitivity resides, instead of including all available days in the analysis. This helps to determine whether SM conditions lingering in the transitional regime are all that is necessary to ensure a strong coupling. In addition, this filters out nonlinearity in the SM-LE relationship contributed by

merely crossing the threshold. Details of the data sets, the method for critical value detection, and mutual information are described in section 2. Section 3 presents the results and conclusions are presented in section 4.

2. Methods

2.1. Data

This study targets data sets that provide daily global fields of both surface SM and LE. Multi-year daily fields are used from climate models, reanalyses, and observationally-based data sets to maximize statistical robustness. Table 1 summaries the data sets used in this study. Multiple climate models participating in the Coupled Model Intercomparison Project Phase 6 (CMIP6) are included in this study. We select the models that have a historical run in which both daily SM and LE fields were available online (https://esgf-node.llnl.gov/search/cmip6/) at the time of our analysis. 30 years of output spanning 1986-2015 from a total of 11 models has been used.

CMIP6 Models (1986-2015)					
Model	Abbreviation	exp_id	DOI	Resolution	Surface soil layer
CESM2	CESM2	r11i1p1f1	10.22033/ESGF/CMIP6.2185	0.9375° lat x1.25° lon	top 10 cm
CanESM5	CanESM5	r1i1p1f1	10.22033/ESGF/CMIP6.1303	2.8125° lat x 2.8125° lon	
HadGEM3-GCM1-MM	HadGEM3		10.22033/ESGF/CMIP6.420	0.56° lat x 0.83° lon	
INM-CM5-H	INM		10.22033/ESGF/CMIP6.1423	0.5° lat x 0.66° lon	
EC-Earth-Veg	EC-Earth		10.22033/ESGF/CMIP6.642	0.7° lat x 0.7° lon	
GFDL-CM4	GFDL		10.22033/ESGF/CMIP6.1402	1° lat x 1.25° lon	
IPSL-CM6A-LR	IPSL		10.22033/ESGF/CMIP6.5195	1.25° lat x 2.5° lon	
MIROC6	MIROC6		10.22033/ESGF/CMIP6.9121	1.4° lat x 1.4° lon	
SAM0-UNICON	SAM0		10.22033/ESGF/CMIP6.7791	0.9375° lat x 2.5° lon	
MRI-ESM2-0	MRI		10.22033/ESGF/CMIP6.621	1.125° lat x 1.125° lon	
AWI-ESM-1-1-LR-g3	AWI		10.22033/ESGF/CMIP6.359	1.875° lat x 1.875° lon	
Reanalyses (1986-2015)					
MERRA2	MERRA2		10.5067/RKPHT8KC1Y1T	0.5° lat x 0.625° lon	top 5 cm
ERA5	ERA5		10.1002/qj.3803	31 km x 31 km	top 7 cm
Satellite-based Observations (Apr 2015-Oct 2020)					
SMAP Level 4	SMAP		10.5067/B59DT1D5UMB4	9 km x 9 km	top 5 cm

Table 1. Sources of gridded global SM and LE data used in this study.

Reanalysis data comes from two sources. First, the NASA Modern-Era Retrospective Analysis for Research and Applications-2 (MERRA2) provides hourly fields of variables at a resolution of 0.5° latitude x 0.625° longitude (GMAO 2015).

In MERRA-2, a major feature is that the land is forced by gridded observed precipitation, instead of model-generated precipitation. This strategy makes observed precipitation the main driver of SM, of which time series have been shown to have better agreement with independent observational datasets (Reichle et al. 2017a). Second, the ECMWF Reanalysis v5 (ERA5) provides hourly grids of a variety of variables at a resolution of ~31km (Hersbach et al. 2020). Daily fields covering the period 1986-2015 are calculated from both reanalyses by averaging in time relative to UTC. ERA5 assimilates satellite SM data while MERRA2 does not, although this difference is not a focus of this study.

To provide a perspective from observationally-based products, the NASA Soil Moisture Active Passive (SMAP) mission (Entekhabi et al. 2010) Level-4 Soil Moisture (L.4_SM) is used. SMAP L4 (Reichle et al. 2017) assimilates the SMAP observations into the same land surface model used in MERRA2, but completely uncoupled from the atmospheric model; it is also forced by observed precipitation. Although only a relatively short time period is available, SMAP L4 provides complete global coverage of SM and LE in space and time, and thus is ideal for this analysis. In contrast, the Level-3 product of SMAP, available at the same spatial resolution and composed only of quality-controlled observations of SM from orbit, provides coverage over only about one third of the global land surface per day and lacks LE estimates. SMAP L4 is assumed here to be the closest to representing real conditions among these data sets. However, considering the short period of data as well as the involvement of a land model in the data assimilation process, SMAP L4 is rather like another reanalysis product and serves as an additional "ensemble member" when we present composite results.

We note that the variable called "surface soil moisture" used in this study represents different thicknesses of soil layer wetness in different data sets. In CMIP6, surface soil moisture represents wetness in the top 10cm of the model soil column whereas it represents the top 5cm for MERRA2, 7cm for ERA5, and 5cm in SMAP. Theoretically, this induces a deviation in the SM as well as its variability in time, and thus may affect slightly the magnitude of the critical SM values between products. Nevertheless, the resulting translation of SM-LE patterns does not greatly affect the detection of SM regimes. All analysis is done at the native resolution of each data set except for SMAP. In an attempt to compensate for the smaller sample size attributed to the shorter period in SMAP, analysis is done on a scaled up grid (18km x18km)

consisting of 2x2 grid cells (9km x 9km), assuming that the heterogeneity of land cover type and topography at this larger domain is not sufficient to induce a stark change in how LE behaves with SM variations. Trading space for time, this simulates a quadrupling of the time series length to ~22 years.

2.2. Critical value detection

We employ the approach proposed by Schwingshackl et al. (2017) to define the wilting point and critical SM thresholds. Theoretically, as wilting point (WP) and critical soil moisture (CSM) separate the SM into dry, transitional, and wet regimes, a total of five candidate segmented regressions can result depending on which regimes are detected in the data. Four levels of complexity are tested to find the best fit at each location. We name these five candidates using three-digit binary numbers consisting of 0 or 1 to indicate the absence or presence of (in order) the dry, transitional and/or wet regime. These possible segmented regression candidates are illustrated in Fig 1a and are listed below:

- No dependency on SM: indicating SM either never reaches values above the WP or never falls below the CSM (candidates 100 or 001). Practically, locations identified with no SM-LE dependency are found mostly over energy-limited regions such as rainforests, high latitudes, and alpine locations where soils are almost always wet or frozen. Rare cases of no SM-LE dependency can be found at coastal regions dominated by maritime air where SM nevertheless spans a wide range of values. Thus, we treat all cases without SM-LE dependency as candidate 001.
- One-segment regression consisting solely of a segment with positive slope: indicating SM always lies in the transitional regime. (Candidate 010)
- Two-segment regression consisting of a constant segment followed by a segment with positive slope: indicating SM spans the WP. (Candidate 110)
- Two-segment regression consisting of a segment with positive slope followed by a constant segment: indicating SM spans the CSM. (Candidate 011)
- Three segment regression consisting of a segment with positive slope between two constant segments: indicating SM spans all regimes. (Candidate 111)

Since a more complex statistical model will normally have a better performance, we use Bayesian information criterion (BIC) for fitted candidate selection (Schwarz

1978) to introduce a penalty term by considering the number of the parameters in the model to avoid overfitting:

$$BIC = n\ln(RSS/n) + k\ln(n) \quad \text{Eq1}$$

where n is the sample size, RSS is the residual sum of squares, and k is the number of model parameters. The model with the lowest BIC is selected. However, based on an empirical criterion (Kass and Raftery 1995), if the difference of BIC between a model and its one level simpler model is less than 10, the simpler model will be selected.

Figure 1b displays an example in which the candidate 111 has been selected as the best fit at a MERRA2 grid cell in the tropics (10°N, 20°E). In this case, both WP and CSM are detected, indicating SM crosses all SM regimes, although kernel density indicates that many days lie near the wilting point. The color scheme in panel (a) is used hereafter to indicate the five candidates for SM-LE relationships.

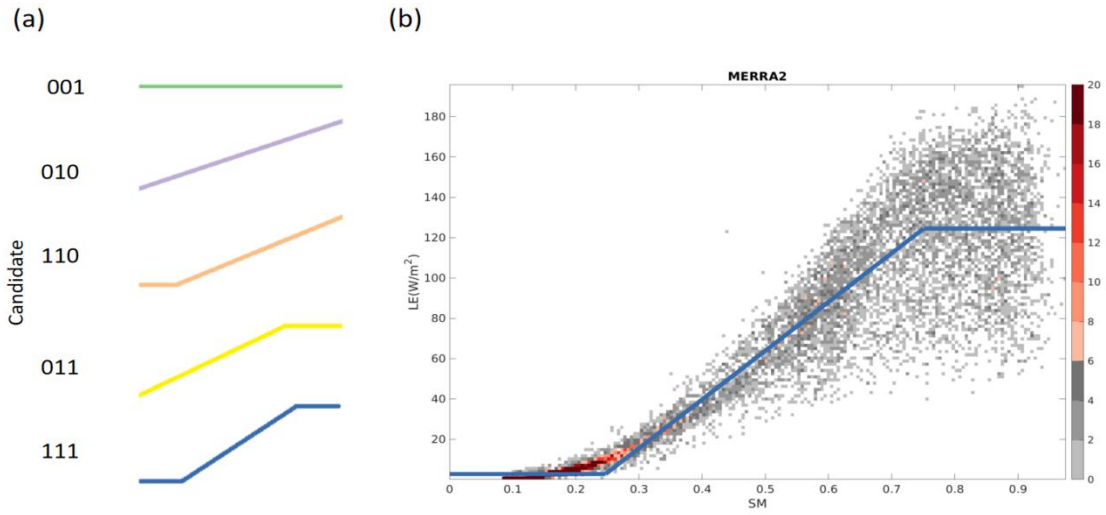

Figure 1. (a) Five segmented regression candidates used to fit the data. (b) An example displays that the MERRA2 SM-LE data at (10°N, 20°E) is best fitted by the candidate 111. Color in the scatter plot shows the density of data in each interval spanning 0.01 SM value (unitless) and 1 W/m^2 LE value.

2. 3. Normalized mutual information

The dependency between SM and LE in the transitional regime is quantified by normalized mutual information (NMI):

$$NMI(SM;LE) = \frac{I(SM;LE)}{H(LE)} \quad \text{Eq2.1}$$

$$H(X) = -\sum p(x)\log_2 p(x) \quad \text{Eq2.2}$$

$$I(X;Y) = \sum p(x,y)\log_2\left(\frac{p(x,\mathrm{y})}{p(x)\mathrm{p(y)}}\right) \quad \text{Eq2.3}$$

H is the Shannon Entropy (Shannon 1948) which uses the probability distribution function $p(x)$ of a single random variable X to measure X's uncertainty (Eq2.2). In this study, a logarithm with base 2 is used and yields results in the unit of bits. The probability distribution of daily values of variables is examined and is cut by intervals with fixed bin width so that summation of $p(x)$ across a finite number of bins is executable. Mutual information *I(X;Y)* (Cover and Thomas 1991) measures the reduction in uncertainty of a random variable X by the knowledge of another random variable Y (Eq2.3). As a result, *NMI(SM;LE)* is interpreted as the fraction of variability of the LE that is explained by variations in SM (Eq2.1).

$$\underline{I(X;Y)} = I(X;Y) - I(X;Y)' \quad \text{Eq3.1}$$

$$I(X;Y)' = I(X;Y') \quad \text{Eq3.2}$$

$$Y' = Y - \hat{Y} \quad \text{Eq3.3}$$

$$\hat{Y} = b + \sum_i a_i X_i \quad \text{Eq3.4}$$

$$NMI = \underline{NMI} + NMI' = \frac{I(X;Y)}{H(Y)} + \frac{I(X;Y)'}{H(Y)} \quad \text{Eq3.5}$$

Mutual information *I(X;Y)* can be decomposed as linear information *I(X;Y)* and nonlinear information *I(X;Y)′* (Smith 2015). In Eq3.1, *I(X;Y)* is quantified as the difference between *I(X;Y)* and *I(X;Y)′*; their linkage with NMI is shown in Eq3.5. Nonlinear information *I(X;Y)′* is obtained by calculating mutual information between X and Y' (Eq3.2). Y' is a nonlinear residual term calculated by the following procedure: a linear regression model (Eq3.4) is fitted to the time series to calculate the residual of the Y by Eq3.3. Then, quantile normalization is applied to Y' based on the quantile of the Y. This ensures the equivalence of total entropy of Y and Y' so that *I(X;Y)* and *I(X;Y)′* are comparable. More detail of the procedure can be found in Smith (2015).

We note that the fitted line with positive slope used in the linear segmented regression to determine WT and CSM may be different from the fitted line used to decompose the mutual information into linear and nonlinear components for the transitional regime. The nonlinearity of SM-LE dependency resulting from such a two-step method ensures that the nonlinear SM-LE relationship can be contributed by

properties other than the threshold behavior. This helps to classify whether the nonlinear SM-LE relationship found in our previous work is mostly attributable to the transitions of sensitivity of LE to SM induced by WT and CSM.

2.4. Workflow

All analysis is done for each grid cell in each data set except for SMAP, which has an effective grid cell domain consisting of 2x2 actual grid cells. Total values of all daily timesteps of SM and LE are used to detect soil moisture regimes by breakpoint analysis. The wilting point (WP) or critical soil moisture (CSM) thresholds (or both), if detected, are recorded simultaneously for further dependency analysis.

After critical point detection, we remove variability in the original timeseries having frequencies lower than 1/365 days using a high-pass filter. Then, concatenated time series grouped by calendar month spanning the whole period are constructed. For instance, daily data for each June is connected from each June 30th of one year to June 1st of the next year to produce the multi-year June timeseries. Discontinuities do not negatively affect the NMI calculations. Such a month-by-month analyses enables us to avoid artificial dependencies which would be introduced if the data distribution were modified by other approaches for removing seasonality.

Daily time steps of SM and LE when SM is below WP or above CSM are removed from the timeseries. Thus, only data in the SM transitional regime is used for dependency calculations. Since the retained data size is different between different locations and products within any calendar month, subsampling is used to provide a bootstrap estimation of uncertainty. When the sample size of a grid cell for the specific month is larger than 500 days (around half of the original time series), we randomly subsample 300 days from the constructed time series. Fixed binning with 10^2 bins is used to compute the two-dimensional probability distribution functions constructed from SM and LE and to obtain *NMI* and its decomposition. This is repeated 100 times and averages of *NMI*, $\underline{NMI}$, and *NMI′*, named as *mNMI*, $m\underline{NMI}$, and *mNMI′* respectively, are obtained. We note that though the choice for the number of bins affects the magnitude of mutual information and the ratio of information partitioned into nonlinearity, it does so systematically such that it does not affect the general spatial patterns of *mNMI*, $m\underline{NMI}$, and *mNMI′* so long as the same number of bins are used everywhere. In thus study, we focus on the comparison among the products instead of comparison between the information components within a single

product. Thus, sensitivity of information content to the number of bins does not affect the interpretation of our results.

2.5. Significance testing

Statistical significance is tested in different ways for total, linear, and nonlinear mutual information on each grid cell and each calendar month. For total mutual information, a shuffled surrogates method with the null hypothesis that no total dependency exists is applied on *mNMI*. Once a value of *mNMI* is obtained by the workflow described in section 2.4, daily values of SM and LE selected only from days when SM in dry and wet regimes are removed are randomly permuted by breaking original SM-LE pairs and then randomly matching them. Each permutation of the timeseries yields a new estimate of *mNMI*. By repeating the process 30 times, a probability distribution of randomized values of mNMI as well as its mean μ and standard deviation σ are obtained. Observed mNMI is 99% statistically significant if it is larger than μ+3σ.

For nonlinearity, the identical procedure and null hypothesis are used to obtain the significance of *mNMI′*. The actual *mNMI′* is compared to the distribution of *mNMI* computed from the shuffled surrogates method with *Y′* timeseries (which is LE here; the linear fit is subtracted after each permutation). An observed *mNMI′* larger than μ+3σ indicates a statistically significant dependence at the 99% confidence level and such a dependence only by the nonlinear relationships.

Statistical significance of the linear component is tested for the bootstrap mean of $\underline{I(X;Y)}$. A criterion value of the correlation coefficient ρ_c representing 99% significance is found for given pairs of random variables. With such a Gaussian distributed assumption, the criterion for the bootstrap mean of $\underline{I(X;Y)}$, $\underline{MIC}$, is calculated by Eq4.

$$\underline{MIC} = -\frac{1}{2} log(1 - \rho_c^{\,2}) \qquad \text{Eq4}$$

The workflow and significant testing are almost identical to that described in Hsu and Dirmeyer (2021) where more discussion on the seasonality issue, bin size sensitivity, and validation of the significant testing can be found.

3. Results

3. 1. Global soil moisture regimes

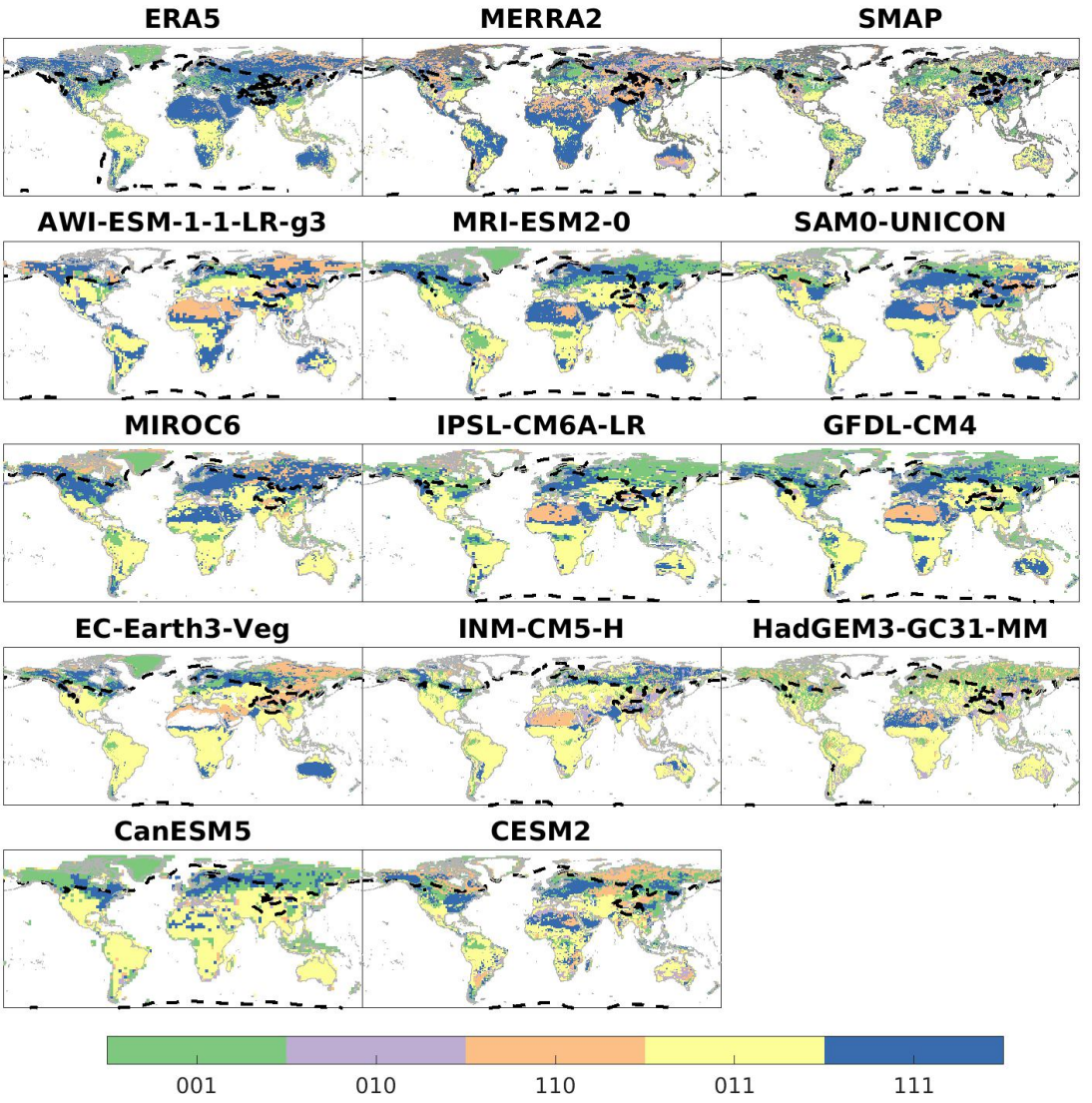

Figure 2. The most likely SM:LE candidate at each grid cell. Dashed black line represents the boundary where the climatological annual mean 2-meter air temperature in that dataset is 0˚C. Color coding is as in Figure 1(a).

Figure 2 shows the elected candidate that best fits the SM-LE daily data for each grid cell in each dataset. Candidates 001 and 011 occupy the tropics across all products, as also indicated by the mode of candidate in the equatorial regions (Figure 3a). The prevalence of candidate 011 outside the deep tropics reveals that SM is not always above the CSM but can fall into the SM transitional regime. Such a characteristic varies among the products as only around 50% of the products agree that the elected candidate is the same as the mode (Figure 3b) and the discrepancy is particularly large over the maritime continent. To quantify the discrepancy of detected

SM regimes, we have devised an index δ to quantify the degree of disagreement. δ is calculated by Eq5:

$$\delta = |a-x| + |b-y| + |c-z| \quad \text{Eq5}$$
$$\delta_{dry} = |a-x|$$
$$\delta_{tran} = |b-y|$$
$$\delta_{wet} = |c-z|$$

a, *b*, and *c* are the left bit, center bit and right bit respectively of the candidate value for any product; *x, y,* and *z* represent same digits as *a*, *b*, and *c* but for the mode of candidates among all products.

Besides measuring the uniformity of the selected candidate as that in Figure 3b, δ considers the inherent differences between the regimes included in each candidate. For example, the δ between candidates 111 and 001 is 2, which is larger than that between candidates 111 and 011; a selected candidate 111 which represents SM spanning full regimes is closer to candidate 011 than candidate 001, since candidate 001 means only wet regime is detected. The summation of δ across all products measures the degree of consensus, and is shown in Figure 3c. Figures 3d, 3e, and 3f are the degree of consensus among the products in detection of dry, transitional, and wet regime, respectively. Compared to the rest of the world, the summation of δ is moderate over the tropical regions As shown in Figure 3d,e,f, consensus is greatest for dry and wet regimes, but whether SM lies in the transitional regime is often in dispute among the analyzed products.

Semiarid regions are dominated by candidates 011 and 111 (Figure 2 and 3a). Among the CMIP6 models, there is a prominently different width of territory occupied by candidate 111 in the Sahel, whereas in the reanalyses, candidate 111 dominates over North Africa. This implies a different degree of distinction between dry and wet seasons, which might be attributed to the character of the simulated West African monsoon within the climate models. In the SMAP product, candidate 111 has a relatively narrow band in the Sahel. In other monsoon regions, although candidate 111 is detected in the reanalyses (e.g. Mexico, India, and Australia), most CMIP6 models show a lack of a dry regime leading frequently to candidate 011 (Figure 3a). Semiarid regions located in temperate zones have larger discrepancies of the elected candidate. In addition to the North American Great Plains and regions with a humid continental climate, agreement barely reaches 50% over South Africa, Europe, sections of South America and southern Australia. In these regions, even though the

transitional regime is typically detected without disagreement (Figure 3a&e), whether SM also routinely lies in either the dry or wet regime is disputed even between the reanalyses.

Arid regions, counterintuitively, are dominated by candidates 110 and 111 (Figures 2 and 3a). This may indicate that the land surface in deserts can occasionally be moistened beyond the WP or even, it appears, the CSM. Whether SM normally lies in the wet regime is disputed (Figure 3f) not only among the climate models but between the reanalyses; the full SM regime (111) is generally found in ERA5 over the Sahara, Arabian Peninsula, Australian outback, and Chile but only fragmentarily in MERRA2 (Figure 2). For SMAP, candidate 111 is found sporadically over those regions. CanESM5 and MIROC6 stand out among CMIP6 models as many arid regions are occupied by candidate 011, indicating an unrealistically wet climate in the desert.

In high latitudes and alpine regions, candidate 001 encompasses much of the area, whereas strong disagreement can be seen among the products (Figures 2, 3a and 3b). In those regions, as indicated by the 0˚C dashed contour in Figure 2, surface soils may be frozen much of year and thus the coupling between SM and LE is interrupted. Consequently, one should keep in mind that a lack of sensitivity between SM and LE does not necessarily mean that SM is always in the wet regime. Furthermore, the low available energy from solar radiation at high latitudes can lead to a low value of CSM and thus whether a wet regime is detected can also be sensitive to radiative conditions that themselves can be affected by albedo, land cover type, and a variety of soil/biological parameters. Over Siberia, most climate models show diametrically opposed results to the reanalyses and SMAP: SM in the CMIP6 models can frequently lie in the wet regime, but not in reanalyses or SMAP.

Overall, the hydroclimate reflected by SM-LE behavior is shown to have fairly low consensus among the products. Large differences among detected SM regimes between products is seen over arid regions. Over desert, whether SM can frequently cross the CSM seems disputable. A deeper investigation of the precipitation frequency, the physics of surface soil water retention and drydown in the dry climate regions is needed to resolve the discrepancy. Strong divergences are also found over alpine regions and areas with a subarctic climate, where solid-liquid phase change of SM might play a crucial role. Meanwhile, good agreement is reached within semiarid regions implying that, as observed in the real world, climate models simulate strong

seasonal variation of wetness. The spatial patterns are much more heterogeneous in SMAP and the reanalyses than in the climate models. In addition to resolution differences, this may be attributed by a more complex surface conditions in the real world affecting the assimilated observational data stream that are not fully represented in the parameterizations within climate models, or by the effects of the data assimilation process, which introduces extra variability when adding observational increments of SM and/or near surface atmospheric conditions that can impact the SM-LE relationship.

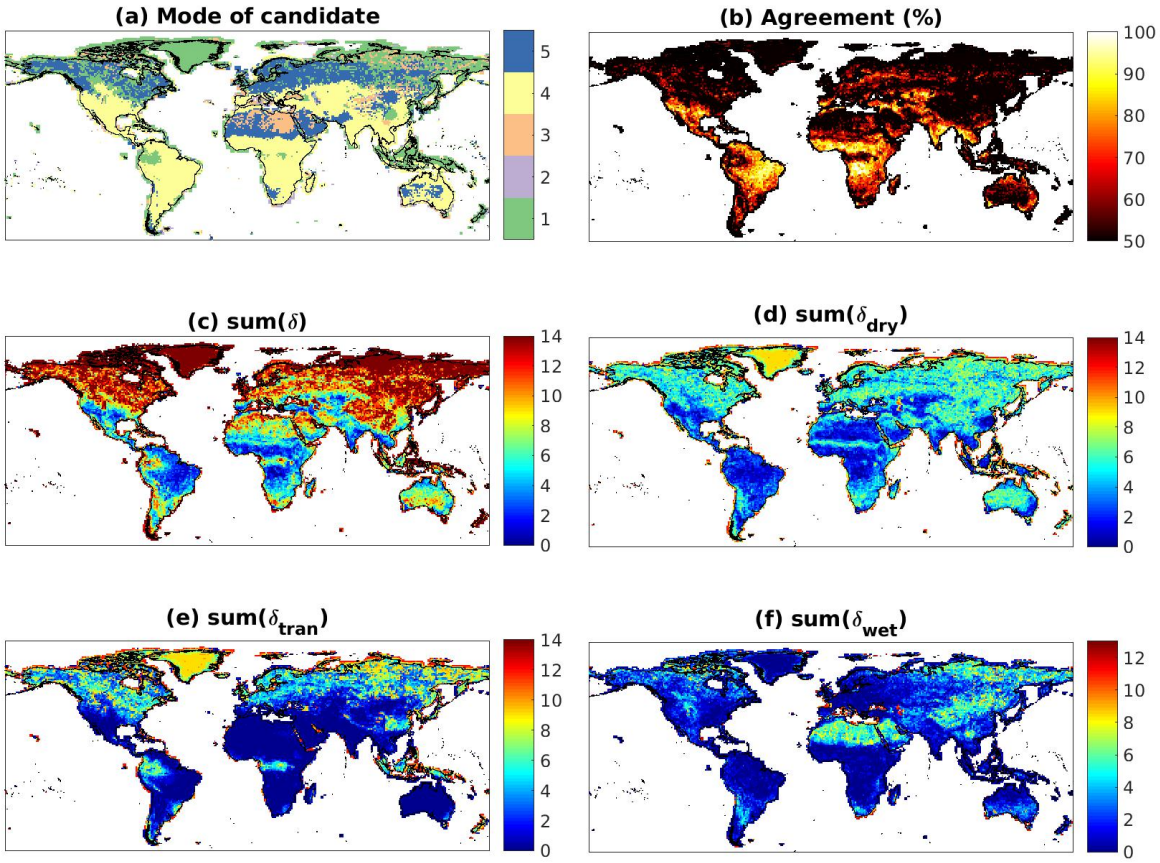

Figure 3. (a) Mode of the candidate among 14 analyzed products. (b) Agreement of the elected candidate calculated as the fraction of the products that vote for the same candidate as the mode. (c) Consensus of soil regime quantified by summation of degree of disagreement δ calculated by eq2. (d),(e),(f) are same as (c) but for δ_{dry}, δ_{tran}, δ_{wet}, respectively

Figure 4 shows the probability density functions (PDF) of SM over regions dominated by each candidate (colored curves) and the combined climatological SM (white curve) over the world for each data product (the grid cells with climatological 2-m temperature <0°C are masked out). Different patterns of climatological SM PDF

are seen among products and even the two reanalyses do not agree on whether SM is bimodally distributed. Such discrepancies reveal an essential difference of model-simulated moisture fields among different land surface models, as reported by Koster et al. (2009), still exists in current climate models. A relatively moisture-limited world is portrayed by SMAP as well as AWI-ESM and IPSL.

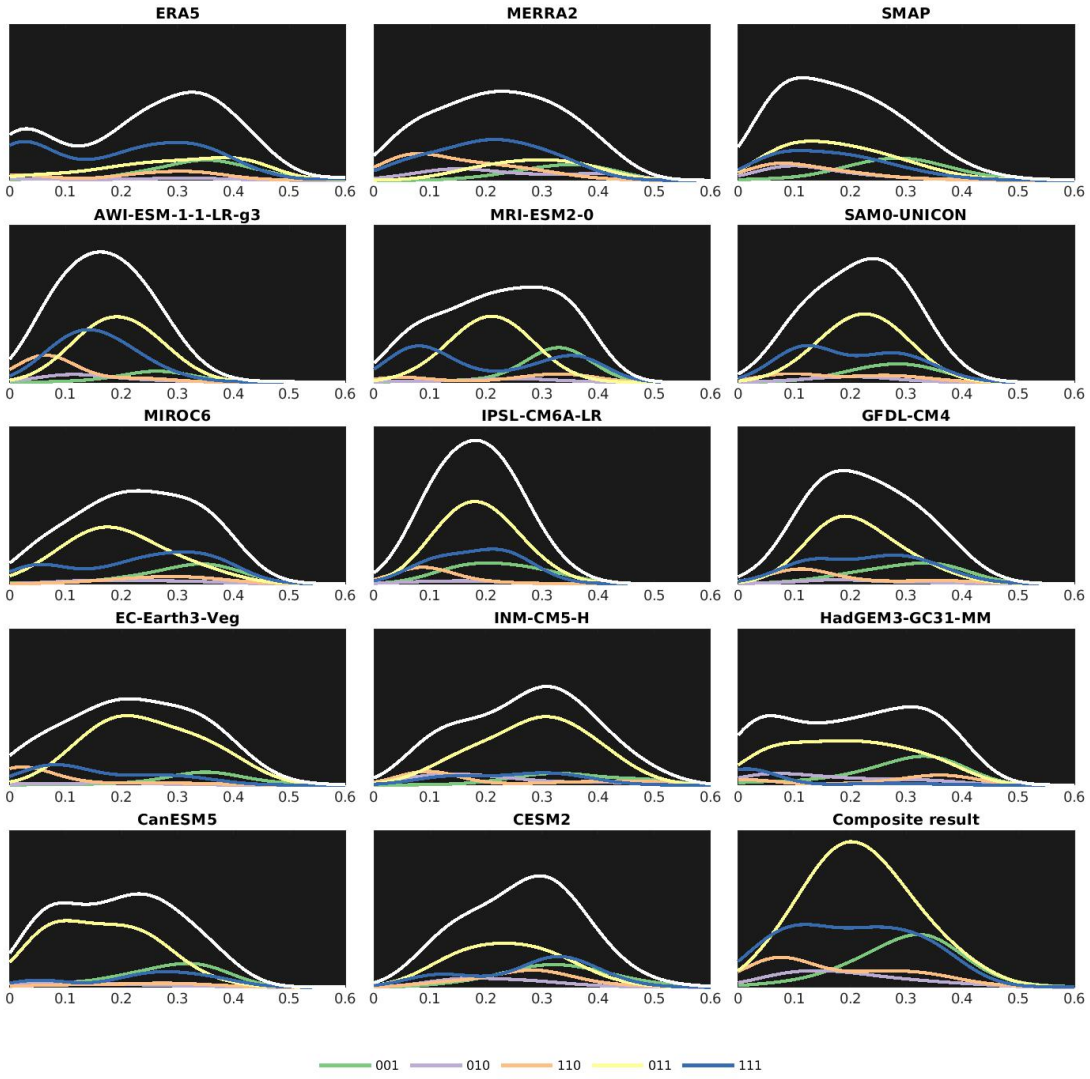

Figure 4. Probability distribution function of SM (volumetric water content, m^3/m^3) over the locations governed by each candidate (colored lines) and global SM distribution (white line). Regions in where climatological 2-m temperature < 0°C are excluded in the analysis. All pdfs are estimated by kernel density estimation with fixed bin (0.01 m^3/m^3 intervals).

As to the SM regime, in general, candidate 001 is found mostly in wet soil conditions and candidate 110 is more common over drier regions. Candidate 111 is found across a wide range of climatological wetness conditions. Strong discrepancies are seen between the climate models and the reanalyses in the area under the curve of

each candidate. In most climate models, candidate 011 accounts for the highest proportion of land area. Even though locations with SM spanning the full set of regimes are detected more in the reanalysis, the relative proportions of each candidate are similar. The dominance of candidate 011 seen in the climate models could be related to an unrealistic homogeneity of land surface properties. The relative rareness of dry regimes in climate models might be attributable to biases in radiation, precipitation, precipitation frequency, or the soil water retention in their land surface models.

3.2. Global SM-LE coupling

We examine the coupling strength between SM and LE only when daily SM values are in the SM transitional regime. This ensures a positive SM-LE dependency that implies that a change in LE is principally due to a change in SM. Focusing just on June-July-August (JJA), Figures 5, 6, and 7 display the total, linear, and nonlinear contributions to the mean normalized mutual information (mNMI) between SM-LE respectively.

Locations having at least 500 days when SM is within the transitional regime are found over most of the world during JJA. Though a positive SM-LE relationship exists everywhere, the total dependency of SM-LE measured by normalized mutual information exhibits strong spatial variability, as can be seen among all analyzed products (Figure 5). Several studies that have investigated coupling strength using the full range of SM showed regions with strong coupling are mostly over semi-arid regions. Logically, this has been attributed to the fact that SM usually lies within transitional regime in those locations. Other regions spend fewer days in the transitional regime, but they also appear to have a weaker SM-LE relationship even while in the transitional regime. Thus, our results still find these transition zones are strongly coupled compared to the rest of the world.

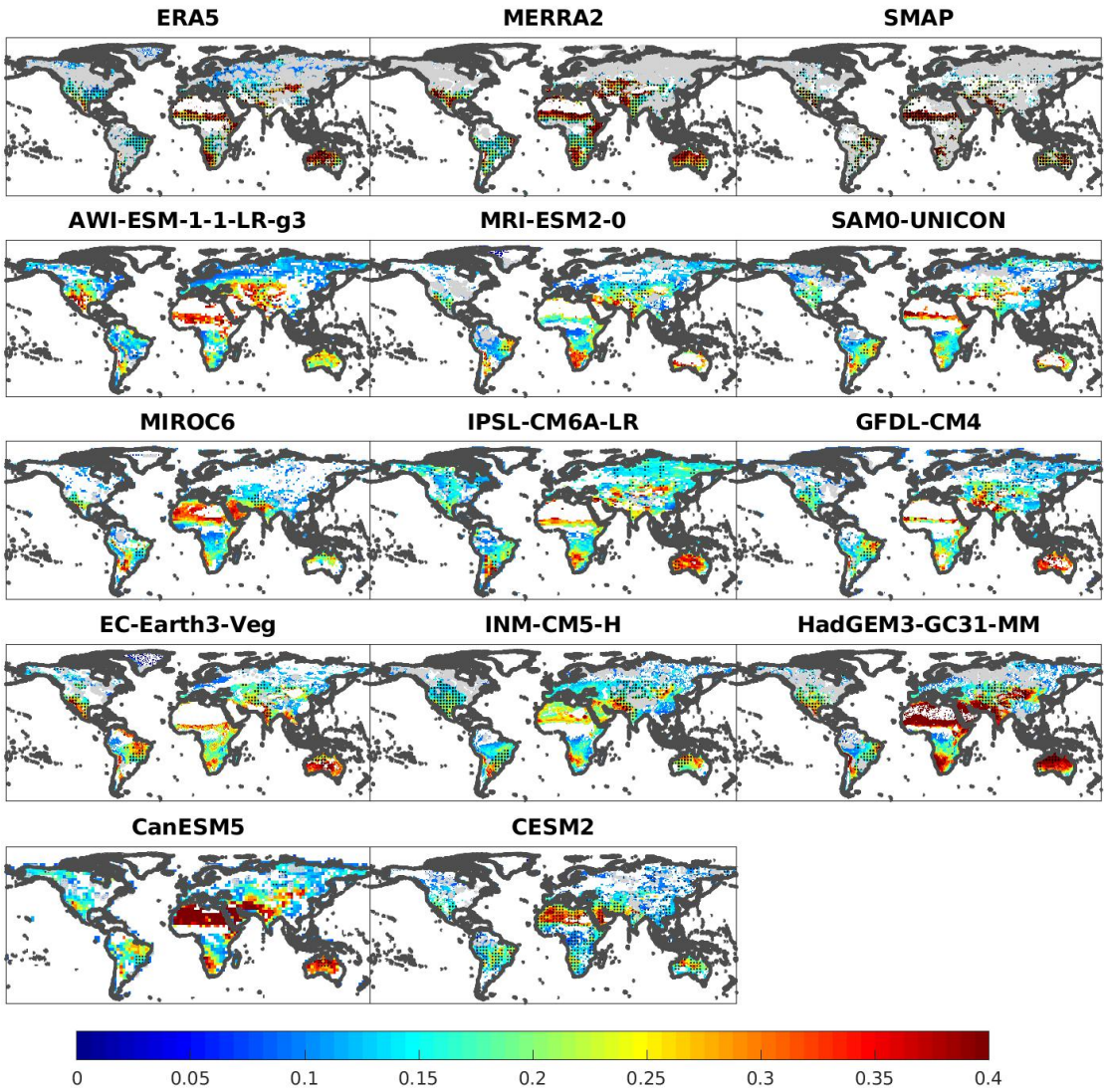

Figure 5. Average of JJA mean normalized mutual information (mNMI, units: bits) between SM and LE within the transitional regime. Locations where values of all three analyzed months are statistically significant (p-value<0.01) are dotted. Shaded grey areas are statistically insignificant for any month and white land areas are not included in the analysis due to insufficient days (<500) with SM in the transitional regime within the analyzed period. For SMAP, each sample is treated as a day; each analyzed set of 2x2 grid cells spanning 6 years yields ~720 days for applying the analysis during boreal summer months.

Regions with strong coupling are in better agreement, in terms of latitudinal distribution, among ERA5, MERRA2, and SMAP than that among the CMIP6 models. This can be seen prominently by comparing the position and width of the strongly-coupled band in the Sahel region of West Africa. In the tropics, though SM can still affect LE when SM falls into transitional regime, the dependency is relatively weak or statistically insignificant, as is particularly clear in MERRA2 and several of the CMIP6 models. On the other hand, dry areas such as the Sahara, Arabian

Peninsula, Western Australia, and Chile, when significant, have strong SM-LE coupling.

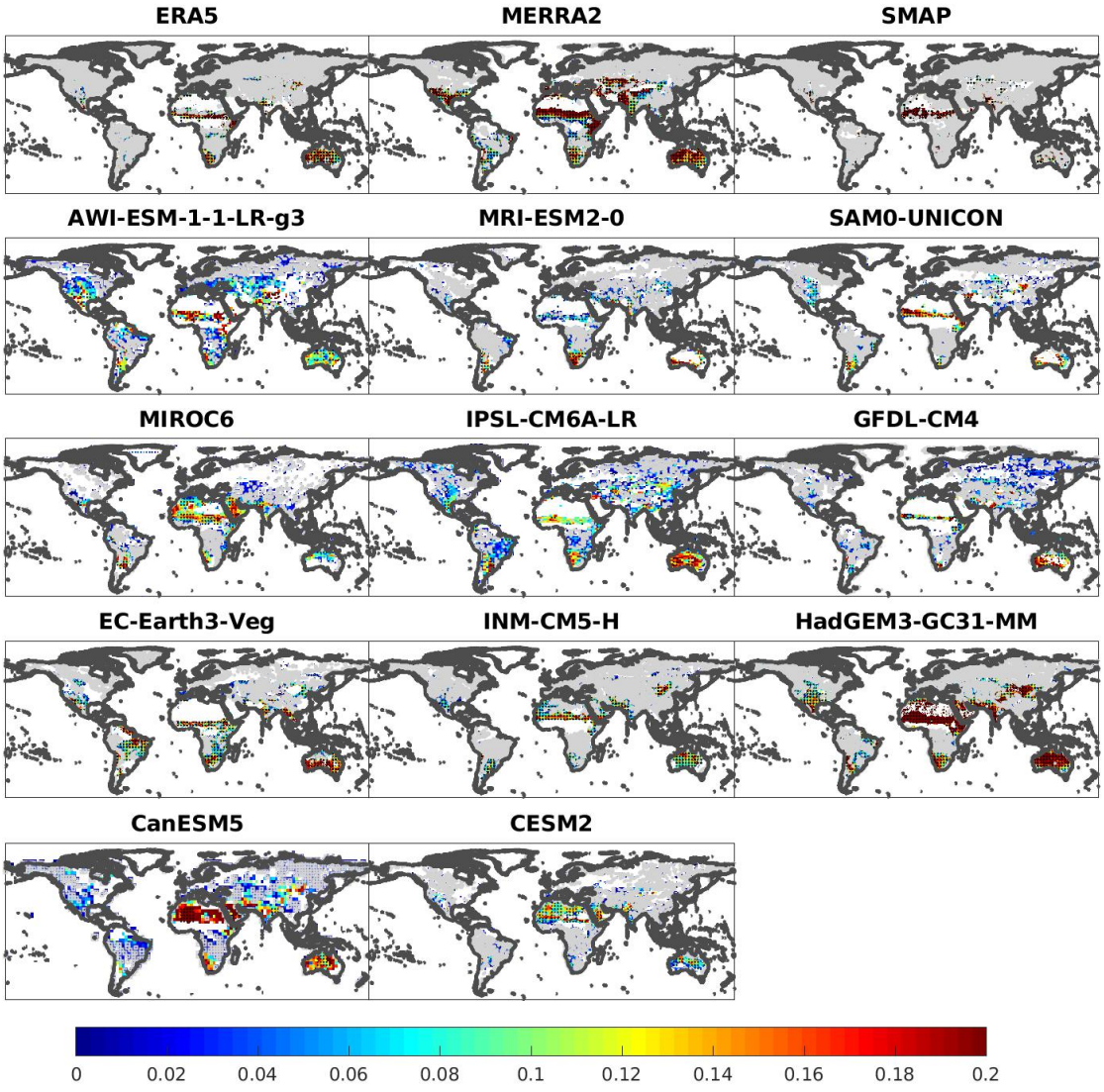

Figure 6. Same as Figure 5 but for linear mNMI.

Similar patterns are found for linear mNMI among the analyzed products (Figure 6). Strong linear SM-LE dependency is mostly found over semiarid regions, e.g., the Sahel, Southern Africa, and the Great Plains. A few climate models (CanESM5, CESM2, MIROC6) show a strong linear SM-LE coupling over arid regions (Figure 6). The pattern of strong linear coupling corresponds to the "hot spots" identified by previous studies using metrics involving a linear statistical framework (e.g., Koster et al. 2004; Dirmeyer 2011; Hsu and Dirmeyer 2021) applied on the full range of SM. Much of the world contains a sufficient number of days with

SM values in the transitional regime, and the linear dependency in most regions is statistically insignificant. Again, this reveals that the frequency of transitional SM values is a necessary condition but not the deciding factor to determine if the land-atmosphere coupling strength is strong.

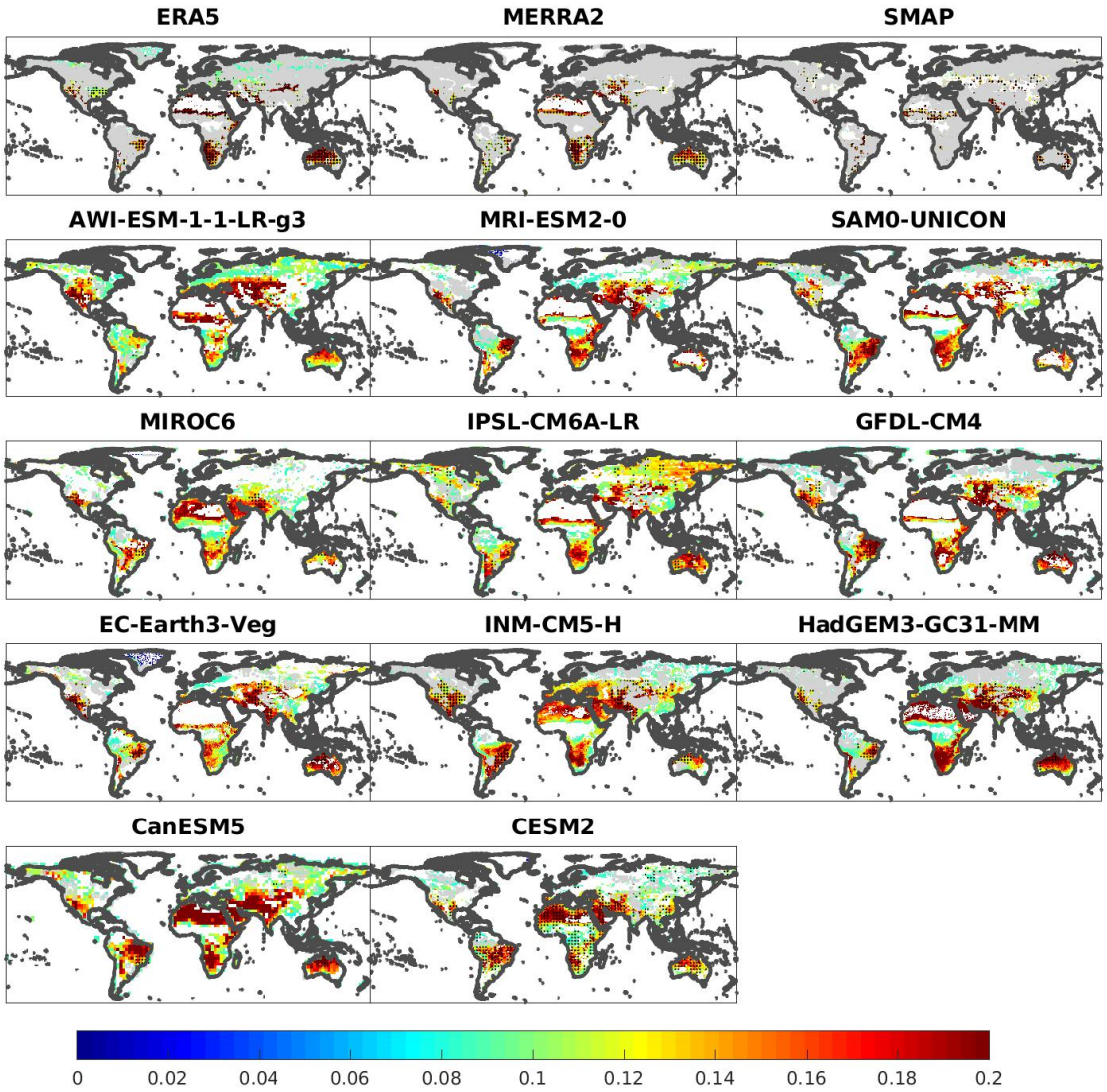

Figure 7. Same as Figure 5 but for nonlinear mNMI.

The nonlinear component shows a broadly similar but lower magnitude pattern compared to total *mNMI* in most data products (Figure 7). The nonlinear SM-LE dependency in most locations tends to be statistically significant in the CMIP6 models while only "hot spots" regions bear strong nonlinearity in the observationally-constrained data sets. The cause for this discrepancy is unclear and needs further examination. Arid regions such as North Africa emerge in the analysis for a few climate models (INM, CanESM5, CESM2, MIROC6) and all of them

suggest that nonlinearity in dry regions is strong. This could be attributed to the abundant energy and weak water retention capacity over these regions — once a precipitation event occurs, it induces a spike in SM and thus strong LE. Combined with a short drydown period, this typically leads to sporadic high values of LE and thus total mutual information is less dominated by linear dependencies.

4. Discussion and conclusions

Daily fields of surface SM and LE from climate models, reanalyses, and satellite-based products, are used to assess the coupling strength constrained within the SM transitional regime where most sensitivity exists. The transitional regime is constrained by the SM wilting point (WP) and a critical value (CSM) above which LE ceases to increase with increasing SM. Five candidate segmented regressions based on a conventional SM-heat flux conceptual framework (Schwingschackl et al. 2017) are statistically derived to determine the prominent SM regimes at each location in each data set. The method detects where changes in feedback regimes exist, indicated by the detection of WP and/or CSM, and where the positive slope of the SM-LE segment indicates the transitional regime that is critical to physical processes linking land and atmosphere via the SM-LE relationship.

We find that the robustness of the representative values of CSM is not as great as that of WP. CSM can vary with different environmental conditions as implied in the analysis of Haghighi et al. (2018). Unlike WP, CSM is not a single value of SM but a range of SM whose exact value can depend on other meteorological factors. Nevertheless, the impact of this uncertainty is neglectable as the key element is detection of any CSM point separating clear transitional and wet SM segments.

An index δ is proposed to quantify the degree of discrepancy among products in detecting SM regimes. The spatial distribution of detected SM regimes is found to vary strongly among the data sets, although for certain regions of the globe there is good agreement, such as in the subtropics and semi-arid regions. Though the causes of most of these disagreements can be inferred, further studies examining the relationship between LE and SM in different datasets will help to determine the reasons for the low consensus.

The degree of consensus in SM regimes among climate models can act as a confidence score when using multiple climate models to explore extremes in current

and/or projected climate. Shifts in SM regimes reflect a "change of gears" in land-atmosphere coupling, and thus the impacts of extremes. For example, when conditions switch from the transitional regime into the dry regime, disconnection of SM from LE and a stronger sensitivity of sensible heat flux to SM are implied (Dirmeyer et al. 2021; Benson and Dirmeyer 2021). Accordingly, if the existence of such a shift in SM regime lacks consensus among climate models, it could degrade the robustness of assessments related to processes or extremes that involve land-atmosphere interactions.

Days with SM in the transitional regime are used to evaluate the dependency of LE on SM across climate models and observationally-constrained products. A limitation is set that any given calendar month must contain at least 500 days in the data sets with SM in the transitional regime during the analysis period; much of the world passes this test during boreal summer. Generally, spatial patterns of mNMI and its decomposition are similar among the data products, although a smaller percentage of grid cells are found to be statistically significant in observationally-based datasets (ERA5, MERRA2, and SMAP). The SM-LE dependency, especially the linear component, within the transitional regime is found to be relatively strong over semiarid regions. Due to the universally positive dependency of LE to SM within the transitional regime, one might expect that much of the analyzed area would show strong SM-LE coupling. However, the strongly coupled areas found here remain limited in extent, like the land-atmosphere interaction "hot spots" in the previous studies that used data spanning the full range of SM values. This implies that the reason SM and LE are strongly coupled in semi-arid regions is not merely due to the preponderance of time SM values spend in the transitional regime.

We hypothesize the reasons that induce the coupling pattern seen in Figure 5. Besides the variability of SM and available energy, wind speed and atmospheric moisture deficit can also affect the rate of LE. Additionally, subsurface SM and air temperature can affect transpiration through biophysical effects on plants (Sellers et al. 1986). Among these factors, the magnitude of atmospheric moisture deficit might be the most important. Our results show that even though high latitudes and rainforest regions have sufficient days with SM in the transitional regime, SM and LE there are not strongly coupled. Following the Clausius-Clapeyron relationship, in high latitudes the normally low atmospheric moisture deficit results from low near surface air temperature, while in rainforest regions it is due to the year-round high humidity.

Small atmospheric moisture deficits limit SM-LE coupling during the course of the day even when SM is in the sensitive transitional regime.

Though strong differences exist in the distribution of SM regimes among the data products, the regions with strong SM-LE coupling are fairly consistent and similar to the previously identified “hot spots”. Combining the findings of this study, the low consensus of SM regimes yet similar patterns of SM-LE dependency among the data sets, yields the inference that despite a strong dispersion of local hydroclimates among the data sets, attributable to diverse potential factors such as monsoon extent or the physics of soil water retention, the inherent physics of how LE reacts to SM variability is well represented by land models. Recent studies have examined the relative role of net radiation to SM in determining surface heat fluxes (Haghighi et al 2018; Hsu and Dirmeyer 2021). A further investigation of how daily variability of variables such as wind speed, atmospheric moisture deficit, and air temperature compare to the relative importance of SM variations in determining LE can help to clarify the source or absence of locally strong coupling.

The framework here to determine any location’s span of SM regimes and critical SM values as well as associated coupling strength has potential applications. For forecasting, diagnosing the position and transition of SM among wet, transitional, and dry regimes enables inference of when and where land-atmosphere feedbacks, which play a crucial role in extremes such as heat waves and drought, may become important. Getting these transitions right, as well as slope and degree of nonlinearity within the sensitive regime, would be an indication of correct process representation and should improve model skill. Meanwhile, the disagreement index δ can be used on climate projections to examine the credibility of shifts in terrestrial hydrology in different scenarios.

Finally, this study is confined to the SM-LE relationship because of the lack of availability of daily fields of sensible heat flux from most of the CMIP6 models. Given that it has been increasingly common for studies to apply their own frameworks on long-term daily data sets of land surface variables and several new features of SM-surface heat flux relationships have been discovered, we encourage CMIP modeling groups to provide complete daily fields of surface heat fluxes relevant to surface water and energy balances for both historical and projected simulations.

***Acknowledgements**:*

This work was supported by the National Aeronautics and Space Administration (80NSSC20K1803). We are grateful to the developers of the data products used in this study.

Data Availability Statement

ERA5 (Hersbach et al. 2020, doi:10.1002/qj.3803) was downloaded from the Copernicus Climate Change Service (C3S) Climate Data Store. MERRA-2 was downloaded from Global Modeling and Assimilation Office (GMAO 2015, doi:10.5067/RKPHT8KC1Y1T). SMAP L4 (Reichle et al. 2017, doi:10.5067/B59DT1D5UMB4) data was downloaded from NASA National Snow and Ice Data Center. CMIP6 data was downloaded from https://esgf-node.llnl.gov/search/cmip6/.

References

(1) Benson, D. O.; Dirmeyer, P. A. Characterizing the Relationship between Temperature and Soil Moisture Extremes and Their Role in the Exacerbation of Heat Waves over the Contiguous United States. *Journal of Climate* **2021**, *34* (6), 2175–2187. https://doi.org/10.1175/JCLI-D-20-0440.1.

(2) Bonan, G. B. Forests and Climate Change: Forcings, Feedbacks, and the Climate Benefits of Forests. *Science* **2008**, *320* (5882), 1444–1449. https://doi.org/10.1126/science.1155121.

(3) Boucher, O.; Denvil, S.; Levavasseur, G.; Cozic, A.; Caubel, A.; Foujols, M.-A.; Meurdesoif, Y.; Cadule, P.; Devilliers, M.; Ghattas, J.; Lebas, N.; Lurton, T.; Mellul, L.; Musat, I.; Mignot, J.; Cheruy, F. IPSL IPSL-CM6A-LR Model Output Prepared for CMIP6 CMIP, 2018. https://doi.org/10.22033/ESGF/CMIP6.1534.

(4) Cover, T. M.; Thomas, J. A. ELEMENTS OF INFORMATION THEORY. 774.

(5) Danabasoglu, G. NCAR CESM2 Model Output Prepared for CMIP6 CMIP, 2019. https://doi.org/10.22033/ESGF/CMIP6.2185.

(6) Denissen, J. M. C.; Teuling, A. J.; Reichstein, M.; Orth, R. Critical Soil Moisture Derived From Satellite Observations Over Europe. *J. Geophys. Res. Atmos.* **2020**, *125* (6). https://doi.org/10.1029/2019JD031672.

(7) Dirmeyer, P. A. The Terrestrial Segment of Soil Moisture-Climate Coupling: SOIL MOISTURE-CLIMATE COUPLING. *Geophys. Res. Lett.* **2011**, *38* (16), n/a-n/a. https://doi.org/10.1029/2011GL048268.

(8) Dirmeyer, P. A.; Balsamo, G.; Blyth, E. M.; Morrison, R.; Cooper, H. M. Land-Atmosphere Interactions Exacerbated the Drought and Heatwave Over Northern Europe During Summer 2018. *AGU Advances* **2021**, *2* (2). https://doi.org/10.1029/2020AV000283.

(9) Diro, G. T.; Sushama, L.; Martynov, A.; Jeong, D. I.; Verseghy, D.; Winger, K. Land-atmosphere Coupling over North America in CRCM5. *J. Geophys. Res. Atmos.* **2014**, *119* (21). https://doi.org/10.1002/2014JD021677.

(10)

EC-Earth Consortium (EC-Earth). EC-Earth-Consortium EC-Earth3-Veg Model Output Prepared for CMIP6 CMIP, 2019. https://doi.org/10.22033/ESGF/CMIP6.642.

(11) Eltahir, E. A. B. A Soil Moisture-Rainfall Feedback Mechanism: 1. Theory and Observations. *Water Resour. Res.* **1998**, *34* (4), 765–776. https://doi.org/10.1029/97WR03499.

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

(13) Findell, K. L.; Eltahir, E. A. B. An Analysis of the Soil Moisture-Rainfall Feedback, Based on Direct Observations from Illinois. *Water Resour. Res.* **1997**, *33* (4), 725–735. https://doi.org/10.1029/96WR03756.

(14) Fischer, E. M.; Seneviratne, S. I.; Vidale, P. L.; Lüthi, D.; Schär, C. Soil Moisture–Atmosphere Interactions during the 2003 European Summer Heat Wave. *Journal of Climate* **2007**, *20* (20), 5081–5099. https://doi.org/10.1175/JCLI4288.1.

(15) Gevaert, A. I.; Miralles, D. G.; Jeu, R. A. M.; Schellekens, J.; Dolman, A. J. Soil Moisture-Temperature Coupling in a Set of Land Surface Models. *J. Geophys. Res. Atmos.* **2018**, *123* (3), 1481–1498. https://doi.org/10.1002/2017JD027346.

(16) Guillod, B. P.; Orlowsky, B.; Miralles, D. G.; Teuling, A. J.; Seneviratne, S. I. Reconciling Spatial and Temporal Soil Moisture Effects on Afternoon Rainfall. *Nat Commun* **2015**, *6* (1), 6443. https://doi.org/10.1038/ncomms7443.

(17) Guo, H.; John, J. G.; Blanton, C.; McHugh, C.; Nikonov, S.; Radhakrishnan, A.; Rand, K.; Zadeh, N. T.; Balaji, V.; Durachta, J.; Dupuis, C.; Menzel, R.; Robinson, T.; Underwood, S.; Vahlenkamp, H.; Bushuk, M.; Dunne, K. A.; Dussin, R.; Gauthier, P. P.; Ginoux, P.; Griffies, S. M.; Hallberg, R.; Harrison, M.; Hurlin, W.; Lin, P.; Malyshev, S.; Naik, V.; Paulot, F.; Paynter, D. J.; Ploshay, J.; Reichl, B. G.; Schwarzkopf, D. M.; Seman, C. J.; Shao, A.; Silvers, L.; Wyman, B.; Yan, X.; Zeng, Y.; Adcroft,

A.; Dunne, J. P.; Held, I. M.; Krasting, J. P.; Horowitz, L. W.; Milly, P. C. D.; Shevliakova, E.; Winton, M.; Zhao, M.; Zhang, R. NOAA-GFDL GFDL-CM4 Model Output, 2018. https://doi.org/10.22033/ESGF/CMIP6.1402.

(18) Herold, N.; Kala, J.; Alexander, L. V. The Influence of Soil Moisture Deficits on Australian Heatwaves. *Environ. Res. Lett.* **2016**, *11* (6), 064003. https://doi.org/10.1088/1748-9326/11/6/064003.

(19) Hersbach, H.; Bell, B.; Berrisford, P.; Hirahara, S.; Horányi, A.; Muñoz‐Sabater, J.; Nicolas, J.; Peubey, C.; Radu, R.; Schepers, D.; Simmons, A.; Soci, C.; Abdalla, S.; Abellan, X.; Balsamo, G.; Bechtold, P.; Biavati, G.; Bidlot, J.; Bonavita, M.; Chiara, G.; Dahlgren, P.; Dee, D.; Diamantakis, M.; Dragani, R.; Flemming, J.; Forbes, R.; Fuentes, M.; Geer, A.; Haimberger, L.; Healy, S.; Hogan, R. J.; Hólm, E.; Janisková, M.; Keeley, S.; Laloyaux, P.; Lopez, P.; Lupu, C.; Radnoti, G.; Rosnay, P.; Rozum, I.; Vamborg, F.; Villaume, S.; Thépaut, J. The ERA5 Global Reanalysis. *Q. J. R. Meteorol. Soc.* **2020**, *146* (730), 1999–2049. https://doi.org/10.1002/qj.3803.

(20) Hirschi, M.; Mueller, B.; Dorigo, W.; Seneviratne, S. I. Using Remotely Sensed Soil Moisture for Land–Atmosphere Coupling Diagnostics: The Role of Surface vs. Root-Zone Soil Moisture Variability. *Remote Sensing of Environment* **2014**, *154*, 246–252. https://doi.org/10.1016/j.rse.2014.08.030.

(21) Hirschi, M.; Seneviratne, S. I.; Alexandrov, V.; Boberg, F.; Boroneant, C.; Christensen, O. B.; Formayer, H.; Orlowsky, B.; Stepanek, P. Observational Evidence for Soil-Moisture Impact on Hot Extremes in Southeastern Europe. *Nature Geosci* **2011**, *4* (1), 17–21. https://doi.org/10.1038/ngeo1032.

(22) Hsu, H.; Dirmeyer, P. A. Nonlinearity and Multivariate Dependencies in the Terrestrial Leg of Land‐Atmosphere Coupling. *Water Res.* **2021**, *57* (2). https://doi.org/10.1029/2020WR028179.

(23) Hsu, H.; Lo, M.-H.; Guillod, B. P.; Miralles, D. G.; Kumar, S. Relation between Precipitation Location and Antecedent/Subsequent Soil Moisture Spatial Patterns: Precipitation-Soil Moisture Coupling. *J. Geophys. Res. Atmos.* **2017**, *122* (12), 6319–6328. https://doi.org/10.1002/2016JD026042.

(24)
Koster, R. D. Regions of Strong Coupling Between Soil Moisture and Precipitation. *Science* **2004**, *305* (5687), 1138–1140. https://doi.org/10.1126/science.1100217.
(25)
Koster, R. D.; Milly, P. C. D. The Interplay between Transpiration and Runoff Formulations in Land Surface Schemes Used with Atmospheric Models. *J. Climate* **1997**, *10* (7), 1578–1591. https://doi.org/10.1175/1520-0442(1997)010<1578:TIBTAR>2.0.CO;2.
(26)
Koster, R. D.; Sud, Y. C.; Guo, Z.; Dirmeyer, P. A.; Bonan, G.; Oleson, K. W.; Chan, E.; Verseghy, D.; Cox, P.; Davies, H.; Kowalczyk, E.; Gordon, C. T.; Kanae, S.; Lawrence, D.; Liu, P.; Mocko, D.; Lu, C.-H.; Mitchell, K.; Malyshev, S.; McAvaney, B.; Oki, T.; Yamada, T.; Pitman, A.; Taylor, C. M.; Vasic, R.; Xue, Y. GLACE: The Global Land–Atmosphere Coupling Experiment. Part I: Overview. *Journal of Hydrometeorology* **2006**, *7* (4), 590–610. https://doi.org/10.1175/JHM510.1.
(27)
Liu, D.; Wang, G.; Mei, R.; Yu, Z.; Gu, H. Diagnosing the Strength of Land–Atmosphere Coupling at Subseasonal to Seasonal Time Scales in Asia. *Journal of Hydrometeorology* **2014**, *15* (1), 320–339. https://doi.org/10.1175/JHM-D-13-0104.1.
(28)
Lorenz, R.; Pitman, A. J.; Hirsch, A. L.; Srbinovsky, J. Intraseasonal versus Interannual Measures of Land–Atmosphere Coupling Strength in a Global Climate Model: GLACE-1 versus GLACE-CMIP5 Experiments in ACCESS1.3b. *Journal of Hydrometeorology* **2015**, *16* (5), 2276–2295. https://doi.org/10.1175/JHM-D-14-0206.1.
(29)
Miralles, D. G.; van den Berg, M. J.; Teuling, A. J.; de Jeu, R. A. M. Soil Moisture-Temperature Coupling: A Multiscale Observational Analysis: SOIL MOISTURE-TEMPERATURE COUPLING. *Geophys. Res. Lett.* **2012**, *39* (21), n/a-n/a. https://doi.org/10.1029/2012GL053703.
(30)
Miralles, D. G.; Gentine, P.; Seneviratne, S. I.; Teuling, A. J. Land-Atmospheric Feedbacks during Droughts and Heatwaves: State of the Science and Current Challenges: Land Feedbacks during Droughts and Heatwaves. *Ann. N.Y. Acad. Sci.* **2019**, *1436* (1), 19–35. https://doi.org/10.1111/nyas.13912.
(31)

Reichle, R. H.; Liu, Q.; Koster, R. D.; Draper, C. S.; Mahanama, S. P. P.; Partyka, G. S. Land Surface Precipitation in MERRA-2. *Journal of Climate* **2017**, *30* (5), 1643–1664. https://doi.org/10.1175/JCLI-D-16-0570.1.

(32) Reichle, Rolf; Lannoy, Gabrielle De; Koster, Randal; Crow, Wade; Kimball, John. SMAP L4 Global 3-Hourly 9 Km Surface and Rootzone Soil Moisture Geophysical Data, Version 3, 2017. https://doi.org/10.5067/B59DT1D5UMB4.

(33) Ridley, J.; Menary, M.; Kuhlbrodt, T.; Andrews, M.; Andrews, T. MOHC HadGEM3-GC31-MM Model Output Prepared for CMIP6 CMIP, 2019. https://doi.org/10.22033/ESGF/CMIP6.420.

(34) Santanello, J. A.; Dirmeyer, P. A.; Ferguson, C. R.; Findell, K. L.; Tawfik, A. B.; Berg, A.; Ek, M.; Gentine, P.; Guillod, B. P.; van Heerwaarden, C.; Roundy, J.; Wulfmeyer, V. Land–Atmosphere Interactions: The LoCo Perspective. *Bulletin of the American Meteorological Society* **2018**, *99* (6), 1253–1272. https://doi.org/10.1175/BAMS-D-17-0001.1.

(35) Santanello, J. A.; Peters-Lidard, C. D.; Kumar, S. V. Diagnosing the Sensitivity of Local Land–Atmosphere Coupling via the Soil Moisture–Boundary Layer Interaction. *Journal of Hydrometeorology* **2011**, *12* (5), 766–786. https://doi.org/10.1175/JHM-D-10-05014.1.

(36) Schwarz, G. Estimating the Dimension of a Model. *Ann. Statist.* **1978**, *6* (2). https://doi.org/10.1214/aos/1176344136.

(37) Schwingshackl, C.; Hirschi, M.; Seneviratne, S. I. Quantifying Spatiotemporal Variations of Soil Moisture Control on Surface Energy Balance and Near-Surface Air Temperature. *Journal of Climate* **2017**, *30* (18), 7105–7124. https://doi.org/10.1175/JCLI-D-16-0727.1.

(38) Sellers, P. J.; Mintz, Y.; Sud, Y. C.; Dalcher, A. A Simple Biosphere Model (SIB) for Use within General Circulation Models. *J. Atmos. Sci.* **1986**, *43* (6), 505–531. https://doi.org/10.1175/1520-0469(1986)043<0505:ASBMFU>2.0.CO;2.

(39) Semmler, T.; Danilov, S.; Rackow, T.; Sidorenko, D.; Barbi, D.; Hegewald, J.; Sein, D.; Wang, Q.; Jung, T. AWI AWI-CM1.1MR Model

Output Prepared for CMIP6 CMIP, 2018. https://doi.org/10.22033/ESGF/CMIP6.359.

(40) Seneviratne, S. I.; Corti, T.; Davin, E. L.; Hirschi, M.; Jaeger, E. B.; Lehner, I.; Orlowsky, B.; Teuling, A. J. Investigating Soil Moisture–Climate Interactions in a Changing Climate: A Review. *Earth-Science Reviews* **2010**, *99* (3–4), 125–161. https://doi.org/10.1016/j.earscirev.2010.02.004.

(41) Seneviratne, S. I.; Lüthi, D.; Litschi, M.; Schär, C. Land–Atmosphere Coupling and Climate Change in Europe. *Nature* **2006**, *443* (7108), 205–209. https://doi.org/10.1038/nature05095.

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

(43) Smith, R. A Mutual Information Approach to Calculating Nonlinearity: Measuring Nonlinearity with Mutual Information. *STAT* **2015**, *4* (1), 291–303. https://doi.org/10.1002/sta4.96.

(44) Swart, N. C.; Cole, J. N. S.; Kharin, V. V.; Lazare, M.; Scinocca, J. F.; Gillett, N. P.; Anstey, J.; Arora, V.; Christian, J. R.; Jiao, Y.; Lee, W. G.; Majaess, F.; Saenko, O. A.; Seiler, C.; Seinen, C.; Shao, A.; Solheim, L.; von Salzen, K.; Yang, D.; Winter, B.; Sigmond, M. CCCma CanESM5 Model Output Prepared for CMIP6 CMIP, 2019. https://doi.org/10.22033/ESGF/CMIP6.1303.

(45) Takemura, T. MIROC MIROC6 Model Output Prepared for CMIP6 AerChemMIP, 2019. https://doi.org/10.22033/ESGF/CMIP6.9121.

(46) Volodin, E.; Mortikov, E.; Gritsun, A.; Lykossov, V.; Galin, V.; Diansky, N.; Gusev, A.; Kostrykin, S.; Iakovlev, N.; Shestakova, A.; Emelina, S. INM INM-CM5-0 Model Output Prepared for CMIP6 CMIP, 2019. https://doi.org/10.22033/ESGF/CMIP6.1423.

(47) Yukimoto, S.; Koshiro, T.; Kawai, H.; Oshima, N.; Yoshida, K.; Urakawa, S.; Tsujino, H.; Deushi, M.; Tanaka, T.; Hosaka, M.; Yoshimura, H.; Shindo, E.; Mizuta, R.; Ishii, M.; Obata, A.; Adachi, Y. MRI MRI-ESM2.0 Model Output Prepared for CMIP6 CMIP, 2019. https://doi.org/10.22033/ESGF/CMIP6.621.

(48) Zhang, J.; Wang, W.-C.; Wei, J. Assessing Land-Atmosphere Coupling Using Soil Moisture from the Global Land Data Assimilation System

and Observational Precipitation. *J. Geophys. Res.* 2008, *113* (D17), D17119. https://doi.org/10.1029/2008JD009807.

(49) Bayes Factors: Journal of the American Statistical Association: Vol 90, No 430 https://www.tandfonline.com/doi/abs/10.1080/01621459.1995.10476572 (accessed 2021 -10 -01).

(50) GES DISC Dataset: MERRA-2 tavg1_2d_lnd_Nx: 2d,1-Hourly,Time-Averaged,Single-Level,Assimilation,Land Surface Diagnostics V5.12.4 (M2T1NXLND 5.12.4) https://disc.gsfc.nasa.gov/datasets/M2T1NXLND_5.12.4/summary (accessed 2021 -10 -01).