Functional data decomposition reveals unexpectedly strong soil moisture–precipitation coupling over the Great Plains

Yifu Gao[1], Runze Li[1], Efi Foufoula-Georgiou[1], Jasper A. Vrugt[1*]

[1]Department of Civil and Environmental Engineering, University of California, Irvine, California, USA

Abstract

Soil moisture–precipitation coupling (SMPC) plays a critical role in Earth's water and energy cycles but remains difficult to quantify due to synoptic-scale variability and the complex interplay of land–atmosphere processes. Here, we apply high-dimensional model representation (HDMR) to functionally decompose the structural, correlative, and cooperative contributions of key land–atmosphere variables to precipitation. Benchmark tests confirm that HDMR overcomes limitations of commonly used correlation and regression approaches in isolating direct versus indirect effects. For example, analysis of gross primary productivity using a light-use-efficiency model shows that linear regression underestimates the temperature effect, while HDMR captures it accurately. Applying HDMR to CONUS404 reanalysis data reveals that morning soil moisture explains up to 40% of the variance in summertime afternoon precipitation over the Great Plains, more than double prior estimates. On days with afternoon rainfall (12-hour totals of 4.7-8.2 mm), first-order SM effects can boost precipitation by up to 8 mm under wet conditions, with an additional 3 mm from second-order interactions involving temperature and moisture. By capturing real-world co-variability and higher-order effects, HDMR provides a physically grounded, data-driven framework

*Corresponding Author: jasper@uci.edu

for diagnosing land–atmosphere coupling. These results underscore the need for more nuanced, interaction-aware data analysis methods in climate modeling and prediction.

1   Introduction

The interaction between soil moisture (SM) and precipitation is a critical driver of Earth's surface water and energy cycles [1]. This SM-precipitation coupling (SMPC) modulates near-surface processes and exerts control on energy partitioning [2], boundary layer dynamics [3, 4], and mesoscale circulation [5], thereby influencing regional extremes of precipitation [6, 7] and droughts [8]. SMPC occurs across a continuum of spatiotemporal scales, spanning distances from a few to several thousands of kilometers and extending from diurnal cycles to seasonal patterns [9, 10]. Moreover, SMPC exhibits substantial regional variability in both its strength and sign, attributed to the sensitivity of evapotranspiration to SM and of atmospheric conditions to latent heat fluxes [11]. An in-depth understanding of the coupling between SM and precipitation is essential for accurate weather forecasting and climate modeling, especially in the context of global warming and ongoing land-cover and land-use changes.

SMPC mechanisms can have either positive or negative effects on precipitation. These effects are categorized as positive (wet soil) or negative (dry soil) outcomes. As illustrated in Figure 1, a wet soil supports larger latent heat fluxes (or evapotranspiration) in a SM-limited regime [12], increasing the air's moisture content, moist static energy (MSE) [13] and evaporative fraction (EF) [2]. The triggered convection initiation and shallow planetary boundary layer (PBL) development [14] promote convective cloud formation and precipitation. Drier soils on the contrary have higher surface temperatures, a larger temperature gradient, and convective triggering potential (CTP) [15]. Enhanced thermal updrafts in turn reduce convection inhibition (CIN) and promote air parcels to reach the lifting condensation level (LCL) to form precipitation [16, 17]. At the mesoscale, both of these feedback mechanisms primarily manifest on a diurnal basis, especially during the midday hours of the warmer seasons. Precipitation

recycling is a more straightforward SMPC mechanism but has a more prolonged, non-local, and large-scale impact on the atmosphere [9, 18].

Over the past two decades, many data-driven and simulation-based studies have investigated the strength, sign, and governing mechanisms or pathways of SMPC [15, 2, 16, 17, 19, 4]. However, there is still ongoing debate about the extent and causal influence of SM on afternoon precipitation, particularly in hotspot regions such as the Central and Southeastern United States and the Sahel [2, 16, 20]. One major factor influencing SMPC estimates is the resolution and quality of SM data [21, 22]. Another important, yet underexplored, factor is the variability in how coupling strength is defined and computed across different studies. Most previous work quantifies SMPC using simple statistical metrics or model diagnostics, each with its own set of limitations, as discussed in earlier studies [1]. Specifically, linear (Pearson) [23, 4] or partial correlation coefficients [24] have been widely used to quantify the direct associations between SM and precipitation or convection initiation. These analyses are often preceded by Principal Component Analysis, which is applied to isolate the leading modes of variability in SM, precipitation, and temperature [25, 7]. However, correlation analysis fails to detect nonlinear and multivariate relationships among land–atmosphere variables. Although the coefficients of a multiple linear regression model can provide some insights into coupling strength [26, 27], such models are inherently linear and additive. They do not distinguish between direct and indirect effects, nor can they capture nonlinear interactions or multivariate dependencies among land-atmosphere variables and precipitation. Composite and Bayesian analyses offer some improvement by examining precipitation patterns under varying SM conditions and classifications [2, 28], but these approaches also fall short in disentangling direct effects from cooperative effects of land–atmosphere variables. Causal inference may help dial in on the direct relationship between SM and precipitation occurrence [29] but ignores higher-order synergies and interactions within the land–atmosphere system [30].

Global and regional climate models provide valuable causal insights into SMPC. These models can simulate the effects of SM perturbations on precipitation. This sensitivity is commonly used as a
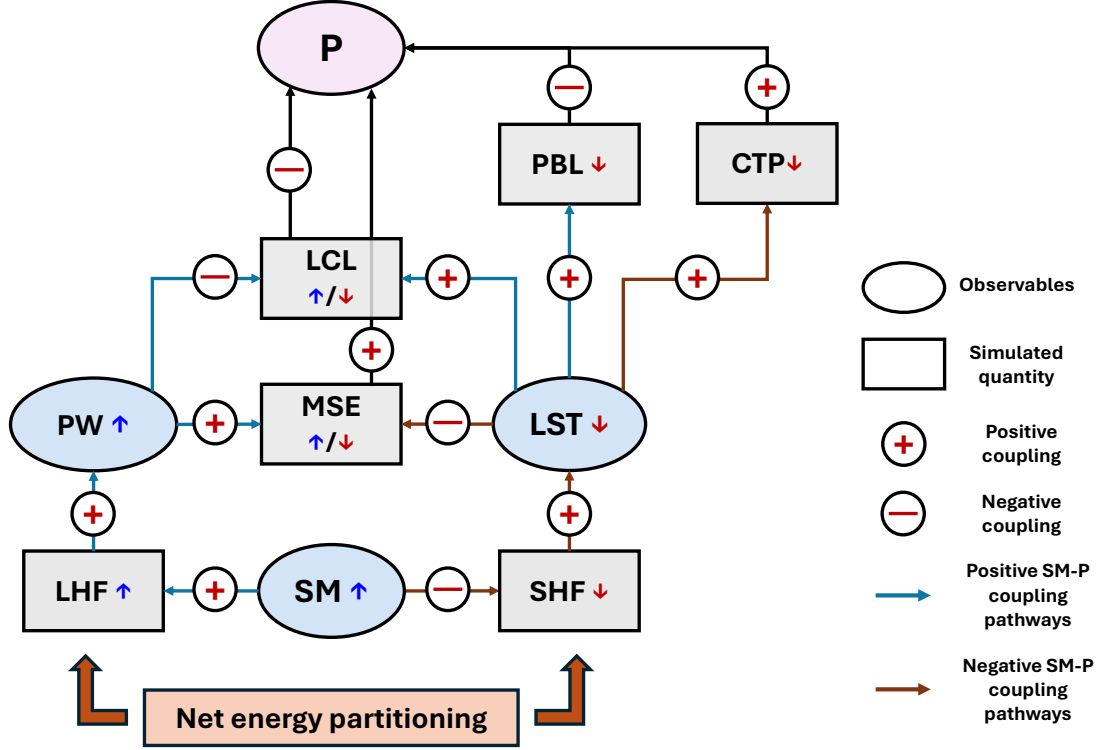
Figure 1: Schematic illustration of positive ("wet-soil effect") and negative ("dry-soil effect") soil moisture–precipitation coupling (SMPC). Key variables include soil moisture (SM), latent heat flux (LHF), sensible heat flux (SHF), land surface temperature (LST), precipitable water (PW), moist static energy (MSE), lifting condensation level (LCL), planetary boundary layer (PBL), and convective triggering potential (CTP). The arrows and interactions demonstrate how wet and dry soil conditions influence boundary-layer moisture, development, and precipitation.

proxy for the strength of land–atmosphere coupling [28, 31, 32, 33] or, alternatively, as the fraction of precipitation variance explained by SM, based on comparisons between reference and prescribed SM climatologies [34, 35]. However, such coupling diagnostics or inferred causal relationships are subject to considerable uncertainties, primarily due to the physical parameterizations of convective and boundary-layer processes within climate models [36, 37, 38]. Moreover, artificially imposed SM states used in model experiments often fail to capture realistic antecedent conditions (e.g., the memory effects of prior precipitation) thereby reducing the analysis to a one-way feedback that oversimplifies the full complexity of land–atmosphere interactions. Finally, because these diagnostics rely on controlled perturbations that cannot be reproduced in observational settings, direct validation of the modeled causal relationships is infeasible [1].

State-of-the-art data and sensitivity analysis methods offer new opportunities to analyze high-

dimensional multivariable datasets from both models and observations in the search for robust SMPC signals across spatiotemporal scales. Here, we illustrate the application of one such method – High-Dimensional Model Representation (HDMR) [39, 40, 41] – to CONUS404 reanalysis data. HDMR is a generalization of Sobol′ [42] finite multivariable function expansion to dependent variables, and decomposes the multivariate relationships and coupling among land–atmosphere variables into a hierarchical set of first-, second-, and higher-order response/component functions that capture direct and indirect effects of land–atmosphere variables to precipitation. HDMR's ability to parse out the structural, correlative, and cooperative contributions of land-atmosphere variables to precipitation exceeds the capabilities of commonly used correlation and regression methods and offers deeper insights into the multiple different pathways through which SM modulates precipitation in both modeling and observational frameworks. Our case studies demonstrate HDMR's potential for inferring causative relationships, quantifying the strength and sign of SMPC across CONUS and diagnosing land-atmosphere coupling in weather and climate models.

2   High Dimensional Model Representation

Most correlation and regression-based methods capture only direct and linear relationships between SM and precipitation, thereby overlooking the contributions of bivariate and higher-order cooperative effects (e.g., compound contribution of SM and PW) and internal coupling (e.g., SM-LST) of land-atmosphere variables [1]. HDMR, on the contrary, decomposes the SMPC into a hierarchy of component functions and uses an analysis of covariance (ANCOVA) to separate direct (structural) from indirect (correlative and cooperative) effects (see Figure 2). This enables a more systematic, multivariate and multivariable evaluation of how SM influences precipitation.

Let $\mathbf{x} = (x_1, \ldots, x_d)^\top$ be a vector of d input variables (e.g., morning SM and other land-atmosphere variables in Table 1) that explain afternoon precipitation or target variable y. HDMR decomposes the
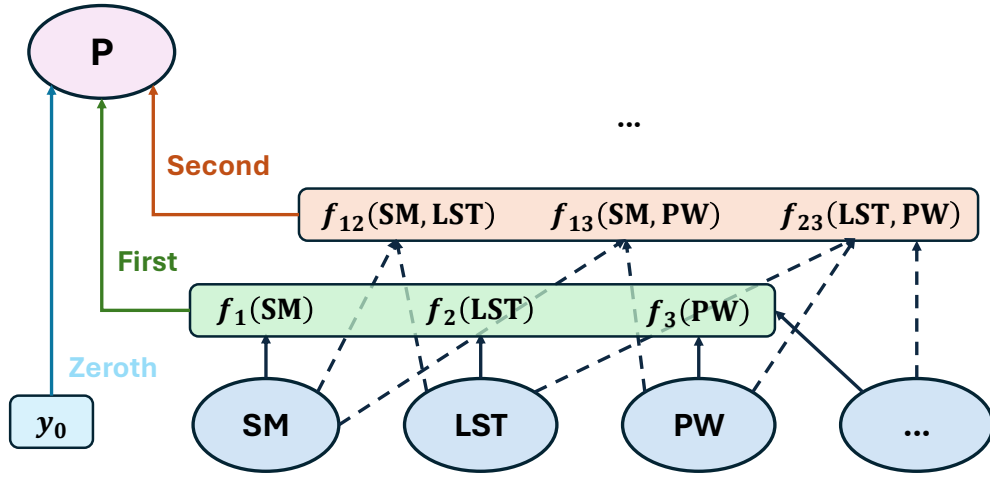
Figure 2: Schematic overview of the HDMR second-order decomposition of afternoon precipitation as a function of antecedent SM, LST and PW.

afternoon precipitation y into a sum of hierarchical component functions [40]:

$$y = y_0 + \sum_{i=1}^{n_1} f_i(x_i) + \sum_{1 \leq i < j \leq d}^{n_2} f_{ij}(x_i, x_j) + \sum_{1 \leq i < j < k \leq d}^{n_3} f_{ijk}(x_i, x_j, x_k)$$

$$+ \cdots + f_{12\ldots d}(x_1, x_2, \ldots, x_d) + \epsilon, \tag{1}$$

where y_0 is the mean 12-hr accumulated precipitation between noon and midnight (local time, and on days that experience afternoon rainfall) in units of mm and $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$ is a zero-mean normally distributed residual with constant variance, σ_ϵ^2. The first-order terms, $f_i(x_i)$, capture the individual contribution of each input variable, while the second- and higher-order terms such as $f_{ij}(x_i, x_j), \ldots,$ characterize cooperative contributions of two or more variables to y. In many physical systems, third- and higher-order terms are often negligible [43, 41]. Therefore, we retain only the $n_{12} = n_1 + n_2 = d + d(d-1)/2$ first- and second-order component functions:

$$y = y_0 + \sum_{u=1}^{n_{12}} f_u(\mathbf{x}_u) + \epsilon, \tag{2}$$

where index u runs over both individual (d) and pairwise $d(d-1)/2$ terms, and $\mathbf{x}_u$ denotes the subvector corresponding to the individual (x_i) and pairwise (x_i, x_j) components of the input vector

$\mathbf{x} = (x_1, \ldots, x_d)^\top$. The component functions f_u quantify the individual and bivariate contributions of the land-atmosphere variables to precipitation.

To successfully delineate the direct and indirect (cooperative/correlative) effects of the land-atmosphere variables, the component functions must satisfy a so-called *relaxed* vanishing condition [44]

$$\int_0^1 w_u(\mathbf{x}_u) f_u(\mathbf{x}_u) \mathrm{d}x_i = 0, \tag{3}$$

where $w_u(\mathbf{x}_u)$ is the joint probability density function of the variables in $\mathbf{x}_u$. (3) enforces hierarchical orthogonality of the component functions. This condition ensures a unique functional decomposition that properly distinguishes between structural, correlative, and cooperative contributions of the input variables [40, 41]. By definition, the relaxed vanishing condition requires each second-order function $f_{ij}(x_i, x_j)$ to be orthogonal to its associated first-order counterparts $f_i(x_i)$ and $f_j(x_j)$. This orthogonality ensures that each second-order HDMR term captures new, interactive information about the target variable y not already represented by lower-order terms. Only then can we successfully unravel the dynamic web of land-atmosphere variables.

Each component function is represented by a set of orthogonal polynomials $\phi_r(x_i)$ (see Materials and Methods 5.3.1). The corresponding expansion coefficients are estimated via D-MORPH regression (see Materials and Methods 5.3.2) and ensure hierarchical orthogonality. Next, the contribution of each component function $f_u(\mathbf{x}_u)$ to the total variance $\mathrm{Var}[y]$ of y, is quantified by *structural* and *correlative* sensitivity indices [45]:

$$S_u^{\mathrm{a}} = \frac{\mathrm{Var}[f_u(\mathbf{x}_u)]}{\mathrm{Var}[y]} \quad \text{and} \quad S_u^{\mathrm{b}} = \frac{\mathrm{Cov}[f_u(\mathbf{x}_u), \sum_{m \neq u}^{n_{12}} f_m(\mathbf{x}_m)]}{\mathrm{Var}[y]}. \tag{4}$$

Their sum defines the coupling index S_u of $f_u(\mathbf{x}_u)$:

$$S_u = S_u^{\mathrm{a}} + S_u^{\mathrm{b}}. \tag{5}$$

The cooperative coupling index of x_i sums up all higher-order coupling indices associated with variable x_i, $\sum_{j \neq i} S_{ij}(x_i, x_j)$. A closely related measure, the total effect [46]:

$$S_i^{\mathrm{T}}(x_i) = S_i(x_i) + \sum_{j \neq i} S_{ij}(x_i, x_j), \tag{6}$$

quantifies the total explained variance of y by variable x_i through its direct (structural) and indirect (correlative due to dependencies with other variables and cooperative due to nonlinear interactions) effects. HDMR shares elements with information-theoretic approaches [47, 48], which decompose mutual information into unique, redundant, and synergistic components. This information partitioning, however, assumes knowledge of the marginal and joint probability density functions (PDFs) of the variables, which can be challenging for high-dimensional datasets. In contrast, HDMR decomposes the target variance into structural, correlative, and cooperative contributions without requiring an underlying PDF, making it computationally more efficient and scalable to large datasets. Below, we present HDMR results for three case studies of increasing complexity. Details of the data, models and methods used in these studies are found in the Materials and Methods.

3 Results

3.1 Case I: A Simple Bivariate Function

We consider a simple bivariate function $y = x_1 + x_2 + x_1 x_2$ and compare the results of HDMR, correlation analysis, and multiple linear regression analysis for two different cases. In Exp1A x_1 and x_2

are independent, whereas these two variables are negatively correlated in Exp1B with $r_{x_1,x_2} = -0.6$ (see Materials and Methods 5.2.1 for details on the properties of x_1 and x_2). Figure 3 presents the results of our analysis for Exp1A (left) and Exp1B (right) where in each panel, the large bars depict the total fractional variance of y explained by x_1 (blue) and x_2 (orange) according to HDMR, correlation analysis, and linear Regression. For HDMR, the smaller stacked bars in dark blue and red portray the first-order structural and correlative contributions, respectively. The smaller purple and pink bars in Exp1B denote the cooperative effect of the second-order component function $f_{12}(x_1, x_2)$. In Exp1A with
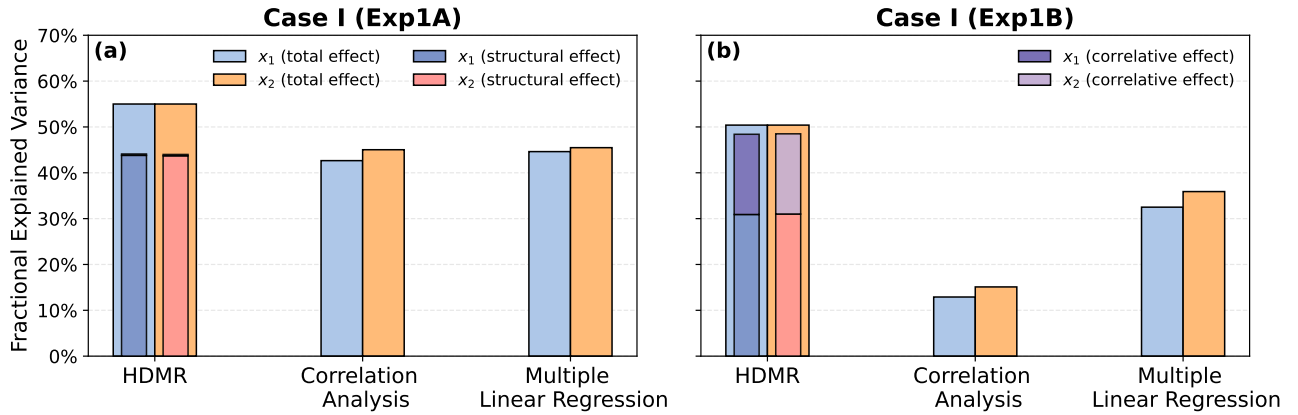

Figure 3: Fractional contributions of x_1 and x_2 to variance of y according to HDMR, Correlation analysis, and Linear Regression for Exp1A (left) and Exp1B (right). The large bars (blue for x_1 and orange for x_2) represent each method's total explained variance. For HDMR, the stacked (smaller) bars show the first-order structural and correlative contributions, respectively. The cooperative (interactive) effect of x_1 and x_2 equals the net difference between the HDMR total and the sum of structural and correlative effects.

uncorrelated inputs, HDMR's total effects for x_1 and x_2 are about 55% in line with their equal impact on y in (10). The structural effect of 43.8% reflects each variable's individual contribution to y. The correlative contribution of 0.3% for both variables is negligible (not visible in bars) and testifies to the independence of x_1 and x_2. The interaction term $S_{12} = S_1^{\mathrm{T}} - S_1^{\mathrm{a}} - S_1^{\mathrm{b}}$ of about 10.9% accounts for the remainder. The sum of the total effect $\sum S_i^{\mathrm{T}}$ does not necessarily equal 1 because the interaction term S_{12} is counted twice in (6). If we compute the normalized total effect $S_i^{\mathrm{TN}}(x_i)$ instead:

$$S_i^{\mathrm{TN}}(x_i) = S_i(x_i) + \frac{1}{2}\sum_{i \neq j} S_{ij}(x_i, x_j), \tag{7}$$

then $S_1^{\mathrm{TN}}(x_1) = S_2^{\mathrm{TN}}(x_2) = 43.8 + 11/2 = 49.3\%$. This matches the theoretical expected contribution of 50% for x_1 and x_2 in (10). Correlation analysis and multiple linear regression report noticeably lower values of 42.6% and 45.5% using (23)-(24), respectively. These two numbers are in agreement with HDMR's structural effects for x_1 and x_2 but omit the interaction altogether. This merely illustrates that correlation and regression-based methods overlook the contribution of the product term $x_1 x_2$ and do not adequately capture cooperative effects on y.

The right panel shows that correlation $r_{x_1,x_2} = -0.6$ among variables x_1 and x_2 drastically reduces the fractional explained variance by these two variables according to correlation analysis and multiple linear regression. This correlation does not affect HDMR-based estimates of the total effects of x_1 and x_2. Both are locked in at 50% and split into a structural part of 30.9% and correlative component of 17.5%. This time, the interaction term explains only 2% of the variance of y. This is noticeable less than in the uncorrelated case and this reduction is due to the negative correlation $r_{x_1,x_2} = -0.6$ of x_1 and x_2. These results reiterate HDMR's ability to successfully decompose the variance of y in the presence of variable correlation and interaction. Correlation and regression methods fail in doing so.

3.2 Case II: A Light-Use-Efficiency (LUE) Model

We use a light use efficiency (LUE) model to simulate data of gross primary productivity (GPP, gC/m^2/day) as a function of SM, net radiation R_{n}, 2-m air temperature $T_{2\mathrm{m}}$, and wind speed u. Details are presented in the Materials and Methods 5.2.2. We use this simulated data set to quantify the contributions of $x_1 = \mathrm{SM}$, $x_2 = R_{\mathrm{n}}$, $x_3 = T_{2\mathrm{m}}$ and $x_4 = u$ to GPP. All three methods (HDMR, correlation analysis, and multiple linear regression) are in agreement (see Figure 4a) on the importance of the four input variables in explaining the variance of GPP. Net radiation exerts the largest control on GPP, followed by 2-m air temperature, SM and wind speed. The effects of SM and wind speed are relatively minimal. Despite assigning similar rankings to the input variables, correlation analysis and multiple linear regression substantially underestimate the explained GPP variance by the different variables. This

is particularly true for T_{2m} or x_3. Correlation analysis (22.8%) and multiple linear regression (19.3%) substantially underestimate the temperature contribution to GPP. According to HDMR, the explained GPP variance by T_{2m} is 50.7%. To understand these differences, Fig. 4b examines the structural and
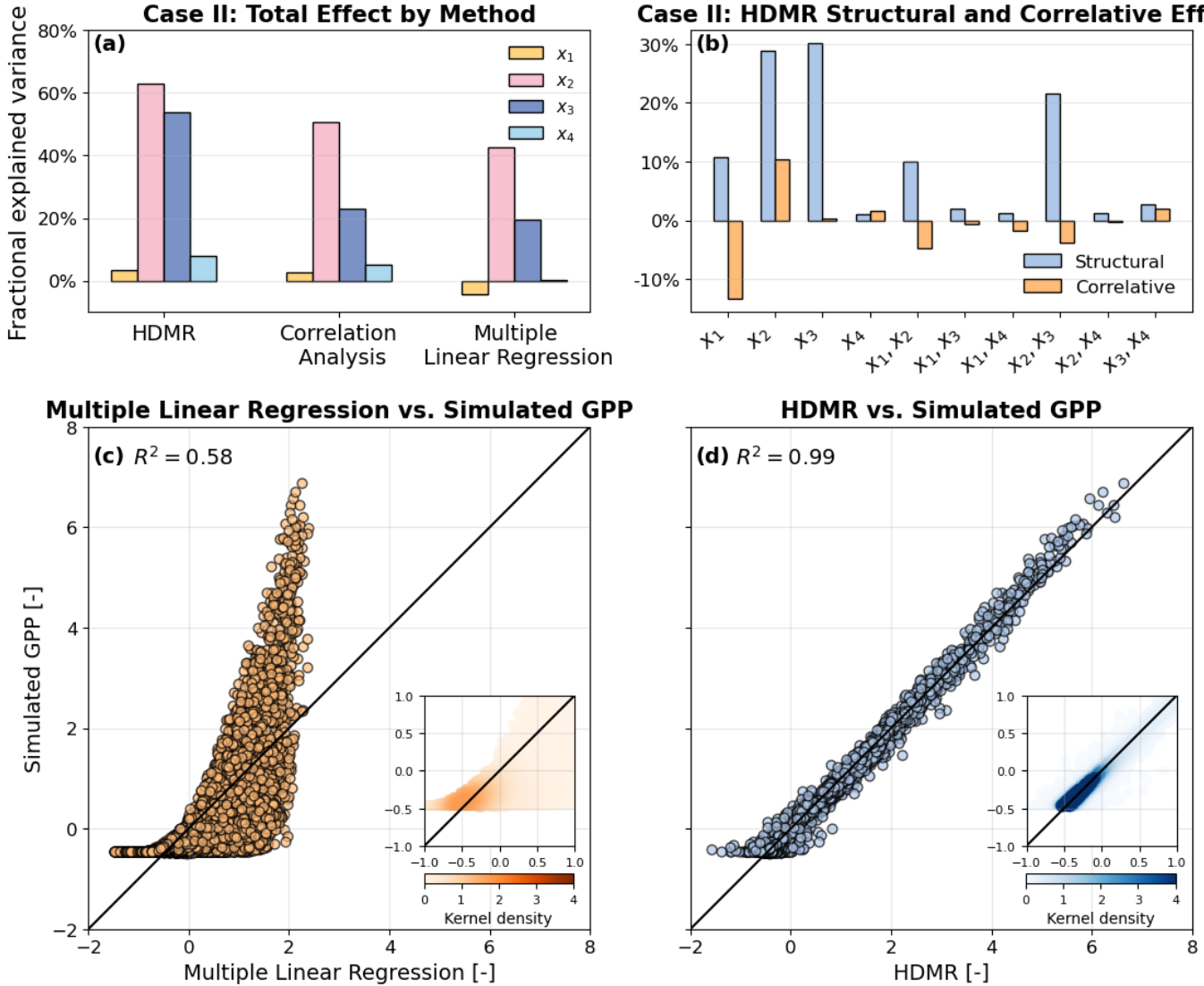

Figure 4: Top row: (a) total fractional variance of GPP explained by soil moisture x_1, net radiation x_2, 2-m air temperature x_3 and wind speed x_4 according to HDMR, correlation analysis, and multiple linear regression, and (b) structural and correlative effects of HDMR component functions. Bottom row: scatter plots of (c) multiple linear regression and (d) HDMR predicted GPP values (x-axis) against their simulated values (standardized) from the LUE model, with the 1:1 reference line. The small insets in (c,d) present density estimates of a Gaussian kernel for the [-1,1] data range (Scott bandwidth = 0.22).

correlative contributions of the HDMR component functions. The structural effect of T_{2m} of about 30% is already about 10% larger than the total explained GPP variance by correlation analysis and multiple linear regression. This result underscores HDMR's effectiveness in capturing nonlinear dependencies

that are missed by linear methods. In addition, several second-order component functions contribute significantly to GPP, indicating strong interactive effects among the variables. Notably, the joint influence of radiation and temperature (x_2, x_3) explains nearly 20% of the variance in GPP, an effect that correlation and regression methods fail to capture. The structural and correlative contributions of the component functions jointly explain 99% of the GPP variance. This inspires confidence in the HDMR functional decomposition of the LUE model. Further evidence for this is presented in Fig. 4d which shows an excellent agreement ($R^2 = 0.99$) of HDMR predicted and LUE simulated GPP values. Multiple linear regression, on the contrary, explains a much smaller fraction ($R^2 = 0.58$) of the GPP data (see Fig. 4c).

3.3   Case III: Soil Moisture–Precipitation Coupling (SMPC)

3.3.1   Hot Spots of SMPC

We now apply HDMR to CONUS404 reanalysis data to quantify the SMPC strength and identify hotspot regions across the Conterminous United States, focusing on the summer months (April–September) from 2012 to 2021. Data preparation is described in sections 5.1 and 5.2.3 of the Materials and Method. In short, we examine the 12-hour accumulated precipitation (12:00–24:00 local time), as function of x_1 morning SM (07:00 local time) and five auxiliary land–atmosphere variables including x_2 LST, x_3 PW, x_4 LAI, and the x_5 horizontal and x_6 vertical wind speeds at 10 meter. Synoptic effects were removed by relinquishing samples with antecedant rainfall in the 24 hours leading up to the afternoon events.

Figure 5a shows HDMR-derived estimates of the total variance in afternoon precipitation over CONUS that is explained by SM alone. We benchmark these results against correlation analysis (Fig. 5b) and multiple linear regression (5c), as these two data analysis methods are commonly used in SMPC studies [9, 23, 33, 26, 49].

All three methods identify the Great Plains as a region of relatively strong SM influence on precipita-
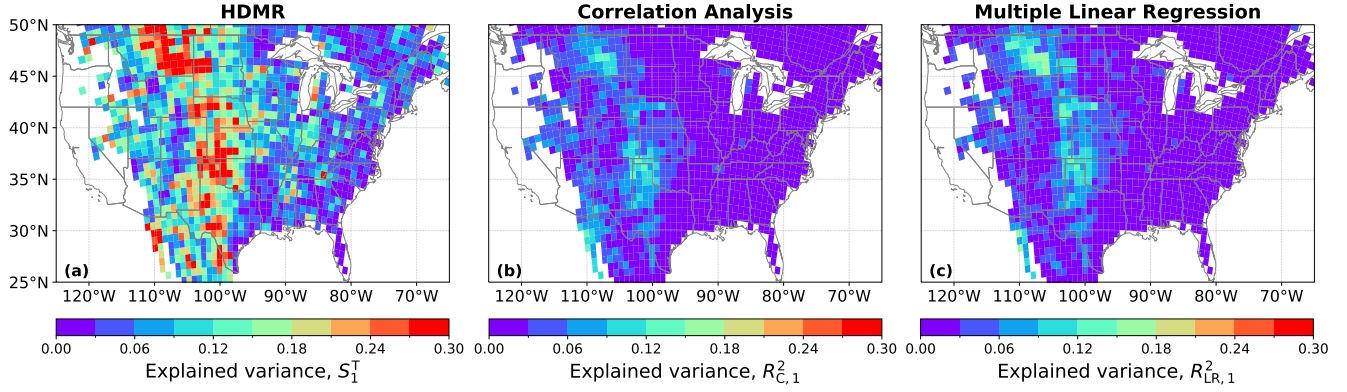
Figure 5: Total fractional variance of afternoon accumulated precipitation (from noon to midnight) during the warm season (April-September, 2012–2021) explained by morning SM across CONUS: (a) HDMR, (b) correlation analysis, and (c) multiple linear regression. Only grid cells with $5,000$ or more valid data samples are shown in color.

tion. The three methods differ substantially in their magnitudes of the explained precipitation variance. According to HDMR, SM explains between 0–40% of the summer convective precipitation variance over much of the central and northern Great Plains. This is much larger than the 5-15% derived from correlation analysis and multiple linear regression. Physically, SM plays a particularly influential role in regions where evapotranspiration is water-limited. In these transitional zones, moderate anomalies in soil wetness can have a considerable impact on surface energy fluxes and boundary-layer moisture, thereby influencing convective initiation and rainfall patterns [1]. In contrast, humid regions such as the Southeastern US are characterized by SM abundance and operate under energy-limited conditions, where precipitation is primarily controlled by atmospheric dynamics than by local SM variability or land surface processes [1]. This regime is reflected in the low SM-based explained variance seen in Figure 5. Nevertheless, SM may still influence the frequency of afternoon rainfall events through a positive feedback between evaporative fraction and precipitation [2]. Similarly, the mountainous western US also exhibits relatively weak SM impacts on precipitation, with the notable exception of regions influenced by monsoonal surges (e.g., New Mexico), aligning with previous studies [50, 51].

Our results are in strong qualitative agreement with GLACE findings [35], but demonstrate a much stronger signal over the Great Plains. According to HDMR, morning SM explains up to 40% of summertime precipitation variance in this region. This signal is much stronger than the approximately

16% of Koster et al. [35]. Two methodological differences likely explain this discrepancy. First, Koster et al. used a global climate model with relatively coarse resolution and simplified convection schemes, whereas our analysis is based on CONUS404 – a high-resolution WRF-derived reanalysis data set that much better captures and describes mesoscale convective systems, boundary-layer dynamics, and precipitation patterns. Second, the GLACE simulations quantify SMPC strength ($\Delta\Omega$) by prescribing soil moisture climatologies. This "local-sensitivity" approach decouples the land-atmosphere feedback, suppressing the natural co-variability among land surface and atmospheric variables. In contrast, HDMR is a variance-based global sensitivity analysis method that allows all land-atmosphere variables to vary simultaneously. It partitions precipitation variance into direct (unique) and indirect (shared) contributions from individual drivers and their interactions. As a result, HDMR provides a more physically consistent and substantially higher estimate of SM control on warm-season precipitation.

3.3.2 Structural, Correlative, and Cooperative Coupling Effects

The results in Fig. 5 raise an important question: Why does HDMR outperform correlation analysis and multiple linear regression in decomposing precipitation variance? Is this a result of model complexity, specifically, the large number of estimable expansion coefficients in the hierarchical polynomial component functions of the HDMR functional decomposition? To investigate this, we retain the exact same polynomial basis but estimate the expansion coefficients using ordinary least squares (OLS) rather than the HDMR-specific D-MORPH regression. The OLS estimator does not enforce the relaxed vanishing condition in (3), which is critical for disentangling first- and higher-order structural and correlative effects. For a representative grid cell in the northern Great Plains (46.2°N, 106.0°W), the OLS-derived SM component function is

$$f_1^{\mathrm{OLS}}(x_1) = 0.122\phi_1(x_1) + 0.102\phi_2(x_1) + 0.068\phi_3(x_1), \tag{8}$$

yielding a structural coupling index of $S_1^a = 3\%$. Using the same basis functions, D-MORPH regression produces:

$$f_1^{\text{HDMR}}(x_1) = 0.433\phi_1(x_1) + 0.320\phi_2(x_1) + 0.212\phi_3(x_1), \tag{9}$$

with a substantially higher coupling index of $S_1^a = 34\%$. While both functions explain the same proportion of variance in precipitation ($R^2 = 57\%$), the OLS parameterization of the SM component function obscures the direct SM effect. In contrast, hierarchical orthogonality enforced by D-MORPH regression uniquely partitions the variance into structural, correlative, and cooperative effects (see Figure 6), making the physical role of SM explicit. These findings highlight that HDMR's superiority lies not in a richer basis, but in its ability to delineate compound influences. This is a prerequisite for mechanistic interpretation of land-atmosphere coupling. To further unravel how SM, LST, and PW individually
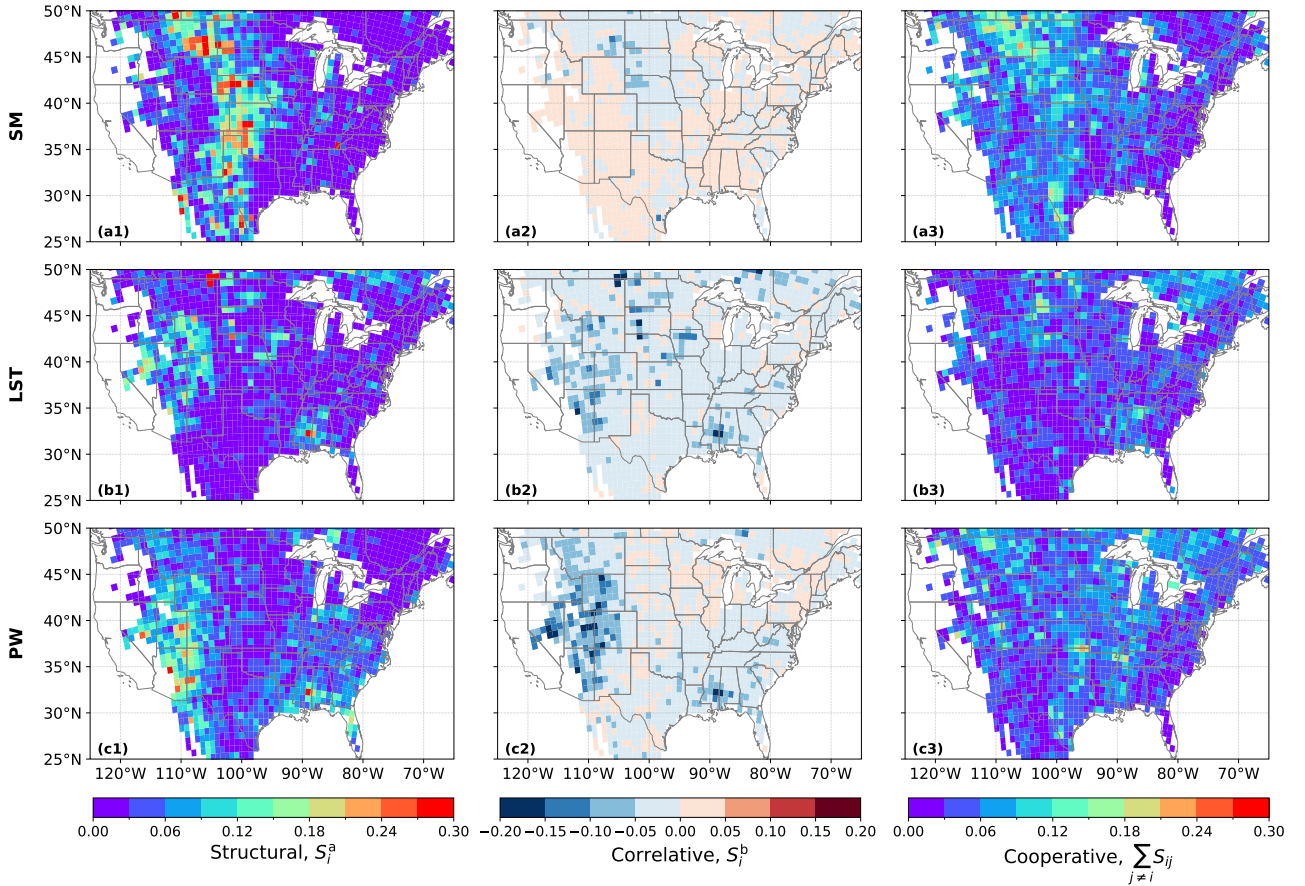

Figure 6: HDMR results for CONUS: Structural (a1-c1), correlative (a2-c2), and cooperative (a3-c3) coupling indices of (a) SM, (b) LST, and (c) PW.

and jointly contribute to precipitation, we analyze their structural, correlative, and cooperative HDMR

indices (Figure 6). LAI, u_{10}, and v_{10} have only a minimal contribution to the afternoon precipitation (see Figure S1) and are not discussed further. Provided that SM's total effect in the Great Plains (Figure 5, left panel) can exceed 30%, its structural contribution (Fig. 6a1) makes up almost 80% of that fraction, indicating a direct and "stand-alone" SM impact on precipitation variability. Correlative SM contributions are generally small and possibly even somewhat negative (Fig. 6a2). This finding demonstrates that once the direct contribution of SM to precipitation is accounted for, correlation with other inputs do not systematically amplify precipitation. Yet, for certain hot spots SM displays a moderate cooperative effect (Fig. 6a3) with other variables, underscoring SM's nontrivial interactions with other land-atmosphere variables. Indeed, the nonzero total coupling indices S_{12} and S_{13} in the Northwestern US (see Figure S2) indicate that morning SM, surface heating, and atmospheric humidity jointly influence surface moisture fluxes, CIN, and evaporative cooling, thereby potentially modulating afternoon precipitation.

The roles of LST and PW differ markedly from SM, especially in mountainous and coastal regions. Their structural indices approach 10–25% over the Rocky mountains (Fig. 6b1-c1) and highlights the importance of abundant surface heating and atmospheric moisture for boundary-layer destabilization and moist convection. Yet, some regions exhibit negative correlative indices (Fig. 6b2-c2), where high temperatures coincide with conditions that suppress local rainfall potentially due to dry soils and reduced CAPE [52]. PW shows a structural effect of up to 25% in the Southern Rockies and Gulf Coast (Fig. 6c1), where moisture-rich air enhances convective potential [53]. However, in areas frequently affected by large-scale moisture advection, negative correlative contributions diminish PW's net influence. The cooperative effects (Fig. 6b3-c3) of LST and PW are generally modest across CONUS, with localized enhancements along the Southeastern coast, the Southern Rockies, and the Northeastern US-regions where compound processes such as surface heating and tropospheric moisture advection more strongly influence afternoon precipitation Overall, the total effects of LST and PW (Figure S3) are comparatively smaller and generally below 15% across CONUS. This reaffirms the dominant control of SM on warm-

season precipitation and its characteristic hot spots within transitional climate zones.

3.3.3   The Functional SMPC

The most pronounced SMPC signal is observed over the Great Plains (Fig. 5a), and, thus, we further examine this region to help unlock the physical mechanisms that govern the feedback between SM and precipitation. Fig. 7a displays the first-order SM component functions, $f_1(x_1)$, for three representative grid cells in the northern (46.2°N, 106.0°W, blue line), central (37.7°N, 98.5°W, orange line), and southern (30.4°N, 100.1°W, green line) Great Plains. The mean 12-hour afternoon rainfall for the analyzed days (i.e., y_0) at the three grid cells is 4.7, 8.0, and 8.2 mm, respectively. For a cell in the northern Great Plains (NPG), we display the second-order component functions of variable pairs SM, LST and SM, PW designated $f_{12}(x_1, x_2)$ and $f_{13}(x_1, x_3)$ in Figs. 7b and 7c, respectively. The two bivariate interaction terms have relatively high total coupling indices ($S_{12} = 8.2\%$ and $S_{13} = 9.6\%$), compared to other SM-related second-order terms (see Figure S2). This warrants a closer inspection.
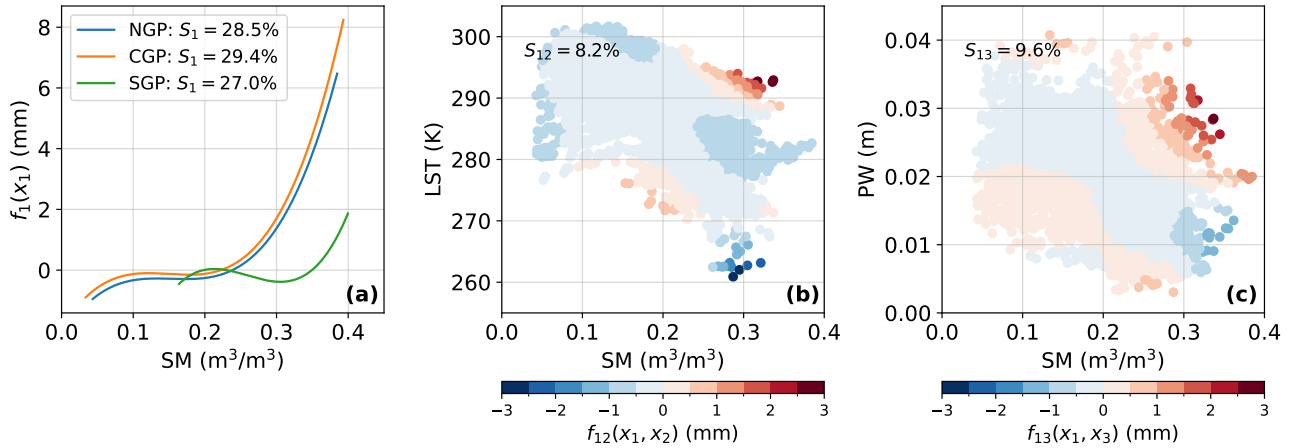

Figure 7: HDMR results of SMPC in the Great Plains: (a) First-order contribution of SM, $f_1(x_1)$, to afternoon precipitation for three representative grid cells located in the northern (46.2°N, 106.0°W, blue line), central (37.7°N, 98.5°W, orange line), and southern (30.4°N, 100.1°W, green line) Great Plains denoted NGP, CGP, and SGP, respectively, (b,c) Contribution of second-order component functions, $f_{12}(x_1, x_2)$ (SM–LST) and $f_{13}(x_1, x_3)$ (SM–PW) to precipitation for the NGP grid cell. Coupling indices S_1, S_{12}, and S_{13} quantify the fraction of precipitation variance explained by their respective first- and second-order component functions.

In Fig. 7a, the first-order SM component functions exhibit a roughly piecewise relationship between soil wetness and precipitation at the three Great Plains sites, highlighting a distinct "wet-soil advantage".

17

12-hour afternoon precipitation shows minimal sensitivity to soil wetness when SM drops below 0.2 cm^3/cm^3 for the NGP and CGP, and below 0.3 cm^3/cm^3 for the SGP. Above these thresholds, increasing SM leads to enhanced 12-hour precipitation, with peak increases of approximately 6, 8, and 2 mm for the NGP, CGP, and SGP, respectively. These results align closely with the spatial structure of SMPC shown in Fig. 5a1. As expected, the strongest SMPC signals are in the northern and central US, whereas farther south (e.g., in Texas), moisture advection from the Gulf of Mexico and stronger capping inversions likely suppress the local influence of SM on afternoon rainfall [54].

The second-order SM-LST component function $f_{12}(x_1, x_2)$ for the NGP (Fig. 7b) indicates that SM-LST coupling can modulate precipitation by up to 3 mm when wet soils coincide with high surface temperatures. While the first-order SM component function quantifies the direct contribution of soil water availability to precipitation, the interaction term $f_{12}(x_1, x_2)$ reveals how co-occurring high SM and LST jointly enhance evapotranspiration and convective development. Strong solar heating over moist surfaces amplifies moisture fluxes and boundary-layer growth, thereby boosting precipitation. Conversely, low LST suppresses surface fluxes and reduces precipitation, even when SM is abundant. Fig. 7c shows that SM-PW interactions can contribute up to 3 mm of additional precipitation when soils are wet (SM > 0.30 cm^3/cm^3) and atmospheric moisture is moderately high ($2 < \text{PW} < 3$ cm). This cooperative effect between SM and PW underscores how moist soils, in tandem with elevated tropospheric humidity, can reduce convective inhibition and promote more efficient rainfall production once storms initiate [55]. Accordingly, while LST and PW alone do not dominate precipitation variability at this NGP location (Fig. 6b,c), their joint interactions with SM explain nearly 10% of the summertime rainfall variance. Thus, HDMR reveals not only the direct "wet-soil advantage" but also key compound effects arising from the co-variability of surface temperature and atmospheric humidity.

4   Discussion and Conclusions

Soil moisture precipitation coupling (SMPC) exerts strong control on Earth's water and energy cycles, yet is difficult to observe, examine, and quantify due to synoptic effects and the complex and intricate web of land-atmosphere variables involved. Correlation analysis and regression-based methods have become the standard in analyzing SMPCs but are unable to adequately characterize nonlinear and cooperative (compound) effects. In this paper, we introduced functional data decomposition by means of high-dimensional model representation (HDMR) to delineate the structural, correlative, and cooperative contributions of key land-atmosphere variables to precipitation. HDMR [40] is a generalization of Sobol's functional decomposition [42] to dependent (correlated) input variables. Both HDMR and Sobol's method trace their roots to the analysis of variance (ANOVA), which decomposes output variability into contributions from individual inputs and their interactions. While Sobol's approach assumes input independence, HDMR relaxes this assumption, making it suitable for more complex, real-world datasets with correlated drivers.

A synthetic benchmark experiment with a simple bivariate function showed that HDMR does not suffer the limitations of commonly used correlation and regression analysis methods in distinguishing between direct and indirect (correlative) effects. Data analysis of gross primary productivity (GPP) from a light-use-efficiency model demonstrated that commonly used regression methods substantially underestimate the temperature contribution to GPP, while HDMR accurately captured this effect. Next, continental-scale analysis of CONUS404 reanalysis data confirmed the presence of SMPC hot spots in the central and northern Great Plains. Our regional SMPC characterization aligns closely with earlier modeling studies [34, 35], however, the HDMR-based functional decomposition can explain up to 40% of the variance in summertime afternoon rainfall. This is a substantially higher fraction than the $\sim 16\%$ reported both by Koster et al. [35] and by our linear regression analysis. This difference in the SM coupling strength is attributable to three main factors:

(i) We use CONUS404 – a 4 km WRF-based reanalysis data product – which explicitly resolves mesoscale convective and boundary-layer processes, preserving the pathways through which SM influences precipitation.

(ii) HDMR is a covariance-based global sensitivity analysis method that allows all land–atmosphere variables to vary jointly, instead of varying one at a time. This yields physically consistent estimates of coupling strength.

(iii) HDMR enforces hierarchical orthogonality among component functions, allowing it to retain and disentangle structural, correlative, and cooperative effects. Such signals are often missed by simpler data analysis methods.

HDMR results further demonstrate that on summer days with afternoon rainfall (12-hour totals of 4.7 to 8.2 mm), first-order SM effects can increase precipitation by up to 8 mm when soils are sufficiently wet. Second-order effects associated with warm and moist conditions can potentially contribute another 3 mm.

Distinct roles of land surface temperature (LST) and precipitable water (PW) emerged in mountainous and coastal regions, where surface heating and atmospheric moisture supply usually outweigh local SM effects. While both LST and PW contribute moderately to afternoon precipitation in areas such as the Rocky Mountains and Gulf Coast, their cooperative interactions with SM are particularly variable over the northern Great Plains, highlighting the dynamic nature of land–atmosphere coupling. Collectively, the HDMR results provide a unified framework for understanding how morning SM competes with thermal and moisture conditions in shaping afternoon precipitation. This holistic perspective is particularly valuable in transitional climate zones, where SM variability can exert substantial control over surface energy partitioning and rainfall intensity.

As with any higher-order polynomial expansion, the HDMR functional decomposition requires sub-

stantial data resources. To ensure robust estimates of coupling strength, the number of data samples N must exceed sufficiently the number of expansion coefficients l [41]. This condition will be met with simulated or satellite-based datasets, but HDMR's data requirements increase rapidly with the number of input variables d and order of decomposition, potentially requiring sparse basis representations or regularization techniques. It is also important to note that, although the polynomial component functions in HDMR can capture nonlinear relationships, they are empirical representations of observed (or simulated) input–output behavior rather than explicit mechanistic models. As such, the causal pathways inferred through HDMR must ultimately be validated using process-based modeling and experimental studies.

Our SMPC analysis focuses on a specific diurnal lag, namely, 7:00 am land-atmosphere conditions on same-day afternoon precipitation. However, SMPC is not constrained to a single timescale or lag structure [56, 57, 58]. Future research can explore wetness-dependent lag dynamics to account for slower, long-memory couplings, particularly under dry conditions where delayed responses may be more pronounced.

We conclude that HDMR offers valuable new insights into SMPC processes and represents a powerful addition to the suite of data analysis tools used by hydrometeorologists and hydroclimatologists. By disentangling the direct, correlative, and cooperative contributions to precipitation variability, HDMR enables a more nuanced diagnosis of land–atmosphere coupling, potentially guiding future improvements in parameterizations of convection and boundary layer processes in weather and climate models.

5 Data, Materials, and Software Availability

5.1 CONUS404 Dataset

The publicly available CONUS404 dataset [59, 60], provides hourly hydroclimate reanalysis at a 4 km spatial resolution over the conterminous United States (CONUS) for water years 1980–2021. This data offers a realistic platform for examining how morning SM and other factors influence afternoon precipitation. We focus on warm-season diurnal SMPC processes by selecting data from April to September over the 2012–2021 period, analyzing afternoon-to-midnight (12:00–24:00 local time) precipitation. Morning conditions of SM and other land–atmosphere variables are used as antecedent states within our SMPC analysis framework. CONUS404 was generated by dynamically downscaling fifth-generation
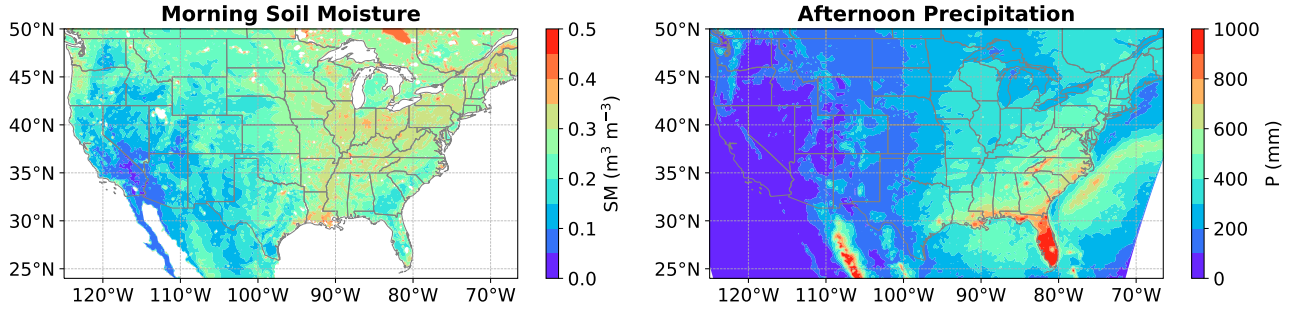

Figure 8: Ten-year time-averaged values of (a) morning (07:00 local time) soil moisture (SM) and (b) cumulative afternoon precipitation (12:00–24:00 local time) from the CONUS404 reanalysis dataset for the warm seasons (April–September) of 2012–2021. Data are upscaled from 4 km to 16 km using 4×4 averaging.

ECMWF reanalysis (ERA5) using version 3.9.1.1 of the Weather Research and Forecasting (WRF) Model. The primary physical configurations of the CONUS404 reanalysis include the Thompson microphysics scheme [61], Yonsei University planetary boundary layer scheme [62], Rapid Radiative Transfer Model for General Circulation Models [63], and Noah-MP land surface model [64]. A key advancement over its predecessor, CONUS1 [65], is the Miguez-Macho–Fan groundwater scheme [66, 67], which substantially reduces biases in surface energy fluxes and convective precipitation. Figure 8 displays the 10-year (2012–2021) mean morning soil moisture (SM) and mean afternoon precipitation (12:00–24:00 local time) during the warm season (April–September).

Table 1 list the CONUS404 variables used in this study. Soil moisture (SM), surface skin temperature (LST), precipitable water (PW), leaf area index (LAI), and wind components (u_{10} and v_{10}) are used as input variables to HDMR and used to predict afternoon precipitation P. These variables characterize key aspects of the land–atmosphere system, including the thermodynamic environment, evapotranspiration, and moisture availability, all of which influence convective triggering and precipitation formation [1, 51]. Wind speed and direction capture the potential for moisture advection, which can modulate SPMC [68]. Case studies and data preprocessing steps are discussed in the next section.

Table 1: Land-atmosphere variables of CONUS404 used in this study

Variables	Symbol	Units
Grid-scale cumulative precipitation	P	mm
Soil moisture	SM	cm^3/cm^3
Surface skin temperature	LST	K
2-m air temperature	T_{2m}	K
Precipitable water	PW	m
Leaf area index	LAI	-
U-component of wind speed at 10 m	u_{10}	m/s
V-component of wind speed at 10 m	v_{10}	m/s

5.2 Case Studies and Experimental Design

We provide the information needed for each case study so that our analyses and numerical experiments can be independently replicated. Specifically, we explain how HDMR's input data were prepared and processed. In each study, HDMR takes as input an $N \times d$ matrix $\mathbf{X}$, consisting of N samples of the input variables $x_1, \ldots, x_d$ and an $N \times 1$ vector $\mathbf{y} = (y^{(1)}, \ldots, y^{(N)})^\top$, containing the corresponding values of the target variable. Prior to executing our HDMR toolbox in MATLAB (or Python), we standardize the entries of both $\mathbf{X}$ and $\mathbf{y}$. As a result, each column of $\mathbf{X}$ (i.e., each input variable) has zero mean and unit standard deviation. This preprocessing step promotes consistency across HDMR trials and case studies.

5.2.1 Case I: A Simple Bivariate Function

To verify that HDMR recovers variance partitions under both uncorrelated and correlated inputs, we analyze the scalar-valued function

$$y = x_1 + x_2 + x_1 x_2 + e. \tag{10}$$

with $d = 2$ input variables, x_1 and x_2, which are sampled from a bivariate normal distribution, $\mathbf{x} \sim \mathcal{N}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with unit mean $\boldsymbol{\mu} = (1\ 1)^\top$ and 2×2 covariance matrix, $\boldsymbol{\Sigma}$. The noise term, $e \sim \mathcal{N}(0, \sigma_e^2)$ is small compared to $\mathrm{Var}[y]$. We draw at random two data sets, $\mathbf{X}_1$ (Exp1A) and $\mathbf{X}_2$ (Exp1B), of $N = 5,000$ samples using the covariance matrices:

$$\boldsymbol{\Sigma}_1 = \begin{bmatrix} 1.0 & 0.0 \\ 0.0 & 1.0 \end{bmatrix} \qquad \boldsymbol{\Sigma}_2 = \begin{bmatrix} 1.0 & -0.6 \\ -0.6 & 1.0 \end{bmatrix}. \tag{11}$$

The variables x_1 and x_2 are independent in data set $\mathbf{X}_1$ of Exp1A, whereas they exhibit a negative correlation of -0.6 in $\mathbf{X}_2$ of Exp1B.

The function output y is linearly dependent on x_1 and x_2 individually but also their product (i.e., the interaction term). In Exp1A with uncorrelated inputs, HDMR reduces to Sobol's method with correlative coupling indices equal to zero. Exp1B demonstrates how correlation among input variables influences the partitioning of structural, correlative, and cooperative effects.

5.2.2 Case II: Light Use Efficiency Model

The Light Use Efficiency (LUE) model simulates gross primary productivity (GPP) in units of grams of carbon per square meter per day:

$$\mathrm{GPP} = \varepsilon_0 R_n \cdot \mathrm{PAR} \cdot f(\mathrm{SM}) \cdot g(T_{2m}) \cdot h(u), \tag{12}$$

where $\varepsilon_0 = 1.8$ (gC/MJ) is the baseline (maximum) LUE, R_n (W/m²) signifies the net radiation, PAR = 0.45 is the unitless fraction of R_n that is photosynthetically active radiation and $f(\text{SM})$, $g(T_{2m})$, and $h(u)$ are dimensionless response functions of SM, 2-meter air temperature and wind speed, respectively. These functions are given below:

$$f(\text{SM}) = \begin{cases} 0 & \text{if SM} < \text{SM}_w \\ \dfrac{\text{SM} - \text{SM}_w}{\text{SM}_f - \text{SM}_w} & \text{if SM}_w \le \text{SM} < \text{SM}_f \\ 1 & \text{if SM} \ge \text{SM}_f \end{cases} \tag{13}$$

$$g(T_{2m}) = \exp[-(T_{2m} - 25)^2/100] \tag{14}$$

$$h(u) = \{1 + \exp[-0.5(u - 2)]\}^{-1}, \tag{15}$$

where $\text{SM}_w = 0.1$ and $\text{SM}_f = 0.3$ are the soil's wilting point and field capacity in units of cm³/cm³, respectively.

We use the CONUS404 reanalysis dataset and extract hourly values of SM, T_{2m}, u_{10}, and v_{10} for a grid cell located at $38.80°\text{N}, 97.15°\text{W}$, in the central US. Next, we compute the wind speed $u = (u_{10}^2 + v_{10}^2)^{1/2}$ (m/s) and time series of R_n (W/m²) using Text S1. Then, we evaluate Eq. (12) and simulate a yearly dataset comprised of $N = 8{,}760$ hourly GPP values capturing diurnal to seasonal variability in ecosystem productivity. The standardized LUE model input variables make up the columns of the $N \times 4$ input matrix $\mathbf{X}$, where $x_1 = \text{SM}$, $x_2 = R_n$, $x_3 = T_{2m}$ and $x_4 = u$. The simulated GPP values are standardized and stored in an $N \times 1$ vector $\mathbf{y}$. This $\{\mathbf{X}, \mathbf{y}\}$ dataset serves as input to HDMR to estimate the structural, correlative, and interactive contributions of SM, R_n, T_{2m}, and u to plant productivity.

5.2.3 Case III: Soil Moisture-Precipitation Coupling (SMPC)

This study investigates soil moisture–precipitation coupling over the contiguous US. To address this objective, we implement the following CONUS404 data preparation steps. First, to balance computational feasibility and preserve mesoscale variability, the original 4 km CONUS404 outputs are aggregated to 16 km × 16 km grid cells. Next, we remove data of synoptic systems [2, 69], as they can obscure local soil moisture impacts. Large-scale stratiform events, for example, can yield a misleading positive correlation between antecedent SM and subsequent rainfall [70]. Therefore, we focus attention on precipitation events in the months of April-September of 2012-2021 that satisfy the following two criteria [71], (i) the start time, t_{start}, of afternoon rainfall is between 12:00-14:00 local time, (ii) in the antecedant period between $t_{\text{start}} - 24$ and $t_{\text{start}} - 1$, the grid cell is rain-free. For all grid cells that satisfy the above two criteria, we compute the 12-hr cumulative rainfall between t_{start} and $t_{\text{start}} + 11$ hr. This rainfall total, expressed in millimeters, constitutes a single entry $y^{(i)}$ in the data vector $\mathbf{y}$. The corresponding morning values at 07:00 local time of SM and the other auxiliary land–atmosphere variables listed in Table 1 serve as the explanatory variables for $y^{(i)}$. These morning conditions are stored in the ith row of input matrix $\mathbf{X}$, denotes as $\mathbf{x}^{(i)} = (x_1, \ldots, x_6) = (\text{SM}, \text{LST}, \text{PW}, \text{LAI}, u_{10}, v_{10})$.

The above data preprocessing steps reduce the number of valid rain-initiation events to only a few hundred for many grid cells across the central US. To satisfy HDMR data requirements, we combine data from 5×5 blocks of 16×16 km grid cells in $\mathbf{X}$ and $\mathbf{y}$, corresponding to aggregated areas of 80×80 km. This upscaling in spatial resolution substantially increases the number of data points N, ensuring robust estimation of the polynomial expansion coefficients in the HDMR functional decomposition. As a result, we obtain a single set of coupling metrics for each 80×80 km block. Figure 9 displays the sample size N of the HDMR input matrix $\mathbf{X}$ and the $N \times 1$ data vector $\mathbf{y}$ for each aggregated block across CONUS. Blocks with $N < 5,000$ are discarded from the HDMR analysis (e.g., the southwestern US) as the warm-season precipitation is not frequent enough to yield robust SMPC signals.
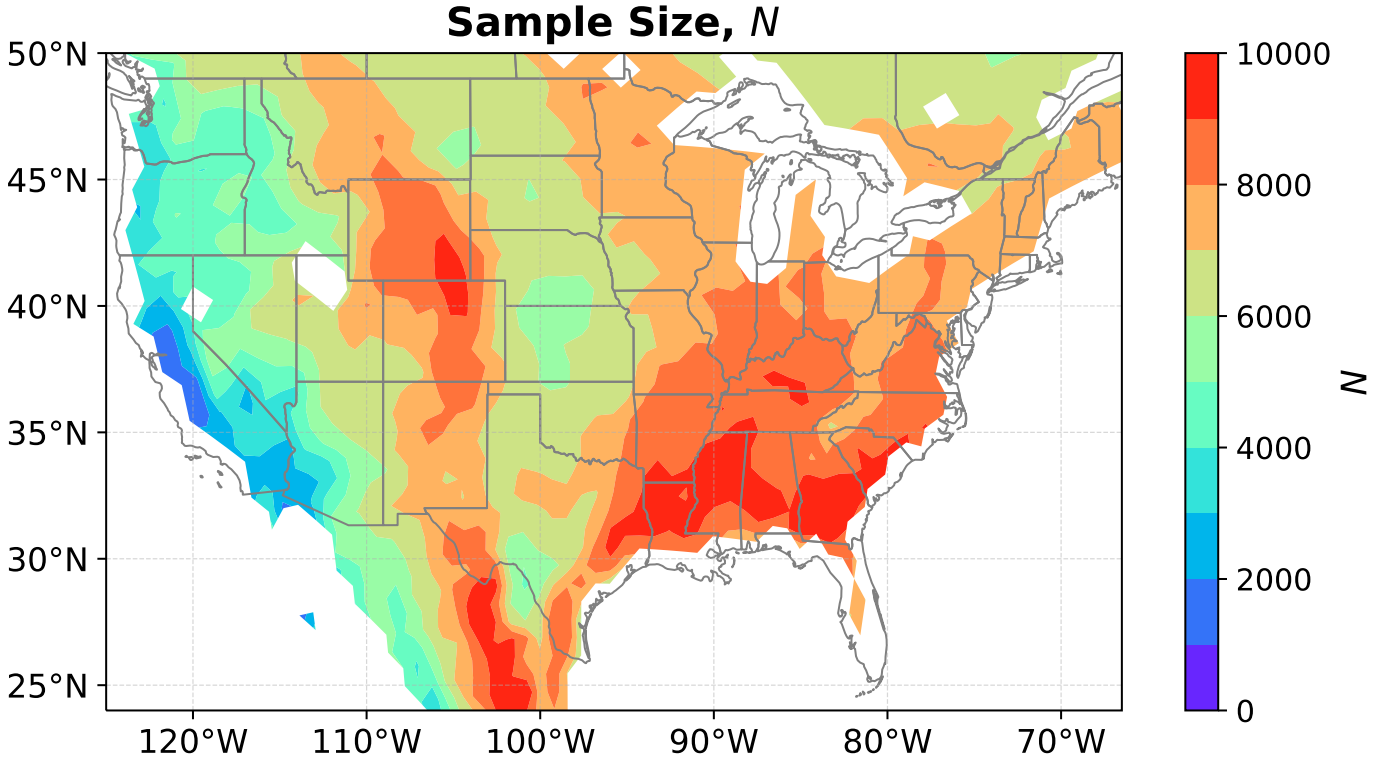
Figure 9: Contour map of sample size N used in the HDMR analysis of SMPC, based on a 10-year excerpt (2012–2021) of the CONUS404 dataset for warm-season months (April–September).

Note that we do not include net radiation (R_{n}) as an explanatory variable in our HDMR analysis because it is a byproduct of LST and other land–atmosphere variables. Including R_{n} would confound the inference of the individual and cooperative effects of LST on afternoon precipitation [51].

5.3   HDMR Method

5.3.1   Polynomial Construction of Component Functions

We follow Li and Rabitz [40] and construct each component function as a sum of linear multiples of orthonormalized polynomial functions. Specifically, for a single dimension x_i, we define a family of polynomials:

$$\phi_1(x_i) = a_1 x_i + a_0, \qquad\qquad \text{degree } p = 1 \qquad\qquad (16\text{a})$$

$$\phi_2(x_i) = b_2 x_i^2 + b_1 x_1 + b_0, \qquad\qquad \text{degree } p = 2 \qquad\qquad (16\text{b})$$

$$\phi_3(x_i) = c_3 x_i^3 + c_2 x_i^2 + c_1 x_1 + c_0 \qquad\qquad \text{degree } p = 3. \tag{16c}$$

We select $p = 3$ for our case studies. The values of coefficients a, b and c are derived from Gram-Schmidt orthonormalization. This process builds an orthonormal basis of polynomial functions on the unit interval with respect to a chosen weighting function. Each HDMR component function then appears as a linear combination of these orthonormalized polynomial functions of degrees 1 to p

$$f_i(x_i) = \sum_{r=1}^{p} \alpha_r^{(i)i} \phi_r(x_i) \tag{17a}$$

$$f_{ij}(x_i, x_j) = \sum_{r=1}^{p} \left[\alpha_r^{(ij)i} \phi_r(x_i) + \alpha_r^{(ij)j} \phi_r(x_j) \right]$$

$$+ \sum_{r=1}^{p} \sum_{s=1}^{p} \beta_{rs}^{(ij)ij} \phi_r(x_i) \phi_s(x_j), \tag{17b}$$

where the extended bases of the second-order component functions will help satisfy the vanishing condition in (3). Parenthesized superscripts in α and β indicate which component function (or functions) they belong to, while non-parenthesized superscripts refer to their position in the input vector, $\mathbf{x}$.

If we substitute (17a) and (17b) into (1) we obtain a closed-form expression for the relationship between land-atmosphere variables $\mathbf{x} = (x_1, \ldots, x_d)^\top$ and 12-hr accumulated precipitation y. This expression has $l = d \cdot p + d(d-1)(2p + p^2)/2$ unknown expansion coefficients and its different terms correspond to the marginal and cooperative effects of the input variables to precipitation. Specifically, the first-order SM component function, $f_1(x_1) = \sum_{r=1}^{p} \alpha_r^{(1)1} \phi_r(x_1)$, quantifies the sign and magnitude (in units of mm) of the SM contribution (direct effect) to precipitation.

5.3.2 D-MORPH Regression

Recall that $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(N)}$ is a $N \times d$ matrix with N different samples of the input variables $x_1, \ldots, x_d$ and $y^{(1)}, \ldots, y^{(N)}$ is a $N \times 1$ vector with corresponding values of the target variable, then we can write

(2) in matrix form $\boldsymbol{\Phi}\mathbf{c} = \mathbf{b}$ and yield

$$
\boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{\phi}(\mathbf{x}^{(1)})^\top \\ \vdots \\ \boldsymbol{\phi}(\mathbf{x}^{(N)})^\top \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} y^{(1)} - y_0 \\ \vdots \\ y^{(N)} - y_0 \end{bmatrix}, \tag{18a}
$$

where $\boldsymbol{\phi}(\mathbf{x})^\top$ is a $1 \times l$ design vector with orthonormalized polynomial functions of (17) (and products thereof) evaluated at their respective entries of $\mathbf{x}$ and arranged in appropriate order, $\mathbf{c}$ is a $l \times 1$ coefficient vector with values of α and β, and the $N \times 1$ vector $\mathbf{b}$ stores differences between measured/simulated $y^{(i)}$'s and the mean value y_0 of the training samples, $i = (1, \dots, N)$.

To offer some protection against underdetermined problems $N < l$ or a rank-deficient design matrix, we remove duplicate entries of the basis functions (i.e., $\phi_r(x_i)$) of the first- and second-order component functions. This reduced system is easier to solve in practice [40]. First, we determine the least squares values $\widehat{\mathbf{c}}_{\mathrm{ls}}$ of the expansion coefficients

$$
\widehat{\mathbf{c}}_{\mathrm{ls}} = (\boldsymbol{\Phi}^\top \boldsymbol{\Phi})^\dagger \mathbf{d}, \tag{19}
$$

where the $l \times (l - d \cdot p)$ matrix $(\boldsymbol{\Phi}^\top \boldsymbol{\Phi})^\dagger$ is the generalized pseudo inverse of the $l \times l$ Gramian matrix, $\mathbf{G} = \boldsymbol{\Phi}^\top \boldsymbol{\Phi}$, which satisfies all four Moore-Penrose conditions [72, 73] and whose redundant rows (first $d \cdot p$ rows of the first-order basis functions) are removed and $\mathbf{d}$ is the $(l - d \cdot p) \times 1$ vector $\boldsymbol{\Phi}^\top \mathbf{b}$ without the first $d \cdot p$ rows. Diffeomorphic Modulation (dm) under Observable Response Preserving Homotopy (D-MORPH) regression [39] enforces hierarchical orthogonality of the component functions in pursuit of the optimum coefficients

$$
\widehat{\mathbf{c}}_{\mathrm{dm}} = \mathbf{V}_{l-r}(\mathbf{U}_{l-r}^\top \mathbf{V}_{l-r})\mathbf{U}_{l-r}^\top \widehat{\mathbf{c}}_{\mathrm{ls}}, \tag{20}
$$

where $\mathbf{U}_{l-r}$ and $\mathbf{V}_{l-r}$ equal the last $l - r$ columns of the $l \times l$ matrices $\mathbf{U}$ and $\mathbf{V}$ determined from singular value decomposition $\mathbf{PB} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top$ of the product of a $l \times l$ projection matrix $\mathbf{P} = \mathbf{I}_l - \mathbf{G}$ and $l \times l$ constraint matrix $\mathbf{B}$ of inner products of the orthonormalized polynomials. Matrix $\mathbf{B}$ enforces the relaxed vanishing condition of (3) [39], $\mathbf{I}_l$ is the $l \times l$ identity matrix and r signifies the number of

nonzero singular values.

Our HDMR implementation uses bootstrapping to quantify the uncertainty of the inferred component functions and coupling indices [41]. However, the resulting bootstrap confidence intervals are consistently very narrow, and therefore we do not present them in this paper. This is primarily due to the relatively high ratio of the number of samples N in each 16×16 km grid cell to the number of expansion coefficients l. For example, with $d = 6$ explanatory variables, the third-order $p = 3$ polynomial expansion in our HDMR analysis has $l = d \cdot p + d(d-1)(2p+p^2)/2 = 243$ expansion coefficients. As we only use grid cells with $N > 5,000$, the ratio of N and l exceeds 20, and the confidence intervals of the coupling indices are negligibly small.

5.4   Other Methods

We benchmark the HDMR results against two other commonly used methods. For each column (input variable) of matrix $\mathbf{X}$ we compute Pearson's linear correlation coefficient, $r_{x_i,y}$, between x_i and y [74]

$$r_{x_i,y} = \frac{\sum_{j=1}^{N}(x_i^{(j)} - m_{x_i})(y^{(j)} - m_y)}{\sqrt{\sum_{j=1}^{N}(x_i^{(j)} - m_{x_i})^2}\sqrt{\sum_{j=1}^{N}(y^{(j)} - m_y)^2}}, \tag{21}$$

where m_{x_i} is the sample mean of x_i and $i = 1, \ldots, d$. Then, we also use multivariate linear regression

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_d x_d + \epsilon, \tag{22}$$

to explain the standardized target variable y from the standardized input variables $x_1, \ldots, x_d$ of matrix $\mathbf{X}$. Values of the intercept β_0 and multiplicative coefficients $\beta_1, \ldots, \beta_d$ are determined using ordinary least squares (OLS). The OLS estimator $\widehat{\boldsymbol{\beta}} = (\widehat{\beta}_0, \ldots, \widehat{\beta}_d)^\top$ is equal to $\widehat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$.

To compare the results of the two methods to the coupling indices of HDMR, we must turn the $r_{x_i,y}$'s

and β_i's into measures of the normalized explained variance:

$$R^2_{i,\mathrm{C}} = r^2_{x_i,y} \tag{23}$$

$$R^2_{i,\mathrm{LR}} = \frac{\beta_i^2 s_i^2 + \beta_i s_i \sum_{j \neq i} r_{x_i,x_j} \beta_j s_j}{\mathrm{Var}[y]}, \tag{24}$$

where $R^2_{\mathrm{C},i}$ and $R^2_{\mathrm{LR},i}$ are the explained variances of y by x_i according to correlation analysis and multivariate linear regression, respectively, r_{ij} is the correlation coefficient of x_i and x_j, and s_i and s_j denote the sample standard deviations of x_i and x_j, respectively. This unifies the output of correlation analysis and linear regression with the covariance-based coupling indices of HDMR, allowing direct comparisons.

Data, Materials, and Software Availability

The CONUS404 dataset is available from the National Center for Atmospheric Research (NCAR) Research Data Archive, as described by Rasmussen et al. [60]. MATLAB and Python implementations of the HDMR toolbox can be downloaded from `https://github.com/jaspervrugt/HDMR_EXT`. Post-processing software will be archived on Zenodo along with the final version of the CONUS404-derived dataset.

Acknowledgements

Samueli School of Engineering, UC Irvine.

References

[1]   Sonia I. Seneviratne, Thierry Corti, Edouard L. Davin, Martin Hirschi, Eric B. Jaeger, Irene Lehner, Boris Orlowsky, and Adriaan J. Teuling. "Investigating soil moisture–climate interactions in a changing climate: A review". In: *Earth-Science Reviews* 99.3 (2010), pp. 125–161. ISSN: 0012-8252. DOI: `10.1016/j.earscirev.2010.02.004`. URL: `https://www.sciencedirect.com/science/article/pii/S0012825210000139`.

[2]   K. L. Findell, P. Gentine, B. R. Lintner, and C. Kerr. "Probability of afternoon precipitation in eastern United States and Mexico enhanced by high evaporation". In: *Nature Geoscience* 4.7 (2011), pp. 434–439. DOI: `10.1038/ngeo1174`.

[3]   M. B. Ek and A. A. M. Holtslag. "Influence of soil moisture on boundary layer cloud development". In: *Journal of hydrometeorology* 5.1 (2004), pp. 86–99. DOI: `10.1175/1525-7541(2004)005<0086:IOSMOB>2.0.CO;2`.

[4]   T. W. Ford, J. Steiner, B. Mason, and S. M. Quiring. "Observation-Driven Characterization of Soil Moisture-Precipitation Interactions in the Central United States". In: *Journal of Geophysical Research: Atmospheres* 128.12 (2023), e2022JD037934. DOI: `10.1029/2022JD037934`.

[5]   C. M. Taylor, A. Gounou, F. Guichard, P. P. Harris, R. J. Ellis, F. Couvreux, and M. De Kauwe. "Frequency of Sahelian storm initiation enhanced over mesoscale soil-moisture patterns". In: *Nature Geoscience* 4.7 (2011), pp. 430–433. DOI: `10.1038/ngeo1173`.

[6]   P. A. Levine, J. T. Randerson, S. C. Swenson, and D. M. Lawrence. "Evaluating the strength of the land–atmosphere moisture feedback in Earth system models using satellite observations". In:

Hydrology and Earth System Sciences 20.12 (2016), pp. 4837–4856. DOI: 10.5194/hess-20-4837-2016.

[7]   X. Dong, H. Chen, Y. Zhou, P. Hsu, and W. Zhang. "Disentangling the Complexities of How Underlying Surface Thermal Factors Influence July Precipitation in Eastern China". In: *Journal of Climate* 37.19 (2024), pp. 5105–5129. DOI: 10.1175/JCLI-D-23-0748.1.

[8]   E. Bevacqua, O. Rakovec, D. L. Schumacher, R. Kumar, S. Thober, L. Samaniego, S. I. Seneviratne, and J. Zscheischler. "Direct and lagged climate change effects intensified the 2022 European drought". In: *Nature Geoscience* (2024), pp. 1–8. DOI: 10.1038/s41561-024-01559-2.

[9]   H. M. Duerinck, R. J. Van der Ent, N. C. Van de Giesen, G. Schoups, V. Babovic, and P. J. F. Yeh. "Observed soil moisture–precipitation feedback in Illinois: A systematic analysis over different scales". In: *Journal of Hydrometeorology* 17.6 (2016), pp. 1645–1660. DOI: 10.1175/JHM-D-15-0032.1.

[10] W. Liu, Q. Zhang, C. Li, L. Xu, and W. Xiao. "The influence of soil moisture on convective activity: a review". In: *Theoretical and Applied Climatology* 149.1-2 (2022), pp. 221–232. DOI: 10.1007/s00704-022-04046-z.

[11] Z. Guo, P. A. Dirmeyer, R. D. Koster, Y. C. Sud, G. Bonan, K. W. Oleson, E. Chan, D. Verseghy, P. Cox, C. T. Gordon, et al. "GLACE: the global land–atmosphere coupling experiment. Part II: analysis". In: *Journal of Hydrometeorology* 7.4 (2006), pp. 611–625. DOI: 10.1175/JHM511.1.

[12] M.I. Budyko. *Climate and Life.* Vol. 18. International Geophysics. Academic Press, 1974.

[13] Elfatih A. B. Eltahir. "A soil moisture–rainfall feedback Mechanism: 1. Theory and observations". In: *Water Resources Research* 34.4 (1998), pp. 765–776. DOI: 10.1029/97WR03499. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/97WR03499. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/97WR03499.

[14] C. M. Taylor. "Detecting soil moisture impacts on convective initiation in Europe". In: *Geophysical Research Letters* 42.11 (2015), pp. 4631–4638. DOI: 10.1002/2015GL064030.

[15] K. L. Findell and E. A. B. Eltahir. "Atmospheric controls on soil moisture–boundary layer interactions. Part I: Framework development". In: *Journal of Hydrometeorology* 4 (2003a), pp. 552–569. DOI: 10.1175/1525-7541(2003)004<0552:ACOSML>2.0.CO;2.

[16] C. M. Taylor, R. A. M. de Jeu, . Guichard, P. P. Harris, and W. A. Dorigo. "Afternoon rain more likely over drier soils". In: *Nature* 489.7416 (2012), pp. 423–426. DOI: 10.1038/nature11377.

[17] T. W. Ford, S. M. Quiring, O. W. Frauenfeld, and A. D. Rapp. "Synoptic conditions related to soil moisture-atmosphere interactions and unorganized convection in Oklahoma". In: *Journal of Geophysical Research: Atmospheres* 120.22 (2015), pp. 11–519.

[18] S. Cropper, K. Solander, B. D. Newman, O. A. Tuinenburg, A. Staal, J. J. E. Theeuwen, and C. Xu. "Comparing deuterium excess to large-scale precipitation recycling models in the tropics". In: *npj Climate and Atmospheric Science* 4.1 (2021), p. 60. DOI: 10.1038/s41612-021-00217-3.

[19] E. D. Rappin, R. Mahmood, U. S. Nair, and R. A. Pielke Sr. "Land–atmosphere interactions during GRAINEX: planetary boundary layer evolution in the presence of irrigation". In: *Journal of Hydrometeorology* 23.9 (2022), pp. 1401–1417. DOI: 10.1175/JHM-D-21-0160.1.

[20] B. P. Guillod, B. Orlowsky, D. G. Miralles, A. J. Teuling, and S. I. Seneviratne. "Reconciling spatial and temporal soil moisture effects on afternoon rainfall". In: *Nature communications* 6.1 (2015), p. 6443. DOI: 10.1038/ncomms7443.

[21] C. M. Taylor, C. E. Birch, D. J. Parker, N. Dixon, F. Guichard, G. Nikulin, and G. M. S. Lister. "Modeling soil moisture-precipitation feedback in the Sahel: Importance of spatial scale versus convective parameterization". In: *Geophysical Research Letters* 40.23 (2013), pp. 6213–6218. DOI: 10.1002/2013GL058511.

[22] S. Yuan, Y. Wang, S. M. Quiring, T. W. Ford, and A. L. Houston. "A sensitivity study on the response of convection initiation to in situ soil moisture in the central United States". In: *Climate Dynamics* 54 (2020), pp. 2013–2028. DOI: `10.1007/s00382-019-05098-0`.

[23] J Welty and X Zeng. "Does soil moisture affect warm season precipitation over the southern Great Plains?" In: *Geophysical Research Letters* 45.15 (2018), pp. 7866–7873. DOI: `10.1029/2018GL078598`.

[24] H. Hu, L. R. Leung, and Z. Feng. "Early warm-season mesoscale convective systems dominate soil moisture–precipitation feedback for summer rainfall in central United States". In: *Proceedings of the National Academy of Sciences* 118.43 (2021), e2105260118. DOI: `10.1073/pnas.2105260118`.

[25] H. Chen, W. Zhan, B. Zhou, F. Teng, J. Zhang, and Y. Zhou. "Impact of nonuniform land surface warming on summer anomalous extratropical cyclone activity over East Asia". In: *Journal of Geophysical Research: Atmospheres* 124.19 (2019), pp. 10306–10320. DOI: `10.1029/2018JD030165`.

[26] S. Zhou, A. P. Williams, B. R. Lintner, K. L. Findell, T. F. Keenan, Y. Zhang, and P. Gentine. "Diminishing seasonality of subtropical water availability in a warmer world dominated by soil moisture–atmosphere feedbacks". In: *Nature communications* 13.1 (2022), p. 5756. DOI: `10.1038/s41467-022-33473-9`.

[27] G. Wang, R. Fu, Y. Zhuang, P. A. Dirmeyer, J. A. Santanello, G. Wang, K. Yang, and K. McColl. "Influence of lower-tropospheric moisture on local soil moisture–precipitation feedback over the US Southern Great Plains". In: *Atmospheric Chemistry and Physics* 24.6 (2024), pp. 3857–3868. DOI: `10.5194/acp-24-3857-2024`.

[28] H. Su, Z. Yang, R. E. Dickinson, and J. Wei. "Spring soil moisture-precipitation feedback in the Southern Great Plains: How is it related to large-scale atmospheric conditions?" In: *Geophysical Research Letters* 41.4 (2014), pp. 1283–1289. DOI: `10.1002/2013GL058931`.

[29] S. Tuttle and G. Salvucci. "Empirical evidence of contrasting soil moisture–precipitation feedbacks across the United States". In: *Science* 352 (2016), pp. 825–828.

[30] Lu Li, Wei Shangguan, Yi Deng, Jiafu Mao, JinJing Pan, Nan Wei, Hua Yuan, Shupeng Zhang, Yonggen Zhang, and Yongjiu Dai. "A causal inference model based on random forests to identify the effect of soil moisture on precipitation". In: *Journal of Hydrometeorology* 21.5 (2020), pp. 1115–1131. DOI: `10.1175/JHM-D-19-0209.1`.

[31] J. D. Fast, L. K. Berg, Z. Feng, F. Mei, R. Newsom, K. Sakaguchi, and H. Xiao. "The impact of variable land-atmosphere coupling on convective cloud populations observed during the 2016 HI-SCALE field campaign". In: *Journal of Advances in Modeling Earth Systems* 11.8 (2019), pp. 2629–2654. DOI: `10.1029/2019MS001727`.

[32] J. Wei, J. Zhao, H. Chen, and X. Liang. "Coupling between land surface fluxes and lifting condensation level: mechanisms and sensitivity to model physics parameterizations". In: *Journal of Geophysical Research: Atmospheres* 126.5 (2021), e2020JD034313. DOI: `10.1029/2020JD034313`.

[33] S. Zhou, A. P. Williams, B. R. Lintner, A. M. Berg, Y. Zhang, T. F. Keenan, B. I. Cook, S. Hagemann, S. I. Seneviratne, and P. Gentine. "Soil moisture–atmosphere feedbacks mitigate declining water availability in drylands". In: *Nature Climate Change* 11.1 (2021), pp. 38–44. DOI: `10.1038/s41558-020-00945-z`.

[34] Randal D. Koster, Paul A. Dirmeyer, Zhichang Guo, Gordon Bonan, Edmond Chan, Peter Cox, C. T. Gordon, Shinjiro Kanae, Eva Kowalczyk, David Lawrence, Ping Liu, Cheng-Hsuan Lu, Sergey Malyshev, Bryant McAvaney, Ken Mitchell, David Mocko, Taikan Oki, Keith Oleson, Andrew Pitman, Y. C. Sud, Christopher M. Taylor, Diana Verseghy, Ratko Vasic, Yongkang Xue, and Tomohito Yamada. "Regions of strong coupling between soil moisture and precipitation". In: *Science* 305.5687 (2004), pp. 1138–1140. DOI: `10.1126/science.1100217`. eprint: `https:`

//www.science.org/doi/pdf/10.1126/science.1100217. URL: https://www.science.org/doi/abs/10.1126/science.1100217.

[35] R. D. Koster, Y. C. Sud, Z. Guo, P. A. Dirmeyer, G. Bonan, K. W. Oleson, E. Chan, D. Verseghy, P. Cox, H. Davies, et al. "GLACE: the global land–atmosphere coupling experiment. Part I: overview". In: *Journal of Hydrometeorology* 7.4 (2006), pp. 590–610. DOI: 10.1175/JHM510.1.

[36] Y. Huang, S. T. Siems, M. J. Manton, and G. Thompson. "An evaluation of WRF simulations of clouds over the Southern Ocean with A-Train observations". In: *Monthly Weather Review* 142.2 (2014), pp. 647–667. DOI: 10.1175/MWR-D-13-00128.1.

[37] G. L. Stephens, T. L'Ecuyer, R. Forbes, A. Gettelmen, J. C. Golaz, A. Bodas-Salcedo, K. Suzuki, P. Gabriel, and J. Haynes. "Dreary state of precipitation in global models". In: *Journal of Geophysical Research: Atmospheres* 115.D24 (2010). DOI: 10.1029/2010JD014532.

[38] R. Heinze, A. Dipankar, C. C. Henken, C. Moseley, O. Sourdeval, S. Trömel, X. Xie, P. Adamidis, F. Ament, H. Baars, et al. "Large-eddy simulations over Germany using ICON: A comprehensive evaluation". In: *Quarterly Journal of the Royal Meteorological Society* 143.702 (2017), pp. 69–100. DOI: 10.1002/qj.2947. eprint: https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.2947. URL: https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.2947.

[39] G. Li and H. Rabitz. "D-MORPH regression: application to modeling with unknown parameters more than observation data". In: *Journal of Mathematical Chemistry* 48 (2010), pp. 1010–1035. DOI: 10.1007/s10910-010-9722-2.

[40] G. Li and H. Rabitz. "General formulation of HDMR component functions with independent and correlated variables". In: *Journal of Mathematical Chemistry* 50.1 (2012), pp. 99–130. DOI: 10.1007/s10910-011-9898-0.

[41] Y. Gao, A. Sahin, and J. A. Vrugt. "Probabilistic sensitivity analysis with dependent variables: Covariance-based decomposition of hydrologic models". In: *Water Resources Research* 59.4 (2023),

e2022WR032834. DOI: `10.1029/2022WR032834`. eprint: `https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2022WR032834`. URL: `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2022WR032834`.

[42]   I. M. Sobol′. "Sensitivity estimates for nonlinear mathematical models". In: *Mathematical Modelling and Computational Experiment* 1.4 (1993), pp. 407–414.

[43]   H. Rabitz and Ö. F. Ali ̧s. "General foundations of high-dimensional model representations". In: *Journal of Mathematical Chemistry* 25.2 (1999), pp. 197–233. DOI: `10.1023/A:1019188517934`.

[44]   G. Hooker. "Generalized functional ANOVA diagnostics for high-dimensional functions of dependent variables". In: *Journal of Computational and Graphical Statistics* 16.3 (2007), pp. 709–732. DOI: `10.1198/106186007X237892`.

[45]   G. Li, H. Rabitz, P. E. Yelvington, O. O. Oluwole, F. Bacon, C. E. Kolb, and J. Schoendorf. "Global sensitivity analysis for systems with independent and/or correlated inputs". In: *The Journal of Physical Chemistry A* 114.19 (2010), pp. 6022–6032. DOI: `https://doi.org/10.1021/jp9096919`.

[46]   T. Homma and A. Saltelli. "Importance measures in global sensitivity analysis of nonlinear models". In: *Reliability Engineering & System Safety* 52.1 (1996), pp. 1–17. DOI: `10.1016/0951-8320(96)00002-6`.

[47]   Virgil Griffith and Christof Koch. "Quantifying Synergistic Mutual Information". In: *Guided Self-Organization: Inception.* Ed. by Mikhail Prokopenko. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014, pp. 159–190. ISBN: 978-3-642-53734-9. DOI: `10.1007/978-3-642-53734-9_6`. URL: `https://doi.org/10.1007/978-3-642-53734-9_6`.

[48]   A. E. Goodwell and P. Kumar. "Temporal information partitioning: Characterizing synergy, uniqueness, and redundancy in interacting environmental variables". In: *Water Resources Research* 53.7 (2017), pp. 5920–5942. DOI: `10.1002/2016WR020216`.

[49] C. Zhang, Q. Li, X. Liang, L. Dong, B. Xie, W. Li, and C. Sun. "Impacts of land surface processes on summer extreme precipitation in Eastern China: Insights from CWRF simulations". In: *Atmospheric Research* 314 (2025), p. 107783. DOI: 10.1016/j.atmosres.2024.107783.

[50] Y. Wang and S. M. Quiring. "Impact of soil moisture initializations on WRF-simulated North American monsoon system". In: *Journal of Geophysical Research: Atmospheres* 126.4 (2021), e2020JD033858. DOI: 10.1029/2020JD033858.

[51] Y. Gao, C. Guilloteau, E. Foufoula-Georgiou, C. Xu, X. Sun, and J. A. Vrugt. "Soil moisture-cloud-precipitation feedback in the lower atmosphere from functional decomposition of satellite observations". In: *Geophysical Research Letters* 51.22 (2024), e2024GL110347. DOI: 10.1029/2024GL110347.

[52] J. Yin, J. D. Albertson, J. R. Rigby, and A. Porporato. "Land and atmospheric controls on initiation and intensity of moist convection: CAPE dynamics and LCL crossings". In: *Water Resources Research* 51.10 (2015), pp. 8476–8493. DOI: 10.1002/2015WR017286.

[53] K. E. Kunkel, S. E. Stevens, L. E. Stevens, and T. R. Karl. "Observed climatological relationships of extreme daily precipitation events with precipitable water and vertical velocity in the contiguous United States". In: *Geophysical Research Letters* 47.12 (2020), e2019GL086721. DOI: 10.1029/2019GL086721.

[54] K. L. Findell and E. A. B. Eltahir. "Atmospheric controls on soil moisture–boundary layer interactions. Part II: Feedbacks within the continental United States". In: *Journal of Hydrometeorology* 4.3 (2003b), pp. 570–583. DOI: 10.1175/1525-7541(2003)004<0570:ACOSML>2.0.CO;2.

[55] Y. Cheng, P. W. Chan, X. Wei, Z. Hu, Z. Kuang, and K. A. McColl. "Soil moisture control of precipitation reevaporation over a heterogeneous land surface". In: *Journal of the Atmospheric Sciences* 78.10 (2021), pp. 3369–3383. DOI: 10.1175/JAS-D-21-0059.1.

[56] T. Stacke and S. Hagemann. "Lifetime of soil moisture perturbations in a coupled land–atmosphere simulation". In: *Earth System Dynamics* 7.1 (2016), pp. 1–19. DOI: 10.5194/esd-7-1-2016.

[57] Y. Wang, J. Yang, Y. Chen, P. De Maeyer, Z. Li, and W. Duan. "Detecting the causal effect of soil moisture on precipitation using convergent cross mapping". In: *Scientific reports* 8.1 (2018), p. 12171. DOI: 10.1038/s41598-018-30669-2.

[58] C. M. Taylor, C. Klein, and B. L. Harris. "Multiday soil moisture persistence and atmospheric predictability resulting from sahelian mesoscale convective systems". In: *Geophysical Research Letters* 51.20 (2024), e2024GL109709. DOI: 10.1029/2024GL109709.

[59] R. M. Rasmussen, F. Chen, C. H. Liu, K. Ikeda, A. Prein, J. Kim, T. Schneider, A. Dai, D. Gochis, A. Dugger, et al. "CONUS404: The NCAR–USGS 4-km long-term regional hydroclimate reanalysis over the CONUS". In: *Bulletin of the American Meteorological Society* 104.8 (2023a), E1382–E1408. DOI: 10.1175/BAMS-D-21-0326.1.

[60] R. M. Rasmussen, F. Chen, C. Liu, K. Ikeda, A. Prein, J. Kim, T. Schneider, A. Dai, D. Gochis, A. Dugger, Y. Zhang, A. Jaye, J. Dudhia, C. He, M. Harrold, L. Xue, S. Chen, A. Newman, E. Dougherty, R. Abolafia-Rozenzweig, N. Lybarger, R. Viger, K. Dunne, K. Rasmussen, and G Miguez-Macho. *CONUS404: Four-kilometer long-term regional hydroclimate reanalysis over the conterminous United States (ver. 2.0, December 2023): U.S. Geological Survey data release [Dataset].* 2023b. DOI: 10.5066/P9PHPK4F.

[61] G. Thompson and T. Eidhammer. "A study of aerosol impacts on clouds and precipitation development in a large winter cyclone". In: *Journal of the atmospheric sciences* 71.10 (2014), pp. 3636–3658. DOI: 10.1175/JAS-D-13-0305.1.

[62] S. Y. Hong, Y. Noh, and J. Dudhia. "A new vertical diffusion package with an explicit treatment of entrainment processes". In: *Monthly weather review* 134.9 (2006), pp. 2318–2341. DOI: 10.1175/MWR3199.1.

[63] M. J. Iacono, J. S. Delamere, E. J. Mlawer, M. W. Shephard, S. A. Clough, and W. D. Collins. "Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models". In: *Journal of Geophysical Research: Atmospheres* 113.D13 (2008). DOI: 10.1029/2008JD009944.

[64] G. Y. Niu, Z. L. Yang, K. E. Mitchell, Fei. Chen, M. B. Ek, M. Barlage, A. Kumar, K. Manning, D. Niyogi, E. Rosero, et al. "The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements". In: *Journal of Geophysical Research: Atmospheres* 116.D12 (2011). DOI: 10.1029/2010JD015139.

[65] C. Liu, K. Ikeda, R. Rasmussen, M. Barlage, A. J. Newman, A. F. Prein, F. Chen, L. Chen, M. Clark, A. Dai, et al. "Continental-scale convection-permitting modeling of the current and future climate of North America". In: *Climate Dynamics* 49 (2017), pp. 71–95. DOI: 10.1007/s00382-016-3327-9.

[66] G. Miguez-Macho, Y. Fan, C. P. Weaver, R. Walko, and A. Robock. "Incorporating water table dynamics in climate modeling: 2. Formulation, validation, and soil moisture simulation". In: *Journal of Geophysical Research: Atmospheres* 112.D13 (2007). DOI: 10.1029/2006JD008112.

[67] M. Barlage, F. Chen, R. Rasmussen, Z. Zhang, and G. Miguez-Macho. "The importance of scale-dependent groundwater processes in land-atmosphere interactions over the central United States". In: *Geophysical Research Letters* 48.5 (2021), e2020GL092171. DOI: 10.1029/2020GL092171.

[68] P. Ji and X. Yuan. "Spatially varying effect of soil moisture-atmosphere feedback on spring streamflow under future warming in China". In: *Communications Earth & Environment* 5.1 (2024), p. 518. DOI: 10.1038/s43247-024-01701-3.

[69] S. E. Tuttle and G. D. Salvucci. "Confounding factors in determining causal soil moisture-precipitation feedback". In: *Water Resources Research* 53.7 (2017), pp. 5531–5544. DOI: 10.1002/2016WR019869.

[70] L. Alfieri, P. Claps, P. D'Odorico, F. Laio, and T. M. Over. "An analysis of the soil moisture feedback on convective and stratiform precipitation". In: *Journal of Hydrometeorology* 9.2 (2008), pp. 280–291. DOI: `10.1175/2007JHM863.1`.

[71] I. N. Williams. "Evaluating soil moisture feedback on convective triggering: Roles of convective and land-model parameterizations". In: *Journal of Geophysical Research: Atmospheres* 124.1 (2019), pp. 317–332. DOI: `10.1029/2018JD029326`.

[72] R. Penrose. "A generalized inverse for matrices". In: *Proceedings of the Cambridge Philosophical Society* 51.3 (1955), pp. 406–413. DOI: `10.1017/S0305004100030401`.

[73] G. H. Golub and C. F. Van Loan. *Matrix computations (3rd ed.)* Baltimore: Johns Hopkins, 1996, pp. 257–258. ISBN: 978-0-8018-5414-9.

[74] Karl Pearson. "Note on Regression and Inheritance in the Case of Two Parents". In: *Proceedings of the Royal Society of London* 58 (1895), pp. 240–242. DOI: `10.1098/rspl.1895.0041`. URL: `https://doi.org/10.1098/rspl.1895.0041`.

[75] L. Zotarelli, M. D. Dukes, C. C. Romero, K. W. Migliaccio, and K. T. Morgan. "Step by step calculation of the Penman-Monteith Evapotranspiration (FAO-56 Method)". In: *Institute of Food and Agricultural Sciences. University of Florida* 8 (2010).

Supporting Information

Text S1. Calculation of Net Radiation from CONUS404 Datasets

Net radiation, R_n [W/m^2], can be derived from five inputs in the CONUS404 dataset: the downward shortwave flux, R_s^{in} [W/m^2], the downward longwave flux, R_l^{in} [W/m^2], the surface emissivity, ε_s [-], the land-surface albedo, α [-], and the land-surface temperature, LST [K]. Using the Stefan–Boltzmann constant $\sigma = 5.67 \times 10^{-8}\,\mathrm{W\,m^{-2}\,K^{-4}}$, we express the outgoing longwave flux via the standard blackbody emission formula

$$R_l^{out} = \sigma \varepsilon_s\, LST^4. \tag{25}$$

Following Zotarelli et al. [75], the net radiation then becomes

$$R_n = \left(R_s^{in} - \alpha R_s^{in}\right) - \left(R_l^{in} - R_l^{out}\right). \tag{26}$$

In other words, this formula accounts for the net shortwave term $\left(R_s^{in}\right.$ minus its reflection due to α and the net longwave term R_l^{in} minus the outgoing R_l^{out}, yielding a physically grounded estimate of R_n. We use this approach for the second case study in the main text, where net radiation is one of the meteorological drivers in the simplified Light Use Efficiency (LUE) model.
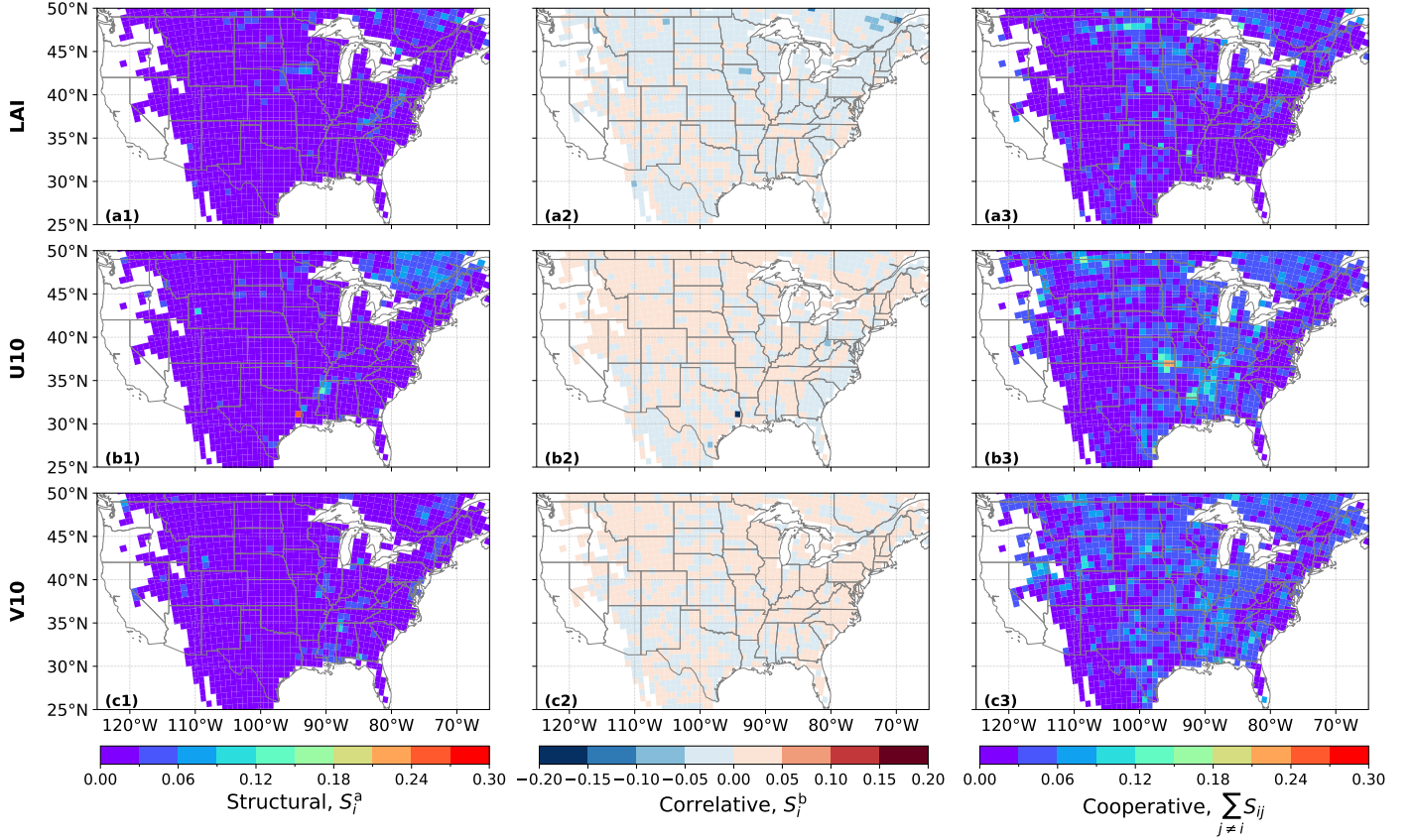
Figure S1: Partition of (a) LAI, (b) $\mathbf{u}_{10}$, and (c) $\mathbf{v}_{10}$ into structural, correlative, and cooperative contributions. Each panel (e.g., a1–a3) parallels Figure 8 of the main text, but for variables with minimal overall impact on precipitation. Only grid blocks with at least 5,000 valid rain-initiation events are shown (same for the next two figures).
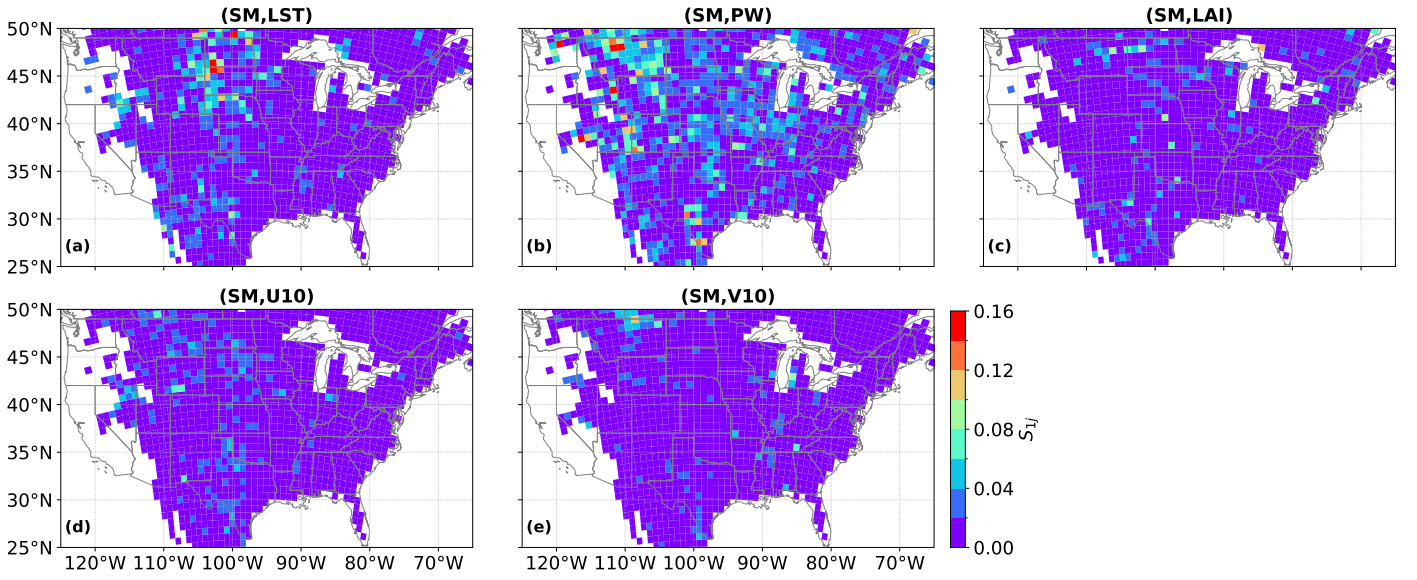
Figure S2: Total coupling indices S_{1j} showing cooperative effects between SM and each other variable. Panels (a)–(e) respectively depict the SM–LST, SM–PW, SM–LAI, SM–$\mathbf{u}_{10}$, and SM–$\mathbf{v}_{10}$ cooperative contributions. These maps indicate where morning SM co-varies with surface temperature, atmospheric moisture, vegetation, or wind fields to influence afternoon precipitation beyond their individual (first-order) roles.
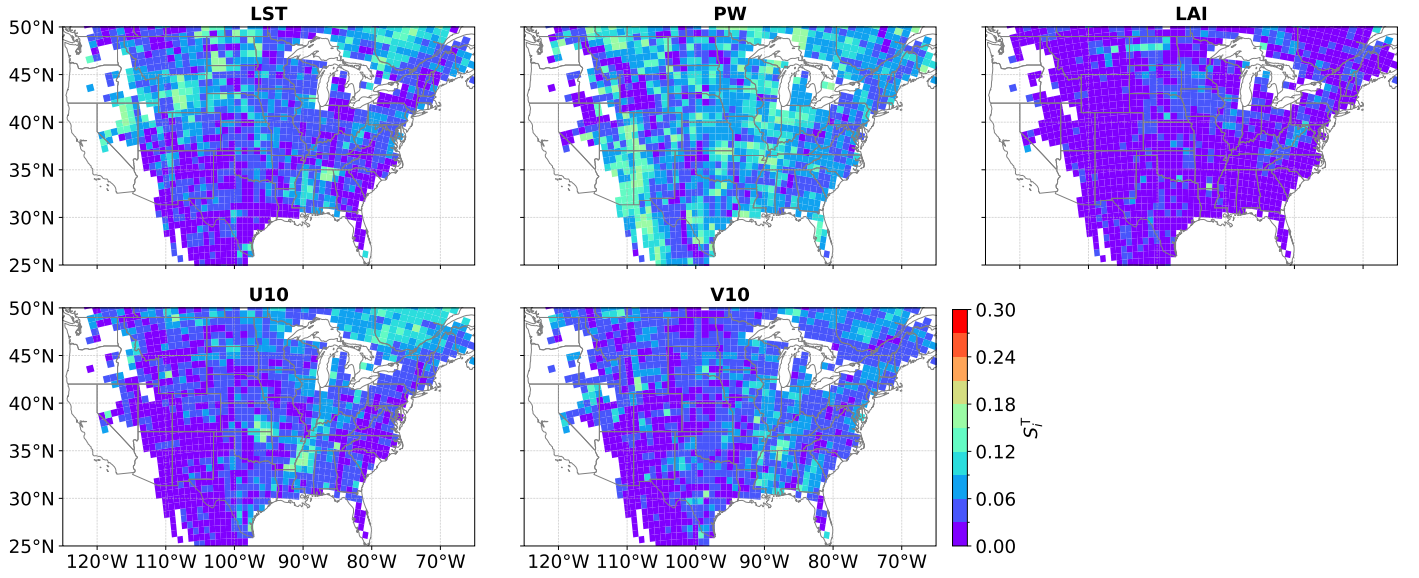
Figure S3: Total fractional variance of afternoon precipitation explained by each non-SM variable across the CONUS. Panels (a)–(e) present the HDMR-derived total effect of (a) LST, (b) PW, (c) LAI, (d) $\mathbf{u}_{10}$, and (e) $\mathbf{v}_{10}$, similar to Figure 7 of the main text (which focuses on SM).