AGU ADVANCING EARTH AND SPACE SCIENCES

JGR Atmospheres

RESEARCH ARTICLE
10.1029/2026JD046616

Special Collection:
Land-atmosphere coupling: measurement, modelling and analysis

Key Points:
- High Dimensional Model Representation disentangles structural, correlative, and cooperative effects in soil moisture-precipitation coupling
- Observations reveal wet-soil coupling in the eastern Great Plains and dry-soil, thermally driven coupling over southern Oklahoma
- Convection-permitting simulations reproduce the spatial sign of coupling but systematically underestimate the coupling strength

Supporting Information:
Supporting Information may be found in the online version of this article.

Correspondence to:
Y. Gao,
gaoyifu@nuist.edu.cn

Citation:
Gao, Y., Li, R., Chen, H., Foufoula-Georgiou, E., & Vrugt, J. A. (2026). Systematic underestimation of nonlinear and synergistic soil moisture-precipitation coupling in convection-permitting models. *Journal of Geophysical Research: Atmospheres*, *131*, e2026JD046616. https://doi.org/10.1029/2026JD046616

Received 6 FEB 2026
Accepted 18 AUG 2026

Author Contributions:
Conceptualization: Yifu Gao, Runze Li
Data curation: Yifu Gao, Runze Li
Formal analysis: Yifu Gao
Funding acquisition: Haishan Chen, Efi Foufoula-Georgiou, Jasper A. Vrugt
Investigation: Yifu Gao, Runze Li, Haishan Chen, Efi Foufoula-Georgiou, Jasper A. Vrugt
Methodology: Yifu Gao, Jasper A. Vrugt
Resources: Haishan Chen, Efi Foufoula-Georgiou, Jasper A. Vrugt

Systematic Underestimation of Nonlinear and Synergistic Soil Moisture-Precipitation Coupling in Convection-Permitting Models

Yifu Gao[1,2,3], Runze Li[3], Haishan Chen[1,2], Efi Foufoula-Georgiou[3,4], and Jasper A. Vrugt[3,4]

[1]State Key Laboratory of Climate System Prediction and Risk Management/Key Laboratory of Meteorological Disaster, Ministry of Education/Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters, Nanjing University of Information Science and Technology, Nanjing, China, [2]School of Atmospheric Sciences, Nanjing University of Information Science and Technology, Nanjing, China, [3]Department of Civil and Environmental Engineering, University of California, Irvine, CA, USA, [4]Department of Earth System Science, University of California, Irvine, CA, USA

Abstract Soil moisture-precipitation coupling critically shapes Earth's water and energy cycles, yet it remains difficult to quantify because land-atmosphere interactions are nonlinear, multivariate, and state-dependent. This study applies the High-Dimensional Model Representation (HDMR) framework to observation-based data sets over the contiguous United States to diagnose temporal SM-precipitation amount coupling, defined here as the influence of morning land-atmosphere states on same-day afternoon rainfall. HDMR decomposes structural, correlative, and cooperative controls of morning precursors and shows that traditional linear approaches, such as correlation analysis, underestimate coupling strength, explaining only 6%–8% of precipitation variance in regions where HDMR identifies soil moisture (SM) contributions of up to about 20%. The first-order decomposition reveals a direct wet-soil advantage in the eastern Great Plains, while convective available potential energy (CAPE) over Texas and land-surface temperature over Louisiana dominate in separate instability- and surface-temperature-controlled regimes, respectively. Beyond direct effects, the strongest second-order signal is a localized SM-CAPE interaction over southern Oklahoma, where dry soils combined with high CAPE enhance afternoon precipitation through a thermodynamic-triggering pathway. Thus, positive and negative SM effects can coexist over the Great Plains within temporal precipitation-amount coupling through distinct physical pathways. Finally, using the observation-based HDMR diagnostics as a benchmark, we evaluate the convection-permitting CONUS404 simulation. CONUS404 qualitatively reproduces the main functional structures of SMPC, including the wet-soil first-order response and dry-soil-high-CAPE synergy, but systematically underestimates their coupling strength. These findings underscore the need for nonlinear diagnostics to improve the representation of sub-daily land-atmosphere coupling in weather and climate models.

Plain Language Summary Soil moisture plays a critical role in local weather by controlling the release of heat and water vapor into the atmosphere. However, because these land-atmosphere interactions are complex, standard statistical tools often struggle to measure them accurately. We apply a nonlinear functional decomposition method to observations across the United States to quantify how morning soil conditions influence afternoon rainfall. We then use these results to evaluate whether convection-permitting models accurately simulate these relationships. We find that soil moisture explains up to 20% of rainfall variability compared to only 6%–8% estimated by traditional methods. While we identify distinct regions where either wet or dry soils favor rainfall, the convection-permitting model correctly simulates the patterns but systematically underestimates their strength. These results suggest that current models undervalue the land surface's influence on precipitation. Capturing these complex interactions is essential for improving precipitation forecasts and climate projections.

1. Introduction

The interaction between soil moisture (SM) and precipitation is a critical driver of Earth's surface water and energy cycles (Seneviratne et al., 2010). This SM-precipitation coupling (SMPC) modulates near-surface processes and exerts control on energy partitioning (Findell et al., 2011), boundary layer dynamics (Ek & Holtslag, 2004; Ford et al., 2023), and mesoscale circulation (Taylor et al., 2011), thereby influencing regional

Software: Yifu Gao, Jasper A. Vrugt
Supervision: Haishan Chen, Efi Foufoula-Georgiou, Jasper A. Vrugt
Validation: Yifu Gao, Runze Li, Efi Foufoula-Georgiou, Jasper A. Vrugt
Visualization: Yifu Gao, Runze Li
Writing – original draft: Yifu Gao
Writing – review & editing: Yifu Gao, Runze Li, Haishan Chen, Efi Foufoula-Georgiou, Jasper A. Vrugt

extremes of precipitation (Dong et al., 2024; Levine et al., 2016) and droughts (Bevacqua et al., 2024). 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 (Duerinck et al., 2016; W. Liu et al., 2022). 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 (Guo et al., 2006). An in-depth understanding of the coupling between SM and precipitation is essential for improving precipitation forecasts, enhancing drought prediction, and refining climate change projections, especially in the context of global warming and ongoing land-cover and land-use changes (Qiao et al., 2023).

SMPC mechanisms can have either positive (wet soil) or negative (dry soil) effects on precipitation (Santanello et al., 2007). A wet soil supports larger latent heat fluxes (or evapotranspiration) in a SM-limited regime (Budyko, 1974), increasing the air's moisture content, moist static energy (MSE) (Eltahir, 1998) and evaporative fraction (EF) (Findell et al., 2011). The triggered convective initiation and shallow planetary boundary layer (PBL) development (Taylor, 2015) promote convective cloud formation and precipitation. A dry soil on the contrary has higher surface temperature, a larger temperature gradient, and convective triggering potential (CTP) (Findell & Eltahir, 2003a). Enhanced thermal updrafts in turn reduce convection inhibition and promote air parcels to reach the lifting condensation level (LCL) to form precipitation (Ford, Quiring, et al., 2015; Taylor et al., 2012). At the mesoscale, these feedback mechanisms typically operate on diurnal to multi-day timescales, with precipitation most often developing during the midday to afternoon hours of the warm season.

Over the past two decades, numerous data-driven and simulation-based studies have investigated the strength, sign, and governing pathways of SMPC (Findell & Eltahir, 2003a; Ford, Quiring et al., 2015; Ford et al., 2023; Rappin et al., 2022; Sun et al., 2025), yet a long-standing debate remains regarding the exact nature of these feedbacks in hotspot regions like the Central United States and the Sahel (Ferguson et al., 2012; Findell et al., 2011; Gaal & Kinter III, 2021; Guillod et al., 2015; Taylor et al., 2012). Crucially, an emerging consensus indicates that these apparent discrepancies stem not only from model physics uncertainties (Heinze et al., 2017; Huang et al., 2014; Jiang & Wang, 2023; Lee & Hohenegger, 2024; Stephens et al., 2010) and data resolutions (Yuan et al., 2020; A. Wang et al., 2025), but also from differences in the definitions used to define "coupling" (Guillod et al., 2015; Lee et al., 2026; Santanello et al., 2018). Generally, SMPC research can be structured along a two-dimensional conceptual matrix: (a) *temporal* versus *spatial* coupling frameworks, and (b) *precipitation amount* versus *precipitation probability* metrics. This taxonomy yields four distinct categories of coupling: (a) temporal SM-precipitation amount coupling, (b) spatial SM-precipitation amount coupling, (c) temporal SM-precipitation probability coupling, and (d) spatial SM-precipitation probability coupling. Many landmark observational studies focus heavily on the spatial and probabilistic dimensions (Types c and d). For instance, Taylor et al. (2012) isolated a dominant negative spatial coupling (Type d), demonstrating a clear geographic preference for afternoon convective initiation over drier antecedent soil patches. This spatial preference was further shown by Hsu et al. (2017) to be strongly dependent on background soil wetness. Similarly, Guillod et al. (2015) reconciled seemingly contradictory behaviors by showing that while spatial coupling can be negative (Type d), the underlying temporal coupling is often positive (Type c) by placing precipitation occurrence in the context of day-to-day variations. More recently, Lee et al. (2026) confirmed that global storm-resolving simulations can successfully reproduce these spatial and temporal preferences (Types c and d) at convection-permitting scales. These studies provide important insight into precipitation occurrence and preferred location, but they address coupling definitions that differ from the one analyzed here. In this study, we focus on Type a coupling: temporal SM-precipitation amount coupling. Specifically, we quantify how morning SM and other land-atmosphere states explain the variance of same-day afternoon precipitation within local analysis blocks.

Once the coupling target is defined, the method used to compute coupling strength introduces an additional limitation. Most previous work quantifies SMPC using simple statistical metrics or model diagnostics, each with its own limitations (Seneviratne et al., 2010). Specifically, linear Pearson correlations (Welty & Zeng, 2018) or multiple linear regressions (MLR) (Zhou et al., 2022) have been used to link SM to precipitation quantities. However, these metrics fail to detect nonlinear relationships and struggle to disentangle the signal of SM from dominant confounding variables such as land-surface temperature (LST), convective available potential energy (CAPE), or large-scale synoptic circulation (Chen et al., 2023; Song et al., 2019). Within modeling frameworks, such Type a coupling is frequently diagnosed via artificial SM perturbation experiments (Koster et al., 2004; Zhou et al., 2021). While useful, such controlled sensitivities evaluate a non-intervenable system under non-equilibrium states, making direct verification against real-world observations impossible (Van Leeuwen et al., 2021). To date, a framework that can systematically decompose the nonlinear, multivariate, and interactive variance contributions of land-atmosphere states to temporal precipitation amounts under realistic covariations remains lacking.

In this paper, we address this gap by introducing a comprehensive framework to quantify the nonlinear dependence of afternoon precipitation amounts on morning SM and auxiliary land-atmosphere variables using HDMR (Li & Rabitz, 2010, 2012). Grounded firmly in Type A (temporal amount) coupling, HDMR generalizes the analysis of variance (ANOVA) to dependent input factors by generating a strict mathematical superposition of first-, second-, and higher-order component functions. This decomposition isolates the structural, correlative, and cooperative contributions of SM and other variables to precipitation, providing an interpretable and nonlinear extension of conventional correlation or regression analysis. Unlike perturbation-based sensitivity experiments, HDMR explores the entire input space simultaneously, ensuring that derived coupling diagnostics reflect realistic covariations among land-atmosphere processes. To demonstrate the merits of the proposed method, we first show HDMR's capacity to parse out structural, correlative, and cooperative effects through an illustrative numerical example. We then apply HDMR to observational data sets, using Soil Moisture Active Passive (SMAP) measurements and Multi-Radar Multi-Sensor (MRMS) precipitation, to analyze the nonlinear coupling between morning SM and afternoon precipitation across the Continental United States (CONUS). We train the model using data from 2016 to 2019 and test on 2020, identifying SMPC hot spots that are consistent across the train-test periods. The results also reveal distinct spatial regimes where precipitation is dominated by CAPE or LST instead of SM. Benchmarking against a Random Forest (RF) model further demonstrates that HDMR effectively captures both linear and nonlinear dependencies. Finally, as a diagnostic application, we evaluate the CONUS404 convection-permitting simulations and find that, although the model qualitatively reproduces the observation-based spatially varying sign of SMPC, it systematically underestimates the magnitude of these couplings, highlighting HDMR's capability for diagnosing land-atmosphere coupling in climate models toward targeted improvements.

The remainder of this paper is organized as follows. Section 2 describes the observational data sets, including satellite- and radar-based SM and precipitation, as well as the selected auxiliary land-atmosphere variables used for HDMR decomposition. Section 3 outlines the HDMR framework and the derivation of the ANOVA-inspired coupling strength, along with the experimental design for the numerical case study and the main SMPC analysis. Section 4 presents the analyses of observed SMPC hot spots over the central United States and benchmarking against RF. The performance of the convection-permitting CONUS404 simulation is evaluated against observations with respect to SMPC, illustrating HDMR's diagnostic potential for land-atmosphere coupling in climate models. Finally, Section 5 summarizes the main findings and discusses future research directions.

2. Data

To investigate the diurnal coupling between SM and precipitation, we focus on the CONUS as our study domain. The analysis spans the warm seasons (April through September) from 2016 to 2020, a period defined by the overlapping availability of our primary satellite and radar-based data sets. Specific details regarding data preprocessing, including spatial regridding to a common resolution and temporal synchronization, are provided in Section 3.2.

2.1. SMAP L4 Product

The Level 4 SM product (L4_SM) from the Soil Moisture Active Passive (SMAP) mission provides global, 3-hourly estimates of surface and root-zone SM at a 9-km resolution (Reichle et al., 2022). Although SMAP's active radar ceased operations in July 2015, its passive L-band (1.4 GHz) radiometer continues to provide robust measurements of land surface brightness temperatures. The L4_SM algorithm assimilates these brightness temperatures from both the descending (≈6:00 a.m.) and ascending (≈6:00 p.m.) local solar time satellite passes into the NASA catchment land surface model, which is configured on a global 9-km EASE-Grid 2.0. This study utilizes the 3-hourly, time-averaged geophysical product (SPL4SMGP), which reports SM $\left(m^3\ m^{-3}\right)$ for the top

Table 1
Description of the Major CONUS404 Variables Used in This Study

Parameter	Symbol	Units
Accumulated grid-scale precipitation	P	mm
Soil moisture	SM	$m^3\ m^{-3}$
Convective available potential energy	CAPE	$J\ kg^{-1}$
Surface skin temperature	LST	K
Precipitable water	PW	m
Leaf area index	LAI	–

0–5 cm soil layer. Many studies have confirmed the accuracy and reliability of SMAP/L4 (Koster et al., 2018; Reichle et al., 2017; Tavakol et al., 2019).

2.2. MRMS Gauge-Corrected QPE

For precipitation, we mainly rely on the MRMS gauge-corrected Quantitative Precipitation Estimation (QPE) product (Smith et al., 2016; J. Zhang et al., 2016), available since around 2015. MRMS integrates 3D volume-scan data from 146 S-band dual-polarization Weather Surveillance Radar-1988 Doppler (WSR-88D) radars across the conterminous United States (CONUS) and roughly 30 C-band single-polarization radars operated by Environment and Climate Change Canada in Canada, and merges these observations into a seamless 3D radar mosaic at very high resolution (1-km, 2-min).

These radar data are combined with key environmental information including surface and wet-bulb temperatures, winds, and relative humidity from the NWP model, along with lightning and about 7,000 rain gauges from the Hydrometeorological Automated Data System. The gauge information is essential for correcting systematic radar biases and improving the quantitative accuracy of the precipitation estimates, which MRMS uses to produce a suite of severe-weather diagnostics and QPE products. When aggregated to hourly or longer timescales, MRMS generally provides more accurate precipitation estimates, both in magnitude and timing, than satellite products (Habibi et al., 2021; Moazami & Najafi, 2021; White & Nelson, 2024). Owing to its high spatiotemporal resolution and the gauge-correction step, MRMS is well-suited as a reference precipitation data set in our analysis of diurnal SMPC.

2.3. CONUS404 Data Set

We resorted to CONUS404 data set (Rasmussen et al., 2023a, 2023b) for auxiliary land-atmosphere variables as HDMR inputs. CONUS404 is a convection-permitting hydroclimate reanalysis, which provides hourly atmospheric and hydrologic variables at a 4 km spatial resolution over the CONUS for water years 1980–2021. The CONUS404 data set was produced by dynamically downscaling the fifth-generation ECMWF reanalysis using the Weather Research and Forecasting Model (v3.9.1.1). The primary physical configurations include the Thompson microphysics scheme (G. Thompson & Eidhammer, 2014), the Yonsei University (YSU) PBL scheme (Hong et al., 2006), the Rapid Radiative Transfer Model for General Circulation Models (RRTMG) (Iacono et al., 2008), and the Noah multiparameterization land surface model (Noah-MP) (Niu et al., 2011). An important enhancement over its predecessor, CONUS1 (C. Liu et al., 2017), is the incorporation of the Miguez-Macho-Fan groundwater scheme (MMF) (Barlage et al., 2021; Miguez-Macho et al., 2007), which significantly reduces biases in surface fluxes and convective precipitation, key factors in resolving mesoscale convective systems.

Morning conditions of LST, CAPE, precipitable water (PW), and leaf area index (LAI) serve as antecedent states in our SMPC analysis framework. Besides, we extract the SM and precipitation data for a model diagnosis through comparing the SMPC inferred from observations (SMAP/MRMS) and climate model simulations (CONUS404). Table 1 summarizes the auxiliary variables used in this study. These fields capture critical aspects of the surface-atmosphere system, such as evapotranspiration, thermodynamic environment, and moisture availability, all of which influence convective triggering and precipitation formation (Gao et al., 2024; Seneviratne et al., 2010). In Section 3.2, we elaborate on our steps for data preprocessing.

3. Methods

3.1. High-Dimensional Model Representation

SMPC involves multiple and interconnected pathways, which are difficult to quantify by traditional correlation and regression analyses. These methods typically focus on direct, linear relationships, often neglecting critical compound effects (e.g., joint SM-PW contributions) and internal couplings (e.g., SM-LST interactions)

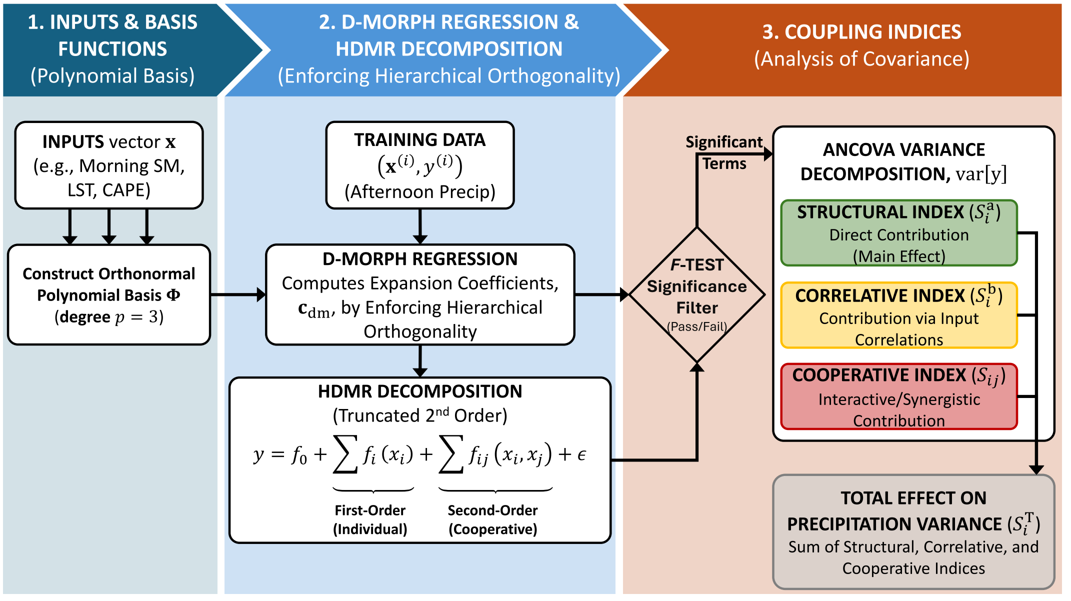

Figure 1. Schematic of the High-Dimensional Model Representation (HDMR) analysis workflow. Input vectors (morning soil moisture (SM) and auxiliary variables) are first standardized and serve as input variables for orthonormal polynomial basis functions. These basis functions, combined with the expansion coefficients estimated by D-MORPH regression, construct the first- and second-order component functions of HDMR. Finally, the component functions are integrated to derive coupling indices, quantifying the structural, correlative, and cooperative effects of SM and other variables on afternoon precipitation.

(Seneviratne et al., 2010). By contrast, HDMR transforms this complex web of couplings into a hierarchy of component functions and employs analysis of covariance (ANCOVA) to isolate direct (structural) from indirect (correlative and cooperative) effects. This allows HDMR to provide a more systematic, multi-variable assessment of how SM influences precipitation. Figure 1 provides a schematic overview of the HDMR methodology.

3.1.1. HDMR Formulation

Let $\mathbf{x} = (x_1, \ldots, x_d)^\top$ denote the vector of d input variables (morning SM and other land-atmosphere fields in Table 1) governing an output $y = f(\mathbf{x})$, which we define as afternoon precipitation. HDMR builds on the finite multivariable function expansion of Sobol (1993) and decomposes precipitation into a sum of hierarchical component functions (Li & Rabitz, 2012):

$$y = y_0 + \sum_{i=1}^{n_1} f_i(x_i) + \sum_{1\le i<j\le d}^{n_2} f_{ij}(x_i, x_j) + \sum_{1\le i<j<k\le d}^{n_3} f_{ijk}(x_i, x_j, x_k) + \cdots + f_{12\ldots d}(x_1, x_2, \ldots, x_d) + \epsilon, \tag{1}$$

where y_0 is the mean function output and $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$ denotes the normally distributed residual with zero mean and a constant variance, σ_ϵ^2. The first-order terms, $f_i(x_i)$, capture the individual contribution of each input variable, while the second-order terms and higher-order terms, $f_{ij}(x_i, x_j), \ldots$, represent cooperative contributions among two or more variables. In typical physical systems, third-order and higher-order terms are often negligible (Falchi et al., 2018; Gao et al., 2023; Kucherenko et al., 2011; Rabitz & Aliş, 1999; Shereena & Rao, 2019; H. Wang et al., 2017). Thus, we truncate the series by retaining only up to second-order components:

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

where $n_{12} = d + \frac{d(d-1)}{2}$ is the total number of first- and second-order component functions, the index u runs over both individual (d) and pairwise $\left[\frac{d(d-1)}{2}\right]$ terms, and $\mathbf{x}_u$ is the vector corresponding to the individual (x_i) and pairwise (x_i, x_j) components of the input vector $\mathbf{x} = (x_1, \ldots, x_d)^{\top}$. In our implementation, y_0 signifies the mean 12-hr afternoon precipitation in units of mm, and the component functions f_u quantify the individual and bivariate contributions of the land-atmosphere variables to precipitation.

Note that these component functions portray the nonlinear dependence of precipitation on SM and auxiliary land-atmosphere variables through orthonormal polynomials (see Text S2 in Supporting Information S1). Capturing this nonlinearity is essential, as land-atmosphere interactions are typically threshold-driven processes, characterized by sharp transitions between moisture-limited and energy-limited regimes that linear metrics fail to resolve (Koster et al., 2004; Seneviratne et al., 2010).

3.1.2. Hierarchical Orthogonality

In Sobol's original formulation, strictly enforcing orthogonality (the "vanishing condition") between component functions facilitates an ANOVA style interpretation of input importance (Sobol, 1993). While the vanishing condition can be easily achieved by random sampling approaches, it disregards correlation among physical variables. To accommodate input correlation, Hooker (2007) proposed a *relaxed* vanishing condition

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

where $w_u(\mathbf{x}_u)$ is the joint probability density function of the variables in $\mathbf{x}_u$. Equation 3 is both necessary and sufficient for *hierarchical orthogonality* among the component functions, ensuring a unique functional decomposition that exactly distinguishes between structural, correlative, and cooperative contributions of the input variables (Gao et al., 2023; Li & Rabitz, 2012). 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 guarantees that each HDMR term captures new or interactive information not already contained in lower-order terms, thereby achieving a more accurate representation of complex, correlated processes. The component functions are constructed using the extended bases orthonormal polynomials with up to degree $p = 3$ and associated linear expansion coefficients, $\mathbf{c}_{\mathrm{dm}}$. The total number of the expansion coefficients is $l = d \cdot p + \frac{1}{2}d(d-1)(2p + p^2)$. D-MORPH regression (Li & Rabitz, 2010) enforces hierarchical orthogonality of the component functions in pursuit of the optimum expansion coefficients. Further details on component function construction and D-MORPH regression are described in Text S2 in Supporting Information S1.

3.1.3. Coupling Indices: Analysis of Covariance (ANCOVA)

To ensure that only robust component functions contribute to the coupling diagnostics, we evaluate the statistical significance of each first- and second-order HDMR term using an F-test. For a given component function $f_d(x_d)$, significance is assessed by comparing the sum of squared residuals of the HDMR expansion with and without this term. Let SSR_1 denote the residual sum of squares for the truncated expansion $y = y_0 + \sum_{i=1}^{d-1} f_i(x_i)$ with $l_1 = (d-1)p$ coefficients, and let SSR be the corresponding quantity for the full expansion including $f_d(x_d)$ with $l = l_1 + p$ coefficients. The F-statistic

$$F = \frac{(\mathrm{SSR}_1 - \mathrm{SSR})/(l - l_1)}{\mathrm{SSR}_1/(n - l_1)}, \tag{4}$$

is compared against the critical value

$$F_{\mathrm{crit}} = F_{\mathcal{F}}^{-1}(1 - \alpha | l - l_1,\, n - l_1), \tag{5}$$

where $F_{\mathcal{F}}^{-1}(p_\alpha | \nu_1, \nu_2)$ is the quantile function of the Fisher-Snedecor distribution, α is the chosen significance level, and $p_\alpha = 1 - \alpha$. If $F < F_{\mathrm{crit}}$, the corresponding component function is treated as statistically insignificant. In such cases, we set its structural, correlative, and cooperative coupling indices (described below) to zero. This filtering ensures that only HDMR components that demonstrably improve the model fit contribute to the variance and covariance decomposition, and prevents spurious coupling signals from arising due to sampling noise or weak relationships.

The subscript notation introduced earlier facilitates a variance-based decomposition of the model output, y (after subtracting its mean, y_0), following Li et al. (2010). Recall that

$$\mathrm{Var}[y] = \mathbb{E}\left[(y - \mathbb{E}(y))^2\right] = \int_{\mathbb{K}^d} w(\mathbf{x})\left(y - y_0\right)^2 \mathrm{d}\mathbf{x}. \tag{6}$$

where $w(\mathbf{x})$ is the joint probability density function of the input variables and $\mathbb{K}^d$ is the unit hypercube of the rescaled inputs $\left(\mathbf{x} \in \mathbb{K}^d = [0,1]^d\right)$. Substituting the truncated HDMR expansion (up to second order) into this expression yields

$$\mathrm{Var}[y] = \int_{\mathbb{K}^d} w(\mathbf{x})\left(\sum_{u=1}^{n_{12}} f_u(\mathbf{x}_u) + \epsilon\right)^2 \mathrm{d}\mathbf{x}, \tag{7}$$

which is equivalent to

$$\mathrm{Var}[y] = \int_{\mathbb{K}^d} w(\mathbf{x})\left(\sum_{u=1}^{n_{12}} f_u(\mathbf{x}_u)\right)^2 \mathrm{d}\mathbf{x} + 2\int_{\mathbb{K}^d} w(\mathbf{x}) \sum_{u=1}^{n_{12}} \left(f_u(\mathbf{x}_u)\cdot\epsilon\right) \mathrm{d}\mathbf{x} + \mathbb{E}\left[\epsilon^2\right]. \tag{8}$$

Because the least squares procedure ensures each component function f_u is orthogonal to the residual, $\epsilon \sim \mathcal{N}\left(0, \sigma_\epsilon^2\right)$, the middle term drops out:

$$\int_{\mathbb{K}^d} w(\mathbf{x})\left(f_u(\mathbf{x}_u)\cdot\epsilon\right) \mathrm{d}\mathbf{x} = 0. \tag{9}$$

And Equation 8 simplifies to (Li et al., 2010)

$$\mathrm{Var}[y] = \int_{\mathbb{K}^d} w(\mathbf{x})\left(\sum_{u=1}^{n_{12}} f_u(\mathbf{x}_u)\right)^2 \mathrm{d}\mathbf{x} + \mathbb{E}\left[\epsilon^2\right]. \tag{10}$$

Expanding the square of the sum then yields both *variance* and *covariance* contributions of the individual component functions:

$$\mathrm{Var}[y] = \sum_{u=1}^{n_{12}} \mathrm{Var}\left[f_u(\mathbf{x}_u)\right] + \sum_{u=1}^{n_{12}} \mathrm{Cov}\left[f_u(\mathbf{x}_u), \sum_{m\neq u} f_m(\mathbf{x}_m)\right] + \mathbb{E}\left[\epsilon^2\right]. \tag{11}$$

To measure how each component function $f_u(\mathbf{x}_u)$ influences the total output variance, Li et al. (2010) define the *structural* and *correlative* indices by normalizing the variance and covariance terms, respectively:

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

Their sum gives the total coupling index for the uth component function:

$$S_u = S_u^{\mathrm{a}} + S_u^{\mathrm{b}}. \tag{13}$$

For a pair of variables x_i and x_j, its cooperative coupling index $S_{ij}(x_i, x_j)$ quantifies the synergistic contribution of their joint state to precipitation variance, while a related measure, the total-effect index (Homma & Saltelli, 1996),

$$
\begin{aligned}
S_i^{\mathrm{T}}(x_i) &= S_i(x_i) + \sum_{j \neq i} S_{ij}(x_i, x_j) \\
&= S_i^{\mathrm{a}}(x_i) + S_i^{\mathrm{b}}(x_i) + \sum_{j \neq i} S_{ij}(x_i, x_j)
\end{aligned}
\tag{14}
$$

quantifies, for example, the total contribution of SM to precipitation through direct (structural) and indirect (correlative and cooperative) effects.

To aid physical interpretation, the HDMR indices can be viewed as quantifying distinct pathways by which a predictor contributes to afternoon precipitation variability. The structural coupling index, S_i^{a}, measures the direct, stand-alone contribution of a variable, whereas the correlative coupling index, S_i^{b}, measures the portion of its apparent influence arising from covariance with other predictors. By virtue of hierarchical orthogonality, the cooperative index, S_{ij}, quantifies the additional contribution from the joint state of two variables that is not captured by the corresponding first-order effects alone. In addition to these variance-based indices, the HDMR component functions provide physical insight into the form of the coupling itself. For example, the first-order function, $f_i(x_i)$, describes how precipitation responds to anomalies in a given predictor and can reveal whether the coupling is monotonic, threshold-like, or saturating. A threshold-like $f_1(x_1)$, for instance, indicates a regime in which precipitation is relatively insensitive to SM below a certain level, but increases rapidly once that threshold is exceeded, thereby helping diagnose the underlying mechanism of SMPC. More detailed interpretation of the functional shapes of the HDMR component functions is provided in the Results section.

3.2. Experimental Design

We apply HDMR to address the main objective of this study, which is to disentangle the structural, correlative, and cooperative controls of SMPC across CONUS and their respective strengths. All input samples are standardized to have zero mean and unit standard deviation. Before applying HDMR, all data sets are collocated to a common 16 km analysis grid defined from the aggregated CONUS404 grid. CONUS404 fields are spatially aggregated from the native 4 km grid to 16 km by averaging all native grid cells within each analysis grid cell. MRMS precipitation is collocated to the same 16 km grid by area-averaging the native 1 km precipitation fields within each analysis grid cell. SMAP SM is collocated to this grid using linear interpolation based on Delaunay triangulation, with no extrapolation outside the convex hull of available SMAP grid cells. The 16 km grid is used to maintain consistent spatial support between the observation-based and model-based experiments and to reduce the computational burden of the CONUS-scale HDMR analysis.

An essential prerequisite for detecting and studying SMPC is to mitigate the impact of synoptic systems (Findell et al., 2011; Tuttle & Salvucci, 2017), which can obscure local SM impacts. Large-scale stratiform events, for example, often yield a misleading positive correlation between antecedent SM and subsequent rainfall (Alfieri et al., 2008). To address this, we extract only those days during the 5-year record (2016–2020, April–September) meeting the following criteria:

1. *Afternoon Rainfall Initiation*: Precipitation begins between 12:00–15:00 local time on a day (labeled t_{start}).
2. *Negligible Prior Rainfall*: From $t_{\mathrm{start}} - 24$ hrs to $t_{\mathrm{start}} - 1$ hrs, the cumulative precipitation of a grid cell is below 0.5 mm.
3. *Precipitation Accumulation*: The 12-hr total between t_{start} and $t_{\mathrm{start}} + 11$ hrs serves as a single outcome $y^{(i)}$ in the data vector.
4. *Convective Environment*: We restrict the analysis to events with relatively large CAPE (> 200 J kg^{-1}) to select afternoon rainfall events that are more likely convective.

The choice of CAPE threshold is consistent with prior studies that use CAPE of 100–200 J kg^{-1} to identify environments favorable for deep moist convection (R. L. Thompson et al., 2007; Kolendowicz et al., 2017; Taszarek et al., 2018). We use CAPE rather than CTP because it is directly available from CONUS404 and more directly related to parcel buoyancy and the pre-convective instability of the atmosphere, whereas CTP is primarily a lower-tropospheric profile diagnostic commonly used in combination with humidity constraints in the Findell-Eltahir framework (Findell & Eltahir, 2003a). For each valid event, we extract morning conditions (7:00 local time) of SM and auxiliary land-atmosphere variables from Table 1 to construct the 5×1 input vector:

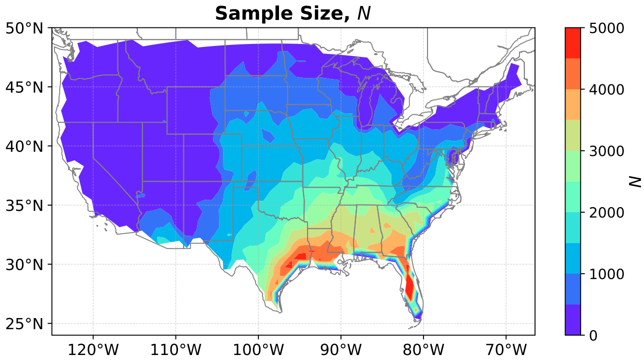

Figure 2. Sample size N for High-Dimensional Model Representation analysis of Soil moisture-precipitation coupling over CONUS determined from Soil Moisture Active Passive/Multi-Radar Multi-Sensor products with a 5-year record (2016–2020, April–September).

$$\mathbf{x} = (x_1, \ldots, x_5)^{\top} = [\text{SM, CAPE, LST, PW, LAI}]^{\top}. \tag{15}$$

The above filtering steps mitigate the influence of large-scale synoptic forcing but reduce the number of usable rain-initiation days per 16 km × 16 km grid cell, typically to a few hundred in the Central US. Since HDMR requires a sufficiently large sample size, we do not fit the model separately at each 16 km × 16 km grid cell. Instead, we aggregate neighboring cells into 5 × 5 blocks, corresponding to 80 km × 80 km regions, and pool all valid event samples from the 25 constituent cells into a single data set. One HDMR decomposition is then performed for each block, yielding one set of coupling metrics and component functions representative of the pooled regional relationship within that 80 km × 80 km area. This gives a total sample size of $N = \sum_{i=1}^{25} N_i$, where N_i denotes the number of valid events in each cell, and ensures sufficient observations for reliable estimation of the polynomial expansion coefficients in Equation (S6) in Supporting Information S1.

After assembling each block's $N \times 6$ HDMR input matrix $\mathbf{X}$ and $N \times 1$ data vector $\mathbf{b}$, we map the sample size N in Figure 2. Blocks with $N < 1{,}000$ are discarded from the HDMR analysis, reflecting the western US, where warm-season convective precipitation is infrequent to reveal robust diurnal SMPC signals. Our preprocessing steps above ensure that the final data set is both representative of local SM-precipitation interactions and sufficiently large to support HDMR's higher-order expansions. To enhance the credibility of the derived regional couplings, we benchmark the predictive skill of the HDMR expansion against RF and Multiple Linear Regression (MLR). MLR is included as a simple linear benchmark because conventional SMPC diagnostics often rely on correlation- or regression-based approaches (Welty & Zeng, 2018; Zhou et al., 2021; G. Wang et al., 2024), whereas RF provides a flexible nonlinear benchmark that is widely used in atmospheric and climate applications (de Burgh-Day & Leeuwenburg, 2023; O'Gorman & Dwyer, 2018). The RF model configuration, comprising 300 trees with a maximum depth of 100, is selected via a five-fold cross-validation grid search over tree counts of $\{100, 300, 500\}$ and depths of $\{10, 50, 100\}$. Other hyperparameters are maintained at standard default values.

We define the main experiment (ExpA) by training HDMR on a 4-year (2016–2019) period using SMAP SM and MRMS precipitation. This yields the optimal expansion coefficients and corresponding component functions that represent the dominant multi-year coupling patterns, which are then evaluated against the independent 2020 data. Again, we only diagnose SMPC for blocks with more than 1,000 valid event samples to ensure robust estimation of the expansion coefficients. As shown in our previous HDMR sensitivity analysis (Gao et al., 2023), stable estimation requires the ratio between sample size and the number of unknown expansion coefficients, N/l, to exceed about 5. In the present setup, five predictors give $l = 165$, so the retained blocks typically satisfy $N/l \approx 6$, indicating that the 4-year training period provides sufficient sample support for stable HDMR fitting. By focusing our analysis on regions where the signal remains significant during the 2020 testing period, we ensure that the identified coupling mechanisms are persistent physical features rather than artifacts of overfitting the training data. Moreover, we include a second experiment, ExpB, which replaces the observational training data with simulated SM and precipitation from the CONUS404 data set. This experiment serves as a diagnostic test of the climate model's ability to reproduce the observation-based SMPC patterns identified in ExpA.

3.2.1. Benchmarking Metrics and Preliminary Case Study

To compare HDMR results with simpler methods, we compute the Pearson correlation (r_i) and MLR (β_i) coefficients for each input x_i:

$$r_i = \frac{\sum_{m=1}^{N} \left(x_i^{(m)} - \overline{x_i}\right) \left(y^{(m)} - \bar{y}\right)}{\sqrt{\sum_{m=1}^{N} \left(x_i^{(m)} - \overline{x_i}\right)^2} \sqrt{\sum_{m=1}^{N} \left(y^{(m)} - \bar{y}\right)^2}} \tag{16}$$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_d x_d + \epsilon, \tag{17}$$

Journal of Geophysical Research: Atmospheres 10.1029/2026JD046616

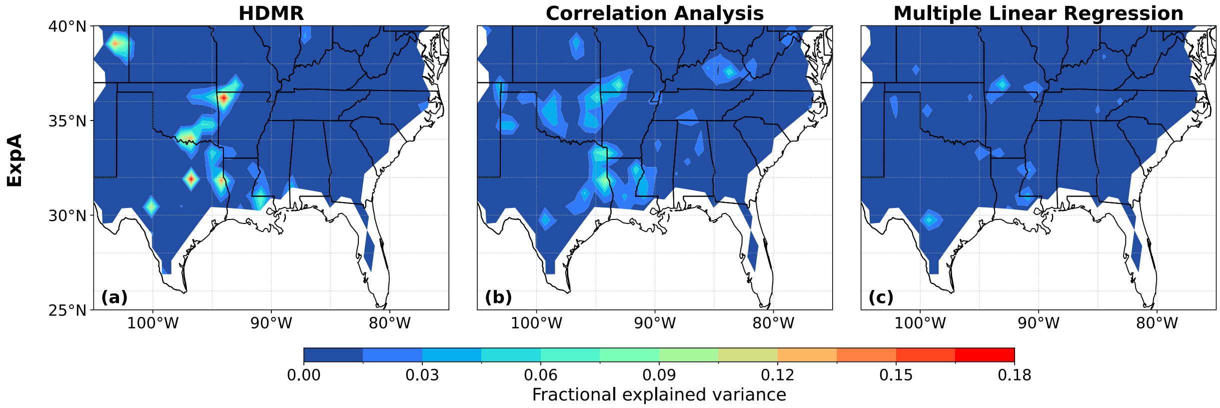

Figure 3. Spatial distribution of the total fractional variance of warm season (April–September 2020) afternoon precipitation (Multi-Radar Multi-Sensor Quantitative Precipitation Estimation) explained by morning soil moisture (Soil Moisture Active Passive L4) across the CONUS. The columns correspond to the (a) High-Dimensional Model Representation, (b) correlation, and (c) MLR methods.

where i and m are variable and sample indices, respectively, and β_i are estimated by ordinary least squares. Since HDMR's coupling indices are inherently variance-based (see Section 3.1.3), we transform r_i and β_i into measures of explained variance:

$$R^2_{\mathrm{C},i} = r_i^2 \tag{18}$$

$$R^2_{\mathrm{LR},i} = \frac{\beta_i^2 \sigma_i^2 + \beta_i \sigma_i \sum_{j \neq i} r_{ij} \beta_j \sigma_j}{\mathrm{var}[y]}, \tag{19}$$

where $R^2_{\mathrm{C},i}$ and $R^2_{\mathrm{LR},i}$ denote the explained variance of y by x_i using correlation coefficient and linear regression, respectively, r_{ij} is the correlation coefficient between x_i and x_j, and σ_i and σ_j refer to the standard deviation of x_i and x_j, respectively. This unifies the output of correlation/regression with the variance-based perspective of HDMR, allowing direct comparisons between these approaches.

As a preliminary benchmark, we apply HDMR to a simplified Light-Use Efficiency model to quantify gross primary productivity sensitivity to four meteorological drivers (SM, R_n, $T_{2\mathrm{m}}$, and u). Because the synthetic GPP response is generated from a prescribed analytical equation in which the relevant nonlinear controls are known a priori, this example provides a controlled reference for evaluating whether HDMR can recover structural, correlative, and cooperative effects. As presented in Text S1 and Figure S1 in Supporting Information S1, HDMR recovers the structural, correlative, and cooperative effects more faithfully than linear methods. This provides additional support for its application to the more complex SMPC analysis that follows.

4. Results

4.1. Hot Spots of SMPC

We now apply HDMR to identify and quantify SMPC hot spots across the CONUS during the warm months of 2020 (April-September). Preprocessing steps mitigate synoptic effects through removing days with prior rainfall (see Section 3.2) to focus on afternoon events. The primary target variable is the 12-hr accumulated precipitation (12:00-24:00 local time), and the input vector includes morning (07:00 local time) SM plus four auxiliary land-atmosphere variables: CAPE, LST, PW, and LAI.

Figure 3 compares the total variance of afternoon precipitation explained by morning SM during the test period (2020), as estimated by HDMR, correlation analysis, and MLR trained on the 2016–2019 data set. We intentionally compare HDMR against correlation and MLR, as these linear approaches are frequently employed in SMPC studies (Duerinck et al., 2016; Welty & Zeng, 2018; Zhou et al., 2021, 2022; C. Zhang et al., 2025). In the reference experiment (ExpA), both HDMR and correlation analysis identify the Central United States, specifically eastern Oklahoma, southwestern Missouri, and eastern Texas, as the primary region of strong coupling

(Figures 3a and 3b). Within these statistically significant SM hot spots, this pattern is physically consistent with transitional climate zones where evapotranspiration is water-limited; moderate soil wetness anomalies can substantially alter surface fluxes and boundary-layer moisture, thereby modulating convective initiation (Seneviratne et al., 2010; Teuling et al., 2009). In contrast, as shown in the next section, regions outside these SM-dominated hot spots are more strongly controlled by CAPE or LST, while SM provides limited additional explanatory information for afternoon precipitation variability during the 2020 test period. A localized secondary area of elevated explained variance, reaching about 20%, also appears over eastern Colorado. We do not interpret this feature as a primary SMPC hotspot in the same sense as the more spatially coherent central Great Plains signals. Nevertheless, its location is physically plausible because afternoon convective initiation east of the Rockies can be influenced by terrain-related mesoscale convergence, as noted in previous studies (Gaal & Kinter III, 2021). We avoid further mechanistic interpretation for this localized signal and focus instead on the more coherent hot spots over eastern Texas and the Oklahoma-Arkansas/Missouri region. In contrast to HDMR and correlation analysis, MLR generally struggles to isolate the specific impact of SM from other land-atmosphere variables, failing to identify coherent spatial patterns of coupling (Figure 3c).

While the spatial identification of hot spots is broadly consistent between HDMR and correlation analysis, the magnitude of the coupling differs substantially. HDMR attributes 6%–20% of the precipitation variance to SM over key hot spots (reaching ∼20% in eastern Oklahoma), whereas correlation analysis suggests a more modest influence of 6%–8%. Furthermore, HDMR identifies strong coupling over southern Oklahoma, a feature less apparent in the correlation maps. As illustrated in our case study, this discrepancy likely stems from the ability of HDMR to capture nonlinear, synergistic effects-such as the interaction between SM and CAPE, which linear methods miss. We elaborate on these sources of variance later by presenting the first- and second-order component functions. Conversely, over eastern Texas and Louisiana, HDMR detects limited significant coupling (filtered by the F-test), whereas correlation analysis indicates a weak signal (3%–6%). This suggests that while a linear association may exist, the relationship does not meet the statistical significance threshold required by the HDMR framework in this energy-limited regime.

Extended results for the 2016–2019 training period are provided in Figure S2 in Supporting Information S1, while the corresponding component functions and their functional classifications for the 2020 test period are presented and discussed later in the next section. To rigorously isolate consistent signals, regions exhibiting statistically significant coupling during training but failing to reach significance in the 2020 test period are masked (set to zero). This constraint ensures a robust train-test comparison focused strictly on areas exhibiting temporally consistent coupling signals. Consequently, our inferred hot spots demonstrate remarkable spatial coherence across time periods (2016–2019 vs. 2020). While the spatial patterns are robust, minor fluctuations in coupling magnitude reflect the interannual variability of the underlying hydroclimatic drivers. Furthermore, these spatial patterns share strong coherence with foundational observational and modeling benchmarks (Ford et al., 2023; Gao et al., 2024; Hu et al., 2021; Koster et al., 2004). The difference in coupling magnitude, however, highlights the critical influence of methodological choices, data resolution, and the physical scale of the coupling processes.

4.2. First-Order Land-Atmosphere Control on Precipitation

To disentangle the distinct functional roles of land-atmosphere variables, we examine the HDMR decomposition shown in Figure 4. Because the contributions of PW and LAI are negligible (Figure S3 in Supporting Information S1), we focus on the three primary drivers: SM, LST, and CAPE. The spatial distribution of the first-order structural coupling index, S_1^a, for SM (Figure 4a) closely resembles its total-effect pattern in Figure 3, indicating that in several localized regions SM exerts a robust and largely independent control on afternoon precipitation variability. A pronounced SMPC hot spot appears in the east of the Great Plains transition zone, especially near the Texas-Louisiana and Oklahoma-Arkansas intersections and eastern Texas, where the first-order SM contribution reaches up to about 13% of the afternoon precipitation variance. This result confirms that, in these regions, SM alone can explain a substantial fraction of the convective precipitation variability.

To provide a more systematic interpretation of the first-order SM effect, the second column of Figure 4 presents a functional classification of the first-order SM component functions within the region delineated by the S_1^a map. The classification procedure is described in Text S3 in Supporting Information S1. Briefly, each first-order component function is interpolated onto a common standardized-anomaly grid and classified using amplitude, end-to-end change, and slope behavior across the anomaly range. The colored square markers are plotted at the

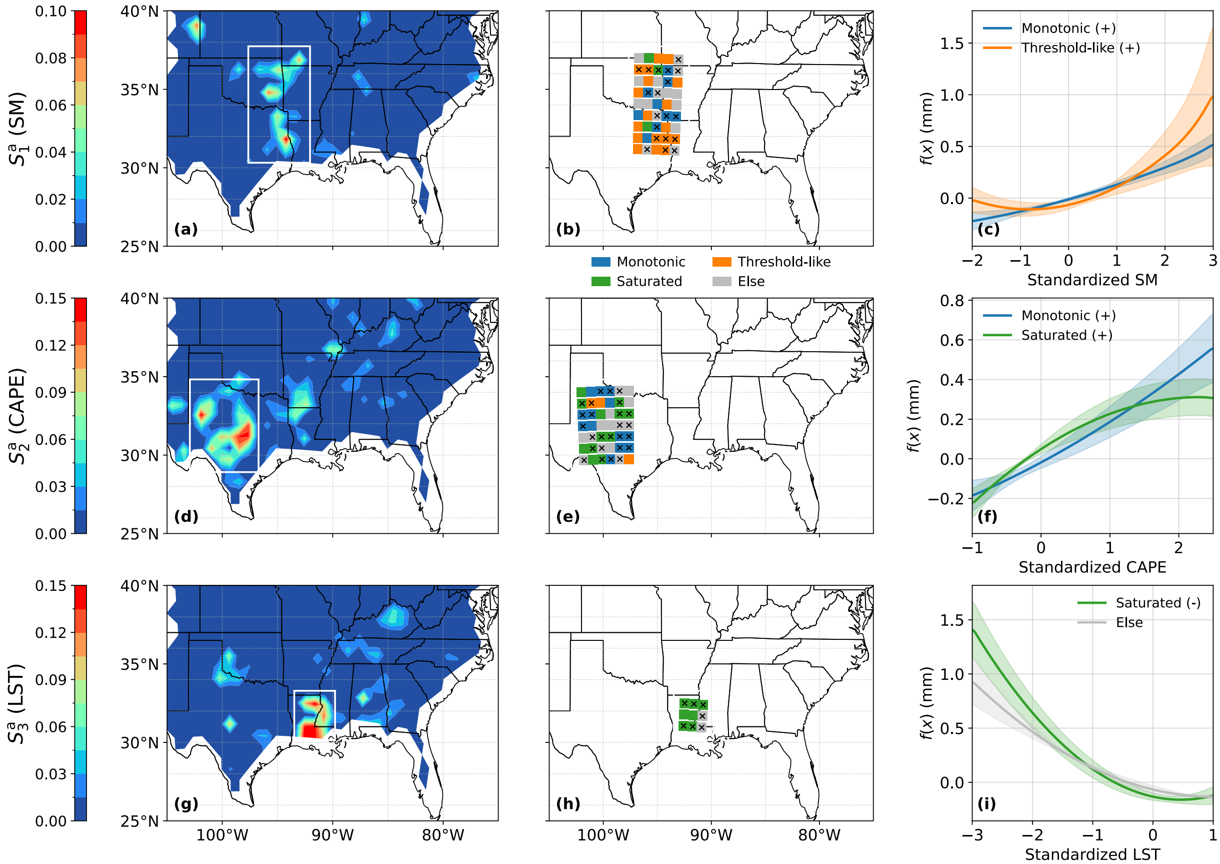

Figure 4. Decomposed first-order controls of morning land-atmosphere variables, soil moisture (SM), convective available potential energy (CAPE), and land-surface temperature (LST), on afternoon precipitation during April-September 2020. Panels (a–c) show the results for SM, panels (d–f) for CAPE, and panels (g–i) for LST. The first column shows the first-order structural coupling indices, S_i^a, representing the direct contributions of each variable to precipitation variance. The second column presents the functional classification of the significant first-order component functions within the regions delineated in the first column; only 80 km × 80 km blocks with statistically significant first-order effects according to the F-test are marked by crosses. The third column shows the regional-mean first-order component functions, f_i, for the two dominant functional types of each variable, summarizing the primary modes of direct land-atmosphere control.

geographic centers of the corresponding 80 km × 80 km HDMR analysis blocks defined on the CONUS404-derived 16 km grid. Therefore, they may not appear perfectly aligned in longitude-latitude coordinates. Only the 80 km × 80 km blocks with statistically significant first-order SM effects according to the F-test are denoted by cross markers. Although the SM first-order component function types exhibit noticeable regional variability, the significant functions are dominated by two types: threshold-like and monotonic. The third column then summarizes these dominant behaviors using the regional mean f_1 for each functional type, providing a more objective characterization of how SM modulates afternoon precipitation across the hot-spot region. The threshold-like SM response is most prevalent over eastern and northeastern Texas. As shown by the orange curve in Figure 4c, precipitation is relatively insensitive to anomalously dry soils (standardized SM < 0), but once SM becomes positive, afternoon precipitation increases rapidly. From the regional-mean perspective, this wet-soil enhancement can increase afternoon rainfall by up to about 1.0 mm across eastern Texas. In contrast, the monotonic SM response is more characteristic of the Texas-Louisiana and Oklahoma-Arkansas intersection regions. The blue curve in Figure 4c shows an approximately linear relationship with an intercept close to zero, implying a nearly proportional increase in precipitation with increasing SM, with a slope of roughly 0.16 mm per one standard deviation of SM. Thus, these results indicate that the first-order SM effect on afternoon precipitation features a clear wet-soil advantage and is organized into threshold-like and monotonic behaviors representing the primary modes of direct SM control. To further illustrate the robustness of these regional-mean functional regimes, Figure 5 shows the clustered individual first-order component functions underlying the dominant functional classes, together with their corresponding regional-mean curves.

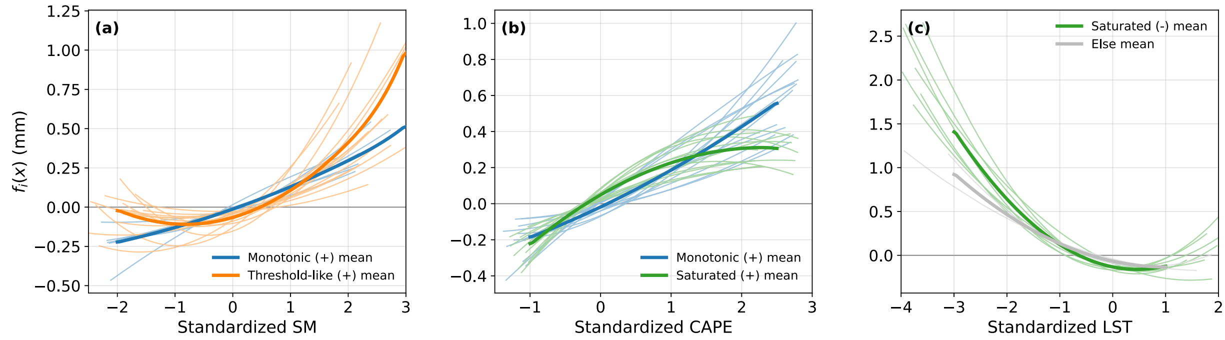

Figure 5. Clustered individual first-order component functions for the dominant functional classes of (a) soil moisture, (b) convective available potential energy, and (c) land-surface temperature. In each panel, light-colored curves denote the individual significant 80 km × 80 km blocks belonging to the selected functional classes, while the darker curves show the corresponding regional-mean first-order component functions also presented in the last column of Figure 4. This figure highlights the extent to which the regional-mean responses are supported by recurrent local functional behavior rather than by a few isolated blocks.

A different picture emerges for CAPE, whose first-order effect is concentrated across the Texas transition zone (Figure 4d). In this region, morning CAPE accounts for up to about 18% of the afternoon precipitation variance, underscoring the importance of atmospheric preconditioning in areas frequently influenced by strong capping inversions (Findell & Eltahir, 2003b; Yin et al., 2015). The functional classification in Figure 4e indicates that significant CAPE responses are mainly organized into two positive functional types: monotonic and saturating. The monotonic positive response is most evident over central and northwestern Texas, where the regional-mean curve in Figure 4f shows an approximately linear increase in afternoon precipitation with increasing CAPE. This behavior indicates that, in these conditionally unstable environments, greater morning buoyancy systematically increases the likelihood or intensity of afternoon convection once the cap is breached. In contrast, the saturated positive response is more apparent over southwestern and northeastern Texas. The corresponding mean function increases rapidly from low to moderate CAPE anomalies, but changes little under very large CAPE anomalies. This saturation suggests that beyond a certain level of atmospheric instability, additional CAPE alone no longer produces a proportional precipitation increase, likely because convective realization also depends on triggering, moisture availability, and inhibition. Quantitatively, increases in morning CAPE translate to an additional 0.3–0.6 mm of afternoon rainfall.

The first-order LST effect exhibits a structurally different behavior and is most prominent over the humid subtropical environment of Louisiana (Figure 4g), where morning LST explains up to about 15% of the afternoon precipitation variance. The functional classification in Figure 4h shows that the significant LST responses are dominated by a saturated negative type. The regional-mean function in Figure 4i indicates that below-average LST enhances afternoon precipitation by up to about 1.5 mm, whereas warmer conditions exert little to no direct influence. This negative and saturating dependence is consistent with an energy-limited regime in which lower morning LST may act as a proxy for higher EF, greater near-surface moisture availability, and a lower LCL, all of which favor earlier or deeper convection (Ford et al., 2023). In addition, reduced morning LST may reflect radiative cooling associated with cloud cover or shallow frontal environments, which are common ahead of organized convection in the lower Mississippi Valley (Su et al., 2023). Thus, unlike the wet-soil advantage identified for SM or the positive instability control represented by CAPE, the LST response primarily captures a cool-surface, moisture-favorable regime in which lower morning surface temperature is linked to enhanced afternoon precipitation.

4.3. Second-Order Land-Atmosphere Control on Precipitation

Beyond the first-order or direct effects, we next examine second-order interactions to identify cases in which the influence of SM depends on the concurrent state of the atmosphere. In contrast to the broader spatial coherence of the first-order responses, the second-order effects are more localized and variable across different pairs of land-atmosphere variables, indicating conditional synergistic regimes rather than widespread controls. Figure 6 summarizes these effects with a focus on the SM-CAPE interaction, which shows the strongest spatial coherence and physical basis. Panel (a) reproduces the total-effect map and delineates the southern Oklahoma subregion where a clear discrepancy emerges between the total-effect and first-order patterns. This emphasizes the presence

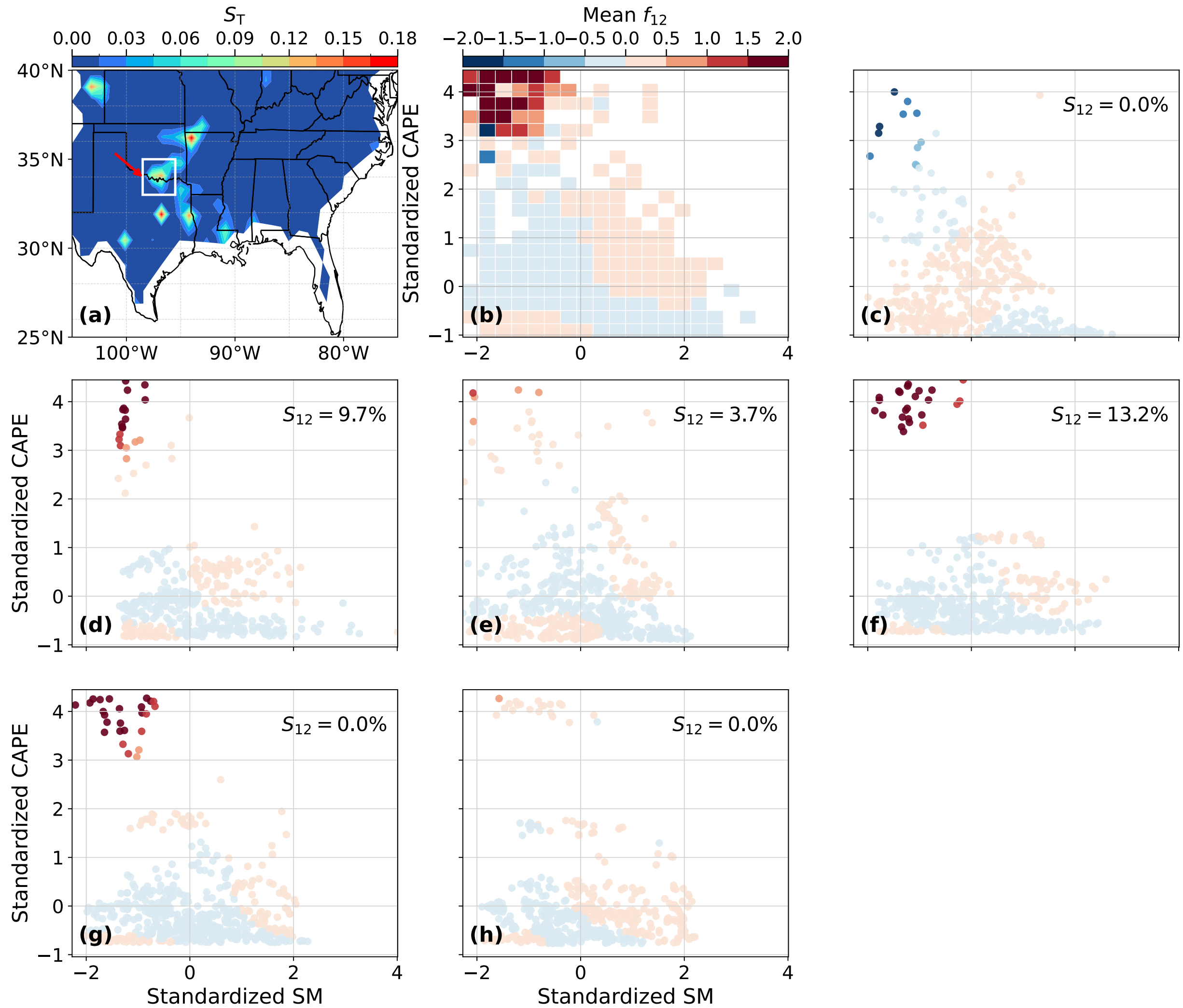

Figure 6. Second-order interaction between morning soil moisture (SM) and convective available potential energy in controlling afternoon precipitation. Panel (a) shows the total-effect of SM on afternoon precipitation and delineates the southern Oklahoma subregion where second-order effects are prominent. Panel (b) presents the regional-mean SM-CAPE interaction component, f_{12}, obtained by bin-averaging samples from the six 80 km × 80 km blocks within this subregion. Panels (c–h) show the sample-level second-order component for each individual block. The annotation S_{12} denotes the explained variance of precipitation by f_{12} in each block.

of an additional interaction-driven contribution beyond the direct effect alone. Panel (b) shows the regional mean interaction component obtained by bin-averaging samples from the six 80 km × 80 km blocks within this subregion, while panels (c)-(h) display the sample-level scatter of the second-order component for each block.

In southern Oklahoma, the SM-CAPE interaction represents the strongest and most spatially organized second-order pattern. Three of the six blocks pass the F-test and explain 3.7%–13.2% of the afternoon precipitation variance. Although some block-to-block variability remains in panels (c)-(h), the functional shapes are qualitatively consistent across most of the subregion and indicate a robust dry-soil-high-CAPE synergistic regime. Specifically, afternoon precipitation is enhanced when below-normal SM co-occurs with elevated CAPE, whereas the interaction weakens or disappears when CAPE is low. The regional-mean pattern in panel (b) highlights this behavior clearly: CAPE provides the thermodynamic potential for convection, while dry soil contributes the triggering mechanism through enhanced sensible heating and more vigorous boundary layer growth, allowing parcels to access the stored instability aloft. In this sense, the interaction term captures a conditional regime in which the realization of convective potential depends on land-surface forcing, rather than CAPE alone. This SM-CAPE result helps explain an important feature of the Texas-southern Great Plains transition zone. While the first-order analysis already showed that local land and atmospheric states can independently organize afternoon precipitation, the second-order analysis indicates that in southern Oklahoma part of

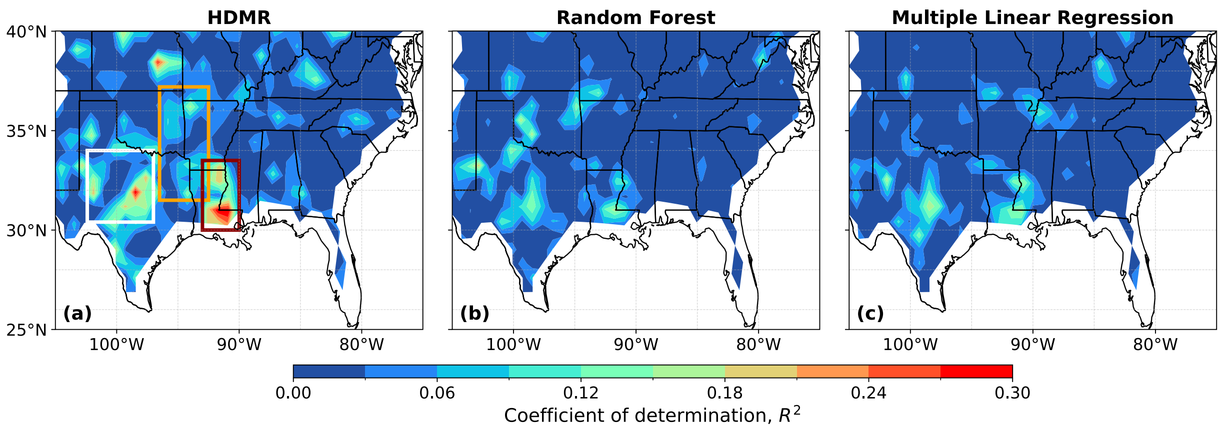

Figure 7. Spatial distribution of model performance (R^2) for (a) High-Dimensional Model Representation, (b) Random Forest, and (c) MLR. All models utilize the same vector of morning precursors: soil moisture (SM), land-surface temperature (LST), CAPE, PW, and leaf area index. The colored rectangles in (a) delineate the different functional regimes identified in this study, where precipitation is primarily governed by SM (orange, 92.5°W–97°W, 32°N–37°N), LST (red box, 90°W–93°W, 30°N–33°N), and convective available potential energy (white, 97°W–103°W, 30°N–34°N).

the precipitation variability arises specifically from the joint occurrence of dry soils and a convectively favorable atmosphere. Thus, the strongest second-order effect is not simply an extension of the first-order SM signal, but an additional synergistic mode that becomes active only when the atmospheric instability is sufficiently large.

HDMR also identifies SM-LST and SM-PW second-order effects (Figures S4-S5 in Supporting Information S1), but these are localized and less significant compared to the SM-CAPE regime. The component functions show consistent shapes across several blocks, suggesting physically meaningful but spatially scattered patterns. The SM-LST interaction exhibits a contrasting pattern to SM-CAPE, but points to the same underlying mechanism of surface-heating control over eastern Texas. Extremely wet soil combined with low LST suppresses precipitation by reducing sensible heating and limiting boundary layer growth (LaFleur et al., 2023; Matus et al., 2023). The SM-PW interaction shows a localized moisture-limited regime over the Oklahoma-Arkansas intersection sub-region, in which local SM becomes more influential when atmospheric moisture is low, whereas its marginal effect weakens under high-PW conditions (Ford, Rapp, et al., 2015). However, most blocks fail to pass the second-order significance test during the 2020 test period. We therefore interpret these SM-PW and SM-LST effects as localized, pair-dependent modulations rather than robust regional second-order regimes, indicating that first-order effects remain the dominant controls over most hotspot regions.

Finally, to validate the fidelity of the derived couplings, we benchmark the predictive skill of the HDMR expansion (Equation 2) against RF and MLR (Figure 7). HDMR consistently outperforms the competing methods, achieving coefficients of determination (R^2) between 15% and 25% across the key coupling hot spots. The R^2 values are spatially coherent, aligning with the distinct functional regimes highlighted in Figure 7a: the transition zones dominated by SM (orange box), the energy-limited regions controlled by LST (red box), and the capped environments driven by CAPE (white box). In addition, a localized LAI contribution appears over southeastern Kansas (also presented in Figure S3 in Supporting Information S1). Previous studies suggest that vegetation may locally modulate precipitation variability through its influence on surface energy partitioning and boundary layer development (Gao et al., 2024; Lauwaet et al., 2009). On the contrary, both RF and MLR moderately resolve these signals, typically explaining less than 15% of the variance. Importantly, our intent is not to position HDMR as a superior predictive model. Its F-test-based postprocessing deliberately filters out statistically insignificant component functions, which is not feasible in a standard operational prediction. Instead, this comparison underscores HDMR's strength as an interpretive framework: it retains only those nonlinear, individual, and cooperative structures that are supported by the training-testing data, thus extracting robust SMPC regions. The resulting component functions therefore provide a physically grounded and transparent characterization of how morning land-surface and atmospheric states jointly shape afternoon precipitation, lending confidence to the SMPC patterns identified in this study.

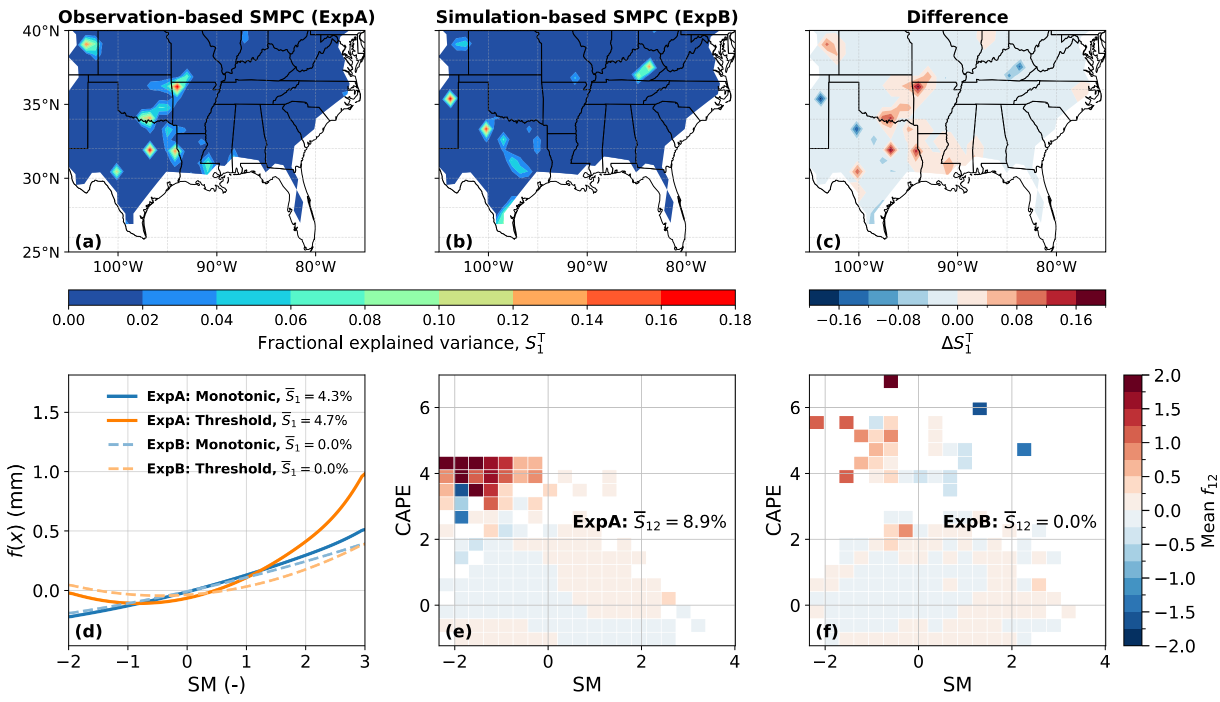

Figure 8. Evaluation of simulated Soil moisture-precipitation coupling against the observation-based baseline. Panels (a–c) compare the total-effect (S_1^T) of morning soil moisture (SM) on afternoon precipitation, showing the observation-based result derived from Soil Moisture Active Passive L4/Multi-Radar Multi-Sensor (ExpA), the convection-permitting CONUS404 simulation (ExpB), and their difference, respectively. Panel (d) compares the regional-mean first-order SM component functions for the two dominant functional classes, monotonic and threshold-like, for ExpA and ExpB. Panels (e–f) show the bin-averaged second-order interaction component between SM and convective available potential energy for ExpA and ExpB, respectively, summarizing the dry-soil-high-CAPE synergistic regime at the subregional scale. Text annotations of $\bar{S}_1$ and $\bar{S}_{12}$ in the bottom row indicate the regional-mean explained variance of precipitation by the first-order SM effect and the second-order SM-CAPE interaction, respectively, computed over the subregions delineated in Figures 4a and 6a.

4.4. SMPC Diagnosis of Climate Model Simulations

To demonstrate HDMR's utility as a model diagnostic tool, we evaluate the representation of SMPC within the convection-permitting CONUS404 simulation (ExpB). By replicating the experimental design of ExpA, but replacing the training data with WRF-simulated SM and precipitation for 2016–2019, we assess whether the model reproduces the observed coupling patterns relative to the observation-based baseline. Figures 8a–8c compares the total-effect (S_1^T) of morning SM on afternoon precipitation. A pronounced discrepancy is evident: CONUS404 systematically underestimates coupling strength across the central United States. Most notably, the prominent coupling hot spot observed near the Texas-Louisiana and Oklahoma-Arkansas intersections in the observation-based analysis is largely absent in the simulation, where the model yields near-zero total effects over much of the same region.

This underestimation is further reflected in the first-order component functions. Instead of comparing individual selected blocks, Figure 8d summarizes the regional mean first-order SM response for the two dominant functional classes identified in the observations, namely the monotonic and threshold-like types. The simulation reproduces the qualitative shapes of these two SM response modes reasonably well, indicating that CONUS404 captures the basic form of the direct SM influence on convection. However, the coupling magnitude is substantially weaker, especially under wet-soil conditions. For the threshold-like regime, the observation-based analysis (ExpA) indicates that positive SM anomalies can enhance afternoon precipitation by about 1.0 mm at the regional scale, whereas the corresponding increase in ExpB is only about 0.4 mm. More broadly, although the model retains the qualitative functional structure of the first-order SM effect, its explained variance is greatly reduced and often becomes statistically insignificant, consistent with the weak total-effect values in panels (a)-(c).

A similar conclusion emerges for the second-order interaction between SM and CAPE. Figures 8e and 8f show the bin-averaged f_{12} surfaces for ExpA and ExpB, respectively, summarizing the regional SM-CAPE interaction discussed earlier. In ExpA, the dominant pattern reflects a dry-soil-high-CAPE synergistic regime, in which elevated CAPE provides the convective potential while dry soil contributes the triggering through enhanced surface heating; under this joint regime, afternoon precipitation can be enhanced by about 2.0–3.0 mm in the regional mean. CONUS404 reproduces the same qualitative structure, indicating that the model captures the physical sign of the interaction. However, the simulated f_{12} surface is clearly weaker in amplitude, implying a reduced synergistic enhancement of precipitation (0–1.0 mm) and a diminished contribution to precipitation variability. Thus, both the first-order and second-order diagnostics point to the same conclusion: CONUS404 captures the broad functional character of the observed land-atmosphere coupling, but substantially underestimates its magnitude, particularly in the regimes where SM anomalies most strongly modulate afternoon convection. This finding corroborates recent studies suggesting that convection-permitting models may suppress the sensitivity of convection to land-surface anomalies (Lee & Hohenegger, 2024; Qin et al., 2025), reflecting uncertainties in coupled land-atmosphere modeling, including land-surface flux partitioning and boundary layer processes.

5. Discussion and Conclusions

Accurately quantifying the coupling between SM and precipitation is fundamental for improving hydroclimate predictability, yet it remains challenged by the inherent nonlinearity and multivariate nature of land-atmosphere interactions. This study applies the HDMR framework to both observation-based data sets over CONUS and the convection-permitting CONUS404 simulation, decomposing the structural, correlative, and cooperative controls of morning precursors on afternoon rainfall. We first use the observation-based analysis to identify the dominant physical regimes of SMPC, and then evaluate whether CONUS404 reproduces these coupling patterns and strengths. We detail our main findings as follows.

5.1. Reconciling SMPC Strength via HDMR Decomposition

A primary contribution of this study is the definition of comprehensive coupling indices and component functions derived from HDMR, which together quantify nonlinear and synergistic SMPC. Our analysis demonstrates that traditional linear metrics, including correlation analysis and MLR, consistently underestimate coupling strength, attributing only 6%–8% of precipitation variance to SM in regions where HDMR identifies total contributions of up to about 20%. In particular, HDMR shows that the first-order structural SM effect alone can explain up to about 13% of afternoon precipitation variance in localized hot spots across the eastern Great Plains transition zone. Rather than acting as a continuous linear driver, SM primarily operates through organized functional regimes: a threshold-like wet-soil response over eastern and northeastern Texas, where precipitation increases rapidly once SM exceeds its mean state, and a monotonic response near the Texas-Louisiana and Oklahoma-Arkansas intersections, where precipitation increases nearly proportionally with SM. By isolating these structural effects, HDMR reveals a stronger and more physically interpretable SMPC signal than conventional linear approaches.

5.2. Distinct First-Order Physical Regimes

The first-order HDMR decomposition reveals that afternoon precipitation is governed by spatially distinct direct controls from SM, CAPE, and LST. These controls do not represent a single uniform SMPC behavior, but instead reflect different land-atmosphere regimes across the central United States. The first regime is a moisture-limited transition-zone regime, where SM exerts a direct positive influence on afternoon precipitation. This regime is most evident over eastern Texas and near the Texas-Louisiana and Oklahoma-Arkansas intersections. The regional-mean functions indicate that wet SM anomalies can enhance afternoon precipitation by up to about 1.0 mm in the threshold-like regime, with a monotonic increase of roughly 0.16 mm per standard deviation of SM in the more linear regime.

A second regime is governed primarily by atmospheric instability. Across Texas, morning CAPE explains up to about 18% of afternoon precipitation variance and exhibits monotonic and saturating positive responses. The monotonic response indicates that increasing morning buoyancy favors afternoon convection, whereas the saturating response suggests that additional CAPE produces diminishing precipitation increases once instability is already large. This behavior is consistent with a capped or conditionally unstable environment, where convective realization also depends on triggering, moisture availability, and inhibition. Quantitatively, increases in morning CAPE translate to an additional 0.3–0.6 mm of afternoon rainfall.

A third regime appears over the humid subtropical environment of Louisiana, where LST becomes the dominant first-order control and explains up to about 15% of afternoon precipitation variance. The LST response is dominated by a saturated negative form: below-average LST enhances afternoon precipitation by up to about 1.5 mm, whereas warmer conditions exert little direct influence. This behavior is consistent with a cool-surface, moisture-favorable regime, in which lower morning LST may indicate higher EF, lower LCL, or cloud/shallow-frontal preconditioning favorable for later convection.

5.3. The Localized Synergistic Effects

Beyond direct effects, our results show that second-order interactions provide additional but more localized modulation of SMPC. The clearest and most spatially organized cooperative signal is the SM-CAPE interaction over southern Oklahoma. Three of the six 80 km × 80 km blocks pass the *F*-test, with the interaction explaining 3.7%–13.2% of afternoon precipitation variance. The corresponding component functions reveal a coherent dry-soil-high-CAPE synergistic regime: elevated CAPE provides the thermodynamic potential for convection, while dry soils enhance sensible heating and boundary-layer growth, helping parcels access the stored instability aloft. Thus, this second-order term captures a conditional regime in which convective realization depends on the joint state of land-surface forcing and atmospheric instability, rather than on either variable alone.

HDMR also detects localized SM-PW and SM-LST interactions, but these are much less spatially coherent than the SM-CAPE regime. Their component functions suggest physically plausible moisture-limited and surface-heating-related modulation, respectively; however, most blocks fail to pass the significance test during the 2020 test period. We therefore treat these effects as localized, pair-dependent modulations, while first-order effects remain dominant over most hotspot regions.

Taken together, the first- and second-order results reveal that within Type A coupling (temporal precipitation amount coupling), positive and negative SM effects can coexist through different physical pathways. The first-order SM effect primarily captures a direct moisture supply pathway (a wet-soil advantage), whereas the second-order SM-CAPE interaction reveals a conditional thermodynamic-triggering pathway (a dry-soil advantage) that emerges exclusively under high-instability conditions. This finding is conceptually consistent with previous studies of precipitation probability and spatial preference, which show that the apparent sign of SMPC depends on background soil wetness, spatial heterogeneity, and coupling definition (Guillod et al., 2015; Hsu et al., 2017; Lee et al., 2026; Taylor et al., 2012). However, those studies primarily address precipitation probability and/or spatial preference (Types c and d), whereas our analysis quantifies temporal precipitation amount coupling (Type a). Therefore, the comparison is intended to highlight consistent physical mechanisms rather than to imply a direct mathematical equivalence between different coupling metrics.

5.4. Implications for Convection-Permitting Modeling

Finally, the application of HDMR as a diagnostic tool reveals critical deficiencies in state-of-the-art climate modeling. While the convection-permitting CONUS404 simulation qualitatively reproduces the bimodal sign of the coupling (positive in the eastern Great Plains, negative in southern Oklahoma), it suffers from a systematic intensity bias. The model significantly dampens the functional response of precipitation to land-surface anomalies, underestimating the magnitude of both the wet-soil advantage and the dry-soil thermal trigger. This suggests that, despite convection-permitting resolution, processes controlling land-atmosphere coupling are not fully resolved or constrained, introducing uncertainty in land-surface flux partitioning, boundary-layer turbulence, entrainment, and the resulting balance between surface and atmospheric forcing. Correcting this dampening is crucial, as underestimated coupling strength may affect the model's ability to represent drought persistence, heatwave amplification, and sub-daily feedbacks between surface moisture, boundary-layer growth, and convective triggering under a warming climate.

We note that this model diagnosis is conducted on a common 16 km analysis grid to ensure consistent spatial support between the observation-based and model-based experiments. This choice is necessary because the observational SM product has a coarser native resolution than CONUS404. However, the spatial resolution of the analysis may influence the diagnosed magnitude of SMPC (Yuan et al., 2020). Thus, the reported underestimation of SMPC strength in CONUS404 should be interpreted within this 16 km grid framework. A native-resolution (4 km) assessment of CONUS404 would require a redesigned collocation, downscaling/remapping, and sampling strategy to address the mismatch between the 4 km model grid and the coarser SMAP SM resolution, and remains an important direction for future work.

We acknowledge that the temporal scope of this study (2016–2020) is constrained by the availability of high-quality gauge-corrected radar observations (MRMS), and the regional SMPC is validated only for the 2020 test year, limiting a direct assessment of interannual and interdecadal variability. However, the robust detection of nonlinear and synergistic coupling regimes during this period provides a critical benchmark for disentangling and quantifying the complex SMPC. Our SMPC analysis focuses on a specific diurnal lag, namely, 7:00 a.m. land-atmosphere conditions on same-day afternoon precipitation. However, SMPC is not constrained to a single timescale or lag structure (Stacke & Hagemann, 2016; Y. Wang et al., 2018; Zhao et al., 2021; Taylor et al., 2024). 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.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Availability Statement

The SMAP L4 product is obtained from the National Snow and Ice Data Center (Reichle et al., 2022), and the CONUS404 data set from the National Center for Atmospheric Research Research Data Archive (Rasmussen et al., 2023b). The MRMS Gauge-Corrected QPE is available through the Iowa Environmental Mesonet at https://mesonet.agron.iastate.edu/archive/mrms.php. The HDMR software used for the sensitivity analysis is publicly available from its GitHub repository at https://github.com/jaspervrugt/HDMR. After installing the HDMR package, the MATLAB scripts developed for this study can be used to reproduce the analyses presented here. These study-specific scripts, together with the final processed SMAP, MRMS, and CONUS404 data sets, have been archived in Zenodo (Gao et al., 2026) under a Creative Commons Attribution 4.0 International license.

Acknowledgments

Yifu Gao wishes to thank the support of the Henry Samueli Fellowship provided during the last year of his PhD studies by the Department of Civil and Environmental Engineering at the Samueli School of Engineering at UC Irvine.

References

Alfieri, L., Claps, P., D'Odorico, P., Laio, F., & Over, T. M. (2008). An analysis of the soil moisture feedback on convective and stratiform precipitation. *Journal of Hydrometeorology*, *9*(2), 280–291. https://doi.org/10.1175/2007JHM863.1

Barlage, M., Chen, F., Rasmussen, R., Zhang, Z., & Miguez-Macho, G. (2021). The importance of scale-dependent groundwater processes in land-atmosphere interactions over the central United States. *Geophysical Research Letters*, *48*(5), e2020GL092171. https://doi.org/10.1029/2020GL092171

Bevacqua, E., Rakovec, O., Schumacher, D. L., Kumar, R., Thober, S., Samaniego, L., et al. (2024). Direct and lagged climate change effects intensified the 2022 European drought. *Nature Geoscience*, *17*(11), 1–8. https://doi.org/10.1038/s41561-024-01559-2

Budyko, M. (1974). *Climate and life* (Vol. 18). Academic Press.

Chen, Y., Wang, A., & Feng, G. (2023). The relative contribution of large-scale circulation and local soil moisture to summer precipitation over Asian mid-low latitudes. *Climate Dynamics*, *60*(7), 2097–2112. https://doi.org/10.1007/s00382-022-06402-1

de Burgh-Day, C. O., & Leeuwenburg, T. (2023). Machine learning for numerical weather and climate modelling: A review. *Geoscientific Model Development*, *16*(22), 6433–6477. https://doi.org/10.5194/gmd-16-6433-2023

Dong, X., Chen, H., Zhou, Y., Hsu, P., & Zhang, W. (2024). Disentangling the complexities of how underlying surface thermal factors influence July precipitation in Eastern China. *Journal of Climate*, *37*(19), 5105–5129. https://doi.org/10.1175/JCLI-D-23-0748.1

Duerinck, H. M., Van der Ent, R. J., Van de Giesen, N. C., Schoups, G., Babovic, V., & Yeh, P. J. F. (2016). Observed soil moisture–precipitation feedback in Illinois: A systematic analysis over different scales. *Journal of Hydrometeorology*, *17*(6), 1645–1660. https://doi.org/10.1175/JHM-D-15-0032.1

Ek, M. B., & Holtslag, A. A. M. (2004). Influence of soil moisture on boundary layer cloud development. *Journal of Hydrometeorology*, *5*(1), 86–99. https://doi.org/10.1175/1525-7541(2004)005⟨0086:IOSMOB⟩2.0.CO;2

Eltahir, E. A. B. (1998). A soil moisture–rainfall feedback mechanism: 1. Theory and observations. *Water Resources Research*, *34*(4), 765–776. https://doi.org/10.1029/97WR03499

Falchi, A., Minisci, E., Kubicek, M., Vasile, M., & Lemmens, S. (2018). HDMR-based sensitivity analysis and uncertainty quantification of GOCE aerodynamics using DSMC. *Stardust Final Conference*, 301–323. https://doi.org/10.1007/978-3-319-69956-1_18

Ferguson, C. R., Wood, E. F., & Vinukollu, R. K. (2012). A global intercomparison of modeled and observed land–atmosphere coupling. *Journal of Hydrometeorology*, *13*(3), 749–784. https://doi.org/10.1175/JHM-D-11-0119.1

Findell, K. L., & Eltahir, E. A. B. (2003a). Atmospheric controls on soil moisture–boundary layer interactions. Part I: Framework development. *Journal of Hydrometeorology*, *4*, 552–569. https://doi.org/10.1175/1525-7541(2003)004⟨0552:ACOSML⟩2.0.CO;2

Findell, K. L., & Eltahir, E. A. B. (2003b). Atmospheric controls on soil moisture–boundary layer interactions. Part II: Feedbacks within the continental United States. *Journal of Hydrometeorology*, *4*(3), 570–583. https://doi.org/10.1175/1525-7541(2003)004⟨0570:ACOSML⟩2.0.CO;2

Findell, K. L., Gentine, P., Lintner, B. R., & Kerr, C. (2011). Probability of afternoon precipitation in eastern United States and Mexico enhanced by high evaporation. *Nature Geoscience*, *4*(7), 434–439. https://doi.org/10.1038/ngeo1174

Ford, T. W., Steiner, J., Mason, B., & Quiring, S. M. (2023). Observation-Driven characterization of soil moisture-precipitation interactions in the central United States. *Journal of Geophysical Research: Atmospheres*, *128*(12), e2022JD037934. https://doi.org/10.1029/2022JD037934

Ford, T. W., Quiring, S. M., Frauenfeld, O. W., & Rapp, A. D. (2015). Synoptic conditions related to soil moisture-atmosphere interactions and unorganized convection in Oklahoma. *Journal of Geophysical Research: Atmospheres*, *120*(22), 11–519. https://doi.org/10.1002/2015jd023975

Ford, T. W., Rapp, A. D., Quiring, S. M., & Blake, J. (2015). Soil moisture–precipitation coupling: Observations from the Oklahoma mesonet and underlying physical mechanisms. *Hydrology and Earth System Sciences*, *19*(8), 3617–3631. https://doi.org/10.5194/hess-19-3617-2015

Gaal, R., & Kinter III, J. L. (2021). Soil moisture influence on the incidence of summer mesoscale convective systems in the US Great Plains. *Monthly Weather Review*, *149*(12), 3981–3994. https://doi.org/10.1175/MWR-D-21-0140.1

Gao, Y., Guilloteau, C., Foufoula-Georgiou, E., Xu, C., Sun, X., & Vrugt, J. A. (2024). Soil moisture-cloud-precipitation feedback in the lower atmosphere from functional decomposition of satellite observations. *Geophysical Research Letters*, *51*(22), e2024GL110347. https://doi.org/10.1029/2024GL110347

Gao, Y., Sahin, A., & Vrugt, J. A. (2023). Probabilistic sensitivity analysis with dependent variables: Covariance-based decomposition of hydrologic models. *Water Resources Research*, *59*(4), e2022WR032834. https://doi.org/10.1029/2022WR032834

Gao, Y., Vrugt, J. A., Foufoula-Georgiou, E., Chen, H., & Li, R. (2026). Contributions of land-atmosphere variables to warm-season afternoon precipitation over CONUS: HDMR functional decomposition of observational and simulated datasets (Version 1.0) [Software]. *Zenodo*. https://doi.org/10.5281/zenodo.20605840

Guillod, B. P., Orlowsky, B., Miralles, D. G., Teuling, A. J., & Seneviratne, S. I. (2015). Reconciling spatial and temporal soil moisture effects on afternoon rainfall. *Nature Communications*, *6*(1), 6443. https://doi.org/10.1038/ncomms7443

Guo, Z., Dirmeyer, P. A., Koster, R. D., Sud, Y. C., Bonan, G., Oleson, K. W., et al. (2006). GLACE: The global land–atmosphere coupling experiment. Part II: Analysis. *Journal of Hydrometeorology*, *7*(4), 611–625. https://doi.org/10.1175/JHM511.1

Habibi, H., Awal, R., Fares, A., & Temimi, M. (2021). Performance of Multi-Radar Multi-Sensor (MRMS) product in monitoring precipitation under extreme events in Harris County, Texas. *Journal of Hydrology*, *598*, 126385. https://doi.org/10.1016/j.jhydrol.2021.126385

Heinze, R., Dipankar, A., Henken, C. C., Moseley, C., Sourdeval, O., Trömel, S., et al. (2017). Large-eddy simulations over Germany using ICON: A comprehensive evaluation. *Quarterly Journal of the Royal Meteorological Society*, *143*(702), 69–100. https://doi.org/10.1002/qj.2947

Homma, T., & Saltelli, A. (1996). Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering and System Safety*, *52*(1), 1–17. https://doi.org/10.1016/0951-8320(96)00002-6

Hong, S. Y., Noh, Y., & Dudhia, J. (2006). A new vertical diffusion package with an explicit treatment of entrainment processes. *Monthly Weather Review*, *134*(9), 2318–2341. https://doi.org/10.1175/MWR3199.1

Hooker, G. (2007). Generalized functional ANOVA diagnostics for high-dimensional functions of dependent variables. *Journal of Computational & Graphical Statistics*, *16*(3), 709–732. https://doi.org/10.1198/106186007X237892

Hsu, H., Lo, M., Guillod, B. P., Miralles, D. G., & Kumar, S. (2017). Relation between precipitation location and antecedent/subsequent soil moisture spatial patterns. *Journal of Geophysical Research: Atmospheres*, *122*(12), 6319–6328. https://doi.org/10.1002/2016JD026042

Hu, H., Leung, L. R., & Feng, Z. (2021). Early warm-season mesoscale convective systems dominate soil moisture–precipitation feedback for summer rainfall in central United States. *Proceedings of the National Academy of Sciences*, *118*(43), e2105260118. https://doi.org/10.1073/pnas.2105260118

Huang, Y., Siems, S. T., Manton, M. J., & Thompson, G. (2014). An evaluation of WRF simulations of clouds over the Southern Ocean with A-Train observations. *Monthly Weather Review*, *142*(2), 647–667. https://doi.org/10.1175/MWR-D-13-00128.1

Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., & Collins, W. D. (2008). Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *Journal of Geophysical Research*, *113*(D13). https://doi.org/10.1029/2008JD009944

Jiang, Y., & Wang, G. (2023). A new approach to soil initialization for studying subseasonal land-atmosphere interactions. *Journal of Advances in Modeling Earth Systems*, *15*(11), e2023MS003822. https://doi.org/10.1029/2023MS003822

Kolendowicz, L., Taszarek, M., & Czernecki, B. (2017). Atmospheric circulation and sounding-derived parameters associated with thunderstorm occurrence in Central Europe. *Atmospheric Research*, *191*, 101–114. https://doi.org/10.1016/j.atmosres.2017.03.009

Koster, R. D., Dirmeyer, P. A., Guo, Z., Bonan, G., Chan, E., Cox, P., et al. (2004). Regions of strong coupling between soil moisture and precipitation. *Science*, *305*(5687), 1138–1140. https://doi.org/10.1126/science.1100217

Koster, R. D., Liu, Q., Mahanama, S. P. P., & Reichle, R. H. (2018). Improved hydrological simulation using SMAP data: Relative impacts of model calibration and data assimilation. *Journal of Hydrometeorology*, *19*(4), 727–741. https://doi.org/10.1175/JHM-D-17-0228.1

Kucherenko, S., Feil, B., Shah, N., & Mauntz, W. (2011). The identification of model effective dimensions using global sensitivity analysis. *Reliability Engineering and System Safety*, *96*(4), 440–449. https://doi.org/10.1016/j.ress.2010.11.003

LaFleur, A. T., Tanamachi, R. L., Dawson, D. T., & Turner, D. D. (2023). Factors affecting the rapid recovery of CAPE on 31 March 2016 during VORTEX-Southeast. *Monthly Weather Review*, *151*(6), 1459–1477. https://doi.org/10.1175/MWR-D-22-0051.1

Lauwaet, D., van Lipzig, N. P. M., & De Ridder, K. (2009). The effect of vegetation changes on precipitation and Mesoscale Convective Systems in the Sahel. *Climate Dynamics*, *33*(4), 521–534. https://doi.org/10.1007/s00382-009-0539-2

Lee, J., & Hohenegger, C. (2024). Weaker land–atmosphere coupling in global storm-resolving simulation. *Proceedings of the National Academy of Sciences*, *121*(12), e2314265121. https://doi.org/10.1073/pnas.2314265121

Lee, J., Warnau, S., Lim, K. S., Benedict, I., & van Heerwaarden, C. (2026). Spatial and temporal preferences for afternoon precipitation over morning soil moisture in global storm-resolving simulations. *Geoscience Letters*, *13*(1), 13. https://doi.org/10.1186/s40562-026-00461-6

Levine, P. A., Randerson, J. T., Swenson, S. C., & Lawrence, D. M. (2016). Evaluating the strength of the land–atmosphere moisture feedback in Earth system models using satellite observations. *Hydrology and Earth System Sciences*, *20*(12), 4837–4856. https://doi.org/10.5194/hess-20-4837-2016

Li, G., & Rabitz, H. (2010). D-MORPH regression: Application to modeling with unknown parameters more than observation data. *Journal of Mathematical Chemistry*, *48*(4), 1010–1035. https://doi.org/10.1007/s10910-010-9722-2

Li, G., & Rabitz, H. (2012). General formulation of HDMR component functions with independent and correlated variables. *Journal of Mathematical Chemistry*, *50*(1), 99–130. https://doi.org/10.1007/s10910-011-9898-0

Li, G., Rabitz, H., Yelvington, P. E., Oluwole, O. O., Bacon, F., Kolb, C. E., & Schoendorf, J. (2010). Global sensitivity analysis for systems with independent and/or correlated inputs. *The Journal of Physical Chemistry A*, *114*(19), 6022–6032. https://doi.org/10.1021/jp9096919

Liu, C., Ikeda, K., Rasmussen, R., Barlage, M., Newman, A. J., Prein, A. F., et al. (2017). Continental-scale convection-permitting modeling of the current and future climate of North America. *Climate Dynamics*, *49*(1–2), 71–95. https://doi.org/10.1007/s00382-016-3327-9

Liu, W., Zhang, Q., Li, C., Xu, L., & Xiao, W. (2022). The influence of soil moisture on convective activity: A review. *Theoretical and Applied Climatology*, *149*(1–2), 221–232. https://doi.org/10.1007/s00704-022-04046-z

Matus, S. A., Dominguez, F., & Ford, T. W. (2023). Land and atmosphere conditions prior to extreme Great Plains low-level jets. *Journal of Hydrometeorology*, *24*(4), 783–800. https://doi.org/10.1175/JHM-D-22-0045.1

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

Moazami, S., & Najafi, M. R. (2021). A comprehensive evaluation of GPM-IMERG V06 and MRMS with hourly ground-based precipitation observations across Canada. *Journal of Hydrology*, *594*, 125929. https://doi.org/10.1016/j.jhydrol.2020.125929

Niu, G. Y., Yang, Z. L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., et al. (2011). The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *Journal of Geophysical Research*, *116*(D12), D12109. https://doi.org/10.1029/2010JD015139

O'Gorman, P. A., & Dwyer, J. G. (2018). Using machine learning to parameterize moist convection: Potential for modeling of climate, climate change, and extreme events. *Journal of Advances in Modeling Earth Systems*, *10*(10), 2548–2563. https://doi.org/10.1029/2018MS001351

Qiao, L., Zuo, Z., Zhang, R., Piao, S., Xiao, D., & Zhang, K. (2023). Soil moisture–atmosphere coupling accelerates global warming. *Nature Communications*, *14*(1), 4908. https://doi.org/10.1038/s41467-023-40641-y

Qin, H., Alvee, F. M., Leung, L. R., Sun, X., & Wang, G. (2025). Performance of convection-permitting and convection-parameterized models in reproducing the extreme precipitation intensity relationship with surface conditions. *Journal of Climate*, *1*(aop), 4611–4623. https://doi.org/10.1175/JCLI-D-24-0589.1

Rabitz, H., & Aliş, Ö. F. (1999). General foundations of high-dimensional model representations. *Journal of Mathematical Chemistry*, *25*(2), 197–233. https://doi.org/10.1023/A:1019188517934

Rappin, E. D., Mahmood, R., Nair, U. S., & Pielke Sr., R. A. (2022). Land–atmosphere interactions during GRAINEX: Planetary boundary layer evolution in the presence of irrigation. *Journal of Hydrometeorology*, *23*(9), 1401–1417. https://doi.org/10.1175/JHM-D-21-0160.1

Rasmussen, R. M., Chen, F., Liu, C., Ikeda, K., Prein, A., Kim, J., et al. (2023b). CONUS404: Four-kilometer long-term regional hydroclimate reanalysis over the conterminous United States (ver. 2.0, December 2023 [Dataset]. *U.S. Geological Survey data release*. https://doi.org/10.5066/P9PHPK4F

Rasmussen, R. M., Chen, F., Liu, C. H., Ikeda, K., Prein, A., Kim, J., et al. (2023a). CONUS404: The NCAR–USGS 4-km long-term regional hydroclimate reanalysis over the CONUS. *Bulletin of the American Meteorological Society*, *104*(8), E1382–E1408. https://doi.org/10.1175/BAMS-D-21-0326.1

Reichle, R. H., De Lannoy, G., Koster, R. D., Crow, W. T., Kimball, J. S., & Liu, Q. (2022). SMAP L4 global 3-hourly 9 km EASE-Grid surface and root zone soil moisture analysis update version 7 [Dataset]. *NASA National Snow and Ice Data Center Distributed Active Archive Center*. https://doi.org/10.5067/LWJ6TF5SZRG3

Reichle, R. H., De Lannoy, G. J. M., Liu, Q., Koster, R. D., Kimball, J. S., Crow, W. T., et al. (2017). Global assessment of the SMAP Level-4 surface and root-zone soil moisture product using assimilation diagnostics. *Journal of Hydrometeorology*, *18*(12), 3217–3237. https://doi.org/10.1175/JHM-D-17-0130.1

Santanello, J. A., Friedl, M. A., & Ek, M. B. (2007). Convective planetary boundary layer interactions with the land surface at diurnal time scales: Diagnostics and feedbacks. *Journal of Hydrometeorology*, *8*(5), 1082–1097. https://doi.org/10.1175/JHM614.1

Santanello, J. A., Jr., Dirmeyer, P. A., Ferguson, C. R., Findell, K. L., Tawfik, A. B., Berg, A., et al. (2018). Land–atmosphere interactions: The LoCo perspective. *Bulletin of the American Meteorological Society*, *99*(6), 1253–1272. https://doi.org/10.1175/BAMS-D-17-0001.1

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

Shereena, O. A., & Rao, B. N. (2019). HDMR-Based bayesian structural System identification. In *Recent advances in structural engineering* (Vol. 1, pp. 453–464). Springer. https://doi.org/10.1007/978-981-13-0362-3_36

Smith, T. M., Lakshmanan, V., Stumpf, G. J., Ortega, K. L., Hondl, K., Cooper, K., et al. (2016). Multi-Radar Multi-Sensor (MRMS) severe weather and aviation products: Initial operating capabilities. *Bulletin of the American Meteorological Society*, *97*(9), 1617–1630. https://doi.org/10.1175/BAMS-D-14-00173.1

Sobol, I. M. (1993). Sensitivity estimates for nonlinear mathematical models. *Mathematical Modelling in Civil Engineering*, *1*(4), 407–414.

Song, F., Feng, Z., Leung, L. R., Houze Jr., R. A., Wang, J., Hardin, J., & Homeyer, C. R. (2019). Contrasting spring and summer large-scale environments associated with mesoscale convective systems over the US Great Plains. *Journal of Climate*, *32*(20), 6749–6767. https://doi.org/10.1175/JCLI-D-18-0839.1

Stacke, T., & Hagemann, S. (2016). Lifetime of soil moisture perturbations in a coupled land–atmosphere simulation. *Earth System Dynamics*, *7*(1), 1–19. https://doi.org/10.5194/esd-7-1-2016

Stephens, G. L., L'Ecuyer, T., Forbes, R., Gettelmen, A., Golaz, J. C., Bodas-Salcedo, A., et al. (2010). Dreary state of precipitation in global models. *Journal of Geophysical Research*, *115*(D24). https://doi.org/10.1029/2010JD014532

Su, T., Li, Z., & Zheng, Y. (2023). Cloud-surface coupling alters the morning transition from stable to unstable boundary layer. *Geophysical Research Letters*, *50*(5), e2022GL102256. https://doi.org/10.1029/2022GL102256

Sun, J., Yang, K., He, X., Wang, G., Wang, Y., Yu, Y., & Lu, H. (2025). Causal pathways underlying global soil moisture–precipitation coupling. *Nature Communications*, *16*(1), 8935. https://doi.org/10.1038/s41467-025-63999-7

Taszarek, M., Brooks, H. E., Czernecki, B., Szuster, P., & Fortuniak, K. (2018). Climatological aspects of convective parameters over Europe: A comparison of era-interim and sounding data. *Journal of Climate*, *31*(11), 4281–4308. https://doi.org/10.1175/JCLI-D-17-0596.1

Tavakol, A., Rahmani, V., Quiring, S. M., & Kumar, S. V. (2019). Evaluation analysis of NASA SMAP L3 and L4 and SPoRT-LIS soil moisture data in the United States. *Remote Sensing of Environment*, *229*, 234–246. https://doi.org/10.1016/j.rse.2019.05.006

Taylor, C. M. (2015). Detecting soil moisture impacts on convective initiation in Europe. *Geophysical Research Letters*, *42*(11), 4631–4638. https://doi.org/10.1002/2015GL064030

Taylor, C. M., de Jeu, R. A. M., Guichard, F., Harris, P. P., & Dorigo, W. A. (2012). Afternoon rain more likely over drier soils. *Nature*, *489*(7416), 423–426. https://doi.org/10.1038/nature11377

Taylor, C. M., Gounou, A., Guichard, F., Harris, P. P., Ellis, R. J., Couvreux, F., & De Kauwe, M. (2011). Frequency of Sahelian storm initiation enhanced over mesoscale soil-moisture patterns. *Nature Geoscience*, *4*(7), 430–433. https://doi.org/10.1038/ngeo1173

Taylor, C. M., Klein, C., & Harris, B. L. (2024). Multiday soil moisture persistence and atmospheric predictability resulting from sahelian mesoscale convective systems. *Geophysical Research Letters*, *51*(20), e2024GL109709. https://doi.org/10.1029/2024GL109709

Teuling, A. J., Hirschi, M., Ohmura, A., Wild, M., Reichstein, M., Ciais, P., et al. (2009). A regional perspective on trends in continental evaporation. *Geophysical Research Letters*, *36*(2). https://doi.org/10.1029/2008GL036584

Thompson, G., & Eidhammer, T. (2014). A study of aerosol impacts on clouds and precipitation development in a large winter cyclone. *Journal of the Atmospheric Sciences*, *71*(10), 3636–3658. https://doi.org/10.1175/JAS-D-13-0305.1

Thompson, R. L., Mead, C. M., & Edwards, R. (2007). Effective storm-relative helicity and bulk shear in supercell thunderstorm environments. *Weather and Forecasting*, *22*(1), 102–115. https://doi.org/10.1175/WAF969.1

Tuttle, S. E., & Salvucci, G. D. (2017). Confounding factors in determining causal soil moisture-precipitation feedback. *Water Resources Research*, *53*(7), 5531–5544. https://doi.org/10.1002/2016WR019869

Van Leeuwen, P. J., DeCaria, M., Chakraborty, N., & Pulido, M. (2021). A framework for causal discovery in non-intervenable systems. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *31*(12), 123128. https://doi.org/10.1063/5.0054228

Wang, A., Chen, Y., Li, M., Gao, Y., & Ma, X. (2025). A review of soil moisture data in China: Definitions, sources, and applications in hydrometeorology research. *Bulletin of the American Meteorological Society*, *106*(6), E1108–E1129. https://doi.org/10.1175/BAMS-D-24-0197.1

Wang, G., Fu, R., Zhuang, Y., Dirmeyer, P. A., Santanello, J. A., Wang, G., et al. (2024). Influence of lower-tropospheric moisture on local soil moisture–precipitation feedback over the US Southern Great Plains. *Atmospheric Chemistry and Physics*, *24*(6), 3857–3868. https://doi.org/10.5194/acp-24-3857-2024

Wang, H., Chen, L., Ye, F., & Chen, L. (2017). Global sensitivity analysis for fiber reinforced composite fiber path based on D-MORPH-HDMR algorithm. *Structural and Multidisciplinary Optimization*, *56*(3), 697–712. https://doi.org/10.1007/s00158-017-1681-9

Wang, Y., Yang, J., Chen, Y., De Maeyer, P., Li, Z., & Duan, W. (2018). Detecting the causal effect of soil moisture on precipitation using convergent cross mapping. *Scientific Reports*, *8*(1), 12171. https://doi.org/10.1038/s41598-018-30669-2

Welty, J., & Zeng, X. (2018). Does soil moisture affect warm season precipitation over the southern Great Plains? *Geophysical Research Letters*, *45*(15), 7866–7873. https://doi.org/10.1029/2018GL078598

White, P., & Nelson, P. A. (2024). Evaluation of sub-hourly MRMS quantitative precipitation estimates in mountainous terrain using machine learning. *Water Resources Research*, *60*(12), e2024WR037437. https://doi.org/10.1029/2024WR037437

Yin, J., Albertson, J. D., Rigby, J. R., & Porporato, A. (2015). Land and atmospheric controls on initiation and intensity of moist convection: CAPE dynamics and LCL crossings. *Water Resources Research*, *51*(10), 8476–8493. https://doi.org/10.1002/2015WR017286

Yuan, S., Wang, Y., Quiring, S. M., Ford, T. W., & Houston, A. L. (2020). A sensitivity study on the response of convection initiation to in situ soil moisture in the central United States. *Climate Dynamics*, *54*(3–4), 2013–2028. https://doi.org/10.1007/s00382-019-05098-0

Zhang, C., Li, Q., Liang, X., Dong, L., Xie, B., Li, W., & Sun, C. (2025). Impacts of land surface processes on summer extreme precipitation in Eastern China: Insights from CWRF simulations. *Atmospheric Research*, *314*, 107783. https://doi.org/10.1016/j.atmosres.2024.107783

Zhang, J., Howard, K., Langston, C., Kaney, B., Qi, Y., Tang, L., et al. (2016). Multi-Radar Multi-Sensor (MRMS) quantitative precipitation estimation: Initial operating capabilities. *Bulletin of the American Meteorological Society*, *97*(4), 621–638. https://doi.org/10.1175/BAMS-D-14-00174.1

Zhao, J., Wang, A., & Wang, H. (2021). Soil moisture memory and its relationship with precipitation characteristics in China Region. *Chinese Journal of Atmospheric Sciences*, *45*(4), 799–818. https://doi.org/10.3878/j.issn.1006-9895.2007.20149

Zhou, S., Williams, A. P., Lintner, B. R., Berg, A. M., Zhang, Y., Keenan, T. F., et al. (2021). Soil moisture–atmosphere feedbacks mitigate declining water availability in drylands. *Nature Climate Change*, *11*(1), 38–44. https://doi.org/10.1038/s41558-020-00945-z

Zhou, S., Williams, A. P., Lintner, B. R., Findell, K. L., Keenan, T. F., Zhang, Y., & Gentine, P. (2022). Diminishing seasonality of subtropical water availability in a warmer world dominated by soil moisture–atmosphere feedbacks. *Nature Communications*, *13*(1), 5756. https://doi.org/10.1038/s41467-022-33473-9

References From the Supporting Information

Golub, G. H., & Van Loan, C. F. (2013). *Matrix computations*. JHU press.

Penrose, R. (1955). A generalized inverse for matrices. *Proceedings of the Cambridge Philosophical Society*, *51*(3), 406–413. https://doi.org/10.1017/S0305004100030401

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

Supporting Information for "Systematic Underestimation of Nonlinear and Synergistic Soil Moisture-Precipitation Coupling in Convection-Permitting Models"

Yifu Gao[1,2,3 *], Runze Li[3], Haishan Chen[1,2],Efi Foufoula-Georgiou[3,4], and

Jasper A. Vrugt[3,4]

[1] State Key Laboratory of Climate System Prediction and Risk Management/Key Laboratory of Meteorological Disaster, Ministry of Education/Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters, Nanjing University of Information Science and Technology, Nanjing, Jiangsu, China.

[2] School of Atmospheric Sciences, Nanjing University of Information Science and Technology, Nanjing, Jiangsu, China.

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

[4]Department of Earth System Science, University of California, Irvine, California, USA.

Contents of this file

---

[*]Corresponding author:

gaoyifu@nuist.edu.cn

Introduction

This Supporting Information contains additional details and figures that complement our main manuscript. Specifically:

- **Text S1** provides a benchmark validation using a synthetic Light Use Efficiency model, where HDMR accurately reconstructs ecosystem productivity ($R^2 = 97.6\%$) and correctly identifies nonlinear drivers that linear methods miss. This confirms the framework's ability to disentangle complex structural and cooperative interactions, establishing confidence for its application to the more complex problem of precipitation coupling.

- **Text S2** details the mathematical formulation of HDMR, where component functions are approximated using orthonormal polynomials. The unknown expansion coefficients are solved using Diffeomorphic Modulation (D-MORPH) regression, a specialized least-squares method that enforces hierarchical orthogonality and vanishing conditions. This ensures the resulting decomposition is unique and physically interpretable, satisfying the mathematical requirements for accurate variance attribution.

- **Text S3** illustrates the functional classification procedure for first-order component functions. Each first-order component function is interpolated onto a common standardized-anomaly grid and classified using amplitude, end-to-end change, and slope behavior across the anomaly range.

- **Figures S1 to S3** illustrate extended HDMR results not shown in the main text, including the minimal contributions of LAI and wind speed/direction, the second-order cooperative indices for SM with other variables, and the total fractional variance explained by each non-SM variable (LST, PW, LAI, u_{10}, v_{10}) across the CONUS.

These materials provide further insight into how different land–atmosphere variables contribute to afternoon precipitation variability, reinforcing or extending our primary findings on the dominant soil moisture–precipitation coupling (SMPC) across the Great Plains.

Text S1. Case Study: Simplified Light Use Efficiency Model

We employ a simplified Light Use Efficiency (LUE) model to estimate gross primary productivity (GPP) [gC/m/day]. Formally, we write:

$$\mathrm{GPP} = \varepsilon_0 \mathrm{R_n} \cdot \mathrm{PAR} \cdot f(\mathrm{SM}) g(\mathrm{T_{2m}}) h(\mathrm{u}), \tag{S1}$$

where ε_0 is the baseline (maximum) LUE [gC/MJ], $\mathrm{R_n}$ refers to the net radiation [$\mathrm{W/m^2}$], PAR denotes the fraction of net radiation that is photosynthetically active radiation [-]. $f(\mathrm{SM})$, $g(\mathrm{T_{2m}})$, and $h(\mathrm{u})$ are dimensionless nonlinear functions that impose SM stress, temperature dependence, and wind-speed effects, respectively:

$$f(\mathrm{SM}) = \begin{cases} 0 & \text{if } \mathrm{SM} < \mathrm{SM_w} \\ \dfrac{\mathrm{SM} - \mathrm{SM_w}}{\mathrm{SM_f} - \mathrm{SM_w}} & \text{if } \mathrm{SM_w} \leq \mathrm{SM} < \mathrm{SM_f} \\ 1 & \text{if } \mathrm{SM} \geq \mathrm{SM_f} \end{cases} \tag{S2}$$

$$g(\mathrm{T_{2m}}) = \exp\left[-\left(\frac{\mathrm{T_{2m}} - 25}{10} \right)^2 \right], \tag{S3}$$

$$h(\mathrm{u}) = \frac{1}{1 + \exp\left[-0.5(\mathrm{u} - 2)\right]}, \tag{S4}$$

where $\mathrm{SM_w}$ and $\mathrm{SM_f}$ are the soil wilting point [$\mathrm{m^3\,m^{-3}}$] and field capacity [$\mathrm{m^3\,m^{-3}}$], assumed to be 0.1 and 0.3, respectively, and u is determined by $\sqrt{\mathrm{u}_{10}^2 + \mathrm{v}_{10}^2}$. This simplified LUE model directly incorporates SM and other land-surface conditions as a dominant control on GPP.

To implement HDMR, we select one grid cell (97.15°N, 38.80°W) in the Central US from the CONUS404 dataset. We extract June–September hourly data during the 2020 for the four relevant variables $(x_1, \ldots, x_4)^\top = [\mathrm{SM}, \mathrm{R_n}, \mathrm{T_{2m}}, \mathrm{u}]^\top$, plus a prescribed ε_0 (1.8 gC/MJ) and PAR fraction (0.45). Note that $\mathrm{R_n}$ is not available directly from CONUS404. We simulate the $\mathrm{R_n}$ following the method of Zotarelli et al. (2010). The June–September period provides an hourly input matrix $\mathbf{X}$ of size $N \times 4$. We split these samples chronologically into a 70% training set followed by a 30% testing set. Standardization is applied only after the split, using statistics from the training period. Substituting the standardized inputs into Equation (S1) yields the corresponding hourly GPP record, which captures the diurnal to seasonal variability of ecosystem productivity.

We then feed **X** and the synthetic GPP series into the HDMR framework, aiming to quantify the structural, correlative, and cooperative contributions of SM, R_n, T_{2m}, and u to plant productivity. Such an approach highlights the multi-variable relationships that more basic correlation or regression methods may overlook, particularly under nonlinear or coupled conditions. As we will show next, these simpler methods can miss crucial interactive contribution, for instance, how R_n and T_{2m} jointly modulate GPP, thereby underestimating the scope of land-atmosphere interactions that ultimately shape ecosystem productivity.

All three methods (HDMR, correlation, and MLR) agree that R_n exerts the greatest control on GPP. However, significant discrepancies arise regarding the magnitude of this control and the role of secondary variables. Given that the HDMR model achieves a near-perfect reconstruction of the observed GPP time series (Figure S1c), we regard its variance attribution as the most accurate characterization of input importance. Against this benchmark, the limitations of linear methods become evident. MLR generally underestimates the drivers, capturing only half of the wind effect (7.8% vs. 16.5% in HDMR) and completely failing to detect the influence of LST (0% vs. 9.2% in HDMR). Conversely, while simple correlation aligns better with HDMR regarding wind and temperature, it fails to capture the contribution of SM (0% vs. 7.1% in HDMR). To probe the source of these discrepancies, we dissect HDMR's structural and correlative components on the top-right panel of Figure S1. SM, for example, exhibits a sizable first-order structural effect ($S_1^a = 10.3\%$) and a strong second-order structural interaction with net radiation ($S_{12}^a = 9.6\%$), but these contributions are partially offset by negative correlative effects arising from its correlation with other variables ($S_1^b = -3.8\%$ and $S_{12}^b = -6.8\%$). Such compensating nonlinear and cooperative influences cannot be represented by correlation or linear regression, explaining why these simpler methods mischaracterize the true importance of SM and other drivers.

These discrepancies highlight fundamental methodological limitations. Simple correlation is not a Global Sensitivity Analysis (GSA) method. It characterizes only pairwise associations, rendering it unsuitable for quantitative multivariate variance attribution in complex systems. Similarly, the failure of linear regression is confirmed by the predictive performance shown in Figure S1c. While HDMR achieves a near-perfect coefficient of determination ($R^2 = 97.6\%$), MLR performs worse ($R^2 = 64.5\%$) and exhibits systematic bias, underestimating diurnal peaks and overestimating nocturnal lows. This confirms that linear assumptions misattribute GPP variance to SM and wind, whereas HDMR provides the physically faithful decomposition required to identify the true meteorological controls.

Text S2. Polynomial Construction of Component Functions and D-MORPH Regression

We represent each component function with a set of orthogonal polynomial expansions following Li and Rabitz (2012). Specifically, for a single dimension x_i, we define a family of polynomials:

$$\begin{array}{ccc} \phi_1(x_i) = a_1 x_i + a_0 & \phi_2(x_i) = b_2 x_i^2 + b_1 x_1 + b_0 & \phi_3(x_i) = c_3 x_i^3 + c_2 x_i^2 + c_1 x_i + c_0 \\ \text{degree } p = 1 & \text{degree } p = 2 & \text{degree } p = 3, \end{array} \tag{S5}$$

where $p = 3$ is selected (Gao et al., 2023) and 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

$$f_i(x_i) = \sum_{r=1}^{p} \alpha_r^{(i)i} \phi_r(x_i) \tag{S6a}$$

$$f_{ij}(x_i, x_j) = \sum_{r=1}^{p} \left[\alpha_r^{(ij)i} \phi_r(x_i) + \alpha_r^{(ij)j} \phi_r(x_j)\right] + \sum_{r=1}^{p} \sum_{s=1}^{p} \beta_{rs}^{(ij)ij} \phi_r(x_i) \phi_s(x_j), \tag{S6b}$$

where the extended bases of the second-order component functions will help satisfy the vanishing condition in Equation (3). The use of extended bases has implications for our index notation of the coefficients. Here, 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 equations (S6a) and (S6b) into equation (1) we get a closed-form expression for the relationship between the land-atmosphere variables $\mathbf{x} = (x_1, \ldots, x_d)^\top$ and precipitation y with total number of unknown expansion coefficients equal to $l = d \cdot p + 12d(d-1)(2p+p^2)$. The terms of this explicit relationship account for the marginal and cooperative effects of the input variables in determining the magnitude and sign of SMPC. 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/hr) of the individual contribution of SM to precipitation.

Let $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(N)}$ represent the N training samples of the input vector and $y^{(1)}, \ldots, y^{(N)}$ denote the corresponding precipitation, then we can write equation (2) in matrix form $\mathbf{\Phi c} = \mathbf{b}$ and yield

$$\mathbf{\Phi} = \begin{bmatrix} \boldsymbol{\phi}(\mathbf{x}^{(1)})^\top \\ \vdots \\ \boldsymbol{\phi}(\mathbf{x}^{(N)})^\top \end{bmatrix} \qquad \text{and} \qquad \mathbf{b} = \begin{bmatrix} y^{(1)} - y_0 \\ \vdots \\ y^{(N)} - y_0 \end{bmatrix}, \tag{S7a}$$

where $\boldsymbol{\phi}(\mathbf{x})^\top$ is a $1 \times l$ design vector with orthonormalized polynomial functions of equation (S6) (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 the measured/simulated $y^{(i)}$ and mean y_0 precipitation for each training sample, $i = (1, \ldots, N)$.

To estimate the expansion coefficients, we determine the corresponding least squares values $\widehat{\mathbf{c}}_{\text{ls}}$

$$\widehat{\mathbf{c}}_{\text{ls}} = (\mathbf{\Phi}^\top \mathbf{\Phi})^\dagger \mathbf{d}, \tag{S8}$$

where the $l \times (l - d \cdot p)$ matrix $(\mathbf{\Phi}^\top \mathbf{\Phi})^\dagger$ is the generalized pseudo inverse of the $l \times l$ Gramian matrix, $\mathbf{G} = \mathbf{\Phi}^\top \mathbf{\Phi}$, which satisfies all four Moore-Penrose conditions (Penrose, 1955; Golub & Van Loan, 2013) 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 $\mathbf{\Phi}^\top \mathbf{b}$ without the first $d \cdot p$ rows. Diffeomorphic Modulation (dm) under Observable Response Preserving Homotopy (D-MORPH) regression (Li & Rabitz, 2010) enforces hierarchical orthogonality of the component functions in pursuit of the optimum coefficients

$$\widehat{\mathbf{c}}_{\mathrm{dm}} = \mathbf{V}_{l-r}(\mathbf{U}_{l-r}^{\top}\mathbf{V}_{l-r})\mathbf{U}_{l-r}^{\top}\widehat{\mathbf{c}}_{\mathrm{ls}}, \tag{S9}$$

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 $\mathbf{PB} = \mathbf{U\Sigma V}^{\top}$ 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 polynomial functions. This latter matrix $\mathbf{B}$ enforces the relaxed vanishing condition of Equation (3) (Li & Rabitz, 2010), and matrix $\mathbf{I}_l$ is the $l \times l$ identity matrix and r is the number of nonzero singular values.

Text S3. Functional classification of first-order component functions

To facilitate physical interpretation of the first-order HDMR component functions, we classify each statistically significant $f_i(x_i)$ into a small number of functional types. For each variable, the raw component function values are first sorted by the standardized anomaly x_i, duplicate x_i values are averaged, and the resulting curve is interpolated onto a common anomaly grid. A short moving-average smoothing is then applied to reduce small-scale sampling noise while preserving the overall functional shape.

The classification is based on simple shape descriptors of the interpolated curve. We compute the total amplitude,

$$A_i = \max f_i - \min f_i, \tag{S10}$$

the end-to-end change,

$$\Delta_i = f_i(x_{\max}) - f_i(x_{\min}). \tag{S11}$$

We also calculate the fraction of positive and negative local slopes and the mean slopes over the lower, middle, and upper thirds of the anomaly range. Curves with amplitudes below the 25th percentile of all amplitudes are assigned to the "else" category.

The remaining curves are classified as follows. A monotonic function requires all local slopes to have the same sign, with a monotonic fraction of 1.00. A threshold-like function requires a weak response at low or middle anomalies but a stronger response at high anomalies: the high-anomaly slope must exceed 0.15 of the local maximum slope and be at least 1.20 times larger than the earlier slopes. A saturated function requires a strong lower- or middle-anomaly response followed by flattening at high anomalies: the curve must be mostly monotonic, with a monotonic fraction of at least 0.75, and the high-anomaly slope must be no larger than 0.60 of the earlier slope magnitude. Each classified curve is also assigned a sign according to the direction of its end-to-end response, yielding monotonic (+/−), threshold-like (+/−), saturated (+/−), and else categories.

This classification is intended as a compact diagnostic of the dominant physical form of the first-order response rather than a separate statistical model. It allows the regional patterns of $f_i(x_i)$ to be summarized in terms of interpretable functional regimes, such as monotonic wet-soil enhancement, threshold-like responses, or saturated surface-temperature effects.

References

Gao, Y., Sahin, A., & Vrugt, J. A. (2023). Probabilistic sensitivity analysis with dependent variables: Covariance-based decomposition of hydrologic models. *Water Resources Research*, *59*(4), e2022WR032834. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2022WR032834` doi: 10.1029/2022WR032834

Golub, G. H., & Van Loan, C. F. (2013). *Matrix computations.* JHU press.

Li, G., & Rabitz, H. (2010). D-MORPH regression: application to modeling with unknown parameters more than observation data. *Journal of Mathematical Chemistry*, *48*, 1010–1035. doi: 10.1007/s10910-010-9722-2

Li, G., & Rabitz, H. (2012). General formulation of HDMR component functions with independent and correlated variables. *Journal of Mathematical Chemistry*, *50*(1), 99-130. doi: 10.1007/s10910-011-9898-0

Penrose, R. (1955). A generalized inverse for matrices. *Proceedings of the Cambridge Philosophical Society*, *51*(3), 406-413. doi: 10.1017/S0305004100030401

Zotarelli, L., Dukes, M. D., Romero, C. C., Migliaccio, K. W., & Morgan, K. T.

(2010). Step by step calculation of the Penman-Monteith Evapotranspiration (FAO-56 Method). *Institute of Food and Agricultural Sciences. University of Florida*, *8*.

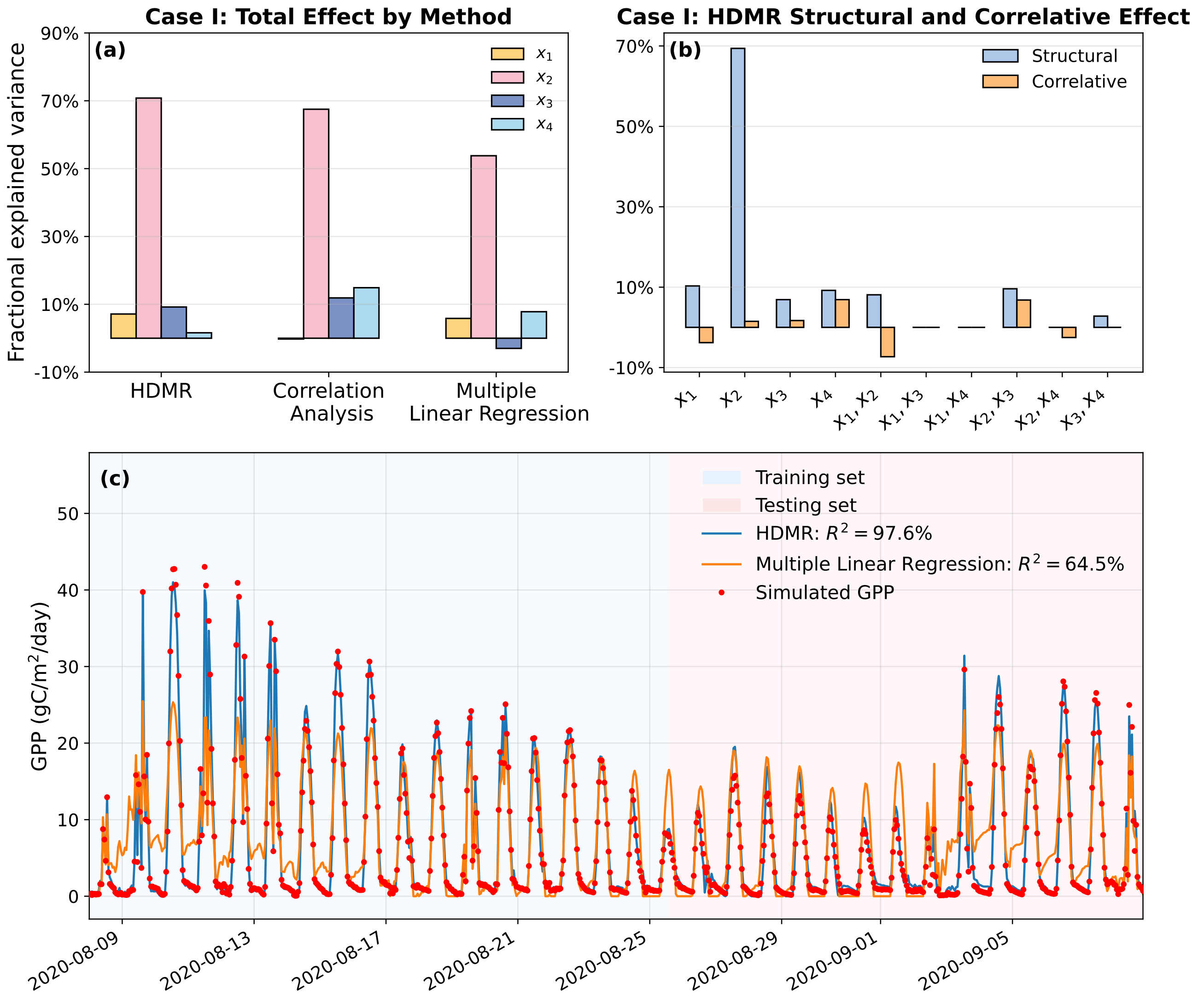

Figure S1. Top row (test period): (a) total fractional variance of GPP explained by four variables $(x_1, \ldots, x_4)^\top = [\mathrm{SM}, \mathrm{R_n}, \mathrm{T_{2m}}, \mathrm{u}]^\top$ under HDMR, correlation, and linear regression methods, and (b) the partition of structural versus correlative effects derived from the HDMR framework. Bottom row: (c) time series comparison of hourly GPP generated by Equation (S1) (red dots, serving as the reference) against predictions from HDMR (solid blue line) and MLR (solid orange line). Shaded blue and orange regions denote the training and testing periods, respectively.

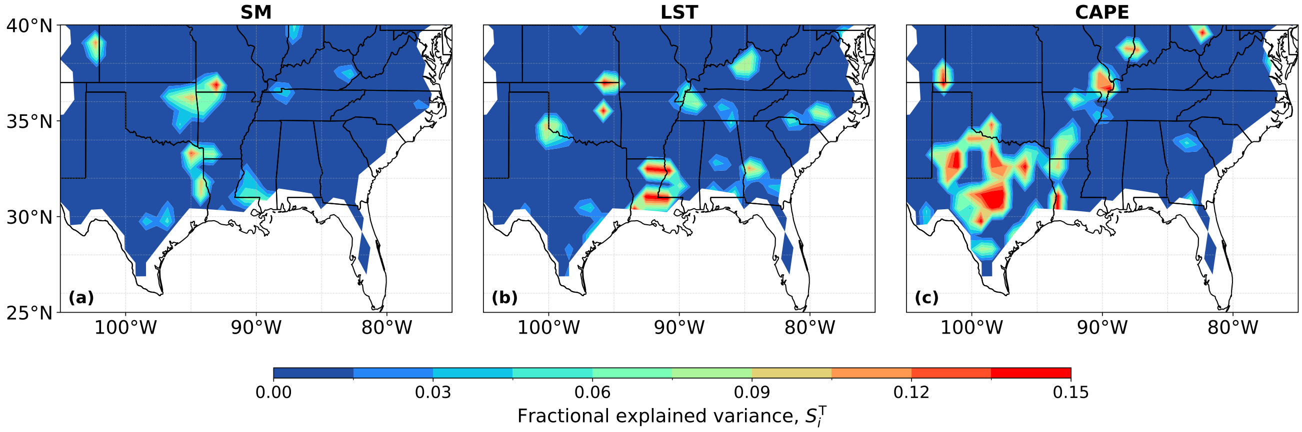

Figure S2. Spatial distribution of the total fractional variance, S_i^{T}, of warm season (April–September 2016-2019, training period) afternoon precipitation explained by morning SM, CAPE, and LST.

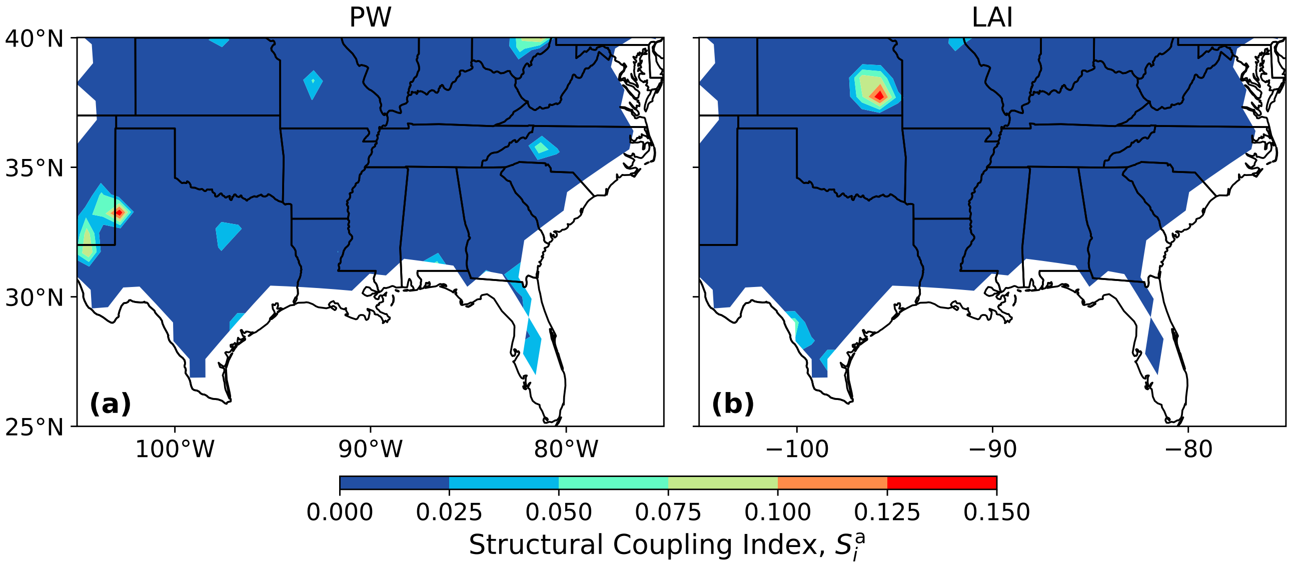

Figure S3. Spatial distribution of the structural coupling index, S_i^a, of PW and LAI, representing the fractional explained variance of warm season (April–September 2020, testing period) afternoon precipitation explained by the direct effects of these two variables.