Geophysical Research Letters

RESEARCH LETTER
10.1029/2022GL101942

Finding Causal Gateways of Precipitation Over the Contiguous United States

Xueli Yang[1], Zhi-Hua Wang[1], Chenghao Wang[2,3], and Ying-Cheng Lai[4,5]

[1]School of Sustainable Engineering and the Built Environment, Arizona State University, Tempe, AZ, USA, [2]School of Meteorology, University of Oklahoma, Norman, OK, USA, [3]Department of Geography and Environmental Sustainability, University of Oklahoma, Norman, OK, USA, [4]School of Electricity, Computer and Energy Engineering, Arizona State University, Tempe, AZ, USA, [5]Department of Physics, Arizona State University, Tempe, AZ, USA

Key Points:

- We use the convergent cross mapping algorithm (CCM), based on embedding theory, for causality inference in this study
- CCM is used to detect causal influence in precipitation perturbations among different climate regions of US
- The Ohio Valley region emerges as a causal gateway of moisture transport and propagation of regional precipitation anomalies in the US

Supporting Information:

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

Correspondence to:

Z.-H. Wang,
zhwang@asu.edu

Citation:

Yang, X., Wang, Z.-H., Wang, C., & Lai, Y.-C. (2023). Finding causal gateways of precipitation over the contiguous United States. *Geophysical Research Letters*, *50*, e2022GL101942. https://doi.org/10.1029/2022GL101942

Received 30 OCT 2022
Accepted 4 FEB 2023

Author Contributions:

Conceptualization: Zhi-Hua Wang, Ying-Cheng Lai
Data curation: Xueli Yang, Chenghao Wang
Formal analysis: Xueli Yang
Funding acquisition: Zhi-Hua Wang
Investigation: Xueli Yang
Methodology: Xueli Yang, Zhi-Hua Wang, Ying-Cheng Lai
Project Administration: Zhi-Hua Wang
Software: Xueli Yang
Supervision: Zhi-Hua Wang
Validation: Xueli Yang

Abstract Identifying regions that mediate regional propagation of atmospheric perturbations is important to assessing the susceptibility and resilience of complex hydroclimate systems. Detecting the regional gateways through causal inference, can help unravel the interplay of physical processes and inform projections of future changes. In this study, we characterize the causal interactions among nine climate regions in the contiguous United States using long-term (1901–2018) precipitation data. The constructed causal networks reveal the cross-regional propagation of precipitation perturbations. Results show that the Ohio Valley region acts as an atmospheric gateway for precipitation and moisture transport in the U.S., which is largely regulated by the regional convective uplift. The findings have implications for improving predicative capacity of hydroclimate modeling of regional precipitation.

Plain Language Summary Successful detection of causality in complex systems is important to unraveling the underlying mechanisms of system dynamics. The dynamic interactions in Earth's climate system are often nonlinear, weakly or moderately coupled, and essentially non-separable, which renders conventional approaches of causal inference, such as statistical correlation or Granger causality, infeasible or ineffective. Here we applied the convergent cross mapping method to detect causal influence among different climate regions in the contiguous U.S. in response to precipitation perturbations. The results of our study show that the Ohio Valley region, as an atmospheric convergence zone, acts as a regional gateway and mediator for the long-term precipitation perturbations in the U.S. The temporal evolution of causal effect and susceptibility exhibits superposition of climate variability at various time scales, highlighting the impact of prominent climate variabilities such as El Niño–Southern Oscillation on the dynamics of causality.

1. Introduction

The Earth system comprises numerous nonlinear subsystems that interact with each other dynamically in a complex way. Understanding the interactions and the underlying causal mechanisms of nonlinear components is of crucial importance to tasks such as refining physical schemes in Earth system models, reducing model biases and uncertainties, and improving weather predictions and climate projections (Shepherd, 2014). Conventional statistical approaches such as correlation- and regression-based methods have been widely used to topology of the Earth system, especially connectivity over long spatial distance, known as *teleconnections* (Boers et al., 2019). However, conventional statistics-based methods are often unable to unravel the true causal mechanisms (Pearl & Mackenzie, 2018; Runge, Bathiany, et al., 2019). In addition, spurious correlations between variables are common even in simple nonlinear systems (Mysterud et al., 2001). As a result, for natural systems ubiquitously governed by nonlinear dynamics, the causation inferred from linear correlations can be misleading or erroneous (Good et al., 2015; Sugihara & May 1990).

In the literature, the classical Granger causality (GC) paradigm has been prevailing for identifying causation in time series (Granger, 1969). However, the GC approach can be prone to significant errors (McCann et al., 1998; Sugihara et al., 2012) when applied to dynamics systems consisting of weakly or moderately coupled subsystems, especially when the interactions are forced by shared external, strong drivers that can lead to apparent synchrony (Moran, 1953). Another fundamental difficulty of GC is the requirement of extensive length of time series to generate meaningful causal inference.

Visualization: Xueli Yang, Chenghao Wang
Writing – original draft: Xueli Yang, Zhi-Hua Wang, Chenghao Wang, Ying-Cheng Lai
Writing – review & editing: Xueli Yang, Zhi-Hua Wang, Chenghao Wang, Ying-Cheng Lai

To overcome the limitations of GC, researchers have developed and tested causality algorithms specifically suitable for nonlinear dynamic systems with moderate coupling, the convergent cross mapping (CCM) method being a representative one (Jiang et al., 2016; Kretschmer et al., 2016; Runge, 2018; Runge et al., 2015; Runge, Bathiany, et al., 2019; Runge, Nowack, et al., 2019; Sugihara et al., 2012). The CCM method is based on the classic Taken's delay-coordinate embedding theory for reconstructing the phase space of the underlying nonlinear system from time series (Deyle & Sugihara, 2011; Kantz & Schreiber, 1997; Lai & Ye, 2003; Packard et al., 1980; Sauer et al., 1991; Sugihara & May 1990; Takens, 1981). The CCM and other similar causal inference methods have been successfully applied to detect dynamic causality in Earth's hydroclimate system (Ombadi et al., 2020; Shi et al., 2022; Wang et al., 2018; Yang, Wang, Wang, et al., 2022).

In this paper, we adopt the CCM framework to detect and quantify hydroclimatic causal interactions among different regions over the contiguous United States (CONUS) from long-term observational precipitation data sets. We find that the Ohio Valley region acts as a causal gateway for regional precipitation and atmospheric transport in the CONUS, which is largely regulated by the regional convective uplift. This finding will help to improve the predictive capacity of hydroclimate modeling by incorporating the causal inference in dynamic processes. In addition, with the increasing availability of data from measurements and climate models, causal inference in climate systems will facilitate the development of data-driven and system-based frameworks for integrated Earth system research (Fan et al., 2021; Wang & Wang, 2020). One example is that the time evolution of regional causality contains signals resulting from multi-scale climate variability, which has the potential to serve as early warning signs to presage critical transitions in complex hydroclimate systems (Yang, Wang, & Wang, 2022).

2. Methods

2.1. Data Retrieval and Treatment

In this study, we retrieved the monthly mean precipitation over the period 1901–2018 (1,416 months) from Climatic Research Unit (CRU) Time-Series (TS) version 4.03 (https://catalogue.ceda.ac.uk/uuid/10d3e3640f004c578403419aac167d82), produced by the Center for Environmental Data Analysis (CEDA) Archive. The $0.5° \times 0.5°$ spatially gridded data set covers the spatial domain of global land surfaces except Antarctica, with in total 3,288 grid cells over the entire CONUS. The observational data are anormalized using 1961–1990 monthly averages for each gridcell (Harris et al., 2020). Individual grid points, if there are too close to each other, often contain similar information (monthly precipitation herein) and may not be sufficiently representative to interpret potential propagating perturbations. To better represent distinct sub-regional processes, we aggregate all grids within each of the nine climatically consistent regions (or climate regions) in the CONUS following NOAA's division, as shown in Figure 2a to obtain the average regional-scale time series (Kretschmer et al., 2016). These climatic regions are Northwest (NW), West (WE), Southwest (SW), Northern Rockies and Plains (NRP), South (SO), Upper Midwest (UM), Ohio Valley (OV), Southeast (SE), and Northeast (NE). The division of climate regions is defined by the National Centers for Environmental Information of NOAA (https://www.ncdc.noaa.gov/monitoring-references/maps/us-climate-regions.php). There are previous studies revealing that spatial coherence patterns existed after the dimension reduction of the long-term gridded monthly precipitation based on methods such as principal components analysis (PCA) (Karl & Koscielny, 1982; Vejmelka et al., 2015; Walsh et al., 1982). The spatial coherence patterns or the corresponding nine principal components (PC) identified are consistent with a series of documented cyclone trajectories (Walsh et al., 1982; Zishka & Smith, 1980). In addition, the nine identifiable patterns of drought (calculated from averaged monthly precipitation and temperature) are delineated in the United States performed by PCA, with each PC is characterized by a distinct annual oscillation of monthly precipitation (Karl & Koscielny, 1982). The nine identifiable components are similar to the climatic division defined by the NOAA. We then removed the seasonal cycle by subtracting monthly averages from the regional temperature time series during the study period to minimize the impact of seasonality. The detrended time series of temperature anomalies are used for subsequent causality analysis.

2.2. Convergent Cross Mapping Algorithm

The CCM is based on simple projection (Sugihara et al., 2012), a nearest-neighbor algorithm that involves kernel density estimation of nearby points on the reconstructed or shadow manifolds (Takens, 1981). To assess the potential causation between two climate regions, we construct two shadow manifolds using time-lagged

coordinates of historical precipitation series (Deyle & Sugihara, 2011; Kantz & Schreiber, 1997; Lai & Ye, 2003; Packard et al., 1980; Sauer et al., 1991; Sugihara & May 1990; Takens, 1981) and estimate precipitation anomalies in one region using information from another region. A shadow manifold can be reconstructed using the delay-coordinate embedding method. In particular, for a scalar time series $X(t)$ of length L from a specific climate region, an E-dimensional time-delayed vector $\boldsymbol{x}(t) = [X(t), X(t\text{-}\tau), \ldots, X(t-(E-1)\tau)]$ can be formed from $t = 1 + (E-1)\tau$ to $t = L$, with τ and E the time delay and embedding dimension, respectively, to construct the shadow manifold M_X (likewise for $\boldsymbol{y}(t)$ and M_Y for a scalar time series of $Y(t)$).

The cross-mapping estimate of $Y(t)$, denoted as $\hat{Y}(t)|M_X$, is based on a simple projection of the $E+1$ nearest neighbors of vector $\boldsymbol{x}(t)$ in the manifold M_X. Here $E+1$ is the minimum number of data points required for a bounded simplex in the E-dimensional space. The time indices of those E+1 neighbors $\boldsymbol{x}(t_1)$, $\boldsymbol{x}(t_2)$,…, $\boldsymbol{x}(t_{E+1})$ (from closest to farthest) in M_X are used to identify the corresponding putative neighbors in Y, that is, $Y(t_1)$, $Y(t_2)$,…, $Y(t_{E+1})$. The cross-mapping estimate of $Y(t)$ is then determined using the weighted average as

$$\hat{Y}(t)|M_X = \sum_{i=1}^{E+1} w_i(t) \cdot Y(t_i), \tag{1}$$

where the weight vector $w_i(t)$ is estimated by

$$w_i(t) = \frac{u_i(t)}{\sum_{j=1}^{E+1} u_i(t)}, \tag{2}$$

with

$$u_i(t) = \exp\left\{-\frac{d[x(t), x(t_i)]}{d[x(t), x(t_1)]}\right\}, \tag{3}$$

and $d[x(t), x(t_i)]$ the Euclidean distance between the two vectors $x(t)$ and $x(t_i)$ in M_X, which can be calculated as the length of the line segment between points $x(t)$ and $x(t_i)$ in an E-dimensional Euclidean space.

If $X(t)$ and $Y(t)$ are dynamically coupled and causally related, the nearest neighbors of M_X should identify the time indices of corresponding nearest neighbors on the attractor manifold of M_Y, and vice versa (Sugihara et al., 2012). Consequently, $\hat{Y}(t)|M_X$ should converge to $Y(t)$, and $\hat{X}(t)|M_Y$ to $X(t)$. To measure the causality from Y to X, the correlation coefficient $\rho_{Y|M_X}$ between the original $Y(t)$ and the cross-mapping estimate $\hat{Y}(t)|M_X$ will be used, which is defined by

$$\rho_{Y|M_X} = \frac{\mathbf{E}\left\{[Y(t) - \mu_Y] \cdot \left[\hat{Y}(t)|M_X - \mu_{\hat{Y}}\right]\right\}}{\sigma_Y \sigma_{\hat{Y}}}, \tag{4}$$

where $\mathbf{E}$, μ, and σ are the statistical expectation, average, and standard deviation, respectively.

A stronger causal influence of $Y(t)$ on $X(t)$ indicates that $X(t)$ contains "more" of $Y(t)$, thereby making more accurate the prediction of $Y(t)$ with information about $X(t)$. Thus, a larger value of the correlation coefficient $\rho_{Y|M_X}$ signifies a stronger dynamical causal influence of $Y(t)$ on $X(t)$. Practically, an empirical threshold specified with different significance levels using, for example, t-test, can be set to determine if $Y(t)$ has causal influence on $X(t)$ (Jiang et al., 2016). If $\rho_{Y|M_X} \leq 0$, then $Y(t)$ has no causal influence on $X(t)$. Likewise, the correlation coefficient $\rho_{X|M_Y}$ is a measure of the possible causal influence of $X(t)$ on $Y(t)$.

2.3. The Reconstruction of Phase-Space Dynamical System

To start with the CCM causal inference, the time delay τ and embedding dimension E are the two key parameters in reconstructing the phase space of a nonlinear dynamical system. Empirically, the delay time can be chosen as the average oscillation period of the underlying times series (corresponding to the unit value in a discrete-time map) (Grassberger & Procaccia, 1983, 2004; Lai et al., 1996; Lai & Lerner, 1998; Lai & Ye, 2003). Here we choose time delay τ as 1. The choice is reasonable because, for the precipitation time series, the dependency among atmospheric interactions typically decays within a month (von Storch & Zwiers, 2001). For a proper

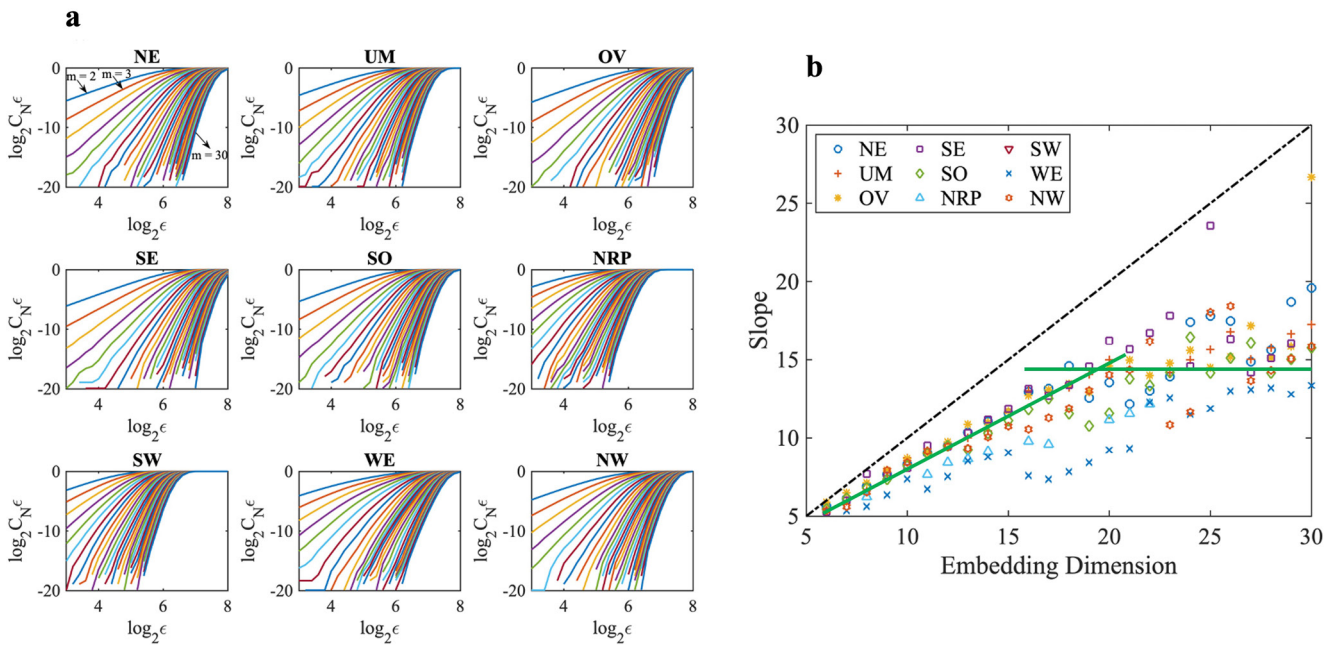

Figure 1. Determination of the embedding dimension from the GP-based correlation integral and dimension for the hydroclimate system over the CONUS using the aggregated time series of regional precipitation anomalies. (a) Correlation integral on a logarithm scale with different embedding dimension $E = 1, \ldots, 30$ for nine climate regions. (b) The slope values $\log C_N(\epsilon)$ versus $\log(\epsilon)$ as a function of embedding dimension for nine climate regions. The slope increases with E and then reaches an approximate plateau value for $E \geq 17$, as shown by the two green solid lines in (b).

choice of the embedding dimension, we use the standard Grassberger-Procaccia (GP) correlation integral and dimension algorithm (Grassberger & Procaccia, 2004).

For scalar time-series data, computations of the correlation can be carried out in the reconstructed phase space. The *correlation integral* is taken as the fraction of pairs of points on the attractor in the phase space of a nonlinear system within a hypersphere of radius ε, which can be calculated approximately as:

$$C_N(\varepsilon) = \frac{1}{N(N-1)} \sum_{j=1}^{N} \sum_{i=j+1}^{N} \Theta(\varepsilon - \|\boldsymbol{x}_i - \boldsymbol{x}_j\|), \tag{5}$$

For $N \to \infty$, the correlation dimension is given by

$$D_2 = \lim_{\varepsilon \to 0} \lim_{N \to \infty} \frac{\log C_N(\varepsilon)}{\log \varepsilon}. \tag{6}$$

To resolve the dynamics in the underlying system that generates the scalar time series, the dimension E of the reconstructed phase space must be sufficiently large. For a given value of the embedding dimension E, D_2 can be determined according to Equation 6. Since the intrinsic dimension of the underlying attractor is not known *a prior*, it is necessary to systematically increase the value of E to calculate a series of values for D_2. For an infinite, noiseless time series, the estimated dimension value D_2 increases with E but plateaus for $E > \left[\overline{D_2}\right] + 1$. For finite and noisy time series, the value of E required for D_2 to plateau is likely to be higher. For a completely stochastic system that is intrinsically infinitely dimensional, the estimated D_2 will *never* plateau, no matter how large E is. This line of reasoning emphasizes the need to estimate D_2 from a systematic set of E values (Ding et al., 1993).

Results of correlation integrals from the aggregated time series of precipitation anomalies for the nine climate regions in the CONUS are shown in Figure 1a. We use the least squares fitting method to determine the slope for the most linear part of each curve in Figure 1a (Lai & Ye, 2003). Slope values of $\log C_N(\varepsilon)$ versus $\log\varepsilon$, as functions of embedding dimension are shown in Figure 1b. Statistically, the slope increases with the embedding dimension and plateaus when $E_c \geq 17$, justifying the use of $E = 17$ in the CCM causality analysis. Note that the E value for each climate region is slightly different. To guarantee the complete reconstruction of the attractor

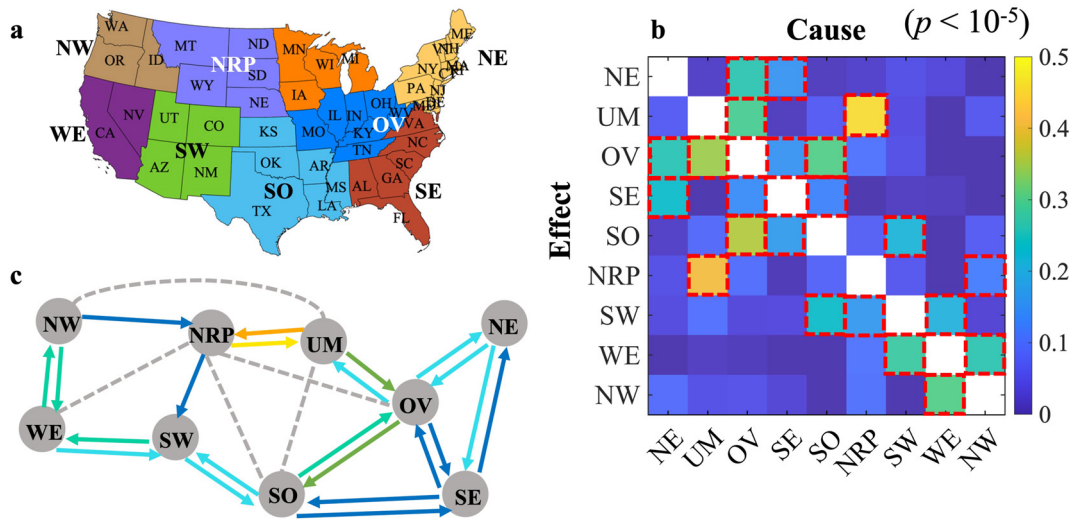

Figure 2. Detecting and quantifying causal interactions of the hydroclimate system over the CONUS. (a) The nine climatically consistent regions within the CONUS. (b) CCM results of all successful detection of significant causal interactions (red dashed squares) among the nine climate regions. The causal interactions are represented by a causally weighted directed network (no self-links) with embedding parameter values $E = 17$ and $\tau = 1$. (c) The reconstructed directed causal network from the results in (b), where the solid arrows indicate the directed links while the gray dashed lines represent the spurious unidirectional links due to a common driver or transitivity effect from pairwise association as determined by the cross correlation. The colors of the solid arrows specify the strength of the causal interactions as defined by the color bar in (b). The statistically significant spurious links are determined by comparing the undirected network from the pairwise cross correlation with the CCM inferred causal networks.

manifolds for all climate regions, we take the value that exceeds the maximum plateau in the nine climate regions as the embedding dimension E for the underlying climate system in this study.

2.4. Quantification of Causal Effect and Susceptibility for Climate Regions in the U.S.

In the directed causal matrices ($\rho_{Y|M_X}$ and $\rho_{X|M_Y}$) among all pairs of climate regions, we take the mean along each column as a measure of the average causal effect (ACE) to estimate the causal effect that a climate region R has on all other climate regions, whereas the mean along each row as the average causal susceptibility (ACS) is used to measure the sensitivity of a climate region to perturbations from other parts of the system (Runge et al., 2015). For region R, The ACE and ACS are calculated separately by

$$\mathrm{ACE}_R(t) = \frac{1}{N_R - 1} \sum_{i \neq R} \rho_{X_R(t)|M_i}, \tag{7}$$

$$\mathrm{ACS}_R(t) = \frac{1}{N_R - 1} \sum_{i \neq R} \rho_{X_i(t)|M_R}, \tag{8}$$

where N_R is the total number of climate regions. Furthermore, to investigate the long-term trend of ACE and ACS, we compute their running averages using a sliding window of size w as

$$\overline{\mathrm{ACE}}_{R,k} = \frac{1}{w} \sum_{j=k-(w-1)/2}^{k+(w-1)/2} \mathrm{ACE}(j), \tag{9}$$

$$\overline{\mathrm{ACS}}_{R,k} = \frac{1}{w} \sum_{j=k-(w-1)/2}^{k+(w-1)/2} \mathrm{ACS}(j). \tag{10}$$

where the time window is $[k - (w - 1)/2, k + (w - 1)/2]$, with k the center of the window. In a given region, a larger ACE value means a stronger causal effect of that region to mediate the precipitation climatology in other regions. Likewise, a higher ACS value signifies that the region is more susceptible to precipitation perturbations from other regions.

3. Results

3.1. Cross-Regional Causality in the CONUS

The causal networks constructed based on CCM identify statistically significant causal effects of precipitation anomalies between various pairs of adjacent nodes (climate regions), as shown in Figure 2b, where several indirect and spurious paths/links as revealed by the conventional cross correlations (e.g., the link between NW and UM) have been removed. The direct causal interactions are generally much stronger than the indirect ones. Note that, unlike the symmetrical pairwise association estimated using cross correlation, the causally weighted directed network is asymmetric. Of particular importance is the causal dependency between climate regions UM and NRP, which is statistically significant in both directions (with causation strengths greater than 0.4). Albeit being asymmetrically bidirectional, these significant causal interactions suggest mutually coupled precipitation dynamics in these two climate regions. Climate region NRP is relatively weakly (but still significantly) connected to NW and SW as indicated by the unidirectional links running from NW to NRP and those from NRP to SW. As shown in Figure 2c, the estimated causal network is sparser than the pairwise correlation network, as the latter often includes spurious links, especially teleconnections due to common forcing (e.g., El Niño-Southern Oscillation or ENSO) in the coupled climatic system (Runge, Bathiany, et al., 2019; Runge, Nowack, et al., 2019).

Additional analyses suggest the robustness of these causal interactions among climate regions to the selection of spatial aggregation and time lag. For spatial aggregation, we conduct a state-level causal analysis by aggregating the gridded precipitation (anomaly) data for each state, and the state-level causal patterns are generally consistent with results for the nine climate regions (see Figure S1 in Supporting Information S1). For time lag, we perform a similar causal analysis but with a 1-month lag for precipitation time series. The results agree with those in Figure 2b (see Figure S2 in Supporting Information S1). In addition to the significance test used in Figure 2b, we also evaluate causal interactions using the bootstrap method. Figures S3 and S4 in Supporting Information S1 summarize the causal results based on 50 bootstrap resampling. The consistency between different trails further demonstrates the robustness of the causal interactions shown in Figure 2b.

3.2. Regional Causal Effect and Susceptibility

To quantify the importance of various climate regions in spreading and mediating perturbations in the reconstructed causal, weighted, and directed network, we measure the causal influence of precipitation anomalies in one region on another using ACE and ACS indices defined in Equations 7 and 8. Figures 3a and 3b show the values of ACE and ACS averaged using 15-year sliding windows for the nine CONUS climate regions. Note that the size of the moving window cannot be too small or too large, as the cross-mapping causality estimate generally increases with time-series length (library) until reaching a plateau (Figure 3 in Sugihara et al., 2012). A 15-year sliding window is selected in this study mainly because it reveals robust structures of causality interaction among climate regions, as suggested by our sensitivity analysis (see Figure S5 in Supporting Information S1).

The distribution of ACE and ACS over all 15-year moving windows is shown in Figure 3e, where region OV and NRP have the largest ACE and ACS. The results indicate that the Ohio Valley region has manifestly the most significant causal effect on other regions and susceptibility among all CONUS regions, signaling that it acts as a regional gateway for propagating precipitation perturbations in the CONUS. The fact that Ohio Valley is a critical region in influencing hydrological processes and moisture propagation is consistent with the previous findings (Karl & Koscielny, 1982; Konapala & Mishra, 2017; Walsh et al., 1982). There are several plausible underlying mechanisms. First, the Ohio Valley is characterized by the leading principle component of winter precipitation (Walsh et al., 1982) and the third principle component of drought severity index (Karl & Koscielny, 1982), exhibiting the highest winter moisture variability in the United States. Second, the Ohio Valley has the strongest geostrophic wind components (Walsh et al., 1982). Third, the Ohio Valley is significantly affected by ENSO conditions in terms of precipitation (Zhang et al., 2010) and temperature extremes (Gershunov & Barnett, 1998). As a result, the high climate variability in Ohio Valley and its teleconnection with ENSO events are likely to be responsible for the strong causal effect and large susceptibility observed in this region (Konapala & Mishra, 2017).

In addition to ENSO, other climatic variability may also play some roles in regulating the causal links in the CONUS precipitation network through teleconnection, such as the potential Arctic amplification on mid-latitude summer circulation (Coumou et al., 2018) or the influence of Northern Pacific Oscillation (NPO) on the circulation and precipitation in the CONUS (Gershunov & Barnett, 1998). Moreover, a previous analysis of monthly

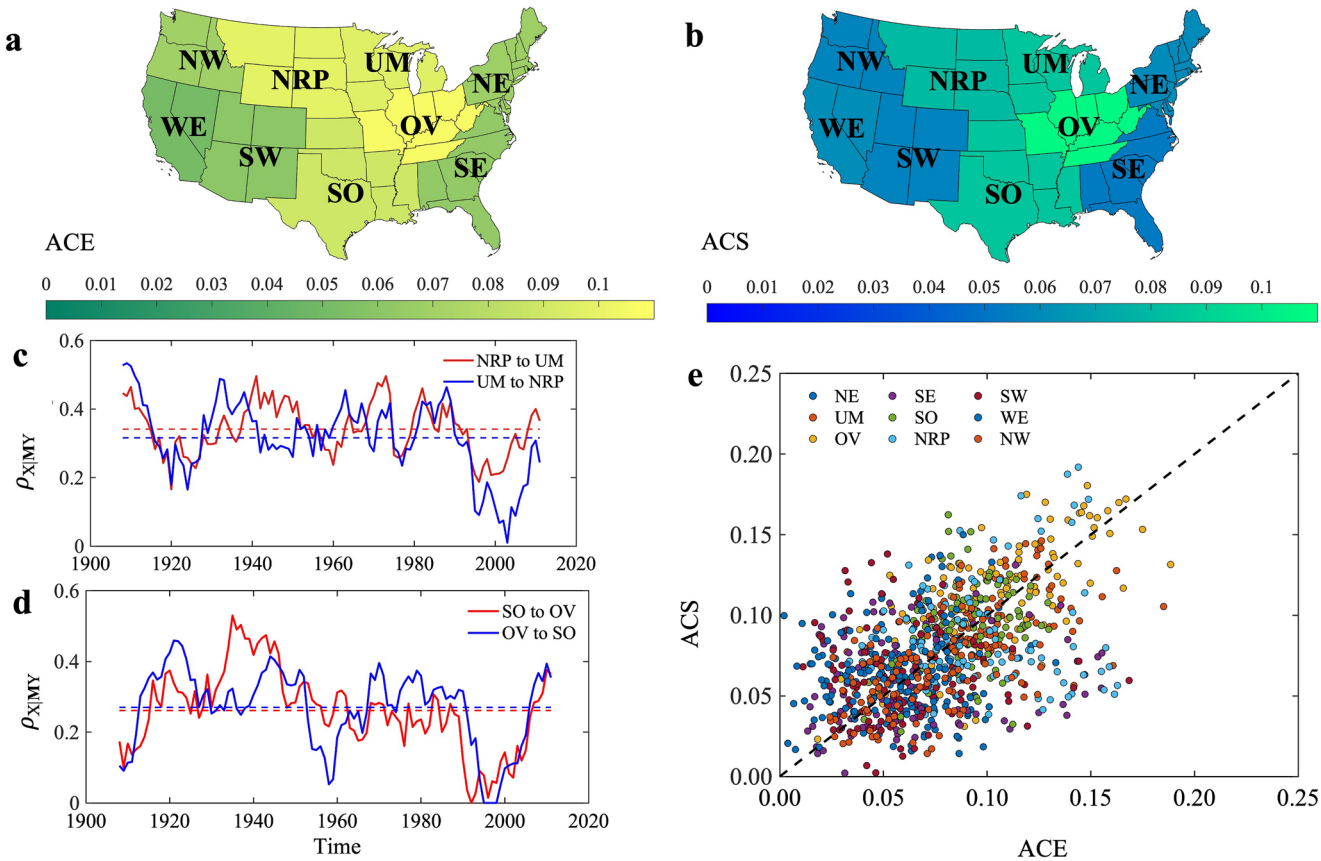

Figure 3. Measuring causal effect in the dynamical network of precipitation in the CONUS. (a) and (b) Long-term averaged causal effect (ACE) and averaged causal susceptibility (ACS) for each climate region. (c) Evolution of the strength of the CCM causality over time (with a 15-year sliding window) between two adjacent regions: NRP (Northern Rockies and Plains) and UM (Upper Midwest). (d) Time evolution of the CCM causality strength between the South and Ohio Valley. The horizontal dashed lines in red or blue in (c) and (d) represent the mean values of CCM causality strength. (e) ACE versus ACS over all 15-year sliding windows for each climate region.

precipitation identified a strong association between geostrophic wind components and sea-level pressure anomalies in the central and eastern United States where spatial coherence is manifest in Ohio Valley, Great Lake regions, and Northern Plains (Walsh et al., 1982). This is also supported by more recent causality analysis showing that regions with large ACE and ACS values correspond to major atmospheric convergence zones (Runge et al., 2015). Regions with strong geostrophic wind generate strong uplifts that integrate incoming perturbations at the surface and transport them vertically into the higher troposphere, which can influence other regions via atmospheric downdrafts, signaling strong causal effect and susceptibility, as shown in Figures 2a and 2b (the brighter zones).

3.3. Temporal Variability of Causality Between Pairs of Climate Regions

We further evaluate the temporal variability of causality strength of all climate components to assess if the causal dependencies are contemporaneous or cyclic). Figures 3c and 3d exemplify the causality strength over time (calculated from a sliding window) for the climate components with a strong ability to spread perturbations and with a high susceptibility to be causally influenced by others (NRP, UM, SO, and OV). Further analysis about the frequency or the periodicity of the time-varying causality is carried out using the method of empirical mode decomposition (EMD), which is a data-adaptative technique that decomposes a time series signal into rotational components of different frequencies, or the intrinsic mode functions (IMFs), where each IMF represents an oscillation mode embedded in the data (Huang et al., 1998; Huang & Wu, 2008).

For illustration, we apply the EMD method to the causality variability for two regions in Figure 3c, and the collection of IMFs after decomposition is shown in Figure 4 c_1, c_2, c_3, and c_4 are the four locally non-overlapping

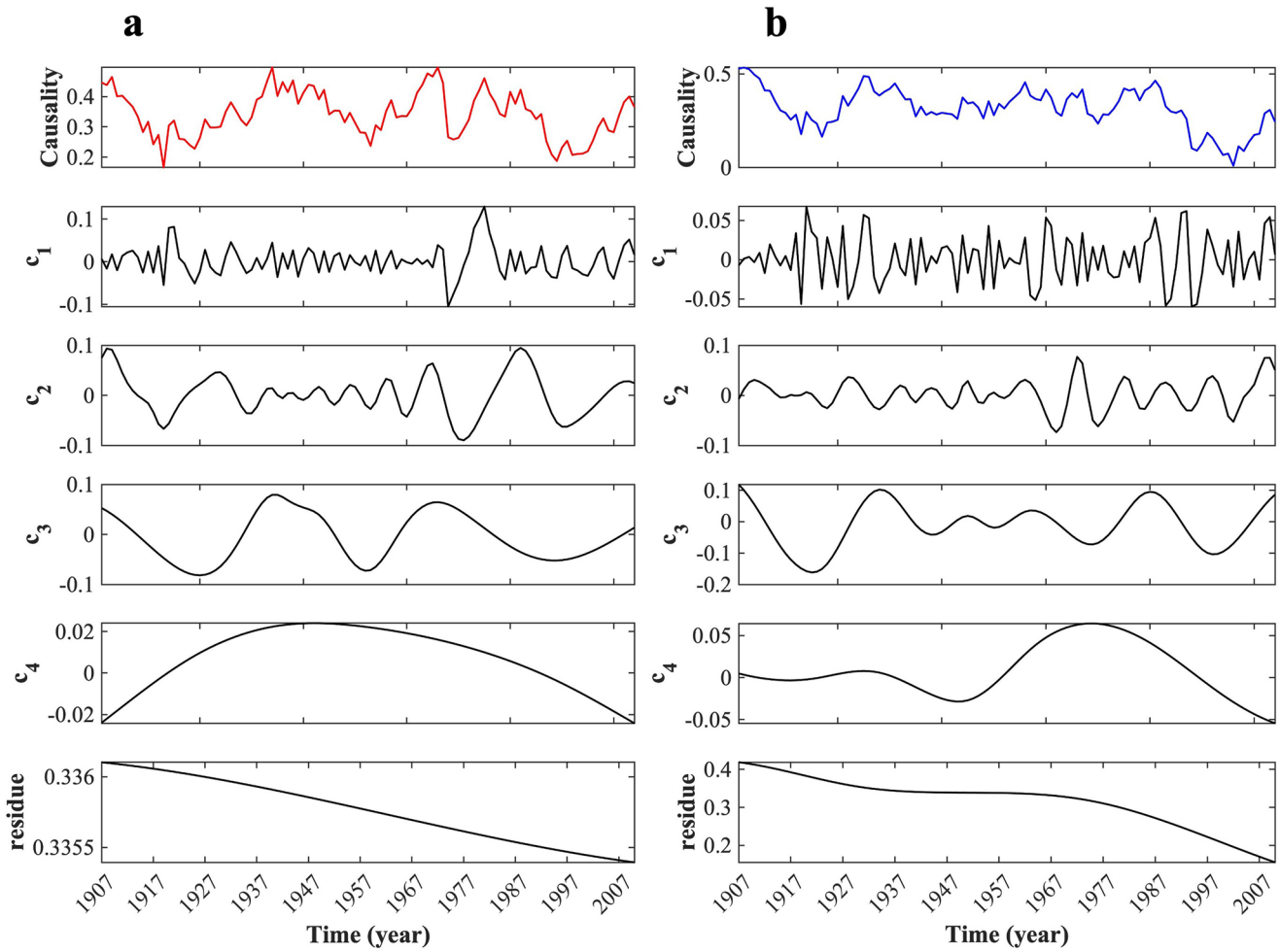

Figure 4. The collection of intrinsic mode functions (IMFs) decomposed by the EMD method for the time-varying causality in two regions in Figure 3c. The top panel in (a) is the causality variability from the region NRP to UM. The top panel in (b) is the causality variability from region UM to NRP. The four components from c_1 to c_4 correspond to the four IMFs with varying frequencies. The residue represents the general trend of the time-varying causality.

time scale components, while the residue time series signifies the general trend of the causality variability. It is noteworthy that the temporal variability of causality strength exhibits a strong periodicity from interannual (c_1 and c_2) to interdecadal recurrence (c_3 and c_4). Interannual periodicity can be attributed to the influence of the low-frequency variability inherent in the climate system (Ghil & Lucarini, 2020). Examples of such low-frequency oscillations, for example, ENSO or NPO, and their connection to the causality inference in CONUS precipitation are discussed above. In comparison, interdecadal periodicity appears to be linked with the oscillations in the global ocean's thermohaline circulation and its coupling to the atmosphere (Ghil & Lucarini, 2020)

4. Discussion

Climate changes in recent years have resulted in extreme weather in many regions worldwide. The western U.S. has been experiencing extremely severe drought, with no ending in sight. A key to mitigating the unprecedented drought lies in accurate knowledge about the causal links in the precipitation climatology and identification of the major climate regions, that is, regional gateways, that exhibit significant causality. Physically, such gateways are due to strong climate variability, atmospheric convergence, and/or pressure anomalies. Applying the CCM algorithm to the monthly precipitation time series for more than a century enables us to obtain an unambiguous network picture of the causal relations among the major climate regions in the CONUS. A quantitative assessment of the causal relations reveals that the Ohio Valley region effectively serves as a regional mediator for precipitation in other regions, where its bidirectional causal influence is regulated by the regional convective

uplift. This finding has identified, for the first time, the possible dynamical driving force of the precipitation activities in the CONUS.

The time evolution of causality influence and susceptibility among different climate regions of the CONUS uncovered here helps reveal the long-term trend of the precipitation dynamics. It is plausible, from the findings of this study, that the temporal variability of causality is a result of the synthesis of climate variability on multiple scales, ranging from annual cycles (e.g., trade winds) to decadal variation of planetary oscillators (e.g., ENSO and NPO). While much research effort has been devoted to investigating the relationship between low-frequency oscillators and the regional and global hydrological processes (e.g., precipitation, drought, and evaporation), research remains scarce in identifying the direct causal inferences of these contributions and their relative roles/strength in modulating the complex hydrologic dynamics. Our work partially fills this knowledge gap.

An ongoing challenge in the field is that time series based on causality inference in the Earth system science often assumes Gaussian noise (Runge, Bathiany, et al., 2019), whereas the distributions of precipitation in climate are often non-Gaussian. Another outstanding issue is that attractors constructed from real-world data are only low-dimensional approximations of the dynamics occurring in higher dimensions, while the degree of convergence is also limited as a result of observational error and process noise (Sugihara & May 1990). The "curse" of high dimensionality in the complex hydrological system could lead to a less accurate causal detection (Runge, Nowack, et al., 2019). For example, the CCM framework assumes causal sufficiency, which requires the absence of unobserved common drivers. In practice, with the technical assumptions being relaxed, the method may result in unreliable estimates of causations (Runge, Bathiany, et al., 2019). Previously, it was found that introducing proper noises, especially asymmetric noises, into the time series has the benefit of enhancing the detectability of directed dynamical influences in complex systems (Jiang et al., 2016). Exploiting this beneficial role of noise in detecting and characterizing causality from various climate data is worth pursuing.

5. Concluding Remarks

Our results of causality analysis of CONUS precipitation are promising as it not only identifies the regional mediators of the dynamics and propagation of moisture (anomalies) in the United States, but also has the potential to be extended to analyzing other hydroclimatic variables, especially those which are subject to anthropogenic influence and modulate the emergence of future climate patterns. Examples include using the CCM method to unravel the causal impact of anthropogenic emissions of heat, moisture, and greenhouse gases on the future evolution of complex hydroclimate systems with a focus on the occurrence of climatic extremes such as flooding, droughts, or mega heatwaves. For instance, the observed decrease in drought severity over the central United States during the second half of the twentieth century seems to be primarily driven by variability associated with tropical sea surface temperature (Shin & Sardeshmukh, 2011), which is largely attributable to the anthropogenic carbon emission. Detection of such causal relations is of paramount importance to informing and helping policy makers to develop and implement more sustainable strategies for mitigating climatic risks and extreme events faced by the humanity (Eyring et al., 2019). Causal inference also stands out as a powerful tool for detecting the potential critical, and often catastrophic, transitions in Earth and climate systems as both are believed to evolve toward unprecedented and irreversible changes due to anthropogenic stressors. Finding the causal relationship in the Earth system could enable us to pin down the crucial players, that is, tipping elements, of future critical transitions, as well as to help decision makers to find countermeasures to mitigate or even reverse the system tipping (Lenton et al., 2008).

Data Availability Statement

The gridded data set for 1901–2018 is the Climatic Research Unit (CRU) Time-Series (TS) version 4.03, which is archived by the Center for Environmental Data Analysis (CEDA) and is publicly available at https://catalogue.ceda.ac.uk/uuid/10d3e3640f004c578403419aac167d82. The division of climate regions is defined by the National Centers for Environmental Information of NOAA at https://www.cpc.ncep.noaa.gov/products/analysis_monitoring/regional_monitoring/regions.shtml.

Acknowledgments
This study is based upon work supported by the U.S. National Science Foundation (NSF) under Grant AGS-1930629 and CBET-2028868, and the National Aeronautics and Space Administrations (NASA) under Grant 80NSSC20K1263. YCL was supported by the Office of Naval Research (ONR) under Grant N00014-21-1-2323 and by the Air Force Office of Scientific Research (AFOSR) under Grant FA9550-21-1-0438. The authors thank Dr. Junjie Jiang for discussions about the CCM algorithm.

References

Boers, N., Goswami, B., Rheinwalt, A., Bookhagen, B., Hoskins, B., & Kurths, J. (2019). Complex networks reveal global pattern of extreme-rainfall teleconnections. *Nature*, *566*(7744), 373–377. https://doi.org/10.1038/s41586-018-0872-x

Coumou, D., Di Capua, G., Vavrus, S., Wang, L., & Wang, S. (2018). The influence of Arctic amplification on mid-latitude summer circulation. *Nature Communications*, *9*(1), 2959. https://doi.org/10.1038/s41467-018-05256-8

Deyle, E. R., & Sugihara, G. (2011). Generalized theorems for nonlinear state space reconstruction. *PLoS One*, *6*(3), e18295. https://doi.org/10.1371/journal.pone.0018295

Ding, M., Grebogi, C., Ott, E., Sauer, T., & Yorke, J. A. (1993). Plateau onset for correlation dimension: When does it occur? *Physical Review Letters*, *70*(25), 3872–3875. https://doi.org/10.1103/PhysRevLett.70.3872

Eyring, V., Cox, P. M., Flato, G. M., Gleckler, P. J., Abramowitz, G., Caldwell, P., et al. (2019). Taking climate model evaluation to the next level. *Nature Climate Change*, *9*(2), 102–110. https://doi.org/10.1038/s41558-018-0355-y

Fan, J., Meng, J., Ludescher, J., Chen, X., Ashkenazy, Y., Kurths, J., et al. (2021). Statistical physics approaches to the complex Earth system. *Physics Reports*, *896*, 1–84. https://doi.org/10.1016/j.physrep.2020.09.005

Gershunov, A., & Barnett, T. P. (1998). Interdecadal modulation of ENSO teleconnections. *Bulletin of the American Meteorological Society*, *79*(12), 2715–2726. https://doi.org/10.1175/1520-0477(1998)079<2715:IMOET>2.0.CO;2

Ghil, M., & Lucarini, V. (2020). The physics of climate variability and climate change. *Reviews of Modern Physics*, *92*(3), 035002. https://doi.org/10.1103/RevModPhys.92.035002

Good, P., Lowe, J. A., Andrews, T., Wiltshire, A., Chadwick, R., Ridley, J. K., et al. (2015). Nonlinear regional warming with increasing CO_2 concentrations. *Nature Climate Change*, *5*(2), 138–142. https://doi.org/10.1038/nclimate2498

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Journal of the Econometric Society*, *37*(3), 424–438. https://doi.org/10.2307/1912791

Grassberger, P., & Procaccia, I. (1983). Characterization of strange attractors. *Physical Review Letters*, *50*(5), 346–349. https://doi.org/10.1103/PhysRevLett.50.346

Grassberger, P., & Procaccia, I. (2004). Measuring the strangeness of strange attractors. In *The theory of Chaotic attractors* (pp. 170–189). Springer. https://doi.org/10.1007/978-0-387-21830-4_12

Harris, I., Osborn, T. J., Jones, P., & Lister, D. (2020). Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Scientific Data*, *7*(1), 109. https://doi.org/10.1038/s41597-020-0453-3

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., et al. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, *454*(1971), 903–995. https://doi.org/10.1098/rspa.1998.0193

Huang, N. E., & Wu, Z. (2008). A review on Hilbert-Huang transform: Method and its applications to geophysical studies. *Reviews of Geophysics*, *46*(2). https://doi.org/10.1029/2007RG000228

Jiang, J.-J., Huang, Z.-G., Huang, L., Liu, H., & Lai, Y.-C. (2016). Directed dynamical influence is more detectable with noise. *Scientific Reports*, *6*(1), 24088. https://doi.org/10.1038/srep24088

Kantz, H., & Schreiber, T. (1997). *Nonlinear time series analysis* (first. ed.). Cambridge University Press.

Karl, T. R., & Koscielny, A. J. (1982). Drought in the United States: 1895–1981. *Journal of Climatology*, 2(4), 313–329. https://doi.org/10.1002/joc.3370020402

Konapala, G., & Mishra, A. (2017). Review of complex networks application in hydroclimatic extremes with an implementation to characterize spatio-temporal drought propagation in continental USA. *Journal of Hydrology*, *555*, 600–620. https://doi.org/10.1016/j.jhydrol.2017.10.033

Kretschmer, M., Coumou, D., Donges, J. F., & Runge, J. (2016). Using causal effect networks to analyze different arctic drivers of midlatitude winter circulation. *Journal of Climate*, *29*(11), 4069–4081. https://doi.org/10.1175/JCLI-D-15-0654.1

Lai, Y.-C., David, L., & Hayden, R. (1996). An upper bound for the proper delay time in chaotic time-series analysis. *Physics Letters A*, *218*(1–2), 30–34. https://doi.org/10.1016/0375-9601(96)00408-2

Lai, Y.-C., & Lerner, D. (1998). Effective scaling regime for computing the correlation dimension from chaotic time series. *Physica D: Nonlinear Phenomena*, *115*(1–2), 1–18. https://doi.org/10.1016/S0167-2789(97)00230-3

Lai, Y.-C., & Ye, N. (2003). Recent developments in chaotic time series analysis. *International Journal of Bifurcation and Chaos*, *13*(06), 1383–1422. https://doi.org/10.1142/S0218127403007308

Lenton, T. M., Held, H., Kriegler, E., Hall, J. W., Lucht, W., Rahmstorf, S., & Schellnhuber, H. J. (2008). Tipping elements in the Earth's climate system. *Proceedings of the National Academy of Sciences*, *105*(6), 1786–1793. https://doi.org/10.1073/pnas.0705414105

McCann, K., Hastings, A., & Huxel, G. R. (1998). Weak trophic interactions and the balance of nature. *Nature*, *395*(6704), 794–798. https://doi.org/10.1038/27427

Moran, P. A. P. (1953). The statistical analysis of the Canadian Lynx cycle. *Australian Journal of Zoology*, 1(3), 291–298. https://doi.org/10.1071/zo9530291

Mysterud, A., Stenseth, N. C., Yoccoz, N. G., Langvatn, R., & Steinheim, G. (2001). Nonlinear effects of large-scale climatic variability on wild and domestic herbivores. *Nature*, *410*(6832), 1096–1099. https://doi.org/10.1038/35074099

Ombadi, M., Nguyen, P., Sorooshian, S., & Hsu, K. (2020). Evaluation of methods for causal discovery in hydrometeorological systems. *Water Resources Research*, *56*(7), e2020WR027251. https://doi.org/10.1029/2020WR027251

Packard, N. H., Crutchfield, J. P., Farmer, J. D., & Shaw, R. S. (1980). Geometry from a time series. *Physical Review Letters*, *45*(9), 712–716. https://doi.org/10.1103/PhysRevLett.45.712

Pearl, J., & Mackenzie, D. (2018). *The book of why: The new science of cause and effect*. Basic Books.

Runge, J. (2018). Causal network reconstruction from time series: From theoretical assumptions to practical estimation. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *28*(7), 075310. https://doi.org/10.1063/1.5025050

Runge, J., Bathiany, S., Bollt, E., Camps-Valls, G., Coumou, D., Deyle, E., et al. (2019). Inferring causation from time series in Earth system sciences. *Nature Communications*, *10*(1), 2553. https://doi.org/10.1038/s41467-019-10105-3

Runge, J., Nowack, P., Kretschmer, M., Flaxman, S., & Sejdinovic, D. (2019). Detecting and quantifying causal associations in large nonlinear time series datasets. *Science Advances*, *5*(11), eaau4996. https://doi.org/10.1126/sciadv.aau4996

Runge, J., Petoukhov, V., Donges, J. F., Hlinka, J., Jajcay, N., Vejmelka, M., et al. (2015). Identifying causal gateways and mediators in complex spatio-temporal systems. *Nature Communications*, *6*(1), 8502. https://doi.org/10.1038/ncomms9502

Sauer, T., Yorke, J. A., & Casdagli, M. (1991). Embedology. *Journal of Statistical Physics*, *65*(3–4), 579–616. https://doi.org/10.1007/bf01053745

Shepherd, T. G. (2014). Atmospheric circulation as a source of uncertainty in climate change projections. *Nature Geoscience*, *7*(10), 703–708. https://doi.org/10.1038/ngeo2253

Shi, H., Zhao, Y., Liu, S., Cai, H., & Zhou, Z. (2022). A new perspective on drought propagation: Causality. *Geophysical Research Letters*, *49*(2), e2021GL096758. https://doi.org/10.1029/2021GL096758

Shin, S.-I., & Sardeshmukh, P. D. (2011). Critical influence of the pattern of Tropical Ocean warming on remote climate trends. *Climate Dynamics*, *36*(7–8), 1577–1591. https://doi.org/10.1007/s00382-009-0732-3

Sugihara, G., May, R., Ye, H., Hsieh, C.-H., Deyle, E., Fogarty, M., & Munch, S. (2012). Detecting causality in complex ecosystems. *Science*, *338*(6106), 496–500. https://doi.org/10.1126/science.1227079

Sugihara, G., & May, R. M. (1990). Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature*, *344*(6268), 734–741. https://doi.org/10.1038/344734a0

Takens, F. (1981). Detecting strange attractors in fluid turbulence. In D. Rand & L. S. Young (Eds.), *Dynamical systems and turbulence* (pp. 366–381). Springer-Verlag.

Vejmelka, M., Pokorná, L., Hlinka, J., Hartman, D., Jajcay, N., & Paluš, M. (2015). Non-random correlation structures and dimensionality reduction in multivariate climate data. *Climate Dynamics*, *44*(9–10), 2663–2682. https://doi.org/10.1007/s00382-014-2244-z

von Storch, H., & Zwiers, F. W. (2001). *Statistical analysis in climate research*. Cambridge University Press.

Walsh, J. E., Richman, M. B., & Allen, D. W. (1982). Spatial coherence of monthly precipitation in the United States. *Monthly Weather Review*, *110*(4), 272–286. https://doi.org/10.1175/1520-0493(1982)110<0272:SCOMPI>2.0.CO;2

Wang, C., & Wang, Z. H. (2020). A network-based toolkit for evaluation and intercomparison of weather prediction and climate modeling. *Journal of Environmental Management*, *268*, 110709. https://doi.org/10.1016/j.jenvman.2020.110709

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

Yang, X., Wang, Z. H., & Wang, C. (2022). Critical transitions in the hydrological system: Early-warning signals and network analysis. *Hydrology and Earth System Sciences*, *26*(7), 1845–1856. https://doi.org/10.5194/hess-26-1845-2022

Yang, X., Wang, Z. H., Wang, C., & Lai, Y. C. (2022). Detecting the causal influence of thermal environments among climate regions in the United States. *Journal of Environmental Management*, *322*, 116001. https://doi.org/10.1016/j.jenvman.2022.116001

Zhang, X., Wang, J., Zwiers, F. W., & Groisman, P. Y. (2010). The influence of large-scale climate variability on winter maximum daily precipitation over North America. *Journal of Climate*, *23*(11), 2902–2915. https://doi.org/10.1175/2010JCLI3249.1

Zishka, K. M., & Smith, P. J. (1980). The climatology of cyclones and anticyclones over North America and surrounding ocean environs for January and July, 1950–77. *Monthly Weather Review*, *108*(4), 387–401. https://doi.org/10.1175/1520-0493(1980)108<0387:TCOCAA>2.0.CO;2