

Time variability and periodicities of cross-regional hydroclimatic causation in the contiguous United States

Xueli Yang¹, Zhi-Hua Wang^{1*}, Qi Li² and Ying-Cheng Lai^{3,4}

¹*School of Sustainable Engineering and the Built Environment, Arizona State University, Tempe, AZ 85287, USA*

²*School of Civil and Environmental Engineering, Cornell University, Ithaca, NY 14853, USA*

³*School of Electricity, Computer and Energy Engineering, Arizona State University, Tempe, AZ 85287, USA*

⁴*Department of Physics, Arizona State University, Tempe, AZ 85287, USA*

* Corresponding author. Email: zhwang@asu.edu. Tel: +1-480-727-2933; Fax: +1-480-965-0557

Abstract

Identifying and understanding various causal relations are fundamental to climate dynamics for improving the predictive capacity of Earth system modeling. In particular, causality in Earth systems has manifest temporal periodicities, like physical climate variabilities. To unravel the characteristic frequency of causality in climate dynamics, we develop a data-analytic framework based on a combination of causality detection and Hilbert spectral analysis, using longterm temperature and precipitation dataset in contiguous United States. Using the Huang-Hilbert transform, we identify the intrinsic frequencies of cross-regional causality for precipitation and temperature, ranging from interannual to interdecadal time scales. In addition, we analyze the spectra of the physical climate variabilities, including El Niño-Southern Oscillation and Pacific Decadal Oscillation. It is found that the intrinsic causal frequencies are positively associated with the physics of the oscillations in the global climate system. The proposed methodology provides fresh insights into the causal connectivity in Earth's hydroclimatic system and its underlying mechanism as regulated by the characteristic low-frequency variability associated with various climatic dynamics.

Keywords: *Convergent cross mapping; Empirical mode decomposition; Hilbert-Huang transform; Hydroclimate systems; Low-frequency variability*

1. Introduction

The Earth's climate system involves a complex interplay of dynamic processes that are closely coupled and more importantly, causally influence each other. An important characteristic of the climate system is the coexistence and nonlinear interactions of multiple subsystems, processes, and scales (Ghil and Lucarini 2020). Successful identification of causality provides an approach to understand the complex dynamics and the underlying physical mechanisms of the climate system (Runge et al. 2019a; van Nes et al. 2015). Methods based on pairwise association, such as correlation-based measures, are common tools for analyzing the relationships among the variables in the climate system, but the statistical-association based methods are often inadequate for infer causality (Runge et al. 2019b; Gao et al. 2022). To detect causality from time series, a celebrated framework is the Granger causality (GC) method (Granger 1969). In climate sciences, the GC method has been widely used in applications such as identifying the temperature and wind patterns (McGraw and Barnes 2018) or climate teleconnections (Silva et al. 2021). Mathematically, the GC method is based on the assumption that the underlying dynamical system is decriable as a linear stochastic process. For nonlinear and nonseparable complex systems, the GC method can be ineffective in distinguishing causation from simple correlation (Hannart et al., 2016).

Recent years have witnessed the use of a number of nonlinear and complex dynamical systems-based algorithms of causality detection in Earth system sciences. For example, Hannart et al. (2016) proposed a causal counterfactual theory for the attribution of weather and climate-related events, which makes a valuable contribution to theoretical tools for understanding the climate crisis from Hasselmann's programme (Lucarini and

Chekroun, 2023). In addition, a framework of causality detection based on deterministic dynamics has been proposed and evaluated for nonlinear dynamic systems based on the convergent cross mapping (CCM) method (Sugihara et al. 2012; Runge et al. 2015; Kretschmer et al. 2016; Runge et al. 2019a; Leng et al. 2020). The fundamental idea of CCM is that when causation is unilateral, the information of the cause can be inferred from that of its effect (Sugihara et al. 2012). In addition, the method involves the property of convergence, viz. the skill of cross-mapped estimates improves with the length of time series, that distinguishes it from simple correlation.

The CCM method has been applied to earth and climate systems for detecting the causal feedback and relationships among climate variables (van Nes et al. 2015) and atmospheric interactions (Runge et al. 2015). Quite recently, advances in data collection, archiving, and availability have greatly benefited causal inference in Earth system sciences (Runge et al. 2019a, b). For example, data-intense methods such as those based on complex network analysis (Boers et al. 2019; Tsonis et al. 2008; Wang and Wang, 2020), critical transitions (Wang et al. 2020; Yang et al. 2022a), and machine learning (Li et al. 2022) have been introduced.

The dynamical behavior of the climate system is much more complicated than those associated with equilibrium or periodic oscillations (Ghil 2002). Relatively small changes in external forcing arising from various anthropogenic and/or natural sources can contribute to large climate variability on a wide spectrum of temporal and spatial scales (Lucarini et al. 2014; Ghil and Lucarini 2020). Critical mechanisms involved in climate variability and periodicity include air-sea interaction, equatorial wave dynamics, and radiative forcing by the seasonal cycle (Bjerknes, 1969). In general, there are three kinds of

natural climate variability (Ghil 2002): (1) those driven by a purely periodic force such as the daily and the seasonal cycle and orbital forcing, or the solar irradiance, (2) those due to the nonlinear complex interplay of feedbacks within the climate system, and (3) those that are related to the random stochastic fluctuations in physical or chemical forcing, such as the volcanic eruptions or weather fluctuations in the climate system. Among the current debate on the relative roles of natural and anthropogenic variability, *interdecadal* and *centennial* climate oscillations are of pivotal importance and have attracted a great deal of interest (Dijkstra and Ghil 2005).

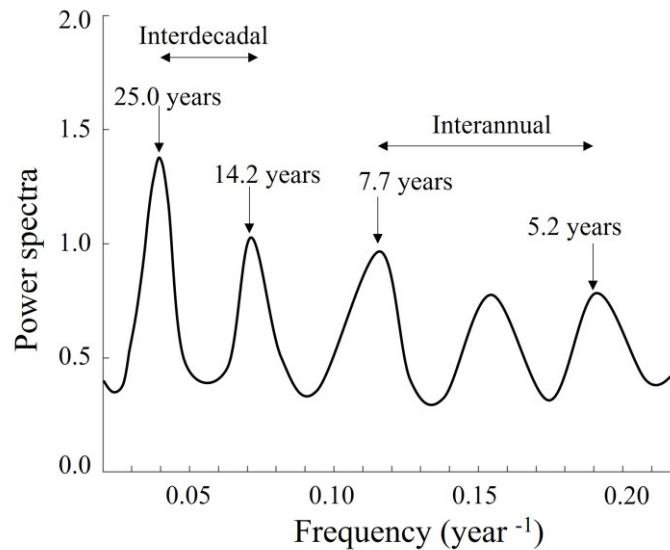


Figure 1. A schematic illustration of the power spectra of climate variability. Shown is the spectrum of the 335-year Central English Temperature record and the tentative physical causes of the spectrum peaks (after Plaut et al. 1995).

To quantify the temporal variability in the climate system, the traditional tools such as Fourier spectrum or the singular spectrum methods were used to analyze climate oscillations (Ghil and Vautard 1991; Ghil et al. 2002), where the spectral peaks, for

example, can be related to certain forcing mechanisms (Ghil and Lucarini 2020; Ghil et al. 2002; Plaut et al. 1995). **Figure 1** presents an example: the spectrum of the 335-year Central English Temperature record and the tentative physical causes of the spectral peaks (Plaut et al. 1995). The climate variability on multiple timescales can be identified by a singular spectrum, with each peak in the spectrum potentially attributable to a different physical mechanism (Ghil and Lucarini 2020; Ghil et al. 2002; Plaut et al. 1995). For example, the climate oscillations with interannual (7 to 8-year) and interdecadal (14- and 25-year) periods are likely correlated to the North Atlantic’s wind-driven and thermohaline circulation, respectively (Dijkstra and Ghil 2005; Plaut et al. 1995). The peak with a period around 5.5 year can be attributed to the effect of low-frequency El Niño-Southern Oscillation (ENSO) (Ghil and Robertson, 2000). Fourier transform has been the standard approach to analyzing the frequency distributions of signals. However, the Fourier analysis is most suited for linear and stationary systems. Difficulties can arise when the system generating the signal is nonlinear or non-stationary (Huang et al. 1998; Huang and Wu 2008).

For signals from complex, nonlinear or chaotic, and non-stationary dynamical systems, the methodology of empirical mode decomposition (EMD) in combination with Hilbert transform and instantaneous frequency-based analysis have been demonstrated to be effective for identifying the significant “physical frequencies” of the underlying system (Huang et al. 1998; Huang and Wu 2008; Yalçinkaya and Lai 1997; Lai 1998; Huang et al. 2017). To appreciate this, recall a defining feature of a sinusoidal signal: in its time evolution a local maximum is followed by a zero, then by a local minimum, and by a local maximum again, and so on. Physically, such a signal is the projection of a proper rotation

in the plane onto an arbitrary one-dimensional axis, where the term “proper” is referred to as the existence of a unique center of rotation. For signals corresponding to a proper rotation, the Hilbert transform can be performed to yield a complex analytic signal for which the frequency of rotation in the complex plane can be identified unambiguously. Complex signals from the real world often do not possess such a proper structure of rotation, so a direct application of the Hilbert transform will lead to ambiguities with misleading results.

A remedy is then to first decompose it into a small number of modes, each corresponding to a proper rotation, the so-called empirical modes (Huang et al. 1998). For each empirical mode, Hilbert transform is well defined and can be carried out to yield the specific frequency of rotation of the corresponding analytic signal in the complex plane. Note that the frequencies so obtained are physically meaningful because of the underlying rotation structure, and they differ fundamentally from the frequencies of the Fourier transform associated with various harmonics. Because EMD typically yields a comparatively small number of modes, the number of distinct physical frequencies is equally small, in contrast to the mathematical Fourier frequencies that can often be infinite in number. In literature, the combination of empirical-mode decomposition and Hilbert transform is often referred to as Hilbert-Huang transform (HHT) (Huang and Wu 2008). The major limitation of EMD is that it is empirically based, and lacks sound theoretical foundation as compared to, e.g. Fourier transform (Boashash, 2016).

The Earth system and climatic dynamics are undoubtedly highly nonstationary and nonlinear, rendering appropriate use of the HHT for analyzing the time variations and detecting the dominant physical frequencies. In the past, the HHT has been used to study a

wide variety of climatological phenomena such as the diurnal cycle, front passages, or seasonal cycles (Duffy 2004; Huang and Wu 2008). Moreover, HHT has been applied to capture episodic phenomena such as snowmelt and heavy precipitation events (Duffy 2004), near-Hermean environment (Alberti et al., 2021), extreme sea level events (Alberti et al., 2023), as well as to identify possible links between river flow variability and the global climate regime fluctuations (Massei and Fournier 2012). These existing studies highlighted that HHT can be an effective, though hitherto underexplored, tool to detect climate variability and its underlying climatic events in Earth system.

To analyze the temporal variability and identify the possible periodicities of hydroclimatic causation in the contiguous United States (CONUS), we first apply the convergent crossing mapping (CCM) method to quantify the causality interactions among different climate regions of the U.S., based on the long-term monthly precipitation and temperature time series. As presented in our recent studies (Yang et al. 2022b, 2023a), our network causality analysis revealed that the Ohio Valley acts as a regional atmospheric gateway in mediating the propagation of temperature perturbations. We then invoke HHT to analyze the causality signal, which allows several distinct physical frequencies underlying the temporal variations of the hydroclimatic causation to be detected. The HHT analysis reveals that the hydroclimatic causality possesses characteristic periodicities, ranging from interannual to interdecadal scales. Surprisingly, the observed periodicities are quite distinct, enabling us to identify the underlying physical causes and the mechanisms as regulated by characteristic low-frequency variability associated with various climatic dynamics.

2. Method

2.1. Dataset retrieval and processing

In this study, we use the monthly mean near-surface precipitation (P) and temperature (T) over the period 1901–2018, obtained from Climatic Research Unit (CRU) Time-Series (TS) version 4.03 produced by the Center for Environmental Data Analysis (CEDA) Archive, to assess the periodicity of hydroclimate causation in CONUS. We first anomalized the spatially gridded ($0.5^\circ \times 0.5^\circ$) temperature and precipitation time series by removing the long-term trends (Harris et al., 2020). The anomalization process is conducted by subtracting the annual trends from the original time series, e.g. the anomaly of precipitation or temperature in January is obtained by subtracting the average of 118 January P/T values over the study period of 1901-2018. This way, we minimize the impact of high frequency (subseasonal to annual) variability in subsequent analyses.

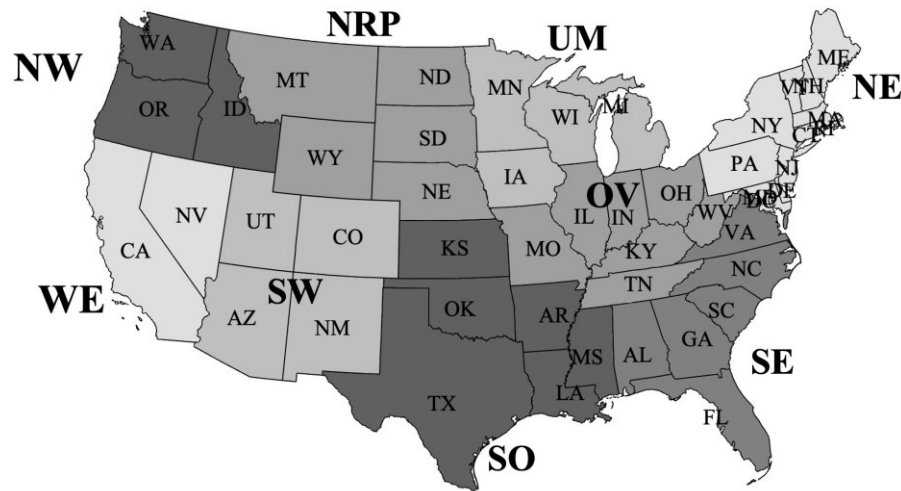


Figure 2. The map of nine climate regions in the contiguous United States, including Northwest (NW), West (WE), Southwest (SW), Northern Rockies and Plains (NRP), South (SO), Upper Midwest (UM), Ohio Valley (OV), Southeast (SE), and Northeast (NE).

The time series of gridded anomalies is then aggregated (averaged over gridcells) into nine climatic regions following the Geographical Reference Maps of National Centers for Environmental Information of National Oceanic and Atmospheric Administration (NOAA). These climatic regions, as shown in **Fig. 2**, include Northwest (NW), West (WE), Southwest (SW), Northern Rockies and Plains (NRP), South (SO), Upper Midwest (UM), Ohio Valley (OV), Southeast (SE), and Northeast (NE), which are consistent with the nine principal components identified using the principal component analysis (PCA) method (Vejmelka et al., 2015; Runge et al., 2015). The subseasonal and seasonal periodicity is also removed by subtracting the monthly averages from the time series of the aggregated temperature and precipitation anomalies in nine regions. This treatment removes the high-frequency variability in the hydroclimate system, which also renders the application of CCM method more tractable.

2.2. Causal inference by convergent cross mapping

Mathematically, the CCM algorithm for causal inference is developed based on the classic delay-coordinate embedding theory for nonlinear time series analysis (Takens, 1981), and a nearest-neighbors algorithm for reconstruction of the cross-mapping estimate (Sugihara and May 1990, Sugihara et al. 2012). This method is particularly suitable for detecting the directed causal influence in moderately-coupled nonlinear dynamic systems. Despite its prior applications (e.g. see Runge et al. 2015; van Nes et al. 2015; Yang et al., 2022b), the potential of CCM in detecting causal relations in Earth's climate systems is hitherto under-explored.

The fundamental idea of the CCM algorithm is that if causality exists between a pair of generic variables, represented by two time series $X(t)$ and $Y(t)$ respectively, then the cross-mapping dynamics of one variable can be reconstructed from the information of the other. Technically, we can first construct a shadow manifold M_X from $X(t)$ by re-arranging the original time series into a lagged-coordinate vector, denoted as $\mathbf{x}(t) = [X(t), X(t-\tau), \dots, X(t-(E-1)\tau)]$, where τ and E are the time delay and embedding dimension, respectively. A shadow manifold $\mathbf{y}(t)$ and M_Y can be constructed likewise. Using the shadow manifold, we then construct their cross-mapping estimates. For example, the cross-mapping $\hat{Y}(t) | M_X$ can be constructed using a simple projection of the $E+1$ nearest neighbors of vector $\mathbf{x}(t)$ in the manifold of M_X , with exponentially weighted distances. Mathematically, time indices of the $E+1$ points on M_X are used to identify the corresponding neighbors in Y , i.e., $Y(t_1), Y(t_2), \dots, Y(t_{E+1})$. Thus, the cross-mapping estimate of $\hat{Y}(t) | M_X$ is calculated as,

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

where $w_i(t)$ are the weighting coefficients given by

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

with

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

where $d[\mathbf{x}(t), \mathbf{x}(t_i)]$ is the Euclidean distance between $\mathbf{x}(t)$ and $\mathbf{x}(t_i)$ in M_X .

It follows that the causality from Y to X is measured by the correlation coefficient

$\rho_{Y|M_X}$ between the original $Y(t)$ and the cross-mapping estimate $\hat{Y}(t) | M_X$, given by

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

where $\mathbf{E}$, μ , and σ are the statistical expectation, average, and standard deviation, respectively. A larger value $\rho_{Y|M_X}$ implies a stronger casual influence, while Y is not causal to X if $\rho_{Y|M_X} \leq 0$. Likewise, the causality from X to Y can be quantified using the same procedure by constructing the cross-mapping estimate $\hat{X}(t) | M_Y$ and finding its correlation to the original data series of $X(t)$ as illustrated above.

In addition, the accuracy of the CCM algorithm depends largely on the proper choice of the time delay τ and the embedding dimension E , which can be determined using the correlation integral and dimension method (Grassberger and Procaccia 1983; Lai and Ye 2003). More specifically, our previous analysis shows that $\tau = 1$ and $E = 3$ for temperature (Yang et al. 2022b), and $\tau = 1$ and $E = 17$ for precipitation (Yang et al. 2023a), respectively (Results regarding choice of E and τ are also included in Supplementary **Figs. 1** and **2** in the Supplementary Information). It can be interpreted as that the time lag of one month is used to construct the lagged-coordinate embeddings of precipitation and temperature, while the suitable dimension of the shadow manifolds is 3 for temperature and 17 for precipitation for causal inference. Applying the CCM algorithm to the temperature and precipitation dataset in CONUS enables us to generate the directed causal *network* of nine nodes, each representing one climate region. Furthermore, two additional indices are defined to measure the causal effect and causal sensitivity for individual climate regions, viz. the average causal *effect* (*ACE*) and the average causal *susceptibility*

(ACS). ACE and ACS for a given region R are calculated by averaging each column and rows of the adjacency matrix of the causality network, respectively (Runge et al., 2015):

$$ACE_R(t) = \frac{1}{N_R - 1} \sum_{i \neq R} \rho_{X_R(t) \| M_i}, \quad (5)$$

$$ACS_R(t) = \frac{1}{N_R - 1} \sum_{i \neq R} \rho_{X_i(t) \| M_R}, \quad (6)$$

where N_R is the total number of climate regions. In this study, we also compute their running averages to study their temporal variation using a sliding window of size w , as

$$\overline{ACE}_{R,k} = \frac{1}{w} \sum_{j=k-(w-1)/2}^{k+(w-1)/2} ACE(j), \quad (7)$$

$$\overline{ACS}_{R,k} = \frac{1}{w} \sum_{j=k-(w-1)/2}^{k+(w-1)/2} ACS(j). \quad (8)$$

where k is the time (year) centered at each moving window $[k-(w-1)/2, k+(w-1)/2]$. After sensitivity tests (Yang et al., 2023a), we use a moving window $w = 15$ years in this study. It is understood that a larger ACE value signals stronger causal effect of the climatic region of interest to mediate the thermal (temperature) or moisture (precipitation) perturbations in other regions. Similarly, larger ACS values mean that the region is more *susceptible* to the influence caused by other regions.

2.3. The empirical mode decomposition

The EMD method is a data-adaptive technique that decomposes a time series signal into rotational components of different frequencies, termed as the intrinsic mode function (IMF) (Huang et al. 1998; Huang and Wu 2008). Each IMF represents an oscillation mode embedded in the data and an IMF can have time-varying amplitude and frequency.

Intuitively, an IMF is a function with the same number of extrema and zero crossings, whose envelopes are symmetric with respect to zero (Huang et al. 1998). More specifically, an IMF is a function satisfying the two conditions: (1) the number of extrema and the number of zero-crossings in the data time series must either be equal or differ at most by one (IMF is sinusoidal-like function); and (2) the mean value of the envelopes defined by the local maxima and minima is zero at any point. The decomposition method due to Huang (Huang et al. 1998) identifies the upper and lower envelopes defined by the local maxima and minima, respectively, through a shifting process. All the local maxima are connected by a cubic spline line as the upper envelop, and the same procedure applies for the local minima. The procedure is repeated until the upper and lower envelopes cover all the data points in between. The difference between the mean m_1 of envelopes and the original time series gives the first component h_1 :

$$h_1 = X(t) - m_1, \quad (9)$$

Ideally, h_1 should be an IMF. However, there is no guarantee that the two defining conditions of an IMF would be met by h_1 , as new extrema can be generated since changing the local zero from a rectangular to a curvilinear system. The shifting process was introduced to solve this problem by repeating the process in Eq. (9) k times until the resulting signal component meets the two IMF conditions:

$$h_{1k} = h_{1(k-1)} - m_{1k}, \quad (10)$$

which gives the first IMF component c_1 :

$$c_1 = h_{1k}. \quad (11)$$

The criterion for the sifting process to stop is determined by limiting the size of the normalized squared difference between two successive shifting processes, defined as,

$$SD_k = \frac{\sum_{t=0}^T |h_{1(k-1)}(t) - h_{1k}(t)|^2}{\sum_{t=0}^T h_{1(k-1)}^2(t)}. \quad (12)$$

The typical value of this squared difference SD can be set as a small fraction (Huang et al. 1998). In our study, we choose $SD = 0.2$.

The first IMF component c_1 so obtained contains the finest time scale or the shortest period component of the signal. By extracting c_1 from the original causality signal $C(t)$, we can get the residue of the first component r_1 :

$$r_1 = C(t) - c_1. \quad (13)$$

Since the information of longer period component can be contained in the residue, r_1 is then treated as the new data, subject to the same sifting process. This procedure can be repeated on all the subsequent residues to yield

$$r_1 - c_2 = r_2, \dots, r_{n-1} - c_n = r_n, \quad (14)$$

where c_n is the n th component, and r_n is the n th residue. The process is stopped when c_n or r_n becomes adequately small from which no more IMF can be extracted. The decomposition into n -empirical modes of the data and residue r_n can be obtained by summing up Eq. (13) and (14),

$$C(t) = \sum_{i=1}^n c_i + r_n. \quad (15)$$

The residue is either a monotonic function or a function with only one extremum that does not contain information to be further decomposed into a physically meaningful oscillatory component. Or equivalently, the final residue represents the temporal *trend* of the quantity (physical or causal) represented in the time series $C(t)$. The original time series can be

reconstructed from the IMFs by successively adding components with increasing frequency.

2.4. Hilbert spectral analysis

After the IMFs are obtained, Hilbert spectral analysis can be carried out, for each IMF, to obtain the instantaneous frequency as function of time. The resulting Hilbert spectrum is a frequency-time distribution of the signal amplitude (or energy), which enables us to identify the localized features of the original data that correspond to natural phenomena and/or admit physical interpretation. The Hilbert transform $H_C(t)$ of a generic time series $C(t)$ is defined as

$$H_C(t) = \frac{P}{\pi} \int_{-\infty}^{+\infty} \frac{C(\tau)}{t - \tau} d\tau, \quad (16)$$

where P is the principal Cauchy value of the integral. The time series $C(t)$ and its Hilbert Transform $H_C(t)$ form the complex conjugate pair, from which an analytical signal $Z(t)$ can be calculated as

$$Z(t) = C(t) + iH_C(t) = a(t) e^{i\theta(t)}, \quad (17)$$

where $i = \sqrt{-1}$ is the imaginary unit and a is the instantaneous amplitude given by

$$a(t) = \sqrt{C^2(t) + H_C^2(t)}, \quad (18)$$

with

$$\theta(t) = \arctan \left[\frac{H_C(t)}{C(t)} \right]. \quad (19)$$

The instantaneous frequency is then defined as

$$w = \frac{d\theta(t)}{dt}, \quad (20)$$

After performing the Hilbert transform on each IMF component j , we have the complex representation of the original signal $C(t)$ as

$$C'(t) = \sum_{j=1}^n a_j(t) e^{i \int w_j(s) ds}, \quad (21)$$

where both the amplitude and the frequency of each component are functions of time. The frequency-time distribution of the amplitude is designated as the Hilbert amplitude spectrum or simply Hilbert spectrum. The time localities of the events are preserved and the instantaneous frequency and energy, rather than the global properties as in the conventional Fourier spectral analysis, are involved. The frequency-time distribution in the Hilbert spectrum is more physically meaningful than that from the Fourier analysis, as spurious harmonics are always generated in the Fourier-based analysis due to nonlinearity and non-stationarity. In the Hilbert spectral analysis, the intrawave frequency modulations not only present a clear physical picture of the motion but are also effective in mitigating the spurious harmonics (Huang et al. 1998).

3. Results and Discussion

3.1. Regional hydroclimatic causal networks in the CONUS

We first construct the pair-wise directed causal networks of hydroclimate among the nine CONUS climate regions for precipitation and temperature anomalies. The results of constructed causal graphs and comparison of precipitation and temperature causation are illustrated in **Fig. 3**, where the causality strength ρ is calculated using Eq. (4). Naturally, the causal networks of precipitation and temperature are asymmetric because of the causal

influence between two regions are directed, i.e., the causal influence of region A on B is not necessarily reciprocated with the same strength. The degree of asymmetry is not very high in **Fig. 3** possibly due to the fact that the causal relation was determined between a pair of like variables (temperature with temperature) in different regions. The asymmetry of causal relation becomes much more manifest between different variables, e.g. between temperature and pressure or carbon fluxes (Yang et al., 2023b; Wang et al., 2024).

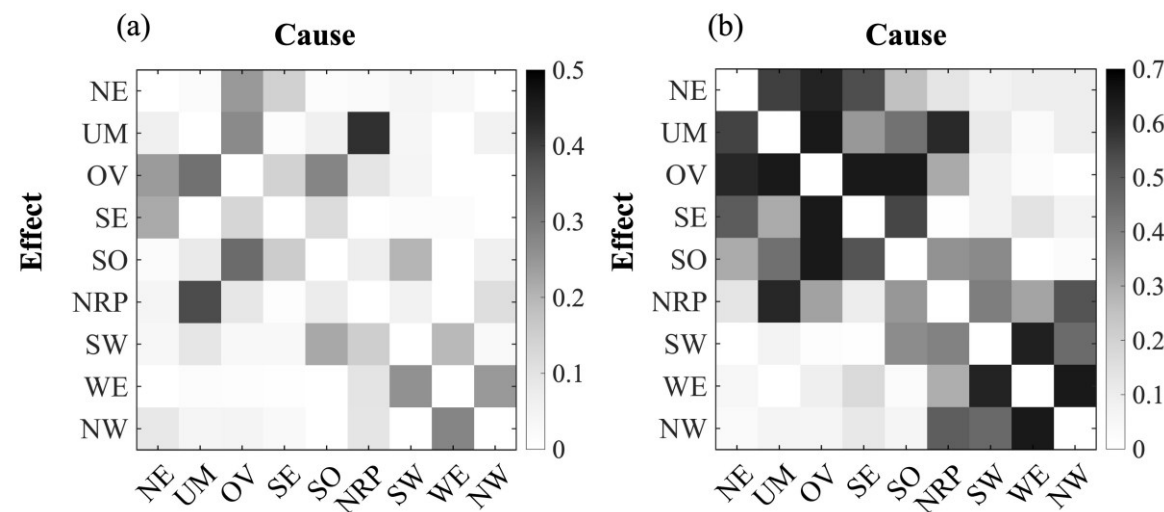


Figure 3. Results of CCM causal analysis. Shown are the directed causal networks for (a) precipitation and (b) temperature, respectively, among climate regions over the CONUS.

The time delay and embedding dimension for constructing the manifold in CCM is $\tau = 1$, $E = 17$ for precipitation and $\tau = 1$, $E = 3$ for temperature. Self-links are removed. The white diagonal cells represent negligible causality strength (with zero or negative values).

A comparison between the two causal graphs suggests that, in general, the cross-regional causation on temperature anomalies is often much stronger than that in precipitation. This is physically meaningful because the transport of heat and propagation of thermal anomalies are likely less disturbed by other confounding variables, such as

humidity, pressure, wind, than the transport of moisture (Yang et al., 2023a, Wang et al., 2024). Furthermore, the formation of precipitation is not completely determined by the atmospheric content of water vapor, but rather depends on many meteorological factors and even atmospheric chemistry, while the thermal environment of region is almost exclusively determined by temperature.

Despite the apparent difference in the strength of causation, there is similarity in the spatial pattern of both causal graphs in **Fig. 3**. The regions that are causally active in both precipitation and temperature include Northern Rockies and Plains, Ohio Valley, and Upper Midwest, all playing important roles in the regulating the cross-regional transport of heat and moisture. This could be potentially attributed to the planetary waves modulated by the Rockies. On the contrary, the Northwest and West regions are less causally influenced by other regions except their mutual causality, such that the two regions form a rather isolated hydroclimate cluster in the CONUS.

3.2. Averaged hydroclimatic causal effect and susceptibility

With the causal networks constructed, we study the cross-regional connectivity of hydroclimatic causation over the entire study period of 118 years. This is done by applying the CCM method with a 15-year sliding window to the precipitation and temperature anomalies in the same time span. The regional causal effect and susceptibility are quantified using the indicators of *ACE* and *ACS*, as defined in Eq. (7) and (8), respectively. The time series of *ACE* and *ACS* are then averaged over the entire period of the data series for each climate region, the results are shown in **Fig. 4**.

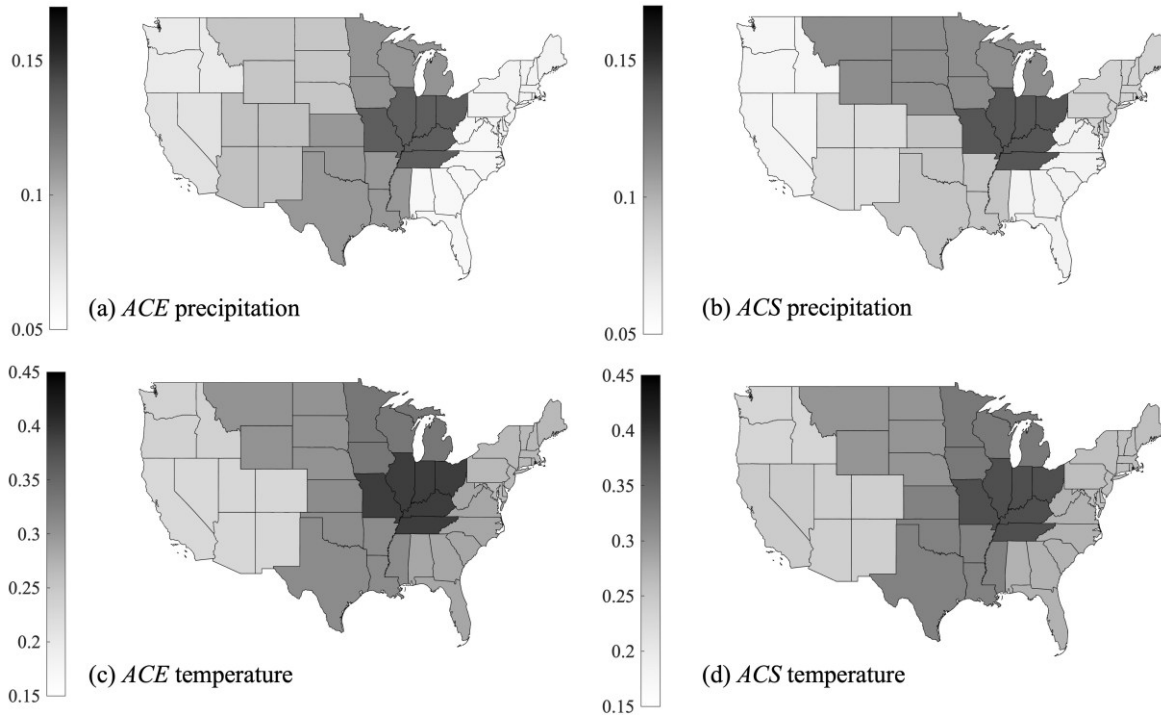


Figure 4. Map distribution of the averaged causal effect and susceptibility for precipitation and temperature over the CONUS. The results are averaged from *ACE* and *ACS* time series over the entire period of 1901-2018.

It can be seen that the Ohio Valley stands out as a regional atmospheric gateway in the CONUS hydroclimatic network, with the highest value of *ACE* and *ACS* in both precipitation and temperature. The fact that the Ohio Valley plays an important role in regulating the transport of heat and moisture in CONUS is consistent with the results of observation and physical modeling reported in the literature. For example, it was found that the Ohio Valley region has the strongest geostrophic wind components in CONUS (Walsh et al. 1982) and is significantly influenced by ENSO conditions (Gershunov and Barnett 1998), along with high climate variability (Konapala and Mishra 2017; Zhang et al. 2010). In addition to cross-regional causal interactions over the nine climate regions, we

also conducted the state-level causality analysis with each state in CONUS as an individual node. The resulted causal graphs of state-level analysis (results are included in **Supplementary Fig. 3** in Supplementary Information) are in good agreement with those in **Fig. 4**, in particular, with all states in the Ohio Valley exhibiting strong causality strength.

3.3. Intrinsic modes of precipitation and temperature

Hydroclimatic causation is time-varying, whose temporal variability can be quite complicated with many local extrema but no zero crossings since the directed graphs are of positive causal strength. We apply the EMD analysis and decompose the time series of *ACE* and *ACS* of the precipitation and temperature causality into a collection of intrinsic modes (IMFs) for each climate region. The EMD method was previously applied to Earth system studies and generated novel and informative results. For example, the decomposition of Vostok temperature from ice cores reveals the role of three Milankovitch cycles that related to the Earth's eccentricity (about 100 ka), axial tilt (about 41 ka), and precession (about 23 ka), which impact the Earth's climate system in a nonlinear fashion (Huang and Wu 2008). The analysis of length-of-day by EMD identified a low-frequency variability (with interannual timescale) that can be related to the El Nino years (Gross et al. 1996; Huang and Wu 2008).

Since Ohio valley represents the causal gateway of hydroclimatic dynamics (**Fig. 4**), we choose it for HHT analysis. The time series of *ACE* and *ACS* and their correspondingly decomposed IMFs and residues are shown in **Figs. 5** and **6**, respectively. All the IMF are obtained from the repeated sifting processes, from which we extracted a total of 4 components (c_1 to c_4), revealing that the causality can be separated into 4 locally non-

overlapping time scale components. Despite the apparent difference in strength (**Fig. 3**), the IMFs of precipitation and temperature causality exhibit similar magnitudes in all modes. The first IMF c_1 has the highest frequency, corresponding to the fastest oscillating component in the time variability of causation. The residues of each panel signify the general trend of the causality variability, after all IMFs are subtracted from the ACE or ACS time series. We perform the EMD analysis for all nine CONUS climatic regions and found similar decomposition of IMFs (see **Supplementary Fig. 4** for the results of decomposition of temperature causal signal for all climate regions) with recurrence ranging from interannual to interdecadal scales.

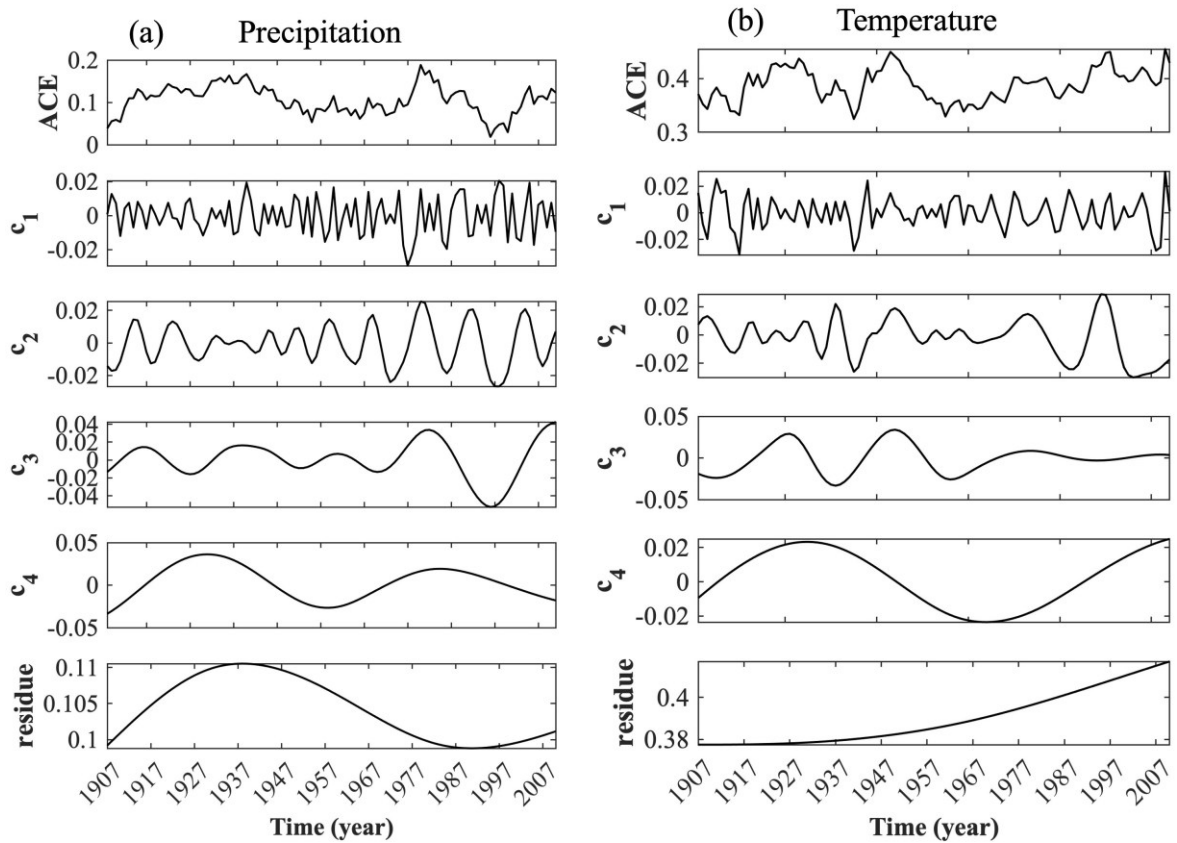


Figure 5. Results of EMD analysis for the time-varying causal effect. Shown are the ACE values in the Ohio Valley region for (a) precipitation and (b) temperature, respectively.

The top panels in (a) and (b) are the original *ACE* time series, while the lower panels from c_1 to c_4 represent decomposed IMFs from high to low frequency. The time series of *ACE* is obtained using moving temporal average with 15-year sliding window through the period of 1901-2018.

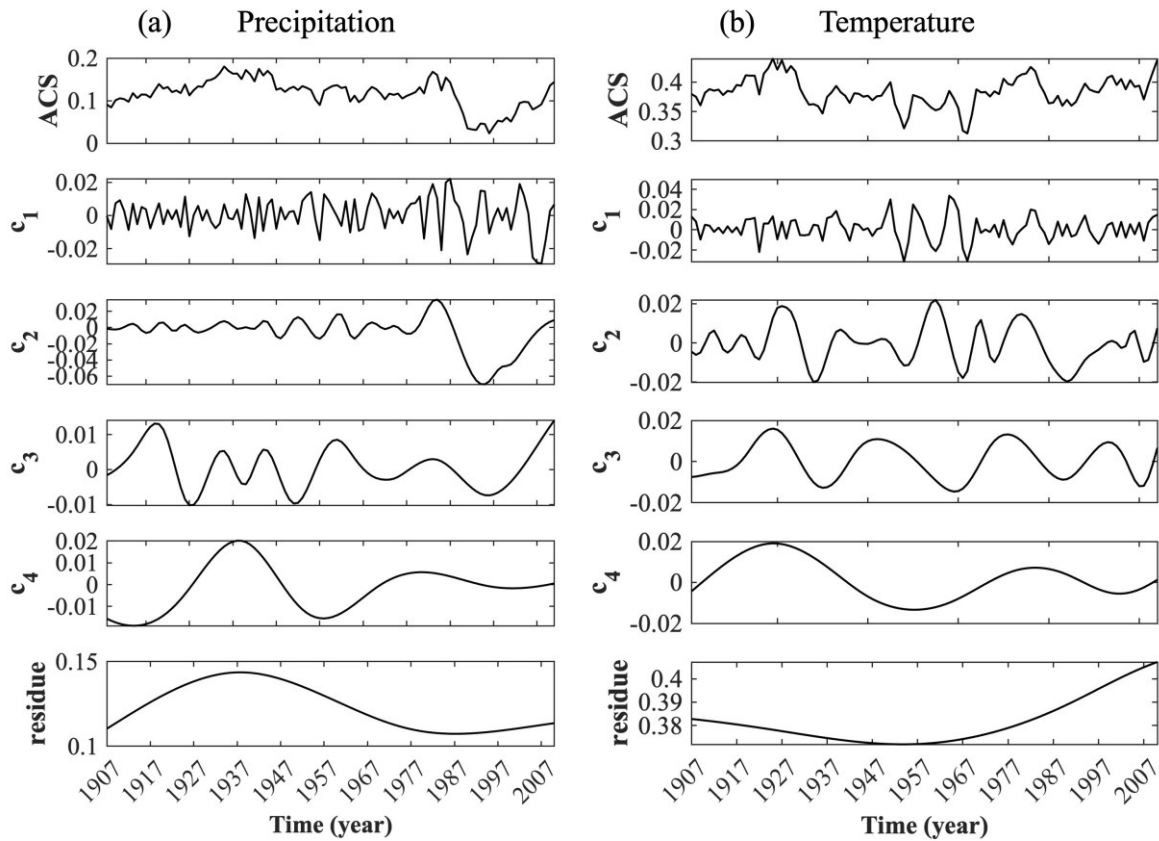


Figure 6. Results of EMD analysis for the time-varying causal effect. Shown are the *ACS* values in the Ohio Valley region for (a) precipitation and (b) temperature, respectively.

The top panels in (a) and (b) are the original *ACS* time series, while the lower panels from c_1 to c_4 represent decomposed IMFs from high to low frequency. A 15-year moving average is used here, same as in **Fig. 5**.

These decomposed IMFs in **Figs. 5** and **6** can be physically related to or interpreted by natural oscillations in Earth's hydroclimatic system reported in the literature. First, the IMF components c_1 and c_2 for both precipitation and temperature causality exhibit significant high-frequency periodicities at the interannual (1-10 years) scales. We believe this mode is related to the ENSO-related phenomena with internal modes of the ocean-atmosphere system in the tropical Pacific (Neelin et al. 1998), or some natural climate variability associated with the North Atlantic Oscillation, the Pacific Decadal Oscillation, or the Arctic Oscillation (Ghil and Lucarini 2020). As illustrated in **Fig. 1**, the natural climate variability with recurrence period of about 5.2 years is often attributable to the remote effect of ENSO low-frequency mode, where the ~ 7.7 -year cycle is likely due to the North Atlantic mode of variability that arises from the Gulf Stream's interannual cycle of meandering and intensification (Plaut et al. 1995). Furthermore, an important component of interannual climate variability, due to El Nino in the tropical Pacific Ocean, appears about every 2-7 years, which is largely characterized by the sea surface temperature anomalies (Ghil and Lucarini 2020; Philander 1983).

For the IMF components c_3 and c_4 , a periodicity of interdecadal scale (20-40 years) is found in both precipitation and temperature causation. This interdecadal mode is closely linked with the climate variability with about 14 and 25 years recurrence as shown in **Fig. 1**, which appears to be associated with the oscillations in the global ocean's thermohaline circulation and its coupling to the atmosphere above (Ghil and Lucarini 2020; Plaut et al. 1995). In addition, results of previous works on a complete decomposition of geopotential height (Coughlin and Tung 2004a, b) extracted a clear 11-year cycles in stratosphere, which can be potentially attributed to the cycle of solar activity and its downward

propagation to the lower troposphere. This could also be a potential source of the interdecadal variability and inherited by the IMF components c_3 and c_4 in cross-regional hydroclimatic causation found here.

To better understand the different IMFs of causality variability, as well as to strengthen the link with physical components in the climate system, we extend the EMD analysis to frequently used indices of climate oscillations, viz. ENSO and the Pacific Decadal Oscillation (PDO). More specifically, we choose ENSO Longitude Index (ELI) from Lawrence Berkely National Laboratory (<https://cascade.lbl.gov/enso-longitude-index-eli/>) and PDO indices from NOAA (<https://www.ncei.noaa.gov/access/monitoring/pdo/>), and extract the time variability during the same time period as the causality analysis. In particular, ELI tracks the average longitude of tropical Pacific deep convection and characterizes the diversity of ENSO in a single index. In addition, ELI accounts for the nonlinear response of deep convection to sea surface temperature (SST) and provides a continuous time series for analyses of ENSO dynamics (Williams and Patricola. 2018). On the other hand, PDO is often described as a long-live El Niño-like pattern of Pacific climate variability (Zhang et al. 1997). The index PDO is based on NOAA's extended reconstruction of SSTs. Extremes in the PDO patterns are marked by widespread variations in the Pacific Basin and North American climate.

The results of EMD decomposition are shown in **Fig. 7**. The four IMFs of both ENSO and PDO indices are similar with those of causality variability in hydroclimate (**Figs. 5 and 6**). In addition, the comparison of their characteristic (mean) periods is summarized in **Table 1**. It is clear that the mean periods of climate oscillations and causality variability are generally in good agreement, suggesting that the causality of

regional climatology in the U.S. has a strong correlation with the physical climate variability, potentially through climate teleconnections. It is also noteworthy that the discrepancy increases for lower frequency, i.e. c_3 and c_4 , which might be due to potential hysteresis effect for long-term climate oscillations to be manifest in causal interactions.

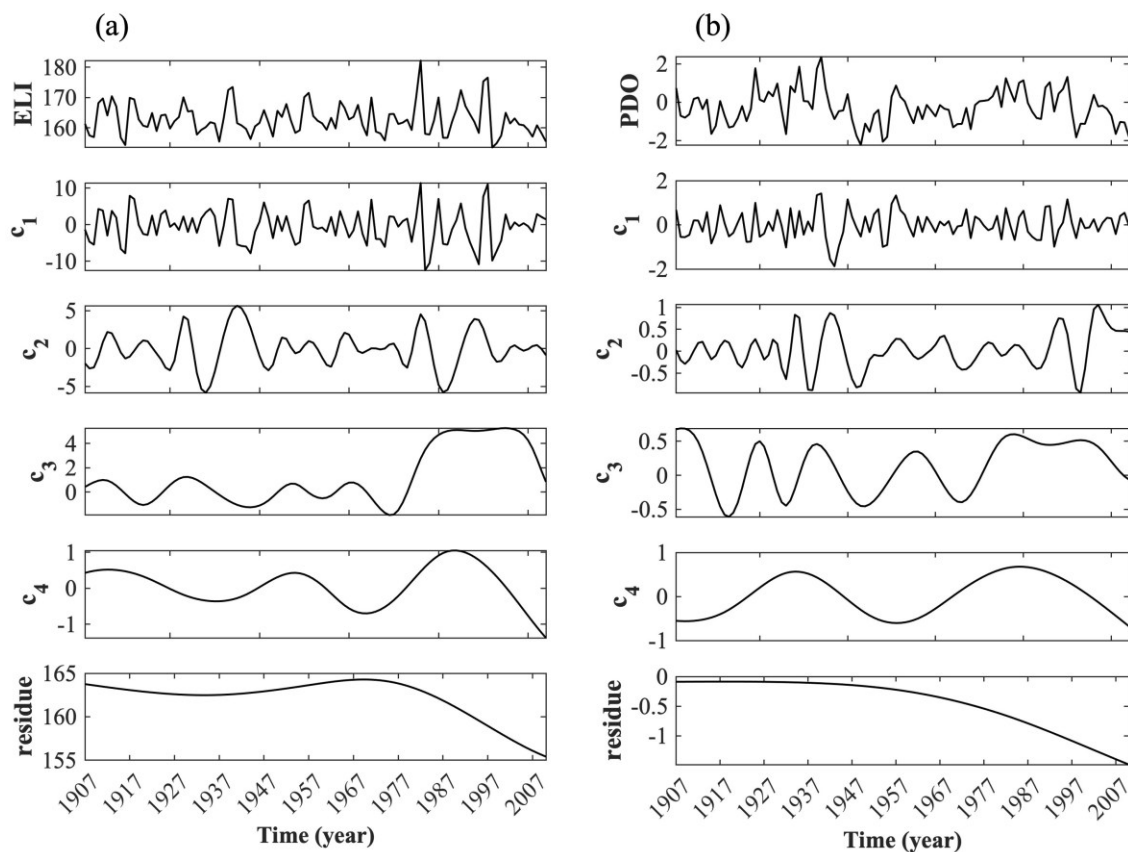


Figure 7. The EMD decompositions of climate oscillations indices in (a) ENSO ELI and (b) POD. c_1 and c_4 are the four IMFs extracted from the time series (top panels).

Table 1. Results of comparison between characteristic (mean) periods of representative climate variability and those of causal variability in U.S. hydroclimate.

Index/mean period (Year)		c_1	c_2	c_3	c_4
Climate variability	ELI	3.5	8	18	39
	PDO	3.4	7.5	18	51

	P (<i>ACE</i>)	3.1	9	22	53
Causal	P (<i>ACS</i>)	3.2	7	16	42
variability	T (<i>ACE</i>)	3.5	9.6	27	40+
	T (<i>ACS</i>)	3.6	10	25	58

3.4. The Hilbert spectrum

To understand the structures of the causality variability on precipitation and temperature in the long-term time series, the detailed information of the signal dispersion and the energy-frequency distribution is required. We calculate the Hilbert spectrum for each decomposed IMF. The results are weighted non-normalized joint amplitude-frequency-time distributions, which can be seen as four overlapping Hilbert spectra (corresponding to each of the 4 IMFs). Unlike Fourier analysis, the HHT analysis preserves the time localities of climate events, based on the *instantaneous*, rather than the global, frequency and energy. As shown in **Fig. 8**, the Hilbert spectrum is highly nodular (localized and spiky) especially in the high frequency range, indicating the causality variability is nonstationary for precipitation or temperature. In addition, the main component is intrawave modulated, signaling that both the precipitation and temperature causations are nonlinear. The presence of fewer components in the Hilbert spectrum as compared to the Fourier spectrum analysis is primarily due to the empirical nature of HHT, suggesting that it often yields more realistic energy-frequency distribution through eliminating spurious harmonics.

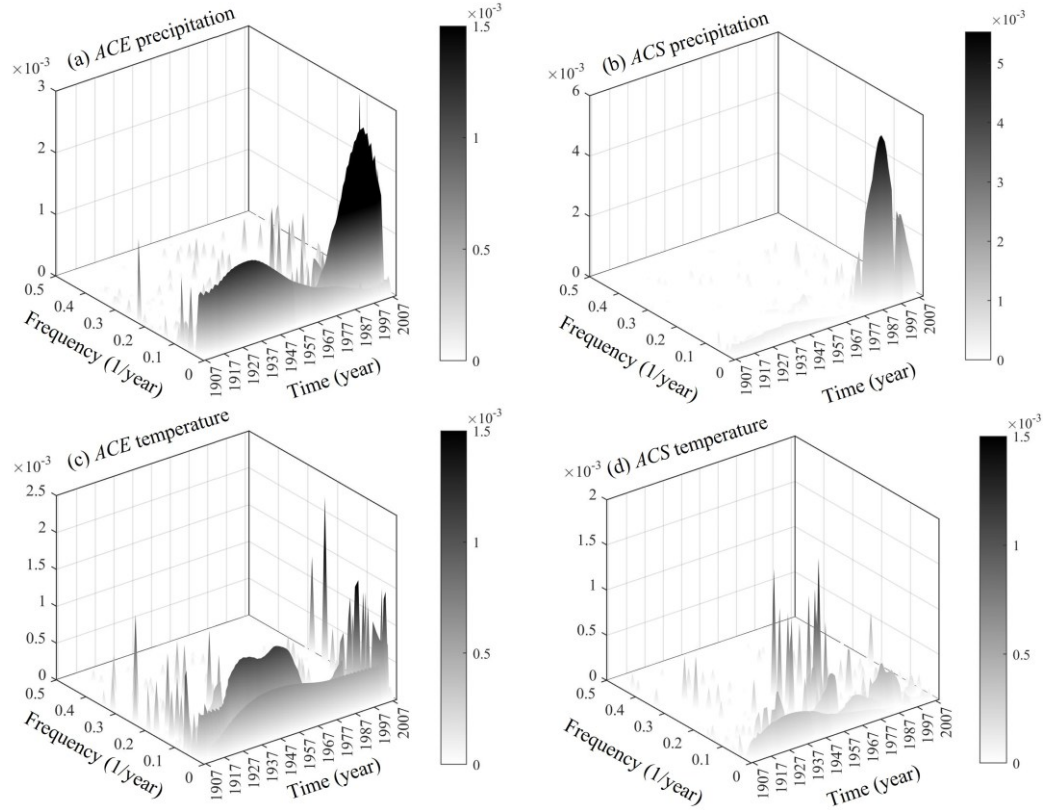


Figure 8. Hilbert spectrum (energy-frequency-time distribution) of the causality variability. Shown are the Hilbert spectra of ACE and ACS for (a, b) precipitation and (c, d) temperature, respectively.

The results in **Fig. 8** also reveal a pattern of energy concentration in the low-frequency range, or with high-amplitude and low-infrequency wave variations. Most energy is concentrated in the low-frequency range with the frequency less than 0.1 year^{-1} . In the temporal dimension, the Hilbert spectrum exhibits sharp energy peaks around year 1930s and 1990s. where the former can be the system response to the change in external forcing in the 1930s when Earth's climate system experienced significant warming together with enhanced climate variability (Diaz and Brandley 1995). The peak around

1990s coincides with a recently discovered potential early-warning signal of drastic changes in system dynamics in the CONUS hydroclimate (Wang et al. 2020; Yang et al. 2022a), where the underlying mechanism of this abrupt change remains obscure though. Furthermore, it is known that ENSO influences the precipitation and temperature along the U.S. East Coast, and the signatures can be identified by peaks in the Hilbert spectrum (Duffy 2004). The Hilbert spectrum from HHT applied to each component of daily Seine river flow detected the concentrating or localized amplitudes around the climate phase change around 1970s, which matches with the increasing amplitude of annual winter-months NAO variability phenomenon (Massei and Fournier 2012).

4. Concluding remarks

In this study, we combined causal inference with spectrum analysis to investigate the dynamics in the complex hydroclimatic system in the CONUS and its intrinsic periodicity. The reconstructed causal networks of precipitation and temperature anomalies indicate that the Ohio Valley region plays a key role in regulating the cross-regional atmospheric connection, while some regions such as Northwest have relatively weak causal interactions except with the West region. The results of HHT analysis reveal that the temporal variability of hydroclimate causation cross different regions of the CONUS contains manifest periodicities, primarily around the interannual and interdecadal scales. The variability of hydroclimatic causality is found to share similar intrinsic modes and characteristic frequencies with the recurrence on climate oscillations, such as ENSO and PDO dynamics. Moreover, the Hilbert spectrum exhibits a bi-modal distribution in the time dimensions, with peaks around 1930s and 1990s, both can be well correlated to the

changes in external forcing or system dynamics. To the best of our knowledge, this is a pioneering study that looks into causal interactions in hydroclimate system and links their variability to oscillations of physical climate indices.

In the broader context, the results of the current study suggest a promising perspective through the potential link between hydroclimatic causation and the physical climate variability. While this study represents an attempt in characterizing the cross-regional causal attribution and its intrinsic variability, the methods developed in this study can be used to infer how causal linkages change over time when being impacted by climate dynamics. Results from such hydroclimatic analysis show a potential to inform cross-regional water resource management that adapts to changing climate dynamics. Moreover, the proposed method can be further improved and incorporated into a holistic framework using multiple data-driven techniques. This toolkit includes, for example, complex network analysis to identify spatial patterns of hydroclimatic causal networks, physical emergence to correlate the peak Hilbert spectrum to abrupt and potentially catastrophic transitions in nonlinear dynamic systems, and machine learning to infer causality from archived earth system data. In particular, in complex systems that involves a wide spectrum of scales in their dynamics, such as the climate system evolves in the spectrum ranging from turbulence to synoptic scales, the cross-scale causation will lead to the *causal emergence* (Hoel et al., 2013) in macro-states from the information in microcosm. Such cross-scale causal interactions need necessarily be driven by big data describing the system evolution, that can be practically handled by machine learning techniques, such as deep neural networks (Marrow et al., 2020). When applied to the hydroclimate system, the method of

causal emergence will shed new lights on the causality in hydroclimate system to physical components, especially the low-frequency (macro-scale) climate variabilities.

Acknowledgement

This work was supported by the U.S. National Science Foundation (NSF) under Grant # AGS-2300548 and CBET-2028868, and by the National Aeronautics and Space Administration (NASA) under grant # 80NSSC20K1263. YCL was supported by the Office of Naval Research under Grant No. N00014-21-1-2323.

Data Availability Statement

The gridded dataset for temperature and precipitation during 1901–2018 is obtained from the Climatic Research Unit (CRU) Time-Series (TS) version 4.03. The dataset is archived by the Center for Environmental Data Analysis (CEDA) and 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.ncdc.noaa.gov/monitoring-references/maps/us-climate-regions.php>. The climate oscillation can be found at <https://cascade.lbl.gov/enso-longitude-index-eli/> for ENSO dynamics and <https://www.ncei.noaa.gov/access/monitoring/pdo/> for PDO dynamics.

References:

- Alberti, T., Milillo, A., Laurenza, M., Massetti, S., Ivanovski, S.L., Ippolito, A., Plainaki, C., De Angelis, E., Mangano, V., Mura, A., Orsini, S., Rispoli, R., 2021. Multiscale features of the near-Hermean environment as derived through the Hilbert-Huang transform. *Front. Phys.* 9, 668098. <https://doi.org/10.3389/fphy.2021.668098>
- Alberti, T., Anzidei, M., Faranda, D., Vecchio, A., Favaro, M., Papa, A., 2023. Dynamical diagnostic of extreme events in Venice lagoon and their mitigation with the MoSE. *Sci. Rep.* 13(1), 10475. <https://doi.org/10.1038/s41598-023-36816-8>
- Boashash, B., 2016. Chapter 4 - Advanced Time-Frequency Signal and System Analysis. In *Time-Frequency Signal Analysis and Processing (Second Edition)*. Academic Press, Oxford, 141-236 pp. <https://doi.org/10.1016/B978-0-12-398499-9.00004-2>
- Bjerknes, J., 1969. Atmospheric teleconnections from the equatorial Pacifica. *Mon. Weather Rev.* 97, 163–172. [https://doi.org/10.1175/1520-0493\(1969\)097<0163:ATFTEP>2.3.CO;2](https://doi.org/10.1175/1520-0493(1969)097<0163:ATFTEP>2.3.CO;2)
- Boers, N., Goswami, B., Rheinwalt, A., Bookhagen, B., Hoskins, B., Kurths, J., 2019. Complex networks reveal global pattern of extreme-rainfall teleconnections. *Nature* 566, 373–377. <https://doi.org/10.1038/s41586-018-0872-x>
- Coughlin, K., Tung, K.K., 2004a. Eleven-year solar cycle signal throughout the lower atmosphere. *J. Geophys. Res. Atmos.* 109. <https://doi.org/10.1029/2004JD004873>
- Coughlin, K.T., Tung, K.K., 2004b. 11-Year solar cycle in the stratosphere extracted by the empirical mode decomposition method. *Adv. Space Res.* 34, 323–329. <https://doi.org/10.1016/j.asr.2003.02.045>
- Diaz, H. F., & Bradley, R. S., 1995. Documenting natural climatic variations: how

different is the climate of the twentieth century from that of previous centuries. *Natural climatic variability on decade-to-century time scales*. Edited by DG Martinson, K. Bruan, M. Ghil, MM Hall, TR Karl, ES Sarachik, S. Sorooshian, and LD Talley. National Academy Press, Washington, DC, 17-31.

Dijkstra, H.A., Ghil, M., 2005. Low-frequency variability of the large-scale ocean circulation: A dynamical systems approach. *Rev. Geophys.* 43.
<https://doi.org/10.1029/2002RG000122>

Duffy, D.G., 2004. The application of Hilbert–Huang transforms to meteorological datasets. *J. Atmos. Ocean. Technol.* 21, 599–611. [https://doi.org/10.1175/1520-0426\(2004\)021<0599:TAOHTT>2.0.CO;2](https://doi.org/10.1175/1520-0426(2004)021<0599:TAOHTT>2.0.CO;2)

Gao, B., Li, M., Wang, J., Chen, Z., 2022. Temporally or spatially? Causation inference in Earth system sciences. *Sci. Bull.* 67, 232–235.
<https://doi.org/10.1016/j.scib.2021.10.002>

Gershunov, A., Barnett, T.P., 1998. Interdecadal modulation of ENSO teleconnections. *Bull. Am. Meteorol. Soc.* 79, 2715–2726. [https://doi.org/10.1175/1520-0477\(1998\)079<2715:IMOET>2.0.CO;2](https://doi.org/10.1175/1520-0477(1998)079<2715:IMOET>2.0.CO;2)

Ghil, M., 2002. Natural climate variability. *Encyclopedia of Global Environmental Change, I*, 544-549.

Ghil, M., Lucarini, V., 2020. The physics of climate variability and climate change. *Rev. Mod. Phys.* 92, 035002. <https://doi.org/10.1103/RevModPhys.92.035002>

Ghil, M., and A.W. Robertson, 2000. Chapter 10 -Solving problems with GCMs: general circulation models and their role in the climate modeling hierarchy, in: Randall,

D.A. (Ed.), International Geophysics, General Circulation Model Development.
Academic Press, pp. 285–325. [https://doi.org/10.1016/S0074-6142\(00\)80058-3](https://doi.org/10.1016/S0074-6142(00)80058-3)

Ghil, M., Vautard, R., 1991. Interdecadal oscillations and the warming trend in global
temperature time series. *Nature* 350, 324–327. <https://doi.org/10.1038/350324a0>

Ghil, M., Allen, M.R., Dettinger, M.D., Ide, K., Kondrashov, D., Mann, M.E., Robertson,
A.W., Saunders, A., Tian, Y., Varadi, F., Yiou, P., 2002. Advanced spectral
methods for climatic time series. *Rev. Geophys.* 40, 3-1-3–41.
<https://doi.org/10.1029/2000RG000092>

Granger, C.W., 1969. Investigating causal relations by econometric models and cross-
spectral methods. *Econometrica* 37, 424–438. <https://doi.org/10.2307/1912791>

Grassberger, P., Procaccia, I., 1983. Characterization of strange attractors. *Phys. Rev. Lett.*
50, 346–349. <https://doi.org/10.1103/PhysRevLett.50.346>

Gross, R.S., Marcus, S.L., Eubanks, T.M., Dickey, J.O., Keppenne, C.L., 1996. Detection
of an ENSO signal in seasonal length-of-day variations. *Geophys. Res. Lett.* 23,
3373–3376. <https://doi.org/10.1029/96GL03260>

Hannart, A., Pearl, J., Otto, F. E. L., Naveau, P., Ghil, M., 2016. Causal counterfactual
theory for the attribution of weather and climate-related events. *Bull. Amer. Met.*
Soc. 97(1), 99-110. <https://doi.org/https://doi.org/10.1175/BAMS-D-14-00034.1>

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

649 Hoel, E. P., Albantakis, L., Tononi, G., 2013. Quantifying causal emergence shows that
 650 macro can beat micro. *Proc. Nat. Acad. Sci. USA* 110(49), 19790-19795.
 651 <https://doi.org/10.1073/pnas.1314922110>
 652 Huang, L., Ni, X., Ditto, W.L., Spano, M., Carney, P.R., Lai, Y.-C., 2017. Detecting and
 653 characterizing high-frequency oscillations in epilepsy: A case study of big data
 654 analysis. *R. Soc. Open Sci.* 4, 160741. <https://doi.org/10.1098/rsos.160741>
 655 Huang, N.E., Wu, Z., 2008. A review on Hilbert-Huang transform: Method and its
 656 applications to geophysical studies. *Rev. Geophys.* 46.
 657 <https://doi.org/10.1029/2007RG000228>
 658 Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.-C., Tung,
 659 C.C., Liu, H.H., 1998. The empirical mode decomposition and the Hilbert spectrum
 660 for nonlinear and non-stationary time series analysis. *Proc. R. Soc. Lond. Ser.*
 661 *Math. Phys. Eng. Sci.* 454, 903–995. <https://doi.org/10.1098/rspa.1998.0193>
 662 Konapala, G., Mishra, A., 2017. Review of complex networks application in hydroclimatic
 663 extremes with an implementation to characterize spatio-temporal drought
 664 propagation in continental USA. *J. Hydrol.* 555, 600–620.
 665 <https://doi.org/10.1016/j.jhydrol.2017.10.033>
 666 Kretschmer, M., Coumou, D., Donges, J.F., Runge, J., 2016. Using causal effect networks
 667 to analyze different Arctic drivers of midlatitude winter circulation. *J. Clim.* 29,
 668 4069–4081. <https://doi.org/10.1175/JCLI-D-15-0654.1>
 669 Lai, Y.-C., 1998. Analytic signals and the transition to chaos in deterministic flows. *Phys.*
 670 *Rev. E* 58, R6911–R6914. <https://doi.org/10.1103/PhysRevE.58.R6911>

671 Lai, Y.-C., Ye, N., 2003. Recent developments in chaotic Time Series Analysis. *Int. J.*
672 *Bifurc. Chaos* 13, 1383–1422. <https://doi.org/10.1142/S0218127403007308>

673 Leng, S., Ma, H., Kurths, J., Lai, Y.-C., Lin, W., Aihara, K., Chen, L., 2020. Partial cross
674 mapping eliminates indirect causal influences. *Nat. Commun.* 11, 2632.
675 <https://doi.org/10.1038/s41467-020-16238-0>

676 Li, P., Xu, T., Wei, S., Wang, Z.-H., 2022. Multi-objective optimization of urban
677 environmental system design using machine learning. *Comput. Environ. Urban*
678 *Syst.* 94, 101796.
679 <https://doi.org/https://doi.org/10.1016/j.compenvurbsys.2022.101796>

680 Lucarini, V., Blender, R., Herbert, C., Ragone, F., Pascale, S., Wouters, J., 2014.
681 Mathematical and physical ideas for climate science. *Rev. Geophys.* 52, 809–859.
682 <https://doi.org/10.1002/2013RG000446>

683 Lucarini, V., Chekroun, M.D., 2023. Theoretical tools for understanding the climate crisis
684 from Hasselmann’s programme and beyond. *Nature Rev. Phys.* 5(12), 744–765.
685 <https://doi.org/10.1038/s42254-023-00650-8>

686 Marrow, S., Michaud, E. J., Hoel, E., 2020. Examining the causal structures of deep neural
687 networks using information theory. *Entropy* 22(12), e22121429.
688 <https://doi.org/10.3390/e22121429>

689 Massei, N., Fournier, M., 2012. Assessing the expression of large-scale climatic
690 fluctuations in the hydrological variability of daily Seine river flow (France)
691 between 1950 and 2008 using Hilbert–Huang Transform. *J. Hydrol.* 448–449, 119–
692 128. <https://doi.org/10.1016/j.jhydrol.2012.04.052>

693 McGraw, M.C., Barnes, E.A., 2018. Memory matters: a case for Granger causality in
694 climate variability studies. *J. Clim.* 31, 3289–3300. [https://doi.org/10.1175/JCLI-D-](https://doi.org/10.1175/JCLI-D-17-0334.1)
695 17-0334.1

696 Neelin, J.D., Battisti, D.S., Hirst, A.C., Jin, F.-F., Wakata, Y., Yamagata, T., Zebiak, S.E.,
697 1998. ENSO theory. *J. Geophys. Res. Oceans* 103, 14261–14290.
698 <https://doi.org/10.1029/97JC03424>

699 Philander, S.G.H., 1983. El Niño Southern Oscillation phenomena. *Nature* 302, 295–301.
700 <https://doi.org/10.1038/302295a0>

701 Plaut, G., Ghil, M., Vautard, R., 1995. Interannual and interdecadal variability in 335 years
702 of central England temperatures. *Science* 268, 710–713.
703 <https://doi.org/10.1126/science.268.5211.710>

704 Runge, J., Nowack, P., Kretschmer, M., Flaxman, S., Sejdinovic, D., 2019a. Detecting and
705 quantifying causal associations in large nonlinear time series datasets. *Sci. Adv.* 5,
706 eaau4996. <https://doi.org/10.1126/sciadv.aau4996>

707 Runge, J., Bathiany, S., Bollt, E., Camps-Valls, G., Coumou, D., Deyle, E., Glymour, C.,
708 Kretschmer, M., Mahecha, M.D., Muñoz-Marí, J., van Nes, E.H., Peters, J., Quax,
709 R., Reichstein, M., Scheffer, M., Schölkopf, B., Spirtes, P., Sugihara, G., Sun, J.,
710 Zhang, K., Zscheischler, J., 2019b. Inferring causation from time series in Earth
711 system sciences. *Nat. Commun.* 10, 2553. [https://doi.org/10.1038/s41467-019-](https://doi.org/10.1038/s41467-019-10105-3)
712 10105-3

713 Runge, J., Petoukhov, V., Donges, J.F., Hlinka, J., Jajcay, N., Vejmelka, M., Hartman, D.,
714 Marwan, N., Paluš, M., Kurths, J., 2015. Identifying causal gateways and mediators

715 in complex spatio-temporal systems. *Nat. Commun.* 6, 8502.
 716 <https://doi.org/10.1038/ncomms9502>
 717 Silva, F.N., Vega-Oliveros, D.A., Yan, X., Flammini, A., Menczer, F., Radicchi, F.,
 718 Kravitz, B., Fortunato, S., 2021. Detecting climate teleconnections with Granger
 719 causality. *Geophys. Res. Lett.* 48, e2021GL094707.
 720 <https://doi.org/10.1029/2021GL094707>
 721 Sugihara, G., May, R., Ye, H., Hsieh, C.-H., Deyle, E., Fogarty, M., Munch, S., 2012.
 722 Detecting causality in complex ecosystems. *Science* 338, 496–500.
 723 <https://doi.org/10.1126/science.1227079>
 724 Sugihara, G., May, R.M., 1990. Nonlinear forecasting as a way of distinguishing chaos
 725 from measurement error in time series. *Nature* 344, 734–741.
 726 <https://doi.org/10.1038/344734a0>
 727 Takens, F., 1981. Detecting strange attractors in fluid turbulence, in: Rand, D., Young,
 728 L.S. (Eds.), *Dynamical Systems and Turbulence*, Lecture Notes in Mathematics.
 729 Springer-Verlag, Berlin, 366–381.
 730 Tsonis, A.A., Swanson, K.L., Wang, G., 2008. On the role of atmospheric teleconnections
 731 in climate. *J. Clim.* 21, 2990–3001. <https://doi.org/10.1175/2007JCLI1907.1>
 732 van Nes, E.H., Scheffer, M., Brovkin, V., Lenton, T.M., Ye, H., Deyle, E., Sugihara, G.,
 733 2015. Causal feedbacks in climate change. *Nat. Clim. Change* 5, 445–448.
 734 <https://doi.org/10.1038/nclimate2568>
 735 Vejmelka, M., Pokorná, L., Hlinka, J., Hartman, D., Jajcay, N. and Paluš, M., 2015. Non-
 736 random correlation structures and dimensionality reduction in multivariate climate

737 data. *Climate Dynamics*, 44, 2663-2682. <https://doi.org/10.1007/s00382-014-2244->
738 z

739 Walsh, J.E., Richman, M.B., Allen, D.W., 1982. Spatial coherence of monthly
740 precipitation in the United States. *Mon. Wea. Rev.* 110, 272–286.
741 [https://doi.org/10.1175/1520-0493\(1982\)110<0272:SCOMPI>2.0.CO;2](https://doi.org/10.1175/1520-0493(1982)110<0272:SCOMPI>2.0.CO;2)

742 Wang, C., Wang, Z.H., 2020. A network-based toolkit for evaluation and intercomparison
743 of weather prediction and climate modeling. *J. Environ. Manage.* 268, 110709.
744 <https://doi.org/10.1016/j.jenvman.2020.110709>

745 Wang, C., Wang, Z.H., Sun, L., 2020. Early warning signals for critical temperature
746 transition. *Geophys. Res. Lett.*, 47, e2020GL088503.
747 <https://doi.org/10.1029/2020GL088503>

748 Wang, Y., Yang, X., Wang, Z.-H., 2024. Causal mediation of urban temperature by
749 geopotential height in U.S. cities. *Sus. Cities Soc.* 100, 105010.
750 <https://doi.org/https://doi.org/10.1016/j.scs.2023.105010>

751 Williams, I.N. and Patricola, C.M., 2018. Diversity of ENSO events unified by convective
752 threshold sea surface temperature: A nonlinear ENSO index. *Geophys. Res.*
753 *Lett.* 45(17), 9236-9244. <https://doi.org/10.1029/2018GL079203>

754 Yalçınkaya, T., Lai, Y.-C., 1997. Phase characterization of chaos. *Phys. Rev. Lett.* 79,
755 3885–3888. <https://doi.org/10.1103/PhysRevLett.79.3885>

756 Yang, X., Wang, Z.H., Wang, C., 2022a. Critical transitions in the hydrological system:
757 early-warning signals and network analysis. *Hydrol. Earth Syst. Sci.* 26(7), 1845-
758 1856. <https://doi.org/10.5194/hess-26-1845-2022>

759 Yang, X., Wang, Z.H., Wang, C., Lai, Y.C., 2022b. Detecting the causal influence of
760 thermal environments among climate regions in the United States. *J. Environ.*
761 *Manag.* 322, 116001. <https://doi.org/10.1016/j.jenvman.2022.116001>

762 Yang, X., Wang, Z.H., Wang, C., Lai, Y.-C., 2023a. Finding causal gateways of
763 precipitation over the contiguous United States, *Geophys. Res. Lett.* 50(4),
764 e2022GL101942. <https://doi.org/10.1029/2022GL101942>

765 Yang, X., Li, P., Wang, Z.H., 2023b. The impact of urban irrigation on the temperature-
766 carbon feedback in U.S. cities. *J. Env. Manag.* 344, 118452.
767 <https://doi.org/https://doi.org/10.1016/j.jenvman.2023.118452>

768 Zhang, X., Wang, J., Zwiers, F.W., Groisman, P.Y., 2010. The influence of large-scale
769 climate variability on winter maximum daily precipitation over North America. *J.*
770 *Clim.* 23, 2902–2915. <https://doi.org/10.1175/2010JCLI3249.1>

771 Zhang, Y., Wallace, J.M. and Battisti, D.S., 1997. ENSO-like interdecadal variability:
772 1900–93. *J. Clim.* 10(5), 1004-1020.