

Estimating Time-Dependent Structures in a Multivariate Causality for Land–Atmosphere Interactions

FEIHONG ZHOU,^a DANIEL FIIFI TAWIA HAGAN,^{a,b} GUOJIE WANG,^a X. SAN LIANG,^{c,d} SHIJIE LI,^{a,e,f} YUHAO SHAO,^a EMMANUEL YEBOAH,^a AND XIKUN WEI^{a,g}

^a Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters, Nanjing University of Information Science and Technology, Nanjing, Jiangsu, China

^b Hydro-Climate Extremes Laboratory, Ghent University, Ghent, Belgium

^c Department of Atmospheric and Oceanic Sciences, and Institute of Atmospheric Sciences, Fudan University, Shanghai, China

^d Division of Frontier Research, Southern Marine Laboratory, Zhuhai, Guangdong, China

^e Department of Remote Sensing, Helmholtz Centre for Environmental Research-UFZ, Leipzig, Germany

^f Remote Sensing Centre for Earth System Research, Leipzig University, Leipzig, Germany

^g Chair of Engineering Hydrology and Water Management, Technical University of Darmstadt, Darmstadt, Germany

(Manuscript received 5 April 2023, in final form 26 October 2023, accepted 29 December 2023)

ABSTRACT: The land surface and atmosphere interaction forms an integral part of the climate system. However, this intricate relationship involves many complicated interactions and feedback effects between multiple variables. As a result, relying solely on traditional linear regression analysis and correlation analysis to distinguish between multivariate complex “driver–response” relations can be challenging, since they do not have the needed asymmetry to establish causality. The Liang–Kleeman (LK) information flow theory provides a strict nonparametric causality measurement for identifying the causality between any given time series, and its recent extension from bivariate to multivariate form provides a powerful tool for causal inference in complex multivariate systems. However, the multivariate LK information flow also assumes stationarity in time and requires a sufficiently long time series to ensure statistical sufficiency. To remedy this challenge, we rely on the square-root Kalman filter to estimate the time-varying form of the multivariate LK information flow causality. The results from theoretical and real-world applications show that the new algorithm provides a valuable tool for characterizing time-varying causal relationships in land–atmosphere interactions, even when the time series are short and highly correlated.

SIGNIFICANCE STATEMENT: Causality in land–atmosphere interactions is generally characterized by seasonal and intraseasonal changes that are usually not captured with commonly used approaches, because most approaches assume the time series are stationary. In this study, we extend the recently proposed multivariate Liang–Kleeman information flow causality (MtvLK) to handle nonstationary systems such as those in land–atmosphere interactions. By considering nonstationarity, we aim to unravel time-varying causal structures that are usually masked out in commonly used methods. Validating the MtvLK with synthetic models showed that the MtvLK is able to obtain the expected causal structures. Furthermore, real-world applications reveal novel findings of the time-varying causal structures between soil moisture, vapor pressure deficit, and the gross primary product.

KEYWORDS: Atmosphere–land interaction; Soil moisture; Kalman filters

1. Introduction

Land–atmosphere interactions play a critical role in the climate system, not only as a key component but also as an essential driver of terrestrial surface water and energy balance (Orth and Seneviratne 2017; Seneviratne et al. 2006, 2010). These interactions are characterized by complex couplings and feedbacks between land and atmospheric states and complex systems that include various natural processes (Schwingshackl et al. 2017). Furthermore, these complex natural processes and interactions play an important role in the exchange of energy and water, which sometimes results in amplifying temperature and humidity anomalies (Humphrey et al. 2021; Zhou et al. 2019), exacerbating soil desiccation and atmospheric aridity (Zhou et al. 2019), leading to heat

waves and droughts (Berg et al. 2016; Ciabatta et al. 2015). They also affect regional convection and precipitation (Tuttle and Salvucci 2016), leading to flooding in anomalously wet times (Saini et al. 2016). Therefore, a good understanding of the complex processes involved in land–atmosphere interactions is vital in improving the prediction of related extreme events (Lemordant et al. 2016). Although land–atmosphere interactions often involve multiple interacting variables, most of the commonly used analysis methods can assess only two variables at a time, making it difficult to account for the impact of confounding variables. In addition, various complex natural processes in land–atmosphere interactions usually vary not only in space (Lintner and Neelin 2009), but also at different time scales ranging from days to months or even years (Baker et al. 2021; Duerinck et al. 2016). It is, therefore, necessary to look beyond bivariate and time-invariable analyses to better understand these complex spatiotemporal characteristics.

Corresponding author: Guojie Wang, gwang@nuist.edu.cn

DOI: 10.1175/JCLI-D-23-0207.1

© 2024 American Meteorological Society. This published article is licensed under the terms of the default AMS reuse license. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

The approaches for quantifying land–atmosphere interactions include physical and numerical simulation experiments that intervene in the interactions of interest under well-controlled conditions and using data science methods for causal inference. In recent years, several data-driven methods have been applied to time series datasets to estimate potential causal dependencies. Traditionally, commonly used approaches are based on correlation and regression analysis. For example, [Duerinck et al. \(2016\)](#) conducted a correlation-based study on the relationship between soil moisture and precipitation in Illinois at different time and spatial scales. Their results showed a strong positive correlation between the state average soil moisture in late spring/early summer and the subsequent state average summer precipitation, while no relationship was found at the daily and weekly scales. [Pal et al. \(2020\)](#) presented the relationship between soil moisture and boundary layer depths (BLDs) after flooding at two weather scales based on regression analysis, effectively improving the prediction of extreme events. Although correlation and regression analysis may help provide causal inferences, they lack the necessary asymmetry to provide causal directions in the interactions between the variables under consideration ([Fu et al. 2022](#); [Novick et al. 2016](#); [Wu et al. 2015](#); [Zhang et al. 2021](#)). As a result, it is difficult to extrapolate the “drive–response” relationship.

The concept of causal inference originated in the early twentieth century ([Wright 1921](#)) and since then several studies have attempted to quantify it. [Granger \(1969\)](#) used a statistical prediction formalism to quantify causal relations. The Granger causality (GC) assumes that for two time series X and Y , X Granger-causes Y if the conditional variance of Y decreases when the knowledge of the past X is provided ([Granger 1969](#)). GC has become one of the most used tools for causal discoveries between time series and has consistently been modified to expand its applications in several fields, including climate science. For instance, [Detto et al. \(2012\)](#) proposed a conditional GC analysis framework, which can eliminate the influence of external forcing and separate the inherent periodic coupling signals of regional eco-climate systems. [Green et al. \(2017\)](#) used this approach to study vegetation–atmosphere feedback mechanism and found that the local feedback mechanism could explain about 30% of surface net radiation and precipitation variance. [Papagiannopoulou et al. \(2017a,b\)](#) developed and applied a nonlinear multivariate Granger causality approach to study the relationship between global vegetation and climate factors and found that water was the main controlling factor of vegetation anomalies, while radiation and temperature had little influence.

Beyond the GC methods, information theory (IT) has been demonstrated to be a rigorous alternative for identifying and quantifying causality ([Goodwell et al. 2020](#)). Since IT measures are based on probability distribution functions (pdfs), they can capture linear and nonlinear relationships between variables. In addition, IT-based methods can be reliably extended to quantify causal relationships between more than two variables ([Goodwell et al. 2020](#)). IT-based methods thus provide a framework within which causality can be considered in Earth system science, providing a more powerful tool for studying complex causal relationships in land–atmosphere interactions. In recent times, [Liang \(2014\)](#) proposed a new causality analysis method based on information flow theory, namely the Liang–Kleeman (LK) information flow

([Liang 2016](#)), which is a strict nonparametric causality measurement form for identifying the causality between any given time series. Compared with the GC, LK causality is easier to compute, since it only involves the sample covariance of time series. Moreover, since it originates from first principles and strict mathematical derivation, it avoids the false causalities sometimes found in GC test and other statistical forms ([Hagan et al. 2019](#)).

Since its development, the LK causality has been successfully applied to causality studies in climate systems. For instance, [Liang \(2014\)](#) used the method to extrapolate the causality between El Niño and the Indian Ocean dipole (IOD). [Stips et al. \(2016\)](#) used this method to study the causality between greenhouse gas emissions and global warming. [Bai et al. \(2018\)](#) also used it to analyze the causes of tropical cyclones in the western North Pacific. More recently, [Tao et al. \(2021\)](#) used the LK causality to quantitatively assess the impacts of climate warming, interdecadal oscillations in the Atlantic and Pacific on global precipitation and their regional differences. However, the original form of the LK assumes stationarity, therefore, it does not account for the time-dependent changes in land–atmosphere interactions. To deal with this problem, [Hagan et al. \(2019\)](#) extended the Liang causality to a time-frequency form by combining the Kalman filter and wavelet analysis and applied it to the interaction between soil moisture and air temperature over China, providing a new tool for studying temporal causal structures in land–atmosphere interactions. However, the approach presented by [Hagan et al. \(2019\)](#) is restricted to causality between only two variables and does not consider the effects of potential confounding variables. Therefore, seeking a causality analysis method that can address nonstationarity in the multivariate causality analysis is of paramount importance for the study of land–atmospheric interactions.

Recently, [Liang \(2021\)](#) extended the LK causality to multivariate time series causality and causal graph reconstruction, although the time series are also assumed to be stationary. [Liang et al. \(2021\)](#) successfully used this method to obtain accurate forecasts of El Niño Modoki 10 years in advance. [Docquier et al. \(2022\)](#) used this method to analyze the effects of several potential climate drivers on Arctic sea ice area and volume, and the reverse effects of sea ice area and volume on these drivers. [Hagan et al. \(2022\)](#) successfully applied it to analyze confounding impacts of soil moisture and vegetation on the causal structure of CO₂ emissions and near-surface mean air temperature couplings for multiple future scenarios of the climate system. Moreover, this theory and methodology, albeit originally born from atmosphere–ocean–climate science, has been widely applied in other disciplines such as quantum mechanics ([Yi and Bose 2022](#)) and neuroscience ([Cong et al. 2023](#)). These studies demonstrate the usefulness of the multivariate LK causality in practical applications. Therefore, this study combines the multivariate LK causality with the Kalman filter to provide a new statistical tool for estimating the time-varying causalities in multivariate land–atmosphere interactions. The rest of the paper is structured as follows: [section 2](#) describes the formalism, [section 3](#) validates the formalism with synthetic models, and [section 4](#) applies it to a real-world case study of vegetation–climate interactions in China. [Section 5](#) presents a summary and discussion of the proposed method and its potential applications.

2. The multivariate time-varying causal inference

a. Estimation of information flow and causality in multivariate time series

Liang (2016) observed that causality is a real physical notion, which can be deduced and quantified by the concept of Liang–Kleeman information flow (LKIF). Liang (2021) realized the causality analysis of multivariate information flow based on a rigorous derivation and explicit formulas for multidimensional information flow has been obtained in a closed form with both deterministic and stochastic systems. The following briefly introduces the derivation process of multivariate information flow method.

Consider a d -dimensional continuous-time stochastic system for $\mathbf{X} = (X_1, \dots, X_d)$

$$d\mathbf{X} = \mathbf{F}(\mathbf{X}, t)dt + \mathbf{B}(\mathbf{X}, t)d\mathbf{W} \quad (1)$$

where $\mathbf{F}$ may be arbitrary nonlinear functions of $\mathbf{X}$ and t , $\mathbf{W}$ is a vector of standard Wiener processes, and $\mathbf{B}$ is the matrix of perturbation amplitudes. The rate of information flowing from X_j to X_i (in nats per unit time) proves to be

$$\begin{aligned} T_{j \rightarrow i} &= -E \left[\frac{1}{\rho_i} \int_{\mathbb{R}^{d-2}} \frac{\partial(F_i \rho_{\mathbf{X}})}{\partial x_i} d\mathbf{x}_{\mathbf{X}} \right] + \frac{1}{2} E \left[\frac{1}{\rho_i} \int_{\mathbb{R}^{d-2}} \frac{\partial^2(g_{ii} \rho_{\mathbf{X}})}{\partial x_i^2} d\mathbf{x}_{\mathbf{X}} \right] \\ &= - \int_{\mathbb{R}^d} \rho_{ji}(x_j | x_i) \frac{\partial(F_i \rho_{\mathbf{X}})}{\partial x_i} d\mathbf{x} + \frac{1}{2} \int_{\mathbb{R}^d} \rho_{ji}(x_j | x_i) \frac{\partial^2(g_{ii} \rho_{\mathbf{X}})}{\partial x_i^2} d\mathbf{x}, \end{aligned} \quad (2)$$

where $d\mathbf{x}_{\mathbf{X}}$ signifies $dx_1, \dots, dx_{i-1}dx_{i+1}, \dots, dx_{j-1}dx_{j+1}, \dots, dx_d$, E is the mathematical expectation, $g_{ii} = \sum_{k=1}^d b_{ik}b_{ik}$, $\rho_i = \rho_i(x_i)$ is the marginal probability density function (pdf) of X_i , ρ_{ji} is the pdf of X_j conditioned on X_i , and $\rho_{\mathbf{X}} = \int \rho(x)dx$.

In Eq. (1), in the case of linear systems where $\mathbf{F}(\mathbf{X}, t) = \mathbf{A}\mathbf{X}$, $\mathbf{A} = (a_{ij})$, Eq. (2) becomes quite simple (Liang 2016):

$$T_{j \rightarrow i} = a_{ij} \frac{\sigma_{ij}}{\sigma_{ii}}, \quad (3)$$

where a_{ij} is the (i, j) th entry of $\mathbf{A}$, and σ_{ij} is the population covariance between X_i and X_j .

Since Eq. (3) is challenging to use in practical applications, we need to obtain a practically applicable formula by estimating Eq. (3). Suppose that the d time series are all equally spaced ($x_i, i = 1, \dots, d$) and all having N data points $[x_i(n), n = 1, 2, \dots, N]$, with $\mathbf{F}(\mathbf{X}, t) = \mathbf{A}\mathbf{X}$, $\mathbf{A} = (a_{ij})$ being a $d \times d$ matrix, and

$$\mathbf{B} = \begin{bmatrix} b_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & b_d \end{bmatrix}$$

being a $d \times d$ diagonal matrix. We take an example of estimating information flow from x_2 to x_1 in a d -dimensional system ($T_{2 \rightarrow 1|3,4,5,\dots,d}$). First, we need to estimate a_{ij} before estimating $T_{2 \rightarrow 1}$; Liang (2014) indicated that when a_{ij} and b_i

are constant, the maximal likelihood estimator (mle) is the least squares solution of the following N algebraic equations:

$$\sum_{j=1}^d a_{1j} x_j(n) = \dot{x}_1(n), \quad n = 1, \dots, N \quad (4)$$

where $\dot{X}_{i,n} = (X_{i,n+k} - X_{i,n})/(k\Delta t)$ is the Euler forward differencing approximation of dX_i/dt , [where k is an integer and $k \geq 1$; $k = 1$ should be set in most of the time to ensure accuracy, except in some cases of deterministic chaos when the sampling at the highest resolution needs to be chosen ($k = 2$); Δt is the time step size]. Following the procedure in Liang (2014) the least squares solution of $(a_{11}, \dots, a_{1d})$, $(\hat{a}_{11}, \dots, \hat{a}_{1d})$, satisfies the algebraic equation

$$\begin{bmatrix} C_{1,1} & \dots & C_{1,v} \\ \vdots & \vdots & \vdots \\ C_{v,1} & \dots & C_{v,v} \end{bmatrix} \begin{bmatrix} \hat{a}_{11} \\ \vdots \\ \hat{a}_{1v} \end{bmatrix} = \begin{bmatrix} C_{1,d1} \\ \vdots \\ C_{v,d1} \end{bmatrix}, \quad (5)$$

where $c_{ij} = 1/N \sum_{n=1}^N [x_i(n) - \bar{x}_i][x_j(n) - \bar{x}_j]$ and $c_{i,dj} = 1/N \sum_{n=1}^N [x_i(n) - \bar{x}_i][\dot{x}_j(n) - \bar{\dot{x}}_j]$ are the sample covariances.

Here, $\hat{a}_{12} = (1/\det \mathbf{C}) \sum_{j=1}^d \Delta_{2j} C_{j,d1}$, where Δ_{ij} are the cofactors. Based on Eq. (3), the information flow from X_2 to X_1 when considering the influence of $X_3, X_4, X_5, \dots, X_d$ on the d -dimensional system is

$$T_{2 \rightarrow 1|3,4,5,\dots,d} = \frac{1}{\det \mathbf{C}} \sum_{j=1}^d \Delta_{2j} C_{j,d1} \frac{C_{12}}{C_{11}}, \quad (6)$$

where $\det \mathbf{C}$ is the determinant of the covariance matrix $\mathbf{C}$, and $C_{j,d1}$ is the sample covariance between x_j and the derived series $\dot{x}_1$. In this case, when $d = 2$, Eq. (6) can be easily reduced to

$$T_{2 \rightarrow 1} = \frac{C_{11} C_{12} C_{2,d1} - C_{12}^2 C_{1,d1}}{C_{11}^2 C_{22} - C_{11} C_{12}^2}, \quad (7)$$

which is obtained in Liang (2014) and frequently used in applications, such as Stips et al. (2016), Bai et al. (2018) and Tao et al. (2021).

b. Kalman filter

The Kalman filter is a time domain filtering algorithm, which takes the minimum mean square error as the best estimation criterion (Kalman 1960). As a learning algorithm, the equations for the Kalman filter fall into two groups: time update equations and measurement update equations. The time update equations are responsible for obtaining the a priori estimates for the next step by projecting the current state and error covariance forward. The measurement update equations are responsible for obtaining an improved a posteriori estimate modified by combining the latest observation information and the prediction information of the current moment (Kalman 1960). So the time update equations can be thought of as predictor equations and the measurement update equations can be considered corrector equations (Kalman 1960).

Because the Kalman filter can identify immediate changes, this would become advantageous when tracking time-varying causality within a system (Havlicek et al. 2010; Punales 2011). Therefore, we use the Kalman filter to estimate the time-varying information flow based on Eq. (6) to estimate the multivariate time-varying causality formalism based on the Liang–Kleeman information flow. The standard Kalman filter is formulated with a linear stochastic system:

$$\mathbf{x}_k = \mathbf{A}\mathbf{x}_{k-1} + \mathbf{B}\mathbf{u}_k + \mathbf{w}_{k-1}, \quad (8)$$

$$\mathbf{y}_k = \mathbf{H}\mathbf{x}_k + \mathbf{v}_k, \quad (9)$$

where $\mathbf{x}$ and $\mathbf{y}$ are the system state and measurement vectors, and $\mathbf{u}$ is the input vector. The parameter $\mathbf{A}$ is the system states transition matrix, $\mathbf{B}$ is the input matrix, and $\mathbf{H}$ is the measurement matrix. Also, $\mathbf{w}$ and $\mathbf{v}$ are zero mean process and measurement noise, respectively, assumed to be independent of each other, with normal probability distributions, and meets the criteria of the following equations:

$$\begin{aligned} E[\mathbf{w}_k] &= 0, \quad E[\mathbf{w}_k \mathbf{w}_i^T] = \begin{cases} \mathbf{Q}, & i = k \\ 0, & i \neq k \end{cases} \\ E[\mathbf{v}_k] &= 0, \quad E[\mathbf{v}_k \mathbf{v}_i^T] = \begin{cases} \mathbf{R}, & i = k \\ 0, & i \neq k \end{cases} \\ E[\mathbf{w}_k \mathbf{v}_i^T] &= 0, \end{aligned} \quad (10)$$

where E represents the expectation, $\mathbf{Q}$ is the covariance matrix of the process noise $\mathbf{w}_k$, and $\mathbf{R}$ is the covariance matrix of the measurement noise $\mathbf{v}_k$.

When the system satisfies Eqs. (8) and (9), the estimation of each time can be obtained by the Kalman filter recursive formula. The specific procedures are as follows:

- 1) According to the state estimates at time step $k - 1$, project the state at time step k by the following equation:

$$\hat{\mathbf{x}}_k^- = \mathbf{A}\hat{\mathbf{x}}_{k-1} + \mathbf{B}\mathbf{u}_k, \quad (11)$$

where $\hat{\mathbf{x}}_k^-$ are the state estimates at time step k and $\hat{\mathbf{x}}_{k-1}$ are the state estimates at time step $k - 1$.

- 2) Define a priori estimate errors as $\mathbf{e}_k^- = \mathbf{x}_k - \hat{\mathbf{x}}_k^-$, and the a priori estimate error covariance is $\mathbf{P}_k^- = E[\mathbf{e}_k^- \cdot \mathbf{e}_k^{-T}]$, so we can project the error covariance at time step k by the equation:

$$\mathbf{P}_k^- = \mathbf{A}\mathbf{P}_{k-1}\mathbf{A}^T + \mathbf{Q}_k, \quad (12)$$

where $\mathbf{P}_k^-$ are the error covariance estimates at time step k and $\mathbf{P}_{k-1}$ are the error covariance estimates at time step $k - 1$.

- 3) Update the Kalman gain at time step k :

$$\mathbf{K}_k = \mathbf{P}_k^- \mathbf{H}^T (\mathbf{H} \mathbf{P}_k^- \mathbf{H}^T + \mathbf{R})^{-1}. \quad (13)$$

- 4) The below equations are used to update the state estimates with measurement $\mathbf{y}_k$ and the Kalman gain $\mathbf{K}_k$ at time step k :

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{K}_k(\mathbf{y}_k - \mathbf{H}\hat{\mathbf{x}}_k^-). \quad (14)$$

- 5) Define a posteriori estimate errors as $\mathbf{e}_k = \mathbf{x}_k - \hat{\mathbf{x}}_k$, and the a posteriori estimate error covariance is $\mathbf{P}_k = E[\mathbf{e}_k \cdot \mathbf{e}_k^T]$, so the equation of update error covariance at time step k is

$$\mathbf{P}_k = (1 - \mathbf{K}_k \mathbf{H}) \mathbf{P}_k^-. \quad (15)$$

Since the calculation of causality by the LKIF is mainly based on the covariance between samples, it is vital to accurately estimate the covariance between the samples. The multivariate time-varying causality formalism based on the LKIF is realized by computing the sample covariance with the resulting covariance of each time step estimated by Kalman filter. However, it is usually difficult to calculate the process noise covariance $\mathbf{Q}$ and the measurement noise covariance $\mathbf{R}$ (Berg et al. 2014); most current studies assume them as constants. However, $\mathbf{Q}$ and $\mathbf{R}$ may change at each time step in practice, thus, we do not assume they are constants here. Our approach is to first use the exponential weighted moving average (EWMA) and unweighted moving average (UWMA) to estimate $\mathbf{Q}$ and $\mathbf{R}$ offline, and then apply the obtained results to the iteration process of Kalman filter, as shown in Eqs. (12) and (13).

In addition, according to Eqs. (11)–(15), it can be seen that the standard Kalman filter essentially uses the Kalman gain $\mathbf{K}_k$ to weighted average estimates and measurements of $\mathbf{y}_k$ to improve the accuracy of the measurement, which determines the weighted component of the measurement when estimating the state estimates. Therefore, an accurate calculation of $\mathbf{K}_k$ is fundamental to improve the accuracy of the calculation. However, correction of the accumulated computational round-off errors from the state variance-covariance computation was done. This is because no feedback existed in the gain loop, creating numerical imprecision within the Kalman filtering process. To remedy this study, we employed the Bierman–Thornton algorithm (Bierman and Thornton 1977), which uses modified Cholesky factors of the state variance–covariance matrix. The detailed description of this algorithm can be found in Hagan et al. (2019). This study uses the square root Kalman filter based on the Bierman–Thornton algorithm to realize the multivariate time-varying LKIF causality formalism.

c. The multivariate time-varying causality formalism based on the Liang–Kleeman information flow

The algorithm begins with the computation of the process and measurement of noise covariances $\mathbf{Q}$ and $\mathbf{R}$ respectively. From there, the filter starts by calculating the Kalman gain $\mathbf{K}$ through the updated covariance matrix $\mathbf{P}$ at each time step. This step is used to compute the causality. The Bierman and Thornton algorithms are used at the measurement and temporal updates to make the calculation more stable. In this way, Eq. (6) can be rewritten as

$$T_{i2 \rightarrow 1|3,4,5,\dots,d} = \frac{1}{\det \mathbf{P}} \sum_{j=1}^d \Delta_{2j} P_{j,d1} \frac{P_{12}}{P_{11}}, \quad (16)$$

where $\mathbf{P}$ is the resulting covariance matrix at each time step of the iteration, Δ_{ij} are the cofactors of the matrix $\mathbf{P} = (P_{ij})$, and

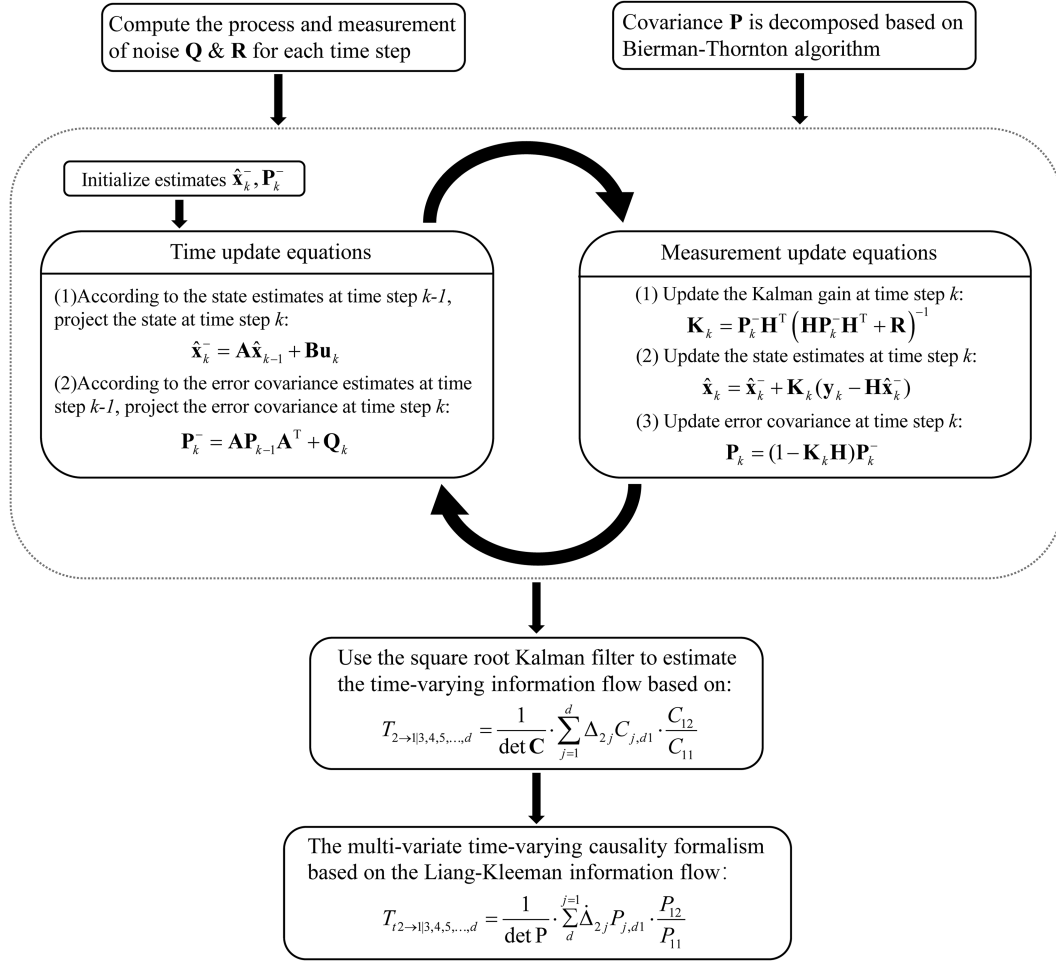


FIG. 1. The framework of MtvLK.

$P_{j,d1}$ are the resulting covariance between X_j and the derived series $\dot{X}_1$.

Since the EMWA and UWMA require choice of a look-back window length to compute the moving average before estimating $\mathbf{Q}$ and $\mathbf{R}$ offline, an information length of the window length is lost to do the forward update of the filter. We propose an effective solution to the problem; the details are given in [appendix B](#). The formalism is henceforth referred to as MtvLK (multivariate time-varying Liang-Kleeman information flow) and will be used throughout the rest of this paper; the framework of MtvLK is shown in [Fig. 1](#). In addition, it is particularly important to note that when there are only two variables in the system, MtvLK will be the same as the framework proposed by [Hagan et al. \(2019\)](#).

3. Simulation studies

To verify the reliability of MtvLK, we set up three idealized experiments with known causalities to examine how well the information flow of the MtvLK captures the preset causal structures. The LKIF in this study is measured in nats per unit time.

a. Synthetic model 1

In land-atmosphere interactions, causal relationships between variables often vary over time instead of remaining constant, as exemplified by the impact of soil moisture on temperature in the Chinese region ([Hagan et al. 2019](#)). The findings of [Hagan et al. \(2019\)](#) indicate that in humid areas of China, the influence of soil moisture on temperature is mostly nonsignificant throughout the year, except for a significant statistical signal during the spring season. Moreover, the information flow between soil moisture and temperature in spring exhibits a trend of initial increase followed by a subsequent decrease. Drawing upon the varying causal relationships observed in this case, we have designed the following synthetic model that incorporates multiple variables:

$$\begin{cases} x_1(t) = 0.35x_1(t-1) + c_{2 \rightarrow 1}(t)x_2(t-1) + \varepsilon_1(t) \\ x_2(t) = 0.35x_2(t-1) + c_{1 \rightarrow 2}(t)x_1(t-1) + \varepsilon_2(t) \\ x_3(t) = 0.35x_3(t-1) + c_{2 \rightarrow 3}(t)x_2(t-1) + \varepsilon_3(t) \end{cases} \quad (17)$$

The first synthetic model is a ternary autoregressive process of $X = [x_1, x_2, x_3]$, as described in [Eq. \(17\)](#). The time series

x_1 , x_2 , and x_3 have the same length of 2000 time points with added white noise of zero means and unit variance $\varepsilon_1(t)$, $\varepsilon_2(t)$, and $\varepsilon_3(t)$. Note that $c_{2 \rightarrow 1}$ and $c_{1 \rightarrow 2}$ represent the influence of x_2 on x_1 and that of x_1 on x_2 , while $c_{2 \rightarrow 3}$ represents the influence of x_2 on x_3 . For $0 < t \leq 500$ and $1500 < t \leq 2000$, all the causalities are set to be 0. For $500 < t \leq 1500$, all the causalities are gradually increased from 0 to 0.5, with an increment of 0.001 per time step when $500 < t \leq 1000$, and all the causalities are gradually decreased from 0.5 to 0, with a decrement of 0.001 per time step when $1000 < t \leq 1500$. Thus, in this system, when $0 < t \leq 500$ and $1500 < t \leq 2000$, no causality exists among x_1 , x_2 , and x_3 . When $500 < t \leq 1500$, causalities exist between x_1 and x_2 and from x_2 to x_3 . The full causal structure of Eq. (17) is shown in Figs. 2a and 2b. In Figs. 2c–j, the blue lines show the realization averages of the absolute value of time-varying information flows, while the red lines are the significant values realizations at a 1% significance level. We infer causality when the absolute information flow values are greater than the significant values. All the results are the average of the absolute values of 1000 realizations at a 1% significance level.

Figures 2c–j are the ensemble means of the time-varying information flow results (in nats per unit time). We added two dashed lines in the figure at the causal relationship turning points of $t = 500$ and $t = 1500$ for a more intuitive display. Additionally, it is worth noting that to more clearly depict the causal relationships each subplot focuses on, we have added a small schematic at the top of each one. In these schematics, the blue variables indicate a variable whose role is being considered within the system, while the gray variables signify those whose roles are not currently being considered. Yellow solid arrows represent predefined causal relationships that have not been tested yet, whereas red solid arrows denote those predefined causal relationships currently under testing. Gray solid arrows pointing toward gray variables indicate causal relationships that are not being considered in the system at present. Meanwhile, red dashed arrows signify causal relationships that are currently being tested but were not predefined.

The results in Figs. 2c–j are consistent with the predefined causalities in Eq. (17), showing that x_1 influences x_2 (Figs. 2c,d) and x_2 influences both x_1 (Figs. 2e,f) and x_3 (Figs. 2g,h) where $500 < t \leq 1500$. Figures 2c and 2d are the absolute value of the time-varying information flows from x_1 to x_2 ($|T_{x_1 \rightarrow x_2}|$) and x_1 to x_2 influenced by x_3 ($|T_{x_1 \rightarrow x_2|x_3}|$), respectively. The results in the two figures show that x_1 influences x_2 for $500 < t \leq 1500$, the causal intensity exhibits an initial increase followed by a subsequent decrease, no causality exists between x_1 and x_2 when $0 < t \leq 500$ and $1500 < t \leq 2000$, and an insignificant influence of x_3 in the multivariate case (Fig. 2f). It is important to emphasize that the absolute value of information flow employed in this study solely serves as a measure of the causal strength between variables. It does not carry any connotations related to positivity or negativity, nor does it imply any other meaning beyond quantifying the strength of the causal relationship. Furthermore, it is crucial to acknowledge that the MtvLK necessitates the selection of a suitable lookback window length for computing the moving average, which is subsequently used in

offline estimation of $\mathbf{Q}$ and $\mathbf{R}$. The choice of this window length can be tailored to meet the specific requirements of the experiment. In this synthetic test, we opted for a window length of 300 to facilitate our investigation. Consequently, the depicted results in Fig. 2 indicate that when causalities change, a relatively brief time step is required for the transition to occur, with the maximum value being attained at approximately $t = 1200$ instead of $t = 1000$. Figures 2e and 2f are the absolute value of the time-varying information flows from x_2 to x_1 ($|T_{x_2 \rightarrow x_1}|$) and x_2 to x_1 influenced by x_3 ($|T_{x_2 \rightarrow x_1|x_3}|$), respectively. The results in the two figures also show that x_2 influences x_1 for $500 < t \leq 1500$, and an insignificant influence of x_3 in the multivariate case (Fig. 2f). Figures 2g and 2h show the absolute value of the time-varying information flow from x_2 to x_3 ($|T_{x_2 \rightarrow x_3}|$) and that of x_2 to x_3 influenced by x_1 ($|T_{x_2 \rightarrow x_3|x_1}|$). Again, the results in the two figures are identical and point to the negligible influence of x_1 in the multivariate case (Fig. 2h). Finally, Figs. 2i and 2j are the absolute value of the time-varying information flow from x_1 to x_3 ($|T_{x_1 \rightarrow x_3}|$) and that of x_1 to x_3 influenced by x_2 ($|T_{x_1 \rightarrow x_3|x_2}|$). Both results correctly show that there are no causalities for $0 < t \leq 500$ and $1500 < t \leq 2000$. However, $|T_{x_1 \rightarrow x_3|x_2}|$ also shows no causalities for $500 < t \leq 1500$ (Fig. 2j) while $|T_{x_1 \rightarrow x_3}|$ shows significant causalities. These results might suggest that $|T_{x_1 \rightarrow x_3}|$ interprets the continuous information flow from x_1 to x_2 , and then x_2 to x_3 as a potential causal link ($|T_{x_1 \rightarrow x_3}|$) when the confounding variable x_2 is not considered (Fig. 2i). All of the findings above exhibit a high degree of consistency with the predetermined causal structure established in this synthetic model. Moreover, the temporal depiction of causation strength effectively captures the dynamics of causal relationships over time.

Additionally, we conducted the same test on all links that did not preset causality, with the results as shown in Fig. C1 (appendix C). Examining the figure, we can clearly see that $|T_{x_3 \rightarrow x_1|x_2}|$ and $|T_{x_3 \rightarrow x_2|x_1}|$ shows no causalities for $0 < t \leq 500$ and $1500 < t \leq 2000$. In addition, although $|T_{x_3 \rightarrow x_1|x_2}|$ and $|T_{x_3 \rightarrow x_2|x_1}|$ exhibit a tendency to initially rise and then decline for $500 < t \leq 1500$, both $|T_{x_3 \rightarrow x_1|x_2}|$ and $|T_{x_3 \rightarrow x_2|x_1}|$ still shows no causalities, which means $|T_{x_3 \rightarrow x_1|x_2}|$ and $|T_{x_3 \rightarrow x_2|x_1}|$ have no causalities for $0 < t \leq 2000$, which is aligning well with our initial expectations.

b. Synthetic model 2

In complex multivariable systems similar to the land-atmosphere interactions, besides the case of homologous influences as in synthetic model 1, there are also many cases where several variables influence one variable simultaneously. Therefore, we generated the following tertiary autoregressive process based on synthetic model 1:

$$\begin{cases} x_1(t) = 0.35x_1(t-1) + c_{2 \rightarrow 1}(t)x_2(t-1) + \varepsilon_1(t) \\ x_2(t) = 0.35x_2(t-1) + \varepsilon_2(t) \\ x_3(t) = 0.35x_3(t-1) + c_{1 \rightarrow 3}(t)x_1(t-1) + c_{2 \rightarrow 3}(t)x_2(t-1) + \varepsilon_3(t). \end{cases} \quad (18)$$

The process of $X = [x_1, x_2, x_3]$, as described in Eq. (18). Here, for $0 < t \leq 500$ and $1500 < t \leq 2000$, $c_{2 \rightarrow 1}$, $c_{2 \rightarrow 3}$ and $c_{1 \rightarrow 3}$ are

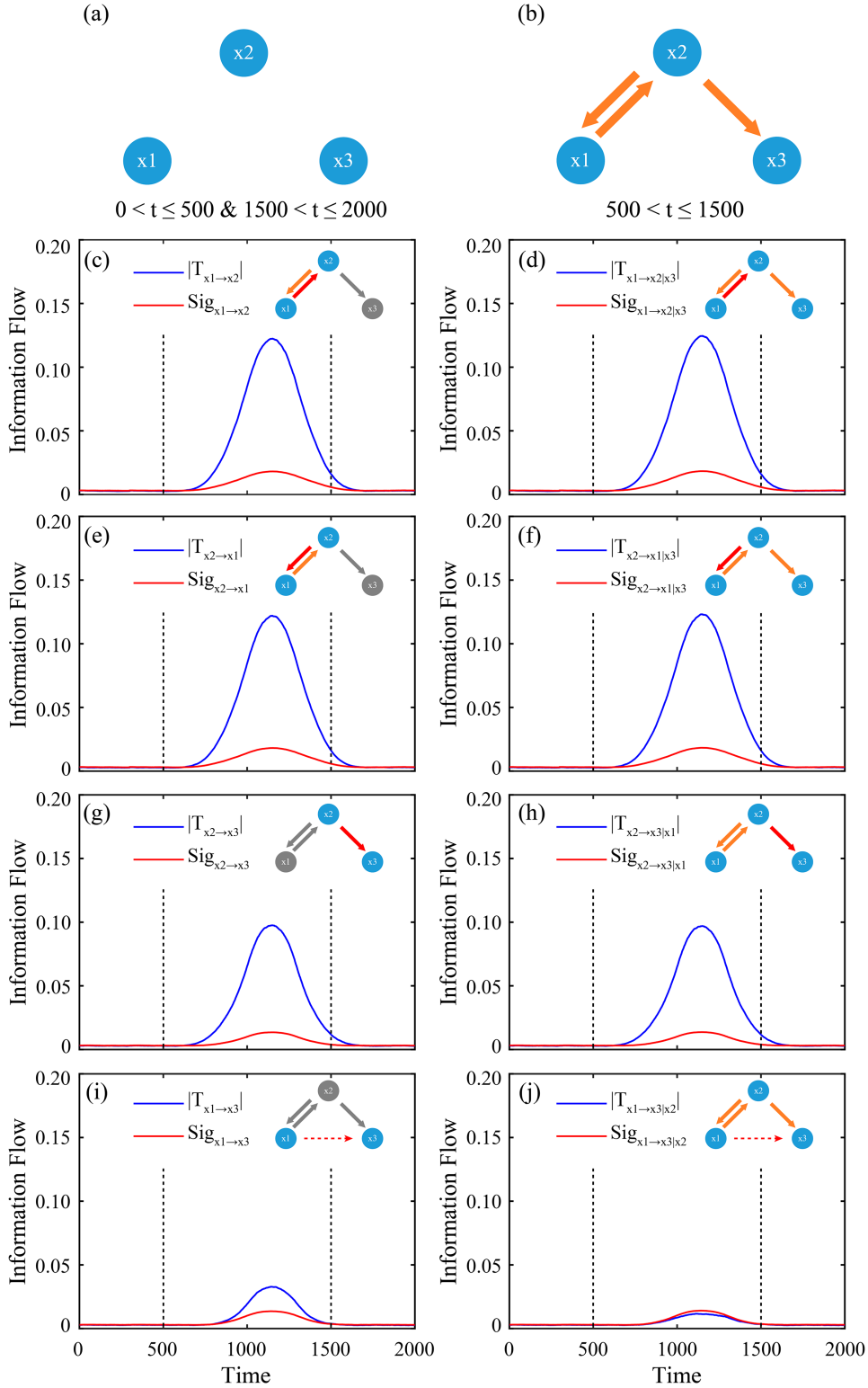


FIG. 2. (a) Schematic diagram of the conceptual model as described in Eq. (17) when $0 < t \leq 500$ and $1500 < t \leq 2000$. (b) As in (a), but for interval $500 < t \leq 1500$. Absolute values of the time-varying information flows among x_1 , x_2 , and x_3 (in nats per unit time): (c) $|T_{x_1 \rightarrow x_2}|$ and $\text{Sig}_{x_1 \rightarrow x_2}$; (d) $|T_{x_1 \rightarrow x_2|x_3}|$ and $\text{Sig}_{x_1 \rightarrow x_2|x_3}$; (e) $|T_{x_2 \rightarrow x_1}|$ and $\text{Sig}_{x_2 \rightarrow x_1}$; (f) $|T_{x_2 \rightarrow x_1|x_3}|$ and $\text{Sig}_{x_2 \rightarrow x_1|x_3}$; (g) $|T_{x_2 \rightarrow x_3}|$ and $\text{Sig}_{x_2 \rightarrow x_3}$; (h) $|T_{x_2 \rightarrow x_3|x_1}|$ and $\text{Sig}_{x_2 \rightarrow x_3|x_1}$; (i) $|T_{x_1 \rightarrow x_3}|$ and $\text{Sig}_{x_1 \rightarrow x_3}$; (j) $|T_{x_1 \rightarrow x_3|x_2}|$ and $\text{Sig}_{x_1 \rightarrow x_3|x_2}$. Dashed lines represent the causal relationship turning points. All the results are the averages of 1000 realizations, and all the significance tests are performed at a 1% significance level.

set to be 0; for $500 < t \leq 1500$, all the causalities are gradually increased from 0 to 0.5, with an increment of 0.001 per time step when $500 < t \leq 1000$, and all the causalities are gradually decreased from 0.5 to 0, with an decrement of 0.001 per time step when $1000 < t \leq 1500$. Thus, in this system, when $0 < t \leq 500$ and $1500 < t \leq 2000$, there are no direct causalities among the x_1 , x_2 , and x_3 ; and when $500 < t \leq 1500$, x_3 is directly influenced not only by x_2 but also by x_1 . The causality among the time series x_1 , x_2 , and x_3 is shown in Figs. 3a and 3b. Similar to synthetic model 1, the window length of 300 was also chosen in this synthetic model to facilitate our investigation. Figures 3c–h are the ensemble means of the absolute value of the time-varying information flows among the x_1 , x_2 , and x_3 (in nats per unit time). All the results are obtained at a 1% significance level. Figures 3c and 3d shows the absolute value of the time-varying information flow from x_2 to x_1 ($|T_{x_2 \rightarrow x_1}|$) and that from x_2 to x_1 influenced by x_3 ($|T_{x_2 \rightarrow x_1|x_3}|$). Both approaches show that x_3 does not affect the causality from x_2 to x_1 since there is no preset feedback from x_3 to x_1 or x_2 . Figures 3e and 3f show the absolute value of the time-varying information flow from x_2 to x_3 ($|T_{x_2 \rightarrow x_3}|$) and that from x_2 to x_3 influenced by x_1 ($|T_{x_2 \rightarrow x_3|x_1}|$), for $500 < t \leq 1500$, where we find that $|T_{x_2 \rightarrow x_3|x_1}|$ is lower than $|T_{x_2 \rightarrow x_3}|$, due to the feedback from x_1 (Barrett and Barnett 2013). Finally, Figs. 3g and 3h show the absolute value of the time-varying informa-

tion flow from x_1 to x_3 ($|T_{x_1 \rightarrow x_3}|$) and that from x_1 to x_3 influenced by x_2 ($|T_{x_1 \rightarrow x_3|x_2}|$). Both cases reproduce the expected causalities as preset in Eq. (18). These results demonstrate the reliability of the MtvLK to unravel time-varying causalities even in complex systems.

Similar to synthetic model 1, we tested all links that did not preset causality in this synthetic model, as shown in Fig. C2 (appendix C). Figure C2 also shows that $|T_{x_1 \rightarrow x_2|x_3}|$, $|T_{x_3 \rightarrow x_1|x_2}|$, and $|T_{x_3 \rightarrow x_2|x_1}|$ all show no causalities for $0 < t \leq 500$ and $1500 < t \leq 2000$, and $|T_{x_1 \rightarrow x_2|x_3}|$, $|T_{x_3 \rightarrow x_1|x_2}|$, and $|T_{x_3 \rightarrow x_2|x_1}|$ all exhibit a tendency to initially rise and then decline for $500 < t \leq 1500$, but there are both no causalities for $500 < t \leq 1500$, which aligns well with our initial expectations too.

c. Synthetic model 3

Synthetic model 1 and synthetic model 2 primarily focus on examining the interactions among three variables, and their study has successfully demonstrated the reliability of the MtvLK method in complex systems. Nevertheless, when investigating land–atmosphere interactions, sometimes it becomes imperative to account for the influence resulting from more than three variables. To address this requirement comprehensively, we generated the following five-level autoregressive process:

$$\begin{cases} x_1(t) = 0.35x_1(t-1) + \varepsilon_1(t) \\ x_2(t) = 0.35x_2(t-1) + c_{1 \rightarrow 2}(t)x_1(t-1) + \varepsilon_2(t) \\ x_3(t) = 0.35x_3(t-1) + c_{2 \rightarrow 3}(t)x_2(t-1) + \varepsilon_3(t) \\ x_4(t) = 0.35x_4(t-1) + c_{2 \rightarrow 4}(t)x_2(t-1) + c_{3 \rightarrow 4}(t)x_3(t-1) + \varepsilon_4(t) \\ x_5(t) = 0.35x_5(t-1) + c_{1 \rightarrow 5}(t)x_1(t-1) + c_{4 \rightarrow 5}(t)x_4(t-1) + \varepsilon_5(t) \end{cases} \quad (19)$$

The process of $X = [x_1, x_2, x_3, x_4, x_5]$, as described in Eq. (19). For $0 < t \leq 500$, $c_{2 \rightarrow 1}$, $c_{2 \rightarrow 3}$, and $c_{1 \rightarrow 3}$ are set to be 0; for $500 < t \leq 1500$, all the causalities are gradually increased from 0 to 0.5, with an increment of 0.0005 per time step; for $1500 < t \leq 2000$, $c_{2 \rightarrow 1}$, $c_{2 \rightarrow 3}$, and $c_{1 \rightarrow 3}$ are set to be 0.5. Thus, in this system, when $0 < t \leq 500$, there are no direct causalities among the time series x_1 , x_2 , x_3 , x_4 , and x_5 as shown in Fig. 4a; when $500 < t \leq 2000$, causalities exist from x_1 to x_2 , x_2 to x_5 , x_2 to x_3 , x_2 to x_4 , x_3 to x_4 , and x_4 to x_5 , where x_4 is directly influenced not only by x_2 but also by x_3 , as shown in Fig. 4b, all the causalities among the variables exhibit an initial gradual increase, followed by a period of stability where they remain unchanged. Figures 4c–h are the ensemble means of the absolute value of the time-varying information flows among the x_1 , x_2 , x_3 , x_4 , and x_5 (in nats per unit time). All the results are obtained at a 1% significance level. For this synthetic model, we have selected a lookback window length of 200 for the experimental. Figures 4c–h show the absolute value of the time-varying information flow from x_1 to x_2 influenced by x_3 , x_4 , and x_5 ($|T_{x_1 \rightarrow x_2|x_3, x_4, x_5}|$), from x_1 to x_5 influenced by x_2 , x_3 , and x_4

($|T_{x_1 \rightarrow x_5|x_2, x_3, x_4}|$), from x_2 to x_3 influenced by x_1 , x_4 , and x_5 ($|T_{x_2 \rightarrow x_3|x_1, x_4, x_5}|$), from x_2 to x_4 influenced by x_1 , x_3 , and x_5 ($|T_{x_2 \rightarrow x_4|x_1, x_3, x_5}|$), from x_3 to x_4 influenced by x_1 , x_2 , and x_5 ($|T_{x_3 \rightarrow x_4|x_1, x_2, x_5}|$), and from x_4 to x_5 influenced by x_1 , x_2 , and x_3 ($|T_{x_4 \rightarrow x_5|x_1, x_2, x_3}|$), respectively. All test results effectively replicate the predetermined time-varying causal relationships within the synthetic model, and the time-varying causal strength among variables consistently aligns with the preset conditions. Consequently, this experiment serves as compelling evidence to affirm the efficacy of MtvLK in complex systems involving more than three variables. It demonstrates the method's ability to accurately identify the time-varying causality between variables within multivariate complex systems.

Similar to synthetic model 1 and synthetic model 2, we tested all links that did not preset causality in this synthetic model; the results are shown in Fig. C3 (appendix C). Figures C3c–n demonstrate that no causal relationships were detected for all the links without predefined causality for $0 < t \leq 2000$. These results further prove the accuracy of MtvLK in multivariate complex systems.

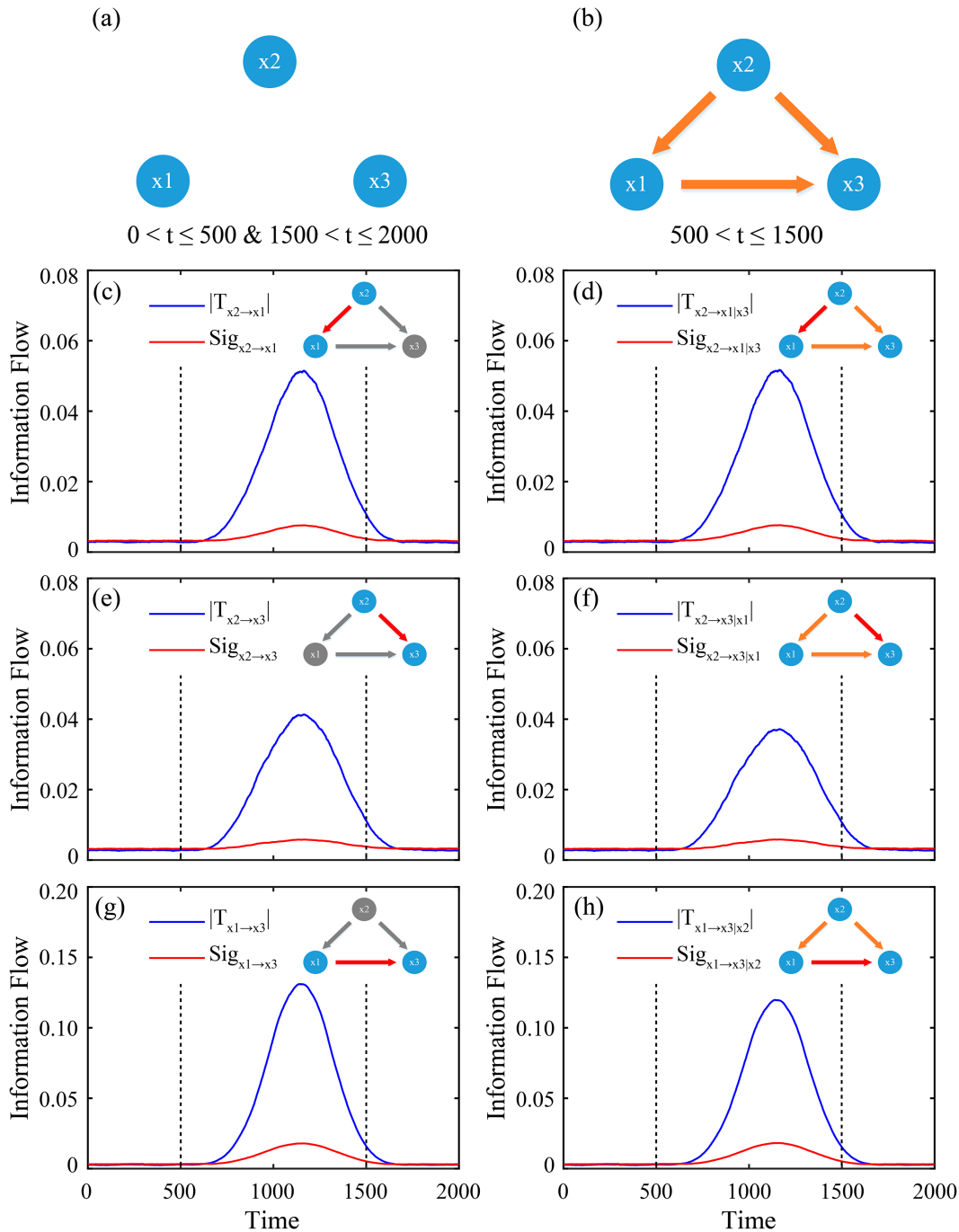


FIG. 3. (a) Schematic diagram of the conceptual model as described in Eq. (18) when $0 < t \leq 500$ and $1500 < t \leq 2000$. (b) As in (a), but for the interval $500 < t \leq 1500$. Absolute values of the time-varying information flows among x_1 , x_2 , and x_3 (in nats per unit time): (c) $|T_{x_2 \rightarrow x_1}|$ and $\text{Sig}_{x_2 \rightarrow x_1}$; (d) $|T_{x_2 \rightarrow x_1|x_3}|$ and $\text{Sig}_{x_2 \rightarrow x_1|x_3}$; (e) $|T_{x_2 \rightarrow x_3}|$ and $\text{Sig}_{x_2 \rightarrow x_3}$; (f) $|T_{x_2 \rightarrow x_3|x_1}|$ and $\text{Sig}_{x_2 \rightarrow x_3|x_1}$; (g) $|T_{x_1 \rightarrow x_3}|$ and $\text{Sig}_{x_1 \rightarrow x_3}$; (h) $|T_{x_1 \rightarrow x_3|x_2}|$ and $\text{Sig}_{x_1 \rightarrow x_3|x_2}$. Dashed lines represent the causal relationship turning points. All the results are the averages of 1000 realizations, and all the significance tests are performed at a 1% significance level.

4. Real applications in land–atmosphere interactions

Vegetation is a life support system for the survival and development of human society. It is a natural link between soil, atmosphere, and water, and has complex interactions with the

climate system (Cui et al. 2020; Feng et al. 2017). Vegetation is part of the terrestrial ecosystem, more specifically, the terrestrial biosphere (N. Chen et al. 2021; Liu et al. 2022), and influences the global carbon and water cycle (Liu et al. 2014; Pravalıe 2018) by regulating carbon balance, alleviating the

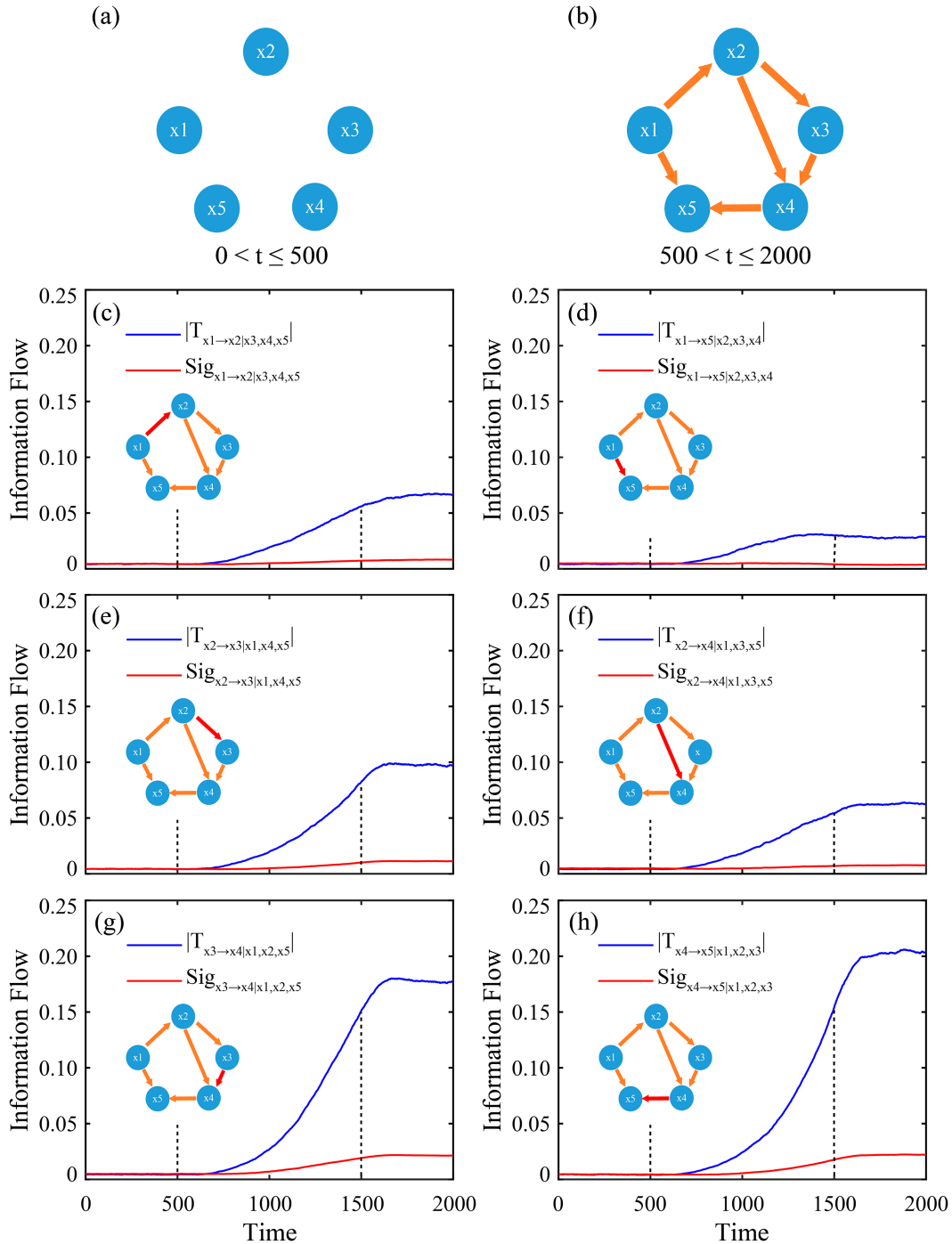


FIG. 4. (a) Schematic diagram of the conceptual model as described in Eq. (19) when $0 < t \leq 500$. (b) As in (a), but for the interval $500 < t \leq 2000$. Absolute values of the time-varying information flows among x_1 , x_2 , x_3 , x_4 , and x_5 (in nats per unit time): (c) $|T_{x_1 \rightarrow x_2|x_3, x_4, x_5}|$ and $\text{Sig}_{x_1 \rightarrow x_2|x_3, x_4, x_5}$; (d) $|T_{x_1 \rightarrow x_5|x_2, x_3, x_4}|$ and $\text{Sig}_{x_1 \rightarrow x_5|x_2, x_3, x_4}$; (e) $|T_{x_2 \rightarrow x_3|x_1, x_4, x_5}|$ and $\text{Sig}_{x_2 \rightarrow x_3|x_1, x_4, x_5}$; (f) $|T_{x_2 \rightarrow x_4|x_1, x_3, x_5}|$ and $\text{Sig}_{x_2 \rightarrow x_4|x_1, x_3, x_5}$; (g) $|T_{x_3 \rightarrow x_4|x_1, x_2, x_5}|$ and $\text{Sig}_{x_3 \rightarrow x_4|x_1, x_2, x_5}$; (h) $|T_{x_4 \rightarrow x_5|x_1, x_2, x_3}|$ and $\text{Sig}_{x_4 \rightarrow x_5|x_1, x_2, x_3}$. Dashed lines represent the causal relationship turning points. All the results are the averages of 1000 realizations, and all the significance tests are performed at a 1% significance level.

increase of greenhouse gas concentration, and reducing soil erosion (Le Quéré et al. 2015; Liu et al. 2017; Melnikova and Sasai 2020; Wang et al. 2021; You et al. 2020). Gross primary production (GPP) is a good proxy for reflecting vegetation growth (Liu et al. 2014; Wang et al. 2021). Many studies have shown that high air temperature and low water availability can significantly inhibit the growth of terrestrial vegetation, resulting in decreasing GPP (H. Chen et al. 2021; Chen et al. 2020; Liu et al. 2020; Zhang et al. 2019). It has been found that soil moisture (SM) and vapor pressure deficit (VPD) are two important driving factors affecting vegetation growth (Kimm et al. 2020; Novick et al. 2016). Therefore, understanding how SM and VPD impact GPP is essential for understanding global or regional terrestrial carbon cycles and predicting future climate change.

In recent years, although the interactions between SM, VPD, and GPP have been investigated in several studies, there are uncertainties about which mainly influences GPP: SM or VPD. SM affects vegetation growth in many ways. It not only determines the maximum amount of water that can be absorbed by plant roots (Sperry et al. 2017) but also affects stomatal conductance of vegetation canopy (Anderegg et al. 2017; Stocker et al. 2018) and controls carbon allocation during vegetation growth (Palacio et al. 2014). Jung et al. (2017) quantified the impact of temperature, solar radiation, and SM on GPP based on statistical regression methods and found that SM was the dominant factor of GPP interannual variation. Liu et al. (2020) pointed out that the reduction in SM caused by global warming accounts for about 70% of the GPP variability in the world. On the other hand, VPD also significantly impacts vegetation: high VPD may increase evapotranspiration, which induces vegetation to close its stomata to reduce leaf water loss and inhibit plant photosynthesis (Li et al. 2022; Oren et al. 1999). Along these lines, it has been proposed that VPD plays a dominant role in influencing GPP (Besnard et al. 2021; Madani et al. 2020; Wang et al. 2014). Yuan et al. (2019) found that global vegetation had a decreasing trend after the 1990s, which was mainly due to the significant increases in VPD, inhibiting vegetation photosynthesis. Madani et al. (2020) pointed out that VPD was replacing temperature as the dominant factor controlling global GPP in recent decades because the increase of VPD would lead to increased vegetation impedance and thus hinder vegetation photosynthesis.

Besnard et al. (2021) observed that about 56% of the global vegetation carbon storage is controlled by temperature and VPD, while the impact of SM accounts for only 29%. Humphrey et al. (2021) also pointed out that the root cause of the above controversy is that the influencing mechanism of climate change on vegetation involves complex land–atmosphere interaction. Although SM dominates the interannual variability of GPP, it mainly affects VPD through atmospheric feedback induced by decreasing soil moisture, and then inhibits vegetation photosynthesis through VPD.

However, previous studies often ignore the influence of land–atmosphere interactions caused by changes in SM. The above review of the debate indicates that the causal relationships among SM, VPD, and GPP are complex, including

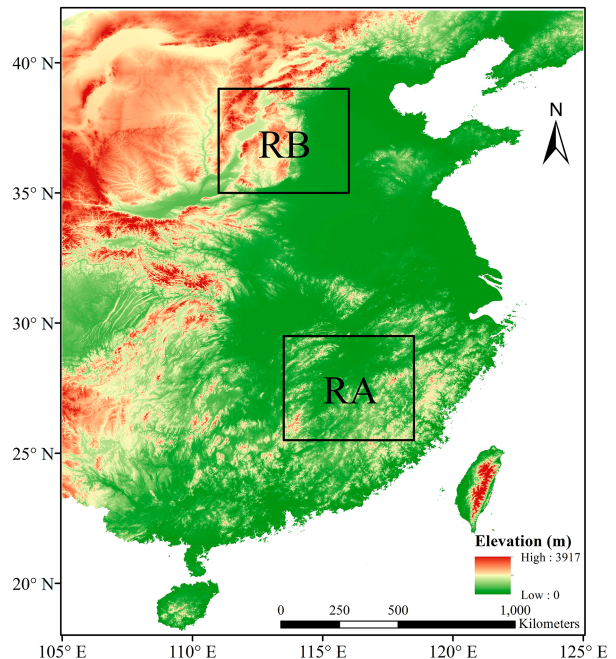


FIG. 5. The study area. RA represents a wet area and RB a partly wet–partly dry climate region.

various direct or indirect causalities. If we analyze the causal relationship between only two variables, the influence of the other variables on the system cannot be considered, and the indirect effects cannot be analyzed, leading to incorrect conclusions. For such a complex system, we must consider the influence of multiple variables on the system simultaneously to analyze the causal relationship accurately. Therefore, in this section, we apply the MtvLK to study the mechanism of SM and VPD in impacting GPP in China, to understand the direct or indirect causal effects of SM and VPD on GPP in China.

The territory of China features a complex topography and a diverse climate. Hagan et al. (2019) selected two typical regions, the Huanan region and Huabei region, for their study, as the climate conditions of these two regions are representatives of strong land–atmosphere interactions. In this study, these two regions are also selected, as shown as region A (RA) and region B (RB) in Fig. 5. The latitudes and longitudes of RA range from 25.5° to 29.5°N, and from 113.5° to 118.5°E, respectively. The climate in this region is humid, with significant precipitation, representing the humid areas of China. RB, located between 35°–39°N and 111°–116°E, on the other hand, is a transitional climate zone between dry and wet climates, representing the semihumid and semiarid regions in China. The GPP we will be using is the ensemble mean of FLUXCOM GPP, which has been widely used to represent the carbon uptake by land ecosystems (Jung et al. 2017). The data are gridded at a horizontal resolution of 0.5°, spanning from January 1979 to December 2018, which the time resolution is the monthly time step. They can be downloaded from <http://www.fluxcom.org/CF-Download/>. Since FLUXCOM GPP is calculated based on ERA5 meteorological driving data, we

also use the SM and VPD from ERA5-Land monthly data to ensure data consistency (available at <https://cds.climate.copernicus.eu>). The weighted average of SM between 0 and 100 cm is used here, with weights determined by the thickness of each layer (0–7, 7–28, 28–100 cm). VPD is estimated with the 2-m air temperature and 2-m dewpoint temperature (Allen et al. 1998). Since the time span of the GPP data differs from that for SM and VPD, we select the overlapping period from January 1981 to December 2018. In addition, we calculate the monthly anomalies for all variables from January 1981 to December 2018, by subtracting the climatology from the original data.

The estimated results may not be accurate enough at the beginning because the Kalman filter is a learning algorithm. Hence, we select the 30-yr (360 months) results from January 1989 to December 2018, with the initial spinup period removed. In addition, we have selected a lookback window length of 90 days for this experiment. Figures 6 and 7 show the absolute value of the information flow rates between SM, VPD, and GPP in South China (RA) and North China (RB) as functions of time from January 1989 to December 2018. The significance tests of each pixel in the two regions are computed at a 5% significance level. The solid blue line shows the absolute value of the average information flow of all the individual pixels in the selected regions, and the solid red line is the mean significance test values for all the pixels. Causality is implied if the solid blue line is higher than the solid red line. In RA, Figs. 6a and 6b represent the time-varying information flow from VPD to SM ($|T_{\text{VPD} \rightarrow \text{SM}}|$) and that from VPD to SM influenced by GPP ($|T_{\text{VPD} \rightarrow \text{SM}|\text{GPP}}|$), respectively. It can be seen that both are almost significant during the whole period and the results are almost the same. These results suggest that causality exists from VPD to SM with potentially very little GPP influence in this region since both results are identical. These results are identical to the toy model results in Figs. 2c and 2d where we also found little influence of the third variable. Additionally, we find a decreasing trend in the strength of the causality from VPD to SM, potentially due to global warming. Figures 6c and 6d show the time-varying information flow from VPD to GPP ($|T_{\text{VPD} \rightarrow \text{GPP}}|$) and that from VPD to GPP influenced by SM ($|T_{\text{VPD} \rightarrow \text{GPP}|\text{SM}}|$), respectively. Both are also significant most of the time, except between 1998 and 2004. Figures 6e and 6f represent the time-varying information flow from SM to GPP ($|T_{\text{SM} \rightarrow \text{GPP}}|$) and that from SM to GPP influenced by VPD ($|T_{\text{SM} \rightarrow \text{GPP}|\text{VPD}}|$), respectively. Clearly, $|T_{\text{SM} \rightarrow \text{GPP}}|$ is barely significant before 2004; but after that, the causality increases significantly from 2004 to 2015. However, $|T_{\text{SM} \rightarrow \text{GPP}|\text{VPD}}|$ is not significant for most of the time, though it is significant in 2004 and around the peak causal period of $|T_{\text{SM} \rightarrow \text{GPP}}|$ (2010–12). The difference in the results might suggest a potential causal structure from SM to GPP as we have identified in Fig. 2b. In this region, the causality from SM to GPP might comprise the influence of VPD on SM (as that from x_2 to x_1 in Fig. 2b) and the influence of VPD on GPP (as that from x_2 to x_3 in Fig. 2b). As a result, when we only consider $|T_{\text{SM} \rightarrow \text{GPP}}|$, we see a similar causality as we find in Fig. 2i, while if the influence of VPD is considered, such as $|T_{\text{SM} \rightarrow \text{GPP}|\text{VPD}}|$, the causal structure undergoes

tremendous changes. Also, as we have found in Fig. 6f, this causal structure changes with time, where VPD might play a mediating role at some time instants (2004, 2010–12), but no role is in play in the years when $|T_{\text{SM} \rightarrow \text{GPP}|\text{VPD}}|$ is insignificant. Nonetheless, it is worth mentioning that, based solely on the revealed causality, one still can hardly tell the underlying mechanism; what one gets are merely the probable outcomes resulting from uncertain causes (Liang 2014). Finally, the impact of SM on VPD is also found to be insignificant in Figs. 6g and 6h.

Dry anomalies could explain the peaking information flow we have found around 2010 through 2014, when anomalous SM conditions would come into play due to increased atmospheric demands, which could, in turn, potentially lead to drought conditions (Figs. 6e,f). Furthermore, during the same period where the SM $\rightarrow$ GPP causality peaks, we also find significant lows in the VPD $\rightarrow$ SM causality (Figs. 6a,b). As noted by Hagan et al. (2019) and Madani et al. (2020), this is characteristic of the region RA.

Next, we analyze the mechanism of how SM and VPD affect GPP in RB. Figures 7a and 7b give the time-varying information flow from VPD to SM ($|T_{\text{VPD} \rightarrow \text{SM}}|$) and that from VPD to SM influenced by GPP ($|T_{\text{VPD} \rightarrow \text{SM}|\text{GPP}}|$), respectively. Both results show significant causalities between 1989 and 2018 over the entire period with and without the influence of VPD, just like that of Figs. 6a and 6b. Similarly, we find a decreasing linear trend in these causalities as shown in Figs. 6a and 6b. Shown in Figs. 7c and 7d are the time-varying information flow from VPD to GPP ($|T_{\text{VPD} \rightarrow \text{GPP}}|$) and that from VPD to GPP influenced by SM ($|T_{\text{VPD} \rightarrow \text{GPP}|\text{SM}}|$). Even though both are significant during the whole period, $|T_{\text{VPD} \rightarrow \text{GPP}}|$ is larger in magnitude than $|T_{\text{VPD} \rightarrow \text{GPP}|\text{SM}}|$, with the causal structure similar to that found in synthetic model 2. This hints at the existence of potential feedback from SM. Here, we find a long-term decreasing linear trend in the information flows. In Figs. 7e and 7f, we see that the information flows from SM to GPP ($|T_{\text{SM} \rightarrow \text{GPP}}|$) and from SM to GPP influenced by VPD ($|T_{\text{SM} \rightarrow \text{GPP}|\text{VPD}}|$) as functions of time. Clearly both are significant throughout, with $|T_{\text{SM} \rightarrow \text{GPP}}|$ larger in magnitude than $|T_{\text{SM} \rightarrow \text{GPP}|\text{VPD}}|$. Just as in Figs. 7c and 7d, the difference in the magnitude suggests a feedback from VPD. Previous studies have shown that both SM and VPD are influencing factors of GPP in semiarid areas although the sensitivity of GPP to VPD is considered higher than the sensitivity of GPP to SM (Humphrey et al. 2021). Based on Figs. 7c–f, we can conclude that SM and VPD both impact GPP over the entire period. In addition, we also find that the information flow from VPD to GPP influenced by SM is higher than the information flow from SM to GPP influenced by VPD. These results agree well with the relevant research conclusions so far. For instance, Novick et al. (2016) highlighted the significant impact of VPD on vegetation ecosystems, surpassing the influence of temperature and soil moisture. Similarly, Madani et al. (2020) suggested that VPD has gradually become the dominant factor controlling global GPP in recent decades, with the main mechanism being the increase in vegetation resistance due to elevated VPD, thereby impeding photosynthesis (Fu et al. 2022).

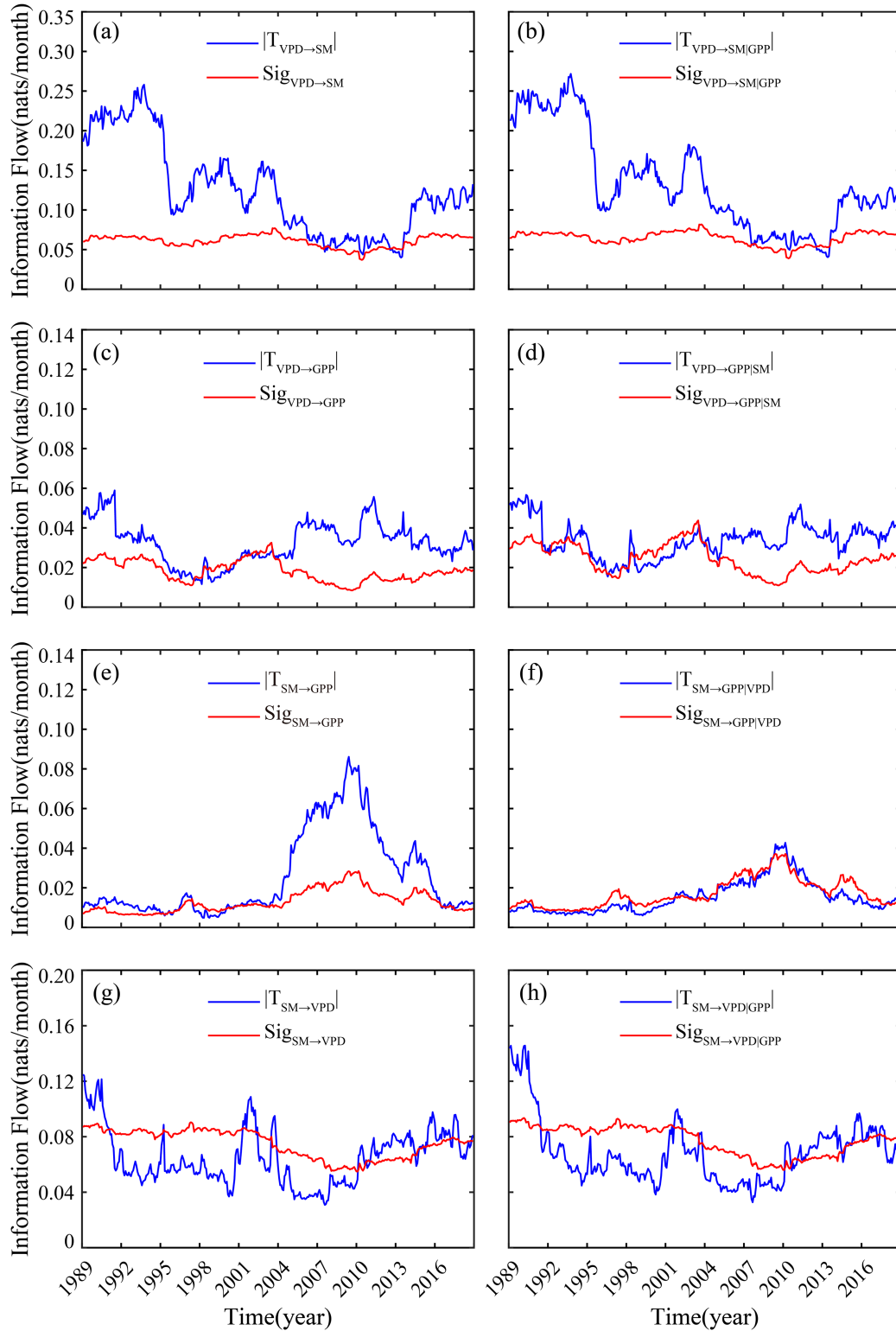


FIG. 6. The absolute values of the time-varying information flows among soil moisture (SM), gross primary productivity (GPP), and vapor pressure deficit (VPD) (in nats per month) for RA: (a) $|T_{VPD \rightarrow SM}|$ and $Sig_{VPD \rightarrow SM}$; (b) $|T_{VPD \rightarrow SM|GPP}|$ and $Sig_{VPD \rightarrow SM|GPP}$; (c) $|T_{VPD \rightarrow GPP}|$ and $Sig_{VPD \rightarrow GPP}$; (d) $|T_{VPD \rightarrow GPP|SM}|$ and $Sig_{VPD \rightarrow GPP|SM}$; (e) $|T_{SM \rightarrow GPP}|$ and $Sig_{SM \rightarrow GPP}$; (f) $|T_{SM \rightarrow GPP|VPD}|$ and $Sig_{SM \rightarrow GPP|VPD}$; (g) $|T_{SM \rightarrow VPD}|$ and $Sig_{SM \rightarrow VPD}$; (h) $|T_{SM \rightarrow VPD|GPP}|$ and $Sig_{SM \rightarrow VPD|GPP}$. All the significance tests are performed at a 5% significance level.

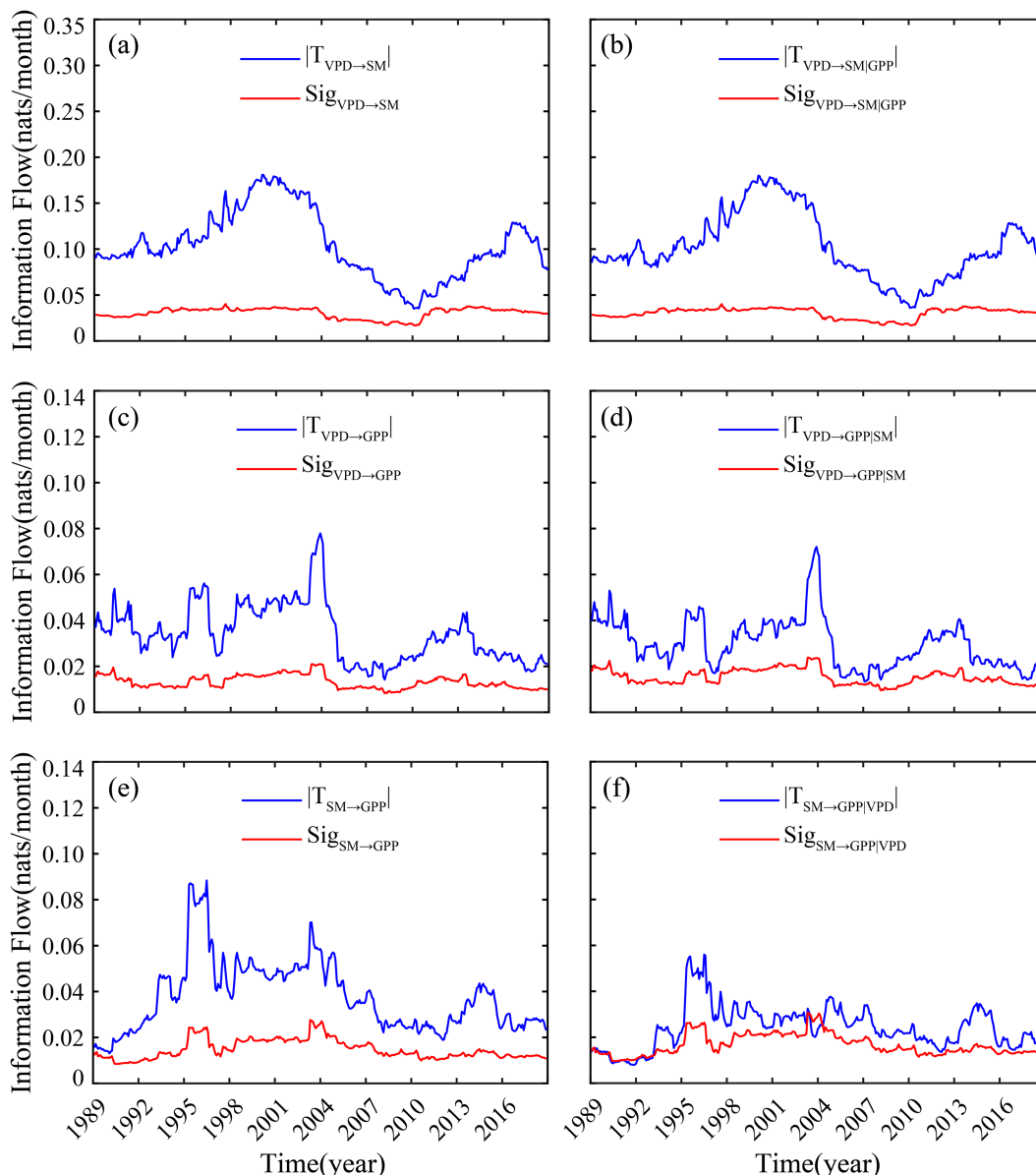


FIG. 7. The absolute values of the time-varying information flows among soil moisture (SM), gross primary productivity (GPP), and vapor pressure deficit (VPD) (in nats per month) for RB: (a) $|T_{\text{VPD} \rightarrow \text{SM}}|$ and $\text{Sig}_{\text{VPD} \rightarrow \text{SM}}$; (b) $|T_{\text{VPD} \rightarrow \text{SM}|\text{GPP}}|$ and $\text{Sig}_{\text{VPD} \rightarrow \text{SM}|\text{GPP}}$; (c) $|T_{\text{VPD} \rightarrow \text{GPP}}|$ and $\text{Sig}_{\text{VPD} \rightarrow \text{GPP}}$; (d) $|T_{\text{VPD} \rightarrow \text{GPP}|\text{SM}}|$ and $\text{Sig}_{\text{VPD} \rightarrow \text{GPP}|\text{SM}}$; (e) $|T_{\text{SM} \rightarrow \text{GPP}}|$ and $\text{Sig}_{\text{SM} \rightarrow \text{GPP}}$; (f) $|T_{\text{SM} \rightarrow \text{GPP}|\text{VPD}}|$ and $\text{Sig}_{\text{SM} \rightarrow \text{GPP}|\text{VPD}}$. All the significance tests are performed at a 5% significance level.

Finally, we have summarized the aforementioned results to facilitate a clearer and more intuitive comparison of the differences between the two regions. First, the causal relationships among SM, VPD, and GPP in RA are sketched in Fig. 8. We have divided the results from 1989 to 2018 into three segments for presentation. The left, center, and right columns in the figure represent the periods January 1989–December 1998, January 1999–December 2008, and January 2009–December 2018, respectively. The red arrow indicates that the absolute value of the information flow (IF) increases in this time range, and the blue arrow indicates that the absolute value of the IF decreases. The thickness of the arrow represents the absolute value of the

IF, a thicker arrow indicates a larger absolute value of the IF. Figures 8a1–c1 and 8a2–c2 represent the causality between VPD to SM and the causality between VPD to SM influenced by GPP, respectively. Figures 8d1–f1 and 8d2–f2 represent the causality between VPD to GPP and the causality between VPD to GPP influenced by SM, respectively. Figures 8g1–i1 and 8g2–i2 represent the causality between SM to GPP and the causality between SM to GPP influenced by VPD, respectively.

As shown in Figs. 8a1–c1 and 8a2–c2, regardless of whether the influence of GPP on the system is considered, the magnitude of causality from VPD to SM generally displays a trend of decreasing from 1989 to 2008 and then increasing from 2009 to

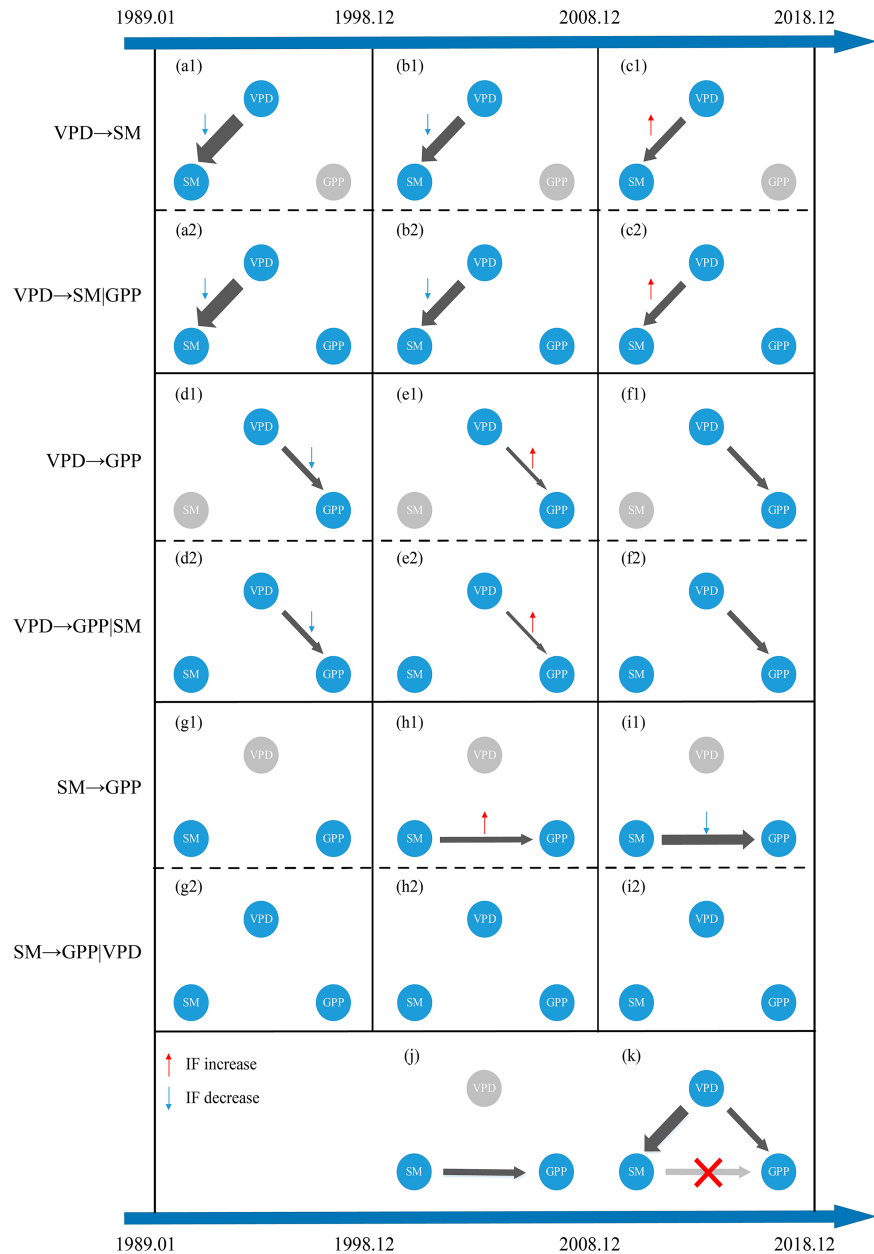


FIG. 8. A schematic of the time-varying causality among SM, VPD, and GPP for RA. Shown are the time periods (left) January 1989–December 1998, (center) January 1999–December 2008, and (right) January 2009–December 2018. (a1)–(c1),(a2)–(c2) The causality between VPD to SM and the causality between VPD to SM influenced by GPP, respectively. (d1)–(f1),(d2)–(f2) The causality between VPD to GPP and the causality between VPD to GPP influenced by SM, respectively. (g1)–(i1),(g2)–(i2) The causality between SM to GPP and the causality between SM to GPP influenced by VPD, respectively. (j) The causality between SM to GPP in 2004–18 and (k) the change of causality when the influence of VPD is taken into account, and the real causal relationships in 2004–18. The red arrow indicates that the absolute value of the IF increases, and the blue arrow indicates that the absolute value of the IF decreases. The thickness of the arrow represents the absolute value of the information flow.

2018. Moreover, there is no significant difference in the causal strength between the two scenarios. Similarly, Figs. 8d1–f1 and 8d2–f2 show that irrespective of considering the influence of SM on the system, the magnitude of causality from VPD to GPP primarily shows a trend of first decreasing and then increasing during 1989–2008 and remains relatively stable within 2009–18, and there is also no significant difference in the causal strength between the two scenarios. Figures 8g1–i1 and 8g2–i2 indicate that only within the range of 1989–98, whether or not the influence of VPD on the system is considered, the causality of SM to GPP is almost negligible, and there is no noticeable difference between them. However, as shown in Figs. 8h1–i1 and 8h2–i2, there are significant differences between the influence of SM on GPP without considering VPD effects and the influence of SM on GPP when considering VPD effects. When not considering the influence of VPD on the system, the magnitude of causality from SM to GPP exhibits a trend of first increasing and then decreasing, as shown in Figs. 8h1–i1. However, when considering the influence of VPD on the system, there is no direct causality between SM to GPP, as shown in Figs. 8h2–i2.

To better demonstrate the influence mechanism of SM and VPD to GPP in RA, we have individually highlighted the most representative findings, as shown in Figs. 8j and 8k. Figure 8j represents the causality between SM to GPP in 1999–2018, Fig. 8k represents the change of causality when the influence of VPD is taken into account, and the real causal relationships in 1999–2018. It represents when VPD is added into the system, a change in causality occurs, so VPD is a common cause of SM and GPP. Thus, although all these three variables are highly correlated, the causalities on SM and GPP look like Fig. 8k in 1999–2018. In addition, we also found the information flow from VPD to GPP influenced by SM is higher than the information flow from SM to GPP influenced by VPD. These results show that MtvLK can detect the influence of confounder(s), and correctly identify causalities in complex multivariate systems.

Similar to Fig. 8, the causal relationships among SM, VPD, and GPP in RB are sketched in Fig. 9. As shown in Figs. 9a1–c1 and 9a2–c2, regardless of whether the influence of GPP on the system is considered, the magnitude of causality from VPD to SM generally shows a trend of increasing first, then decreasing, and then increasing again in three time periods. Moreover, there is no significant difference in the magnitude of causality between the two cases. Figures 9d1–f1 and 9d2–f2 show that regardless of whether or not considering the influence of SM on the system, the magnitude of causality from VPD to GPP is relatively stable from 1989 to 1998, and shows a trend of increasing first and then decreasing from 1999 to 2008. Finally, it also shows a trend of increasing first and then decreasing from 2009 to 2018, and the magnitude of causality from 2009 to 2018 is lower than that from 1999 to 2008. In addition, we also find that when considering the influence of SM on the system, the magnitude of causality from VPD to GPP is lower than that without considering the influence of SM on the system, this hints at the existence of potential feedback from SM. Figures 9g1–i1 and 9g2–i2 represent the causality between SM to GPP without considering the influence of VPD, and the causality between SM to GPP with considering the influence of VPD, respectively. Figures 9g1–i1 show that without considering the influence of

VPD, the magnitude of causality from SM to GPP showed an increasing trend in 1989–98, then showed a decreasing trend in 1999–2008, and finally showed a trend of increasing first and then decreasing in 2009–18. Figures 9g2–i2 show that when considering the influence of VPD, the magnitude of causality from SM to GPP increased first and then decreased from 1989 to 1998, then remained relatively stable from 1999 to 2008, and finally increased first and then decreased again from 2009 to 2018. Additionally, when comparing Figs. 9g1–i1 with Figs. 9g2–i2, it is evident that the direct causality from SM to GPP persists regardless of whether the influence of VPD is considered. However, without taking into account the influence of VPD, the magnitude of causality from SM to GPP is stronger than when the influence of VPD is considered. This suggests that VPD has a significant impact on the causality from SM to GPP.

Finally, we summarized the influence mechanism of SM and VPD to GPP in RB, as shown in Figs. 9j and 9k. Figure 9j shows the causality from SM to GPP in the absence of VPD. It is consistent with the result in RA. When VPD is added into the system, the direct causality from SM to GPP still exists, in contrast to that in RA; besides that, there is also a direct causality from VPD to GPP. Besides these two causal relations, we note that there also exists an indirect causality from VPD to GPP via SM, as shown in Fig. 9k. This real-world problem application shows that SM and VPD influence GPP in the semihumid area. Moreover, the information flow from VPD to GPP influenced by SM is larger in magnitude than that from SM to GPP influenced by VPD. From this, the MtvLK algorithm can correctly identify the causality in highly correlated complex systems and effectively give the magnitude of causality. It hence provides a powerful tool for causal inference for complex Earth system problems.

5. Discussion

Identifying causal relationships and quantifying their strength from time series can help us gain insight into the physical mechanism underlying the observations: a key problem in scientific research. The Liang–Kleeman information flow quantifies the rate of information between systems, showing both the direction and magnitude within a cause-and-effect relation (Liang 2014, 2016, 2021). Since it is rigorously formulated from first principles, it avoids the spurious causality sometimes identified in Granger causality tests and tests with other statistical formalisms. Due to the fact that LK is computed based on the covariance between given time series, the LK is relatively easier to compute compared to other approaches. Moreover, Liang (2018) demonstrated with a nonlinear three-dimensional stochastic gradient system that that LK can still accurately provide potential causal relationships. Nonetheless, the potential limitations should also be studied in future studies. In this study we extended a previous solution to the problem of bivariate LK causality estimation with nonstationary series (Hagan et al. 2019) to the multivariate cases, based on the Kalman filter's learning algorithm, which can identify immediate changes when tracking time-varying information flow within a system. This endeavor yields a time-varying causal structure as seen in Eq. (16). In

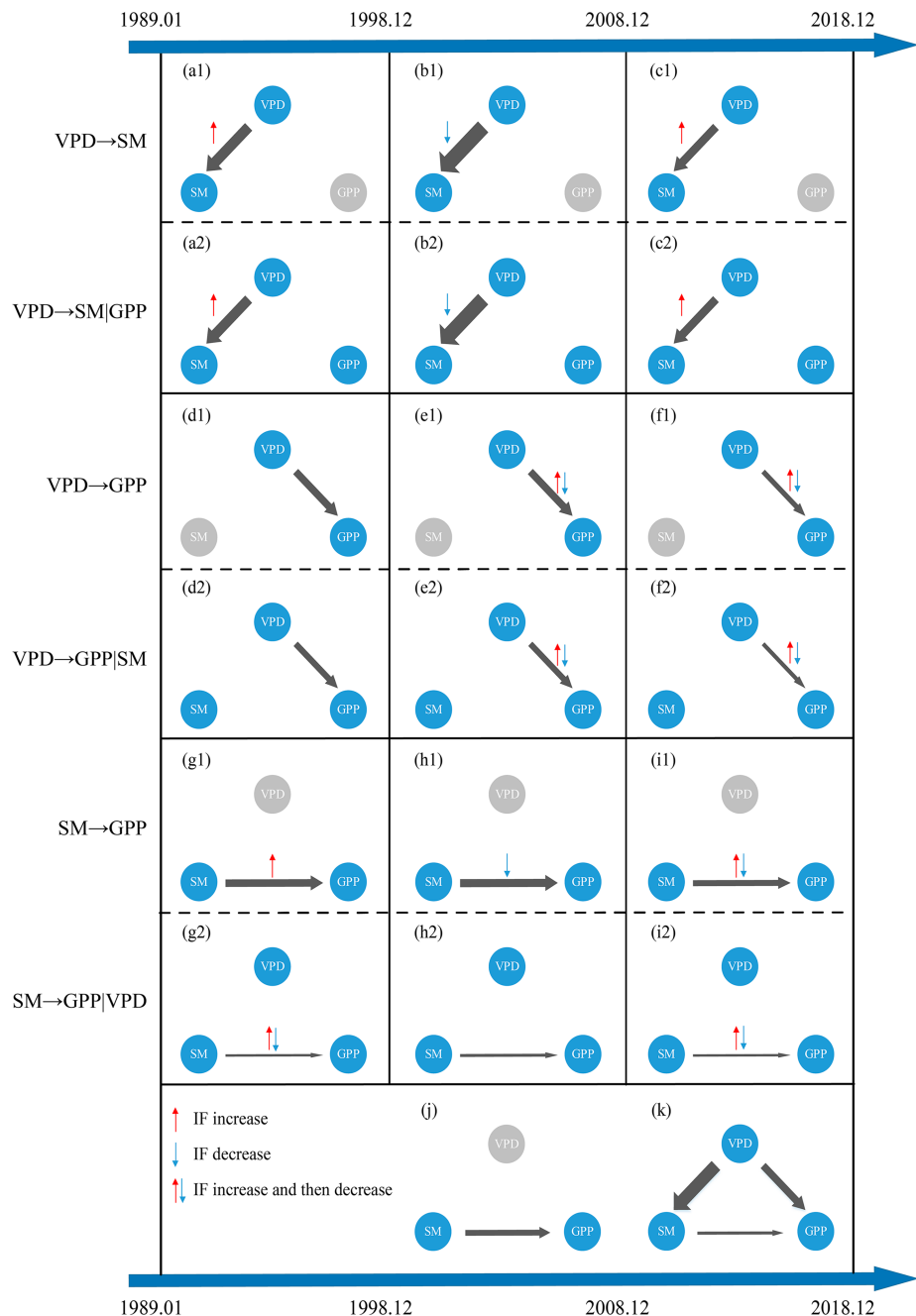


FIG. 9. A schematic of the time-varying causality among the SM, VPD, and GPP for RB. Shown are the time periods (left) January 1989–December 1998, (center) January 1999–December 2008, and (right) January 2009–December 2018. (a1)–(c1),(a2)–(c2) The causality between VPD to SM and the causality between VPD to SM influenced by GPP, respectively. (d1)–(f1),(d2)–(f2) The causality between VPD to GPP and the causality between VPD to GPP influenced by SM, respectively. (g1)–(i1),(g2)–(i2) The causality between SM to GPP and the causality between SM to GPP influenced by VPD, respectively. (j) The inferred causal relation between SM and GPP in the absence of VPD in 1989–2018. (k) The causal relationship among VPD, SM, and GPP in 1989–2018. The red arrow indicates that the absolute value of the IF increases, and the blue arrow indicates that the absolute value of the IF decreases. The thickness of the arrow represents the absolute value of the information flow.

performing the Kalman filter, a lookback window length is needed to calculate the moving average when estimating $\mathbf{Q}$ and $\mathbf{R}$. Here, we use the exponential weighted moving average (EWMA) and unweighted moving average (UWMA), which provide an information length for the window length lost to do the forward update of the filter. We obtained the time-varying information flow without missing data by running the filter backward using the last estimate of the forward recursion for initialization. Then, we combined both the forward and backward runs together. This algorithm is abbreviated as MtvLK.

MtvLK has been validated in three idealized synthetic models, with causality between variables by design varying over time as shown in Figs. 2–4. In addition, MtvLK effectively solves the problem of missing window length information incurred in choosing a lookback window length to compute the moving average when using EMWA and UWMA to estimate $\mathbf{Q}$ and $\mathbf{R}$ offline. Hence there is no missing value at each time step in our results. Additionally, the estimated time-varying causality accurately reproduces the preset causality in the models. This is because Kalman filtering has the ability to detect changes. Therefore, immediate changes along time will be preserved when estimating causal relationships using the MtvLK, which is challenging to achieve in non-time-varying LK information flow and methods based on standard sliding windows. In addition, the principle of nil causality, which appears as a theorem in the LK formula (Liang 2014, 2016, 2021), is also met in the MtvLK estimate, while with some commonly used causal inference approaches this may not always be true. Moreover, even with highly correlated series, MtvLK is still effective, as verified with the toy model (namely, the synthetic model 1): Although the confounding variable x_2 exists in the system, MtvLK can still correctly identify the real causality in the multivariable system.

We have also applied the MtvLK in a real-world problem to investigate how SM and VPD affect GPP in two selected typical regions of China. Our results show that over the humid Huanan region in southeastern China, causality exists from VPD to GPP. While SM is highly correlated with GPP, no direct causality exists between SM and GPP. In the semihumid and semiarid Huabei regions, direct causality exists both from VPD to GPP and from SM to GPP.

There is still more room to improve the MtvLK algorithm. Attention needs to be paid particularly on three issues: First, the computation is rather expensive because of the Kalman filter used here, so the computational efficiency is significantly lower than the original estimation of causality as documented in Liang (2021). The use of EMWA and UWMA to estimate $\mathbf{Q}$ and $\mathbf{R}$ would need further improvement to avoid defining a time window length. A more objective way to define a window length would provide a better estimation for the probability density function of the causal structure. Finally, future studies on MtvLK should look into incorporating a frequency domain as was done by Hagan et al. (2019) to show causalities at different time scales which is important to isolate long- and short-causal impacts. After these optimizations, a fully optimized MtvLK will represent a method with immense potential. Not only can it shed new light on the intricate

time–frequency causality relationships in land–atmosphere interactions, but it can also be applied to analyze complex causal relationships in various other fields, such as finance, further enhancing our understanding of causality issues across diverse domains.

Acknowledgments. This research was funded by the National Natural Science Foundation of China (42275028) and the Sino-German Cooperation Group Program (GZ1447). We are also grateful to all data providers who made their data available for use in this study.

Data availability statement. All code required to replicate this study can be obtained from <https://github.com/FeihongZhou/MtvLK.git>. All data used for this study are publicly available online. The FLUXCOM GPP data are provided by Max Planck Institute for Biogeochemistry at <http://www.fluxcom.org/CF-Download/>. The ERA5-Land data are provided by the European Centre for Medium-Range Weather Forecasts at <https://cds.climate.copernicus.eu>.

APPENDIX A

Solution for Missing Window Length Information Length

Consider a d -dimensional system. First, we compute the time-varying information flow from x_2 to x_1 based on Eq. (16), denoted as $T_{2 \rightarrow 1}$. Aiming at the problem of missing values in the initial lookback window, we reverse the time series of x_1 and x_2 , denoted as x'_1 and x'_2 , respectively. Then we compute the time-varying information flow from x'_1 and x'_2 , denoted as $T'_{1' \rightarrow 2'}$. Next we reverse the obtained result $T'_{1' \rightarrow 2'}$, and denote it as $T''_{1' \rightarrow 2'}$. This step reverses the missing value in the initial lookback window to the tail of the sequence. Finally, we replace the missing value of the initial lookback window in $T_{2 \rightarrow 1}$ with the corresponding region in $T''_{1' \rightarrow 2'}$ to obtain the final result without missing values. It is important to highlight that although this method effectively addresses the issue of missing values in the initial backtracking window, we contend that it does not provide a comprehensive explanation of the underlying physical mechanisms involved in the backtracking window. This limitation arises from the fact that the method calculates based on reverse-engineered time series, essentially employing backward data computation to move forward. Therefore, while the information flow values obtained from this method can serve as a reference for discerning causal relationships between variables in the backtracking window, they are inadequate for effectively elucidating the associated physical mechanisms.

APPENDIX B

Significance Test

For the statistical tool such as that in Eq. (16), significance testing enables us to better understand the robustness of the derived result. When N is large, the information flow is approximately normally distributed around their true values with

variances $(C_{12}/C_{11})^2 \sigma_{a_{12}}^2$, which is derived from maximum likelihood estimation. Denote $\theta = (f, a_{11}, a_{12}, \dots, a_{1d}, b_1)$,

$$\mathbf{I}_{ij} = -\frac{1}{N} \sum_{n=1}^N \frac{\partial^2 \log p(\mathbf{X}_{n+1} | \mathbf{X}_n; \hat{\theta})}{\partial \theta_i \partial \theta_j}, \quad (\text{B1})$$

where $\mathbf{I}$ is a $(d+2) \times (d+2)$ matrix, namely, the Fisher information matrix. The inverse $(\mathbf{I})^{-1}$ is the covariance matrix of θ , which has $\sigma_{a_{12}}^2$; ρ is the marginal probability of $\{\mathbf{X}_n\}$, a Markov process. For details, refer to Liang (2021).

APPENDIX C

Addition Testing of the Synthetic Model

a. Synthetic model 1

To further verify the accuracy of MtvLK in complex multivariate system, we tested all the links in synthetic model 1 that did not preset causality by using the same conditions

outlined in section 3. The time-varying information flow results (in nats per unit time) are presented in Fig. C1. The results show that $|T_{x3 \rightarrow x1|x2}|$ and $|T_{x3 \rightarrow x2|x1}|$ are both no causalities for $0 < t \leq 2000$. These results align with the design in Eq. (17).

b. Synthetic model 2

The time-varying information flow results (in nats per unit time) that did not preset causality in synthetic model 2 are shown in Fig. C2. The results show that $|T_{x1 \rightarrow x2|x3}|$, $|T_{x3 \rightarrow x1|x2}|$, and $|T_{x3 \rightarrow x2|x1}|$ are all no causalities for $0 < t \leq 2000$. These results align with the design in Eq. (18).

c. Synthetic model 3

The time-varying information flow results (in nats per unit time) that did not preset causality in the synthetic model 3 are shown in Fig. C3. Figures C3c–n show that all the test links show no causalities for $0 < t \leq 2000$. These results align with the design in Eq. (19).

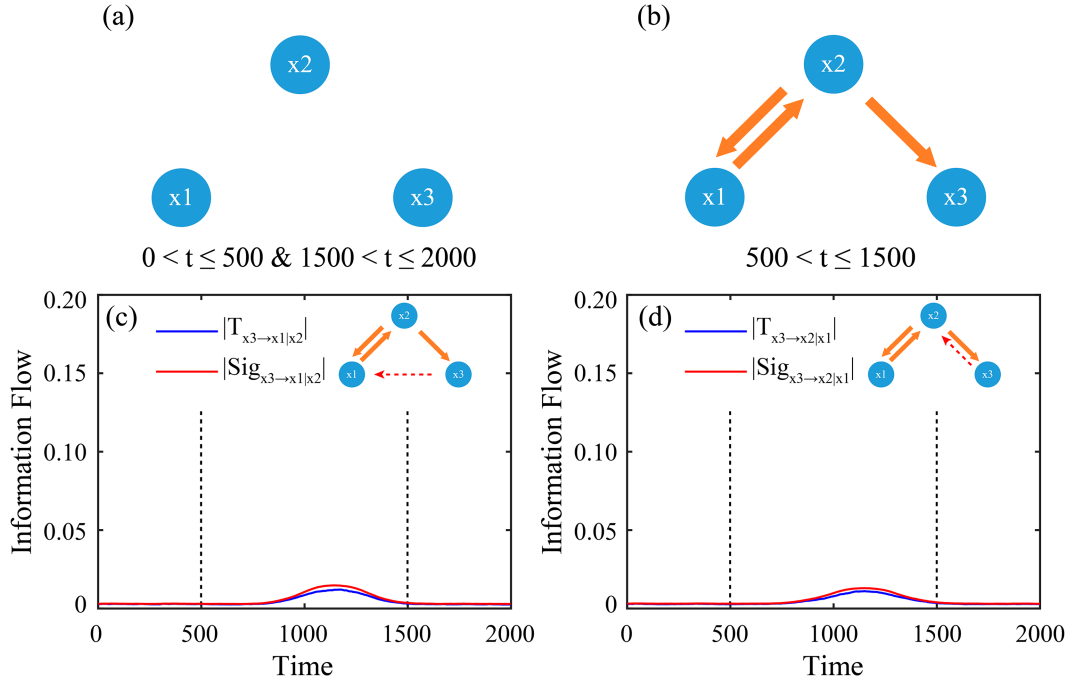


FIG. C1. (a) Schematic diagram of the conceptual model as described in Eq. (17) when $0 < t \leq 500$ and $1500 < t \leq 2000$. (b) As in (a), but for interval $500 < t \leq 1500$. Absolute values of the time-varying information flows among x_1 , x_2 , and x_3 (in nats per unit time): (c) $|T_{x3 \rightarrow x1|x2}|$ and $|Sig_{x3 \rightarrow x1|x2}|$; (d) $|T_{x3 \rightarrow x2|x1}|$ and $|Sig_{x3 \rightarrow x2|x1}|$. Dashed lines represent the causal relationship turning points. All the results are the averages of 1000 realizations, and all the significance tests are performed at a 1% significance level.

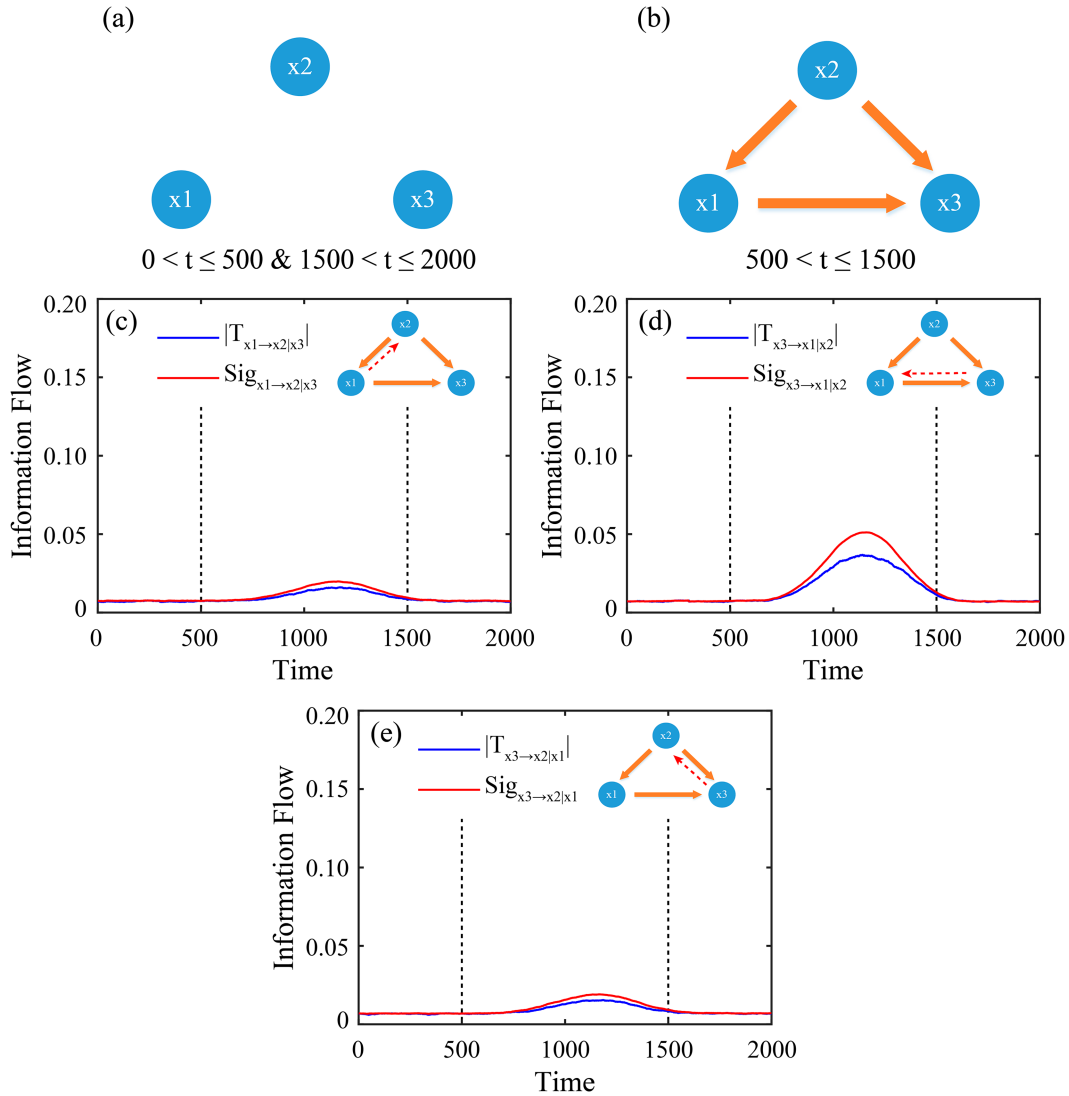


FIG. C2. (a) Schematic diagram of the conceptual model as described in Eq. (18) when $0 < t \leq 500$ and $1500 < t \leq 2000$. (b) As in (a), but for the interval $500 < t \leq 1500$. Absolute values of the time-varying information flows among x_1 , x_2 , and x_3 (in nats per unit time): (c) $|T_{x1 \rightarrow x2|x3}|$ and $\text{Sig}_{x1 \rightarrow x2|x3}$; (d) $|T_{x3 \rightarrow x1|x2}|$ and $\text{Sig}_{x3 \rightarrow x1|x2}$; (e) $|T_{x3 \rightarrow x2|x1}|$ and $\text{Sig}_{x3 \rightarrow x2|x1}$. Dashed lines represent the causal relationship turning points. All the results are the averages of 1000 realizations, and all the significance tests are performed at a 1% significance level.

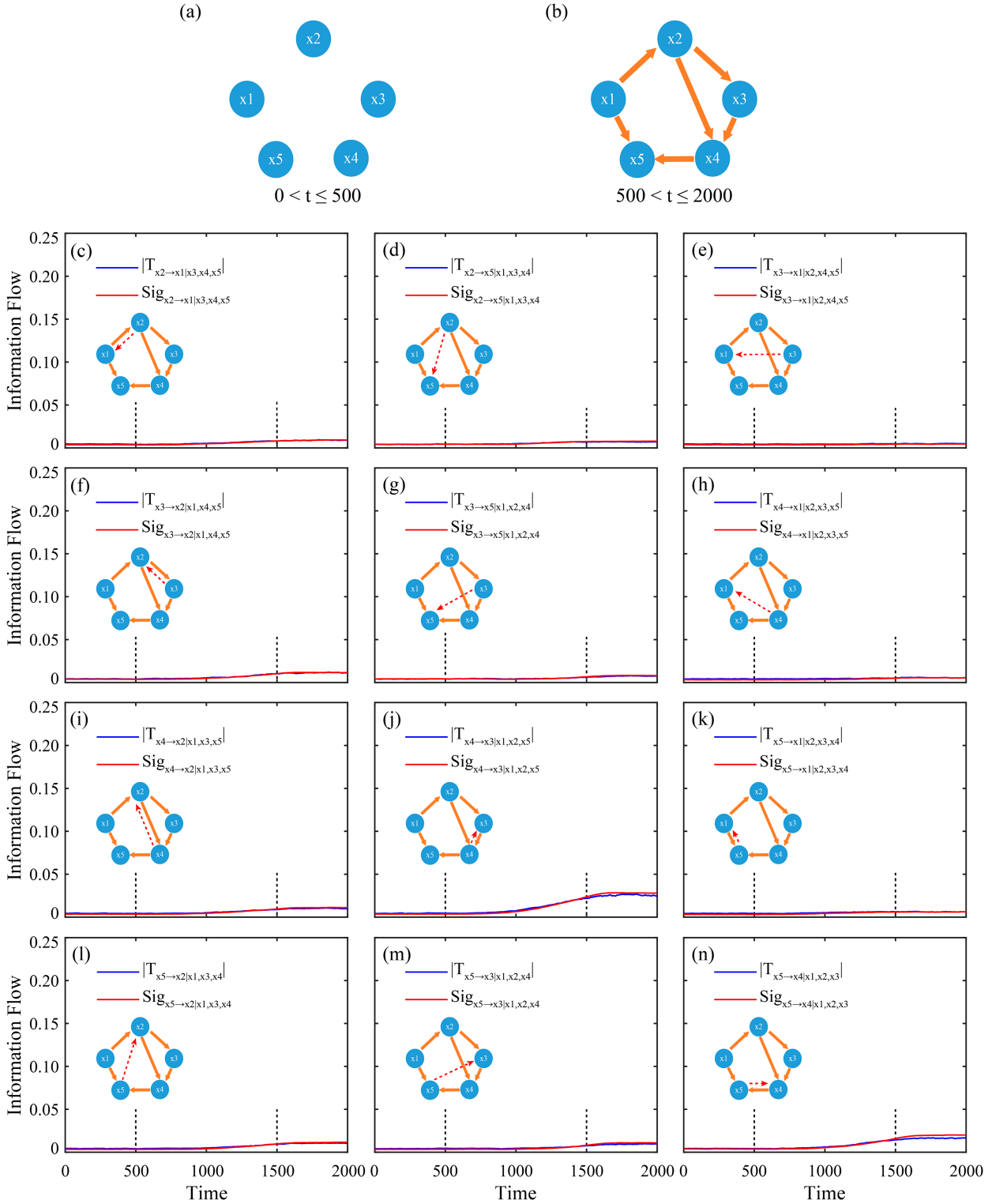


FIG. C3. (a) Schematic diagram of the conceptual model as described in Eq. (19) when $0 < t \leq 500$. (b) As in (a), but for the interval $500 < t \leq 2000$. Absolute values of the time-varying information flows among x_1 , x_2 , x_3 , x_4 , and x_5 (in nats per unit time): (c) $|T_{x_2 \rightarrow x_1|x_3,x_4,x_5}|$ and $\text{Sig}_{x_2 \rightarrow x_1|x_3,x_4,x_5}$; (d) $|T_{x_2 \rightarrow x_5|x_1,x_3,x_4}|$ and $\text{Sig}_{x_2 \rightarrow x_5|x_1,x_3,x_4}$; (e) $|T_{x_3 \rightarrow x_1|x_2,x_4,x_5}|$ and $\text{Sig}_{x_3 \rightarrow x_1|x_2,x_4,x_5}$; (f) $|T_{x_3 \rightarrow x_2|x_1,x_4,x_5}|$ and $\text{Sig}_{x_3 \rightarrow x_2|x_1,x_4,x_5}$; (g) $|T_{x_3 \rightarrow x_5|x_1,x_2,x_4}|$ and $\text{Sig}_{x_3 \rightarrow x_5|x_1,x_2,x_4}$; (h) $|T_{x_4 \rightarrow x_1|x_2,x_3,x_5}|$ and $\text{Sig}_{x_4 \rightarrow x_1|x_2,x_3,x_5}$; (i) $|T_{x_4 \rightarrow x_2|x_1,x_3,x_5}|$ and $\text{Sig}_{x_4 \rightarrow x_2|x_1,x_3,x_5}$; (j) $|T_{x_4 \rightarrow x_3|x_1,x_2,x_5}|$ and $\text{Sig}_{x_4 \rightarrow x_3|x_1,x_2,x_5}$; (k) $|T_{x_5 \rightarrow x_1|x_2,x_3,x_4}|$ and $\text{Sig}_{x_5 \rightarrow x_1|x_2,x_3,x_4}$; (l) $|T_{x_5 \rightarrow x_2|x_1,x_3,x_4}|$ and $\text{Sig}_{x_5 \rightarrow x_2|x_1,x_3,x_4}$; (m) $|T_{x_5 \rightarrow x_3|x_1,x_2,x_4}|$ and $\text{Sig}_{x_5 \rightarrow x_3|x_1,x_2,x_4}$; (n) $|T_{x_5 \rightarrow x_4|x_1,x_2,x_3}|$ and $\text{Sig}_{x_5 \rightarrow x_4|x_1,x_2,x_3}$. Dashed lines represent the causal relationship turning points. All the results are the averages of 1000 realizations, and all the significance tests are performed at a 1% significance level.

REFERENCES

- Allen, R. G., L. S. Pereira, D. Raes, and M. Smith, 1998: Crop evapotranspiration—Guidelines for computing crop water requirements. FAO Irrigation and Drainage Paper 56, 326 pp., https://www.scscourt.org/complexcivil/105cv049053/volume3/172618e_5xagwax8.pdf.
- Anderegg, W. R. L., and Coauthors, 2017: Plant water potential improves prediction of empirical stomatal models. *PLOS ONE*, **12**, e0185481, <https://doi.org/10.1371/journal.pone.0185481>.
- Bai, C., R. Zhang, S. Bao, X. S. Liang, and W. Guo, 2018: Forecasting the tropical cyclone genesis over the northwest Pacific through identifying the causal factors in cyclone–climate interactions. *J. Atmos. Oceanic Technol.*, **35**, 247–259, <https://doi.org/10.1175/JTECH-D-17-0109.1>.
- Baker, J. C. A., and Coauthors, 2021: An assessment of land–atmosphere interactions over South America using satellites, reanalysis, and two global climate models. *J. Hydrometeorol.*, **22**, 905–922, <https://doi.org/10.1175/JHM-D-20-0132.1>.
- Barrett, A., and L. Barnett, 2013: Granger causality is designed to measure effect, not mechanism. *Front. Neuroinfo.*, **7**, 6, <https://doi.org/10.3389/fninf.2013.00006>.
- Berg, A., B. R. Lintner, K. L. Findell, S. Malyshev, P. C. Loikith, and P. Gentine, 2014: Impact of soil moisture–atmosphere interactions on surface temperature distribution. *J. Climate*, **27**, 7976–7993, <https://doi.org/10.1175/JCLI-D-13-00591.1>.
- , and Coauthors, 2016: Land–atmosphere feedbacks amplify aridity increase over land under global warming. *Nat. Climate Change*, **6**, 869–874, <https://doi.org/10.1038/nclimate3029>.
- Besnard, S., and Coauthors, 2021: Global sensitivities of forest carbon changes to environmental conditions. *Global Change Biol.*, **27**, 6467–6483, <https://doi.org/10.1111/gcb.15877>.
- Bierman, G. J., and C. L. Thornton, 1977: Numerical comparison of Kalman filter algorithms: Orbit determination case study. *Automatica*, **13**, 23–35, [https://doi.org/10.1016/0005-1098\(77\)90006-1](https://doi.org/10.1016/0005-1098(77)90006-1).
- Chen, H., and Coauthors, 2021: Soil drying weakens the positive effect of climate factors on global gross primary production. *Ecol. Indic.*, **129**, 107953, <https://doi.org/10.1016/j.ecolind.2021.107953>.
- Chen, N., and Coauthors, 2020: The compensation effects of post-drought regrowth on earlier drought loss across the Tibetan Plateau grasslands. *Agric. For. Meteorol.*, **281**, 107822, <https://doi.org/10.1016/j.agrformet.2019.107822>.
- , and Coauthors, 2021: Divergent impacts of atmospheric water demand on gross primary productivity in three typical ecosystems in China. *Agric. For. Meteorol.*, **307**, 108527, <https://doi.org/10.1016/j.agrformet.2021.108527>.
- Ciabatta, L., L. Brocca, C. Massari, T. Moramarco, S. Puca, A. Rinollo, S. Gabellani, and W. Wagner, 2015: Integration of satellite soil moisture and rainfall observations over the Italian territory. *J. Hydrometeorol.*, **16**, 1341–1355, <https://doi.org/10.1175/JHM-D-14-0108.1>.
- Cong, J., and Coauthors, 2023: Altered default mode network causal connectivity patterns in autism spectrum disorder revealed by Liang information flow analysis. *Hum. Brain Mapp.*, **44**, 2279–2293, <https://doi.org/10.1002/hbm.26209>.
- Cui, J. P., S. Piao, C. Huntingford, X. Wang, X. Lian, A. Chevuturi, A. G. Turner, and G. J. Kooperman, 2020: Vegetation forcing modulates global land monsoon and water resources in a CO₂-enriched climate. *Nat. Commun.*, **11**, 5184, <https://doi.org/10.1038/s41467-020-18992-7>.
- Detto, M., A. Molini, G. Katul, P. Stoy, S. Palmroth, and D. Baldocchi, 2012: Causality and persistence in ecological systems: A nonparametric spectral Granger causality approach. *Amer. Nat.*, **179**, 524–535, <https://doi.org/10.1086/664628>.
- Docquier, D., S. Vannitsem, F. Ragone, K. Wyser, and X. S. Liang, 2022: Causal links between Arctic sea ice and its potential drivers based on the rate of information transfer. *Geophys. Res. Lett.*, **49**, e2021GL095892, <https://doi.org/10.1029/2021GL095892>.
- Duerinck, H. M., R. J. van der Ent, N. C. van de Giesen, G. Schoups, V. Babovic, and P. J.-F. Yeh, 2016: Observed soil moisture–precipitation feedback in Illinois: A systematic analysis over different scales. *J. Hydrometeorol.*, **17**, 1645–1660, <https://doi.org/10.1175/JHM-D-15-0032.1>.
- Feng, H., B. Zou, and J. Luo, 2017: Coverage-dependent amplifiers of vegetation change on global water cycle dynamics. *J. Hydrol.*, **550**, 220–229, <https://doi.org/10.1016/j.jhydrol.2017.04.056>.
- Fu, Z., and Coauthors, 2022: Atmospheric dryness reduces photosynthesis along a large range of soil water deficits. *Nat. Commun.*, **13**, 989, <https://doi.org/10.1038/s41467-022-28652-7>.
- Goodwell, A. E., P. Jiang, B. L. Ruddell, and P. Kumar, 2020: Debates—Does information theory provide a new paradigm for Earth science? Causality, interaction, and feedback. *Water Resour. Res.*, **56**, e2019WR024940, <https://doi.org/10.1029/2019WR024940>.
- Granger, C. W. J., 1969: Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, **37**, 424–438, <https://doi.org/10.2307/1912791>.
- Green, J. K., A. G. Konings, S. H. Alemohammad, J. Berry, D. Entekhabi, J. Kolassa, J.-E. Lee, and P. Gentine, 2017: Regionally strong feedbacks between the atmosphere and terrestrial biosphere. *Nat. Geosci.*, **10**, 410–414, <https://doi.org/10.1038/ngeo2957>.
- Hagan, D. F. T., G. Wang, X. S. Liang, and H. A. J. Dolman, 2019: A time-varying causality formalism based on the Liang–Kleeman information flow for analyzing directed interactions in nonstationary climate systems. *J. Climate*, **32**, 7521–7537, <https://doi.org/10.1175/JCLI-D-18-0881.1>.
- , H. A. J. Dolman, G. Wang, K. T. C. L. K. Sian, K. Yang, W. Ullah, and R. P. Shen, 2022: Contrasting ecosystem constraints on seasonal terrestrial CO₂ and mean surface air temperature causality projections by the end of the 21st century. *Environ. Res. Lett.*, **17**, 124019, <https://doi.org/10.1088/1748-9326/aca551>.
- Havlicek, M., J. Jan, M. Brazdil, and V. D. Calhoun, 2010: Dynamic Granger causality based on Kalman filter for evaluation of functional network connectivity in fMRI data. *Neuroimage*, **53**, 65–77, <https://doi.org/10.1016/j.neuroimage.2010.05.063>.
- Humphrey, V., A. Berg, P. Ciais, P. Gentine, M. Jung, M. Reichstein, S. I. Seneviratne, and C. Frankenberg, 2021: Soil moisture–atmosphere feedback dominates land carbon uptake variability. *Nature*, **592**, 65–69, <https://doi.org/10.1038/s41586-021-03325-5>.
- Jung, M., and Coauthors, 2017: Compensatory water effects link yearly global land CO₂ sink changes to temperature. *Nature*, **541**, 516–520, <https://doi.org/10.1038/nature20780>.
- Kalman, R. E., 1960: A new approach to linear filtering and prediction problems. *J. Basic Eng.*, **82**, 35–45, <https://doi.org/10.1115/1.3662552>.
- Kimm, H., K. Guan, P. Gentine, J. Wu, C. J. Bernacchi, B. N. Sulman, T. J. Griffis, and C. Lin, 2020: Redefining droughts

- for the U.S. Corn Belt: The dominant role of atmospheric vapor pressure deficit over soil moisture in regulating stomatal behavior of maize and soybean. *Agric. For. Meteorol.*, **287**, 107930, <https://doi.org/10.1016/j.agrformet.2020.107930>.
- Lemondant, L., P. Gentile, M. Stefanon, P. Drobinski, and S. Fatichi, 2016: Modification of land-atmosphere interactions by CO₂ effects: Implications for summer dryness and heat wave amplitude. *Geophys. Res. Lett.*, **43**, 10240–10248, <https://doi.org/10.1002/2016GL069896>.
- Le Quéré, C., and Coauthors, 2015: Global carbon budget 2015. *Earth Syst. Sci. Data*, **7**, 349–396, <https://doi.org/10.5194/essd-7-349-2015>.
- Li, S., G. Wang, C. Zhu, J. Lu, W. Ullah, D. F. T. Hagan, G. Kattell, and J. Peng, 2022: Attribution of global evapotranspiration trends based on the Budyko framework. *Hydrol. Earth Syst. Sci.*, **26**, 3691–3707, <https://doi.org/10.5194/hess-26-3691-2022>.
- Liang, X. S., 2014: Unraveling the cause-effect relation between time series. *Phys. Rev. E*, **90**, 052150, <https://doi.org/10.1103/PhysRevE.90.052150>.
- , 2016: Information flow and causality as rigorous notions *ab initio*. *Phys. Rev. E*, **94**, 052201, <https://doi.org/10.1103/PhysRevE.94.052201>.
- , 2018: Causation and information flow with respect to relative entropy. *Chaos*, **28**, 075311, <https://doi.org/10.1063/1.5010253>.
- , 2021: Normalized multivariate time series causality analysis and causal graph reconstruction. *Entropy*, **23**, 679, <https://doi.org/10.3390/e23060679>.
- , F. Xu, Y. Rong, R. Zhang, X. Tang, and F. Zhang, 2021: El Niño Modoki can be mostly predicted more than 10 years ahead of time. *Sci. Rep.*, **11**, 17860, <https://doi.org/10.1038/s41598-021-97111-y>.
- Lintner, B. R., and J. D. Neelin, 2009: Soil moisture impacts on convective margins. *J. Hydrometeorol.*, **10**, 1026–1039, <https://doi.org/10.1175/2009JHM1094.1>.
- Liu, G., Y. Wang, Y. Chen, X. Tong, Y. Wang, J. Xie, and X. Tang, 2022: Remotely monitoring vegetation productivity in two contrasting subtropical forest ecosystems using solar-induced chlorophyll fluorescence. *Remote Sens.*, **14**, 1328, <https://doi.org/10.3390/rs14061328>.
- Liu, L., L. Gudmundsson, M. Hauser, D. Qin, S. Li, and S. I. Seneviratne, 2020: Soil moisture dominates dryness stress on ecosystem production globally. *Nat. Commun.*, **11**, 4892, <https://doi.org/10.1038/s41467-020-18631-1>.
- Liu, Y., A. J. Parolari, M. Kumar, C.-W. Huang, G. G. Katul, and A. Porporato, 2017: Increasing atmospheric humidity and CO₂ concentration alleviate forest mortality risk. *Proc. Natl. Acad. Sci. USA*, **114**, 9918–9923, <https://doi.org/10.1073/pnas.1704811114>.
- Liu, Z. J., L. C. Wang, and S. S. Wang, 2014: Comparison of different GPP models in China using MODIS image and China-FLUX data. *Remote Sens.*, **6**, 10215–10231, <https://doi.org/10.3390/rs61010215>.
- Madani, N., and Coauthors, 2020: Recent amplified global gross primary productivity due to temperature increase is offset by reduced productivity due to water constraints. *AGU Adv.*, **1**, e2020AV000180, <https://doi.org/10.1029/2020AV000180>.
- Melnikova, I., and T. Sasai, 2020: Effects of anthropogenic activity on global terrestrial gross primary production. *J. Geophys. Res. Biogeosci.*, **125**, e2019JG005403, <https://doi.org/10.1029/2019JG005403>.
- Novick, K. A., and Coauthors, 2016: The increasing importance of atmospheric demand for ecosystem water and carbon fluxes. *Nat. Climate Change*, **6**, 1023–1027, <https://doi.org/10.1038/nclimate3114>.
- Oren, R., J. S. Sperry, G. G. Katul, D. E. Pataki, B. E. Ewers, N. Phillips, and K. V. R. Schäfer, 1999: Survey and synthesis of intra- and interspecific variation in stomatal sensitivity to vapour pressure deficit. *Plant Cell Environ.*, **22**, 1515–1526, <https://doi.org/10.1046/j.1365-3040.1999.00513.x>.
- Orth, R., and S. I. Seneviratne, 2017: Variability of soil moisture and sea surface temperatures similarly important for warm-season land climate in the Community Earth System Model. *J. Climate*, **30**, 2141–2162, <https://doi.org/10.1175/JCLI-D-15-0567.1>.
- Pal, S., T. R. Lee, and N. E. Clark, 2020: The 2019 Mississippi and Missouri River flooding and its impact on atmospheric boundary layer dynamics. *Geophys. Res. Lett.*, **47**, e2019GL086933, <https://doi.org/10.1029/2019GL086933>.
- Palacio, S., G. Hoch, A. Sala, C. Körner, and P. Millard, 2014: Does carbon storage limit tree growth? *New Phytol.*, **201**, 1096–1100, <https://doi.org/10.1111/nph.12602>.
- Papagiannopoulou, C., D. G. Miralles, W. A. Dorigo, N. E. C. Verhoest, M. Depoorter, and W. Waegeman, 2017a: Vegetation anomalies caused by antecedent precipitation in most of the world. *Environ. Res. Lett.*, **12**, 074016, <https://doi.org/10.1088/1748-9326/aa7145>.
- , —, S. Decubber, M. Demuzere, N. E. C. Verhoest, W. A. Dorigo, and W. Waegeman, 2017b: A non-linear Granger-causality framework to investigate climate–vegetation dynamics. *Geosci. Model Dev.*, **10**, 1945–1960, <https://doi.org/10.5194/gmd-10-1945-2017>.
- Právalle, R., 2018: Major perturbations in the Earth's forest ecosystems. Possible implications for global warming. *Earth Sci. Rev.*, **185**, 544–571, <https://doi.org/10.1016/j.earscirev.2018.06.010>.
- Punales, A. G., 2011: Time-varying coefficient models and the Kalman filter: Applications to hedge funds. Master's dissertation, Applied Mathematics, Ryerson University, 83 pp., <https://pdfs.semanticscholar.org/e68f/7bf222ccdf8b235a3a15a55738f1bd2d2a14.pdf>.
- Saini, R., G. Wang, and J. S. Pal, 2016: Role of soil moisture feedback in the development of extreme summer drought and flood in the United States. *J. Hydrometeorol.*, **17**, 2191–2207, <https://doi.org/10.1175/JHM-D-15-0168.1>.
- Schwingshackl, C., M. Hirschi, and S. I. Seneviratne, 2017: Quantifying spatiotemporal variations of soil moisture control on surface energy balance and near-surface air temperature. *J. Climate*, **30**, 7105–7124, <https://doi.org/10.1175/JCLI-D-16-0727.1>.
- Seneviratne, S. I., D. Lüthi, M. Litschi, and C. Schär, 2006: Land-atmosphere coupling and climate change in Europe. *Nature*, **443**, 205–209, <https://doi.org/10.1038/nature05095>.
- , T. Corti, E. L. Davin, M. Hirschi, E. B. Jaeger, I. Lehner, B. Orlowsky, and A. J. Teuling, 2010: Investigating soil moisture–climate interactions in a changing climate: A review. *Earth Sci. Rev.*, **99**, 125–161, <https://doi.org/10.1016/j.earscirev.2010.02.004>.
- Sperry, J. S., M. D. Venturas, W. R. L. Anderegg, M. Mencuccini, D. S. Mackay, Y. Wang, and D. M. Love, 2017: Predicting stomatal responses to the environment from the optimization of photosynthetic gain and hydraulic cost. *Plant Cell Environ.*, **40**, 816–830, <https://doi.org/10.1111/pce.12852>.

- Stips, A., D. Macias, C. Coughlan, E. Garcia-Gorriz, and X. San Liang, 2016: On the causal structure between CO₂ and global temperature. *Sci. Rep.*, **6**, 21691, <https://doi.org/10.1038/srep21691>.
- Stocker, B. D., J. Zscheischler, T. F. Keenan, I. C. Prentice, J. Penuelas, and S. I. Seneviratne, 2018: Quantifying soil moisture impacts on light use efficiency across biomes. *New Phytol.*, **218**, 1430–1449, <https://doi.org/10.1111/nph.15123>.
- Tao, L., X. S. Liang, L. Cai, J. Zhao, and M. Zhang, 2021: Relative contributions of global warming, AMO and IPO to the land precipitation variabilities since 1930s. *Climate Dyn.*, **56**, 2225–2243, <https://doi.org/10.1007/s00382-020-05584-w>.
- Tuttle, S., and G. Salvucci, 2016: Empirical evidence of contrasting soil moisture–precipitation feedbacks across the United States. *Science*, **352**, 825–828, <https://doi.org/10.1126/science.aaa7185>.
- Wang, M., S. Wang, J. Zhao, W. Ju, and Z. Hao, 2021: Global positive gross primary productivity extremes and climate contributions during 1982–2016. *Sci. Total Environ.*, **774**, 145703, <https://doi.org/10.1016/j.scitotenv.2021.145703>.
- Wang, X. H., and Coauthors, 2014: A two-fold increase of carbon cycle sensitivity to tropical temperature variations. *Nature*, **506**, 212–215, <https://doi.org/10.1038/nature12915>.
- Wright, S., 1921: Correlation and causation. *J. Agric. Res.*, **20**, 557–585.
- Wu, D., X. Zhao, S. Liang, T. Zhou, K. Huang, B. Tang, and W. Zhao, 2015: Time-lag effects of global vegetation responses to climate change. *Global Change Biol.*, **21**, 3520–3531, <https://doi.org/10.1111/gcb.12945>.
- Yi, B., and S. Bose, 2022: Quantum Liang information flow as causation quantifier. *Phys. Rev. Lett.*, **129**, 020501, <https://doi.org/10.1103/PhysRevLett.129.020501>.
- You, N., J. Meng, L. Zhu, S. Jiang, L. Zhu, F. Li, and L.-J. Kuo, 2020: Isolating the impacts of land use/cover change and climate change on the GPP in the Heihe River Basin of China. *J. Geophys. Res. Biogeosci.*, **125**, e2020JG005734, <https://doi.org/10.1029/2020JG005734>.
- Yuan, W., and Coauthors, 2019: Increased atmospheric vapor pressure deficit reduces global vegetation growth. *Sci. Adv.*, **5**, eaax1396, <https://doi.org/10.1126/sciadv.aax1396>.
- Zhang, F., Q. Quan, F. Ma, D. Tian, Q. Zhou, and S. Niu, 2019: Differential responses of ecosystem carbon flux components to experimental precipitation gradient in an alpine meadow. *Funct. Ecol.*, **33**, 889–900, <https://doi.org/10.1111/1365-2435.13300>.
- Zhang, W., Y. Li, X. Wu, Y. Chen, A. Chen, C. R. Schwalm, and J. S. Kimball, 2021: Divergent response of vegetation growth to soil water availability in dry and wet periods over central Asia. *J. Geophys. Res. Biogeosci.*, **126**, e2020JG005912, <https://doi.org/10.1029/2020JG005912>.
- Zhou, S., and Coauthors, 2019: Land–atmosphere feedbacks exacerbate concurrent soil drought and atmospheric aridity. *Proc. Natl. Acad. Sci. USA*, **116**, 18 848–18 853, <https://doi.org/10.1073/pnas.1904955116>.