Chaos
An Interdisciplinary Journal of Nonlinear Science

RESEARCH ARTICLE | AUGUST 14 2023

Causal interaction in high frequency turbulence at the biosphere–atmosphere interface: Structural behavior

Special Collection: Theory-informed and Data-driven Approaches to Advance Climate Sciences

Leila Constanza Hernandez Rodriguez ; Praveen Kumar

Chaos 33, 083139 (2023)
https://doi.org/10.1063/5.0131468

Chaos ARTICLE pubs.aip.org/aip/cha

Causal interaction in high frequency turbulence at the biosphere–atmosphere interface: Structural behavior

Cite as: Chaos **33**, 083139 (2023); doi: 10.1063/5.0131468
Submitted: 20 October 2022 · Accepted: 12 June 2023 ·
Published Online: 14 August 2023

View Online Export Citation CrossMark

Leila Constanza Hernandez Rodriguez[1,a] and Praveen Kumar[1,2,b]

AFFILIATIONS

[1] Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, Champaign, Illinois 61801, USA
[2] Department of Atmospheric Sciences, University of Illinois at Urbana-Champaign, Champaign, Illinois 61801, USA

Note: This article is part of the Focus Issue, Theory-informed and Data-driven Approaches to Advance Climate Sciences.
[a] **Present address:** Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA.
[b] **Author to whom correspondence should be addressed:** kumar1@illinois.edu

ABSTRACT

High-frequency (e.g., 10 Hz) eddy covariance measurements are typically used to estimate fluxes at the land–atmosphere interface at timescales of 15–60 min. These multivariate data contain information about the interdependency at high frequency between the interacting variables such as wind, humidity, temperature, and CO_2. We use data at 10 Hz from an eddy covariance instrument located at 25 m above agricultural land in the Midwestern US, which offers an opportunity to move beyond the traditional spectral analyses to explore causal dependency among variables. In this study, we quantify the structure of inter-dependencies of interacting variables at high frequency represented by a directed acyclic graph (DAG). We compare DAGs to investigate changes in structural differences in causal interactions. We then apply a distance-based classification and k-means clustering approach to identify the evolution of the causal structure represented by a DAG. Our method selects an unbiased number of clusters of similar structures and characterizes the similarities and differences between them. We explore a range of dynamic behavior using data from a clear sky day and during a solar eclipse in 2017. Our results show well-defined clusters of similar causal dependencies as the system evolves. Our approach provides a methodological framework to understand how causal dependence in turbulence manifests in high-frequency data when represented through a DAG.

At the biosphere–atmosphere interface, multiple hydrometeorological variables, such as wind speed, air temperature, H_2O, and CO_2, influence each other through non-linear interdependencies, where perturbations in one variable can influence the behavior of another variable. Using high frequency (i.e., 10 Hz) observations from a 25 m tall eddy covariance tower and an information theory framework for causal discovery, we infer the causal structure of interdependencies as directed acyclic graphs (DAGs). At the biosphere–atmosphere interface, daytime turbulence promotes land–atmosphere exchange and we hypothesize that different types of causal structures can arise over time. Therefore, we find patterns in the DAGs based on a distance-based classification that characterizes similarities among DAGs and a k-means clustering approach that selects an unbiased number of clusters of similar causal structures. For a clear sky day and during a solar eclipse in 2017, we explore the range of dynamics of the system under changes in incoming solar radiation, a major driver of hydrometeorological processes. Our results show that patterns of similar behavior arise and we observe three well-defined clusters of similar causal structure, mainly characterized by the influence of the thermodynamics on the scalar transport of CO_2 and H_2O and other particular behaviors. This approach allows the exploration of casual interactions and the detection of patterns in turbulence at the biosphere–atmosphere, as the propagation of fluctuations among variables carried by the observation evolves.

I. INTRODUCTION

At the biosphere–atmosphere interface, fast-time-scale turbulent processes are characterized by rapid interactions among variables, such as air temperature, wind speed, water vapor, and CO_2.

Eddy covariance towers, equipped with high-frequency analyzers operating at 10 or 20 Hz, offer a remarkable opportunity to investigate the interactions between turbulent fluctuations of CO_2 and H_2O molar concentrations, as well as temperature and wind velocity components.[1] These towers provide valuable insights into the dynamics at this interface and enable comprehensive exploration of complex relationships among variables. These observations are typically used to estimate the vertical turbulent exchange of momentum, energy, and gases between the biosphere and the atmosphere.[1,2] These high-frequency data have also been used for quality control, using spectral analyses to determine underlying periodicities in the behavior of the observed time-series.[1,3,4] While spectral approaches have provided significant insights into variations in single variables or co-variations among variables,[5–8] they do not reveal the structure of causal dependencies among them. Therefore, here we propose an approach for characterizing causal dependencies, which arise as fluctuations in one variable at any time instigate changes or variations in other linked variables at a future time.[9] Source variables affect response variables creating a causal structure of interdependence, which are often represented as a directed acyclic graph (DAG).[10] We use information flow among the interacting variables to identify patterns of causal structure using DAGs in turbulence using high-frequency data (10 Hz).

Previous efforts to explore complexity in the land–atmosphere interface include the use of time-series measurements of horizontal ocean surface-wind and vertical profiles of time-height winds to investigate emergent features, such as the one detected when scanning space-time images of turbulence.[11] Using measures based on information theory,[12] such as mutual information, evidence of regions of higher complexity have been identified where statistical models of turbulence perform poorly such as at the top of the boundary layer where transitions of stability regime occur.[11] While the location of strong and weak turbulence areas can be inferred, how the dynamic system composed of multiple variables self-organizes to produce such gradients, and consequently, the emergent patterns remains unknown. While a decay in mutual information in ocean surface 10 Hz wind data has been observed, which can be comparable to the occurrence of gravity waves,[11] drawing causal inferences requires the exploration of the dynamics of the multivariate system at the turbulence scale. Recently, Krich *et al.*[13] estimated process graphs using the PCMCI algorithm[14] to test if causal networks are considerably more informative than correlation networks. They confirmed the capacity of causal networks to extract time-lagged dependencies under realistic settings while cautioning against the interpretations depending on the method's assumptions and datasets used to infer causality. In our study, we move one step forward in the study of turbulence to explore the evolution of the dynamic multivariate interdependencies by characterizing the structure of causal interaction and its evolution.

We use the causal structure represented as DAGs to identify patterns of similar interaction and find the range of the variation in which they occur. One way to ensure that we understand the dynamic range of the configuration of causal structure is to perturb the system rapidly and see whether it exposes new modes of variations. In general, we assume that the biosphere–atmosphere system is an open but non-manipulable complex system. It is open given that it exchanges matter, energy, and information with its environment, while non-manipulable implies that it is not easy to intervene in the system to expose causal dependencies. However, a solar eclipse is a natural exogenous intervention on the whole system that allows us to explore a range of causal dependencies, albeit limited to the effect of the rapid decrease and recovery of the incoming solar radiation. Therefore, we explore the evolution of the system's causal structure to determine the range of causal structural behavior for a clear sky day and under the effect of a solar eclipse.

To determine the causal structure and its evolution, we use information flow to characterize temporally asymmetric interdependencies among variables.[15,16] That is, we use information theoretic metrics to identify how fluctuations in one variable instigate a response in a linked variable at a later time. The causal structure then represents the resulting network of interdependencies in the multivariate system that is characterized as DAGs.[16] Recent studies have shown that DAGs are effective in representing causal dependencies in such systems,[17–19] such as for similar interaction patterns and find the range of the variation in which they occur.[15,20–22] However, such multivariate interactions at the turbulence scale at the biosphere–atmosphere interface have not been explored before.

At high frequency, causal graphs in the form of DAGs have the potential to provide a novel understanding of turbulence associated with land–atmosphere exchange. In this study, we divide the high-frequency observations in a sequence of minimal-size windows to contain the sufficient length of data to allow for statistically robust inference of DAGs. Changes in the structure of the DAGs across these contiguous windows are then used to characterize the evolution of the causal structure. We use the approach developed by Runge to estimate the DAGs, where causal associations between variables can be visualized as links in a complex interaction network. Using a distance-based classification between DAGs,[23] we quantify structural differences between pairs of DAGs to identify variations in the evolving structures. In conjunction with a k-means clustering approach, we find patterns of similar causal structure. In this endeavor, information-theoretic metrics[12] play a significant role, offering a compelling approach to investigate causal dependencies between the interacting variables.

In particular, we ask: Are there different types of causal structures embedded in the daytime turbulence of the land–atmosphere exchange? To answer this question, we use the data for the mid-afternoon period for two days. The first day, August 18, 2017, is a clear sky day with a land-surface cover comprising a mosaic of corn and soybean crops at their peak productivity. The second day, August 21, 2017, corresponds to a solar eclipse with maximum obscurity of 94% at the tower location. The same time windows are chosen for both days and they cover the period from onset to clearance of the eclipse. We create DAGs for all 3-min non-overlapping but contiguous time periods for both days and use k-means clustering to find common patterns in the DAGs. This gives us the opportunity to capture and understand the different causal structures that manifest during this period, to see if a rapid decline in solar radiation results in remarkably different causal structures. This allows for assessing if the inferred causal structure is representative of broader dynamical characteristics.

This paper is organized as follows: In Sec. II, we briefly introduce the study site and summarize the basic concepts of information theory and DAGs. Also, we present our novel approach to estimating the evolution of the causal structure using DAGs. In Sec. III, we present the results of the evolution of the causal structure. Section IV discusses the findings on the causal structure of the dynamics of the system at high frequency. Finally, in Sec. V, we provide conclusions for the paper. Several factual and methodological details are included in Appendixes A–F. Appendix A contains a description of the instrumentation used in this study. Appendix C shows an example of how to compare a pair of DAGs using a distance metric. In Appendix D, we test the performance of the distance metric used for clustering. In Appendix E, we determine the number of edges in a fully connected DAG without contemporaneous edges. Finally, in Appendix F, we determine the cross-covariance of the time series used in the analyses to determine a wide range of feasible dependencies, both causal and correlative.

II. THEORETICAL PRELIMINARIES

A. High-frequency eddy covariance data

A 25 m tall eddy covariance flux tower is located in the intensively managed agricultural landscape of the Midwestern US, at 40.155N, 88.578W, Goose Creek Township, Piatt County, Illinois, USA[25,26] [Fig. 1(a)]. Instruments located at 25 m height of the tower measure fluctuations in turbulent motion at 10 Hz, resulting in a set of multivariate high-frequency time series. To measure the fluctuations in turbulent motion at the tower site, we used a Li-7500 Infrared Gas Analyzer manufactured by LiCor Inc. and a CSAT3 Sonic Anemometer by Campbell Scientific Ltd at 25 m height.[25,26] The incoming shortwave radiation (W/m^2) is measured at 15-min intervals [Fig. 1(d)] using a CMR4 Kipp and Zonen Pyranometer located at 10 m height on the same tower.[25,26] In Appendix A, Table I, we include a description of the instrumentation used in this study.

To determine the causal structure of interactions, we use the high-frequency time series of the horizontal wind speed, *WS* (m/s), vertical wind speed, U_z (m/s), air temperature, T (°C), water vapor, H_2O (g/m^3), and, carbon dioxide mass density, CO_2 (kg/m^3). We use the data from a clear sky day, August 18, 2017 [Fig. 1(b)], and from an eclipse day, August 21, 2017, when the tower was in the path of 94% of totality [Fig. 1(c)] and the maximum obscurity occurred at 12:19:53 (HH:MM:SS, U.S. Central Standard Time). We subdivided the data into 3-min non-overlapping contiguous windows. We chose this time interval based on the temporal resolution of our measurements. This interval ensured that we use minimally required data to obtain statistically reliable estimates of the information-theoretic metrics used to estimate the directed

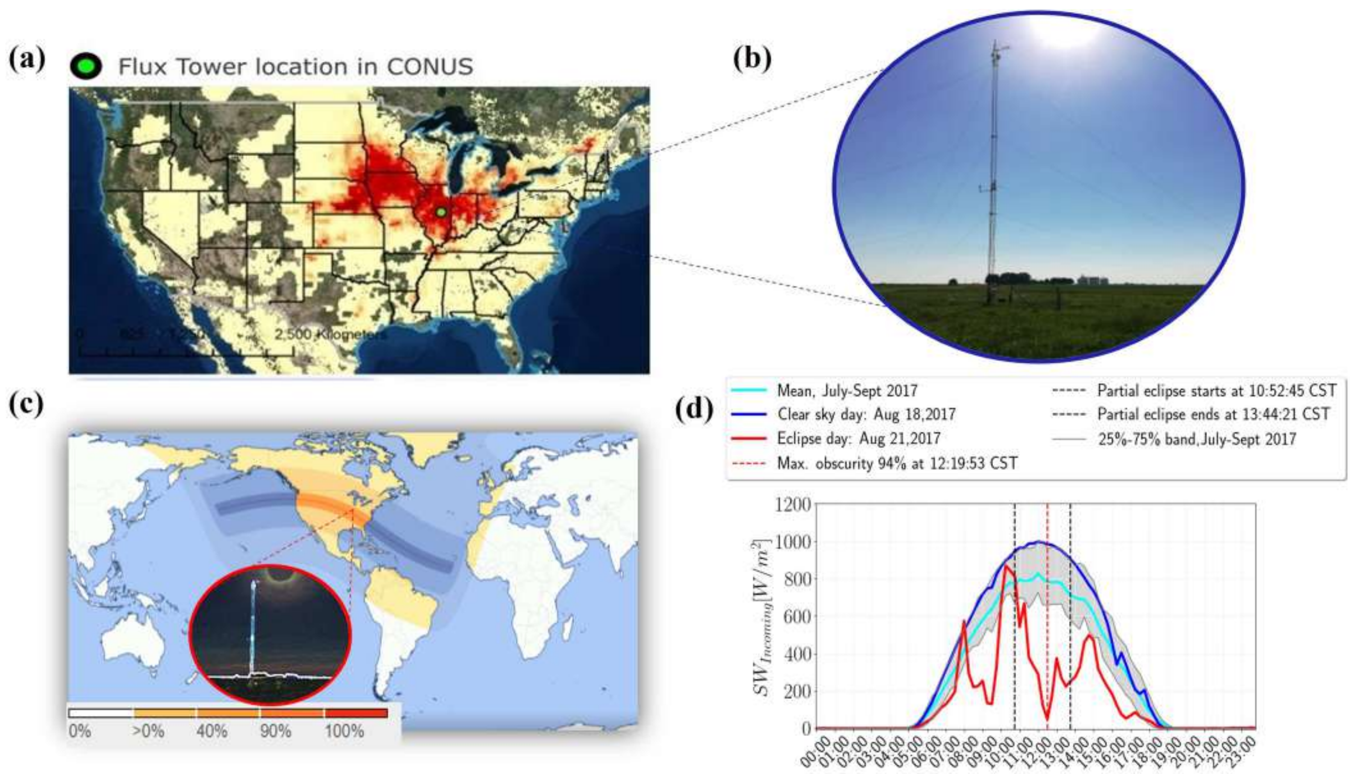

FIG. 1. Time series of high-frequency data from a 25 m tall eddy covariance flux tower offers a unique opportunity to investigate the interdependencies among components of an ecohydrological multivariate system at the biosphere–atmosphere interface. (a) The tower is located in the agricultural Midwestern US (green dot), a region intensively cultivated with corn (high = red, low = light ivory[24]) and other annual crops. (b) On August 18, 2017, the tower observed a mosaic of corn and soybean crops during a clear sky day. (c) On August 21, 2017, the tower was under the 94% band of the totality of a solar eclipse. (d) The attenuation of the solar radiation (red line) due to the eclipse and its contrast, with the clear sky day (blue line).

TABLE I. Hydroclimatic variables measured at Goose Creek flux tower site.

Instrument	Variables measured	Height from land-surface (m)
Gas Analyzer	Ecosystem densities of H_2O and CO_2	25
3D Sonic Anemometer	U_x,U_y, and U_z wind components	25
Temp and RH sensor	Air temperature and relative humidity	25
Pyranometer	Incoming and outgoing shortwave and longwave radiation	10

acyclic graphs (DAGs) (Sec. II B). Our choice led to 1800 data points per variable for the estimation and also we can use the variation between these non-overlapping windows to capture variations in the dynamics. The choice of when a time interval starts is driven by the 3-min interval at the time of maximum obscurity (i.e., 12:19:53 CST on August 21, 2017). The high-frequency data are standardized by subtracting the mean and dividing it by the standard deviation for each variable within the 3-min windows. All estimates are performed independently for each window. In Appendix B, we discuss the measurement error for each variable and the significance of the variation in the high-frequency data.

B. Directed acyclic graphs (DAGs) representation

A DAG for time series is a network structure denoted here as $G = (V, E)$, which consists of a set of vertices V and edges E, where an edge is directed from one vertex to another in a set of ordered pairs $(i, j) \in V \times V$. Alternatively, vertices and edges are called nodes and links, respectively. A DAG is "acyclic" in that the links never form a closed-loop.[10] Two vertices u and v are "connected" if G contains a path from u to v, otherwise, the vertices are "disconnected." If the vertices are additionally connected by a path of length 1, i.e., by a single edge, the vertices are called adjacent. In this study, we use DAGs for time series showing the observed variables, such as temperature, wind speed, and water vapor, as nodes while the links denote interactions among them at lagged times [Figs. 2(e) and 2(f)]. A DAG, which is a graphical representation of the underlying structural causal model,[28] reflects a near-instantaneous state of the dynamics of the system.[31,32]

The interactions are seen as a result of information flow arising when fluctuations in one variable instigate fluctuations in a linked variable.[28] Directed edges between variables in a DAG are interpreted as causal links where the methods used for their estimation measure the statistically significant magnitude of information flow and associated time delays, while at the same time excluding the influence of other indirect influences or common drivers.[28]

C. Identification of DAGs

Quantifying causality was first envisioned by Wiener[33] and later formulated by Granger, anchored on observational data and statistical dependency. For two-time series variables, Granger[34] characterized causality as the reduction in the variance of a target variable Z_t due to the knowledge of a lagged source variable $Y_{t-\tau}$, where $\tau > 0$. This approach accounts for the statistical dependency of one (or multiple) lagged source(s) on a target variable. It can be extended to consider the entire probability distribution using information-theoretic measures such as transfer entropy.[28,29]

In general, information-theoretic metrics provide avenues to quantify information flow among interacting variables.[32] In our study, we used the Tigramite package version 2.0 beta[27–30] for generating the directed acyclic graphs, a Python package that implements a causal discovery algorithm based on the PC algorithm[10] and many conditional independence tests. The choice of the specific conditional independence test is left to the user, and in our case, we use the non-parametric CMI-knn dependency measure for its effectiveness in high-dimensional and non-linear data settings.[35] CMI is estimated by combining the Kozachenko–Leonenko k-nearest-neighbor entropy estimator,[36] which is generalized for CMI,[37] with a nearest-neighbor local permutation scheme. The CMI-knn dependency measure is used to test for the conditional independence between all possible pairs of nodes and test all combinations among them, to identify the links of interactions.[28] The union of edges from the parents and links without including any contemporaneous links comprises the DAG.[28] Therefore, to estimate the joint density, we make no assumptions about the parametric form of the dependencies, and the CMI, $I(X; Y \mid Z)$, between X and Y given Z, is given by[35]

$$I(X; Y \mid Z) = \int p(z) \iint p(x, y \mid z) \log \frac{p(x, y \mid z)}{p(x \mid z) \cdot p(y \mid z)} dxdydz, \tag{1}$$

where p represents the appropriate univariate or joint PDF. This is combined with a shuffle test for statistical independence that is used to generate the distribution under a null hypothesis of statistical independence.[35] The *knn* estimator for the CMI is given by[35]

$$\widehat{I}(X; Y \mid Z) = \psi(k) + \frac{1}{T} \sum_{t=1}^{T} \left[\psi\left(k_{Z,t}\right) - \psi\left(k_{XZ,t}\right) - \psi\left(k_{YZ,t}\right) \right], \tag{2}$$

where ψ is the Digamma function[35] and k is the number of nearest-neighbors, T is the sample length, and $k_{Z,}, k_{XZ}, k_{YZ}$ are the number of neighbors in the respective subspaces.[35] Large k provides a low variance of the estimated CMI, which is important for the significance test.[37]

D. Clustering DAGs

To determine the different types of causal structures, we use clustering of DAGs to find patterns of similar behavior. In our study, each DAG represents the near-instantaneous (3-min average) state of the dynamics of the multivariate system. To compare a pair of DAGs, we use a distance metric proposed by Malmi *et al.*,[23] as briefly summarized here. The Kendall-tau distance $K(G_l, G_m)$ between a pair of DAGs, G_l and G_m, is the number of all concordant and discordant vertex pairs between them. A metric to compare DAGs based

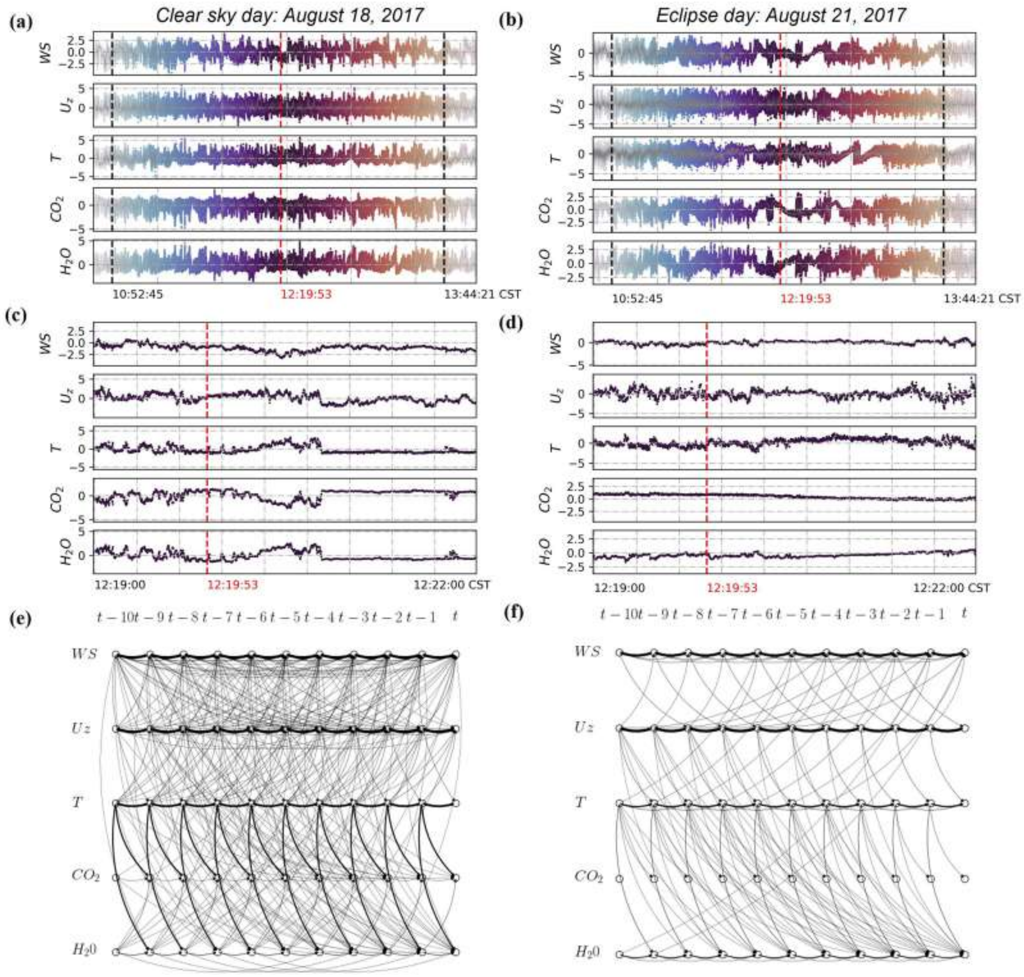

FIG. 2. Time series of standardized 10 Hz data and the directed acyclic graph representation, for five variables: horizontal wind speed (*WS*), vertical wind speed (U_z), air temperature (T), carbon dioxide mass density (CO_2), and water vapor mass density (H_2O). (a) We use data from 10:46:53 to 14:01:53 CST (U.S. Central Standard Time) on August 18, 2017, and (b) the eclipse day, August 21, 2017. The maximum obscurity (red dashed vertical line) occurs at 12:19:53 CST (darker colors indicating proximity to this time). The partial eclipse starts at 10:52:45 CST and ends at 13:44:21 CST (black dashed vertical lines). Times series of 3-min windows for (c) 12:19:00 to 12:22:00 CST for the clear sky day and (d) the eclipse day, the DAG [(e) and (f)] generated using Tigramite.[27–30]

on the configuration of vertices is possible given that all DAGs have the same vertices (nodes), and, therefore, a vertex (link) correspondence between any pair of DAGs is known. For a pair of DAGs, $G_l = (V, E_l)$ and $G_m = (V, E_m)$, two vertices of V are denoted as i and j, and $i < j$ if $i \in V$ precedes $j \in V$. As mentioned before, V is fixed and common between the two DAGs being compared. Two edges connecting the previous vertices are denoted as $e = (i, j)$ and $f = (j, i)$, where f is e with reversed direction. In Malmi *et al.*,[23] the distance measure for comparing DAGs uses parameters p and q to penalize pairs of items where one or both DAGs do not make a clear ordering decision. The values of p and q, where $0 \leq p$, $q \leq 1$, are chosen by the user for a desired balance between sensitivity and specificity. The distance between two DAGs is denoted by $K^{(p,q)}_{i,j}(G_l, G_m)$, which is the sum of penalties associated with vertices i and j, using the fixed parameters p and q, as follows:

$$K^{(p,q)}(G_l, G_m) = \sum_{\substack{(i,j)\in V\times V\\ i<j}} K^{(p,q)}_{i,j}(G_l, G_m). \tag{3}$$

In Appendix C, we provide a detailed description and an example for the estimation of the distance metric K when implementing the penalizations for DAGs for time series. From now on, we omit p and q from the notation by specifying $p = 1$ and $q = 0$ and write $K(G_l, G_m)$ instead of $K^{(p,q)}(G_l, G_m)$. We do not penalize using q, given that the DAGs obtained from our data are sparse and we do not expect to have pairs of identical empty (i.e., fully disconnected) or complete (i.e., fully connected) DAGs. Note that even though the definition of $K(G_l, G_m)$ is based on an ordering of vertices, the final result does not depend on that. To test the performance of the $K(G_l, G_m)$ metric, we evaluated two scenarios that are included in Appendix D.

In this study, we use the distance measure $K(G_l, G_m)$ between DAGs for clustering.[23] Given a set of DAGs, $G_l, \ldots, G_M$, and a candidate centroid $C = (V, E)$, we can estimate the cost of not having the edge (i, j) in the centroid, as $b(i, j)$,[23]

$$b(i,j) = \sum_{m=1}^{M} K_{ij}\left(G_m, G_{\text{empty}}\right), \tag{4}$$

where G_{empty} is a DAG with no edges. Similarly, $w(i, j)$ is the cost of having (i, j) as an edge in the centroid,

$$w(i,j) = \sum_{m=1}^{M} K_{ij}\left(G_m, G_{\text{full}}\right), \tag{5}$$

where G_{full} is a fully connected DAG without any contemporaneous links since they do not reflect causal relations. The estimation of G_{full} is included in Appendix E. This gives us[23]

$$\sum_{m=1}^{M} K(G_m, C) = \sum_{e \in E} w(e) + \sum_{e \notin E} b(e). \tag{6}$$

Now, consider $P = \{e \in V \times V \mid w(e) < b(e)\}$ as the set of all potential edges in the centroid. However, since P may contain cycles, we determine the subset of edges in the centroid E, $E \subseteq P$.[23] For each edge e not included in the centroid, we account for *regret*,[23] $r(e) = b(e) - w(e)$, which is a penalty for not having the edge in C. We can write the *regret* cost as[23]

$$\sum_{m=1}^{M} K(G_m, C) = \sum_{e \in P} w(e) + \sum_{e \notin P} b(e) + \sum_{e \in P \setminus E} r(e), \tag{7a}$$

$$\sum_{e \in P \setminus E} r(e) = \sum_{e \notin P} b(e) - \sum_{e \in P} w(e). \tag{7b}$$

To estimate the centroid of a set of DAGs, we minimize *regret* using a Greedy algorithm.[23] Briefly explained, in the Greedy algorithm, we order the edges based on *regret* cost, where edges with the smallest *regret* go first.[23] Note that we only need to consider edges that have appeared in at least one of the input DAGs, since $w(e) < b(e)$ cannot hold for the other edges. Then, we keep adding edges to the DAG denoted as centroid, in order, ignoring the edges that create cycles. In Malmi *et al.*,[23] Greedy algorithm's approach outperformed another algorithm for which there is a constant approximation guaranteed. Therefore, in our study, we use the Greedy algorithm to estimate the cluster's centroid. For further details in the Greedy approximation of DAG aggregation, we refer the reader to Malmi *et al.*[23]

Our purpose for clustering DAGs is to find groups of similar causal structures, which we consider representative of the near-instantaneous behavior of the system. For clustering DAGs, we use a k-means approach.[38] To update the cluster centroid in each iteration, we use the previously described Greedy algorithm. Given a set of M DAGs $G_l, \ldots, G_M$, we find k clusters $P_1, P_2, \ldots, P_k$, with centroids denoted by $C_1, C_2, \ldots, C_k$, and find[23]

$$\sum_{i=1}^{k} \sum_{G \in P_i} K^{(p,q)}(G, C_i) \tag{8}$$

is minimized.[23] Therefore, from a set of centroids, we group the input DAGs into clusters by minimizing the distance among them. After the clusters are formed, we find the centroid of each cluster. The initialization of the algorithm is random and updated after several restarts using the Greedy algorithm, which updates the centroids.[23] We run five restarts with random initial cluster assignments and select the run which minimizes the clustering cost. A challenge using a k-means type of approach for clustering is the random selection of the initial centroids, which have a significant impact on the results.[39] We select a seed to initialize the Mersenne Twister random number generator,[40] resulting in the same random numbers after each call to the random number generator, guaranteeing the reproducibility of the results. We use the nominal-k method[41] as seed, a rule of thumb based on the number of elements to be clustered, n, as follows:

$$s = \sqrt{n/2}. \tag{9}$$

While the selection of the seed for the Mersenne Twister algorithm used to shuffle is arbitrary, the seed only determines where we start in a very long list of pseudorandom numbers uncorrelated with each other, which is used as a generator.

E. Determining an unbiased number of clusters

One of the challenges of using k-means approach for clustering is the selection of the number of clusters, k, which can be subjective. In Malmi *et al.*,[23] k is an input parameter, which is set along with p and q. However, we require objective criteria for unbiased determination of k to explore causal inference from the system's connectivity. We achieve this by proposing a Dunn index[42] for DAGs and it is calculated as the ratio of the minimum of inter-cluster DAG distances and the maximum of intra-cluster DAG distance using the $K(G_l, G_m)$ metric. This index measures the compactness of the generated clusters by accounting for the maximum intra-cluster distance, that is, the maximum distance between DAGs and the centroid of a cluster. Also, the Dunn index for DAGs considers cluster separation using the inter-cluster distance, that is the minimum distance between clusters of DAGs. Let $\delta(C_i, C_j)$ be the inter-cluster distance, between clusters C_i and C_j, the distance between the clusters centroids. Therefore, the Dunn index (DI) for DAGs for m clusters are defined as follows:

$$DI_m = \frac{\min\limits_{1 \leqslant i < j \leqslant m} \delta(C_i, C_j)}{\max\limits_{1 \leqslant l \leqslant m} \Delta_l}, \tag{10}$$

where Δ_l refers to the intra-cluster distance for each cluster l.

A higher value of DI_m is associated with better clustering. The optimal number of clusters k is chosen as the one associated with the highest value. The Dunn index helps us to identify sets of clusters that are compact, with a small variance between members of the cluster, and well separated, where the means of different clusters are sufficiently far apart, compared to the within-cluster variance.

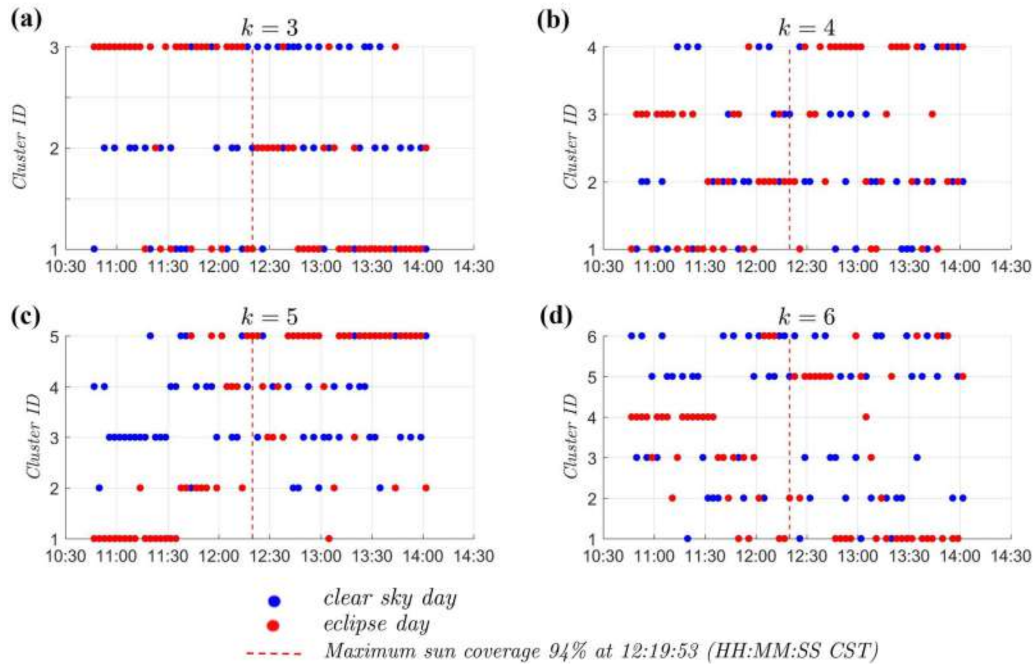

FIG. 3. Time-ordered distribution of clustered DAGs, for a clear sky (blue) and the eclipse (red) day, for different numbers of clusters, k: (a) $k = 3$, (b) $k = 4$, (c) $k = 5$, and (d) $k = 6$.

III. RESULTS

A. Causal structure and their evolution

We use a series of consecutive DAGs to explore the evolution of the causal structure that reflects the temporal interdependencies among the observed variables. We subdivide the 10 Hz multivariate time series from 10:46:53 CST to 14:01:53 CST (Central Standard Time) in non-overlapping contiguous 3-min windows, resulting in 66 windows of 1800 data points per variable for each window, for August 18 and 21, 2017. We estimate the DAG for each window using the Tigramite package version 2.0 beta,[27–30] which is a modified PC algorithm (i.e., Peter and Clark algorithm),[10] with the momentary information transfer (MIT) based conditional independence test to remove the spurious relationship between two nodes that is not causally related, and which may arise due to factors such as measurement error, sampling bias, or the influence of unobserved variables.[17] By doing so, we ensure that the DAG for each window accurately represents the causal relationships among the variables in that window. Each conditional independence test is performed based on 100 samples with a significance level of 0.05. Although only five members of an ensemble would lie at the tail of a distribution under this significance level, we are not making a statement about the absolute certainty of our results. Rather, we are using the significance level as a measure of evidence against the null hypothesis, which is that there is no causal relationship between the variables of interest. For the empirical estimation of all information-theoretic measures, we select the k-nearest neighbor (knn) method.[37,43] In this study, knn is used to compute CMI with the nearest-neighbor parameter $k = 100$. A high value of k reduces the variance of the estimated CMI;[37] however, it has been recommended that k be 5%–50% of the time series length.[17] The maximum Markovian time lag used for establishing the directed edges between lagged variables is set to 9. We selected this value based on the autocorrelation function of the eddy covariance measurements. This value allowed us to capture the dominant temporal dynamics of the system while avoiding overfitting the model. Our analysis focused on fast-time-scale land–atmosphere interactions, which typically occur within seconds to minutes. When each DAG is plotted [Figs. 2(e) and 2(f)], the thickness of each edge represents the coupling strength between two connected nodes, which is given by MIT.[28]

After we estimate the DAGs, we use their sparse adjacency matrices to analyze the evolution of the network structure. To obtain each adjacency matrix, we use the MIT weighted values between each pair of lagged nodes, which are connected when their MIT is larger than a significance threshold that is estimated using the shuffle test as described in Runge *et al.*[28] We extract the adjacency matrices from the DAGs using the NetworkX Python package.[44] After creating a group of adjacency matrices, we estimate the DAG clusters using the k-means algorithm described in Sec. II D. While clustering, each DAG is treated as an element inside a cluster of similar DAGs. Using the distance metric between DAGs, $K(G_l, G_m)$, we estimate different numbers of clusters, k, showing variability in the composition of the DAGs (Fig. 3, red and blue dots, respectively).

We observe that each cluster is a mix of the DAGs from both days. However, when the number of clusters increases, for instance for $k \geq 5$, we observe that a cluster is formed uniquely by DAGs of the early dynamics of the partial eclipse [Figs. 3(c) and 3(d), cluster #1 and #4, respectively]. However, since the selection of the optimum number of clusters, k, can be subjective, we use the Dunn index for DAGs [Eq. (10)] as a quantitative criterion to select the optimal number of clusters. The results from the Dunn Index [Fig. 4(c)] indicate that $k = 3$ offers the most parsimonious classification of the structure of the DAGs obtained from data on both days.

Does the causal structure change over time and are their patterns of connectivity statistically distinguishable? To answer this question, we use the clusters associated with $k = 3$ and we plot the time-ordered distribution of the DAGs for a couple of metrics and indexes (Fig. 5), as described below.

First, we estimate the distance $K(G_l, G_m)$ between each pair of consecutive DAGs in the sequence, i.e., G_{t-1} and G_t for all t [Fig. 5(a)]. Note that G_{t-1} and G_t refer to a DAG at a lagged time $t-1$ with respect to the current DAG at time t. In this case, a larger distance between the pair indicates a rapid change in structure. On the other hand, if the distance decreases between two consecutive DAGs, they become more similar to each other. If there is no change in the distance, the consecutive DAGs are identical in terms of their causal structure. We observe that clusters of DAGs with similar dynamics are identifiable thereby reflecting patterns of similar behavior (Fig. 5, cluster #1 to #3, respectively). In general, the distance between consecutive DAGs is smaller during the eclipse [Fig. 5(a)], meaning that the DAGs are more similar to one another in comparison to the DAGs on the clear sky day. Next, we account for the distance between each DAG and a hypothetical fully connected DAG [Fig. 5(b)]. In this case, a larger distance between them is an indicator of a sparser DAG. On the other hand, the smaller the distance, the denser the DAG is, and, therefore, the more similar to the fully connected DAG. We observed that, in general, the DAGs of the clear sky day have a smaller distance to the fully connected DAG than the ones on the eclipse day. Therefore, the DAGs of the clear sky day are, in general, denser and therefore, more similar to a fully connected DAG. Therefore, in general, they show more interdependencies among variables in comparison to the eclipse day.

To facilitate the recognition of the patterns, we propose a couple of indices to characterize the evolution of the causal structure of DAGs. We consider two types of evolutionary indices: the degree of relative change, D_r, and the degree of connectivity, D_c. For a sequence of consecutive DAGs, the degree of relative change, D_r, is estimated as the ratio between the distance of any two consecutive DAGs, G_{t-1} and G_t, and the number of edges, e, in the lagged DAG, G_{t-1}, as follows:

$$D_r = \frac{K(G_t, G_{t-1})}{e_{G_{t-1}}}. \tag{11}$$

Note that D_r is always positive. It is less than 1 when the distance of any two consecutive DAGs is larger than the number of edges, e, in the lagged DAG, G_{t-1}, and more than 1 otherwise. During the eclipse day [Fig. 5(c)], D_r is less than 1 near the time of the maximum obscurity and thereafter, which coincides with the formation of patterns of similar consecutive DAGs that do not change much in terms of their structure [Fig. 5(c)]. On the other hand, during a clear sky day, D_r is usually larger than 1. Therefore, the interdependencies shown by the DAGs of the clear sky day reflect a more dynamic behavior of the system than during the eclipse day.

The degree of connectivity, D_c, accounts for the ratio of the distance, $K(G_t, G_{\text{full}})$, estimated between each DAG from time series, G_t, and a hypothetical fully connected DAG, G_{full}, and the number

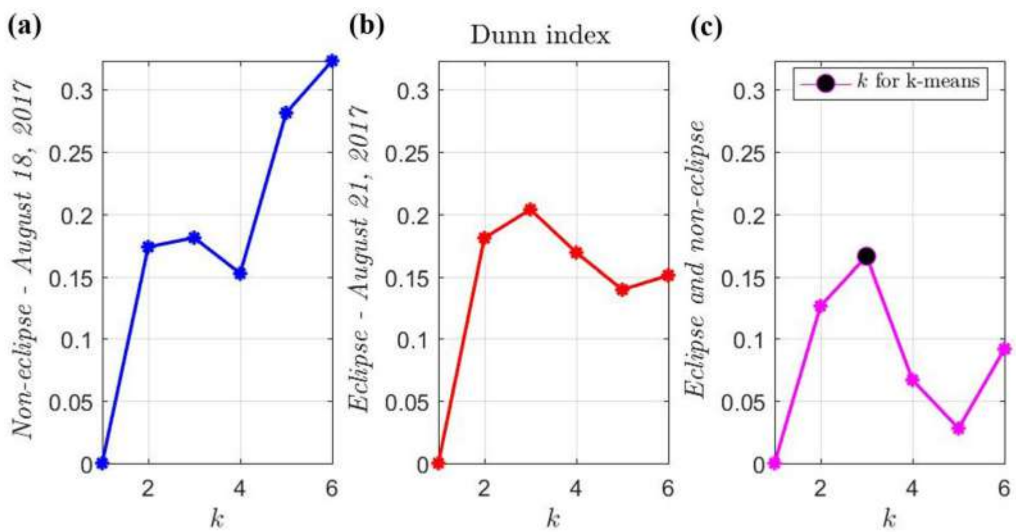

FIG. 4. We use the Dunn index for DAGs to determine the number of clusters, k, for (a) a clear sky day, August 18, 2017, (b) the eclipse day, August 21, 2017, and (c) the combined DAGs from both days. The selected number of clusters $k = 3$ (black filled dot) refers to the maximum Dunn Index when using all DAGs, which is also the maximum and the first local maximum in (b) and (a), respectively.

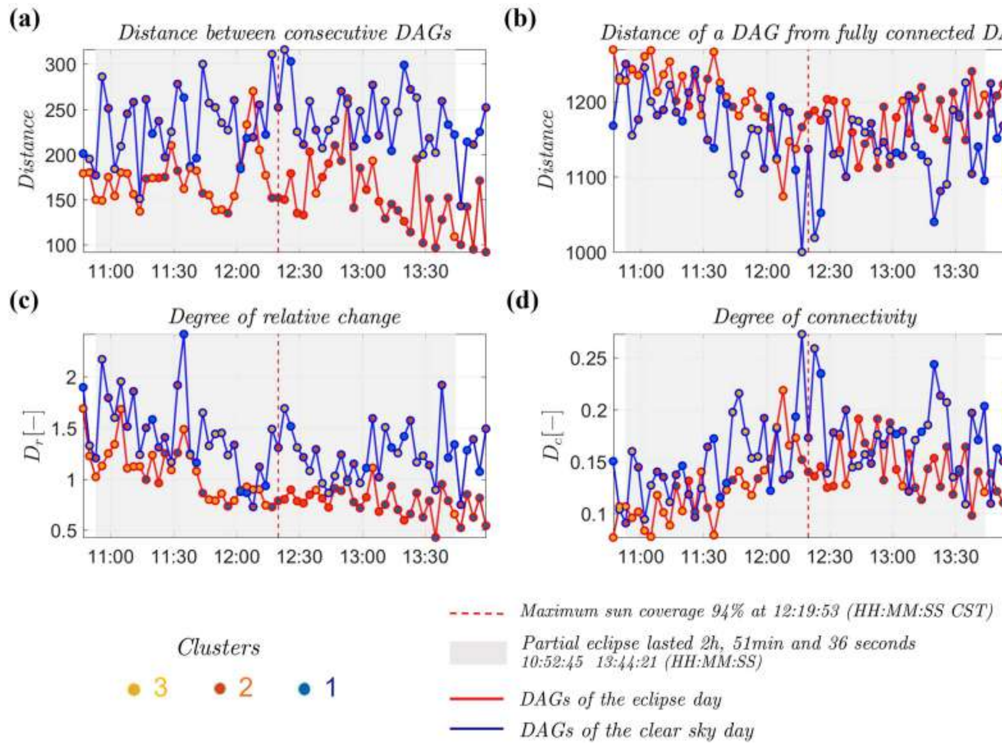

FIG. 5. Metrics and indices show the evolution of the causal structure of the system for the time-ordered clustered DAGs. The filled dots refer to DAGs in cluster #1 (yellow color), cluster #2 (red color), and cluster #3 (blue color). The lines connect DAGs of the same day, for the clear sky (blue) and eclipse day (red). The metrics are the (a) distance between two consecutive DAGs and the (b) distance between each DAG and a hypothetical fully connected DAG. The index refers to the (c) degree of relative change, D_r [Eq. (11)], and the (d) degree of connectivity, D_c [Eq. (12)].

of edges in the fully connected DAG, $e_{G_{\text{full}}}$, as follows:

$$D_c = 1 - \frac{K(G_t, G_{\text{full}})}{e_{G_{\text{full}}}}. \tag{12}$$

The estimation of the number of edges in the fully connected DAG is described in Appendix E. Note that D_c is always positive because the number of edges in G_{full} is always larger than the distance between any pair of DAGs formed by a sparse DAG and a fully connected DAG. Given that G_{full} is a constant, a larger D_c refers to a denser DAG. In other words, the closer D_c is to 1, the better connected the DAG. We observe that the DAGs are sparse, with D_c values ranging from 0.05 and 0.27, approximately. Furthermore, we observe that, in general, the connectivity of the DAGs is lower for the eclipse day when compared to that of a clear sky day [Fig. 5(d)]. Therefore, we find that the perturbation created by the solar eclipse causes a slight decrease in the connectivity associated with the dynamics of the turbulent system with respect to that during a clear sky day, where D_c is generally higher. Moreover, we observe that the degree of connectivity fluctuates within a range of variations, which can be slightly altered and recovered, even under the solar eclipse driven intervention on the whole system.

For the selected number of clusters, $k = 3$, each cluster's centroid can be seen as representing the average behavior within a cluster, which is estimated as described in Sec. II D. However, such centroids are synthetic DAG representations that are not associated with or derived from a time series of observations. For this reason, we look for the DAG from an observed time series that is closest to the centroid of each cluster for the representative behavior associated with the cluster. Moreover, the closest DAGs to the centroid of the clear sky and eclipse day (Fig. 6, blue and red dots above the closest DAG to the centroid, respectively), are useful to determine the dynamics associated within the same cluster.

Some measures of statistical dispersion for each cluster are also useful to characterize the configuration of the clusters (Fig. 6). For example, the average distance of all DAGs within a cluster to the centroid of that cluster is approximately the same (i.e., ≈125), although the number of DAGs inside each cluster varies. Cluster #3 has the largest number of DAGs within the cluster, followed by cluster #1 and #2. In addition, cluster #2 among the three clusters has one DAG from observations that is closest to the synthetic centroid. To plot the synthetic centroids (Fig. 7), we modified the Tigramite[17,27,28,45] plotting script where we use the resulting adjacency matrix of the centroid of each cluster to generate the synthetic DAG. Note that the thickness of the edges in the synthetic DAGs associated with the

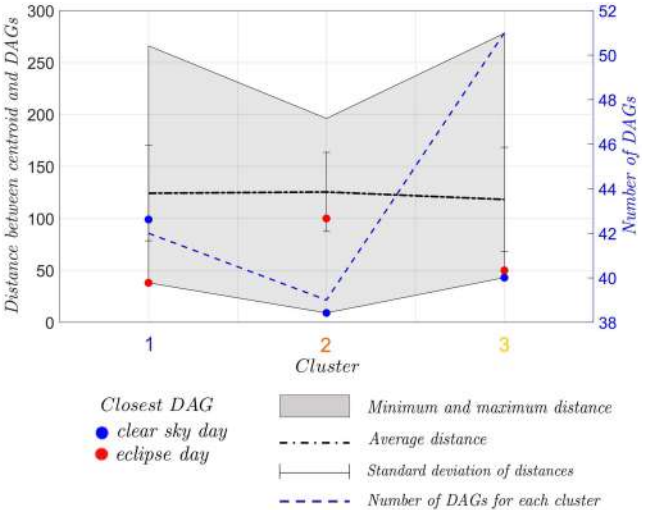

FIG. 6. Distance between the centroid of a cluster and the DAGs inside the cluster. For each cluster, the closest DAG to the centroid (i.e., minimum distance) can either be from a clear sky (blue) or the eclipse (red) day. The minimum and maximum distance between the centroid of a cluster are the closest and farthest DAG (i.e., shadow gray area), along with the average distance (i.e., black dashed-dotted line), the standard deviation (i.e., error bars), and the number of DAGs inside each cluster (i.e., blue dashed line), provide insights of the composition of each cluster that is formed from a combination of DAGs from both days.

centroids does not represent a differential weight on the connectivity between a pair of nodes, but merely the causal connectivity between them. The synthetic centroids [Figs. 7(a)–7(c)] show the main characteristics of the causal dependence among variables for each of the three clusters.

B. Interpreting changes in the causal structure

For the physical interpretation of the dynamics for each cluster, we note that the variables in the DAGs can be classified into three groups based on both their nature and role. First, the horizontal and vertical wind components, *WS* and U_z, reflect the dynamic character of the atmosphere. Second, the temperature, T, reflects the thermodynamic behavior. Third, CO_2 and H_2O are associated with scalar transport at the biosphere–atmosphere interface. Examining the behavior of interactions among these three different groups facilitates the interpretation of causal interactions. We characterize the differences among the clusters using the DAGs from the time series that is closest to the centroid that belongs to each cluster. However, to physically interpret the causal dynamics of each cluster, we need to identify the distinguishable features of each cluster. In the DAGs from time series, the strength of the connectivity between variables is shown by the thickness of the edges linking the source and target variables (Fig. 7), that is, thicker edges refer to stronger dependencies between the lagged stages of the variables.[28,29] The x axis in a DAG shows a lagged time with respect to a current time t expressed by $t - \tau$, where τ is a positive time lag equal to 0.1s. Therefore, each illustrated DAG in Fig. 7 shows the connectivity of the multivariate set for up to one second.

For cluster #3 [Figs. 7(d) and 7(g), for the eclipse and the clear sky day, respectively], the dynamics of the horizontal *WS*, and vertical wind speed, U_z, are largely causally self-dependent. However, the thermodynamics, T, has a weak causal influence on *WS*. However, T strongly influences the scalar transport of CO_2 and H_2O. That is, thermodynamics is the primary causal influence on scalar transport. Furthermore, CO_2 and H_2O also show causal dependence on each other at one-time step lag. This finding is consistent with the strong coupling of photosynthesis and transpiration during the period of high primary productivity. However, it is intriguing that this dependency is manifest in the data at high frequency in spite of the mixing through the large flux footprint associated with measurements at 25 m height.[25]

For cluster #1 [Figs. 7(f) and 7(i), for the eclipse and the clear sky day, respectively], we observe that the dynamics are similar to those as in cluster #3. Similarly, the dynamics of the horizontal *WS*, and vertical, U_z, wind speed are largely causally self-dependent. However, we observe that the vertical wind U_z has some influence on temperature, T, and in the horizontal wind, *WS*, while the thermodynamics, T, is strongly coupled to the scalar transport of H_2O and CO_2.

For cluster #2 [Figs. 7(e) and 7(h)], the dynamics of the horizontal *WS*, and vertical wind speed, U_z, are decoupled from the thermodynamics, T. Therefore, the scalar transport of CO_2 and H_2O depends only on the thermodynamics with weak coupling between them. An additional distinguishable feature in cluster #2 is the weak causal self-dependency of temperature T, which can lead to a complete self-decouple [Fig. 7(e)]. However, even if the temperature shows a causally self-decoupled behavior during the eclipse, it still sustains the dynamics of the scalar transport of H_2O and CO_2. Therefore, cluster #2 is mainly characterized by decoupling between the two main attributes controlling the dynamics of the whole system at turbulent scales: the dynamic influence of the atmosphere given by the wind, and the influence of the thermodynamics, which is strongly associated with the scalar transport.

In this causal inference analysis, we distinguished the predominant role of the influence of the thermodynamics from the influence of the dynamics of the wind on the scalar transport, for both H_2O and CO_2. While the general concept is that the vertical wind is what transports the scalars in the biosphere–atmosphere interface, we here observe that thermodynamics has a more relevant role for the given conditions. At 25 m height, the roughness effect from the surface is not a major control in the turbulent transport of scalars, given that our results suggest that the influence of the thermodynamics on the buoyant movement of scalars is more dominant. In addition, the influence of the wind on temperature was not observed, while the thermodynamics are more likely to occasionally weakly influence the horizontal and vertical wind speed (Fig. 7, cluster #3 and cluster#1, respectively).

We note that the results presented here in the context of considering causal dependencies are a subset of the dependencies that can be found when analyzing time series from observations. Specifically, causal dependencies are carried out by the propagation of fluctuations and are characterized by being temporally asymmetric, meaning that one variable influences the future of another variable asymmetrically in time.[15,16] Other dependencies, which are not causal, include contemporaneous correlative dependencies or those arising

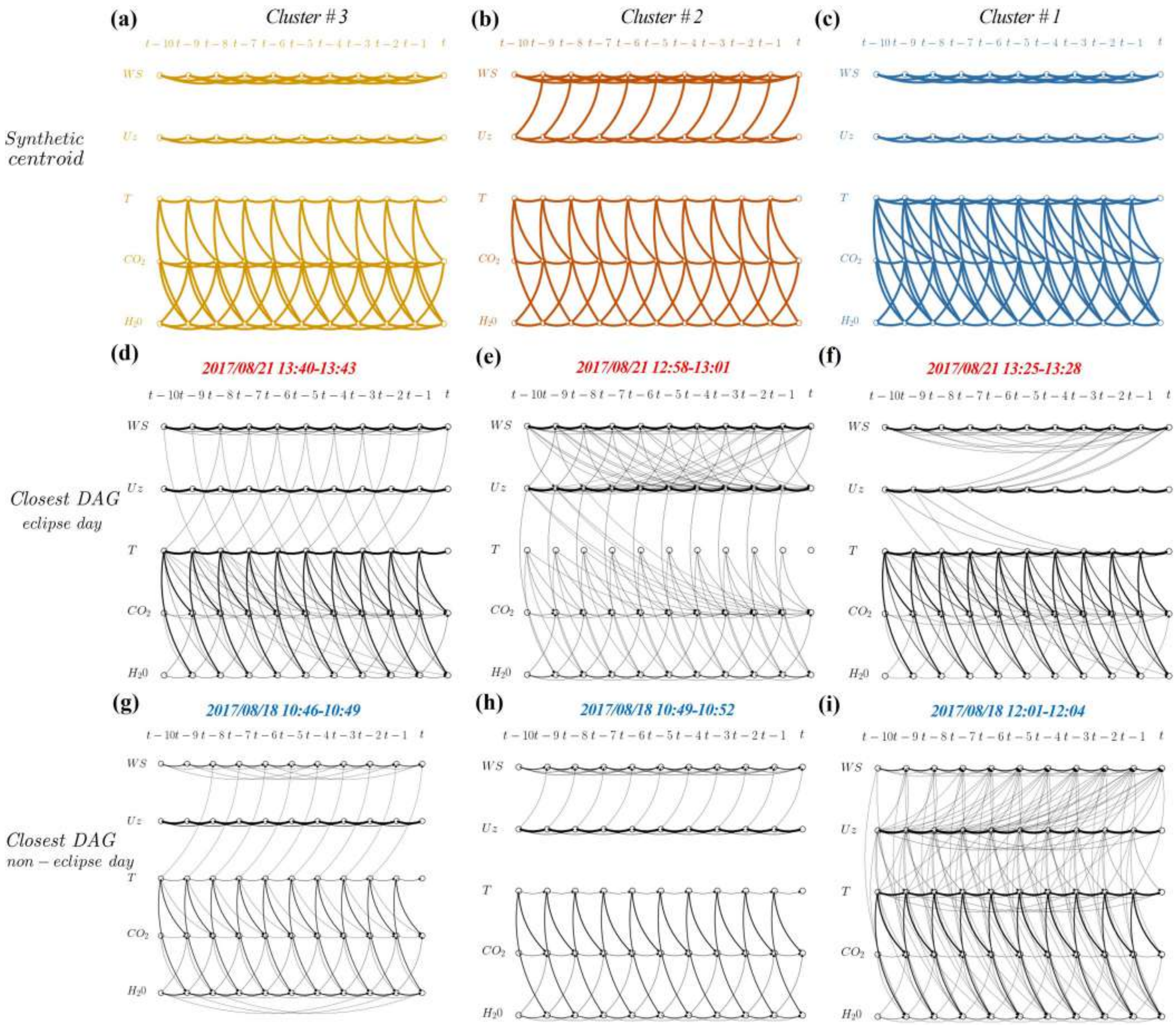

FIG. 7. Synthetic centroids and closest DAGs from time series to each cluster's centroid. The synthetic centroid shows the average of the dynamics of (a) cluster #3 (in yellow), (b) cluster #2 (in red), and (c) cluster #1, in blue. The closest DAGs from time series to each cluster's centroid for the eclipse (d)–(f) and the clear sky day (g)–(i).

due to a common driver. The first of these two are not included as they are deemed non-causal at the time resolution considered here, while the latter are minimized, but not completely eliminated, due to the Markovian assumption associated with DAG construction. For the cases where a variable seems to be self-decoupled in the causal analysis, such as in the case CO_2 and T [Fig. 8(a) and 8(b), respectively], the causal dependencies might be near-instantaneous, meaning that they occur before a 0.1s time lag or that the influence might come from a longer time lag than $\tau = 9$, causing the links to disappear from the causal graph. However, a lagged influence for $\tau > 9$ seems to be unlikely given the rapid nature of the interactions in the system. To further explore this self-decoupling behavior, we can alternatively use the time series to explore cross-covariance analysis to capture correlative dependencies (Figs. 15–21 in Appendix F). Particularly for CO_2 [Fig. 8(c)], the auto-covariance of CO_2 shows a strong instantaneous auto-correlation at time zero that very slowly dies out as the time lag increases. In the case of the self-decoupled T [Fig. 8(d)], we observe a very strong auto-correlation at time zero

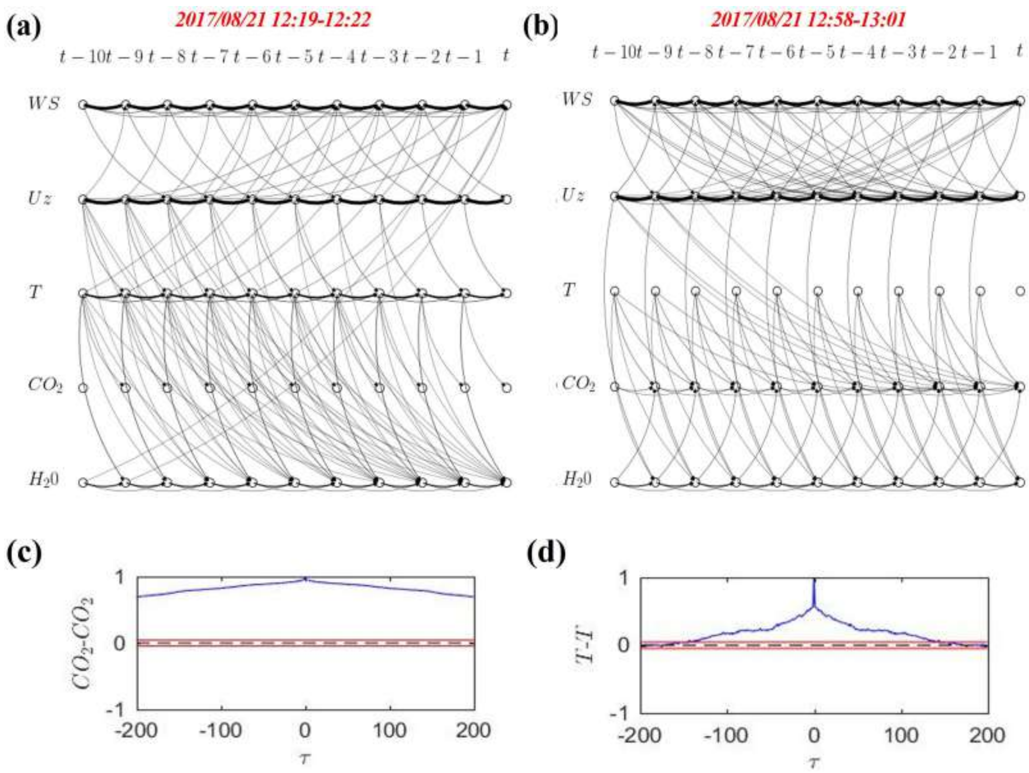

FIG. 8. DAGs that show the apparent absence of causal self-feedback on the dynamics of variables at turbulent scales. A unique case for (a) CO_2 at 12:19–=12:22 on August 21, 2017, and an example of the more frequent case of absence of self-feedback for (b) temperature, as observed at 12:58–13:01 on August 21, 2017. From a non-causal perspective, the auto-covariance of CO_2 [Fig. 8(c)] and T [Fig. 8(d)], shows a very strong auto-correlation at time zero, which dies out after a few seconds.

that rapidly dies out after a few seconds. Therefore, these results suggest that the self-decoupling behavior shown in the DAGs for CO_2 and T is associated with the near-instantaneous nature of the self-dependence.

In addition, we observed a particular behavior in the DAG in Fig. 8(a). In general, during the daytime, we expect the influence of the thermodynamics on the scalar transport of CO_2 and H_2O, as well as from the dynamics of the wind, both from the horizontal, *WS*, and vertical wind speed U_z. However, we would not expect the scalars to influence and modify the dynamics unless they have the ability to modify material properties. That is precisely what we observed in Fig. 8(a). During the eclipse day, our results suggest that the presence of water in a stagnating environment due to the effect of the solar eclipse might have changed the air density in a subtle way, resulting in the temporal lagged influence of H_2O on the horizontal wind speed, *WS*, which we did not expect to observe at this fine resolution.

C. Inferring the creation of patterns of similar behavior

After identifying the distinguishing attributes of each cluster (Sec. III B), we can now infer the behavior of the causal evolution of the system, as observed from the time-ordered distribution of the DAGs for the three clusters (Fig. 9). On the clear sky day [Fig. 9(a)], most of the dynamics first exhibit the behavior of cluster #2, characterized by the decoupling of the wind from the scalar fluxes, H_2O and CO_2, causally driven by thermodynamics. Over time, the coupling structure tends to evolve toward the behavior of cluster #3, characterized by a high, albeit weak connectivity within the variables, with a weak coupling of the wind to the dynamics of the scalar transport of H_2O and CO_2, and stronger control of the thermodynamics. Therefore, we can infer that during a clear sky day, the dynamics of the system tend to easily fluctuate between the states of decoupling and weak coupling [Fig. 9(a)].

On the other hand, during the eclipse day [Fig. 9(b)], many of the DAGs that correspond to the behavior of the cluster #3 occur before the maximum obscurity, characterized by a high albeit weak connectivity of the system, where the scalar transport of H_2O and CO_2 is mainly influenced by thermodynamics but also by the weak influence of the wind. As the solar eclipse reaches the maximum obscurity and immediately after, the system transitions to a cluster #2 behavior, characterized by the decoupling of the dynamics of the wind from the thermodynamics and the scalar transport of H_2O and CO_2. As the partial solar eclipse progresses, the system reaches a cluster #1 behavior, where the main characteristic is the recovery of the influence of vertical wind speed on H_2O and CO_2. Based on these results, we infer that the intervention on the dynamics of the system

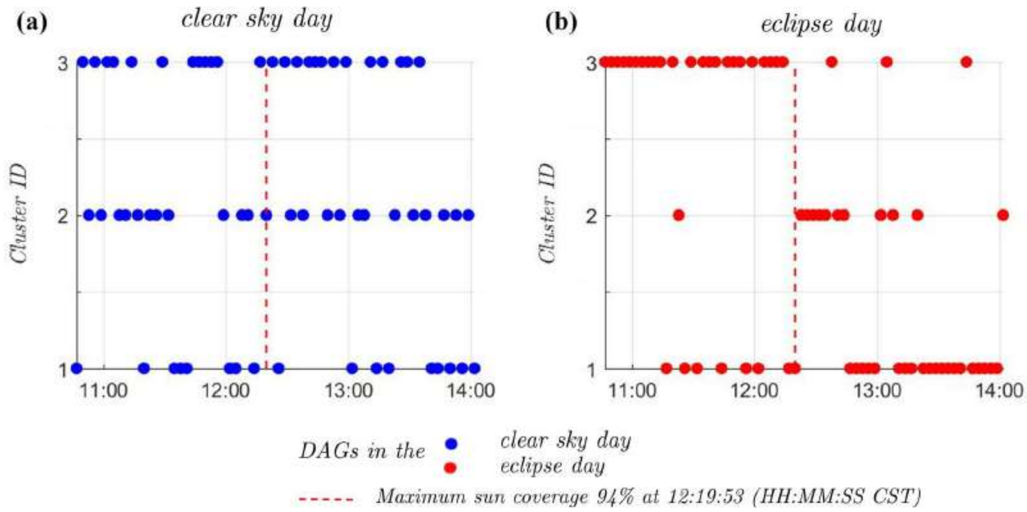

FIG. 9. Causal inference to determine the tendency of the behavior of the system over time. The time-ordered distribution of the clustered DAGs for $k = 3$ shows different patterns of behavior for the (a) clear sky (blue dots) and the (b) eclipse day (red solid dots).

created by the eclipse forced the system to behave differently than during a clear sky day, creating more consistent patterns of similar behavior over time [Fig. 9(b)].

IV. DISCUSSION

Turbulence at the biosphere–atmosphere interface creates a nonlinear system of multivariate interdependencies with a changing structure of coupling among its components. The rapid interactions among variables at high frequency, such as wind speed, air temperature, water vapor, and CO_2, create a network of complex interactions where information flows among variables at lagged times and defines its causal structure. Over time, these networks of causal dependencies propagate fluctuations that produce patterns of similar behavior at high frequency under the influence of a strong intervention of the dynamics of the system. The causal structure of a complex system encodes the evolution of its causal history, similar to how an organism's form is a functional manifestation of its ancestor's evolutionary and its own developmental memories.[46]

In this study, we consider three key aspects to help clarify the interpretation of our results. First, changes in the causal structure represented by the DAGs result from changes in the dependencies among the variables. Second, the interdependencies among the variables arise due to the propagation of fluctuations or variations among them. Third, causal dependencies are generally asymmetric, indicating that the strength of influence of a variable such as $T(t-1) \rightarrow H_2O(t)$ is typically not the same as $H_2O(t-1) \rightarrow T(t)$. An additional aspect pertains to the relevance of using data from a solar eclipse to study the evolution of causal structures in turbulence at the biosphere–atmosphere interface. We use this unique intervention as part of our analysis, not to highlight the behavior of eco-hydrological processes under this particular condition, but because during such a short period, the dynamics of the entire system are subject to the rapid decline and recovery of solar radiation.

Our results show the ability of our approach to determine the patterns of the inherent causal structure in turbulence. The centroids for each cluster of DAGs are representative of the behavior of the time-dependent nonlinear multivariate interactions. Our results show the emergence of patterns of persistent behavior (Fig. 5). Both during a clear sky day and during a solar eclipse, the dynamics maintain a certain order on the causal structure that exhibits particular attributes encapsulated in the cluster classification, whose effects persist for a period to create patterns. We observe that there is a certain stability in the causal structure of interactions that allow for reorganization as the atmospheric conditions change, even during the effect of a strong intervention from the eclipse. This leads us to postulate that while each variable can tell us about its own temporal characteristics [Figs. 2(c) and 2(d), respectively], the high-frequency observations reveal information about the influence of other variables, which together, sustain the system dynamics. The slight reduction in the connectivity of the multivariate system where the wind dynamics tend to decouple from the thermodynamics, and from scalar transport (CO_2 and H_2O) dynamics (Fig. 7, cluster #2), have a strong pattern that lasts longer in comparison to when the system experiences a larger interdependency (Fig. 7, cluster #3, in yellow, and cluster #1, in blue, respectively). Our results suggest that at turbulence scale, the system tends to fluctuate through self-organization in alternate states of high and low connectivity but bounded within a range. Here, we define this as the "system resilience to information flow," which reflects the ability of a system to recover when there has been a disruption or intervention in the typical behavior of the system to ensure the continuity of information flow.

V. CONCLUSIONS

We use the time series of turbulent observations at 10 Hz from a 25 m tall eddy covariance flux tower to explore the evolutionary dynamics of the system at turbulent scales. Non-linear interactions among components in a turbulent and multivariate complex system determine its causal structure, which evolves due to changes in information flow among component variables. The causal structure refers to a network representation that shows the connectivity of the information transfer among components of the system. We explore the evolution of the causal structure during a clear sky day and under the effects of a solar eclipse, a unique natural intervention into the dynamics.

Our approach to exploring the evolution of the causal structure using DAGs allows us to investigate the rapid changes in the interdependencies of turbulence at the biosphere–atmosphere interface. By using 10 Hz data, we have moved beyond the detection of contemporaneous dependencies observed in time-averaged datasets,[13] to explore the lagged influences resulting from self-feedback and cross-dependencies that shape the causal structure. The approach is based on information-theoretic metrics and a k-means type of clustering approach to classifying the DAGs in groups of similar and unique causal structures. When we estimate the evolution of the causal structure of the turbulence system under the influence of the rapid but progressive solar radiation limitation and recovery experienced during a solar eclipse, our results show the formation of behavioral patterns of similar causal structures. We find three well-defined clusters of structural connectivity in the system. Our results show that the multivariate system is likely to experience a reduction of structural connectivity, which is eventually recovered to a higher state of connectivity. We also find the peculiar cases of the decrease of self-feedback of its dynamics, such as for T and CO_2, which are addressed from methodological and phenomenological perspectives.

Our findings have broad implications for the understanding of turbulence systems. We acknowledge that dynamical equations as represented by Navier–Stokes equations have the potential to explain these behaviors, but they have not yet been fully exploited to do so. In the absence of that, we use a structural causal model to determine the patterns of similar dynamics of the multivariate system to explore its evolution using a sequence of DAGs. Each DAG represents the dynamic state of the connectivity of the multivariate system at consecutive times on the overall causal trajectory of the system. Such behavior emerges from the nonlinearities within the system, where the interdependencies among components of the system change over time through self-organization. We observed that the connectivity of the system fluctuates within a stable range of variations, allowing us to define patterns of evolution of its dynamics. Inferring causality to draw conclusions from observations requires some assumptions,[10,47] such as the Markov assumption, faithfulness of the DAGs to represent the dynamics of the system, causal sufficiency, causal stationarity inside each 3-min window, and no contemporaneous causal effects. We acknowledge that the results we obtained from a single tower may not be representative of all ecosystems. However, we selected this tower based on its comprehensive and high-frequency measurements of land–atmosphere interactions, which were suitable for testing the effectiveness of our proposed approach. We also acknowledge that it is essential to investigate the sensitivity of our method to different locations, which we plan to address in our future work by comparing our results from different eddy covariance towers located in various ecosystems. Given that our analysis focused on fast-time-scale land–atmosphere interactions, we acknowledge that other characteristic time scales may also be relevant for other types of turbulent systems, and we plan to explore other scales and atmospheric conditions in our future work. We anticipate that the study of the causal structure of the dynamics of the connectivity of complex systems at high frequency will contribute to the exploration of the dynamics of the information flow for more processes at the turbulence scale and over longer time periods to inform Earth system modeling efforts and prediction.

ACKNOWLEDGMENTS

This research was funded by the NSF (Grant No. EAR-1331906) for the Critical Zone Observatory for Intensively Managed Landscapes *(IMLCZO)*, the NSF (Grant No. OAC-1835834), the NSF (Grant No. EAR-2012850) for the Critical Interface Network for Intensively Managed Landscapes *(CINet)*, and the ARPA-E (Grant No. DE-AR0001225). We thank Peishi Jiang at Pacific Northwest National Laboratory (PNNL) for his support when using the Tigramite software package.[27–30] We also thank Steve Sargent at the Illinois State Geological Survey for the collection of the high-frequency data from the flux tower at Goose Creek.

NOMENCLATURE

CMI	conditional mutual information
CMI*knn*	conditional mutual information based on *knn* estimation
CO_2	carbon dioxide molar concentration (kg/m^3)
CST	Central Standard Time
DAG	directed acyclic graph
DI	Dunn Index
EC	eddy covariance
H_2O	water vapor molar concentration (g/m^3)
HH:MM:SS	hour:minute:second
IT	information theory
knn	k-nearest-neighbor
MIT	momentary information transfer
MIWTR	momentary information weighted transitive reduction
T	air temperature (° C)
TE	transfer entropy
TR	transitive reduction
US	United States of America
U_z	vertical wind speed (m)
WS	horizontal wind speed (m)
WTR	weighted transitive reduction

AUTHOR DECLARATIONS

Conflict of Interest

The authorshave no conflicts to disclose.

Author Contributions

All authors contributed equally to this work.

Leila Constanza Hernandez Rodriguez: Conceptualization (equal); Methodology (equal); Writing – original draft (lead); Writing – review & editing (equal). **Praveen Kumar:** Conceptualization (equal); Methodology (equal); Supervision (lead); Writing – review & editing (equal).

DATA AVAILABILITY

The directed acyclic graphs for time series are estimated using the Tigramite software package version 2.0 beta.[27–30] The distance-based classification between DAGs is estimated using ekQ/dag-comparisons codes https://github.com/ekQ/dag-comparisons.[23] The codes for the evolution of the Causal Structure are available on GitHub, as well as the data used in this study https://github.com/HydroComplexity/CausalStructureEvolution.

APPENDIX A: HYDROCLIMATIC VARIABLES MEASURED AT GOOSE CREEK FLUX TOWER SITE

More than 20 instruments are powered by an off-grid solar panel system and are located in the subsurface, near-surface, and at 10 and 25 m above the surface. Table I describes the instrument name, variables measured, and location and height of the tower for the instruments that measured the high-frequency data (i.e., 10 Hz or 0.1 s). In this study, the high-frequency data are standardized by subtracting the mean and dividing by the standard deviation for each of the 3-min windows. The high-frequency data are stored at the tower in a data logger and is manually collected every month at the tower since April 2016. The data are then uploaded to the *Clowder* web-based data system[48] of the Critical Interface Network for Intensively Managed Landscapes project

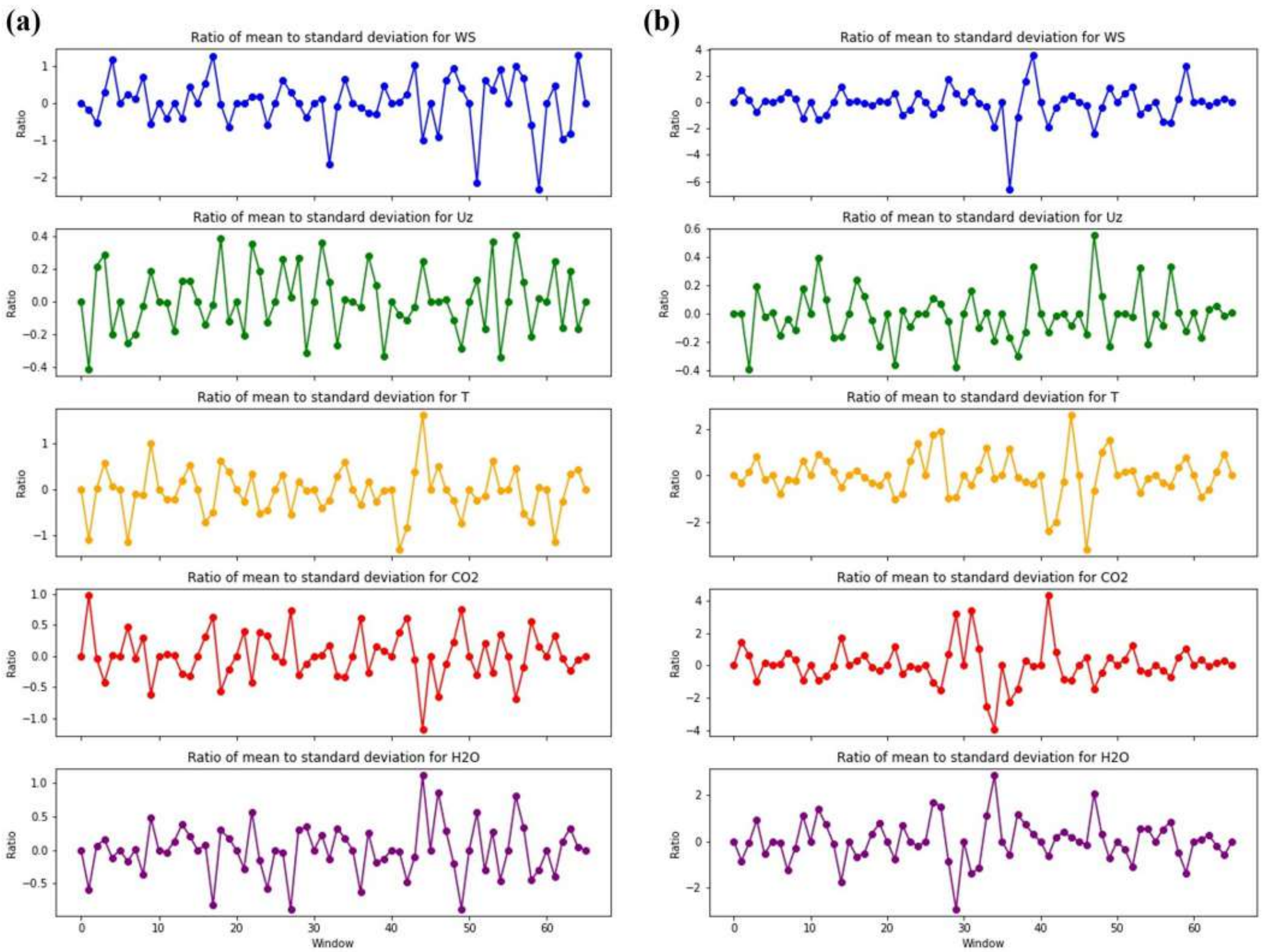

FIG. 10. Ratio between the mean and standard deviation for each variable across 66 windows of analysis for (a) August 18 and (b) August 21, 2017.

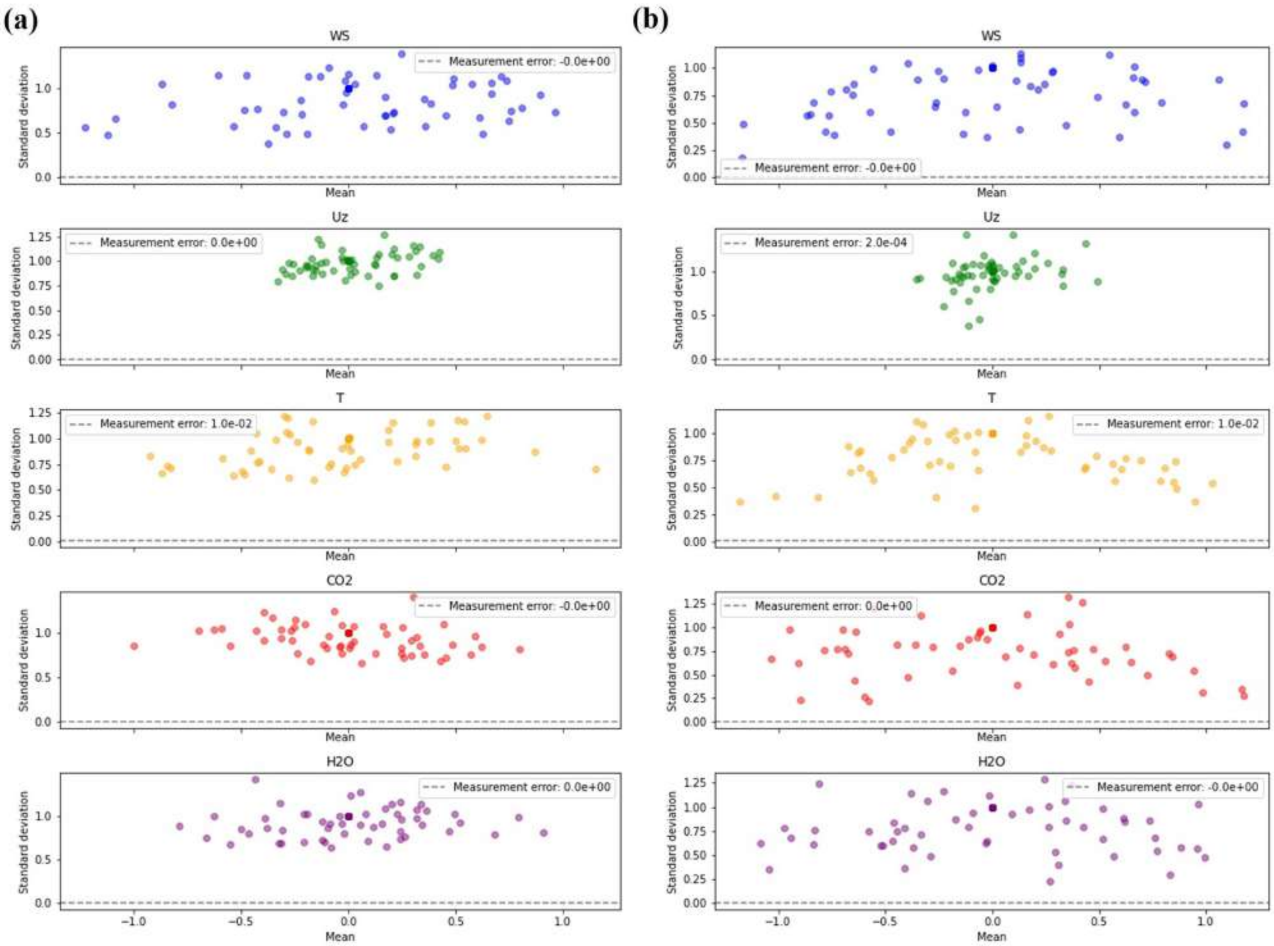

FIG. 11. Comparison of variable standard deviation with expected measurement error across 66 windows of analysis for (a) August 18 and (b) August 21, 2017. Each point represents a single window, the x-coordinate refers to the variable's mean, and the y-coordinate is the variable's standard deviation. The measurement error for each variable is shown as a horizontal line at the level of the measurement error.

(CINet) https://cinet.ncsa.illinois.edu/. For more information about instrumentation and data at this flux tower, as well as other instrumented nearby locations, we refer the reader to Wilson *et al.*[26]

APPENDIX B: MEASUREMENT ERROR AND VARIABILITY IN HIGH-FREQUENCY DATA

Here, we discuss the measurement error for each variable and the significance of the observed variation in the data. We estimate the average ratio between the mean and standard deviation for each variable, for the 66 windows on each day of analysis, August 18 and 21, 2017 [Figs. 10(a) and 10(b), respectively]. This ratio provides an insight into the variability present in the data. A high ratio suggests that there is a significant amount of variation, while a low ratio indicates that the data are more consistent. Additionally, by comparing the standard deviation to the measurement error for each variable on August 18 and 21, 2017 [Figs. 11(a) and 11(b), respectively], we determine if the observed variability is within the expected range of measurement error or is likely due to other factors. We find the expected measurement error for each variable by consulting the documentation of the measurement equipment. The eddy covariance system consists of a CSAT3 three-dimensional sonic anemometer,[49] an LI-7500 open path infrared gas analyzer (IRGA),[50] and a 083E-L temperature and humidity probe.[51] The measurement error for an LI-7500 open path infrared gas analyzer (IRGA) is within ± % of the measured value for CO_2 and H_2O concentration measurements. For the 083E probe, the temperature accuracy is $\pm 0.1\,^{\circ}$C from -50 to $50\,^{\circ}$C. For the CSAT3 anemometer, the measurement error for when the wind vector is within /$\pm 5^{\circ}$ of horizontal is less than $\pm 2\%$ of the measurements. From this analysis, we observe in Fig. 11 that the standard deviation lies outside the measurement error for all variables, suggesting that the variation observed is not just due to measurement error and is significant.

APPENDIX C: AN ILLUSTRATIVE EXAMPLE OF USING THE $K(G_l, G_m)$ METRIC

Here, we provide a detailed description of the cases of penalization and an example for the estimation of the distance metric, $K(G_l, G_m)$, used to compare pairs of DAGs proposed by Malmi *et al.*[23] [Eq. (3)]. The distance metric considers four cases of penalization:

1. $(e \in E_l$ and $e \in E_m)$ or $(f \in E_l$ and $f \in E_m)$, meaning that if both DAGs, G_l and G_m, agree on e, then the distance between the DAGs is zero: $K_{ij}(G_l, G_m) = 0$.
2. $(e \in E_l$ and $f \in E_m)$ or $(e \in E_m$ and $f \in E_l)$, that is if the DAGs G_l and G_m completely disagree on e, then the distance is equal to 1, $K_{ij}(G_l, G_m) = 1$. This case cannot occur for DAGs from time series, as shown with an illustrative example in Appendix C.
3. $(e$ or $f \in E_l$ and $e, f \notin E_m)$ or $(e$ or $f \in E_m$ and $e, f \notin E_l)$, that is an edge between i and j exists in one graph but not in the other. In this case, we set $K_{ij}(G_l, G_m) = p$
4. $e, f \notin E_l$ and $e, f \notin E_m$, that is, there is no edge between i and j in either of the graphs. We set $K_{ij}(G_l, G_m) = q$. Therefore, $K(G, G) = 0$, however, this property does not hold for $q > 0$ as penalty q is induced by edges that are missing from both input graphs. The motivation for setting $q > 0$ is that otherwise, two empty DAGs are as similar as two identical complete DAGs even though one might argue that the latter two have much more in common than the empty DAGs.

To illustrate the distance metric, we consider a pair of toy DAGs, $G_l = (V, E1)$ and $G_m = (V, E2)$ (Fig. 12). As the two DAGs are not identical and have different sets of edges despite sharing the same nodes, we expect the distance K obtained using Eq. (3) to be greater than zero.

To compute the distance between the toy DAGs shown in Fig. 12, we start counting for the number of concordant edges between them. One concordant edge, namely, (a) and (b), connects the same nodes in both DAGs, then K = 0. Next, there is one discordant pair, (c) and (d) in G_l and (d) and (c) in G_m as defined by Malmi *et al.*,[23] which cannot exist in the DAGs for time series of instantaneous data given that only the past can have a causal influence on the future. One pair that is not in the other graph, namely, (a) and (d), results in K = 1, only if p in Eq. (3) is $p = 1$. Now, we look for feasible edges that are missed in both graphs. Given that only non-contemporaneous edges that represent lagged in time influence can be missed, there is only one edge missed in both graphs which is (c) and (b). The rest of the missed edges are contemporaneous edges and we do not account for them in DAGs from time series. Therefore, if $q = 1$ in Eq. (3), then K = 1. The final distance between the pair of toy DAGs is the sum of all previous distances, resulting in K = 2. Notice that to avoid the risk of accounting for the nonexistence of contemporaneous links in both DAGs, a more conservative view is to set $q = 0$. In that case, the difference between the two DAGs is K = 1.

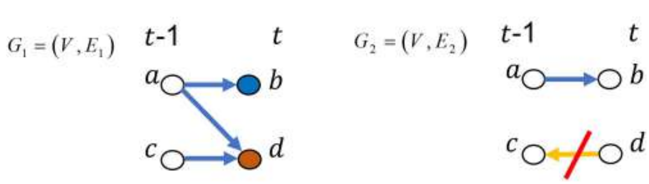

FIG. 12. A pair of toy DAGs related to the example in Appendix C, based on Example 1 in Malmi *et al.*[23] when implemented for DAGs for time series.

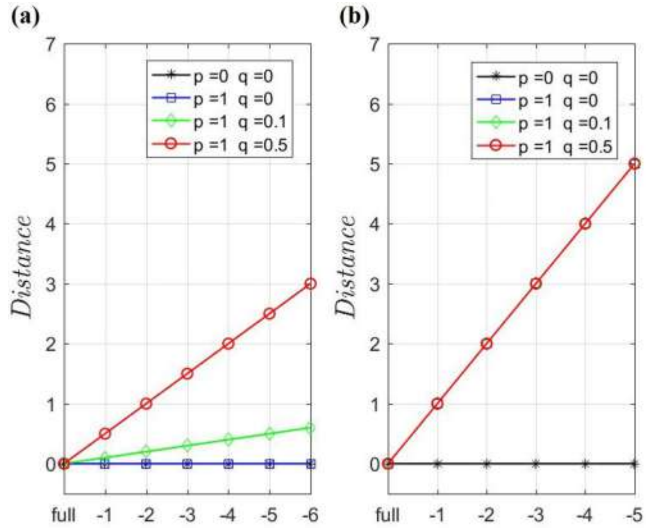

FIG. 13. Testing the distance measure for pairs of (a) identical toy DAGs, $K(G, G)$, while simultaneously removing the same edges in both DAGs, and (b) a fully connected DAG and a DAG which edges are removed increasingly, $K(H_{full}, G)$. We tested four different combinations of p and q values.

APPENDIX D: TESTING THE PERFORMANCE OF THE $K(G_l, G_m)$ METRIC

To test the performance of the $K(G_l, G_m)$ metric, we evaluated two scenarios. First, we estimated the distance between identical toy

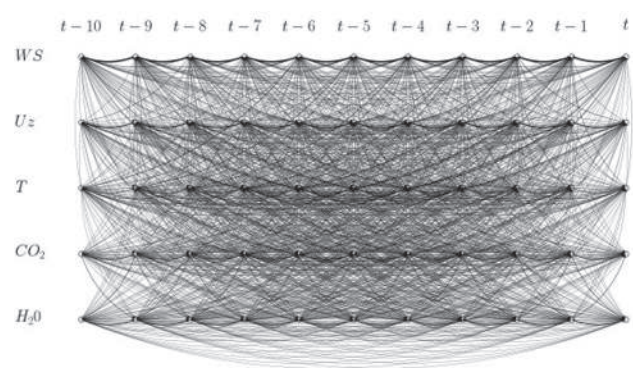

FIG. 14. Fully connected DAG from turbulent time-series, for 5 variables (y axis) and 11 time steps.

DAGs, $K(G, G)$, for different values of p and q [Fig. 13(a)]. In this case, p can take any value without changing the distance, since the same link is removed from both DAGs. On the other hand, q adds to the distance for each link that is removed in both DAGs. In the second scenario, we calculated the distance between a fully connected DAG, G_{full}, and a DAG while its edges have been removed increasingly, for different values of p and q [Fig. 13(b)]. p accounts for the difference in edges between the fully connected DAG and the partially connected DAG, increasing the distance as a linear function of the number of different edges between the DAGs. On the other hand, q is not relevant since the same edge is not being removed from the DAGs simultaneously.

APPENDIX E: ESTIMATING G_{full} WITHOUT CONTEMPORANEOUS LINKS

An important characteristic of the DAGs from time series is that they do not have any contemporaneous links. The reason is that a variable at the current time can only be influenced by the past, whether by itself at a lagged time, by the previous state of another variable, or by its whole causal history. Therefore, when estimating a DAG from a set of turbulent time series, an edge runs from the vertex i to j, only if $i < j$ in time. Then, the maximum number of edges in a fully connected DAG with no contemporaneous links [i.e., $e_{G_{\text{full}}}$ in Eq. (12)], is determined by

$$e_{G_{\text{full}}} = n\frac{m(m-1)}{2} + n(n-1)\frac{m(m-1)}{2}, \tag{E1}$$

where n refers to the number of variables in the DAG, while the number of time step evaluated, including the current time, are denoted by m. An example of a fully connected DAG for $n = 5$ and $m = 11$, is shown in Fig. 14.

APPENDIX F: CROSS-COVARIANCE OF THE TIME SERIES IN THE DAGS OF THE CLUSTER'S CENTROIDS

Here, we determine the cross-covariance of the discrete time series corresponding to the centroids of the clusters shown in Fig. 7. In addition, we include the cross-covariance plots for the time series of 2017/08/21 12:19-12:22 CST (Central Standard Time) to determine the behavior of the self-dependence of CO_2, which is not observed in the causal analysis. The cross-covariance measures the similarity between a time series and a shifted (lagged) time series as a function of the lag. For any two variables X and Y, if the cross-covariance is positive, then larger values of X are associated with larger values of Y and smaller values of X are associated with smaller values of Y. If the covariance is negative, the opposite holds: small values of X are associated with larger values of Y and vice versa. We use the cross-covariance here to determine the time lags for which a pair of time series have any type of association, causal or not, as a function of time lag, τ. We use the generated plots to complement our causal analysis when using DAGs for causal inference in Sec. III B.

We estimate the cross-covariance up to a time lag $\tau = 200$, which corresponds to 20-s for time series of 10 Hz (Figs. 15–21,

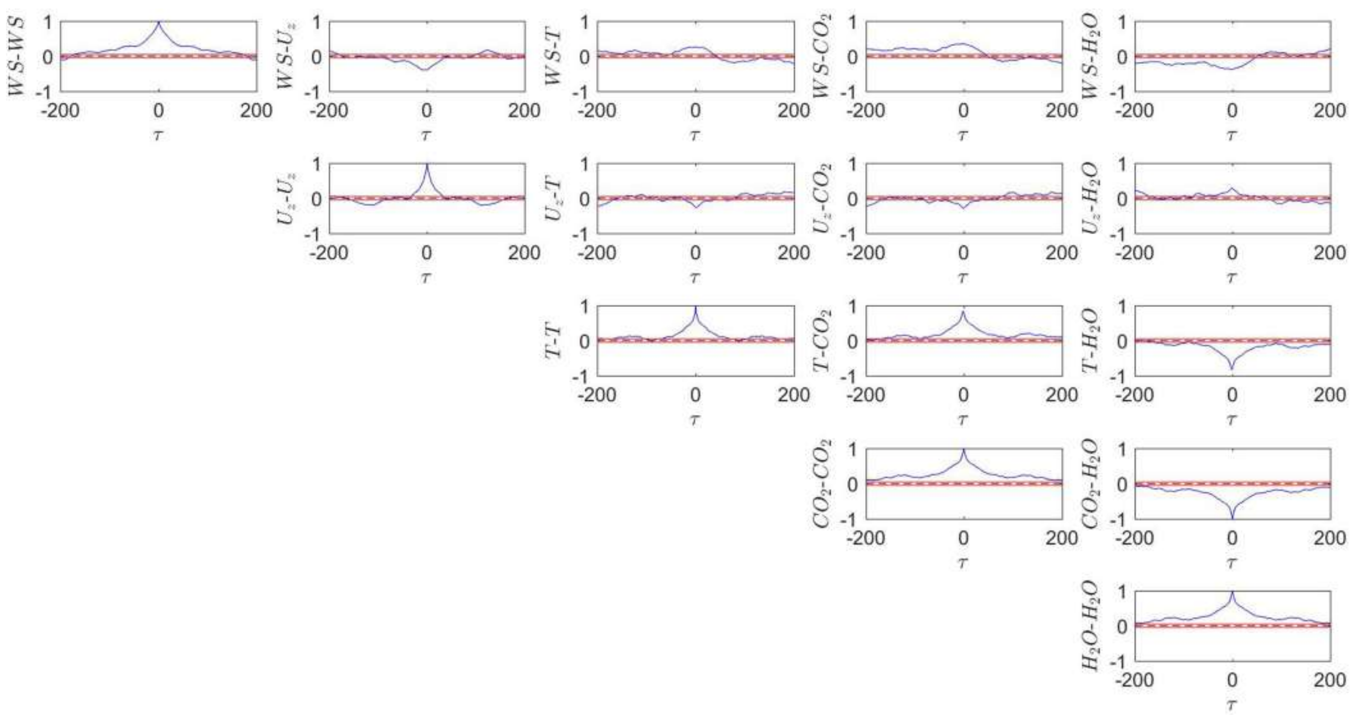

FIG. 15. Cross-covariance for 2017/08/21 13:40–13:43 CST, Cluster #3.

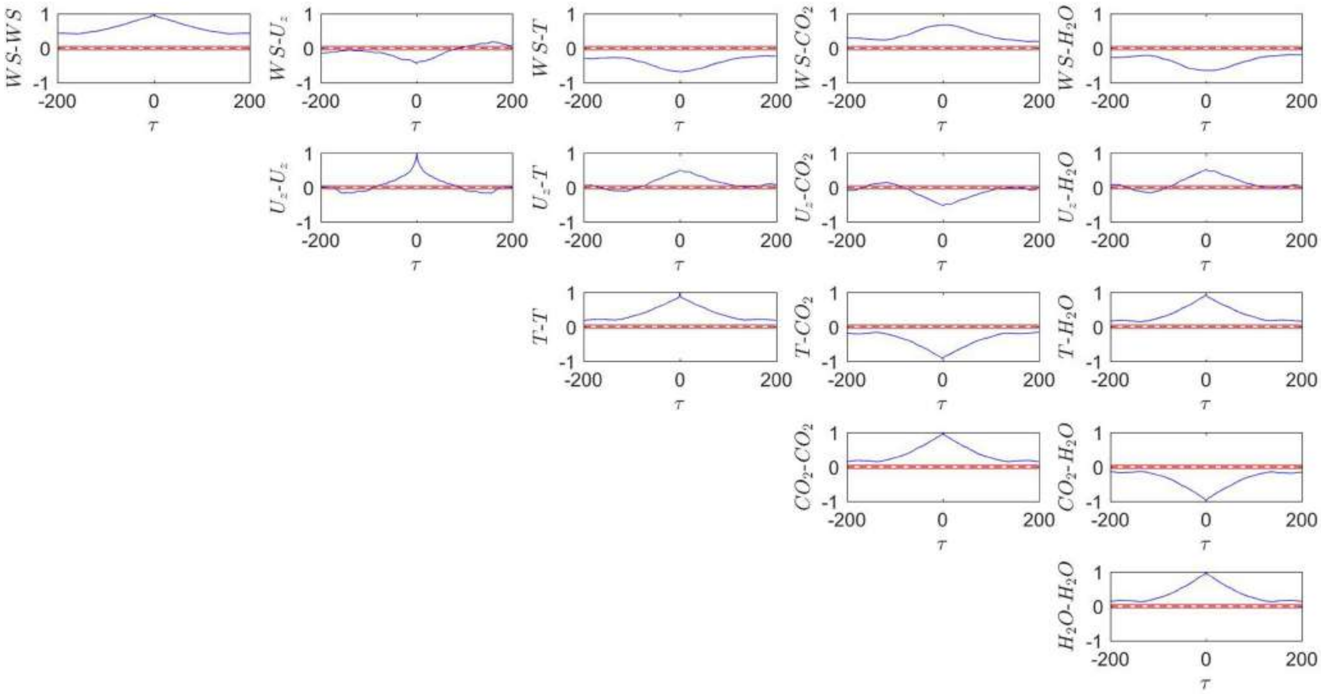

FIG. 16. Cross-covariance for 2017/08/18 10:46–10:49 CST, Cluster #3.

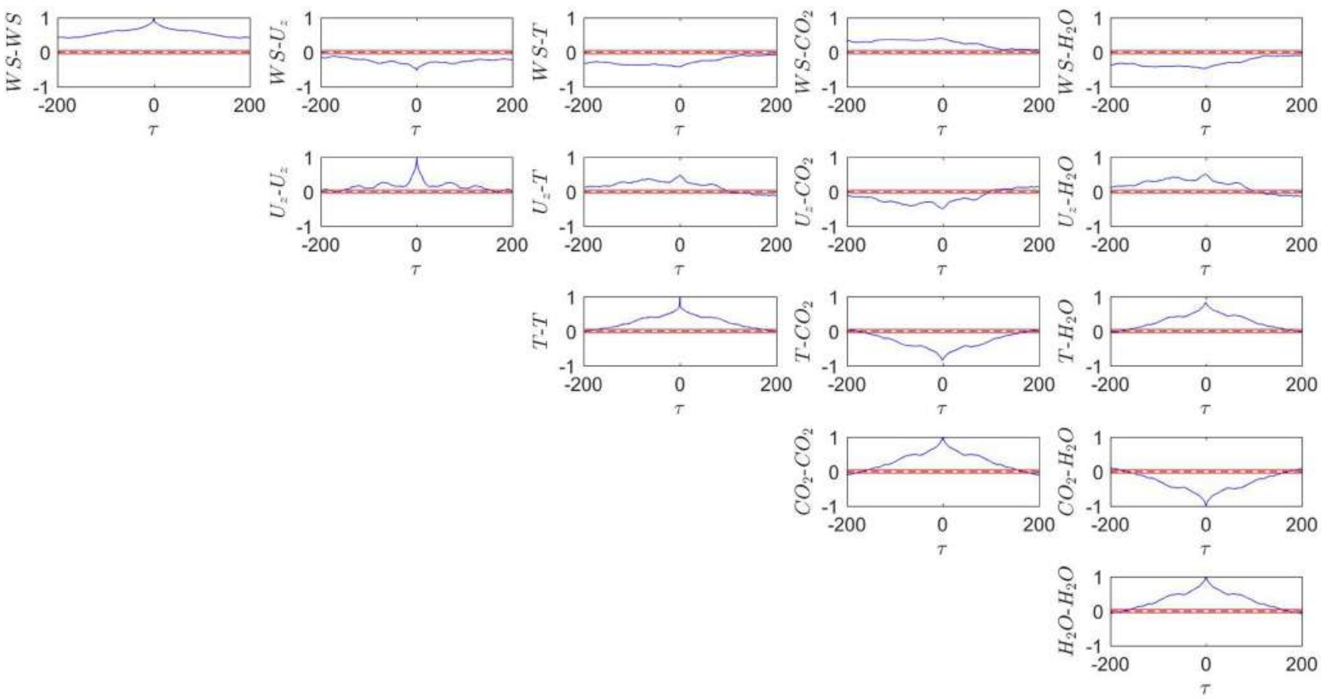

FIG. 17. Cross-covariance for 2017/08/18 10:49–10:52 CST, Cluster #2.

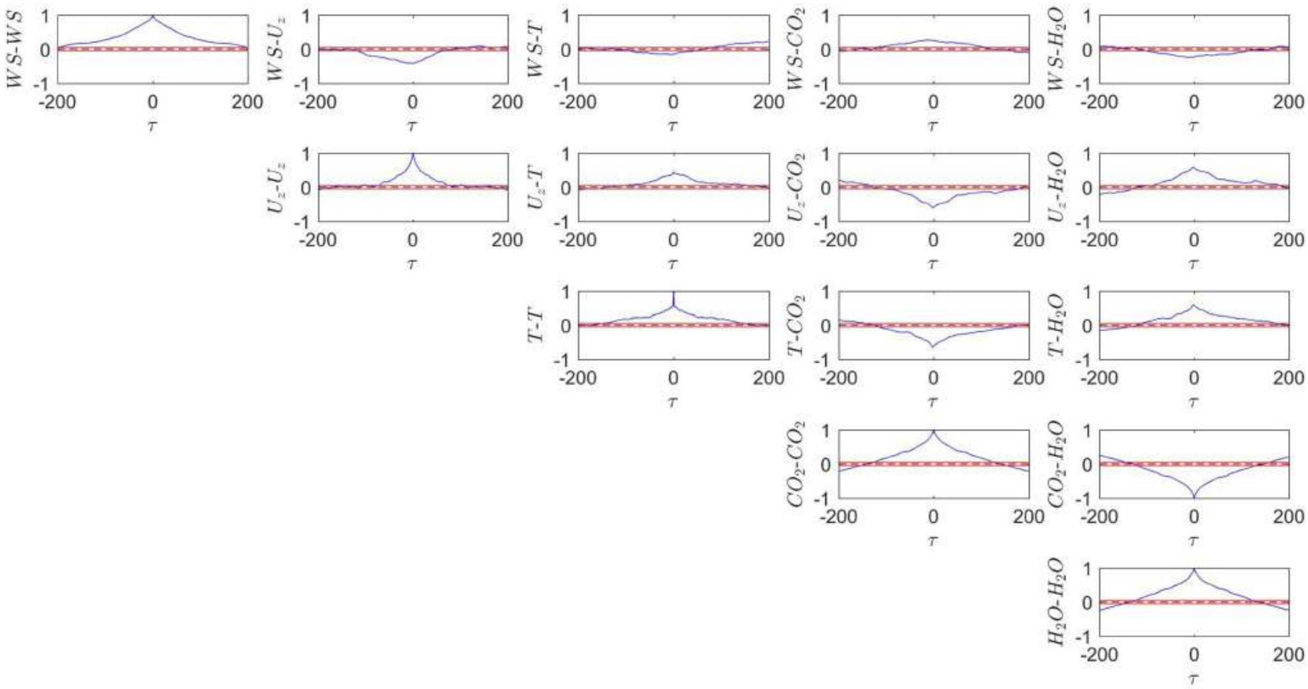

FIG. 18. Cross-covariance for 2017/08/21 12:58–13:01 CST, Cluster #2.

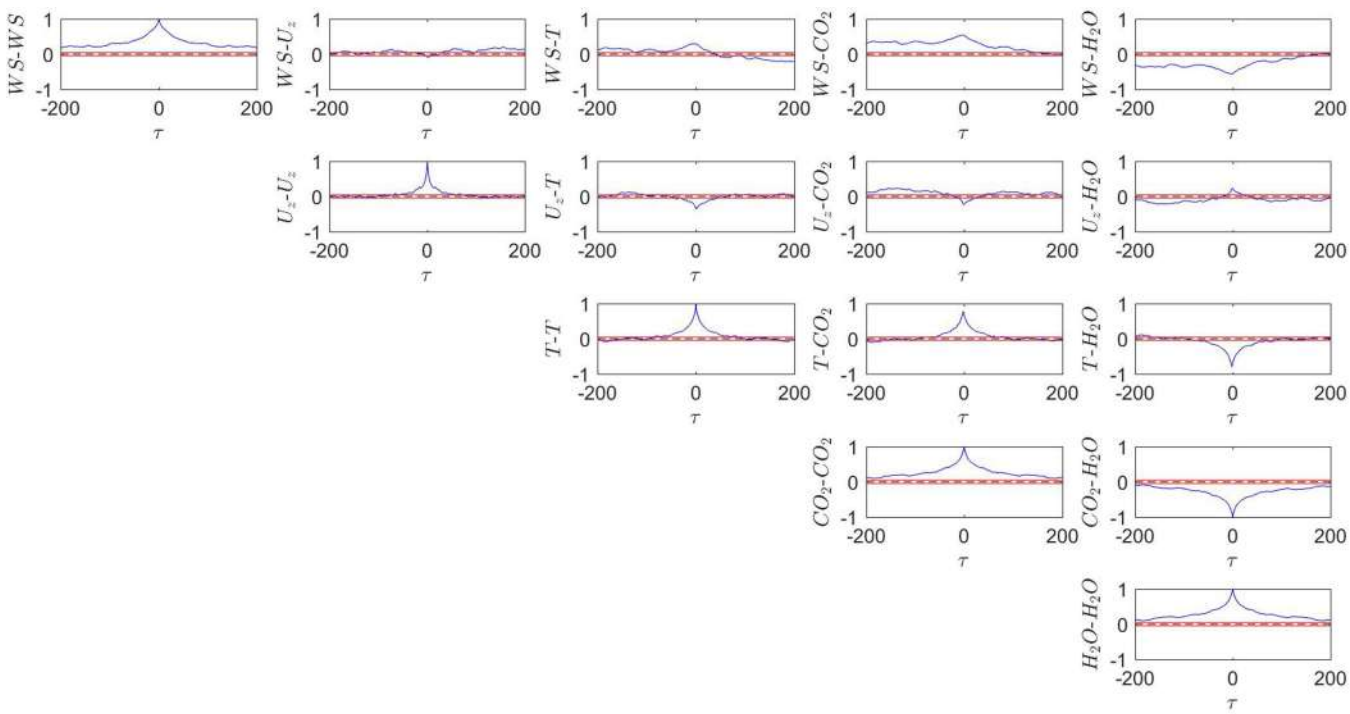

FIG. 19. Cross-covariance for 2017/08/21 13:25–13:28 CST, Cluster #1.

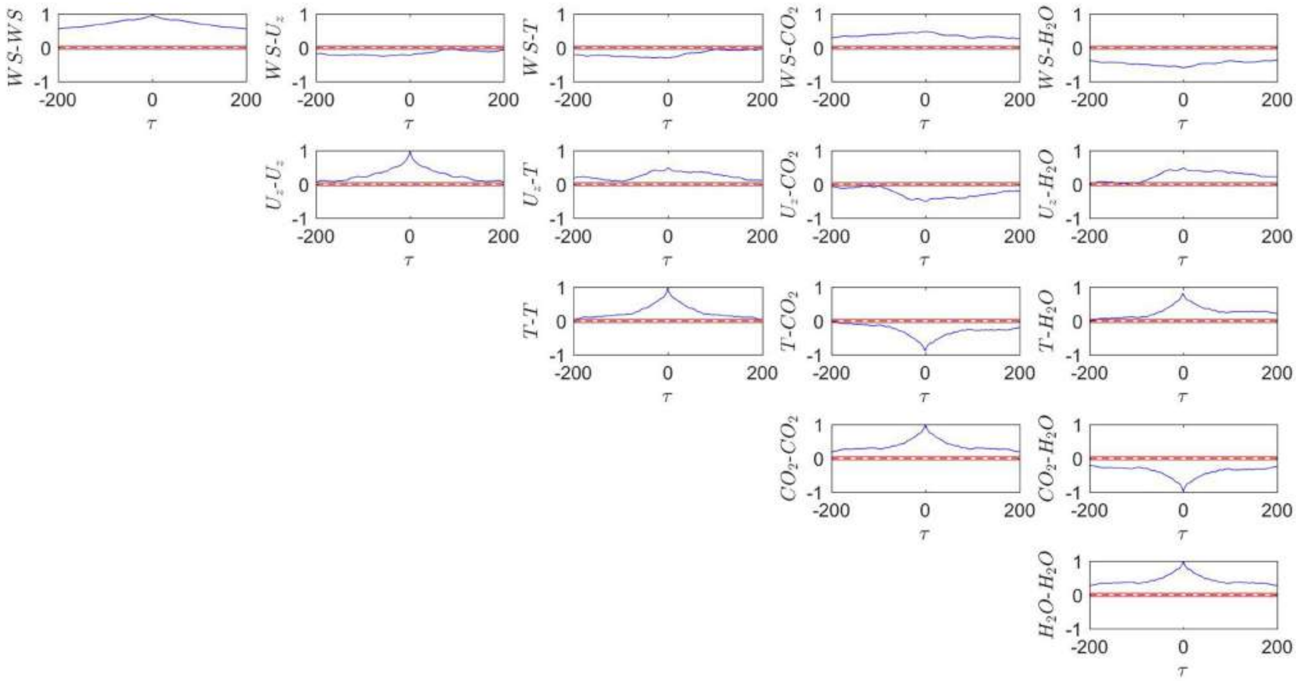

FIG. 20. Cross-covariance for 2017/08/18 12:01–12:04 CST, Cluster #1.

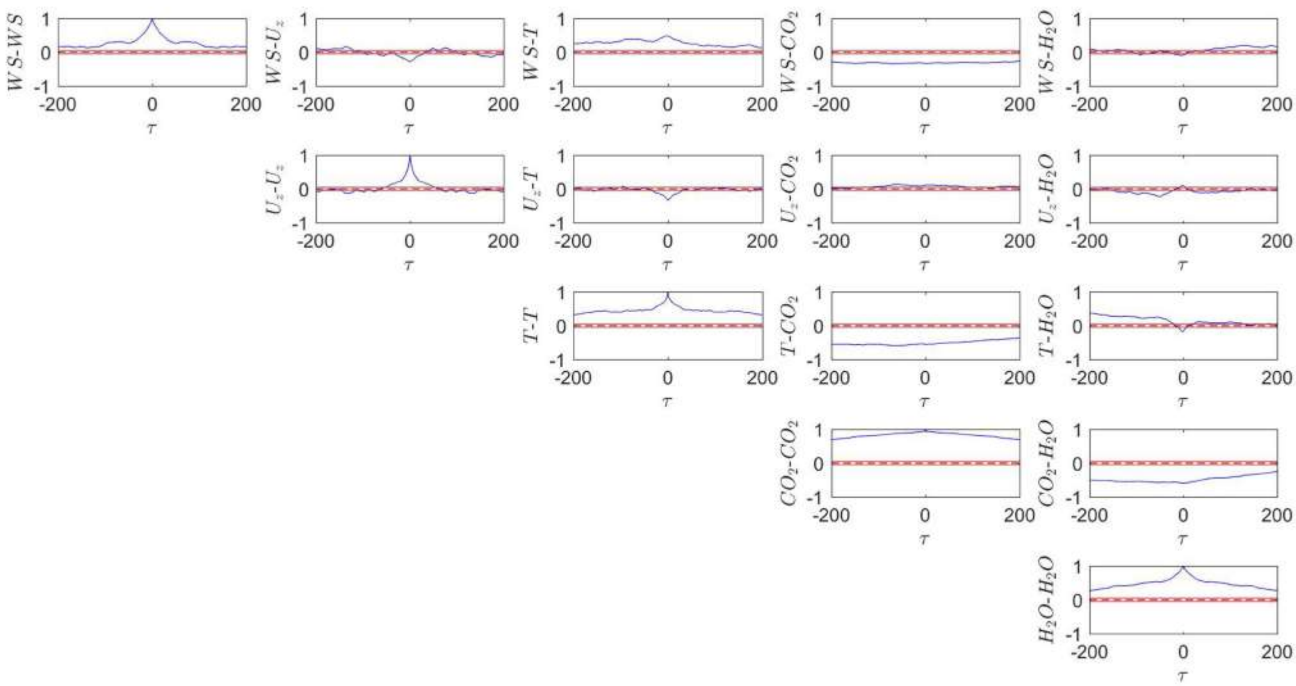

FIG. 21. Cross-covariance for 2017/08/21 12:19–12:22 CST. CO_2 shows a strong instantaneous auto-covariance at time zero that slowly dies out as the time lag increases. This is a unique behavior for CO_2 that is not observed for other windows of analysis.

blue line). We normalize the sequences so that the autocovariance at zero lag equals 1. To determine the confidence intervals for the cross-covariance (Figs. 15–21, red lines above and below the zero line in black), we estimate the lower and upper 95% confidence bounds for the normal distribution discrete time, whose standard deviation is $1/\sqrt{L}$. For a 95%-confidence interval, the critical value is $\sqrt{2}\,\mathrm{erf}^{-1}(0.95) \approx 1.96$ and the confidence interval is

$$\Delta = 0 \pm \frac{1.96}{\sqrt{L}}, \tag{F1}$$

where L is the length of each time series, in our case 1800. In the following plots, the cross-covariance values outside of the 95%-confidence interval (Figs. 15–21, red lines above and below the zero line in black) refer to time series that share a significant cross correlation at different time lags. On the opposite, the time series are not significantly correlated for the time lags where the cross-covariance is inside of the 95%-confidence intervals.

REFERENCES

[1] M. Aubinet, T. Vesala, and D. Papale, *Eddy Covariance: A Practical Guide to Measurement and Data Analysis* (Springer Science & Business Media, 2012).
[2] G. Burba, *Eddy Covariance Method for Scientific, Industrial, Agricultural and Regulatory Applications: A Field Book on Measuring Ecosystem Gas Exchange and Areal Emission Rates* (LI-Cor Biosciences, 2013).
[3] R. B. Stull, *An Introduction to Boundary Layer Meteorology* (Springer Science & Business Media, 2012), Vol. 13.
[4] J. C. Kaimal and J. J. Finnigan, *Atmospheric Boundary Layer Flows: Their Structure and Measurement* (Oxford University Press, 1994).
[5] S. Adarsh and M. J. Reddy, *Multi-Scale Spectral Analysis in Hydrology: From Theory to Practice* (CRC Press, 2021).
[6] D. Vickers and L. Mahrt, "The cospectral gap and turbulent flux calculations," J. Atmos. Ocean. Technol. **20**, 660–672 (2003).
[7] L. H. Koopmans, *The Spectral Analysis of Time Series* (Elsevier, 1995).
[8] G. Padmanabhan and A. R. Rao, "Maximum entropy spectral analysis of hydrologic data," Water Resour. Res. **24**, 1519–1533 (1988).
[9] A. E. Goodwell, P. Jiang, B. L. Ruddell, and P. Kumar, "Does information theory provide a new paradigm for earth science? causality, interaction, and feedback," Water Resour. Res., **115**, e2019WR024940 (2020).
[10] P. Spirtes, C. N. Glymour, R. Scheines, and D. Heckerman, *Causation, Prediction, and Search* (MIT Press, 2000).
[11] A. J. Palmer, C. W. Fairall, and W. Brewer, "Complexity in the atmosphere," IEEE Trans. Geosci. Remote Sens. **38**, 2056–2063 (2000).
[12] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication* (University of Illinois, Urbana, 1949), Vol. 117.
[13] C. Krich, J. Runge, D. G. Miralles, M. Migliavacca, O. Perez-Priego, T. El-Madany, A. Carrara, and M. D. Mahecha, "Estimating causal networks in biosphere–atmosphere interaction with the PCMCI approach," Biogeosciences **17**, 1033–1061 (2020).
[14] J. Runge, P. Nowack, M. Kretschmer, S. Flaxman, and D. Sejdinovic, "Detecting and quantifying causal associations in large nonlinear time series datasets," Sci. Adv. **5**, eaau4996 (2019).
[15] A. E. Goodwell, P. Kumar, A. W. Fellows, and G. N. Flerchinger, "Dynamic process connectivity explains ecohydrologic responses to rainfall pulses and drought," Proc. Natl. Acad. Sci. U.S.A. **115**, E8604–E8613 (2018).
[16] P. Jiang and P. Kumar, "Interactions of information transfer along separable causal paths," Phys. Rev. E **97**, 042310 (2018).
[17] J. Runge, "Quantifying information transfer and mediation along causal pathways in complex systems," Phys. Rev. E **92**, 062829 (2015).
[18] P. Jiang and P. Kumar, "Information transfer from causal history in complex system dynamics," Phys. Rev. E **99**, 012306 (2019).
[19] P. Jiang and P. Kumar, "Bundled causal history interaction," Entropy **22**, 360 (2020).
[20] A. E. Goodwell and P. Kumar, "A changing climatology of precipitation persistence across the United States using information-based measures," J. Hydrometeorol. **20**, 1649 (2019).
[21] A. E. Goodwell and P. Kumar, "Temporal information partitioning networks (TIPNets): A process network approach to infer ecohydrologic shifts," Water Resour. Res. **53**, 5899–5919 (2017).
[22] E. H. Van Nes, M. Scheffer, V. Brovkin, T. M. Lenton, H. Ye, E. Deyle, and G. Sugihara, "Causal feedbacks in climate change," Nature Climate Change **5**, 445–448 (2015).
[23] E. Malmi, N. Tatti, and A. Gionis, "Beyond rankings: Comparing directed acyclic graphs," Data Min. Knowl. Discov. **29**, 1233–1257 (2015).
[24] C. Monfreda, N. Ramankutty, and J. A. Foley, "Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000," Global Biogeochem. Cycles **22**, 1 (2008).
[25] L. C. Hernandez Rodriguez, A. E. Goodwell, and P. Kumar, "Inside the flux footprint: Understanding the role of organized land cover heterogeneity on land-atmosphere exchange fluxes," Agricult. Forest Meteorol. **5**, 39 (2022).
[26] C. G. Wilson, B. Abban, L. L. Keefer, K. Wacha, D. Dermisis, C. Giannopoulos, S. Zhou, A. E. Goodwell, D. K. Woo, Q. Yan, M. Ghadiri, A. Stumpf, M. Pitcel, Y.-F. Lin, L. Marini, B. Storsved, K. Goff, J. Volgelgsand, A. Dere, K. E. Schilling, M. Muste, N. E. Blair, B. Rhoads, A. Bettis, H. Pai, C. Kratt, C. Sladek, M. Wing, J. Selker, T. Scott, H. Lin, P. Kumar, and A. Papanicolaou, "The intensively managed landscape critical zone observatory: A scientific testbed for understanding critical zone processes in agroecosystems," Vadose Zone J. **17**, 1–21 (2018).
[27] J. Runge, V. Petoukhov, J. F. Donges, J. Hlinka, N. Jajcay, M. Vejmelka, D. Hartman, N. Marwan, M. Paluš, and J. Kurths, "Identifying causal gateways and mediators in complex spatio-temporal systems," Nat. Commun. **6**, 1–10 (2015).
[28] J. Runge, J. Heitzig, V. Petoukhov, and J. Kurths, "Escaping the curse of dimensionality in estimating multivariate transfer entropy," Phys. Rev. Lett. **108**, 258701 (2012a).
[29] J. Runge, J. Heitzig, N. Marwan, and J. Kurths, "Quantifying causal coupling strength: A lag-specific measure for multivariate time series related to transfer entropy," Phys. Rev. E **86**, 061121 (2012b).
[30] J. Runge, V. Petoukhov, and J. Kurths, "Quantifying the strength and delay of climatic interactions: The ambiguities of cross correlation and a novel measure based on graphical models," J. Clim. **27**, 720–739 (2014).
[31] F. Rosas, P. A. Mediano, M. Ugarte, and H. J. Jensen, "An information-theoretic approach to self-organisation: Emergence of complex interdependencies in coupled dynamical systems," Entropy **20**, 793 (2018).
[32] P. Jiang, "Causal history analysis of complex system dynamics," Ph.D. thesis (University of Illinois at Urbana-Champaign, 2019).
[33] N. Wiener, *I Am a Mathematician, the Later Life of a Prodigy: An Autobiographical Account of the Mature Years and Career of Norbert Wiener and a Continuation of the Account of his Childhood in Ex-Prodigy* (Doubleday, 1956), Vol. 20.
[34] C. W. Granger, "Investigating causal relations by econometric models and cross-spectral methods," Econometr.: J. Econometric Soc. **37**(3), 424–438 (1969).
[35] J. Runge, "Conditional independence testing based on a nearest-neighbor estimator of conditional mutual information," in *International Conference on Artificial Intelligence and Statistics* (PMLR, 2018), pp. 938–947.
[36] L. Kozachenko and N. N. Leonenko, "Sample estimate of the entropy of a random vector," Probl. Peredachi Inform. **23**, 9–16 (1987).
[37] S. Frenzel and B. Pompe, "Partial mutual information for coupling analysis of multivariate time series," Phys. Rev. Lett. **99**, 204101 (2007).
[38] A. K. Jain, "Data clustering: 50 years beyond k-means," Pattern Recog. Lett. **31**, 651–666 (2010).
[39] R. T. Aldahdooh and W. M. Ashour, "Dimk-means—Distance-based initialization method for k-means clustering algorithm," Int. J. Intell. Syst. Appl. **5**, 2 (2013).
[40] M. Matsumoto and T. Nishimura, "Mersenne twister: A 623-dimensionally equidistributed uniform pseudo-random number generator," ACM Trans. Model. Comp. Simul. **8**, 3–30 (1998).
[41] T. M. Kodinariya and P. R. Makwana, "Review on determining number of cluster in k-means clustering," Int. J. **1**, 90–95 (2013).
[42] J. C. Dunn, "Well-separated clusters and optimal fuzzy partitions," J. Cybernet. **4**, 95–104 (1974).

[43] A. Kraskov, H. Stögbauer, and P. Grassberger, "Estimating mutual information," Phys. Rev. E **69**, 066138 (2004).
[44] A. A. Hagberg, D. A. Schult, and P. J. Swart, "Exploring network structure, dynamics, and function using networkx," in *Proceedings of the 7th Python in Science Conference*, edited by G. Varoquaux, T. Vaught, and J. Millman (Pasadena, CA USA, 2008), pp. 11–15.
[45] J. Runge, P. Nowack, M. Kretschmer, S. Flaxman, and D. Sejdinovic, "Detecting causal associations in large nonlinear time series datasets," arXiv:1702.07007 (2017).
[46] J. P. Crutchfield, "Between order and chaos," Nat. Phys. **8**, 17–24 (2012).
[47] J. Pearl, *Causality* (Cambridge University Press, 2009).
[48] R. Kooper, B. Angelo, L. Marini, P. Kumar, and M. Muste, "Information architecture used to manage multi-domain data analysis in intensively managed landscape—Critical zone observatory," in *American Geophysical Union, Fall Meeting 2016, abstract #IN51C-1867* (2016).
[49] S. Campbell, *Csat3 Three-Dimensional Sonic Anemometer Instruction Manual* (Campbell Scientific, 2017).
[50] LI-COR, "Li-7500 open path CO_2/H_2O analyzer instruction manual," 2003.
[51] S. Campbell, *Model 083E Relative Humidity and Temperature Sensor Instruction manual* (Campbell Scientific, 2016).