A data-driven framework for dimensionality reduction and causal inference in climate fields

Fabrizio Falasca,* Pavel Perezhogin, and Laure Zanna
Courant Institute of Mathematical Sciences
New York University, New York, NY, USA
(Dated: October 17, 2023)

We propose a data-driven framework to simplify the description of spatiotemporal climate variability into few entities and their causal linkages. Given a high-dimensional climate field, the methodology first reduces its dimensionality into a set of regionally constrained patterns. Time-dependent causal links are then inferred in the *interventional* sense through the fluctuation-response formalism, as shown in Baldovin et al. (2020) [1]. These two steps allow to explore how regional climate variability can influence remote locations. To distinguish between true and spurious responses, we propose a novel analytical null model for the fluctuation-dissipation relation, therefore allowing for uncertainty estimation at a given confidence level. Finally, we select a set of metrics to summarize the results, offering a useful and simplified approach to explore climate dynamics. We showcase the methodology on the monthly sea surface temperature field at global scale. We demonstrate the usefulness of the proposed framework by studying few individual links as well as "link maps", visualizing the cumulative degree of causation between a given region and the whole system. Finally, each pattern is ranked in terms of its "causal strength", quantifying its relative ability to influence the system's dynamics. We argue that the methodology allows to explore and characterize causal relationships in high-dimensional spatiotemporal fields in a rigorous and interpretable way.

CONTENTS

I. Introduction 2

II. Framework 3
- A. Partitioning climate fields into regionally constrained patterns 3
  - 1. Graph inference 4
  - 2. Detecting communities 4
- B. Linear response theory and fluctuation-dissipation relation 4
  - 1. General case 5
  - 2. Linear systems and quasi-Gaussian approximation 5
- C. A *null model* for the fluctuation-dissipation relation 5
  - 1. Confidence bounds of the response matrix: numerical estimation 6
  - 2. Confidence bounds of the response matrix: analytical derivation 6
- D. A simple example 7
- E. Metrics 7

III. Data 8

IV. Causality in climate fields 9
- A. Applicability of fluctuation-response theory in climate studies 9
- B. Relation to previous climate studies 9
- C. Application to global sea surface temperature 9
  - 1. Dimensionality reduction and causal inference 9
  - 2. Investigation of few causal interactions 10

V. Conclusions and discussion 13

Acknowledgments 14

Code availability 14

A. Dimensionality reduction in climate. Limitations of current methods and proposal 14
  - 1. Two goals in dimensionality reduction studies 14
  - 2. Few limitations of common methods and proposal 14

B. A *null* model for the Fluctuation-Dissipation relation. Analytical derivation of the confidence bounds 15
  - 1. Notation adopted in this section 15
  - 2. Analytical derivation 16
    - a. Expected value and variance of the response estimator 16
    - b. Computation of each summation 17
    - c. Final result 18

C. Confidence bounds. Numerical vs analytical 18

D. Histograms of each mode $x_i(t)$ in the global SST field 20

References 22

* fabri.falasca@nyu.edu

I. INTRODUCTION

The Earth's climate is a complex dynamical system composed by many interacting components, such as the atmosphere and hydrosphere, and their interactions [2]. Such linkages give rise to nontrivial feedbacks, generating self-sustained spatiotemporal patterns [3, 4]. An example is the El Niño Southern Oscillation (ENSO), a recurrent pattern of natural variability emerging from air-sea interaction in the tropical Pacific Ocean [5, 6]. Other examples include the Asian Monsoon, the Indian Ocean Dipole, and the Atlantic Niño, just to cite a few [7–9]. Such patterns, commonly referred to as *modes of variability*, interact with each other on a vast range of spatial and temporal scales, see for example [10–12]. Spatiotemporal climate dynamics can then be considered a collection of modes of variability and their linkages, or as commonly referred to, a "climate network" [13, 14]. The identification of such a complex array of interactions and the quantification of its response to external forcings (e.g., [15, 16]) is a fundamental (but nontrivial) problem at the root of our understanding of climate dynamics. It requires hierarchies of models, theories, observations, and new tools to analyze and simplify the description of high-dimensional, complex data [4, 17]. In fact, the exponential growth of data from models and observations, together with appropriate and rigorous frameworks, promise new ways to explore and ultimately understand climate dynamics [17]. An important step when "learning" from climate data is to infer meaningful linkages among time series, whether among modes of variability or other features of the system (e.g., global averages). Traditionally, this has been done by quantification of *pairwise* similarities, whether linear or nonlinear (for example [16, 18, 19] and [20], respectively). Such statistical similarities cannot quantify what we refer to as "causality", limiting our ability to discover meaningful mechanisms in high-dimensional dynamical systems such as climate. In the context of dynamical systems, the main idea of causal inference can be informally summarized as follows: given a system $\boldsymbol{x} = [x_1(t), x_2(t), ..., x_N(t)]$ of N time series, where t is a time index, we aim in quantifying (a) to what extent and (b) at what time scales changes in a variable $x_j(t)$ can influence another variable $x_k(t+\tau)$ at later times [1, 21].

This study proposes a scalable framework to (a) coarse grain a spatiotemporal climate field in a set of few patterns and (b) infer the causal links among such entities. Altogether, this allows to study complex, high-dimensional climate dynamics in an interpretable and simplified way.

Causality is a fundamental topic in science ranging from foundational questions in physics and philosophy [22–30] to practical design and implementation of inference algorithms [31]. In the last decades, there has been a great interest in developing new methodologies to infer causal associations directly from data. In the case of time series data, attempts to infer causal connections start from the work of Granger [32], who framed the problem of causal inference in terms of prediction. The main idea of Granger causality was to draw a causal link between two variables x_j and x_k if the past of x_j would enhance the predictability of the future of x_k. Another attempt, coming from the dynamical system literature, was based on the concept of transfer entropy [33, 34]. Crucially, as noted in [1], Granger causality and transfer entropy give similar information and are equivalent for Gaussian variables [35]. In the last decades, new ideas from computer science, mainly driven by Pearl [31, 36], have given us practical ways to design and implement causal tools mainly based on graphical models. Frameworks of such kind have been recently developed in climate science with contributions ranging from the work of Ebert-Uphoff and Deng (2012) in [37] to the newer "PCMCI" method led by Runge et al. (2019) [38]; see [39] for a review. Additionally, the Machine Learning (ML) community is actively interested in causality and applications and we refer to [40] for details on new developments and open problems in "Causal ML".

Recently, it has been noted that linear response theory [41, 42] may serve as a rigorous framework to understand causality in physical systems [1, 21, 43, 44]. The main rationale is that the formalism provides a strategy to compute the change in statistical properties of a physical system after a small perturbation solely from the notion of the *unperturbed* dynamics [44, 45]. This allows to capture causal relations in the *interventional* sense [1, 30, 36], as done typically in physical experiments.

This differs from many commonly employed causal algorithms, such as conditional independence testing [46], Granger causality [32] and transfer entropy [47], by focusing directly on the problem of causal effect estimation [31] rather than causal discovery. Many causal questions in climate can be cast into the paradigm of perturbations and responses as proposed in [1]. Examples of such questions may in fact be: how much do changes in fresh water fluxes in Antarctica affect sea level rise in the North Atlantic? How do changes in sea surface temperature anomalies in the Pacific Ocean affect temperatures in the Indian Ocean? Answering such questions often relies on quantifying the time-dependent "flow of information" along the underlying causal graph rather than discovering the graph itself [1, 44] (see also [48] in the context of information theory). Such difference with causal discovery methods is further explored and discussed in Section II D. On the computational side, causal discovery algorithms such as the one based on conditional independence, do not scale to high-dimensional systems [39, 40]. Differently, linear response theory scales to high-dimensional data and

allows to write rigorous, analytical relations between perturbations and responses.

It should be noted that linear response theory is an active field of research in climate studies [3, 21, 49–56]. Such studies, quantifying long-term, forced changes in climate observables, can be broadly grouped in two approaches [57]: the one pioneered by Leith (1975) [49], making use of the fluctuation-dissipation formalism, and the more general formalism proposed by Ruelle (1998) [43, 58]. This study relates to the approach proposed by Leith [49] by considering (a) stationary fields and (b) impulse perturbations. This statement will be furthered formalized in Section IV.

The extension of the proposal of Baldovin et al. (2020) [1] for studying spatially extended dynamical systems is contingent on two important steps: (i) a methodology to reduce the dimensionality of the system and (ii) a framework for uncertainty estimation. Point (ii) is particularly important when inferring results from real-world data. In such case spurious results are always present.

In this paper, we contribute to (a) dimensionality reduction, (b) linear response theory and (c) causality in climate in the following ways:

i) We introduce a scalable computational strategy to decompose a large spatiotemporal climate fields into a set of few regionally constrained modes. The average time series inside each pattern quantifies the climate variability of specific regions around the world. The time-dependent linkages among such patterns are then inferred through the fluctuation-dissipation relation. This step allows to explore how *local* (i.e. regional) variability can influence *remote* locations.

ii) We propose an analytical *null* model for the fluctuation-dissipation relation. The model assigns confidence bounds to the estimated linear responses, therefore distinguishing between *true* and *spurious* responses. This allows for trustworthy statistical inference from real-world data. The application of this model is general and not limited to climate applications.

iii) We showcase the proposed framework on the monthly sea surface temperature (SST) field at global scale. For this step, we consider a 300 years long, stationary integration of a global coupled climate model. Long-distance linkages in the SST field have been characterized in many previous studies. It therefore offers a good real-world test-bed for the methodology. We show how the proposed framework drastically simplifies the description of such a complex, high-dimensional system in an interpretable and comprehensive way.

The paper is organized as follows: in Sec. II we introduce the proposed framework. The data analyzed are described in Sec. III. The methodology is applied to climate data in Sec. IV. Sec. V concludes the work.

II. FRAMEWORK

A. Partitioning climate fields into regionally constrained patterns

Spatiotemporal chaotic fields can be viewed as dynamical systems $\boldsymbol{x} \in \mathbb{R}^N$ living in a N-dimensional state space [59, 60]. The dimensionality N is theoretically infinite but in practice equal to the number of grid cells used to discretize the longitude, latitude and vertical coordinates (times the total number of variables) [61]. In the case of dissipative chaotic systems, such high-dimensional dynamics is confined on lower-dimensional objects known as "inertial manifolds" or "attractors" [60, 62, 63]. The *effective* dimensionality of the system [64] is then finite and given by the attractor dimension D. This is arguably the case of large scale climate dynamics, where recurrent spatiotemporal patterns, known as modes of variability (e.g., ENSO, monsoon system, Indian Ocean modes [9, 16, 65] etc.) are a manifestation of the low dimensionality of the climate attractor [61, 66].

Here the goal is to coarse grain the original climate field $\boldsymbol{x} \in \mathbb{R}^N$ into a set of very few (order 10) patterns. Crucially, such components should be *regionally constrained* in longitude-latitude space. This comes from the observation that, physically, climate variability can be often thought of as a set of responses driven by *local* perturbations (e.g., warming of the tropics driven by anomalous warming in the eastern Pacific [6]). Methods like δ-MAPS [67, 68] can extract modes of variability. Given climate fields, δ-MAPS first identifies spatially contiguous clusters and then infers a weighted and direct network between such entities based on correlations. The method has proven to be useful in climate studies [16, 19, 69–72] but suffer from few drawbacks: it does not scale well with high-dimensional datasets (i.e., a large number of grid cells) and, depending on the field analyzed, it can be sensitive to one of its parameters.

In Appendix A we discuss strengths and limitations of current dimensionality reduction methods and further motivate our proposal.

In this study, we show that adding a simple constraint to community detection methodologies [73, 74] provide a scalable and practical framework to identify regionally constrained modes of variability in climate fields. The strategy proposed here is based on two main steps: first, given a field $\boldsymbol{x} \in \mathbb{R}^{N,T}$ we infer a graph between its N time series based on both their covariability and distance. We then identify communities in such graph, partitioning

the original data into n components. Each community will consist of sets of highly correlated time series and will serve as proxies for climate modes of variability.

1. Graph inference

Consider a spatiotemporal field saved as a data matrix $\boldsymbol{x} \in \mathbb{R}^{N,T}$, with N time series of length T. Given a pair of time series $x_i(t)$ and $x_j(t)$, scaled to zero mean, we compute their covariance at lag $\tau = 0$, $C_{i,j} = \overline{x_i(t)x_j(t)}$; where $\overline{f}$ stands for the temporal average of function f. An undirected, unweighted graph can then be encoded in adjacency matrix $\boldsymbol{A} \in \mathbb{R}^{N,N}$ as:

$$A_{i,j} = \begin{cases} 1 - \delta_{i,j} & \text{if } C_{i,j} \geq k \text{ and } d(i,j) \leq \eta \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

Where the Kronecker delta $\delta_{i,j}$ allows to remove "self-links".

The parameter k sets the minimum covariance that two time series must be connected. The parameter $d(i,j)$ is the distance between grid cells i and j, and η is a distance threshold. The rationale behind this choice is that we consider two time series $x_i(t)$ and $x_j(t)$ linked to each other if (a) their covariance is larger than a threshold k and (b) if they are relatively close in the spatial domain considered. Importantly, $d(i,j)$ is computed using the Haversine distance, determining the distance between two points (i and j) on a sphere given their longitudes and latitudes. Potentially, such parameters can be specified by the user. However, their optimal values will largely depend on the statistics of the field of interest (e.g., sea surface temperature, cloud fraction) and by the spatial domain considered (e.g., regional or global domains). We therefore propose two heuristics to compute such parameters directly from the data matrix $\boldsymbol{x} \in \mathbb{R}^{N,T}$.

a. Heuristic for parameter k. Given time series $x_i(t)$ and $x_j(t)$: (a) compute covariances $C_{i,j}$, $\forall i,j; i \neq j$ and (b) set k as a high quantile q of the distribution of all covariances $C_{i,j}$. To make this idea feasible in practice, we can approximate such distribution by random sampling S_k pairs of time series $x_i(t)$ and $x_j(t)$ and then computing their covariances. k is then estimated as a high quantile q of the sampled distribution. A pragmatic choice of q is $q = 0.95$ as we observed in different applications that is a good compromise between the identification of a sparse, but not too sparse, graph. The sampling size considered here is $S_k = 10^6$.

b. Heuristic for parameter η. Given time series $x_i(t)$ and $x_j(t)$ embedded at grid point i and j: (a) calculate the Haversine distance $d(i,j)$ between pairs i and j and (b) estimate η as a low quantile of the distribution of all distances $d(i,j)$. As for the parameter k, in practice the distribution of distances can be approximated by random sampling S_η pairs of locations i and j and computing their Haversine distance. We choose $q = 0.15$, with no large sensitivity over such threshold, and $S_\eta = 10^6$.

2. Detecting communities

Sets of highly correlated time series in the original field $\boldsymbol{x} \in \mathbb{R}^{N,T}$ correspond to groups of nodes that are more interconnected to each other than to the rest of the graph, in other words "communities" [74]. Fast and scalable community detection algorithms [75] can be leveraged to reduce the dimensionality of the graph in Eq. 1. In this study, we consider the Infomap methodology [76, 77]. Such method is based on the Map Equation [78] and casts the problem of community detection as an optimal compression problem [77]. Mainly, Infomap exploits the community structure to minimize the description of a random walk on the graph [78]. Such methodology has been found to be the best performing community detection in different benchmarks, such as in [75], and also shown excellent performance in a previous climate study [79].

Finally, given a set of n communities $c = (c_1, c_2, c_3, ...c_n)$ we study their temporal variability as the average over every time series inside. Formally, for each community c_j we define its signal as $X(c_j,t) = \frac{1}{\sum_{i \in c_j} \cos(\theta_i)} \sum_{i \in c_j} x_i(t) \cos(\theta_i)$; where θ_i is the latitude of $x_i(t)$. The term $\cos(\theta_i)$ allows to implement the area-weighted averaging on a uniform longitude-latitude grid. The fluctuation-dissipation response formalism is then leveraged to infer causality among such time series.

In this study, we considered correlation functions rather than covariances, therefore $C_{i,j} = \overline{x_i(t)x_j(t)}$ in Eq. 1 are computed after scaling $x_i(t)$ to unit variance. This was done for qualitative comparison with results obtained through the δ-MAPS framework [67, 68] but covariances can be considered in future work.

B. Linear response theory and fluctuation-dissipation relation

Baldovin et al. (2020) [1], proposed the following physical definition of causality: given a dynamical system $\boldsymbol{x}(t) = [x_1(t), x_2(t), ..., x_N(t)]$ with N time series, each of length T we say that x_j causes x_k, i.e. $x_j \to x_k$, if a small perturbation applied to variable x_j at time $t = 0$, i.e. $x_j(0) \to x_j(0) + \delta x_j(0)$, induces *on average* a change on variable $x_k(\tau)$ at a later time $t = \tau$. We note that [21, 44] pursue close scientific goals, and the proposal in [1] can be viewed as a specific case of the general framework proposed in [21].

1. General case

Consider a Markov process $\boldsymbol{x}(t) = [x_1(t), x_2(t), ..., x_N(t)]$. Each time series $x_i(t)$ is scaled to zero mean. The system is stationary with invariant probability distribution $\rho(\boldsymbol{x})$. We perturb the system $\boldsymbol{x}(t)$ at time $t = 0$ with a small, impulse perturbation $\delta\boldsymbol{x}(0) = [\delta x_1(0), \delta x_2(0), ..., \delta x_N(0)]$. We aim to answer the following question: how does this *external* perturbation $\delta\boldsymbol{x}(0)$ affect the whole system $\boldsymbol{x}(\tau)$ at time $t = \tau$, on average? Formally, we are interested in quantifying the following object:

$$\delta\langle x_k(\tau)\rangle = \langle x_k(\tau)\rangle_p - \langle x_k(\tau)\rangle, \tag{2}$$

where the brackets $\langle x_k(\tau)\rangle$ indicate the ensemble averages of $x_k(\tau)$, i.e. the average over many realizations of the system, and the subscript p specifies the perturbed dynamics. Therefore, Eq. 2 defines the difference between the components $x_k(\tau)$ of the perturbed and unperturbed systems in the *average* sense. Eq. 2 can be used to study changes $\delta\langle\mathcal{O}(x_k(\tau))\rangle$ of a generic observable $\mathcal{O}(x_k(\tau))$ (i.e., a physical measurable quantity, function of the state space vector $\boldsymbol{x}(\tau)$ at time $t = \tau$). To study causality, here we simply consider the identity case $\mathcal{O}(x_k(\tau)) = x_k(\tau)$, see [1].

Under the assumption of a small perturbation $\delta\boldsymbol{x}(0)$ and with $\rho(\boldsymbol{x})$ sufficiently smooth and non-vanishing, the following result holds:

$$R_{k,j}(\tau) = \frac{\delta\langle x_k(\tau)\rangle}{\delta x_j(0)} = -\Big\langle x_k(\tau)\frac{\partial \ln\rho(\boldsymbol{x})}{\partial x_j}\Big|_{\boldsymbol{x}(0)}\Big\rangle. \tag{3}$$

$\boldsymbol{R}(\tau)$ is the linear response matrix and we refer to Section II of Boffetta et al. (2003) [80] for a derivation of Eq. 3. $R_{k,j}(\tau)$ quantifies the response of a variable $x_k(\tau)$ at time $t = \tau$ given a small perturbation $\delta x_j(0)$ applied to variable $x_j(0)$ at time $t = 0$. Eq. 3 is known as the generalized fluctuation-dissipation relation (FDR) and valid for both linear and nonlinear systems [42]. Note that in case of deterministic systems the invariant measure $\rho(\boldsymbol{x})$ is singular almost everywhere on the attractor. Therefore in practice one needs to add Gaussian noise even to deterministic systems in order to "smooth" the probability distribution before applying FDR as proposed here [51].

Eq. 3 is a powerful formula as it allows to compute responses to perturbations solely given the gradients of the probability distribution $\rho(\boldsymbol{x})$ of the *unperturbed* system. However, the functional form of $\rho(\boldsymbol{x})$ is not known a priori and can be highly nontrivial, especially for high-dimensional systems. To overcome such issue, applications often focus on the simpler case of Gaussian distributions (see for example [49, 57]). This is the case of linear systems as shown in the next section.

2. Linear systems and quasi-Gaussian approximation

We now consider a N dimensional stochastic linear process $\boldsymbol{x}(t) = [x_1(t), x_2(t), ..., x_N(t)]$ governed by the following equation:

$$\boldsymbol{x}(t+1) = \boldsymbol{M}\boldsymbol{x}(t) + \boldsymbol{B}\boldsymbol{\xi}(t). \tag{4}$$

The matrix $\boldsymbol{M} \in \mathbb{R}^{N,N}$ specifies the deterministic dynamics of the system. The term $\boldsymbol{\xi} \in \mathbb{R}^N$ with $\xi_i(t) \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$ represents a delta correlated white noise (i.e., $\langle\xi(t)\xi(s)\rangle = \delta_{t,s}$). The matrix $\boldsymbol{B} \in \mathbb{R}^{N,N}$ specifies the amplitude of the noise (i.e., standard deviation).

In this case, the probability distribution $\rho(\boldsymbol{x})$ is Gaussian and Eq. 3 factorizes to:

$$\boldsymbol{R}(\tau) = \boldsymbol{M}^\tau = \boldsymbol{C}(\tau)\boldsymbol{C}(0)^{-1}. \tag{5}$$

Where the covariance function $C_{i,j}(\tau) = \langle x_i(t+\tau)x_j(t)\rangle$ (x_i is assumed to be zero mean). Eq. 5 shows that the response of a linear system to small *external* perturbations is encoded in its covariance functions and can be therefore estimated from its time history [80].

a. Relevance for nonlinear systems. Such form of the FDR has been the one commonly used in climate applications and it is commonly referred to as "quasi-Gaussian approximation" [51, 53, 54, 81, 82]. Importantly, it has been shown that such formula performs well for weakly nonlinear systems. For instance Baldovin et al. (2020) [1] showed remarkably good results when analyzing linear responses in a Langevin equation with a quartic potential. Gritsun et al. (2007) [51] also pointed out how this formula works well also for non-Gaussian systems with second order nonlinearities. Additionally, Eq. 5 has been shown to give reliable results in the case of nonlinear deterministic dynamical systems also in case of finite perturbations, see Fig. 1 in Boffetta et al. (2003) [80]). Furthermore, we will show in Appendix D that the probability distributions considered in this study can be well approximated by Gaussians, further justifying the use of this approximation in our context.

Results presented in this section hold in the sense of ensemble average, therefore covariances $\boldsymbol{C}(\tau)$ and $\boldsymbol{C}(0)$ are computed by averaging over many realizations of the system. This gives rise to an additional complication in real world experiments for which we only have access to a single trajectory.

C. A *null model* for the fluctuation-dissipation relation

In real-world applications we cannot compute ensemble averages. The common way to overcome such problem and reconcile data analysis with theory, is

through the assumption of ergodicity [83]. If the system $\boldsymbol{x}$ is ergodic it holds: $\overline{\mathcal{O}(\boldsymbol{x})} = \langle \mathcal{O}(\boldsymbol{x}) \rangle$ in the limit $T \to \infty$; where $\mathcal{O}(\boldsymbol{x})$ is a general observable, $\overline{\mathcal{O}(\boldsymbol{x})}$ indicates the time average and T is the length of the trajectory $\boldsymbol{x}$.

This is the main assumption behind any work in climate using fluctuation-dissipation theorem (see [53] and references therein). In this case, covariance functions are estimated using temporal averages, e.g. $C_{i,j}(\tau) = \overline{x_i(t+\tau)x_j(t)}$ (x_i is assumed to be zero mean). However, even in this case we are left with the problem of observing the system over a finite time window. Therefore we can always expect *spurious* results when estimating response functions. To the best of our knowledge, a clear statistical test to distinguish between *spurious* and *real* responses in the linear response theory formalism has not been proposed in the literature. Here we fill this void by proposing a *null* model for fluctuation-dissipation relation and derive its analytical solution. We start by proposing a null hypothesis for a general stochastic dynamical system.

a. Null hypothesis. Given a system $\boldsymbol{x}(t) = [x_1(t), x_2(t), ..., x_N(t)]$ it holds $R_{k,j}(\tau) = 0,\ \forall j,k = 1, ..., N;$ with $j \neq k$. In the context of causality this implies that there is no causal link $x_j \to x_k,\ \forall j,k = 1, ..., N; j \neq k$.

b. Null model. Given a process saved as a data matrix $\boldsymbol{x} \in \mathbb{R}^{N,T}$, we define a new process $\tilde{\boldsymbol{x}} \in \mathbb{R}^{N,T}$ simulated by a null model. Every time series in $\boldsymbol{x}$ and $\tilde{\boldsymbol{x}}$ are rescaled to zero mean. The null model takes the following form:

$$
\begin{aligned}
\tilde{\boldsymbol{x}}(t+1) &= \tilde{\boldsymbol{M}}\tilde{\boldsymbol{x}}(t) + \tilde{\boldsymbol{B}}\boldsymbol{\xi}(t) \\
\textit{with } \tilde{\boldsymbol{M}} &= \begin{pmatrix} \phi_1 & 0 & \cdots & 0 \\ 0 & \phi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \phi_N \end{pmatrix}; \\
\tilde{\boldsymbol{B}} &= \begin{pmatrix} \tilde{\sigma}_1 & 0 & \cdots & 0 \\ 0 & \tilde{\sigma}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \tilde{\sigma}_N \end{pmatrix}; \\
\xi_i(t) &\overset{\text{iid}}{\sim} \mathcal{N}(0,1),\ i = 1, ..., N.
\end{aligned}
\tag{6}
$$

Here, ϕ_i is the lag-1 autocorrelation of the "original" time series $x_i(t)$; $\tilde{\sigma}_i = \sigma_i(1-\phi_i^2)$ is the standard deviation of the Gaussian noise, where σ_i is the standard deviation of the "original" time series $x_i(t)$. Therefore, each time series $\tilde{x}_i(t)$ has the same mean, variance and lag-1 autocorrelation of $x_i(t)$, however every pair $\tilde{x}_i(t)$, $\tilde{x}_j(t)$ is now independent. Note that this test is largely inspired by the commonly adopted red noise test in climate analysis [84–87].

The matrix $\tilde{\boldsymbol{M}}$, defining the deterministic evolution, is diagonal; therefore at asymptotic times $T \to \infty$ there is no causal link among variables. However, for finite time windows, the response matrix estimated through time averaged covariance matrices as $\boldsymbol{R}(\tau) = \boldsymbol{C}(\tau)\boldsymbol{C}(0)^{-1}$ will give rise to *spurious* off-diagonal elements. The distribution of responses of the null process $\tilde{\boldsymbol{x}}$ defines confidence bounds of responses of the original process $\boldsymbol{x}$.

To compute the confidence level of the response $R_{k,j}(\tau)$ at each lag τ we first propose a numerical implementation. We then solve the problem analytically for the case $T >> 1$.

1. *Confidence bounds of the response matrix: numerical estimation*

Given a field $\boldsymbol{x} \in \mathbb{R}^{N,T}$, our goal is to provide an estimation of a confidence interval of the response matrix $\boldsymbol{R}(\tau)$ at each lag τ. This can be done as follows:

i) we generate a new process $\tilde{\boldsymbol{x}} \in \mathbb{R}^{N,T}$ using the null model proposed in Eq. 6.

ii) Estimate the response matrix $\boldsymbol{R}(\tau)$ of the null model $\tilde{\boldsymbol{x}}(t)$ for lags $\tau \in [0, \tau_\infty]$.

iii) Repeat the two steps above for B times, (B should be large, $B >> 1$), therefore creating an ensemble of *null* responses.

iv) For each lag τ we obtain a distribution of possible responses generated by the null model. This allows to estimate confidence bounds of responses by computing, for example, low and high quantiles of the distribution, or as chosen in this paper, multiples of its standard deviation.

2. *Confidence bounds of the response matrix: analytical derivation*

We note that the analytical form of the response matrix in the null model in Eq. 6 is trivial and given by $\boldsymbol{R}(\tau) = \boldsymbol{M}^\tau$ with entries $\phi_k^\tau \delta_{k,j}$; $\delta_{k,j}$ being the Kronecker delta. However, estimating responses from time series of finite length, will give rise to spurious results departing from the expected value of $\boldsymbol{M}^\tau$.

In this section we present the analytical probability distribution of the *estimated* (i.e., measured) responses $\boldsymbol{R}(\tau) = \boldsymbol{C}(\tau)\boldsymbol{C}(0)^{-1}$ in the case of time series of finite length generated by the *null* model in Eq. 6. We then refer the reader to Appendix B 2 a for the derivation.

The main assumption is that *null* responses $R_{k,j}(\tau)$ follow a Normal distribution. Therefore the expected value $\mathbb{E}[R_{k,j}(\tau)] = \langle R_{k,j}(\tau) \rangle$ and variance $\mathbb{V}\text{ar}[R_{k,j}(\tau)] =$

$\langle (R_{k,j}(\tau) - \langle R_{k,j}(\tau)\rangle)^2\rangle$ uniquely define the probability distribution $\rho(R_{k,j}(\tau))$. We have:

$$\begin{aligned}\mathbb{E}[R_{k,j}(\tau)] &= \phi_k^\tau \delta_{k,j} \\ \mathbb{V}\mathrm{ar}[R_{k,j}(\tau)] &= \frac{\phi_k^{2\tau}-1}{T} + \frac{2}{T}\Big(\frac{1-\phi_k^\tau\phi_j^\tau}{1-\phi_k\phi_j}\Big) \\ &\quad - \frac{2\phi_k^\tau}{T}\Big(\phi_k\frac{\phi_j^\tau-\phi_k^\tau}{\phi_j-\phi_k}\Big)\,.\end{aligned} \tag{7}$$

Finally, in the case $\phi_k = \phi_j$ we substitute the term $\phi_k \frac{\phi_j^\tau-\phi_k^\tau}{\phi_j-\phi_k}$ with the limit:

$$\lim_{\phi_j\to\phi_k} \phi_k \frac{\phi_k^\tau-\phi_j^\tau}{\phi_k-\phi_j} = \phi_k^\tau \tau. \tag{8}$$

Equation 7 assumes that each time series has been previously normalized to zero mean and unit variance. In the case of non-standardized time series $x_i(t)$ we need to account for contributions coming from the variances σ_i^2. This can be simply done by correcting the equation Eq. 7 as: $(\sigma_k^2/\sigma_j^2)\cdot$Eq. 7 (see also Eq. 15 in [1]).

In this paper, confidence bounds are always defined by $\mathbb{E}[R_{k,j}(\tau)] \pm 3\sqrt{\mathbb{V}\mathrm{ar}[R_{k,j}(\tau)]}$ (i.e., $\pm 3\sigma$ confidence level).

Finally, we note that the analytical confidence bounds proposed in Eq. 7 can potentially overcome an important problem in climate applications of linear response theory. Previous studies such as [51, 54, 82] focused on evaluating the integral $\int_0^\infty d\tau\ \boldsymbol{R}(\tau)$. In practice, the upper bound of the integral needs to be specified by a τ_∞ much larger than the characteristic time of the response. This has been commonly done by considering τ_∞ as low as 30 days, often in order to avoid spurious results for larger values. The confidence bounds proposed in this section can then be leveraged in order to neglect such spurious terms and study responses at longer time scales.

D. A simple example

We test these ideas in the context of a linear Markov model. We choose the same test model used in [1] in order to compare results and show differences between approaches. The system considered is the following:

$$\begin{aligned}\boldsymbol{x}(t+1) &= \boldsymbol{M}\boldsymbol{x}(t) + \boldsymbol{B}\boldsymbol{\xi}(t) \\ \textit{with } \boldsymbol{M} &= \begin{pmatrix} a & \epsilon & 0 \\ a & a & 0 \\ a & 0 & a \end{pmatrix}; \\ \boldsymbol{B} &= \begin{pmatrix} b & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & b \end{pmatrix}; \\ \xi_i(t) &\overset{\text{iid}}{\sim} \mathcal{N}(0,1),\ i=1,2,3.\end{aligned} \tag{9}$$

As in [1], we set $a = 0.5$ and $b = 1$; we then set $\epsilon = 0.04$. Note that here $[x_1, x_2, x_3]$ correspond to $[x, y, z]$ in [1]. In this simple model, a small perturbation applied on variable x_2 would propagate through the system and cause a change first at variable x_1 and then at x_3 [1]. However, a perturbation in x_3 cannot reach either x_1 and x_3, this is clear by looking at the underlying graph in Fig. 1(a). Both these links are correctly captured by the true responses (i.e., $\boldsymbol{M}^\tau$; shown in orange in Fig. 1) with the first nonzero response $R_{3,2}(\tau)$ (i.e., $x_2 \to x_3$) correctly captured at lag $\tau = 2$ and zero responses $R_{2,3}(\tau)$ (i.e., $x_3 \to x_2$) for any τ. As shown in [1], such results could not have been inferred with correlation analysis only.

Let us briefly note here the main conceptual difference between the fluctuation-response formalism and methods for causal discovery. Causal discovery methods used in climate and based on conditional independence such as [46] aim in discovering the underlying causal graph in Fig. 1(a) given time series data. Therefore, the link $x_2 \to x_3$ would not be identified as a causal link. The same holds for Granger causality and transfer entropy [32, 47] as shown in [1]. However, in a physical experiment an intervention over variable x_2 would cause a change in variable x_3. Such "interventional" view of causation is the one considered here and can be correctly captured by linear response theory as shown in Fig. 1(b). We refer to Section IIIA of [1] for an in-depth discussion.

In real-world cases we deal with time series with finite data. We then simulate the system for $T = 10^5$ time steps and estimate the causal links $x_j \to x_k$ with correlation functions (i.e., formula 5 after standardizing each x_i to unit variance) using temporal averages. As expected, in this case our results are affected by spurious terms, see blue lines in Fig. 1. The null model proposed in Eq. 6 is then leveraged to assign confidence bounds to the *estimated* responses. Responses inside the confidence bounds in Fig. 1 can be considered as spurious. The confidence bounds correctly identify the non-zero responses $R_{3,2}(\tau)$ for $\tau = 1$ and large lags as spurious results, see Fig. 1(b). Additionally, the test allows us to disregard the spurious link $x_3 \to x_2$, see Fig. 1(c). All responses $R_{k,j}(\tau)$, i.e. all links $x_j \to x_k$ are reported in Appendix C, Figure 5.

E. Metrics

The framework allows to identify any causal interaction $x_j \to x_k$ given the definition of causality presented in [1]. Given N time series this means $N(N-1)$ time-dependent links. Analyzing all interactions in such network gets rapidly out of hands with larger N; for example $N = 20$ would imply 380 time-dependent links. We then introduce a few metrics to analyze such causal graphs. In [1], the authors proposed a simple "cumulative degree of causation" of each link $x_j \to x_k$ as a Kubo formula [88].

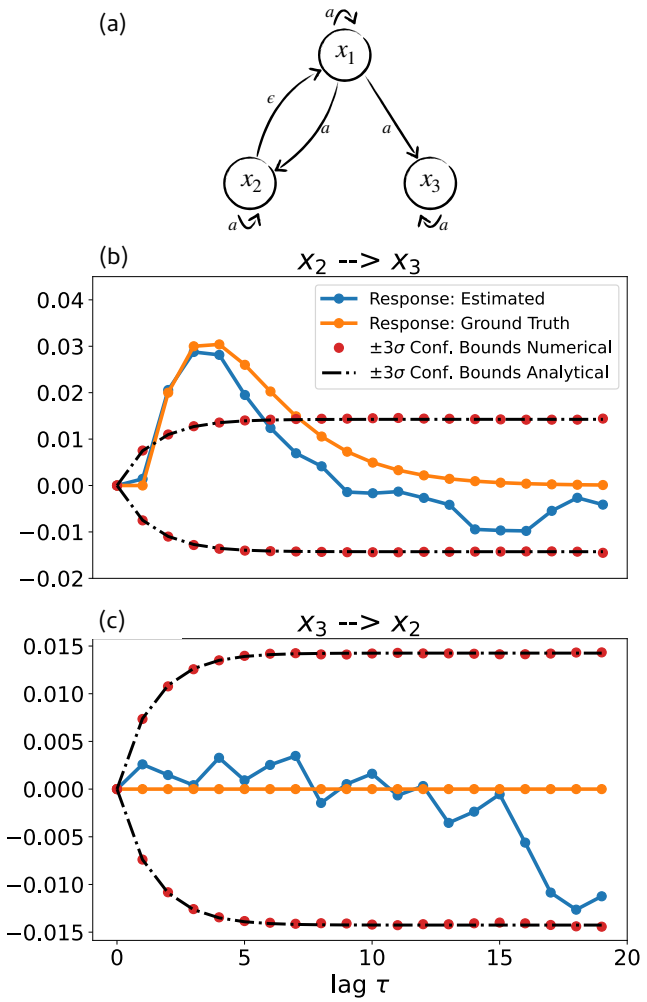

FIG. 1. Panel (a): Graph representing the Markov model in Eq. 9. This is the same simple system considered in Baldovin et al. (2020) [1] (where here $[x_1, x_2, x_3]$ correspond to their $[x, y, z]$ in [1]). Panel (b): response of variable x_3 when perturbing x_2, i.e. testing for link $x_2 \to x_3$. Panel (c): response of variable x_2 when perturbing x_3, i.e. testing for link $x_3 \to x_2$. Here all time series have been rescaled to zero mean and unit variance before computing responses. "Ground truth" of the response is computed as $\boldsymbol{R}(\tau) = \boldsymbol{M}^\tau$. Blue lines are responses estimated through temporal averages: for this step we use a long trajectory of length $T = 10^5$ simulated by system in Eq. 9. Red dots indicate the confidence bounds computed numerically using $B = 10^4$ ensemble members of the *null* model as shown in II C 1. The black dashed line is the analytical solution as in Eq. 7. Confidence bounds are defined correspond to $\pm 3\sigma$. All estimated responses (i.e. blue curves) in between the confidence bounds are here considered as spurious.

Here we consider the same formula while summing over the statistically significant responses $R_{k,j}(\tau^*)$, defined at lags τ^*. We compute respones $R_{k,j}(\tau)$ up to a maximum lag τ_∞; theoretically, the summation would be up to ∞, in practice we choose a τ_∞ much longer than the characteristic time of the response. The "cumulative degree of causation" considered here is then defined as follows:

$$\mathcal{D}_{j\to k} = \sum_{\tau^*}^{\tau_\infty} R_{k,j}(\tau^*) \tag{10}$$

Since responses can be negative and positive, the degree of causation such as in Eq. 10 can be zero even in the presence of causal links. It is therefore useful to consider a modified version of Eq. 10 by summing over the absolute value of responses as follows:

$$\mathcal{D}^*_{j\to k} = \sum_{\tau^*}^{\tau_\infty} \mid R_{k,j}(\tau^*) \mid \tag{11}$$

Eq. 10 (and its slight modification 11) quantifies the time-dependent strength of the causal link $x_j \to x_k$. It therefore allows to identify which variable x_k is influenced the most by perturbations on variable x_j: the largest $\mathcal{D}_{j\to k}$ (in absolute value) the strongest is the link $x_j \to x_k$.

Finally, we rank each variable x_j by defining its "causal strength" as follows:

$$\mathcal{D}_j = \sum_{k=1}^{N} \mathcal{D}^*_{j\to k} \;;\; j \neq k \tag{12}$$

Eq. 12 allows to rank nodes in the climate network in regards to their ability to *causally* influence other nodes. Informally, large values of $\mathcal{D}_j$ would mean that perturbations in x_j will be able to affect a large portion of the system.

Note that in case of comparisons with other datasets, $\mathcal{D}_{j\to k}$, $\mathcal{D}^*_{j\to k}$ can be normalized by $1/\tau_\infty$; $\mathcal{D}_j$ can be normalized by the number of variables as $1/(N-1)$. Furthermore, in case of datasets with different temporal resolutions it is possible to evaluate integrals of the kind $\int_0^{\tau_\infty} R_{k,j}(\tau^*)d\tau^*$ with a simple trapezoidal rule as commonly done in climate applications [51, 54]. These steps are not needed in this study.

Finally, for a given community/mode j identified by the strategy proposed in II A 2, it is possible to plot the cumulative causal links $\mathcal{D}_{j\to k}$ and $\mathcal{D}^*_{j\to k}$ (see Eq. 10 and 11) with any other community k as a map. Given a pattern j will refer to such map as "link map" $\mathcal{D}_{j\to k}$. Similarly, the "causal strength" $\mathcal{D}_j$ of each node j as defined in Eq. 12 can be plotted as a map, referred to as "strength map".

III. DATA

To explore and showcase the proposed causal framework we consider a long, stationary integration of the state-of-the-art coupled climate model GFDL-CM4 [89]. The ocean component of CM4, named MOM6, has an

horizontal grid spacing of 0.25° and 75 vertical layers [90]. The atmospheric/land component is the AM4 model [91, 92] with horizontal grid spacing of roughly 1° and 33 vertical layers. We consider the sea surface temperature field (SST) at global scale. The simulation considered, known as "piControl", is a 650 years long integration with constant CO_2 forcing set to preindustrial level. This allows to focus on a long, stationary climate trajectory. In this work we consider the last 300 years of this simulation. Even with stationary CO_2 forcing, the climate system can display variability at a vast range of time scales coming from the internal dynamics of the system. Importantly, especially at higher latitudes the system can display significant oscillations up to 10–100 years time scales, i.e. "multidecadal oscillations" [93]. Even in a 300 years long run such low frequency oscillations are sampled only a few times. Therefore, to simplify the interpretation of results, in this work we high-pass filter every time series with a cut-off frequency of $f = 1/(10 \text{ years})$ and focus on interannual variability only. Furthermore, the analysis will focus on SST anomalies only, after removing the seasonal cycle. In this study we consider temporal resolution of 1 month as a reasonable time scale to observe propagation of signals among modes of variability at global scale.

IV. CAUSALITY IN CLIMATE FIELDS

A. Applicability of fluctuation-response theory in climate studies

The main theoretical ideas justifying the application of methods in Section II B in climate, trace back at least to the work of Hasselman, K. (1976) [94]. The main intuition of the "Hasselman's program" [4] relies on thinking of processes with enough time scale separation between short and long time scales in terms of Brownian motion. This was first tested by Frankignoul and Hasselman (1977) [95] showing that the statistical properties of sea surface temperature (SST) variability can be in fact explained (at first order) by linear stochastic models with white noise representing the fast atmospheric variability. Such ideas were further explored and convingly demonstrated by Penland, C. (1989) [96] and Penland and Sardeshmukh (1995) [97] and motivated recent work on coupling functions as in [98] and [99].

The aforementioned studies justify the application of concepts introduced in Section II B to explore causality in climate fields. Specifically, this work will focus on the SST fields. Physically, this means that we will make the (rather strong) simplification of considering SST variability as a deterministic process and treat higher-frequency phenomena (e.g., atmospheric variability) as noise as done in [94]. Focusing only on sea surface temperature is however a limitation of this work and should be taken into account when analyzing the results. The extension to a multivariate framework is left for future work.

B. Relation to previous climate studies

We briefly present the main relationship between fluctuation-dissipation response studies investigated in the climate literature [45, 49, 51, 54, 82] and the causality framework explored here. Climate studies focused on studying the response $\delta\langle \boldsymbol{x}(t)\rangle$ of a dynamical system $\boldsymbol{x}$ perturbed by some (infinitesimally small) time-dependent forcing as follows:

$$\delta\langle \boldsymbol{x}(t)\rangle = \int_0^t d\tau\ \boldsymbol{R}(\tau)\delta \boldsymbol{f}(t-\tau). \tag{13}$$

Where $\boldsymbol{R}(t)$ is the linear response operator. In this study we consider stationary fields and *impulse* perturbations and therefore the forcing can be written as a delta function $\delta(t-\tau)$. In such case, Eq. 13 reduces to:

$$\delta\langle \boldsymbol{x}(t)\rangle = \int_0^t d\tau\ \boldsymbol{R}(\tau)\delta(t-\tau) = \boldsymbol{R}(t), \tag{14}$$

and the operator $\boldsymbol{R}(t)$ alone allows to study causal links.

In what follows, responses in Eq. 14 are computed by (a) using the quasi-Gaussian approximation as shown in Eq. 5 and (b) by first standardizing every time series to zero mean and unit variance; therefore the responses considered are computed using correlation functions (rather than covariances), equivalent to Eq. 15 in Baldovin et al. (2020) [1].

C. Application to global sea surface temperature

1. *Dimensionality reduction and causal inference*

We now focus on sea surface temperature (SST) variability at global scale. We consider the latitudinal range 60^oS-60^oN at 1^o resolution accounting for $N = 31141$ time series. The SST field is saved as monthly averages for 300 years for a total of $T = 3612$ time steps. Applying the community detection algorithm without the constraint proposed here, i.e. Eq. 1 without the requirement $d(i,j) \leq \eta$, will result in communities that are not spatially contiguous. This is shown in Fig. 2(a) where the Indian Ocean, eastern Pacific and a part of the Southern Ocean end up in the same pattern. In fact such distant regions can be linked by "teleconnection" patterns; for example at interannual time scales, Indian Ocean variability is forced by the tropical Pacific through an atmospheric wave response to El Niño events [12]. Consequently, variability in such regions is often grouped under the same cluster by community detection or clustering algorithms. In this case it is necessary to further constrain the graph inference step as shown in Eq. 1.

The dimensionality reduction of such graph identifies local and spatially contiguous patterns as shown in Fig. 2(b). Therefore, the additional constraint introduced in Eq. 1 is a simple but important step when coarse graining the system. This step allows to reduce the dimensionality from $N = 31141$ to $N = 19$ time series. Such communities are *regionally constrained*, therefore allowing us to answer the following question: how does the climate system respond to *local* perturbations? To answer such question, we leverage the tools presented in Section II B.

We consider the fluctuation-dissipation relation in its quasi-Gaussian approximation as shown in Eq. 5. In the Appendix, Section D we show that the time series of each community (i.e., mode) follows approximately a Gaussian distribution, therefore justifying the quasi-Gaussian approximation. We infer causality up to a $\tau_\infty = 10$ years and show the causal strength $\mathcal{D}_j$ (Eq. 12) in Fig. 2(c). The strongest mode of variability at interannual time scales is in the tropical Pacific, as expected [6]. Physically, this means that, at interannual time scales, the variability in the tropical Pacific is able to influence a larger part of the world compared to other regions with smaller strength. In what follows we are going to refer to this region as "ENSO region".

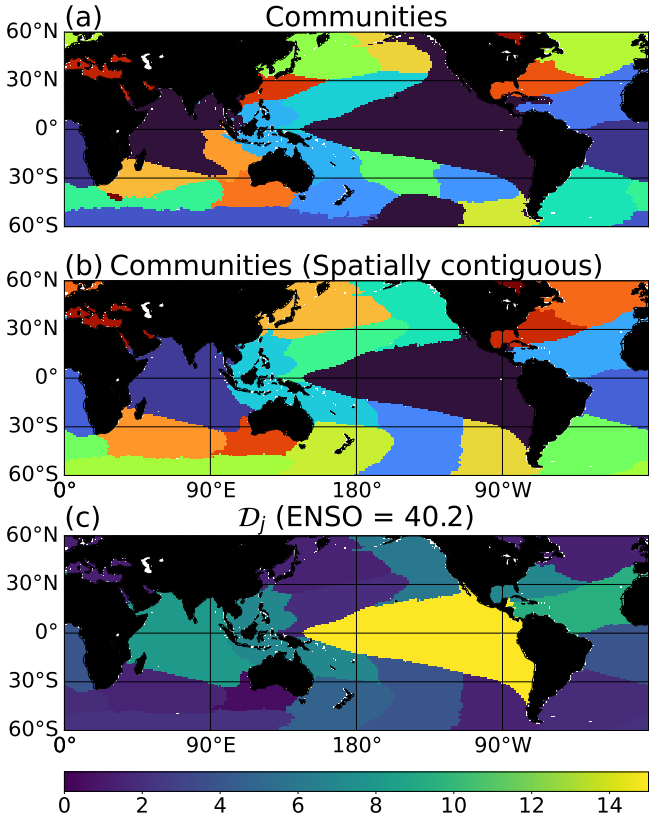

FIG. 2. Community detection of global sea surface temperature in the latitude range [60°S-60°N] and at monthly temporal resolution. Panel (a): an undirected graph is inferred through Eq. 1 but without the proposed constraint $d(i,j) \leq \eta$. Then the community detection method Infomap is applied. Panel (b): same as panel (a) but the undirected graph is inferred through the newly proposed Eq. 1. Panel (c): causal strength as defined by 12. As expected the "ENSO" region is the strongest mode in the inferred causal network. Its strength is reported in the plot title. The response functions are computed up to $\tau_\infty = 10$ years. Only the statistical significant responses contribute to the strength map. Confidence bounds are quantified through Eq. 7 at the $\pm 3\sigma$ level.

2. *Investigation of few causal interactions*

We further analyze the links between three components of the system. Specifically, we focus on the interaction of ENSO, the Indian Ocean (IO) and South Tropical Atlantic (STA). ENSO is known to drive climate variability outside the tropical Pacific through teleconnection patterns and has been studied in many contributions. The way in which Indian and Atlantic variability drive SST in the Pacific has been less appreciated in the past and it is currently debated in the community [100]. Quantification of such linkages is important to better understand climate variability and to improve seasonal forecasting.

During an El Niño phase, the anomalous temperature in the tropical Pacific excites waves in the atmosphere. Such waves, known as eastward-propagating Kelvin and westward-propagating Rossby waves, drive changes in temperature in the whole tropical band [12]. Such causal links are identified in Fig. 3(a,b), with positive responses of both the IO and STA regions to perturbations in the ENSO regions. As expected such positive lead of ENSO is the strongest in magnitude and much larger than the other responses in Fig. 3. Interestingly, we find a (weak) negative link between ENSO and IO in Fig. 3(b) around $\tau = 30$ months, suggesting the emergence of positive (negative) anomalies in the Indian Ocean ~ 3 years after La Niña (El Niño) events. The positive response around 10 years in Fig. 3(b) is here considered as a False Positive.

Fig. 3(c) shows that the positive (negative) anomalies in the STA region, mainly linked to the dynamics of the Atlantic Niño [101] (see also discussion in [17]), leads *on average* to the development of La Niña (El Niño) conditions as recently argued in the literature [102–104].

The IO pattern in our study (see pattern z in Figure 3) mainly identifies what is known as the Indian Ocean Basin (IOB) mode [65]. The IOB mode has been traditionally considered as simply forced by ENSO. Nonetheless, recent studies have revealed how the IOB can also drive ENSO variability. Specifically, it has been demonstrated how a strong IOB warming can in

fact contribute to central Pacific cooling further driving a transition to a La Niña state [100, 105, 106]. Such negative link is correctly identified by the proposed framework (see Fig. 3(d)) but does not show up in correlation-only analyses (see for example Fig. 11(b) in [68]).

As discussed also in [100] these results suggest an increase in potential predictability of ENSO variability when considering the non-local interactions with the Indian Ocean and tropical Atlantic basins.

Finally, in Fig. 4 we show the link maps for four modes: ENSO region, Indian Ocean (IO), South and North Tropical Atlantic (STA and NTA respectively). Such maps show values of $\mathcal{D}_{j\to k}$ (Eq. 10) up to a $\tau_\infty = 6$ months. Fig. 4(a) quantifies the cumulative response of any region given perturbations in the ENSO region. We notice that such map is qualitatively similar to the first Empirical Orthogonal Function of global SST (see for example Fig. 4 in [107]). The framework allows to examine causal linkages from/to any region of the system. Figures 4(b,c,d) show the cumulative degree of causation respectively from IO, STA and NTA regions to any other region in the world. In other words, such maps allow to summarize the cumulative response of the whole globe, given small, local perturbations to any region x_j of choice.

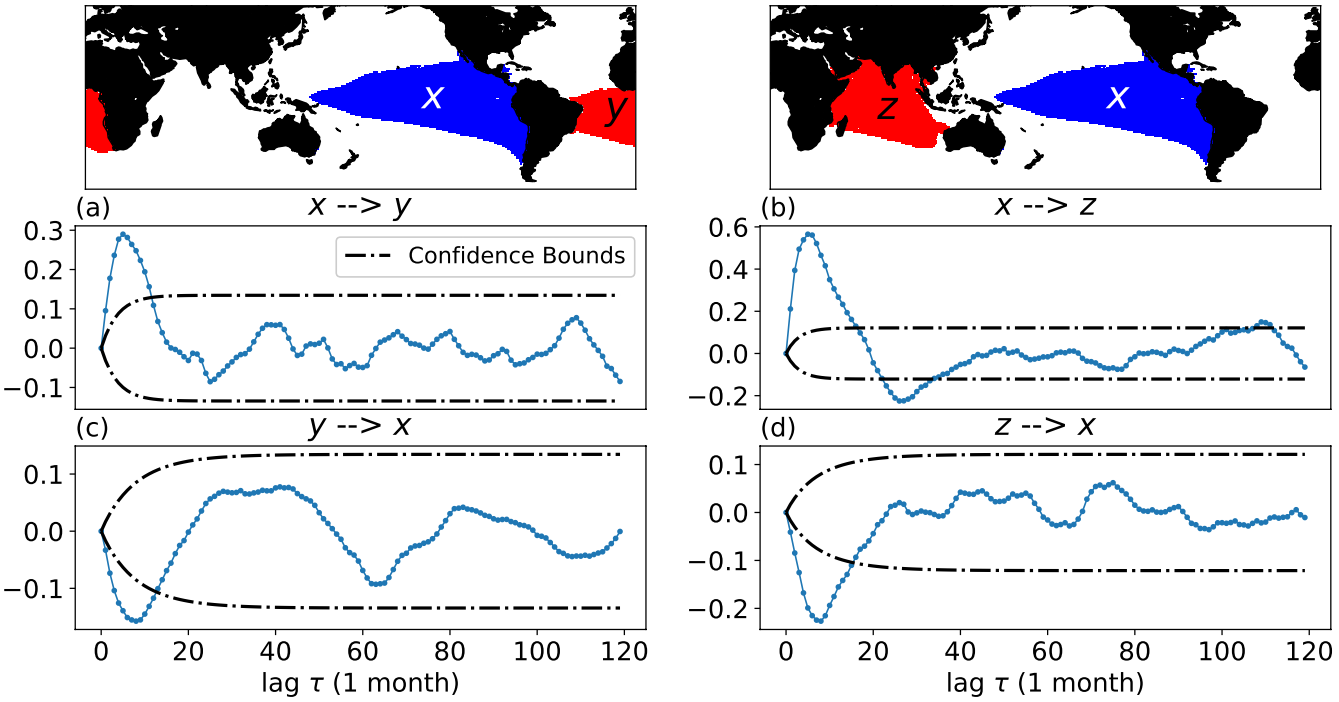

FIG. 3. x: ENSO mode. y: South Tropical Atlantic. z: Indian Ocean. Panel (a,c): causal link $x \to y$ and $y \to x$. Panel (b,d): causal link $x \to z$ and $z \to x$. Response functions have been computed up until $\tau_\infty = 10$ years. Confidence bounds are quantified through Eq. 7 at the $\pm 3\sigma$ level. Responses in between the confidence bounds are here considered as spurious.

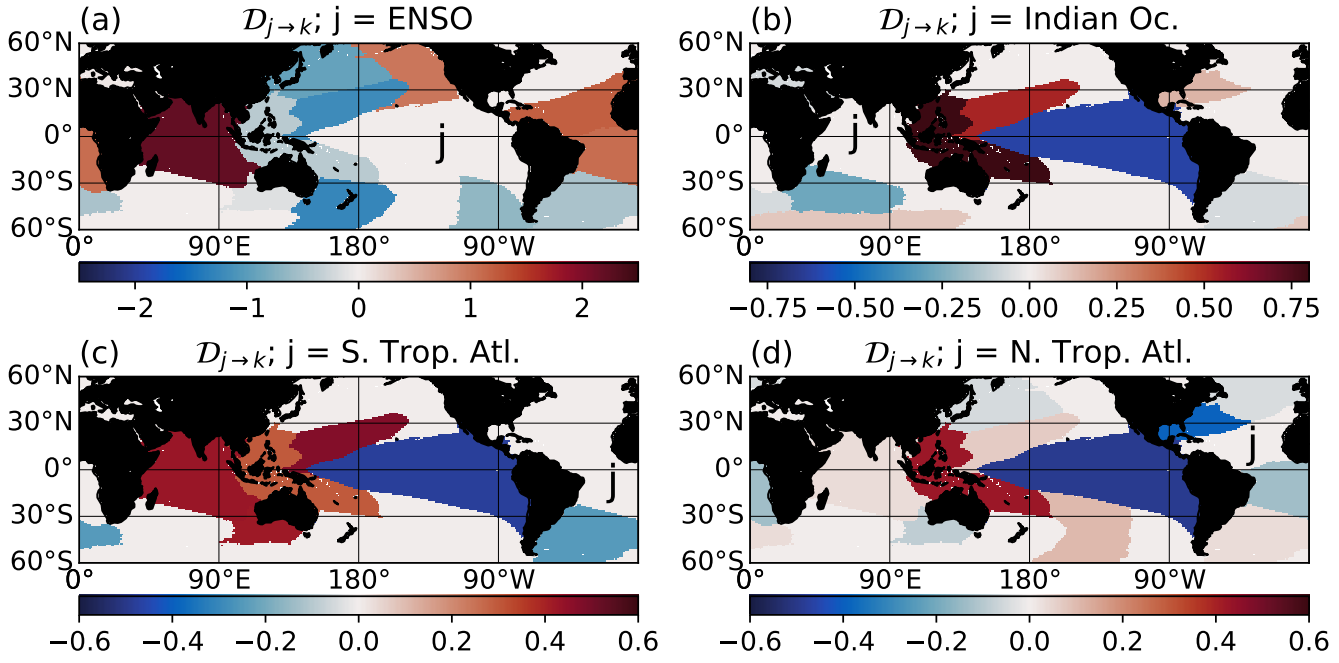

FIG. 4. Link maps $\mathcal{D}_{j\to k}$ for all k, as computed in 10 and considering only up to $\tau_\infty = 6$ months. Regions j considered are ENSO region, Indian Ocean, South and North Tropical Atlantic in panels (a,b,c,d) respectively. The first Empirical Orthogonal Function roughly correspond to the ENSO link map in panel (a). Only the statistical significant responses contribute to the causal link maps. Confidence bounds are quantified through Eq. 7 at the $\pm 3\sigma$ level.

V. CONCLUSIONS AND DISCUSSION

We introduced a novel framework for causal inference in spatiotemporal climate fields. The causal inference step, based upon ideas of Baldovin et al. [1], and in line with [21, 44], frames the problem of causality in the formalism of linear response theory [88]. Here, we further developed these ideas by proposing an analytical *null* model for the fluctuation-dissipation relation. The model allows to distinguish between true and spurious response functions estimated from finite data, with applicability not restricted to climate. The time-dependent causal graph is inferred after coarse-graining the system. This step, based on community detection, is used to reduce the dimensionality of a spatiotemporal field in terms of *regional* "modes" of variability. Such "modes" are defined as regionally constrained sets of time series with large average pairwise correlation. The dimensionality reduction and the causal inference steps allow to study how *local* perturbations can propagate through the system and impact *remote* locations.

We discuss a few important limitations and caveats that may hinder interpretations of results in future studies.

a. The case of hidden variables. The fluctuation-dissipation formalism identifies causal links when we have access to the whole state vector $\boldsymbol{x}$. This is often not the case. This is a problem common to every causal inference method. A "solution" is to include the important variables for the phenomena we want to explain. In this work, we based our analysis on sea surface temperature (SST) building on ideas first proposed by Hasselman, K. (1977) [95] where the fast atmospheric variability can be considered as noise, forcing the (slower) deterministic ocean dynamics. This is a great simplification and should be considered when interpreting results. The question of how many variables are enough to consider the system as Markovian is an old problem with warnings discussed at least since Onsager and Machlup (1953) [108]; see also Section IVB in [1]. Quite interestingly, [1] also showed that applying Takens theorem [109] to reconstruct the state space vector may not always help. The main reason is that Takens embedding theorem, proven for deterministic systems [109], fails for general stochastic processes [1].

b. Computation of the inverse covariance matrix $\boldsymbol{C}(0)^{-1}$. Consider a dynamical system $\boldsymbol{x} \in \mathbb{R}^{N,T}$, N is its dimensionality and T is the length of each time series $x_i(t)$. If $N > T$, the covariance matrix $\boldsymbol{C}(0) \in \mathbb{R}^{N,N}$ will not be full rank, and therefore it will not have an inverse. Generally, the covariance matrix can be ill-conditioned and the computation of the inverse $\boldsymbol{C}(0)^{-1}$ will result in large errors. This point has been described in [51, 54] and more formally in [110, 111] in the context of the fluctuation-response formalism; but it is a general problem in many fields, see for example [112, 113]. Therefore, the proposed framework should be applied for systems $\boldsymbol{x} \in \mathbb{R}^{N,T}$ with $T >> N$, i.e., the number of samples much larger than the dimensionality of the system. As a simple test, when computing responses with the quasi-Gaussian approximation $\boldsymbol{R}(\tau) = \boldsymbol{C}(\tau)\boldsymbol{C}(0)^{-1}$ we recommend to check $\boldsymbol{R}(0) = \boldsymbol{I}$ (at least up to a certain numerical accuracy), $\boldsymbol{I}$ being the Identity matrix. In general, dimensionality reduction steps (as proposed in this paper) allow to reduce the number of time series N to values much smaller than T, allowing for trustworthy computations of $\boldsymbol{C}(0)^{-1}$ [51].

c. Quasi-Gaussian approximation. The quasi-Gaussian approximation considered in this study (see Eq. 5) has been shown to work especially well in many climate applications, see [53] and references therein. However, generally, we suggest checking the data's underlying probability distribution before the analysis. This may be important, especially for paleoclimate applications where climate variability shows a vast range of spectral peaks with no clear time-scale separation. An example is the work shown in [114], where the authors analyzed the causal link between CO_2, temperature (T) and insolation in the last 800 kyr. Distributions of both CO_2 and T in the last 800 kyr are strongly non-Gaussian. The solution was to high-pass filter the data and focus on high-frequency variability, with the hypothesis of slow time scales being linked to the external forcing and faster time scales to the internal system's variability. This was shown to be enough to recover Gaussian distributions [114]. In this work, we have shown that the distributions of the time series analyzed can be reasonably approximated by Gaussians (see Appendix D) justifying the application of the methodology shown in Section II B. A generalization to nonlinear systems is provided by formula 3, as long as the probability distribution $\rho(\boldsymbol{x})$ is known. In specific cases, we note that it may be possible to apply transformations to strongly non-Gaussian fields and still use the quasi-Gaussian approximation explored here. An example is the precipitation field, where a logarithmic scaling can help recover Gaussian-like distribution [115].

The methodology proposed here can be potentially applied to study the dynamics of any climate field, at least given the assumptions and limitations listed above. It serves as a useful, rigorous framework to simplify the description of complex, high-dimensional dynamical systems in terms of few entities and their linkages, aiming to better understand the system's dynamics. Differently from other methods for causal inference adopted in climate, it scales to high-dimensional datasets. Moreover, the method and the proposed *null* model have a clear physical interpretation and can be formalized via analytical formulas. This allows to infer causality avoiding many heuristics and parameters.

The application explored here in Section IV C allowed us to detect well-known links in climate, such as the influence of tropical Pacific variability onto other basins, as well as other linkages, such as the lead of sea surface temperature variability in the Indian Ocean to the Pacific basin, which received less attention in the literature [100]. Additionally, we showed how the "strength maps" and "link maps" as shown in Fig. 2(c) and Fig. 4 summarize cumulative causal interactions across time and space in a comprehensive and interpretable way.

We focused on the sea surface temperature field as the statistics of modes of variability and their linkages in this field have been investigated in many previous studies, therefore offering a good test case for the methodology. Importantly, previous studies have focused on a few modes at a time. Here we showed that the methodology allows to study causal linkages among regions in a comprehensive framework, where all modes of variability and their interactions are studied simultaneously.

Examples of future work range from quantification of drivers of sea level change, such as basin-scale adjustments in the North Atlantic driven by Rossby waves, to studying the evolution of climate modes and their linkages in paleoclimate simulations, with time-dependent orbital and trace-gases forcings (e.g., [16]), to non-local drivers of precipitation. Additionally, the proposed framework offers a way to evaluate new generations of climate models in terms of their emergent causal structure rather than statistical properties only; for example, by assessing the impact of new sub-grid parametrizations onto the large scale dynamics.

ACKNOWLEDGMENTS

FF acknowledges helpful discussions with Marco Baldovin, Simone Contu, Andre Souza, Pedram Hassanzadeh and Aurora Basinski. FF also thanks Martin Rosvall for clarifications on the Infomap methodology at the start of this work. This work was supported in part by NOAA grant NOAA-OAR-CPO-2019-2005530, by the KITP Program "Machine Learning and the Physics of Climate" supported by the National Science Foundation under Grant No. NSF PHY-1748958 and by Schmidt Futures, a philanthropic initiative founded by Eric and Wendy Schmidt, as part of its Virtual Earth System Research Institute (VESRI).

CODE AVAILABILITY

Codes and materials are available at https://github.com/FabriFalasca/Linear-Response-and-Causal-Inference.

Appendix A: Dimensionality reduction in climate. Limitations of current methods and proposal

1. Two goals in dimensionality reduction studies

We note that the use of dimensionality reduction in applications of linear response theory can be leveraged with (at least) two different goals in mind. In the case of very high dimensional systems as for General Circulation Models (GCM), applications of the fluctuation-dissipation response formalism is practically impossible. The usual solution in the climate literature has been to construct response operators in a low-dimensional space spanned by many Empirical Orthogonal Functions (EOFs or Principal Components) [116]; usually, order 10^3 EOFs in order to explain at least 90% of the total variance. Results computed in the low dimensional space are then transformed back to the original space [51, 52, 54]. This computational strategy has been shown to be successful in many applications (see [51, 52]). A second possible goal of dimensionality reduction is to simplify the problem in hand in terms of very few components and apply the linear response formalism directly on those entities. In this case we are interested in studying directly the coarse-grained version of the system. This adds to interpretability and to a first order understanding of the system's dynamics. This second case is the one considered in this paper.

2. Few limitations of common methods and proposal

Traditionally, dimensionality reduction in climate studies is done through Principal Component Analysis (PCA) [116]. PCA, or Empirical Orthogonal Function (EOF) analysis [117] as a useful, first order way to reduce the dimensionality of the system based on the singular value decomposition (see e.g., [118]) of the data matrix. However, the resulting patterns suffer from few drawbacks: first, EOFs are orthogonal by definition. Such constraint hamper their interpretation and make it difficult to distinguish between physical or purely statistical modes [79, 119]. A possible solution has been to rotate the EOFs, such as in [120]. Rotated-EOFs have been found to be sensitive to the rotation criterion, normalizations and number of loadings (see [79, 121]).

Another drawback comes from linearity. Manifold learning algorithms aim in addressing this issue by identifying low-dimensional representations of a high-dimensional system accounting for non-linearities (curved manifolds) [122]. Examples range from the Isomap algorithm [123] to the more recent t-SNE [124], UMAP [125] to the PHATE algorithm [126] and ROCK-PCA [127]. Finally, deep learning tools such as autoencoders can be explored for dimensionality reduction [128] and found applications in climate science [129].

Dependent on the goal in mind (see Section A 1), a possible limitation shared by all these tools when applied to global climate data is that they decompose a field in terms of *global* (in longitude-latitude maps) modes. However, physically, climate dynamics can be often thought of as a set of *remote* connections driven by *local* phenomena (perturbations). Given so, common practice in climate science has been to define "climate indices" as time series averaged in specified regions (i.e., "boxes"). Known examples are the Niño3.4, the Indian Ocean Dipole (IOD) index etc. However, a framework for automated identification of proxies of such indices is needed as the locations of such regions, or "boxes", may be not relevant for the study of future (or past) climates. An example can be found in [16, 130, 131] where the authors showed the emergence of an El Niño-like variability in the Indian Ocean during the Last Glacial Maximum, the last 6000 years and in future projections. In this sense, known indices identified in the current climate are potentially less meaningful in past and future climates.

A method proposed to do so is δ-MAPS [67]. Given a climate fields, δ-MAPS identifies spatially contiguous clusters. The method has proven to be useful in climate studies with applications ranging from model evaluation [68, 69], shifts in climate modes in the last 6000 years [16, 19], sea level budget at regional scale [70], marine ecology [71] and ecosystem dynamics [72]. In the case of relatively low dimensional fields (e.g., global fields at 2° by 2° spatial resolution) δ-MAPS shows excellent performance. However, a known drawback is that it does not scale well with high-dimensional datasets (i.e., large number of grid cells). Additionally, it can show sensitivity to one of its parameters in the domain identification stage, so that often, many exploratory tests are needed to explore sensitivity.

When working with very high dimensional fields, it is often useful to consider fast and scalable algorithms. In the last two decades, climate data analysis have focused on fast methodologies stemming from the complex network literature [74]. An example is the work of [79] where the authors focused on the community detection method "Infomap" [76, 77, 132] to identify communities in the HadISST [133] sea surface temperature dataset. Such methods allow to find patterns that are not necessarily orthogonal. Furthermore, they are fast, memory efficient and scale well with the dimensionality of the dataset. The main issue is that, similar to manifold learning algorithms, community detection algorithms are not constrained to be spatially contiguous.

In this paper we showed that adding a simple constraint on spatial distances is enough to enforce the identification of "local" communities (see Section II A 1). This allows to leverage computationally fast and robust methods such as community detection for dimensionality reduction strategies in climate. Differently from δ-MAPS [67], the identified communities cannot overlap with each other. We find however that conclusions found in previous studies using δ-MAPS (see [68] for example) do not strongly depend on clustering overlapping. The framework proposed here in Section II A is then leveraged as a much simpler (and therefore more robust), practical framework to the problem of identification of *regionally constrained* modes.

Appendix B: A *null* model for the Fluctuation-Dissipation relation. Analytical derivation of the confidence bounds

This work proposes a novel null model for the Fluctuation-Dissipation relation (see 6). In the null model, every variable x_j and x_k is independent, and therefore the expected value of each response $\mathbb{E}[R_{k,j}(\tau)] = 0$ for $j \neq k$ by construction. Nonetheless, estimating such responses by $\boldsymbol{R} = \boldsymbol{C}(\tau)\boldsymbol{C}(0)^{-1}$ (see II B 2) using time series of finite length T simulated by the null model, will give rise to spurious results diverging from the expected value $\mathbb{E}[R_{k,j}(\tau)]$. In Eq. 7 of the main text we showed the analytical probability distribution of $R_{k,j}(\tau)$. The main assumption in this derivation is that responses $R_{k,j}(\tau)$ follow a Normal distribution. Therefore the expected value $\mathbb{E}[R_{k,j}(\tau)]$ and variance $\mathbb{V}\mathrm{ar}[R_{k,j}(\tau)]$ uniquely define the probability density $\rho(R_{k,j}(\tau))$. Here we present the derivation of such formula.

1. Notation adopted in this section

In order to simplify and ease the derivation, it is useful to adopt a simpler and more appropriate statistical formalism. The symbols adopted in this section relate to the ones used in the previous ones as follows: $\mathbb{E}[X] = \langle X \rangle$ represents the expected value of a random variable X. This is equal to the ensemble average considered in the previous sections. Consequently, $\mathbb{V}\mathrm{ar}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2]$ represents the variance of a random variable X. Finally, $\mathbb{C}\mathrm{ov}[X, Y] = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$ represents the covariance of two random variables X and Y. We are going to refer to the null process as $\boldsymbol{x} = [x_1(t), x_2(t), ..., x_N(t)]$ (rather than $\tilde{\boldsymbol{x}}$ as in 6). Finally, each time series $x_j(t)$ is here considered to be scaled to zero mean and unit variance. This step greatly simplifies the derivation. At the end of this section, we provide the general formula for processes that are not unit-variance.

2. Analytical derivation

Consider a long trajectory $\boldsymbol{x} \in \mathbb{R}^{N,T}$ defined by the forward iteration of the *null* model in Eq. 6. The *true* mean, and covariances at lag τ of each individual time series in $\boldsymbol{x}$ are given by $\mathbb{E}[x_j(t)] = 0$ and $\mathbb{E}[x_k(t+\tau)x_j(t)] = \phi_k^\tau \delta_{k,j}$ respectively. Where ϕ_k is the lag-1 autocorrelation of time series $x_k(t)$ and the Kronecker delta $\delta_{k,j}$ differs from zero only in the case $j = k$.

We note that the numerical estimation of both $\boldsymbol{C}(\tau)$ and $\boldsymbol{C}(0)^{-1}$ will lead to spurious terms in $\boldsymbol{R}(\tau)$. We then rewrite the covariance matrix $\boldsymbol{C}(\tau)$ estimated through time averages as a sum of the expected value $\mathbb{E}[\boldsymbol{C}(\tau)]$ plus some small Gaussian residual $\hat{\boldsymbol{C}}(\tau)$ as:

$$\boldsymbol{C}(\tau) = \mathbb{E}[\boldsymbol{C}(\tau)] + \hat{\boldsymbol{C}}(\tau) = \boldsymbol{D}_\phi^\tau + \hat{\boldsymbol{C}}(\tau). \tag{B1}$$

Where $\boldsymbol{D}_\phi^\tau$ is a diagonal matrix with component (i,j) defined as $(\boldsymbol{D}_\phi^\tau)_{i,j} = \phi_i^\tau \delta_{i,j}$. The decomposition (Eq. (B1)) applies to the matrix $\boldsymbol{C}(0)$ as well with $\boldsymbol{D}_\phi^0 = \boldsymbol{I}$ where $\boldsymbol{I}$ is the Identity matrix. The main difficulty is that we are not interested in $\boldsymbol{C}(0)$ but in its inverse $\boldsymbol{C}(0)^{-1}$. By assuming relatively small residuals (true for time series with $T >> 1$), we can approximate an inverse of the estimated covariance matrix $\boldsymbol{C}(0)^{-1}$ using Neumann series [134] as:

$$\boldsymbol{C}(0)^{-1} = (\boldsymbol{I} + \hat{\boldsymbol{C}}(0))^{-1} \approx \boldsymbol{I} - \hat{\boldsymbol{C}}(0). \tag{B2}$$

Where we only retained the first term in the Neumann series. An estimator of the null response $\boldsymbol{R}(\tau) = \boldsymbol{C}(\tau)\boldsymbol{C}(0)^{-1}$ can be then written as

$$\boldsymbol{R}(\tau) = \boldsymbol{C}(\tau)\boldsymbol{C}(0)^{-1} \approx \boldsymbol{C}(\tau) + \boldsymbol{D}_\phi^\tau(\boldsymbol{I} - \boldsymbol{C}(0)). \tag{B3}$$

Where we neglected the term $\hat{\boldsymbol{C}}(\tau)\hat{\boldsymbol{C}}(0)$, a reasonable step in the presence of small residuals, true for time series with length $T >> 1$. To derive the statistical properties of the estimator in Eq. B3, it is useful to rewrite such formula in terms of each component j and k.

$$R_{k,j}(\tau) \approx C_{k,j}(\tau) + \delta_{k,j}\phi_k^\tau - \phi_k^\tau C_{k,j}(0). \tag{B4}$$

The final step is to derive the expected value $\mathbb{E}[R_{k,j}(\tau)]$ and $\mathbb{V}\mathrm{ar}[R_{k,j}(\tau)]$ of Eq. B4, thus uniquely defining the probability distribution of $R_{k,j}(\tau)$, under the assumption of Gaussian statistics.

a. *Expected value and variance of the response estimator*

The expectation of the response estimator proposed in B4 can be derived as

$$\begin{aligned}\mathbb{E}[R_{k,j}(\tau)] &= \mathbb{E}[C_{k,j}(\tau)] + \delta_{k,j}\phi_k^\tau - \phi_k^\tau \mathbb{E}[C_{k,j}(0)] \\ &= \delta_{k,j}\phi_k^\tau + \delta_{k,j}\phi_k^\tau - \phi_k^\tau \delta_{k,j} \\ &= \delta_{k,j}\phi_k^\tau.\end{aligned} \tag{B5}$$

The variance of the response estimator proposed in B4 can be derived as

$$\begin{aligned}\mathbb{V}\mathrm{ar}[R_{k,j}(\tau)] &= \mathbb{V}\mathrm{ar}[C_{k,j}(\tau) - \phi_k^\tau C_{k,j}(0)] \\ &= \mathbb{V}\mathrm{ar}[C_{k,j}(\tau)] + \phi_k^{2\tau}\mathbb{V}\mathrm{ar}[C_{k,j}(0)] \\ &- 2\phi_k^\tau \mathbb{C}\mathrm{ov}[C_{k,j}(\tau), C_{k,j}(0)].\end{aligned} \tag{B6}$$

We remind the reader the following useful equality: the covariance $\mathbb{C}\mathrm{ov}[X,Y]$ of two random variables X and Y can be rewritten as $\mathbb{C}\mathrm{ov}[X,Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$. We now compute the variance of the response estimator in Eq. B6. To do so, we first need to provide an expression to terms $\mathbb{V}\mathrm{ar}[C_{k,j}(\tau)]$ and $\mathbb{C}\mathrm{ov}[C_{k,j}(\tau), C_{k,j}(0)]$. Such terms can be computed as follows:

$$\begin{aligned}\mathbb{V}\mathrm{ar}[C_{k,j}(\tau)] &= \mathbb{E}[C_{k,j}(\tau)C_{k,j}(\tau)] - \delta_{k,j}\phi_k^{2\tau} \\ &= \frac{1}{T^2}\sum_{t',t''=1}^{T} \mathbb{E}[x_k(t'+\tau)x_j(t')x_k(t''+\tau)x_j(t'')] - \delta_{k,j}\phi_k^{2\tau} \\ &= \frac{1}{T^2}\sum_{t',t''=1}^{T} \Big(\mathbb{E}[x_k(t'+\tau)x_k(t''+\tau)]\mathbb{E}[x_j(t')x_j(t'')] \\ &+ \mathbb{E}[x_k(t'+\tau)x_j(t')]\mathbb{E}[x_k(t''+\tau)x_j(t'')] \\ &+ \mathbb{E}[x_k(t'+\tau)x_j(t'')]\mathbb{E}[x_j(t')x_k(t''+\tau)]\Big) - \delta_{k,j}\phi_k^{2\tau} \\ &= \frac{1}{T^2}\sum_{t',t''=1}^{T} \left(\phi_k^{|t'-t''|}\phi_j^{|t'-t''|} + \delta_{k,j}\phi_k^{2\tau} + \delta_{k,j}\phi_k^{|t'+\tau-t''|}\phi_k^{|t'-\tau-t''|}\right) - \delta_{k,j}\phi_k^{2\tau} \\ &= \frac{1}{T^2}\sum_{t',t''=1}^{T} \left(\phi_k^{|t'-t''|}\phi_j^{|t'-t''|} + \delta_{k,j}\phi_k^{|t'+\tau-t''|}\phi_k^{|t'-\tau-t''|}\right).\end{aligned} \tag{B7}$$

$$
\begin{aligned}
\mathbb{C}\mathrm{ov}[C_{k,j}(\tau), C_{k,j}(0)] &= \mathbb{E}[C_{k,j}(\tau)C_{k,j}(0)] - \delta_{k,j}\phi_k^\tau \\
&= \frac{1}{T^2}\sum_{t',t''=1}^{T} \mathbb{E}[x_k(t'+\tau)x_j(t')x_k(t'')x_j(t'')] - \delta_{k,j}\phi_k^\tau \\
&= \frac{1}{T^2}\sum_{t',t''=1}^{T} \Big(\mathbb{E}[x_k(t'+\tau)x_k(t'')]\mathbb{E}[x_j(t')x_j(t'')] \\
&+ \mathbb{E}[x_k(t'+\tau)x_j(t')]\mathbb{E}[x_k(t'')x_j(t'')] \\
&+ \mathbb{E}[x_k(t'+\tau)x_j(t'')]\mathbb{E}[x_j(t')x_k(t'')]\Big) - \delta_{k,j}\phi_k^\tau \\
&= \frac{1}{T^2}\sum_{t',t''=1}^{T} \Big(\phi_k^{|t'+\tau-t''|}\phi_j^{|t'-t''|} + \delta_{k,j}\phi_k^\tau + \delta_{k,j}\phi_k^{|t'+\tau-t''|}\phi_k^{|t'-t''|}\Big) - \delta_{k,j}\phi_k^\tau \\
&= \frac{1}{T^2}\sum_{t',t''=1}^{T} \Big(\phi_k^{|t'+\tau-t''|}\phi_j^{|t'-t''|} + \delta_{k,j}\phi_k^{|t'+\tau-t''|}\phi_k^{|t'-t''|}\Big).
\end{aligned}
\tag{B8}
$$

The computation of Equations B7 and B8 requires to compute the following three terms: $\sum_{t',t''=1}^{T}\phi_k^{|t'-t''|}\phi_j^{|t'-t''|}$, $\sum_{t',t''=1}^{T}\phi_k^{|t'+\tau-t''|}\phi_k^{|t'-\tau-t''|}$ and $\sum_{t',t''=1}^{T}\phi_k^{|t'+\tau-t''|}\phi_j^{|t'-t''|}$. To solve such terms we point out that a summation of type $\sum_{t',t''=1}^{T}(\phi_k\phi_j)^{|t'-t''|}$ will result in T points with value $(\phi_k\phi_j)^0$, $2(T-1)$ points with value $(\phi_k\phi_j)^1$ up to $2(T-t)$ points with value $(\phi_k\phi_j)^t$. The summation can be then rewritten as: $\sum_{t',t''=1}^{T}(\phi_k\phi_j)^{|t'-t''|} = T + \sum_{t=1}^{T-1}(\phi_k\phi_j)^t 2(T-t)$. Similar reasoning can be applied for all the terms above.

b. Computation of each summation

$$
\begin{aligned}
\texttt{Sum(I)}: \sum_{t',t''=1}^{T}\phi_k^{|t'-t''|}\phi_j^{|t'-t''|} &= T + \sum_{t=1}^{T-1}(\phi_k\phi_j)^t 2(T-t) \\
&= \frac{T - T(\phi_k\phi_j)^2 + 2(\phi_k\phi_j)(\phi_k^T\phi_j^T - 1)}{(-1+\phi_k\phi_j)^2}.
\end{aligned}
\tag{B9}
$$

$$
\begin{aligned}
\texttt{Sum(II)}: \sum_{t',t''=1}^{T}\phi_k^{|t'+\tau-t''|}\phi_j^{|t'-\tau-t''|} &= \sum_{t=1-T}^{T-1}\phi_k^{|t+\tau|}\phi_j^{|t-\tau|}(T-\mid t\mid) \\
&= \underbrace{\sum_{t=1}^{T-1}\phi_k^{(t+\tau)}\phi_j^{|t-\tau|}(T-t)}_{\texttt{Sum(a)}} \\
&+ \underbrace{\sum_{t=1-T}^{0}\phi_k^{|t+\tau|}\phi_j^{(-t+\tau)}(T+t)}_{\texttt{Sum(b)}}
\end{aligned}
\tag{B10}
$$

Both summation `Sum(a)` and `Sum(b)` can be further split in sums of simple geometric series:

$$
\begin{aligned}
\texttt{Sum(a)}: \sum_{t=1}^{T-1}\phi_k^{(t+\tau)}\phi_j^{|t-\tau|}(T-t) &= \phi_k^\tau\phi_j^\tau T\sum_{t=1}^{\tau}(\phi_k\phi_j^{-1})^t - \phi_k^\tau\phi_j^\tau\sum_{t=1}^{\tau}(\phi_k\phi_j^{-1})^t\cdot t \\
&+ T\phi_k^\tau\phi_j^{-\tau}\sum_{t=\tau+1}^{T-1}(\phi_k\phi_j)^t - \phi_k^\tau\phi_j^{-\tau}\sum_{t=\tau+1}^{T-1}(\phi_k\phi_j)^t\cdot t.
\end{aligned}
\tag{B11}
$$

$$
\begin{aligned}
\texttt{Sum(b)}: &\sum_{t=1-T}^{0} \phi_k^{|t+\tau|}\phi_j^{(-t+\tau)}(T+t) \\
&= T\phi_k^{-\tau}\phi_j^{\tau}\sum_{t=1-T}^{-\tau}(\phi_k^{-1}\phi_j^{-1})^t \\
&+ \phi_k^{-\tau}\phi_j^{\tau}\sum_{t=1-T}^{-\tau}(\phi_k^{-1}\phi_j^{-1})^t\cdot t \\
&+T\phi_k^{\tau}\phi_j^{\tau}\sum_{t=-\tau+1}^{0}(\phi_k\phi_j^{-1})^t \\
&+ \phi_k^{\tau}\phi_j^{\tau}\sum_{t=-\tau+1}^{0}(\phi_k\phi_j^{-1})^t\cdot t.
\end{aligned}
\tag{B12}
$$

`Sum(a)` and `Sum(b)` are composed by geometric series and can be easily solved.

$$
\begin{aligned}
\texttt{Sum(III)}: &\sum_{t',t''=1}^{T}\phi_k^{|t'+\tau-t''|}\phi_j^{|t'-t''|} \\
&= \sum_{t=1-T}^{T-1}\phi_k^{|t+\tau|}\phi_j^{|t|}(T-\mid t\mid) \\
&= \underbrace{\sum_{t=1}^{T-1}\phi_k^{t+\tau}\phi_j^{t}(T-t)}_{\texttt{Sum(c)}} \\
&+ \underbrace{\sum_{t=1-T}^{0}\phi_k^{|t+\tau|}\phi_j^{-t}(T+t)}_{\texttt{Sum(d)}}
\end{aligned}
\tag{B13}
$$

`Sum(c)` and `Sum(d)` are composed by geometric series and can be easily solved.

$$
\begin{aligned}
\texttt{Sum(c)}: &\sum_{t=1}^{T-1}\phi_k^{t+\tau}\phi_j^{t}(T-t) \\
&= T\phi_k^{\tau}\sum_{t=1}^{T-1}(\phi_k\phi_j)^t - \phi_k\sum_{t=1}^{T-1}(\phi_k\phi_j)^t\cdot t.
\end{aligned}
\tag{B14}
$$

$$
\begin{aligned}
\texttt{Sum(d)}: &\sum_{t=1-T}^{0}\phi_k^{|t+\tau|}\phi_j^{-t}(T+t) \\
&= T\phi_k^{-\tau}\sum_{t=1-T}^{-\tau}(\phi_k^{-1}\phi_j^{-1})^t + \phi_k^{-\tau}\sum_{t=1-T}^{-\tau}(\phi_k^{-1}\phi_j^{-1})^t\cdot t \\
&+ T\phi_k^{\tau}\sum_{t=-\tau+1}^{0}(\phi_k\phi_j^{-1})^t + \phi_k^{\tau}\sum_{t=-\tau+1}^{0}(\phi_k\phi_j^{-1})^t\cdot t.
\end{aligned}
\tag{B15}
$$

`Sum(c)` and `Sum(d)` are composed by geometric series and can be easily solved.

c. *Final result*

We aim in computing the variance of the response estimator $\mathbb{V}\mathrm{ar}[R_{k,j}(\tau)]$ as shown in Eq. B6. We rewrite the expression in function of the three summations `Sum(I)`, `Sum(II)` and `Sum(III)` solved in the previous section.

$$
\begin{aligned}
\mathbb{V}\mathrm{ar}[R_{k,j}(\tau)] = &\frac{1}{T^2}\Big(\texttt{Sum(I)} \\
&+\phi_k^{2\tau}\cdot\texttt{Sum(I)}(\tau=0) \\
&-2\phi_k^{\tau}\cdot\texttt{Sum(III)}\Big) \\
&+\frac{\delta_{k,j}}{T^2}\Big(\texttt{Sum(II)} \\
&+\phi_k^{2\tau}\texttt{Sum(II)}(\tau=0) \\
&-2\phi_k^{\tau}\cdot\texttt{Sum(III)}\Big).
\end{aligned}
\tag{B16}
$$

Where $\texttt{Sum(I)}(\tau=0)$ and $\texttt{Sum(II)}(\tau=0)$ evaluate `Sum(I)` and `Sum(II)` in $\tau=0$.

We focus on the asymptotic case $T >> 1$ and remind the reader that $|\phi_k\phi_j| < 1$. The leading order of the solution is as follows:

$$
\mathbb{V}\mathrm{ar}[R_{k,j}(\tau)] = \frac{\phi_k^{2\tau}-1}{T} + \frac{2}{T}\Big(\frac{1-\phi_k^{\tau}\phi_j^{\tau}}{1-\phi_k\phi_j}\Big) - \frac{2\phi_k^{\tau}}{T}\Big(\phi_k\frac{\phi_j^{\tau}-\phi_k^{\tau}}{\phi_j-\phi_k}\Big).
\tag{B17}
$$

Finally, we note that in the case of $\phi_k = \phi_j$ in Eq. B17 we substitute the term $\phi_k\frac{\phi_j^{\tau}-\phi_k^{\tau}}{\phi_j-\phi_k}$ with the limit:

$$
\lim_{\phi_j\to\phi_k}\phi_k\frac{\phi_k^{\tau}-\phi_j^{\tau}}{\phi_k-\phi_j} = \phi_k^{\tau}\tau.
\tag{B18}
$$

Equation B17 assumes that each time series has been previously normalized to zero mean and unit variance. In the case of non-standardized time series $x_i(t)$ we need to account for contributions coming from the variances σ_i^2. This can be simply done by correcting equation Eq. B17 as: $(\sigma_k^2/\sigma_j^2)\cdot$ Eq. B17 (see also Eq. 15 in [1]).

Appendix C: Confidence bounds. Numerical vs analytical

We consider the system in Eq. 9 and show all the estimated responses $R_{k,j}$, their ground truths and the confidence bounds in Figure 5. Importantly, we compare the analytical confidence bounds presented in 7 with their numerical estimation as shown in Section II C 1. All bounds are set to $\pm 3\sigma$.

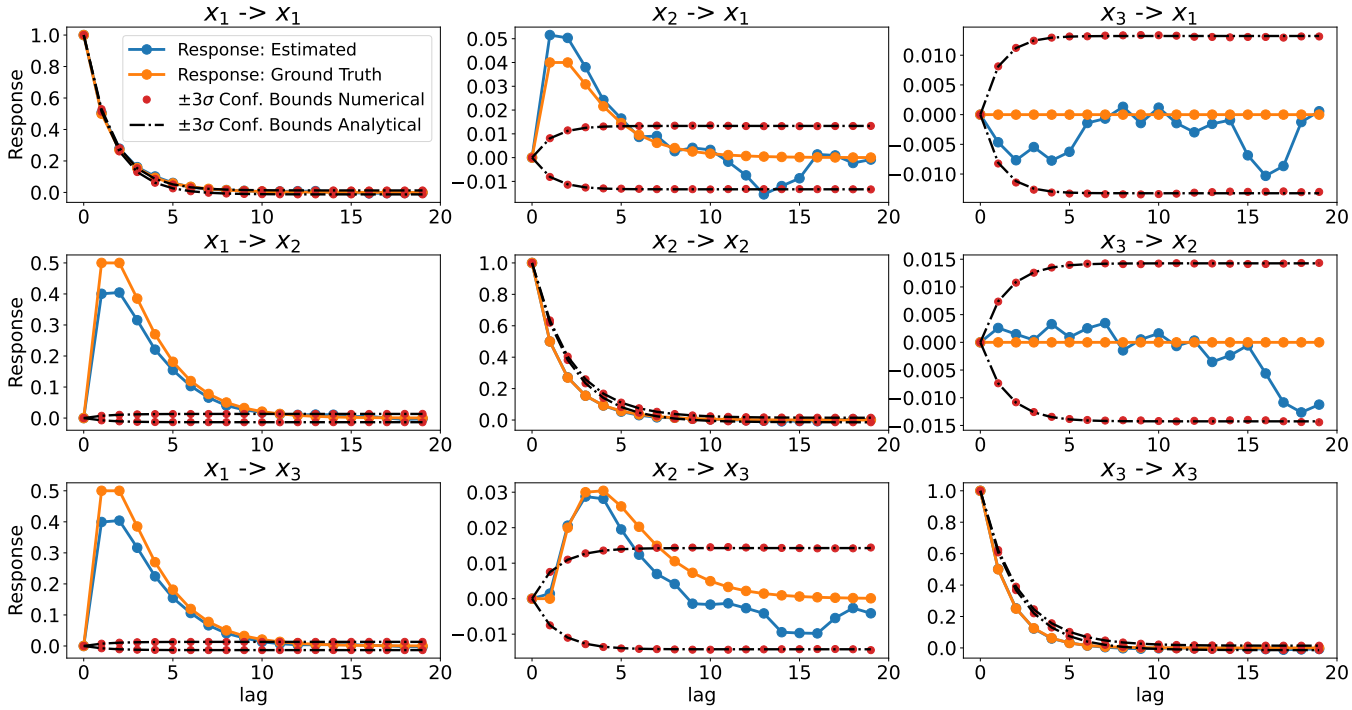

FIG. 5. Comparing the confidence bounds estimated numerically as in Section II C 1 and the analytical solution as shown in Eq. 7 for the simple linear Markov model shown in Eq. 9. Each panel shows the response $R_{k,j}$ representative of the causal link $x_j \to x_k$. "Ground truth" of the response is computed as $\boldsymbol{R}(\tau) = \boldsymbol{M}^\tau$. Blue lines are responses estimated through temporal averages: for this step we use a long trajectory of length $T = 10^5$ simulated by system in Eq. 9. Red dots indicate the confidence bounds computed numerically using $B = 10^4$ ensemble members of the *null* model as shown in II C 1, see Section II C 1. The black dashed line is the analytical solution as in Eq. 7. Confidence bounds are set to $\pm 3\sigma$. All estimated responses (i.e. blue curves) in between the confidence bounds are here considered as spurious.

Appendix D: Histograms of each mode $x_i(t)$ in the global SST field

Histogram of signals $x_i(t)$ defined as shown in Section II A 2 for each community/mode i in the global dataset, see IV C. Each $x_i(t)$ has been first centered to zero mean and than standardized to unit variance. A Gaussian fit is shown in red. The plot shows that the quasi-Gaussian approximation shown in II B 2 is indeed relevant for the system studied.

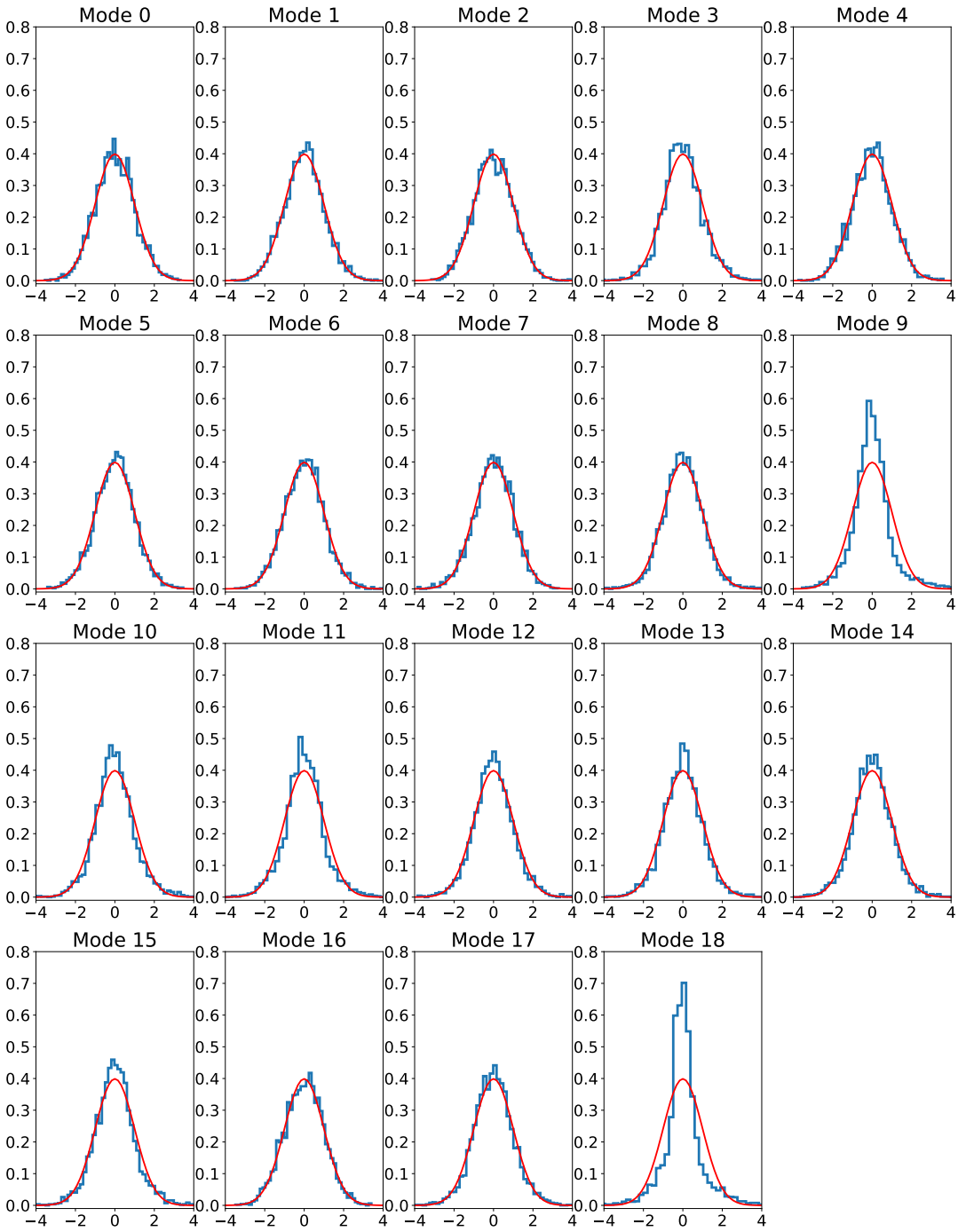

FIG. 6. Probability distributions of each sea surface temperature signal $x_i(t)$ at global scale (see Section IV C as defined in Section II A. Each signal $x_i(t)$ is first centered to zero mean and standardized to unit variance; therefore the x-axis represents degC per standard deviation. Each community is here referred to as "Mode i". A Gaussian fit is shown in red on top of each histogram.

[1] M. Baldovin, F. Cecconi, and A. Vulpiani, Understanding causation via correlations and linear response theory, Physical Review Research **2**, 043436 (2020).

[2] S. Gupta, N. Mastrantonas, C. Masoller, and J. Kurths, Perspectives on the importance of complex systems in understanding our climate and climate change—The Nobel Prize in Physics 2021, Chaos **32**, 052102 (2022).

[3] M. Ghil and V. Lucarini, The physics of climate variability and climate change, Rev. Mod. Phys. **92**, 035002 (2020).

[4] V. Lucarini and M. Chekroun, Hasselmann's program and beyond: New theoretical tools for understanding the climate crisis, arXiv (2023).

[5] S. Philander, El Niño Southern Oscillation phenomena., Nature **302**, 295–301 (1983).

[6] A. Timmermann and et al., El Niño-Southern Oscillation complexity, Nature **559**, 535 (2018).

[7] G. Wang and D. Schimel, Climate change, climate modes, and climate impacts, Annual Review of Environment and Resources **28**, 1 (2003).

[8] A. von der Heydt, P. Ashwin, C. Camp, M. Crucifix, H. Dijkstra, P. Ditlevsen, and T. Lenton, Quantification and interpretation of the climate variability record, Quaternary Research **197**, 103399 (2021).

[9] P. Webster, A. Moore, J. Loschnigg, and R. Leben, Coupled ocean atmosphere dynamics in the Indian Ocean during 1997–98, Nature **401**, 356–360 (1999).

[10] J. M. Wallace and D. S. Gutzler, Teleconnections in the geopotential height field during the Northern Hemisphere winter, Monthly Weather Review **109** (1981).

[11] M. Alexander, I. Bladé, M. Newman, J. Lanzante, N.-C. Lau, and J. Scott, The Atmospheric Bridge: The Influence of ENSO Teleconnections on Air–Sea Interaction over the Global Oceans, Journal of Climate , 2205–2231 (2002).

[12] J. Chiang and A. Sobel, Tropical Tropospheric Temperature Variations Caused by ENSO and Their Influence on the Remote Tropical Climate, Journal of Climate , 2616–2631 (2002).

[13] A. Tsonis, K. Swanson, and P. Roebber, What Do Networks Have to Do with Climate?, Bulletin of the American Meteorological Society **87**, 585–595 (2006).

[14] J. Donges, Y. Zou, N. Marwan, and et al., Complex networks in climate dynamics, Eur. Phys. J. Spec. Top. **174**, 157–179 (2009).

[15] J. Crétat, P. Braconnot, P. Terray, and et al, Mid-Holocene to present-day evolution of the Indian monsoon in transient global simulations, Climate Dynamics **55**, 2761–2784 (2020).

[16] F. Falasca, J. Crétat, A. Bracco, and et al., Climate change in the Indo-Pacific basin from mid- to late Holocene, Climate Dynamics **59**, 753–766 (2022).

[17] A. Bracco, F. Falasca, A. Nenes, and et al., Advancing climate science with knowledge-discovery through data mining, npj Clim Atmos Sci **1**, 20174 (2018).

[18] E. Di Lorenzo, N. Schneider, K. M. Cobb, P. J. S. Franks, K. Chhak, A. J. Miller, J. C. McWilliams, S. J. Bograd, H. Arango, E. Curchitser, T. M. Powell, and P. Rivière, North Pacific Gyre Oscillation links ocean climate and ecosystem change, Geophys. Res. Lett. **35**, L08607 (2008).

[19] F. Falasca, J. Crétat, P. Braconnot, and A. Bracco, Spatiotemporal complexity and time-dependent networks in sea surface temperature from mid- to late Holocene, Eur Phys J Plus , 135:392 (2020).

[20] J. Donges, Y. Zou, N. Marwan, and J. Kurths, The backbone of the climate network, Europhysics Letters **87**, 48007 (2018).

[21] V. Lucarini, Revising and Extending the Linear Response Theory for Statistical Mechanical Systems: Evaluating Observables as Predictors and Predictands, J Stat Phys **173**, 1698–1721 (2018).

[22] D. Hume, *A Treatise of Human Nature* (Oxford University Press, USA, 2001 edited by D. Norton and M. Norton, 1736).

[23] B. Russell, On the notion of cause, Proceedings of the Aristotelian Society 8 **13**, 1 (1913).

[24] D. Bohm, *Causality and Chance in Modern Physics* (Routledge & Kegan Paul and D. Van Nostrand, 1957).

[25] N. Cartwright, Causal laws and effective strategies, In How the Laws of Physics Lie. Oxford: Oxford University Press **13**, 1 (1983).

[26] C. Rovelli, How causation is rooted into thermodynamics, arXiv:2211.00888 doi.org/10.48550/arXiv.2211.00888 (2022).

[27] E. Adlam, Is there causation in fundamental physics? new insights from process matrices and quantum causal modelling, arXiv:2208.02721 https://doi.org/10.48550/arXiv.2208.02721.

[28] J. Ismael, Causation, Free Will, and Naturalism, in *Scientific Metaphysics* (Oxford University Press, 2013).

[29] J. Ismael, Causal content and global laws: Grounding modality in experimental practice, in *The Experimental Side of Modeling* (University of Minnesota Press, 2018) pp. 168–188.

[30] J. Ismael, Reflections on the asymmetry of causation, Interface Focus **12**, 20220081 (2023).

[31] J. Pearl, *Causality: Models, Reasoning, and Inference.* (Cambridge: Cambridge University Press, 2000).

[32] C. Granger, Investigating causal relations by econometric models and cross-spectral methods, Econometrica **37**, 424 (1969).

[33] T. Schreiber, Measuring information transfer, Phys. Rev. Lett. **85**, 461 (2000).

[34] J. Runge, J. Heitzig, V. Petoukhov, and J. Kurths, Escaping the curse of dimensionality in estimating multivariate transfer entropy, Phys. Rev. Lett. **108**, 258701 (2012).

[35] L. Barnett, A. Barrett, and A. Seth, Granger causality and transfer entropy are equivalent for gaussian variables, Phys. Rev. Lett. **103**, 238701 (2009).

[36] J. Pearl, Causal inference in statistics: An overview., Statistics Surveys **3**, 96–146 (2009).

[37] I. Ebert-Uphoff and Y. Deng, Causal discovery for climate research using graphical models, J. Clim. **25**, 5648–5665 (2012).

[38] 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).

[39] G. Camps-Valls, A. Gerhardus, U. Ninad, G. Varando, G. Martius, E. Balaguer-Ballester, R. Vinuesa, E. Diaz,

L. Zanna, and J. Runge, Discovering Causal Relations and Equations from Data, arXiv:2305.13341 https://doi.org/10.48550/arXiv.2305.13341 (2023).
[40] J. Kaddour, A. Lynch, Q. Liu, M. Kusner, and R. Silva, Causal machine learning: A survey and open problems, arXiv:2206.15475v2 https://doi.org/10.48550/arXiv.2206.15475 (2022).
[41] R. Kubo, The fluctuation–dissipation theorem, Rep. Prog. Phys. **29**, 255–284 (1966).
[42] U. Marconi, A. Puglisi, L. Rondoni, and A. Vulpiani, Fluctuation-dissipation: Response theory in statistical physics, Phys. Rep. **461** (2008).
[43] D. Ruelle, A review of linear response theory for general differentiable dynamical systems, Nonlinearity **22**, 855–870 (2009).
[44] U. Tomasini and V. Lucarini, Predictors and predictands of linear response in spatially extended systems, Eur. Phys. J. Spec. Top. **230**, 2813–2832 (2021).
[45] A. J. Majda, R. V. Abramov, and M. J. Grote, *Information Theory and Stochastics for Multiscale Nonlinear Systems* (CRM Monograph Series, American Mathematical Society, 2005).
[46] L. Barnett, A. Barrett, and A. Seth, Granger causality and transfer entropy are equivalent for gaussian variables, Phys. Rev. Lett. **103**, 238701 (2009).
[47] J. Runge, J. Heitzig, V. Petoukhov, and J. Kurths, Investigating causal relations by econometric models and cross-spectral methods, Phys. Rev. Lett. **108**, 258701 (2012).
[48] N. Ay and D. Polani, Information flows in causal networks, Adv. Complex Syst. **11**, 17 (2008).
[49] C. E. Leith, Climate response and fluctuation dissipation, Journal of The Atmospheric Science **32**, 2022–2026 (1975).
[50] V. Lucarini, F. Ragone, and F. Lunkeit, Predicting Climate Change Using Response Theory: Global Averages and Spatial Patterns, J Stat Phys **166**, 1036–1064 (2017).
[51] A. Gritsun and G. Branstator, Climate response using a three-dimensional operator based on the fluctuation–dissipation theorem, Journal of The Atmospheric Science , 2558–2575 (2007).
[52] A. Majda, B. Gershgorin, and Y. Yuan, Low-Frequency Climate Response and Fluctuation–Dissipation Theorems: Theory and Practice , Journal of the Atmospheric Sciences **67**, 1186–1201 (2010).
[53] P. Hassanzadeh and Z. Kuang, The linear response function of an idealized atmosphere. part i: Construction using green's functions and applications, Journal of The Atmospheric Science , 3423–3439 (2016).
[54] P. Hassanzadeh and Z. Kuang, The linear response function of an idealized atmosphere. part ii: Implications for the practical use of the fluctuation–dissipation theorem and the role of operator's nonnormality, Journal of The Atmospheric Science , 3441–3452 (2016).
[55] A. Gritsun, Construction of response operators to small external forcings for atmospheric general circulation models with time periodic right-hand sides, Izvestiya, Atmospheric and Oceanic Physics **6**, 748–756 (2010).
[56] A. Gritsun and V. Dymnikov, Barotropic Atmosphere Response to Small External Actions: Thepry and Numerical Experiments, Izvestiya, Atmospheric and Oceanic Physics **35**, 511 (2010).
[57] H. M. Christensen and J. Berner, From reliable weather forecasts to skilful climate response: A dynamical systems approach., Q. J. R. Meteorological Soc. **145**, 1052–1069 (2019).
[58] D. Ruelle, General linear response formula in statistical mechanics, and the fluctuation-dissipation theorem far from equilibrium, Physics Letters A **245**, 220 (1998).
[59] J. F. Gibson, J. Haclrow, and P. Cvitanović, Visualizing the geometry of state space in plane Couette flow, Journal of Fluid Mechanics **611**, 107–130 (2008).
[60] P. Cvitanović, R. Artuso, R. Mainieri, G. Tanner, and G. Vattay, *Chaos: Classical and Quantum* (ChaosBook.org, Niels Bohr Institute, Copenhagen, 2016).
[61] F. Falasca and A. Bracco, Exploring the Tropical Pacific Manifold in models and observations, Phys. Rev. X **12**, 021054 (2022).
[62] X. Ding, H. Chaté, P. Cvitanović, E. Siminos, and K. A. Takeuchi, Estimating the Dimension of an Inertial Manifold from Unstable Periodic Orbits, Phys. Rev. Lett. **117**, 024101 (2016).
[63] D. Faranda *et al.*, Dynamical proxies of North Atlantic predictability and extremes, Sci. Rep. **7**, 41278 (2017).
[64] J. Theiler, Estimating fractal dimension, J. Opt. Soc. Am. A **7**, 1055 (1990).
[65] S. Klein, B. Soden, and N. Lau, Remote sea surface temperature variations during ENSO: evidence for a tropical atmospheric bridge, J Clim **12**, 917–932 (1999).
[66] B. Dubrulle, F. Daviaud, D. Faranda, L. Marié, and B. Saint-Michel, How many modes are needed to predict climate bifurcations? Lessons from an experiment, Nonlin. Processes Geophys. **29**, 17–35 (2022).
[67] I. Fountalis, C. Dovrolis, A. Bracco, B. Dilkina, and S. Keilholz, δ-MAPS from spatio-temporal data to a weighted and lagged network between functional domain, Appl. Netw. Sci. **3**, 21 (2018).
[68] F. Falasca, A. Bracco, A. Nenes, and I. Fountais, Dimensionality Reduction and Network Inference for Climate Data Using δ-MAPS: Application to the CESM Large Ensemble Sea Surface Temperature, Journal of Advances in Modelling the Earth's System **11**, 1479 (2019).
[69] C. Dalelane, K. Winderlich, and A. Walter, Evaluation of global teleconnections in CMIP6 climate projections using complex networks, Earth Syst. Dynam. **14**, 17–37 (2023).
[70] C. M. L. Camargo, R. E. M. Riva, T. H. J. Hermans, E. M. Schütt, M. Marcos, I. Hernandez-Carrasco, and A. B. A. Slangen, Regionalizing the sea-level budget with machine learning techniques, Ocean Sci. **19**, 17–41 (2023).
[71] L. Novi, A. Bracco, and Falasca, Uncovering marine connectivity through sea surface temperature, Sci Rep **11**, 8839 (2021).
[72] L. Novi and A. Bracco, Machine learning prediction of connectivity, biodiversity and resilience in the Coral Triangle, Commun Biol **5**, 1359 (2022).
[73] A. Lancichinetti and S. Fortunato, Community detection algorithms: A comparative analysis, Phys. Rev. E **80**, 1 –11 (2009).
[74] A. L. Barabási, Network science, Cambridge, UK: Cambridge University Press (2016).
[75] A. Lancichinetti and S. Fortunato, Community detection algorithms: A comparative analysis, Phys. Rev. E

80, 056117 (2009).
[76] M. Rosvall and C. Bergstrom, An information-theoretic framework for resolving community structure in complex networks, Proc. Natl. Acad. Sci. USA **104**, 7327–7331 (2007).
[77] M. Rosvall and C. Bergstrom, Maps of random walks on complex networks reveal community structure, Proc. Natl. Acad. Sci. USA **105**, 1118 –1123 (2008).
[78] M. Rosvall, D. Axelsson, and C. Bergstrom, The map equation, Eur. Phys. J. Spec. Top. **178**, 13–23 (2009).
[79] A. Tantet and H. A. Dijkstra, An interaction network perspective on the relation between patterns of sea surface temperature variability and global mean surface temperature, Earth Syst. Dynam. **5**, 1–14 (2014).
[80] G. Boffetta, G. Lacorata, S. Musacchio, and A. Vulpiani, Relaxation of finite perturbations: Beyond the fluctuation-response relation, CHAOS **13**, 806–811 (2003).
[81] M. J. Ring and R. A. Plumb, The response of a simplified gcm to axisymmetric forcings: Applicability of the fluctuation– dissipation theorem, Journal of The Atmospheric Sciences **65**, 3880–3898 (2008).
[82] A. Majda, B. Gershgorin, and Y. Yuan, Low-frequency climate response and fluctuation–dissipation theorems: Theory and practice, Journal of The Atmospheric Sciences , 1186–1201 (2010).
[83] P. Castiglion, M. Falcioni, A. Lesne, and A. Vulpiani, *Chaos and coarse-graining in statistical mechanics* (Cambridge University Press, 2008).
[84] M. Ghil and S. Childress, Topics in geophysical fluid dynamics: Atmospheric dynamics, dynamo theory, and climate dynamics, Springer **60** (1987).
[85] P. Imkeller and J. V. Storch, Stochastic climate models, Springer Science & Business Media **49** (2001).
[86] M. Allen and L. Smith, Monte Carlo SSA: Detecting irregular oscillations in the Presence of Colored Noise , Journal of Climate **9**, 3373–3404 (1996).
[87] H. A. Dijkstra, E. Hernández-Garcia, C. Masoller, and M. Barreiro, Networks in climate, Cambridge University Press (2019).
[88] R. Kubo, M. Toda, and N. Hashitsume, Statistical mechanics of linear response, in Statistical Physics II (Springer, Berlin, 1991).
[89] I. M. Held, H. Guo, A. Adcroft, J. P. Dunne, L. W. Horowitz, J. Krasting, and et al., Structure and performance of GFDL's CM4.0 climate model, Journal of Advances in Modeling Earth Systems **11**, 3691–3727 (2019).
[90] A. Adcroft, W. Anderson, V. Balaji, C. Blanton, M. Bushuk, C. O. Dufour, and et al., The GFDL global ocean and sea ice model OM4.0: Model description and simulation features, Journal of Advances in Modeling Earth Systems **11**, 3167–3211 (2019).
[91] M. Zhao and Coauthors, The GFDL global atmosphere and land model AM4.0/LM4.0: 1. Simulation characteristics with prescribed SSTs., Journal of Advances in Modeling Earth Systems **10**, 691–734 (2018).
[92] M. Zhao and Coauthors, The GFDL global atmosphere and land model am4.0/LM4.0: 2. Model description, sensitivity studies, and tuning strategies., Journal of Advances in Modeling Earth Systems **10**, 735–769 (2018).
[93] A. Jüling, H. Dijkstra, A. Hogg, and et al, Multidecadal variability in the climate system: phenomena and mechanisms., Eur. Phys. J. Plus **135**, doi.org/10.1140/epjp/s13360-020-00515-4 (2020).
[94] K. Hasselmann, Stochastic climate models part i. theory., Tellus **28**, 473 (1976).
[95] C. Frankignoul and K. Hasselmann, Stochastic climate models, Part II Application to sea-surface temperature anomalies and thermocline variability, Tellus **29**, 289 (1977).
[96] C. Penland, Random Forcing and Forecasting Using Principal Oscillation Pattern Analysis, Monthly Weather Review **117**, 2165 (1989).
[97] C. Penland and P. Sardeshmukh, The optimal growth of tropical sea surface temperature anomalies, Journal of Climate **8**, 1999–2024 (1995).
[98] W. Moon and J. S. Wettlaufer, A unified nonlinear stochastic time series analysis for climate science, Sci. Rep. **7** (2017).
[99] N. Keyes, L. Giorgini, and J. Wettlaufer, Stochastic paleoclimatology: Modeling the epica ice core climate records, arXiv arXiv:2210.00308v1 (2023).
[100] W. Cai and et al., Pantropical climate interactions, Science **363**, eaav4236 (2019).
[101] N. Keenlyside and M. Latif, Understanding Equatorial Atlantic Interannual Variability , Journal of Climate **20**, 131 (2007).
[102] B. Rodríguez-Fonseca, I. Polo, J. García Serrano, T. Losada, E. Mohino, C. Mechoso, and F. Kucharki, Are Atlantic Niños enhancing Pacific ENSO events in recent decades?, Geophysical Research Letters **36**, L20705 (2009).
[103] H. Ding, N. Keenlyside, and M. Latif, Impact of the Equatorial Atlantic on the El Niño Southern Oscillation, Clim Dyn **38**, 1965–1972 (2012).
[104] Y. Ham, J. Kug, J. Park, *et al.*, Sea surface temperature in the north tropical Atlantic as a trigger for El Niño/Southern Oscillation events., Nature Geosci **6**, 112–116 (2013).
[105] T. Izumo, J. Vialard, H. Dayan, M. Lengaigne, and I. Suresh, A simple estimation of equatorial Pacific response from wind stress to untangle Indian Ocean Dipole and Basin influences on El Niño, Clim. Dyn. **46**, 2247–2268 (2016).
[106] K.-J. Ha, J.-E. Chu, J.-Y. Lee, and K.-S. Yun, Interbasin coupling between the tropical Indian and Pacific Ocean on interannual timescale: Observation and CMIP5 reproduction, Clim. Dyn. **48**, 459–475 (2017).
[107] M. Messiè and F. Chavez, Global Modes of Sea Surface Temperature Variability in Relation to Regional Climate Indices, Journal of Climate **24**, 4314–4331 (2011).
[108] L. Onsager and S. Machlup, Fluctuations and irreversible processes, Phys. Rev. **91**, 1505 (1953).
[109] F. Takens, Detecting strange attractors in turbulence, in Dynamical Systems and Turbulence, in *Lect. Notes in Mathematics*, Vol. 898, edited by D. Rand and L. Young (Springer, Berlin,Heidelberg, 1981) p. 21–48.
[110] R. S. Martynov and Y. M. Nechepurenko, Finding the response matrix for a discrete linear stochastic dynamical system, J. Comput. Math. Phys. **44**, 771–781 (2004).
[111] R. S. Martynov and Y. M. Nechepurenko, Finding the response matrix to the external action from a subspace for a discrete linear stochastic dynamical system, Comput. Math. and Math. Phys. **46**, 1155–1167 (2006).
[112] J. Hartlap, P. Simon, and P. Schneider, Why your model parameter confidences might be too optimistic. Unbi-

ased estimation of the inverse covariance matrix, Astronomy and Astrophysics **464**, 399–404 (2007).

[113] M. Yuan, High Dimensional Inverse Covariance Matrix Estimation via Linear Programming, Journal of Machine Learning Researchs **11**, 2261 (2010).

[114] M. Baldovin, F. Cecconi, A. Provenzale, and A. Vulpiani, Extracting causation from millennial-scale climate fluctuations in the last 800 kyr, Scientific Reports **12**, 15320 (2022).

[115] A. Pendergrass and D. Hartmann, Two Modes of Change of the Distribution of Rain, Journal Of Climate **27**, 8357–8371 (2014).

[116] H. Hotelling, Analysis of a complex of statistical variables into principal components, Journal of educational psychology **24(6)**, 417 (1933).

[117] H. v. Storch and F. W. Zwiers, *Statistical Analysis in Climate Research* (Cambridge University Press, 1999).

[118] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning* (Springer,New York, 2013).

[119] D. Dommenget and M. Latif, A cautionary note on the interpretation of eofs, Journal of Climate **15**, 216–225 (2002).

[120] R. Kawamura, A rotated eof analysis of global sea surface temperature variability with interannual interdecadal scales, J. Phys. Oceanogr. **24**, 707–715 (1994).

[121] H. von Storch and F. W. Zwiers, *Regression, in: Statistical Analysis in Climate Research* (Cambridge University Press, 1999b).

[122] L. Saul and S. Roweis, Think Globally, Fit Locally: Unsupervised Learning of Low Dimensional Manifolds, J. Machine Learn. Res. **4**, 119 (2003).

[123] J. Tenenbaum, V. de Silva, and J. Langford, A Global Geometric Framework for Nonlinear Dimensionality Reduction, Science **290**, 2319 (2000).

[124] L. van der Maaten and G. Hinton, Visualizing high-dimensional data using t-SNE, J. Mach. Learn. Res. **9**, 2579–2605 (2008).

[125] L. McInnes, J. Healy, and J. Melville, Umap: Uniform manifold approximation and projection for dimension reduction, arXiv arXiv:1802.03426v3 (2020).

[126] K. Moon, D. van Dijk, Z. Wang, *et al.*, Visualizing structure and transitions in high-dimensional biological data, Nat. Biotechnol. **37**, 1482–1492 (2019).

[127] D. Bueso, M. Piles, and G. Camps-Valls, Nonlinear pca for spatio-temporal analysis of earth observation data, IEEE Transactions on Geoscience and Remote Sensing **58**, 5752 (2020).

[128] K. Lee and K. T. Carlberg, Model reduction of dynamical systems on nonlinear manifolds using deepconvolutional autoencoders, J. Comput. Phys. **404**, 108973 (2020).

[129] S. Shamekh, K. Lamb, Y. Huang, and P. Gentine, Implicit learning of convective organization explainsprecipitation stochasticity, Proceedings of the National Academy of Science **120** (2023).

[130] K. Thirumalai, P. N. DiNezio, J. E. Tierney, M. Puy, and M. Mohtadi, An El Niño Mode in the Glacial Indian Ocean?, Paleoceanography and Paleoclimatology **34**, 1316–1327 (2019).

[131] P. DiNezio, M. Puy, K. Thirumalai, F.-F. Jin, and J. Tierney, Emergence of an equatorial mode of climate variability in the Indian Ocean, Science Advances **6**, eaay7684 (2020).

[132] D. Edler, A. Holmgren, and M. Rosvall, The MapEquation software package (2022).

[133] N. Rayner, D. Parker, E. Horton, C. Folland, L. Alexander, D. Rowell, E. Kent, and A. Kaplan, Global analyses of sea surface temperature, sea ice, and night marineair temperature since the late nineteenth century, JOURNAL OF GEOPHYSICAL RESEARCH **108**, 4407 (2003).

[134] R. Courant and D. Hilbert, *Methods of Mathematical Physics (First English Edition)* (Wiley-VCH Verlag, 2004).