Exploring the Tropical Pacific Manifold in Models and Observations

Fabrizio Falasca[1,2,*] and Annalisa Bracco[2]

[1]*Courant Institute of Mathematical Sciences, New York University, New York, New York 10012, USA*
[2]*School of Earth and Atmospheric Sciences, Georgia Institute of Technology, Atlanta, Georgia 30332, USA*

(Received 8 October 2021; revised 14 March 2022; accepted 20 April 2022; published 8 June 2022; corrected 15 July 2022)

The threat of global warming and the demand for reliable climate predictions pose a formidable challenge because the climate system is multiscale, high-dimensional and nonlinear. Spatiotemporal recurrences of the system hint to the presence of a low-dimensional manifold containing the high-dimensional climate trajectory that could make the problem more tractable. Here we argue that reproducing the geometrical and topological properties of the low-dimensional attractor should be a key target for models used in climate projections. In doing so, we propose a general data-driven framework to characterize the climate attractor and showcase it in the tropical Pacific Ocean using a reanalysis as observational proxy and two state-of-the-art models. The analysis spans four variables simultaneously over the periods 1979–2019 and 2060–2100. At each time t, the system can be uniquely described by a state space vector parametrized by N variables and their spatial variability. The dynamics is confined on a manifold with dimension lower than the full state space that we characterize through manifold learning algorithms, both linear and nonlinear. Nonlinear algorithms describe the attractor through fewer components than linear ones by considering its curved geometry, allowing for visualizing the high-dimensional dynamics through low-dimensional projections. The local geometry and local stability of the high-dimensional, multivariable climate attractor are quantified through the local dimension and persistence metrics. Model biases that hamper climate predictability are identified and found to be similar in the multivariate attractor of the two models during the historical period while diverging under the warming scenario considered. Finally, the relationships between different subspaces (univariate fields), and therefore among climate variables, are evaluated. The proposed framework provides a comprehensive, physically based, test for assessing climate feedbacks and opens new avenues for improving their model representation.

DOI: 10.1103/PhysRevX.12.021054

Subject Areas: Complex Systems
Interdisciplinary Physics
Nonlinear Dynamics

I. INTRODUCTION

A comprehensive analysis of climate variability should account for the multivariate and nonlinear dependencies intrinsic to the system. Quantifying these dependencies is an urgent challenge in climate research [1]. Dynamical system theory offers a pathway to this end, so far successfully applied to high-dimensional turbulent systems [2,3]. At each time step t, the state of a spatiotemporal chaotic system can be viewed as a point in an infinite dimensional state space, parametrized by multiple variables and their spatial dependency. In the case of dissipative chaotic systems, this infinitely dimensional dynamics is confined on a finite dimensional object, commonly known as "inertial manifold" or "attractor" [4,5]. Spatiotemporal chaos can then be seen as a walk on this lower-dimensional inertial manifold [2,3]. Studies of this kind have contributed a quantitative understanding of moderate Reynolds-number turbulence [3,6–8], and similar approaches have been applied to a variety of fields, from computational neuroscience [9–13] to artificial neural networks [14,15] and biophysics [16,17]. Each one of these fields present the challenge of identifying low-dimensional manifolds embedded in very high-dimensional, noisy data [3,5]. Computer scientists have long been interested in identifying low-dimensional manifolds from data as dimensionality reduction tools, a problem commonly known in the literature as "manifold learning" (see, e.g., Ref. [18]). Earth's climate system is indeed a high-dimensional, dissipative dynamical system and its dynamics is expected to be confined to a manifold with lower dimension than the

*fabri.falasca@nyu.edu

full state space [19]. A recent application of manifold learning in climate science has shown the potential of this idea by identifying the multistability properties of an intermediate complexity climate model [20].

In this work, we propose a dynamical systems framework to investigate and compare spatiotemporal climate variability in observations and state-of-the-art climate. The climate system is highly nonlinear and climate models are far from perfect [21]. As a result, climate projections require ensemble simulations to test for sensitivity to initial conditions, and different models [21,22] or strategies [23] to cope with model errors. Intermodel comparisons, such as the Coupled Model Intercomparison Project [24], generate petabytes of data. Mining and quantifying sources of biases, limitations, and ambiguities among these data is fundamental when communicating results to other scientists and policymakers [25,26]. In this context, the dynamical system view that we propose brings several advantages: it allows us to comprehensively study the system accounting for, and quantifying, multivariable dependencies, to assess both mean values and instantaneous properties of each state space point, and to curtail the need for large ensembles in the evaluation of climate model biases and future trajectories. The characterization of the manifold dimensionality of a modeled climate system is indeed largely independent of the realization or ensemble member considered if sufficiently well sampled. For a given climate or time period, the modeled attractor does not change.

The focus of this work is the Pacific Ocean between 20° N and 20° S, where the El Niño Southern Oscillation (ENSO) is the main mode of climate variability. ENSO is an oscillatory mode driving, with its warm and cold phases, El Niño and La Niña, the most dramatic year-to-year variation of Earth's climate system [27]. ENSO affects rainfall patterns, tropical cyclogenesis and the likelihood of droughts and floods, and freshwater availability. ENSO also impacts food security, with cascading effects on health, water, sanitation, education, and overall increased mortality [27–30]. In light of its great societal relevance, the tropical Pacific is a much studied region, and therefore a convenient, well-known, test case. Our approach, however, is general and could be applied anywhere else at both global and regional scales. Model evaluations for ENSO dynamics commonly employ traditional methodologies, such as power spectra and standard deviations [31–33], or more advanced methods such as percolation theory [34] and complex networks [35,36]. Current analyses, however, do not routinely adopt a multivariate approach and quantify only average dynamical properties. Furthermore, the high dimensionality of the system is very seldom considered, and the analysis is often limited to one component, associated to the Niño3.4 index or the first principal component (PC) [37].

Here, we consider a subset of four key variables relevant to the tropical Pacific dynamics: surface temperature, zonal and meridional surface velocities, and outgoing long wave radiation (OLR), and perform our analysis on the ERA5 reanalysis [38] and two state-of-the-art climate models from the Coupled Model Intercomparison Project-phase 6 (CMIP6) [39], the Max Planck Institute (MPI) and EC-Earth models [31,40]. We analyze daily data and their anomalies over two 40-year periods, 1979–2019 and 2060–2100. The two models were chosen because of their output availability and significant differences in their parametrizations, resolutions, and performance [41]. We visualize the high-dimensional inertial manifold using both a linear, principal component analysis (PCA) [37], and a nonlinear, Isomap [42], dimensionality reduction method. Isomap has been chosen among many others (see, e.g., Refs. [43–47]; see also Ref. [48] for a brief review) for its simplicity. Using a nonlinear method has the advantage of identifying the curved manifold itself (intrinsic dimension), rather than the embedding space found by PCA, as discussed in Refs. [12,48].

The time series that we obtain describe the time evolution of modes of climate variability, accounting for more than one variable, and potentially could include all or most model variables; the linear regression of these time series onto each original spatiotemporal field defines the spatial signature of the modes. Most importantly, by comparing projections in the multivariate and univariate representation, we can highlight the role of each variable in the overall system's dynamics and its relation with the other fields.

We then estimate the local properties and stability of the climate attractor of the tropical Pacific in the reanalysis and the models. For a given climate state ζ (i.e., configuration of the Pacific at an instant in time), we evaluate the geometrical properties of the attractor in terms of its local dimension metric $d(\zeta)$ [49,50] and its stability $\theta(\zeta)$ through the inverse of the average persistence of the trajectory around ζ [51,52]. While d roughly quantifies the number of directions the system can evolve from or into, and therefore the number of degrees of freedom required to describe it at that point in time, the persistence quantifies the "stickiness" of the trajectory around each neighborhood in state space, and therefore how predictable the future evolution of that state is, or its predictability potential. These concepts were first introduced in Refs. [53,54] and have been applied so far to univariate fields to explore atmospheric weather regimes [19,49,50,55–61].

Finally, we explore subspaces of the original state space by evaluating local dimension and persistence of each univariate field. The dimension of the modeled manifold is lower than in the observation by construction due to the limits imposed by the model resolution. However, a reliable model should capture the observed relationships among variables, and therefore the scaling of such metrics, to properly represent climate feedbacks.

The data analyzed are described in Sec. II. The analysis framework is presented in Sec. III, followed by the results (Sec. IV). A discussion of our results and their implications and future avenues of research concludes this work.

II. DATA

The ERA5 reanalysis [38] is the observational-based dataset adopted. Produced by the European Centre for Medium-Range Weather Forecasts, ERA5 data are available at a spatial resolution up to 31 km globally. We consider two models, the Max Planck Institute MPI-ESM1.2-HR and EC-Earth3-HR (MPI and EC-Earth hereafter) [31,63]. The resolution of their atmosphere and ocean components is ∼100 km and 0.4° for MPI and ∼40 km and 0.25° for EC-Earth. Both datasets are part of the Coupled Model Intercomparison Project-phase 6 [39] catalog. Their performance according to the scoring analysis by Ref. [41] is in the top five (EC-Earth) and middle range (MPI) among the 37 model configuration tested over the historical period.

The analysis spans the periods 1979–2019 and 2060–2100 at daily frequency. Given that the CMIP6 historical integrations end on December 2014, we concatenated four years from the SSP585 scenario, the "worst-case scenario" with the highest radiative forcing (up to 8.5 W/m^2 in 2100). We further analyzed the SSP585 outputs for the last four decades of the 21st century, 2060–2100.

For each model we analyze four members, randomly chosen, in the historical period and four for EC-Earth and two for MPI (only two are available with daily outputs) in the future. Our discussion, however, focuses on one member, the first in the respective ensembles, in most of our presentation (CMIP6 label: r1i1p1f1 for both models). In the last section of this work (Sec. VI A), we show that the attractor characteristics, as quantified by the local dimension and persistence metrics, do not change as a function of the ensemble member considered. In fact, chaotic trajectories of the same dynamical system are still bounded to live on the same manifold, and manifold properties should not depend on the ensemble member, provided that we sample such object well enough.

The domain of interest is the tropical Pacific in the latitude-longitude range [20° S–20° N, 120° E–70° W]. All datasets are remapped on a coarser grid with resolution of 1° in latitude by 1.5° in longitude; a reasonable step as we are interested in large-scale dynamics. A higher latitudinal resolution is chosen to ensure we are resolving the Rossby wave field [64].

The variables considered are surface temperature (T), zonal and meridional velocities (u, v) at the surface, and outgoing long wave radiation. The temperature field is a driver of variability in the Pacific Ocean, the horizontal velocity vector field quantifies the dynamical response of the system, and the outgoing long wave radiation is a proxy for cloud variability which is key during ENSO. For reproducibility purposes, the variables chosen are temperature at 2m (label: t2m), zonal and meridional velocity at 10m (label: u10 and v10), the top net thermal radiation (label: ttr and equal to the negative of OLR) in ERA5, temperature and velocities at the surface (labels: tas, uas, and vas), and the outgoing long wave flux at the top of the atmosphere (label: rlut) in the models.

A. State space embedding

Given the choice of fields, at each time step t, the tropical Pacific climate is uniquely described by a state space vector defined by $\mathbf{X} = [\mathrm{T}(x,y), \mathrm{u}(x,y), \mathrm{v}(x,y), \mathrm{OLR}(x,y)](t) \in \mathbb{R}^{T,N}$. T is the length of each time series and given that we consider separately two 40-year-long periods at daily temporal resolution, $T = 14\,975$ days. N is the dimensionality of the state space and for the spatial resolution and variables considered, $N = 17\,092$.

III. EARTH'S CLIMATE AS A DYNAMICAL SYSTEM

An important novelty of our approach is that we study the evolution of the highly dimensional climate system focusing on multiple variables simultaneously [3].

Let us consider a spatiotemporal climate system described by m fields $\mathbf{Y}_i$, $i = 1, 2, \ldots, m$, embedded in a grid of size g and spanning a time interval T. For each field $\mathbf{Y}_i$, first we weight all time series by the cosine of their latitude. Each field is then standardized to zero mean and unit variance. If velocity fields are included, it is crucial to standardize the velocity vector and not each single component separately.

At each time step t, the system is uniquely described by a state space vector $\mathbf{X} = [\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_m](t) \in \mathbb{R}^{T,N}$ of dimensionality $N = m \times g$. In this formulation, a single-point trajectory in state space describes the climate system evolution. This trajectory spans a manifold with lower dimension than the full state space because the system is dissipative [19].

Several methodologies can then be applied to investigate such high-dimensional dynamics, as briefly summarized below.

A. Manifold learning

The identification of modes of variability in climate science often relies on principal component analysis [37], commonly referred to as empirical orthogonal functions (EOF) and its variations, such as rotated EOF [65] and extended EOF [66]. In most climate applications, nonlinear components are neglected (for an exception see, e.g., Ref. [43]) and variables are investigated one at a time. However, the climate system is composed by many interacting, covarying variables, and a multivariate approach represents a more rigorous way of quantifying its dynamics. State space embedding for the climate system as a whole is an ill-posed problem, due to the very large number of variables spanning physical, chemical, and biological processes, and not all governing equations are known. For specific problems, on the other hand, a subset of key variables considered together, for example, those for which

we have a good observational record of sort, can offer an in-depth dynamical understanding of the system. To visualize the geometry of the underlying manifold the linear, principal component analysis [37] may still be adopted, or we can rely on nonlinear methods, such as Isomap [42].

1. Principal component analysis

Principal component analysis [37], or empirical orthogonal function analysis [67], is a linear modal decomposition method. Given a spatiotemporal dataset $\mathbf{X} \in \mathbb{R}^{T,N}$ with N time series each of length T, PCA identifies the underlying manifold by fitting hyperplanes in the directions that contain most of the variance. This is achieved by computing the Gram matrix as $\mathbf{G} = [1/(T-1)]\mathbf{X}\mathbf{X}^{\mathrm{T}} \in \mathbb{R}^{T,T}$ [43]. The T eigenvectors $\mathbf{U} \in \mathbb{R}^{T,T}$ of the Gram matrix $\mathbf{G}$ are the principal components of the dataset. Alternatively, it is possible to eigendecompose the covariance matrix $\mathbf{C} = [1/(N-1)]\mathbf{X}^{\mathrm{T}}\mathbf{X} \in \mathbb{R}^{N,N}$. In this case the eigenvectors $\mathbf{V} \in \mathbb{R}^{N,N}$ of $\mathbf{C}$ are spatial patterns and the projection of $\mathbf{V}$ onto $\mathbf{X}$ describes their temporal variability. The decomposition of the Gram matrix returns the same solution of the covariance matrix up to T eigenvectors [43,68]. A third, equivalent alternative, is to eigendecompose the Euclidean distance matrix containing the (Euclidean) distances between each point in $\mathbf{X}$ with the multidimensional scaling (MDS) algorithm [18,69]. Each U_i explains a given variance based on the ratio of its correspondent eigenvalue and the sum of all eigenvalues, i.e., $\lambda_i / \sum_j \lambda_j$. To identify the correspondent spatial patterns is enough to linearly regress each (standardized) PC on the dataset as $[1/(T-1)]\mathbf{U}^{\mathrm{T}}\mathbf{X}$. The low-dimensional projection found by PCA preserves the variances as measured in the high-dimensional input $\mathbf{X}$ [42].

2. Isometric feature mapping (Isomap)

If the manifold is nonlinear Euclidean distances cannot capture the intrinsic distances along the manifold. Isomap is one of the available tools to address this problem (see Ref. [48]) and has been introduced to climate science by Refs. [70,71] for univariate fields. Isomap adds a key step to the MDS algorithm. Given a dataset $\mathbf{X} \in \mathbb{R}^{T,N}$, with N time series each of length T and centered to zero mean, Isomap first identifies the K-nearest neighbors of each point i in the manifold, then it computes the *geodesic* distances $\delta_{i,j}$ between each couple of points i and j by assuming that the manifold is *locally* flat in a radius of K points (see also Appendix A). The geodesic distance matrix $\mathbf{D}_G$ is finally used as input to the MDS algorithm [69]. Given the squared distances $(\mathbf{D}_G^{(2)})_{i,j} = \delta_{i,j}^2$, the double centering matrix is computed as $\mathbf{A} = -\frac{1}{2}\mathbf{J}\mathbf{D}_G^{(2)}\mathbf{J}$, with $\mathbf{J} = \mathbf{I}_T - (1/T)\mathbf{e}\mathbf{e}^{\mathrm{T}}$, where $\mathbf{I}_T$ is the identity matrix of order T and $\mathbf{e} = [1, 1, \ldots, 1]^{\mathrm{T}} \in \mathbb{R}^T$ is a vector of length T containing all ones.

The dimensionality of the dataset is then reduced by finding the eigenvectors of the double centered matrix $\mathbf{A}$ (i.e., solving $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\mathrm{T}}$). Each component V_i is obtained by weighting the ith eigenvector U_i by the square root of its correspondent eigenvalue $\sqrt{\lambda_i}$, as $\mathrm{V}_i = \mathrm{U}_i\sqrt{\lambda_i}$ [42]. Similarly to PCA, the associated spatial patterns can be retrieved by linearly regressing the (standardized) Isomap components on the original dataset as $[1/(T-1)]\mathbf{V}^{\mathrm{T}}\mathbf{X}$.

A downside of using nonlinear dimensionality reduction methods is that the explained variance cannot be directly estimated. The so-called *residual* variance has been proposed as a valid alternative. It is defined as $1 - R(D_M, D_Y)^2$, where D_M is the algorithm's estimate of manifold distances (i.e., Euclidean distance matrix for PCA and geodesic matrix for Isomap), D_Y is the matrix of Euclidean distances in the low-dimensional embedding computed by each algorithm, and R is the Pearson correlation coefficient among all entries of D_M and D_Y [42].

As mentioned, Isomap relies on the parameter K. In our subsequent analysis we set $K = 10$, thus assuming that the manifold is *locally* flat within a radius of 10 days, which is reasonable given the spatial resolution considered and the length of our time series.

B. Dynamical systems metrics: Local dimension and persistence

Quantifying the dimensionality of the inertial manifold allows us to estimate the *effective* degrees of freedom of the dynamical system being investigated, and therefore its complexity [72]. This is a difficult problem for spatiotemporal chaotic systems and an *exact* estimation of such quantity may require knowing the equation of motions, as done in Ding *et al.* [5] for the Kuramoto-Sivashinsky system. For very high-dimensional, noisy datasets, data-driven methods have shown limitations [73], and depending on the problem at hand, certain approaches may be more useful than others.

In the case of climate variability, recent advances at the interface of dynamical system and extreme value theory have opened the possibility to estimate the dimensionality of underlying attractors [54]. For a given dynamical system, the probability of recurrence of a state ζ obeys a generalized Pareto distribution. Locally, such distribution scales with a parameter, shown to be equal to the local dimension $d(\zeta)$. The attractor dimensionality (i.e., Hausdorff dimension) can be retrieved by averaging all d [74,75]. Furthermore, for a climate system it is useful to quantify the tendency of the trajectory to persist in a neighbor of the state space, because this tendency is directly linked to the predictability of that state. The greater the persistence, the higher the predictability. This property can be quantified by introducing the so-called extremal index [51]. The two metrics, local dimension and persistency, are powerful tools to explore high-dimensional dynamics and have been useful to characterize univariate

atmospheric fields in several recent studies (see, e.g., Refs. [49,57,59,61] among others).

Here we briefly present these tools and refer the reader to Ref. [54] for details and rigorous proofs.

We consider the high-dimensional trajectory $\mathbf{X}(t) \in \mathbb{R}^{T,N}$ and for each state $\zeta = \mathbf{X}(\tau)$, with $\tau \in [1, T]$, we define the observable as

$$g(\mathbf{X}(t), \zeta) = -\log\left(\delta(\mathbf{X}(t), \zeta)\right). \tag{1}$$

Here $\delta(x, y)$ represents the Euclidean distance between two vectors x and y and the logarithm further increases the discrimination between close recurrences [49,54]. The minus sign turns minima into maxima for practical convenience; thus the time series g is large when $\mathbf{X}(t)$ is similar to ζ.

We then define a threshold $s(q, \zeta)$ as the qth quantile for each $g(\mathbf{X}(t), \zeta)$ and adopt $q = 0.98$, as in Ref. [49]. The sensitivity to the choice of q value is addressed in Appendix G. The points in $g(\mathbf{X}(t), \zeta)$ that exceed the threshold $s(q, \zeta)$ represent the Poincaré recurrences of ζ and are here referred to as $u(\zeta)$ [54]. Finding recurrences of a state ζ in a neighborhood of radius r is therefore equivalent to find exceedances of $g(\mathbf{X}(t), \zeta)$ over a threshold s, and we refer to a neighborhood of a state ζ as $\Gamma_q(\zeta)$. The Freitas-Freitas-Todd theorem [53], modified by Lucarini *et al.* [54], states that the probability $P(u, \zeta)$ that the dynamics $\mathbf{X}(t)$ returns in a neighborhood $\Gamma_q(\zeta)$ converges to a generalized Pareto distribution [76]:

$$P(u, \zeta) \simeq \exp\left[-\theta(\zeta)\frac{u(\zeta)}{\sigma(\zeta)}\right], \tag{2}$$

where $\sigma(\zeta)$ and $\theta(\zeta)$, the so-called *extremal index* [51,77], are parameters of the distribution.

The *local dimension* $d(\zeta)$ can be computed as $d(\zeta) = [1/\sigma(\zeta)]$ [54]. It relates to the density of state space points in a neighborhood $\Gamma_q(\zeta)$ and roughly quantifies the number of directions the system can evolve from or into. $d(\zeta)$ is therefore linked to the intrinsic *local* predictability of ζ [49] (i.e., the lower $d(\zeta)$ is, the larger the predictability of that state). Reference [54] suggests that the attractor dimension can be computed as the average over all local dimensions $D = \langle d(\zeta) \rangle$, and Ref. [49] has shown that this is indeed the case for the Lorenz system [78].

The *persistence* of the trajectory $\mathbf{X}(t)$ in a neighborhood $\Gamma_q(\zeta)$ is quantified by the extremal index $\theta(\zeta)$ [51,79–81]. Here θ is computed using the methodology proposed by Süveges [79]. Intuitively, $\theta \in [0, 1]$ is linked to the inverse of the mean residence time of $\mathbf{X}(t)$ in $\Gamma_q(\zeta)$, with low values implying higher persistence in the neighborhood. Higher persistence around a state ζ quantifies the tendency of the trajectory to stick in its neighborhood, therefore increasing the potential predictability around ζ. The value of $\theta(\zeta)$ should be divided by the time step Δt [i.e., $\theta(\zeta)/\Delta t$], but in this paper we are in the trivial situation of $\Delta t = 1$ day. For the Lorenz system, lower θ are found in the neighborhoods of the three *unstable* fixed points [49].

The local dimension and persistence metrics are adopted after removing the seasonality and the trend. This allows us to focus on stationary time series.

IV. NONLINEAR AND MULTIVARIATE DIMENSIONALITY REDUCTION

A. First test: Mean state and seasonal cycle

As a proof of concept, we consider the state space evolution in the ERA5, MPI, and EC-Earth in the tropical Pacific retaining trends and seasonal cycle and considering all four fields mentioned. The residual variance after applying the PCA and Isomap algorithms (see Sec. III) on the embedded state space vector $\mathbf{X} = [\mathrm{T}(x, y), \mathrm{u}(x, y), \mathrm{v}(x, y), \mathrm{OLR}(x, y)](t)$ is shown in Fig. 1(a). In both observations and models, the residual variance of the first three components is lower in Isomap; additionally, Isomap saturates faster than PCA. In other words, the 17 092-dimensional trajectory of the tropical Pacific domain lives on a low-dimensional object, which is nonlinear, as verified by comparing Euclidean and geodesic distances for the three datasets (see Appendix A, Fig. 12). The models share strong similarities in their average Euclidean distance, but they differ from the reanalysis in their geodesic components.

Comparing ERA5 and models, the residual variance of the first PCA and Isomap components is higher in the reanalysis, and saturates to a higher value. In EC-Earth, on the other hand, PC1 and Isomap-1 explain a nearly identical amount of variance.

Focusing on the low-dimensional Isomap projections, each point in Fig. 1(b) represents the state of the Pacific system $\mathbf{X} = [\mathrm{T}(x, y), \mathrm{u}(x, y), \mathrm{v}(x, y), \mathrm{OLR}(x, y)](t)$ at a given day. Because of the inclusion of the seasonal cycle and its dominance on the overall variability, the dynamics lives on a torus, which is topologically similar among models and reanalysis. The ERA5 dynamics, however, deviates from the periodic trajectory in correspondence of the 1982–1983 and 1997–1998 El Niño events. No clear deviation occurs in the models at any time, indicating that no ENSO event was capable of modifying the modeled seasonal cycle in the simulations considered [82].

The correlation between Isomap-1 and PC1 is higher than 0.98 independently of the dataset, illustrating that the seasonal cycle variability is a close-to-linear process, as expected. Furthermore, the Pearson correlation across datasets is higher than 0.95 in all cases, indicating that the temporal characteristics of the seasonal cycle are well captured by both models. Large correlations across datasets, however, do not imply similar spatial projections (see Sec. III), which also consider differences in the signals' variance: large regional biases in the representation of the seasonal cycle are verified, as shown in Appendix B, Fig. 13.

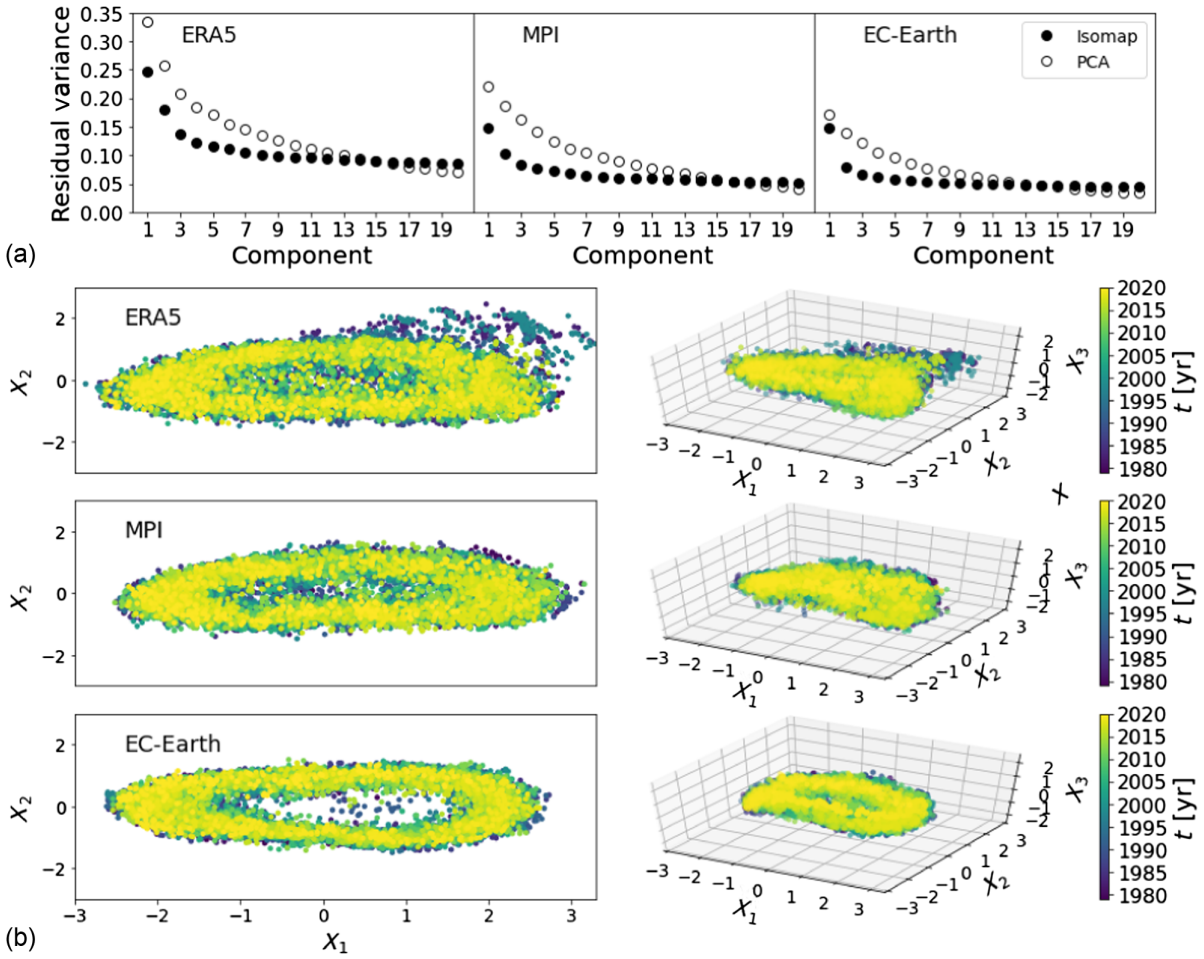

FIG. 1. Dimensionality reduction of the tropical Pacific climate retaining seasonality and trends. (a) Residual variance of PCA and Isomap for ERA5 and the two CMIP6 models (MPI and EC-Earth). (b) Low-dimensional, Isomap projections (first two or three components, respectively) of the three datasets; Isomap components are indicated as X_1, X_2, X_3, respectively. Each point encodes the multivariate state of the tropical Pacific Ocean at a given day from January 1979 to December 2019. Projections have been standardized to unit variance.

B. Variability: ENSO

We next investigate the interannual variability of the tropical Pacific by repeating the analysis after detrending and deseasonalizing the data. PCA and Isomap are again applied to the embedded state space vector $\mathbf{X} = [\mathrm{T}(x,y), \mathrm{u}(x,y), \mathrm{v}(x,y), \mathrm{OLR}(x,y)](t)$ and the residual variances are shown in Fig. 2(a). For ERA5, the first Isomap component explains considerably more variance than the first PC (residual variances are 0.42 and 0.78 for Isomap-1 and PC1, respectively), and the first three Isomap components capture ~70% of the total variance. Higher Isomap components show a faster saturation than PCs in residual variance but never fully saturate, underlying the high dimensionality of the manifold in all three datasets. The PCs residual variance behaves similarly among models and reanalysis, while the Isomap residual variance in ERA5 differs from the models, suggesting that MPI and EC-Earth struggle in capturing the nonlinear topological characteristics of the reanalysis. In Appendix A, Fig. 12 we prove that also the manifold of the anomalies is nonlinear in all cases.

Differences among datasets emerge in the low-dimensional 2D and 3D Isomap projections of the state space trajectory [Fig. 2(b)]. The strong 1982–1983 and 1997–1998 El Niños, followed by the 2015–2016 event, are excursions away from the state space region usually explored by the tropical Pacific trajectory. Neither model replicates such behavior. We cannot exclude that the 40 years examined are outliers in the observational realization, of which we have only one sample. However, we verified that this characteristic disruption of the seasonal cycle is not reproduced in any of the ensemble members analyzed, which may indeed suggest a structural difference between ERA5 and MPI and EC-Earth. The residual variance explained by different components and the state space occupation are similar among models. This is verified despite the different ENSO characteristics quantified by the first Isomap and PC components and their power spectrum

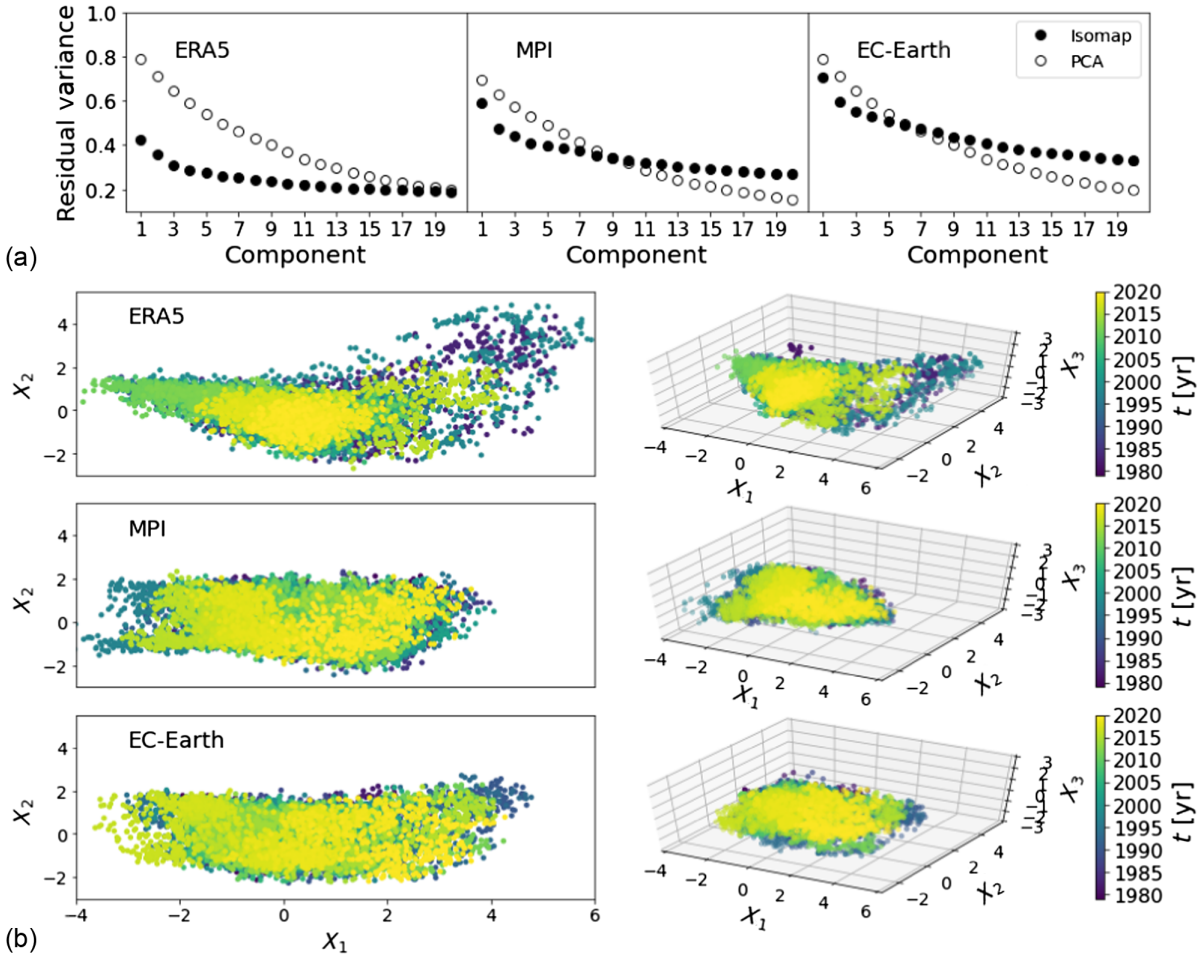

FIG. 2. Dimensionality reduction of the tropical Pacific Ocean anomalies (no linear trend and no seasonal cycle). (a) Residual variance of PCA and Isomap for observations (ERA5) and the two CMIP6 models (MPI and EC-Earth). (b) Low-dimensional, Isomap projections (first two or three components, respectively) of the three datasets; Isomap components are indicated as X_1, X_2, X_3, respectively. Each point encodes the multivariate state of the tropical Pacific Ocean at a given day from January 1979 to December 2019. Projections have been standardized to unit variance.

(see Appendix D, Fig. 15). The spectral analysis identifies the largest, significant peak at ∼3.7 yr in ERA5 and EC-Earth, and at more than ∼8 yr in MPI. In ERA5, PC1 and Isomap-1 projections are well correlated ($r = 0.9$) with largest discrepancies in correspondence of the 1982–1983 and 1997–1998 events. The peak, correspondent to the strong 2015–2016 El Niño event is similarly identified in the PC1 and Isomap-1 components. Therefore, differently from the PC analysis, the Isomap projection implies that the 2015–2016 event was of smaller amplitude and "more linear" in behavior compared to the 1982–1983 and 1997–1998 El Niños. In the models the correlations between PC1 and Isomap-1 projections are higher than in ERA5 ($r = 0.95$ for MPI and 0.94 for EC-Earth).

Spatial patterns are visualized by the regressions of Isomap-1 and shown in Appendix D, Fig. 15. In ERA5, ENSO is characterized by a larger temperature variance on a narrow band in the equatorial eastern Pacific compared to the models, accompanied by a surface wind response consistent with a shift in the convective cell over the central to East Pacific and negative OLR anomalies (or more clouds) in the whole equatorial Pacific. In MPI the ENSO pattern is amplified in the central Pacific, as noted in Ref. [63], and the wind and cloud response is shifted to the western side of the basin. In EC-Earth the ENSO spatial signature in surface temperatures is closer to that of the reanalysis Ref. [31], but the atmospheric response remains biased similarly to MPI, with wind and OLR anomalies not extending sufficiently eastward and a visible double inter-tropical convergence zone bias in cloud distribution.

By the end of the 21st century, ENSO dynamics change in the models in response to the increasing greenhouse gas concentrations of the SSP585 scenario.

The first components—PC1 and Isomap-1—of MPI, resemble the currently observed ENSO, with a spectral peak at ∼3.6 yr but no extreme El Niño events, while in EC-Earth positive and negative events recur periodically every four years (see Appendix D, Fig. 16). ENSO

amplitude increases slightly in MPI, with the largest changes in the OLR field, and substantially in EC-Earth, impacting temperature, zonal velocity, and OLR fields. In EC-Earth, OLR changes the most, with the ENSO-associated anomalies extending over the whole tropical Pacific Ocean. Surface wind changes are concentrated in the zonal direction, and the Isomap-1 regression on the meridional wind velocity is nearly zero.

C. Modes of variability: Multivariate or univariate?

The framework proposed allows for analyzing simultaneously many variables. Here we show the distinctive advantage of doing so.

In Fig. 3 we compare the first Isomap component obtained in the multivariate representation with the univariate case, applying Isomap separately to each field in ERA5. The corresponding figures for the models, looking only at the anomalies, in both current and future periods, can be found in Appendix E, Figs. 18 and 19. Reducing the dimensionality in a multivariate representation corresponds to fitting axis in the direction that maximizes the overall variance. At times, the evolution of a single field reflects the largest fraction of the dataset's variance, leading to small differences between the univariate and multivariate cases. This is the case for the tropical Pacific, where the variability is largely controlled by temperature anomalies.

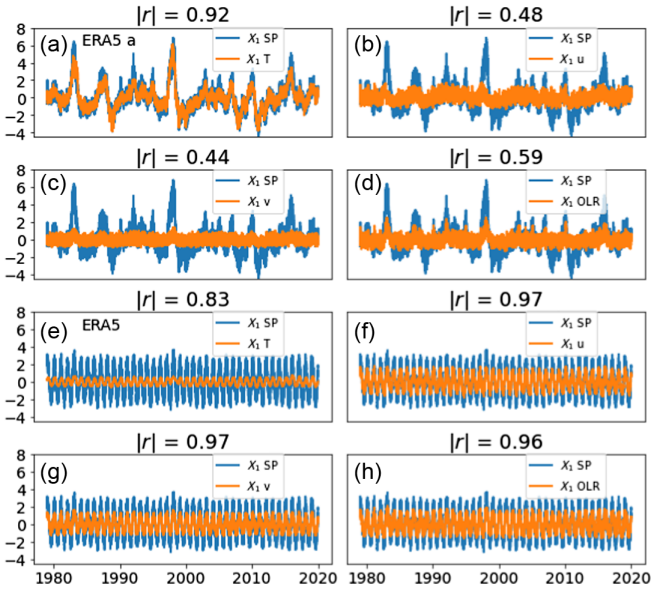

FIG. 3. First Isomap (X_1) component for anomalies (panels a–d) and raw data (panels e–h) in ERA5. In each case components have been standardized so that the *total* variance is equal to 1. Projections in the multivariate case are shown in blue and labeled as "SP" (state space). Projections for each, univariate field are shown in red. Atop of each plot is the correlation coefficient (in absolute value) between the projections in the multivariate and univariate cases.

In general, a multivariate approach helps to identify which variable contributes the most to the dynamics of interest. For example, when considering the seasonal cycle (Fig. 13), the evolution on the torus in correspondence of large El Niño events (1982-1983, 1997-1998, and 2015–2016) does not follow the evolution of temperature, as evident by comparing Isomap-1 in Fig. 3(e) (see also Appendix E, Fig. 17). The multivariate representation largely ignores the temperature contribution in 2015–2016. The seasonal cycle variance is indeed largely dominated by the velocity and OLR fields as shown in Figs. 3(e)–3(h). In other words, the dimensionality reduction with or without embedding highlights the weighted (by variance) relative contribution of each variable to the system dynamics.

Looking again at the anomalies, Fig. 4 displays the correlations between the univariate and multivariate representations of the first three Isomap components in ERA5 and models in the two periods analyzed. A corresponding figure but for the first three PCs can be found in Appendix C, Fig. 14.

For the first component, the temperature field explains the largest part of the variance in both models and observations independently of the period analyzed, with correlations higher than 0.85 in all cases. Both models underestimate the relationship between the zonal wind u and the multivariate case, and MPI shows larger differences than EC-Earth in correspondence of the correlation for meridional velocity v.

Under the SSP585 scenario the relationships among the four variables are nearly invariant in MPI, but change in EC-Earth, especially in the low level wind field, followed by OLR. In EC-Earth the zonal component increases its correlation with the multivariate, temperature-dominated representation, while the opposite is verified in v (see also Appendix D, Fig. 16). The correlations found for the

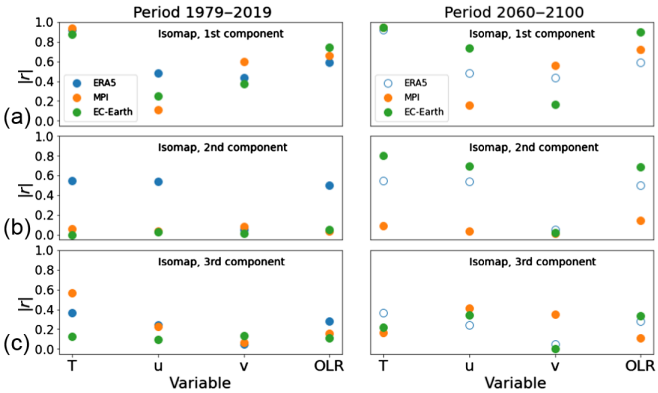

FIG. 4. (a) Correlations (in absolute value) between the first Isomap component of each variable (on the x axis) and the multivariate case in ERA5 and models [signals in Figs. 3(a)–3(d) for ERA5 and Figs. 18 and 19 for MPI and EC-Earth respectively]. Rows (b) and (c) are the same as row (a) but for the second and third Isomap components.

second and third Isomap components [see Figs. 4(b) and 4(d)] further elucidate how the relative role of the variables differs in the modeled versus observed dynamics. Isomap-2, for example, clearly indicates that both models significantly underestimate all correlations between univariate and multivariate representation but for v in historical times. This behavior changes in the future for EC-Earth, that displays even higher correlations than found in ERA5 for T, U, and OLR, but it is unaltered in MPI.

V. LOCAL PROPERTIES OF THE ATTRACTOR

A. Multivariate fields

We quantify local properties of the high-dimensional flow through the local dimension $d(\zeta)$ and persistence $\theta(\zeta)$ metrics introduced in Sec. III. Both metrics are calculated in the state space of the tropical Pacific.

The scatter plots of $d(\zeta)$ versus $\theta(\zeta)$ are shown in Fig. 5. Each point encodes a day in the $d(\zeta)$ versus $\theta(\zeta)$ space and is colored by its respective ENSO value, here defined by the first Isomap component (Appendix D, Figs. 15 and 16). In the observations the two metrics are strongly correlated ($r = 0.81$): days with lower local dimension are characterized by a large mean residence time and higher predictability [low values of $\theta(\zeta)$]. Strong El Niños and, to a lesser extent, La Niñas are characterized by low $d(\zeta)$ and $\theta(\zeta)$ indicating that strong positive and negative ENSO events can be, to a first approximation, interpreted as unstable fixed points of the system. This analysis supports the nonlinear oscillator theoretical framework to explain ENSO dynamics [83] and the asymmetry between El Niños and La Niñas [84]. The asymmetry in the strength of positive and negative events is missed in both models, with the greater predictability of strong El Niños compared to strong La Niñas being reversed in MPI in the historical period. Over the period 1979–2019, the correlation between $d(\zeta)$ and $\theta(\zeta)$ is $r = 0.62$ and 0.72 for MPI and EC-Earth, respectively, therefore lower than in ERA5. Most importantly, the modeled values in the $[d(\zeta), \theta(\zeta)]$ diagram span a smaller region than in the reanalysis.

Indeed, the region characterized by $d(\zeta) > 50$ and $\theta(\zeta) > 0.5$ is very often explored in the reanalysis and is occupied by the majority of ENSO-neutral days, but is seldom visited by both MPI and EC-Earth. In other words, the reanalysis captures higher-dimensional dynamics than the models. These differences are shared by all ensemble members analyzed. This analysis points to a large difference in predictability potential in both models compared to observations, with the models' anomalies being far more predictable in neutral conditions. The difference is further quantified by the histograms of both metrics in Fig. 6 and by the four moments of each distributions summarized in Appendix F.

In the future, both models shift toward greater predictability potential, with lower values of $d(\zeta)$ and $\theta(\zeta)$ (see Fig. 6). This shift to larger predictability in a warmer climate has been termed as the "Hammam effect" and first recognized in model simulations for the sea level pressure at the midlatitudes [85], and in idealized aquaplanet simulations [86]. These changes are subtle in MPI and larger in EC-Earth, especially for persistence. In this model, the regular, periodic behavior of its future ENSO (see Appendix D, Fig. 16) causes the distribution of $\theta(\zeta)$ to shift to lower average and skewness values (Fig. 6 and Appendix F, Table II). In MPI, El Niños become more predictable in the future and have a lower instantaneous dimension than present ones, partially recovering the asymmetry bias found in the historical period.

Figure 7 further visualizes the metrics and the differences between models and observations in the 2D (first and

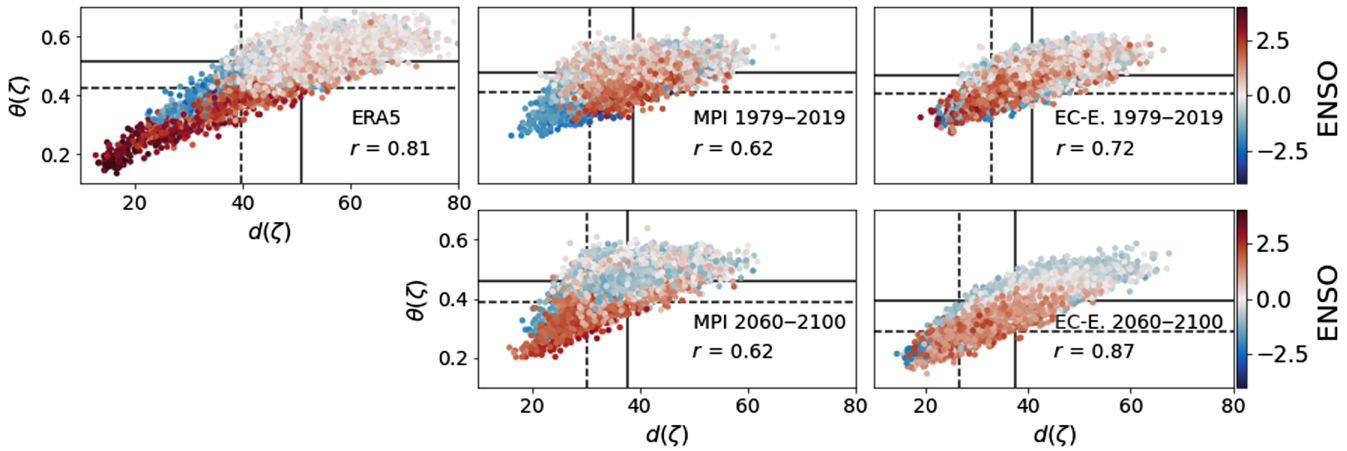

FIG. 5. (a) Scatter plot of stability $\theta(\zeta)$ and local dimension $d(\zeta)$ for observations (ERA5) and the two models (MPI and EC-Earth). Each point represents a day in the period 1979–2019 (top panels) or 2060–2100 (bottom panels). Each point is colored according to its value in the first (standardized) Isomap component, quantifying here the ENSO index and dashed (solid) lines indicate the 0.1 (0.5) quantiles of $d(\zeta)$ and $\theta(\zeta)$. The correlation r between each d and θ is also reported. The computation spans the full 17 092-dimensional state space.

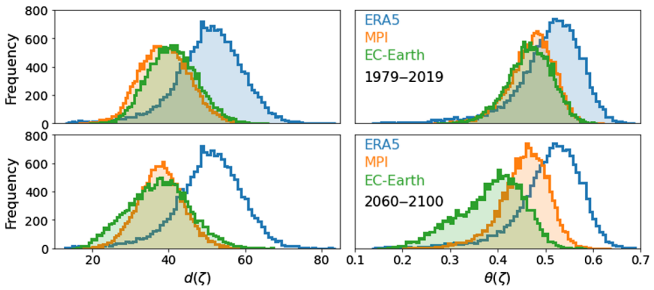

FIG. 6. Histograms of local dimension $d(\zeta)$ and inverse of persistency $\theta(\zeta)$ for ERA5, MPI, and EC-Earth. Top panels: 1979–2019 period. Bottom panels: period 2060–2100 under the SSP585 scenario.

second component) Isomap projections. Both metrics, quantifying local geometry and stability, vary along the manifold, with low values at the outer borders (El Niño and La Niña regions) and high value close to the manifold center.

B. Univariate fields and their scaling

The analysis so far shows that distributions of local dimensions and persistence are biased toward lower values in the modeled climate. A lower average dimension d implies that the models analyzed are inherently less complex; additionally, lower θ imply that anomalies in the Pacific, on average, persist longer than in observations.

Climate models do not solve all scales and therefore their intrinsic dimensionality (i.e., dimension of their inertial manifold) is expected to be lower than in the reanalysis, as suggested in Ref. [62]; however, in spite of large differences in atmospheric horizontal resolution (∼100 and ∼40 km for MPI and EC-Earth, respectively) and in parametrization schemes, these two models have a nearly identical attractor dimensionality (see Appendix F, Table II).

We further investigate the source for this bias by exploring the subspaces defined by each univariate field. While lower dimensional, a model should properly capture the observed scaling among local dimension and persistence of each variable.

We compute local dimension and persistence for each variable separately and quantify distances between ERA5 and models' distributions using the Wasserstein distance metric [87]. Results are presented in terms of pairwise distance matrices. Crucially, we are interested in quantifying how distances between distributions scale among each

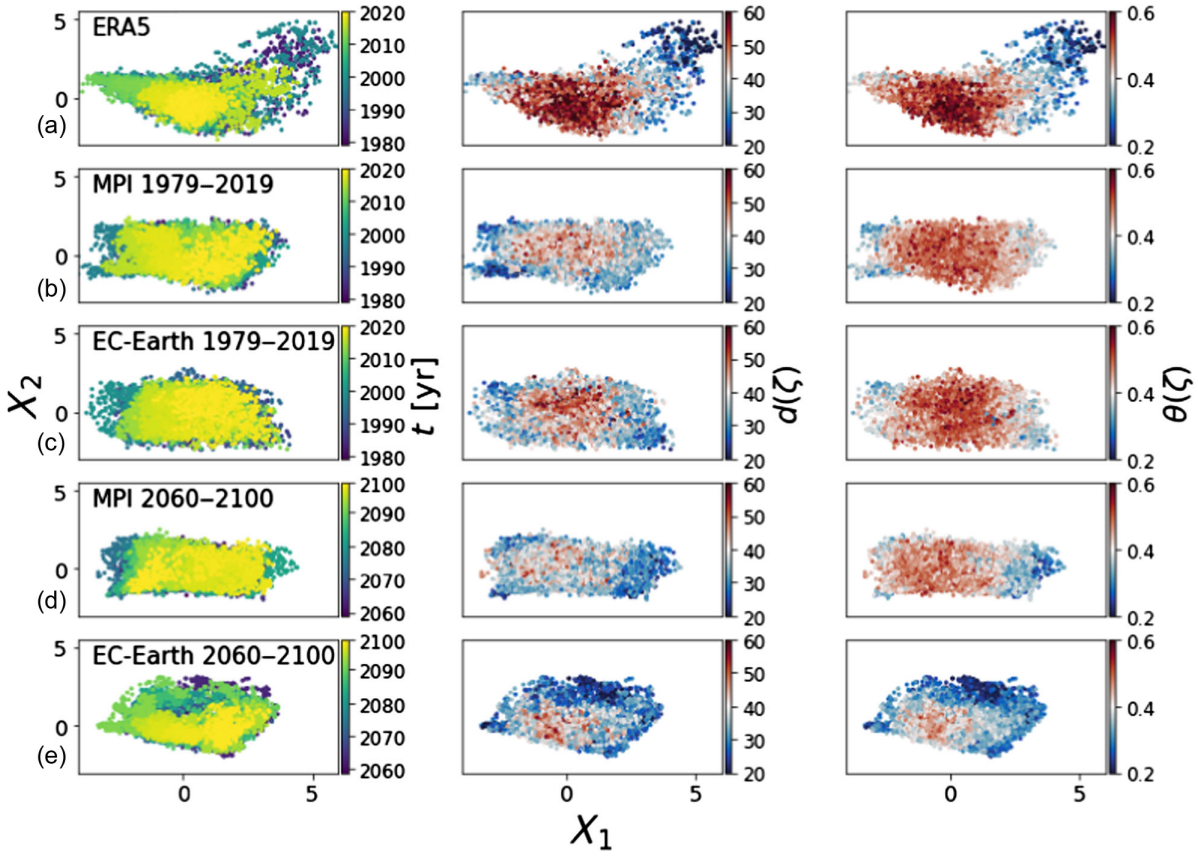

FIG. 7. In each row, from left to right: two-dimensional, Isomap projections of the high-dimensional state space dynamics; Isomap projections with each point colored by its local dimension $d(\zeta)$; Isomap projections with each point colored by its local stability $\theta(\zeta)$. (a)–(e) ERA5 reanalysis, MPI (1979–2019), EC-Earth (1979–2019), MPI (2060–2100), and EC-Earth (2060–2100).

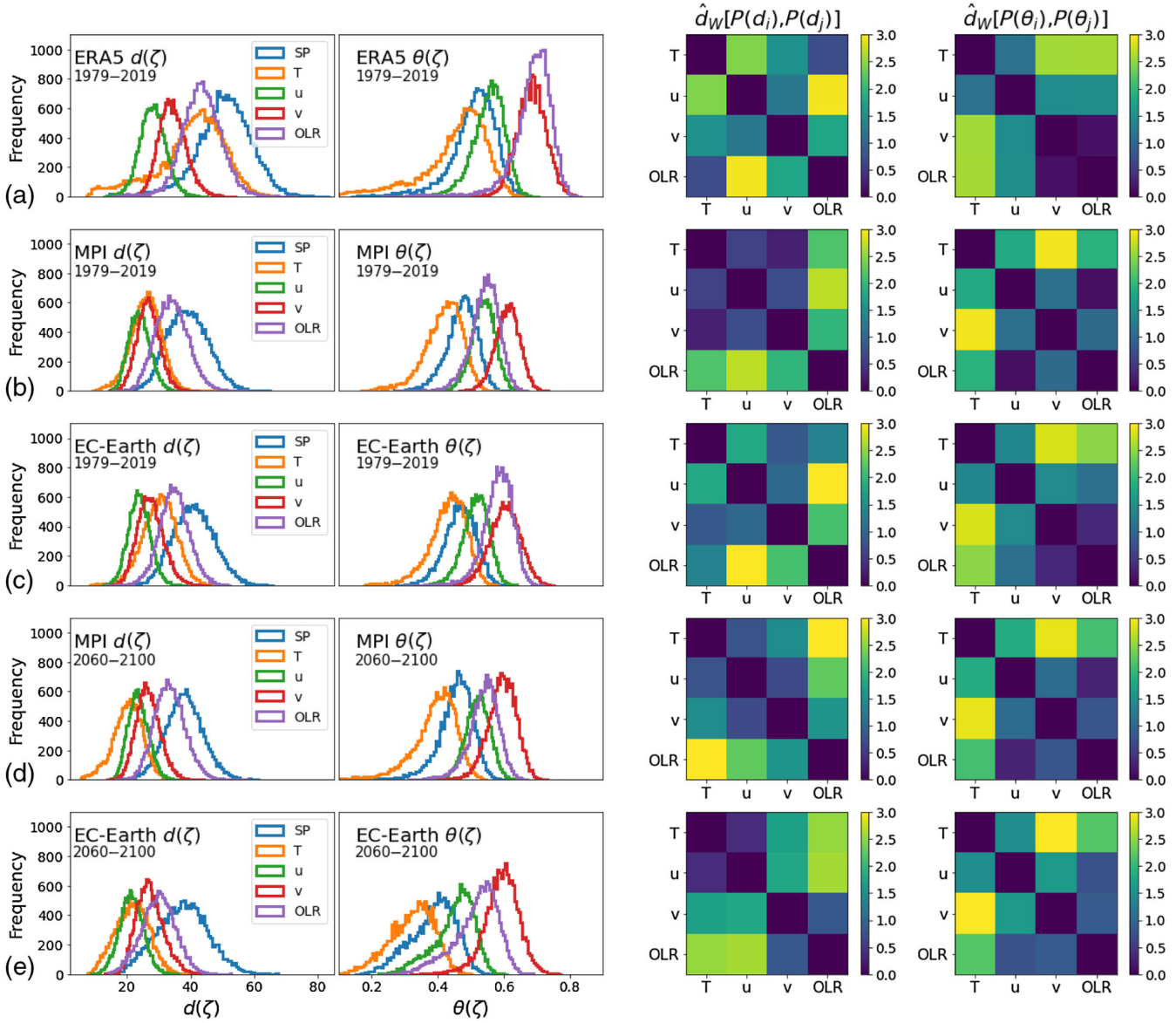

FIG. 8. (a)–(e) First and second columns: histograms of local dimension $d(\zeta)$ and inverse of persistence $\theta(\zeta)$ for the multivariate representation (referred to as state space "SP") and univariate fields (i.e., T, u, v, and OLR). Third and fourth columns: Wasserstein distance between such histograms. For a given dataset, the variance among all distances is one. Rows (a)–(e): ERA5 reanalysis, MPI (1979–2019), EC-Earth (1979–2019), MPI (2060–2100), and EC-Earth (2060–2100).

other, rather than their absolute magnitude; therefore, each pairwise distance matrix is further standardized by its total standard deviation. This analysis is shown in Fig. 8 for present and future periods.

In terms of local dimension, the distributions for ERA5 [Fig. 8(a), left-hand panel] show that the manifold dimensionality of univariate fields is always smaller than the one embedded in a multivariate space, as to be expected. The lowest dimensionality characterizes the zonal velocity, followed by the meridional component, while similar average values (but not tails) are found for temperature and OLR. The strong skewness at low values in the temperature is linked to the most intense ENSO events, as shown previously. In MPI the temperature and velocity fields show similar values of local dimension with almost overlapping distributions between T and v. The OLR field has higher local dimensions. This lack of differentiation among variables is partially corrected in EC-Earth, but T and OLR distributions have different mean values.

By the end of the 21st century, the distributions of local dimension for temperature shift toward lower values in both models, while changes are very limited for the other variables. The increase in intrinsic predictability under the warming scenario around the equator is therefore linked to changes in the surface temperature field alone.

Moving to persistence, distributions in ERA5 display low mean values of θ for temperature and zonal velocity, and more limited predictability (higher θ) for OLR and v. EC-Earth again reproduces the relative scaling and the relative distance among the distributions better than MPI. In the future, the distributions shift to lower values in EC-Earth, especially for temperature, while they remain nearly unchanged in MPI. Pairwise distance matrices, characterizing the degree of similarities among distributions, are then shown in the third column of Fig. 8. They quantify from a dynamical perspective the relative ranking among MPI and EC-Earth noticed in Ref. [41]. EC-Earth is indeed in better agreement with the reanalysis in terms of its representation of the relative contributions of each field to the multivariable distribution.

VI. ROBUSTNESS OF DYNAMICAL SYSTEMS METRICS: INTERNAL VARIABILITY AND RESOLUTION

We conclude the presentation by analyzing the robustness of the dynamical system metrics to the internal variability of the (modeled) system and to the resolution chosen for the analysis.

A. Internal variability

We compute the local dimension and persistence metrics for four ensemble members in EC-Earth and in MPI. For the period 2060–2100 in MPI, we rely only on two members. For simplicity, this analysis focuses only on the temperature variable. Results are shown in Fig. 9. Chaotic trajectories of the same dynamical system are bounded to live on the same manifold and manifold properties are largely independent of the ensemble member,

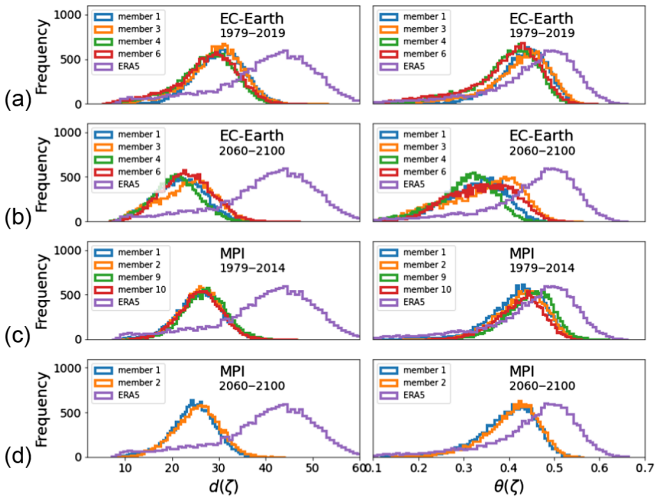

FIG. 9. Histograms of local dimension $d(\zeta)$ and inverse of persistence $\theta(\zeta)$ for the temperature field in EC-Earth and MPI. (a),(b) Results for EC-Earth. (c),(d) Results for MPI. The analysis is performed for the temperature variable.

provided that we sample such object well enough. The similarities among members are further quantified by the first four moments of the distributions in Table I. This analysis further shows that the 40 years considered are sufficient to characterize the attractor.

B. Resolution dependence

To compute the dynamical system metrics we first define an observable as $g(\mathbf{X}(t),\zeta) = -\log(\delta(\mathbf{X}(t),\zeta))$, where $\mathbf{X}(t) \in \mathbb{R}^{T,N}$ represents the high-dimensional trajectory in a state space of dimensionality N (Sec. III). Here the function $\delta(\mathbf{X}(t),\zeta)$ represents the Euclidean distance between a state space point ζ and the trajectory $\mathbf{X}(t)$. Computations of distances in high dimensions are affected by the known "curse of dimensionality" [75], and it is therefore important to check how results differ with the fields' resolution.

Here we focus on the d and θ metrics in the multivariate and univariate representations at two different resolutions.

For ERA5 we consider the following.

(i) The upscaled resolution of 1° in latitude and 1.5° in longitude adopted in most of our analysis. This implies a multivariate embedding in a $N = 17\,092$-dimensional state space and a univariate embedding in a $N = 4273$-dimensional state space.

(ii) A higher resolution with 0.5° in both latitude and longitude. This implies a multivariate embedding in

TABLE I. First four moments of the distributions of local dimensions and persistence. μ, σ, γ, κ are the mean, standard deviation, skewness, and kurtosis of the different histograms, respectively. The subscript d or θ indicates whether the analysis focuses on local dimension or persistence. Note that skewness and kurtosis for a normal distribution equal 0 and 3, respectively. "MPI mi" or "EC-E. mi" indicates the ith member of the respective model ensemble.

	μ_d	σ_d	γ_d	κ_d	μ_θ	σ_θ	γ_θ	κ_θ
1979–2019								
MPI m1	25.72	4.95	−0.26	3.38	0.42	0.06	−0.86	4.35
MPI m2	26.15	4.53	−0.13	3.17	0.43	0.05	−0.55	3.55
MPI m9	26.92	4.63	−0.05	3.22	0.44	0.05	−0.61	3.58
MPI m10	25.98	4.58	−0.11	3.14	0.42	0.05	−0.55	3.28
EC-E. m1	30.05	5.36	−0.3	3.42	0.43	0.06	−0.79	4.2
EC-E. m3	29.23	5.87	−0.49	3.39	0.42	0.07	−0.95	3.89
EC-E. m4	27.22	6.21	−0.55	3.32	0.39	0.07	−0.97	3.86
EC-E. m6	27.34	6.68	−0.58	3.27	0.39	0.08	−1.05	3.91
2060–2100								
MPI m1	24.39	4.83	−0.37	3.43	0.4	0.06	−0.84	3.99
MPI m2	25.26	4.87	−0.2	3.09	0.4	0.06	−0.72	3.55
EC-E. m1	21.54	5.16	0.00	2.67	0.32	0.07	−0.46	2.84
EC-E. m3	22.91	5.87	−0.21	2.50	0.34	0.08	−0.53	2.54
EC-E. m4	20.76	4.89	−0.10	2.90	0.3	0.06	−0.58	3.18
EC-E. m6	23.01	5.43	0.05	2.86	0.34	0.07	−0.26	2.57

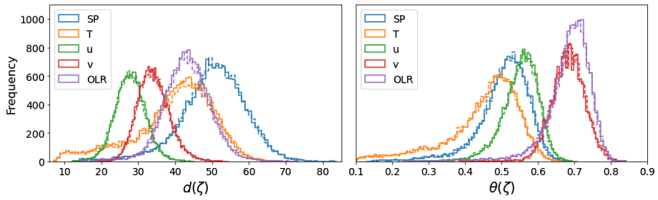

FIG. 10. Histograms of local dimension $d(\zeta)$ and inverse of persistence $\theta(\zeta)$ for the multivariate representation (referred to as state space "SP") and univariate fields (i.e., T, u, v, and OLR) in ERA5. Dashed lines: univariate and SP representations living in $N = 25\,239$ and $N = 100\,956$, respectively. Solid lines: univariate and SP representations living in $N = 4273$ and $N = 17\,092$, respectively.

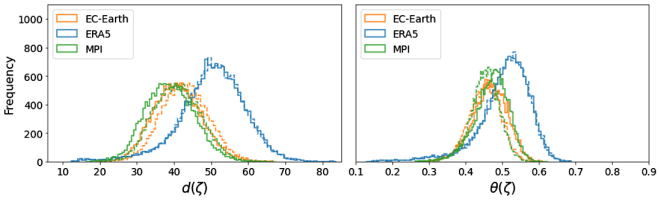

FIG. 11. Histograms of local dimension $d(\zeta)$ and inverse of persistence $\theta(\zeta)$ for the multivariate representation in the two models. Dashed lines: high resolution case, with ERA5, EC-Earth, and MPI state space dimensionality N equal to 100956, 527809, and 28344, respectively. Solid lines: upscaled (same) resolution $N = 17\,092$ for ERA5 and both models.

a $N = 100\,956$ state space and a univariate embedding in a $N = 25\,239$-dimensional state space.

Results are independent of resolution, given that the system has a large-scale imprinting which is captured in both cases (as shown in Fig. 10).

The same is verified in the models, where we compare the multivariate state space dynamics for the upscaled resolution of 1° in latitude and 1.5° in longitude and the original model resolution, for which $N = 52\,780$ for EC-Earth and $N = 28\,344$ for MPI. The analysis is shown in Fig. 11. Differences among models with original or upscaled resolutions are minimal. For reference, we also plot the ERA5 results for the low and high resolution case.

VII. DISCUSSION AND FUTURE WORK

This work introduces a powerful framework stemming from dynamical system theory to investigate climate variability and account for its spatiotemporal and multivariable dependency. The methodology is based on the assumption that the high-dimensional trajectory of the climate system lives on a lower-dimensional manifold [3,19].

Characterizing the topology and quantifying the geometrical properties of the climate attractor, alongside dynamical properties of the trajectories on the manifold, offers a much-needed, robust framework for dimensionality reduction in climate studies. Here we considered the tropical Pacific and four variables that are key to its variability, and explored the high-dimensional system's dynamics in an observational dataset, ERA5, and in realizations of two state-of-the-art climate models, MPI and EC-Earth. The analysis studied the high-dimensional tropical Pacific dynamics through manifold learning algorithms and dynamical system metrics. This provided a first estimate of the dimensionality of the tropical Pacific manifold, which is ∼50 for ERA5, and around 39 in the models.

A first, important result is that in ERA5 the nonlinear algorithm always shows a faster saturation of the residual variance compared to PCA, so that the dynamics can be projected onto fewer dimensional components. Independent of the dataset, the dynamics lives on a torus if the seasonal cycle is included. While the torus is topologically similar among datasets, large excursions in correspondence of El Niño events are found in the observations but are absent in the models. Furthermore, the spatial signature of the seasonal cycle, described by the first Isomap component, is biased in both models. By analyzing the anomalies, we showed that the two models have qualitative and quantitative similarities in the geometrical properties of their manifold, despite different resolutions and different choices in the representation of small-scale, unresolved processes. The PCA residual variance shares similarities among observational and modeled datasets, while the Isomap residual variance differs from that of PCA in the reanalysis while remaining similar in both models. This implies that both MPI and EC-Earth struggle in capturing the nonlinear topological characteristics of the observed manifold, and they do so in a similar way. Differences between the observed and modeled ENSO, on the other hand, are model dependent and larger in MPI for the historical period.

We stress that a key aspect of this work is the inclusion of multiple variables for a more comprehensive and robust quantification of the system's dynamics and feedbacks. The comparison between multivariate and univariate properties of the attractors quantifies the relative contribution of each field and allows for evaluating how this contribution may change over time. In the tropical Pacific the temperature field dominates the variance in both models and reanalysis, but the correlation between each field and the embedded trajectory differs in the models compared to the reanalysis and evolves differently in the two models in the scenario considered. It is also interesting to see how the relative role of the variables differs among the models, while contributing to a similar attractor in the historical periods, when models can be tuned through parameter choices toward the observations.

This work opens the way to evaluating the attractor trajectories over time and comparing them to observations, to better constrain climate sensitivity and the evolution of

climate feedbacks, both imperative to predict the likelihood of tipping points in the system.

By adopting the local geometry and persistence metrics to characterize the attractor's properties, we neatly visualized the day-to-day predictability potential during ENSO events, the El Niño and La Niña asymmetry, and model biases with regard to both aspects. Differences between the attractor in the models and reanalysis are not limited to strong ENSO events. Indeed, the region with both high $d(\zeta)$ and $\theta(\zeta)$, which is the most explored in the reanalysis and corresponds to ENSO-neutral days, is very seldom occupied by the models, implying that the representation of locally unstable motions and their influence on large-scale climate dynamics continues to elude current climate models, and such elusion is not amended in EC-Earth, despite being run at higher resolution than MPI.

These results point to topological (global, in state space) and geometrical (local, in state space) differences between observationally based data and climate model outputs, which can be evaluated considering one simulation, without the need for computationally expensive ensembles. The local scale chaoticity of the climate modeled system remains underestimated in both models, notwithstanding their different resolutions. Furthermore, the relationships among variables, which set their contribution to the global attractor and are fundamental to the evolution of the climate systems, are misrepresented, in different ways, in both models, and more so in MPI. The quantification of these relationships is a key, novel outcome of this work.

The framework we propose can be adopted to evaluate in a straightforward, robust way the impact of parametrizations on the (modeled) climate manifold and therefore assessing their impact on the large-scale dynamics. Most importantly, the analysis sets the stage for manifold learning approaches to climate modeling and climate prediction based not only on small-scale process understanding (machine learning application to subgrid scale parametrizations), but also on the characterization of the global climate system topology and the relations among variables (relational probabilistic models [88]). In the future, novel approaches stemming from data-driven dynamics and control (see, e.g., Refs. [89,90]) could be adopted to learn reduced-order models governing the evolution of the effective degrees of freedom of the system, therefore providing a useful alternative to the traditional partial differential equation approaches adopted in climate science.

A PYTHON implementation of dynamical system metrics is freely available [91]. For the PCA and Isomap algorithm we adopted the implementation in the Scikit-learn library [92]. A Github repository with examples and updates on current work can be found at Ref. [93].

ACKNOWLEDGMENTS

We thank Predrag Cvitanović and the ChaosBook [94] team who influenced many of the ideas in this work. We thank Yoann Robin for the online implementation of dynamical system metrics and Jost von Hardenberg for clarifications on the EC-Earth3, CMIP6 runs. F. F. acknowledges helpful discussions on the topic with Sebastián Ortega, Sylvia Sullivan, and Ilias Fountalis. The authors were supported by the Department of Energy, Regional and Global Model Analysis (RGMA) Program, Grant No. 0000253789. The revision process benefitted from the KITP Program "Machine Learning and the Physics of Climate" supported by the National Science Foundation under Grant No. NSF PHY-1748958. Furthermore, we especially thank the editorial team and two anonymous reviewers for their insightful comments that have greatly improved this paper.

APPENDIX A: IS THE MANIFOLD NONLINEAR? EUCLIDEAN VERSUS GEODESIC DISTANCES

To prove that the tropical Pacific manifold is indeed nonlinear, we compute the Euclidean and geodesic distances between all pairs of points $A, B \in \mathbb{R}^N$, where N is the dimensionality of the state space. Under the assumption that the high-dimensional data live on a low-dimensional object $\mathcal{M} \in \mathbb{R}^d$ (with $d \ll N$), we face two possibilities.

(i) The manifold $\mathcal{M}$ is linear. In this case the Euclidean distances between each pair of points A and B have to be equal to their geodesic distances *along* the manifold.

(ii) The manifold $\mathcal{M}$ is nonlinear (i.e., $\mathcal{M}$ is a curved object). In this case the geodesic distances *along* the manifold between each pair of points A and B are *always* greater than their Euclidean distances. This follows from the simple fact that the Euclidean distance is the shortest distance between two points.

Computing the geodesic distance is an important step of the Isomap algorithm [42].

To compute the geodesic distances *along* the manifold we assume that, while nonlinear, the manifold is *locally* flat around a radius of K points. It follows that the distances between each pair of points inside their K neighborhood is the Euclidean distance δ_E. We can therefore construct a weighted graph such that (a) each point i and j is connected if inside their K neighborhood and (b) their connection is weighted by the distance $\delta_E(i, j)$. The geodesic distance $\delta_G(i, j)$ is then the shortest path between each pair (i, j). For the shortest path computation we adopted the Floyd-Warshall algorithm [95].

We show the result in Fig. 12 for the period 1979–2019. Euclidean and geodesic distances are on the x and y axis,

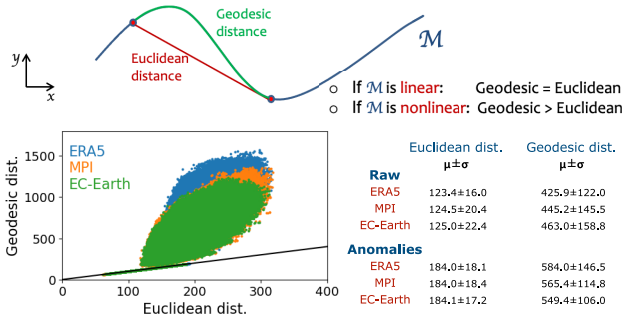

FIG. 12. Top: schematic to explain differences between geodesic (*along* the manifold) and Euclidean distances. In the schematic the state space is two dimensional and the manifold $\mathcal{M}$ is simply a line. Bottom: geodesic (*along* the manifold) and Euclidean distances between each pair of points in each dataset for anomalies (no seasonality and no trends) in the period 1979–2019. Points are above the diagonal, therefore quantifying the intrinsic nonlinearity of the low-dimensional manifold. Distances are computed in the full 17 092-dimensional state space. Mean μ and standard deviation σ for all three datasets are reported in the case of raw and anomalies. Raw data include trends.

respectively. As points are above the diagonal, we can conclude that the manifold is indeed nonlinear for all three datasets analyzed. In this paper, $K = 10$ days. Note of caution: the raw data include trends while the dataset with anomalies has been linearly detrended.

APPENDIX B: SEASONAL CYCLE

In Fig. 13(a), we consider the first Isomap component as described in Sec. IV A. This component represents the seasonal cycle. The Pearson correlation across datasets is always higher than 0.95, indicating that the temporal characteristics of the seasonal cycle are well captured by the two models. Independent of the dataset we find correlations higher than 0.98 between the first components of Isomap and PCA. Similarities across datasets in this first component do not imply similar spatial projections (i.e., the seasonal component of a model may be linked to biased regional processes even if highly correlated to observations). Spatial projections are visualized as the linear regression of an Isomap component onto each variable, and they therefore capture differences in variance. In Fig. 13(b) we show the spatial signature of the first Isomap component for ERA5 and the differences between MPI and EC-Earth. All time series in this analysis are standardized to unit variance. Future studies will focus on differences in the signals' amplitudes.

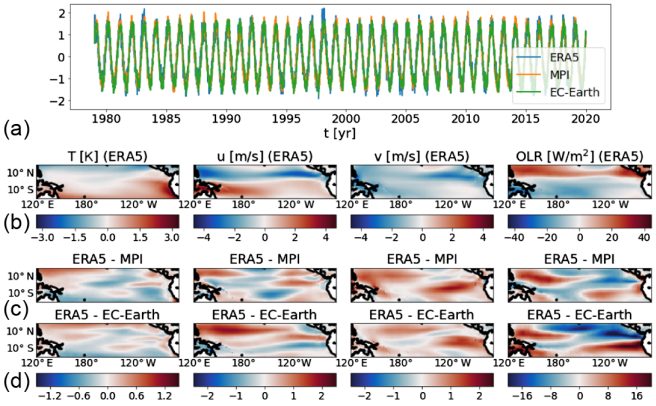

FIG. 13. (a) First (standardized) Isomap component for observations (ERA5) and the two CMIP6 models (MPI and EC-Earth). Each component is standardized to unit variance. Correlations with the first principal component is higher than 0.98 independent of the dataset. (b), top row: linear regression of the first Isomap component onto ERA5. (c,d), bottom rows: differences between spatial projections of ERA5 and the two models. Note that the projections are spatial signatures of a single eigenvector and not four.

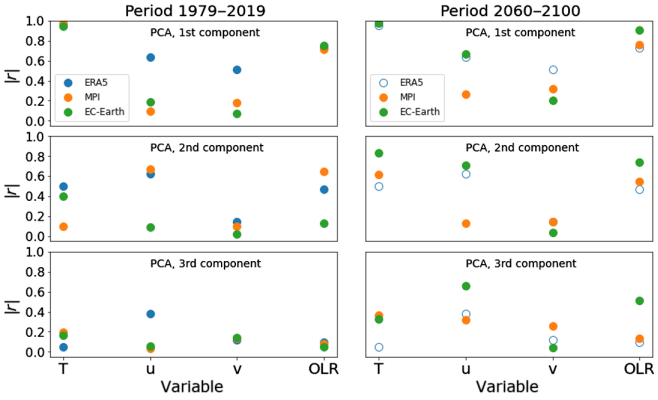

FIG. 14. First row: correlations (in absolute value) between the first PC of each variable (on the x axis) and the multivariate case in ERA5 and models. Second and third rows: same as first row but for the second and third Isomap components.

APPENDIX C: MULTIVARIATE PCA

In Fig. 14 we compare the first three PCs obtained in the multivariate representation with the univariate case.

APPENDIX D: ENSO: PROJECTIONS, SPECTRAL PROPERTIES, AND SPATIAL SIGNATURES

In Figs. 15(a)–15(c), 16(a), and 16(b) we show the first Isomap and PCA components and their Fourier spectra in the periods 1979–2019 and 2060–2100, respectively. Both components have been standardized to unit variance for comparison. Figures 15(d)–15(f), 16(c), and 16(d) show the spatial signature of the Isomap component on the four fields analyzed.

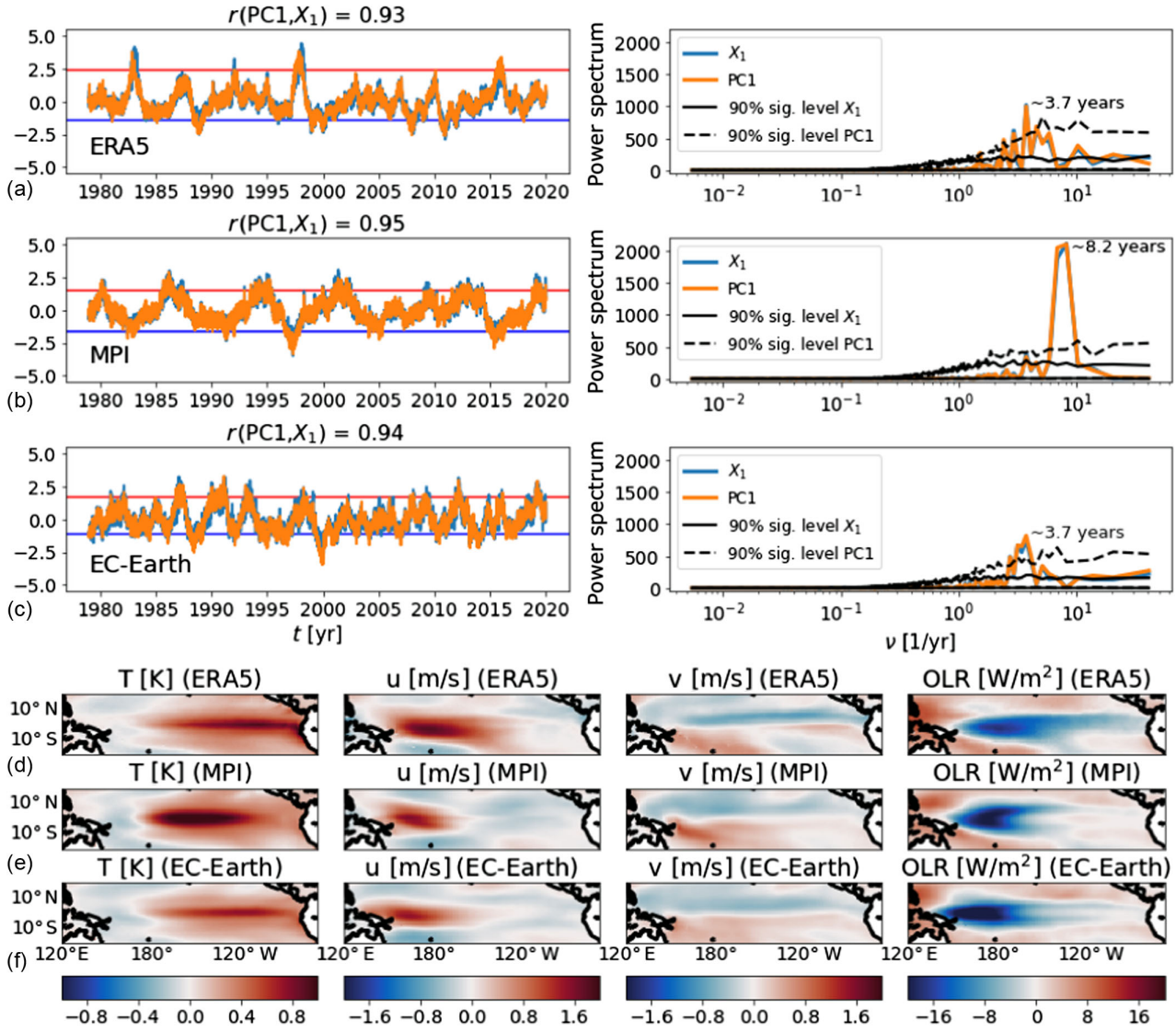

FIG. 15. (a)–(c), left-hand column: first (standardized) Isomap (X_1) and principal component (PC1) for observations (ERA5) and the two CMIP6 models (MPI and EC-Earth). Red (blue) lines represent the median of positive (negative) values in the 0.1 quantile (10th percentile) of the joint PDF of d and θ (see Fig. 5). Right-hand column: correspondent Fourier spectra. The spectral significance has been tested under the null hypothesis of red noise [96–98]. (d)–(f) Projection of the first (embedded) Isomap component on the various datasets.

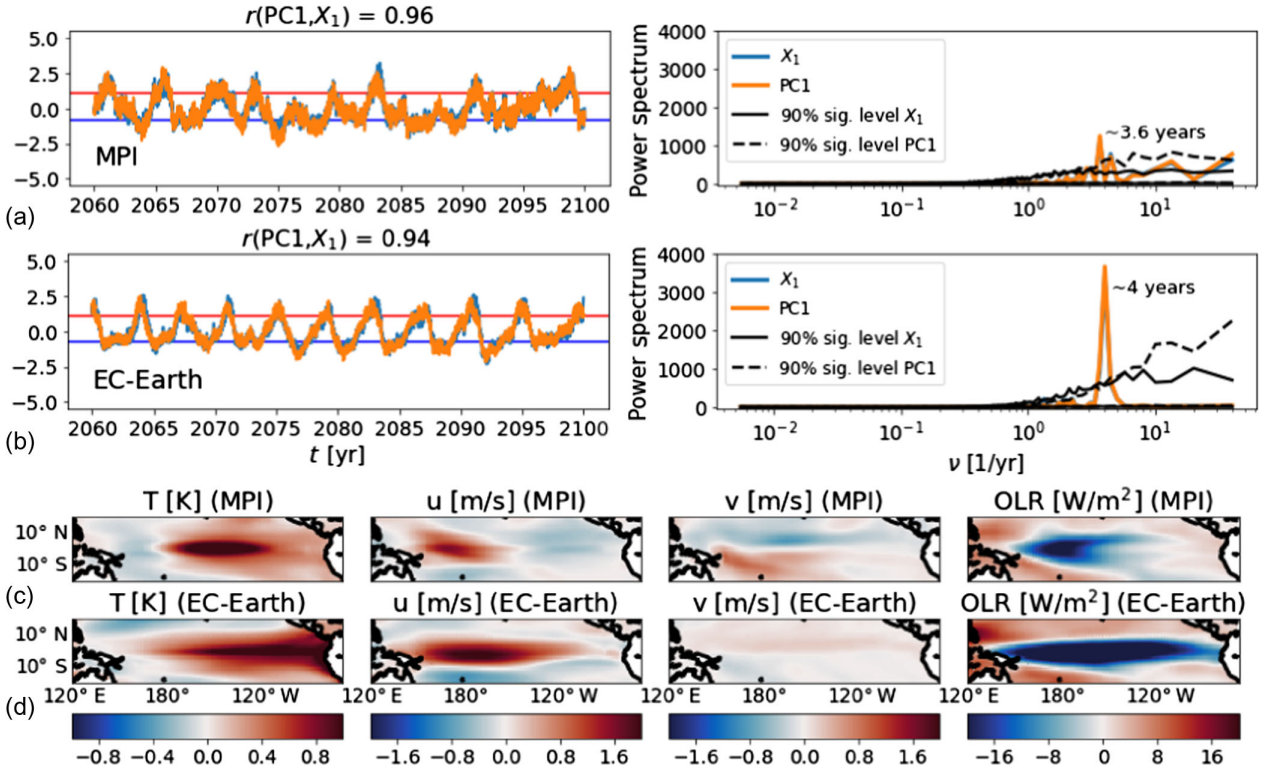

FIG. 16. (a),(b), left-hand column: first (standardized) Isomap (X_1) and principal component (PC1) for the two CMIP6 models (MPI and EC-Earth). Red (blue) lines represent the median of positive (negative) values in the 0.1 quantile (10th percentile) of the joint PDF of d and θ (see Fig. 5). Right-hand column: Fourier spectra. The spectral significance has been tested under the null hypothesis of red noise [96–98]. (c),(d) Projection of the first (embedded) Isomap component on the modeled datasets. The period analyzed is 2060–2100 under the SSP585 scenario.

APPENDIX E: MULTIVARIATE OR UNIVARIATE?

Figure 17 shows the first Isomap for the temperature field in ERA5. The seasonal cycle deviates from its canonical behavior in correspondence to large El Niño events. In Figs. 18 and 19 we compare the first Isomap component obtained in the multivariate representation with the univariate case, for MPI and EC-Earth, respectively.

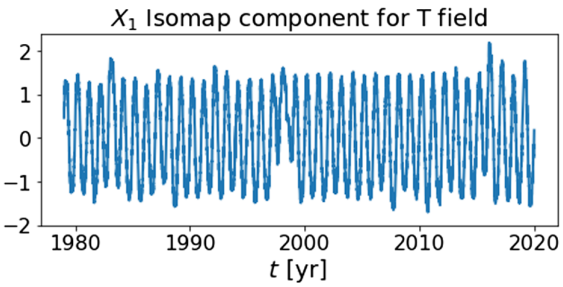

FIG. 17. First Isomap component of temperature anomalies in the ERA5 dataset.

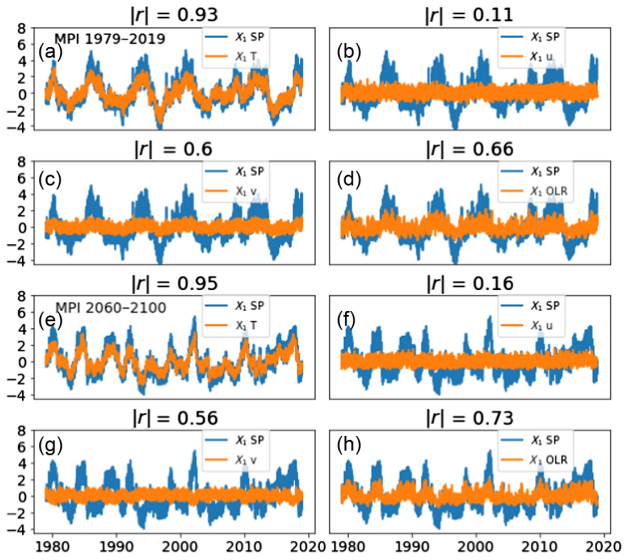

FIG. 18. First Isomap (X_1) component for anomalies in periods 1979–2019 and 2060–2100 in the MPI model in panels (a–d) and (e–h) respectively. In each case components have been standardized so that the *total* variance is equal to 1. Projections in the multivariate case are shown in blue and labeled as "SP" (state space). Projections for each univariate field are shown in red. Atop of each plot we report the correlation coefficient (in absolute value) between the projections in the multivariate and univariate cases.

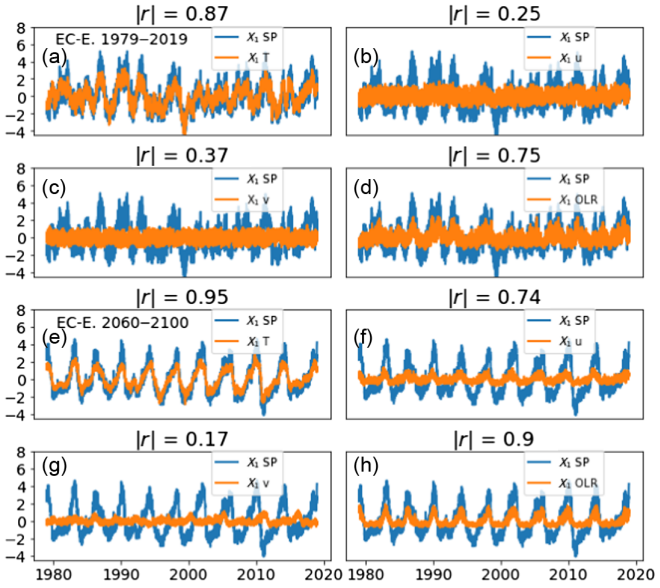

FIG. 19. First Isomap (X_1) component for anomalies in periods 1979–2019 and 2060–2100 in the EC-Earth model in panels (a–d) and (e–h) respectively. In each case components have been standardized so that the *total* variance is equal to 1. Projections in the multivariate case are shown in blue and labeled as "SP" (state space). Projections for each univariate field are shown in red. Atop of each plot we report the correlation coefficient (in absolute value) between the projections in the multivariate and univariate cases.

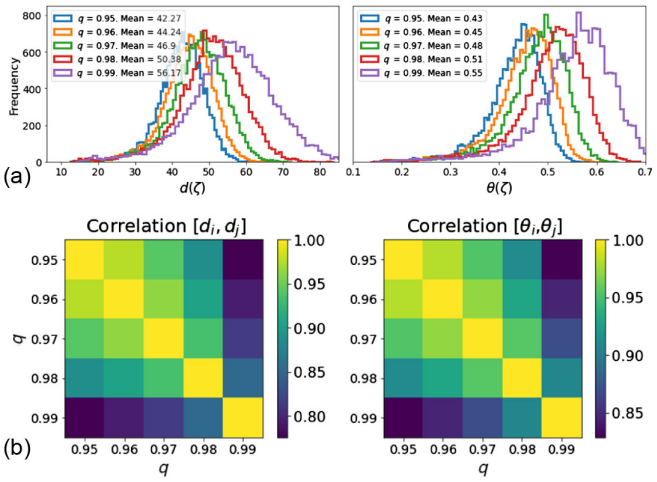

FIG. 20. (a) Histograms of local dimension d and persistence θ identified using different q. (b) Pairwise correlation matrix of d and θ identified using different q. The analysis is performed on the multivariate state space representation.

APPENDIX F: FIRST FOUR MOMENTS OF THE DISTRIBUTIONS OF LOCAL DIMENSIONS AND PERSISTENCE

In Table II, we show the first four moments of the distributions in Fig. 6. The fourth moment (i.e., kurtosis) in ERA5 highlights the presence of larger values in the tails of both the d and θ distributions.

APPENDIX G: DEPENDENCE ON q

We tested the robustness of the dynamical system metrics under the choice of threshold q (see Sec. III) for the multivariate case. Robustness is evaluated for the range $q \in [0.95, 0.99]$. We propose two analyses: first we compute histograms of both d and θ and, second, we compare the temporal variability of the two metrics by looking at their correlation. The analysis is shown in Fig. 20. We first look at the average manifold dimension. Values vary from ∼42 to ∼56. This is quite a small range as we started from a noisy, ∼17 000-dimensional dynamical system. In terms of variability [see Fig. 20(b)]. we see large correlations independently of the d or θ variable. The largest outlier is $q = 0.99$.

We choose the value of $q = 0.98$, as it gives similar results with the $q = 0.95$, 0.96, 0.97 thresholds while still being a very high quantile (preferred as we are quantifying statistics of extremes).

TABLE II. First four moments of the distributions of local dimensions and persistence. μ, σ, γ, κ are the mean, standard deviation, skewness, and kurtosis of the PDFs, respectively. The subscript d or θ indicates the different PDFs. Note that skewness and kurtosis for a normal distribution equal 0 and 3, respectively.

	μ_d	σ_d	γ_d	κ_d	μ_θ	σ_θ	γ_θ	κ_θ
1979–2019								
ERA5	50.38	8.85	−0.59	4.21	0.51	0.07	−1.28	5.92
MPI	38.92	6.45	0.11	2.88	0.47	0.05	−0.58	3.65
EC-E.	40.91	6.41	0.15	2.90	0.47	0.05	−0.32	3.10
2060–2100								
MPI	37.89	5.98	0.06	3.19	0.45	0.05	−0.82	4.42
EC-E.	37.33	8.02	0.07	2.76	0.38	0.07	−0.46	2.75

[1] R. Knutti and M. A. A. Rugenstein, *Feedbacks, Climate Sensitivity and the Limits of Linear Models*, Phil. Trans. R. Soc. A **373**, 20150146 (2015).
[2] E. Hopf, *A Mathematical Example Displaying Features of Turbulence*, Comm. Appl. Math. **1**, 303 (1948).
[3] P. Cvitanović, R. Artuso, R. Mainieri, G. Tanner, and G. Vattay, Chaos: Classical and Quantum, ChaosBook.org, Niels Bohr Institute, Copenhagen (2016), https://physics.gatech.edu/user/predrag-cvitanovic.
[4] C. Foias, G. R. Sell, and R. Temam, *Inertial Manifolds for Nonlinear Evolutionary Equations*, J. Diff. Eq. **73**, 309 (1988).
[5] 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).
[6] J. F. Gibson, J. Halcrow, and P. Cvitanović, *Visualizing the Geometry of State Space in Plane Couette Flow*, J. Fluid Mech. **611**, 107 (2008).

[7] P. Cvitanović, *Recurrent Flows: The Clockwork behind Turbulence*, J. Fluid Mech. **726**, 1 (2013).
[8] F. Christiansen, P. Cvitanović, and V. Putkaradze, *Spatio-temporal Chaos in Terms of Unstable Recurrent Patterns*, Nonlinearity **10**, 55 (1997).
[9] J. Gallego, M. G. Perich, R. H. Chowdhury, S. A. Solla, and L. E. Miller, *Long-Term Stability of Cortical Population Dynamics Underlying Consistent Behavior*, Nat. Neuroscience **23**, 260 (2020).
[10] E. D. Remington, S. W. Egger, D. Narain, J. Wang, and M. Jazayeri, *A Dynamical Systems Perspective on Flexible Motor Timing*, Trends Cognit. Sci. **22**, 938 (2018).
[11] H. Sohn, D. Narain, N. Meirhaeghe, and M. Jazayeri, *Bayesian Computation through Cortical Latent Dynamics*, Neuron **103**, 934 (2019).
[12] M. Jazayeri and S. Ostojic, *Interpreting Neural Computations by Examining Intrinsic and Embedding Dimensionality of Neural Activity*, Curr. Opin. Neurobiol. **70**, 113 (2021).
[13] M. M. Churchland, J. P. Cunningham, M. T. Kaufman, J. D. Foster, P. Nuyujukian, S. I. Ryu, and K. V. Shenoy, *Neural Population Dynamics during Reaching*, Nature (London) **487**, 51 (2012).
[14] U. Cohen, S. Y. Chung, D. D. Lee, and H. Sompolinsky, *Separability and Geometry of Object Manifolds in Deep Neural Networks*, Nat. Commun. **11**, 746 (2020).
[15] C. Stephenson, Suchismita Padhy, A. Ganesh, Y. Hui, H. Tang, and S. Chung, *On the Geometry of Generalization and Memorization in Deep Neural Networks*, in *Proceedings of the International Conference on Learning Representations, Vienna* (2021), https://openreview.net/group?id=ICLR.cc/2021/Conference.
[16] T. Ahamed, A. C. Costa, and G. Stephens, *Capturing the Continuous Complexity of Behaviour in Caenorhabditis elegans*, Nat. Phys. **17**, 275 (2020).
[17] G. Stephens, B. Johnson-Kerner, W. Bialek, and W. S. Ryu, *Dimensionality and Dynamics in the Behavior of C. elegans*, PLoS Comput Biol **4**, e1000028 (2020).
[18] L. Saul and S. Roweis, *Think Globally, Fit Locally: Unsupervised Learning of Low Dimensional Manifolds*, J. Machine Learn. Res. **4**, 119 (2003), https://www.jmlr.org/papers/volume4/saul03a/saul03a.pdf.
[19] M. Ghil and V. Lucarini, *The Physics of Climate Variability and Climate Change*, Rev. Mod. Phys. **92**, 035002 (2020).
[20] G. Margazoglou, T. Grafke, A. Laio, and V. Lucarini, *Dynamical Landscape and Multistability of a Climate Mode*, Proc. R. Soc. A **477**, 20210019 (2021).
[21] L. Smith, *What Might We Learn from Climate Forecasts?*, Proc. Natl. Acad. Sci. U.S.A. **99**, 2487 (2002).
[22] R. Frigg, S. Bradley, H. Du, and L. Smith, *Laplace's Demon and the Adventures of His Apprentices*, Philos. Sci. **81**, 31 (2014).
[23] K. B. Tokarska, M. B. Stolpe, S. Sippel, E. M. Fischer, C. J. Smith, F. Lehner, and R. Knutti, *Past Warming Trend Constrains Future Warming in CMIP6 Models*, Sci. Adv. **6**, eaaz9549 (2020).
[24] B. C. O'Neill, C. Tebaldi, D. P. van Vuuren, V. Eyring, P. Friedlingstein, G. Hurtt, R. Knutti, E. Kriegler, J.-F. Lamarque, J. Lowe, G. A. Meehl, R. Moss, K. Riahi, and B. M. Sanderson, *The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6*, Geosci. Model Dev. **9**, 3461 (2016).
[25] L. Smith and N. Stern, *Uncertainty in Science and Its Role in Climate Policy*, Phil. Trans. R. Soc. A **369**, 4818 (2011).
[26] V. Eyring *et al.*, *Taking Climate Model Evaluation to the Next Level*, Nat. Clim. Change **9**, 102 (2019).
[27] A. Timmermann *et al.*, *El Niño-Southern Oscillation Complexity*, Nature (London) **559**, 535 (2018).
[28] S. Brönnimann, J. Luterbacher, J. Staehelin, T. Svendby, G. Svendby, and T. Svendby, *Extreme Climate of the Global Troposphere and Stratosphere in 1940–42 Related to El Niño*, Nature (London) **431**, 971 (2004).
[29] J.-Y. Yu, Y. Zou, S. T. Kim, and T. Lee, *The Changing Impact of El Niño on US Winter Temperatures*, Geophys. Res. Lett. **39**, L15702 (2012).
[30] S. He, J.-Y. Yu, S. Yang, and S.-W. Fang, *ENSO's Impacts on the Tropical Indian and Atlantic Oceans via Tropical Atmospheric Processes: Observations versus CMIP5 Simulations.*, Clim. Dyn. **54**, 4627 (2020).
[31] R. Döscher *et al.*, *The EC-Earth3 Earth System Model for the Climate Model Intercomparison Project 6*, Geosci. Model Devel. Discuss., **15**, 2973 (2022).
[32] I. M. Held *et al.*, *Structure and Performance of GFDL's CM4.0 Climate Model*, J. Adv. Model. Earth Syst. **11**, 3691 (2021).
[33] P. Sepulchre *et al.*, *IPSL-CM5A2—An Earth System Model Designed for Multi-Millennial Climate Simulations*, Geosci. Model Dev. **13**, 3011 (2020).
[34] Z. Lu, Z. Fu, L. Hua, N. Yuan, and L. Chen, *Evaluation of ENSO Simulations in CMIP5 Models: A New Perspective Based on Percolation Phase Transition in Complex Networks*, Sci. Rep. **8**, 14912 (2018).
[35] I. Fountalis, A. Bracco, and C. Dovrolis, *ENSO in CMIP5 Simulations: Network Connectivity from the Recent Past to the Twenty-Third Century*, Clim. Dyn. **45**, 511 (2015).
[36] F. Falasca, A. Bracco, A. Nenes, and I. Fountalis, *Dimensionality Reduction and Network Inference Using δ-MAPS: Application to the CESM Large Ensemble Sea Surface Temperature*, J. Adv. Model. Earth Syst. **11**, 1479 (2019).
[37] H. Hotelling, *Analysis of a Complex of Statistical Variables into Principal Components*, J. Educ. Psychol. **24**, 417 (1933).
[38] H. Hersbach *et al.*, *The ERA5 Global Reanalysis*, Q. J. R. Meteorol. Soc. **146**, 1999 (2020).
[39] V. Eyring, S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor, *Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) Experimental Design and Organization*, Geosci. Model Dev. **9**, 1937 (2016).
[40] O. Gutjahr, D. Putrasahan, K. Lohmann, J. H. Jungclaus, J.-S. von Storch, N. Brüggemann, H. Haak, and A. Stössel, *Max Planck Institute Earth System Model (MPI-ESM1.2) for the High-Resolution Model Intercomparison Project (HighResMIP)*, Geosci. Model Dev. **12**, 3241 (2019).
[41] J. T. Fasullo, *Evaluating Simulated Climate Patterns from the CMIP Archives Using Satellite and Reanalysis Datasets Using the Climate Model Assessment Tool (CMATv1)*, Geosci. Model Dev. **13**, 3627 (2020).

[42] J. Tenenbaum, V. de Silva, and J. Langford, *A Global Geometric Framework for Nonlinear Dimensionality Reduction*, Science **290**, 2319 (2000).
[43] D. Bueso, M. Piles, and G. Camps-Valls, *Nonlinear PCA for Spatio-Temporal Analysis of Earth Observation Data*, IEEE Trans. Geosci. Remote Sens. **58**, 5752 (2020).
[44] M. Belkin and P. Niyogi, *Laplacian Eigenmaps for Dimensionality Reduction and Data Representation*, Neural Comput. **15**, 1373 (2003).
[45] L. McInnes, J. Healy, and J. Melville, *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction*, arXiv:1802.03426.
[46] L. van der Maaten and G. Hinton, *Visualizing Data Using t-SNE*, J. Machine Learn. Res. **9**, 2579 (2008), https://www.jmlr.org/papers/volume9/vandermaaten08a/vandermaaten08a.pdf.
[47] S. Roweis and L. K. Saul, *Nonlinear Dimensionality Reduction by Locally Linear Embedding*, Science **290**, 2323 (2000).
[48] S. Chung and L. F. Abbott, *Neural Population Geometry: An Approach for Understanding Biological and Artificial Neural Networks*, Curr. Opin. Neurobiol. **70**, 137 (2021).
[49] D. Faranda, G. Messori, and P. Yiou, *Dynamical Proxies of North Atlantic Predictability and Extremes*, Sci. Rep. **7**, 41278 (2017).
[50] G. Messori and D. Faranda, *Technical Note: Characterising and Comparing Different Palaeoclimates with Dynamical Systems Theory*, Climate Past **17**, 545 (2021).
[51] N. R. Moloney, D. Faranda, and Y. Sato, *An Overview of the Extremal Index*, Chaos **29**, 022101 (2019).
[52] T. Caby, D. Faranda, S. Vaienti, and P. Yiou, *On the Computation of the Extremal Index for Time Series*, J. Stat. Phys. **179**, 1666 (2020).
[53] A. C. M. Freitas, J. M. Freitas, and M. Todd, *Hitting Time Statistics and Extreme Value Theory*, Probab. Theory Relat. Fields **147**, 675 (2010).
[54] V. Lucarini *et al.*, *Extremes and Recurrence in Dynamical Systems*, Pure and Applied Mathematics (Wiley, New York, 2016).
[55] G. Messori, N. Harnik, E. Madonna, O. Lachmy, and D. Faranda, *A Dynamical Systems Characterization of Atmospheric Jet Regimes*, Earth Sys. Dyn. **12**, 233 (2021).
[56] P. De Luca, G. Messori, D. Faranda, P. J. Ward, and D. Coumou, *Compound Warm–Dry and Cold–Wet Events over the Mediterranean*, Earth Syst. Dyn. **11**, 793 (2020).
[57] P. De Luca, G. Messori, F. Pons, and D. Faranda, *Dynamical Systems Theory Sheds New Light on Compound Climate Extremes in Europe and Eastern North America*, Q. J. R. Meteorol. Soc. **146**, 1636 (2020).
[58] D. Rodrigues, C. Alvarez-Castro, G. Messori, P. Yiou, Y. Robin, and D. Faranda, *Dynamical Properties of the North Atlantic Atmospheric Circulation in the Past 150 Years in CMIP5 Models and the 20CRv2c Reanalysis*, J. Climate **31**, 6097 (2018).
[59] G. Messori, R. Caballero, and D. Faranda, *A Dynamical Systems Approach to Studying Midlatitude Weather Extremes*, Geophys. Res. Lett. **44**, 3346 (2017).
[60] A. Hochman, S. Scher, J. Quinting, J. Pinto, and G. Messori, *A New View of Heat Wave Dynamics and Predictability over the Eastern Mediterranean*, Earth Syst. Dyn. **12**, 133 (2021).
[61] A. Hochman, P. Alpert, T. Harpaz, H. Saaroni, and G. Messori, *A New Dynamical Systems Perspective on Atmospheric Predictability: Eastern Mediterranean Weather Regimes as a Case Study*, Sci. Adv. **5**, eaau0936 (2019).
[62] 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*, Nonlinear Processes Geophys. **29**, 17 (2022).
[63] W. Müller *et al.*, *A Higher-Resolution Version of the Max Planck Institute Earth System Model (MPI-ESM1.2-HR)*, J. Adv. Model. Earth Syst. **10**, 1383 (2018).
[64] I. Held and P. Phillipps, *Sensitivity of the Eddy Momentum Flux to Meridional Resolution in Atmospheric GCMs*, J. Climate **6**, 499 (1993).
[65] R. Kawamura, *A Rotated EOF Analysis of Global Sea Surface Temperature Variability with Interannual Interdecadal Scales*, J. Phys. Oceanogr. **24**, 707 (1994).
[66] B. C. Weare and J. S. Nasstrom, *Examples of Extended Empirical Orthogonal Function Analysis*, Mon. Weather Rev. **110**, 481 (1982).
[67] H. v. Storch and F. W. Zwiers, *Statistical Analysis in Climate Research* (Cambridge University Press, Cambridge, England, 1999).
[68] C. Bishop, *Pattern Recognition and Machine Learning* (Springer, New York 2006).
[69] I. Borg and P. Groenen, *Modern Multidimensional Scaling: Theory and Applications* (Springer-Verlag, New York, 1997), https://www.amazon.it/Modern-Multidimensional-Scaling-Theory-Applications/dp/1441920463.
[70] I. Ross, P. J. Valdes, and S. Wiggins, *ENSO Dynamics in Current Climate Models: An Investigation Using Nonlinear Dimensionality Reduction*, Nonlinear Processes Geophys. **15**, 339 (2008).
[71] A. Hannachi and A. G. Turner, *Isomap Nonlinear Dimensionality Reduction and Bimodality of Asian Monsoon Convection*, Geophys. Res. Lett. **40**, 1653 (2013).
[72] J. Theiler, *Estimating Fractal Dimension*, J. Opt. Soc. Am. A **7**, 1055 (1990).
[73] E. Altan, S. A. Solla, L. E. Miller, and E. J. Perreault, *Estimating the Dimensionality of the Manifold Underlying Multi-Electrode Neural Recordings*, PLoS Comput. Biol. **17**, e1008591 (2021).
[74] V. Lucarini, D. Faranda, and J. Wouters, *Universal Behaviour of Extreme Value Statistics for Selected Observables of Dynamical Systems*, J. Stat. Phys. **147**, 63 (2012).
[75] F. Pons, G. Messori, M. Alvarez-Castro, and D. Faranda, *Sampling Hyperspheres via Extreme Value Theory: Implications for Measuring Attractor Dimensions*, J. Stat. Phys. **179**, 1698 (2020).
[76] J. Pickands, *Statistical Inference Using Extreme Order Statistics*, Ann. Stat. **3**, 119 (1975), https://projecteuclid.org/journals/annals-of-statistics/volume-3/issue-1/Statistical-Inference-Using-Extreme-Order-Statistics/10.1214/aos/1176343003.full.

[77] M. Leadbetter, *Extremes and Local Dependence in Stationary Sequences*, Theory Relat. Fields **65**, 291 (1983), https://www2.stat.duke.edu/~scs/Projects/Stable/Laptop/Leadbetter1983.pdf.
[78] E. D. Lorenz, *Deterministic Nonperiodic Flow*, J. Atmos. Sci. **20**, 130 (1963).
[79] M. Süveges, *Likelihood Estimation of the Extremal Index*, Extremes **10**, 41 (2007).
[80] C. Ferro and J. Segers, *Inference for Clusters of Extreme Values*, J. R. Stat. Soc. Ser. B **65**, 545 (2003).
[81] R. L. Smith and I. Weissman, *Estimating the Extremal Index*, J. R. Stat. Soc. Ser. B **56**, 515 (1994).
[82] N. Maher, D. Matei, S. Milinski, and J. Marotzke, *ENSO Change in Climate Projections: Forced Response or Internal Variability?*, Geophys. Res. Lett. **45**, 390 (2018).
[83] S.-I. An, S.-K. Kim, and A. Timmermann, *Fokker-Planck Dynamics of the El Niño-Southern Oscillation*, Sci. Rep. **10**, 16282 (2020).
[84] S.-I. An and F.-F. Jin, *Nonlinearity and Asymmetry of ENSO*, J. Climate **17**, 2399 (2004).
[85] D. Faranda, M. Alvarez-Castro, G. Messori, D. Rodrigues, and P. Yiou, *The Hammam Effect or How a Warm Ocean Enhances Large Scale Atmospheric Predictability*, Nat. Commun. **10**, 1316 (2019).
[86] M. Brunetti, J. Kasparian, and C. Vérard, *Co-Existing Climate Attractors in a Coupled Aquaplanet*, Clim. Dyn. **53**, 6293 (2019).
[87] C. Villani, *Optimal Transport: Old and New* (Springer-Verlag, Berlin, 2009).
[88] L. Raedt, K. Kersting, S. Natarajan, and D. Poole, *Statistical Relational Artificial intelligence: Logic, Probability, and Computation*, in *Synthesis Lectures on Artificial Intelligence and Machine Learning* (2016), p. 189, 10.2200/S00692ED1V01Y201601AIM032.
[89] J.-C. Loiseau, B. Noack, and S. L. Brunton, *Sparse Reduced-Order Modelling: Sensor-Based Dynamics to Full-State Estimation*, J. Fluid Mech. **844**, 459 (2018).
[90] K. Champion, B. Lusch, N. Kutz, and S. Brunton, *Data-Driven Discovery of Coordinates and Governing Equations*, Proc. Natl. Acad. Sci. U.S.A. **116**, 22445 (2019).
[91] https://github.com/yrobink/CDSK/tree/master/python.
[92] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, *Scikit-learn: Machine Learning in Python*, J. Machine Learn. Res. **12**, 2825 (2011), https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf.
[93] https://github.com/FabriFalasca/climate-and-dynamical-systems.
[94] https://chaosbook.org/
[95] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, 3rd ed. (The MIT Press, Cambridge, MA, 2009).
[96] *Stochastic Climate Models: Workshop in Chorin, Germany 1999*, edited by P. Imkeller and J. S. V. Storch (Birkhäuser, Berlin, 20012001), Vol. 49, https://books.google.com.np/books?id=qeLHbxNspFwC&printsec=copyright#v=onepage&q&f=false.
[97] H. A. Dijkstra, E. Hernández-Garcia, C. Masoller, and M. Barreiro, *Networks in Climate* (Cambridge University Press, Cambridge, England, 2019).
[98] M. Allen and L. Smith, *Monte Carlo SSA: Detecting Irregular Oscillations in the Presence of Colored Noise*, J. Climate **9**, 3373 (1996).

Correction: The author list in Ref. [12] contained an error and has been fixed.