Variability of SST through Koopman Modes

ANTONIO NAVARRA,[a,b] JOE TRIBBIA,[c] STEFAN KLUS,[d] AND PAULA LORENZO-SÁNCHEZ[a,e]

[a] *Centro Euromediterraneo sui Cambiamenti Climatici, Bologna, Italy*
[b] *Dipartimento di Scienze Geologiche, Biologiche e Ambientali, Universita' di Bologna, Bologna, Italy*
[c] *National Center for Atmospheric Research, Boulder, Colorado*
[d] *School of Mathematical and Computer Sciences, Heriot–Watt University, Edinburgh, United Kingdom*
[e] *Dipartimento di Fisica e Astronomia, Universita' di Bologna, Bologna, Italy*

(Manuscript received 2 June 2023, in final form 26 February 2024, accepted 22 April 2024)

ABSTRACT: The majority of dynamical systems arising from applications show a chaotic character. This is especially true for climate and weather applications. We present here an application of Koopman operator theory to tropical and global sea surface temperature (SST) that yields an approximation to the continuous spectrum typical of these situations. We also show that the Koopman modes yield a decomposition of the datasets that can be used to categorize the variability. Most relevant modes emerge naturally, and they can be identified easily. A difference with other analysis methods such as empirical orthogonal function (EOF) or Fourier expansion is that the Koopman modes have a dynamical interpretation, thanks to their connection to the Koopman operator, and they are not constrained in their shape by special requirements such as orthogonality (as it is the case for EOF) or pure periodicity (as in the case of Fourier expansions). The pure periodic modes emerge naturally, and they form a subspace that can be interpreted as the limiting subspace for the variability. The stationary states therefore are the scaffolding around which the dynamics takes place. The modes can also be traced to the Niño variability and in the case of the global SST to the Pacific decadal oscillation (PDO).

SIGNIFICANCE STATEMENT: We compute the Koopman modes for the tropical and global SST, demonstrating that significant dynamical modes can be identified also in complex high-dimensional datasets with continuous spectra. The Koopman modes are then used to identify the stationary subspace, namely, the limit subspace for the invariant evolution of the system.

KEYWORDS: Sea surface temperature; Interannual variability; Tropical variability; Data science; Support vector machines

1. Introduction

Understanding the dynamical and physical origin of the weather and climate variability has been a primary motivation in the history of atmospheric and climate sciences. The fluctuations originated by the turbulence were very visible, but placing them in a coherent physical paradigm of forces and processes has been elusive. The zero order splitting of the tropospheric midlatitude circulation into a large component represented by the polar vortex and smaller amplitude deviations led to the representation of the turbulence as dominated by instabilities growing on a "basic state" through a well-defined life cycle. The instability model, supported by very significant theoretical formulations, showed remarkable success in discovering time scales and structures that were useful in the understanding of the atmospheric processes. Successive generalizations were developed using more and more realistic basic states (Simmons et al. 1983; Frederiksen 2000), and the original formulation limited to exponential modes (normal modes) was expanded to allow for algebraic growth modes (Farrell 1988, 1989).

However, the linear mode representation was not very appropriate to describe variability at longer time scales. Monthly and seasonal time scales could not really be explained as linear instabilities. Furthermore, the splitting between the basic state, usually some form of time or climatological mean, and the deviation included a substantial arbitrariness of the choice. Starting in the 1980s, the linear mode analysis, while still useful, was reduced to specific applications, and it had exhausted its utility as a primary conceptual tool of the atmosphere/climate variability.

At that time, another methodology became widely used, namely, empirical orthogonal functions (EOFs). Initially introduced by Lorenz (1956), EOFs were shown to be a very effective method to analyze variability in terms of decomposition of the variance. After that, they have been extensively used to analyze observation and simulation data. Several authors (Kutzbach 1970; Lau 1981; Trenberth and Paolino 1981; Quadrelli and Wallace 2004) showed how EOFs can give useful information on the main modes of variability and they were instrumental in discovering the relations with the sea surface temperature (SST), in the tropics and at global level. More complex and sophisticated variants of the EOFs were then derived (e.g., rotated EOFs, complex EOFs, multiparametric EOFs), resulting in extensive and almost ubiquitous

Denotes content that is immediately available upon publication as open access.

Corresponding author: Antonio Navarra, antonio.navarra@cmcc.it

DOI: 10.1175/JCLI-D-23-0335.1

applications [for a review (Hannachi et al. 2007)]. Extensions have been made to the time domain through extended EOF (Lau and Chan 1985) or singular spectrum analysis (Ghil et al. 2002), and nonlinear extensions have also been proposed and used (Giannakis and Majda 2012b,a).

EOFs are powerful and easy to compute and manipulate, but even the more advanced extensions have a significant drawback: they have no obvious relation with the dynamics of the field on which they are calculated. The physical and dynamical nature of EOFs is elusive, and certainly, they do not have any relationship with linear instability modes, and their interpretation as modes is difficult. In particular, they cannot be used to represent the evolution or a drastic simplification of the dynamics, as mentioned by Dommenget and Latif (2002).

However, the old works of Koopman (1931) and Koopman and Neumann (1932) provide an interesting alternative. These papers show that even a nonlinear system can be expressed as a linear system if you are willing to accept a very large number of linear modes. It is a global linearization rather than the local linearization typical of the instability problem, and as such, it does not require a reference basic state. Therefore, it removes the ambiguity of the choice of the reference state and realizes a different representation of the dynamics of the system. In the place of the usual picture of a dynamical system tracing a trajectory in the phase space, Koopman introduced a linear operator that evolves the state, or functions of the state, and he showed that this is always possible if the Koopman operator is allowed to be infinite-dimensional.

Such a theoretical result was largely ignored, until the recent development of powerful machine learning methods to compute the estimates of the Koopman operator from data. The Koopman operator approach (Mezić 2005, 2013; Das and Giannakis 2020; Budišić et al. 2012; Giannakis 2019; Das and Giannakis 2020) shows that for every dynamical system, there is a linear operator acting on a function space whose spectral properties completely characterize the dynamical system.

Several numerical algorithms have been then developed to estimate the Koopman operator, from the early work of Ulam (1960), to the recent *extended dynamic mode decomposition* (EDMD) (Williams et al. 2015a,b; Klus et al. 2016) and the *variational approach of conformation* (VAC) *dynamics* (Noé and Nüske 2013; Nueske et al. 2014). A review of these methods can be found in Klus et al. (2018), and further information can be found in Rowley et al. (2009), Tu et al. (2014), and McGibbon and Pande (2015). These results have allowed the development of practical algorithms that can be used to estimate the Koopman operator from observations and simulation data. The Koopman operator approach belongs to a class of approaches based on operator techniques that have seen several applications in climate problems in recent years. They have been used to investigate and identify phenomena (Froyland et al. 2021; Giannakis et al. 2015), forecast ENSO (Wang et al. 2020; Navarra et al. 2021), and forecast sea ice concentration (Hogg et al. 2020), and for the analysis of tropical convective variability on diurnal to seasonal time scales (Tung et al. 2014).

In a previous paper (Navarra et al. 2021), we have shown that it is possible to achieve the Koopman decomposition for a realistic observed dataset and we can set up a forecasting system with it. In this paper, we want to use the Koopman decomposition to investigate the dynamical significance of the Koopman decomposition and investigate the characteristics of the spectrum and distribution of the eigenvalues with the objective of identifying the most interesting modes. We will also rely on the Koopman modes (Mezić 2013), rather than the composites used in Froyland et al. (2021), so we can obtain the geographical distribution of the dynamical modes that will indicate locally the oscillatory or decaying behavior. The Koopman decomposition is more general than a Fourier analysis as it does not put any structure on the mode: A Fourier analysis will force the decomposition to be nondecaying and orthogonal, but a Koopman decomposition for a periodic system will obtain the same results as a Fourier analysis. Linear inverse model (LIM) methods (Penland and Sardeshmukh 1995; Penland 1996) can be shown to yield an approximation to the Koopman operator (Tu et al. 2014), but the aforementioned data-driven techniques can be extended nonlinearly using suitable kernels, resulting in better estimation of the Koopman operator.

Section 2 will outline the Koopman spectral decomposition, introducing the concept of the stationary modes using the algorithms of Klus et al. (2020). The Koopman modes of the tropical SST will then be shown in section 5, and those of the global ocean, excluding polar areas, will be discussed in section 6. Some discussion and the conclusions will close the paper in section 8.

2. Koopman spectral decomposition

Consider a dynamical system,

$$\frac{d\mathbf{x}}{dt} = \mathbf{H}(\mathbf{x}), \tag{1}$$

where $\mathbf{x}$ is the n-dimensional state and $\mathbf{H}$ is the forcing field representing the dynamics of the process. Usually, the state vector represents directly some physical parameters, like temperature or velocity, but sometimes, it may be that we are really interested in derived quantities, or even that in reality, only derived quantities can be observed. Any derived quantity then turns out to be a function of the state space quantities $\mathbf{x}$; in what follows, we will call these derived quantities *observables*. For instance, in case $\mathbf{x}$ is a velocity, the kinetic energy [$K = (1/2)\mathbf{x}\cdot\mathbf{x}$] is a one-dimensional (scalar) observable as it is its value at a single grid point. In this context, the vector-valued *full-state observable* $\mathbf{g}(\mathbf{x}) = \mathbf{x}$ returns the n-dimensional state vector itself. The value of a general observable $\mathbf{g}(\mathbf{x})$ at time t (where $\mathbf{x}_t$ is the state vector evaluated at time t), starting from $\mathbf{x}_0$ at time zero is given by

$$\mathbf{g}(t, \mathbf{x}_0) = \mathbf{g}(\mathbf{x}_t). \tag{2}$$

The Koopman operator is a linear operator and so it admits a spectral decomposition (Mezic and Surana 2016) that contains all the information on the properties of the dynamical system. The spectral decomposition of a linear operator results in a

couple of eigenvalues λ_i and eigenfunctions $\phi_i(\mathbf{x})$ that separate the time dependence t from the state variables $\mathbf{x}$ dependence.

The Koopman operator K^t will describe the evolution of an observable $\mathbf{g}(\mathbf{x})$ of dimension d as

$$K^t[\mathbf{g}(\mathbf{x}_0)] = \mathbf{g}(\mathbf{x}_t). \tag{3}$$

The Koopman operator can also be expressed in terms of its generator operator $\mathcal{L}$ as $K^t = \exp(\mathcal{L}t)$, with eigenvalues $\exp(\lambda_i t)$, where λ_i are the eigenvalues of the generator operator. In the following, we will refer to the eigenvalues of the generator Koopman operator as "Koopman eigenvalues."

In general, if we have a scalar observable $g(\mathbf{x})$, we can express it using the spectral expansion of the Koopman operator, as (Mezić 2005, 2013)

$$K^t[g(\mathbf{x})] = \overline{g(\mathbf{x})} + \sum_k v_k \phi_k(\mathbf{x}) e^{\lambda_k t} + n(t). \tag{4}$$

We obtain a decomposition for the observable in three parts, the time mean $\overline{g(x)}$, an almost periodic component based on the Koopman eigenfunctions/eigenvalues $[\phi_k(\mathbf{x}), \lambda_k = \sigma_k + i\omega_k]$ and a completely aperiodic component corresponding to the continuous component $n(t)$ of the spectrum. In the case of vector observable $\mathbf{g}(\mathbf{x})$, the decomposition applies to each component of the vector, resulting in coefficients that are also a vector $\mathbf{v}_k$ of the same dimension as the observable:

$$K^t[\mathbf{g}(\mathbf{x})] = \overline{\mathbf{g}(\mathbf{x})} + \sum_k \mathbf{v}_k \phi_k(\mathbf{x}) e^{\lambda_k t} + \mathbf{n}(t). \tag{5}$$

The coefficients $\mathbf{v}_k$ are the Koopman modes, and they represent the projection of each component of the vector observable on the Koopman eigenfunctions.

For measure-preserving systems, the decomposition can be written as

$$K^t[\mathbf{g}(\mathbf{x})] = \overline{\mathbf{g}(\mathbf{x})} + \sum_k \mathbf{v}_k \phi_k(\mathbf{x}) e^{i\omega_k t} + \int_0^1 e^{2\pi i \omega t} dE(\omega) \mathbf{g}(\mathbf{x}), \tag{6}$$

where E is a complex, continuous, operator-valued spectral measure on L^2, which in the last part of the expansion represents the contribution from the continuous part of the spectrum, and all the eigenvalues of the Koopman operator are on the unit circle. The eigenfunctions of the Koopman operator are then orthogonal (Mezić 2013).

The continuous part can then be expressed using a density of states and a projection operator P_ω, $\mu_g(\omega)P_\omega$ such that for the case of the state vector

$$\mathbf{x}(t) = \sum_k \mathbf{v}_k \phi_k e^{2\pi i \omega_k t} + \int_0^1 e^{i\omega t} \mu_x(\omega)(P_\omega \mathbf{x}) d\omega, \tag{7}$$

where the density is connected to the projection operator on the eigenspace corresponding to the interval $[\omega, \omega + d\omega]$.

Quadratic quantities like the variance can also be expressed in terms of the Koopman modes. The full calculation will involve also the continuous spectrum component, but in what follows, we will illustrate the concept using only the discrete eigenvalues that in any case will the result of a numerical calculation,

$$\int \mathbf{x}(t)\mathbf{x}^*(t) d\mathbf{x} = \int \left[\sum_j \mathbf{v}_j \phi_j e^{(\sigma_j + i\omega_j)t}\right]\left[\sum_k \mathbf{v}_k \phi_k e^{(\sigma_k + i\omega_k)t}\right]^* d\mathbf{x} = \sum_{j,k} \mathbf{v}_j \mathbf{v}_k^* e^{(\sigma_j + \sigma_k)t} e^{i(\omega_j - \omega_k)t} \int \phi_j \phi_k^* d\mathbf{x}. \tag{8}$$

In the stationary (or purely oscillatory) case,

$$\int \phi_j \phi_k^* d\mathbf{x} = \delta_{jk}, \tag{9}$$

i.e., the eigenfunctions are orthogonal, and the relation reduces to

$$\int \mathbf{x}(t)\mathbf{x}^*(t) d\mathbf{x} = \sum_k |\mathbf{v}_k|^2. \tag{10}$$

Depending on the physical quantity, it may represent the energy or simply the variance. Under these assumptions, the norm of the Koopman modes gives an indication of their contribution to the variance.

At equilibrium, in general, the Koopman decomposition describes the evolution of the system as quasi-periodic motions together with an aperiodic, chaotic component.

3. Koopman decomposition from data

The Koopman operator decomposition can be obtained from the time series of data, and several techniques are available. We will use the kernelized version of EDMD (Klus et al. 2020) that has been used in a climate application previously (Navarra et al. 2021). The EDMD uses a set of nonlinear functions (features) $\mathbf{F} = [f_1, f_2, \ldots]$, to lift the data through a nonlinear transformation to another data space where a more accurate representation of the Koopman operator can be obtained. The features are then the basis to build various approximations. Features can be constructed from various nonlinear functions, for instance, polynomials, but the explicit computation of features can rapidly become prohibitively expensive. The state space can, for instance, be augmented by adding, e.g., $[f_1^2, f_2^2, \ldots, f_1 f_2, \ldots]$. For a vector with n components, using monomials of order up to k will require computing $\binom{n+k}{k}$ columns to the data matrix, a task practically impossible for any large dataset. Kernels make the process more efficient by computing implicitly the features and providing a Hilbert space, the *reproducing kernel Hilbert space* (RKHS), where the data can be expressed in terms of the feature functions. The value of the features in any vector of the data space $\mathbf{x}$ is obtained from the kernel function as $f_i(\mathbf{x}) = k(\mathbf{x}_i, \mathbf{x})$, and the *representer theorem* then guarantees that any function in the RKHS induced by the kernel can be written as $g(\mathbf{x}) = \sum_i \alpha_i f_i(\mathbf{x})$, where α are the projection of the functions on the features.

The kernels (Scholköpf and Smola 2002) lead to an improved approximation by using a richer set of functions to estimate the Koopman operator, allowing similarity measures more general than the standard scalar product. The mathematical results described in Klus et al. (2016, 2020) and the corresponding algorithms will yield the value of the eigenfunctions on the dataset that can be used for the projections that will result in the Koopman modes for the observable $g(\mathbf{x})$. The procedure will also give us the feature coefficients of the Koopman and Perron eigenfunctions, and therefore, we will be able to compute their value for every vector in the data space. The detailed procedure is described in appendixes A and B.

The estimation of the continuous spectrum with numerical methods is tricky, as the algorithms will always produce discrete eigenvalues. However, we can still estimate the density of states in the spectrum and identify the region where the density is higher. Periodic systems will have distinct eigenvalues, and they will have sharp spectral lines; in a mixing system, there also may exist modes that are close to periodic modes (Cvitanović et al. 2020) that will show up as broadening of the spectral lines.

We will use in the following the Gaussian kernel $\{[1/(\sqrt{2\pi}s)^n]\exp(-\|\mathbf{x}-\mathbf{y}\|^2/2s^2)\}$. The features will then be given by

$$f_i(\mathbf{x}) = \frac{1}{(\sqrt{2\pi}s)^{n/2}}\exp(-\|\mathbf{x}-\mathbf{x}_i\|^2/2s^2). \tag{11}$$

Here, the crucial parameter is the bandwidth s, namely, the influence radius of the data points with respect to each other. Ideally, we would like to have a larger bandwidth in regions where the state density is low and a smaller bandwidth in regions where the state density is high. Such "self-tuning" kernels have been proposed for probability density estimation (Berry and Harlim 2016), but they have not been applied in the present context. The choice of the Gaussian kernel, either self-tuning or not, will generate an infinite-dimensional feature space, enlarging the class of features to an infinite set.

In a previous paper (Navarra et al. 2021), we have chosen a bandwidth that normalized the distances to unity standard deviation and we will follow that choice here. The detailed value of the eigenvalues is sensitive to the bandwidth choice, but the properties of the modes are more robust. There is very little sensitivity to the other free parameter: the Tikhonov regularization parameter.

4. The data and methodology

We will use in this paper the 1870–2020 data from the HadSST dataset (Rayner et al. 2003). They consist of monthly mean SSTs on a 1° × 1° regular grid. The data have been linearly detrended to further isolate the interannual variability of interest, and a running mean of 3 months has been applied to soften the impact of high-frequency variations. The data have then been transformed to anomalies by subtracting a monthly climatology, which eliminates a large portion of the seasonal cycle that would otherwise dominate the Koopman decomposition together with its harmonics.

The data are then unrolled and organized in data matrices,

$$\mathbf{X} = [\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_N], \tag{12}$$

so that every column corresponds to a single sample point unrolled in a long vector.

Prior to the Koopman analysis, the data have been preprocessed through a preliminary EOF analysis to reduce the computational load. The EOF analysis is used as a reduction procedure to reduce the space dimensionality of the data from around 18 000 to a few hundreds. A detailed analysis shows that retaining a number of EOFs equal or greater than the numerically determined rank of the data matrix (around 898 for the tropical case) will yield results identical to those obtained using the gridpoint data without any preliminary EOF reduction. Retaining less than the rank of EOF will start to introduce the differences in the result due to the elimination of some portion of the variance (see more in the *sensitivity analysis* subsection). In what follows, we will use the number of EOFs corresponding to the rank to get faster calculations without the loss of information. It is probably interesting to note that the computational effort is limited, even in the full gridpoint case, and calculations do not require HPC resources but can be done on a workstation.

The kernel EDMD algorithm will provide the coefficients of the eigenfunctions over the features, so that the eigenfunctions will be given by

$$\phi_k(\mathbf{x}) = \sum_i w_{ki} f_i(\mathbf{x}) \text{ (Koopman)}, \tag{13}$$

$$\psi_k(\mathbf{x}) = \sum_i u_{ki} f_i(\mathbf{x}) \text{ (Perron)}. \tag{14}$$

The estimated evolution of the dynamical system described by the dataset is then given by

$$\mathbf{x}(t) = \sum_k \mathbf{v}_k e^{(\sigma_k + i\omega_k)t}\phi_k(\mathbf{x}), \tag{15}$$

$$= \sum_{ik} \mathbf{v}_k e^{(\sigma_k + i\omega_k)t} w_{ki} f_i(\mathbf{x}), \tag{16}$$

where we can see that the state vector can be expressed in terms of the Koopman or Perron modes. If the eigenfunctions were orthogonal, the Koopman mode amplitudes would be the projection of the field grid point by grid point on the eigenfunctions. In general, they are not orthogonal, so we must get them solving a linear system at $t = 0$,

$$\int \mathbf{x}\phi_i^*(\mathbf{x})d\mathbf{x} = \sum_k \mathbf{v}_k \int \phi_k^*(\mathbf{x})\phi_i(\mathbf{x})d\mathbf{x}, \text{ for } i = 1, 2, 3, \ldots. \tag{17}$$

The kernel generalized the relation between data points. We can understand this by considering a linear kernel, which is a polynomial kernel of order 1 with $c = 0$. In this case, the Koopman decomposition is equivalent to a LIM (Tu et al. 2014; Tu 2013), or in more precise terms, it is the LIM that is providing just a linear approximation of the Koopman operator for the dynamical system implied by the data. The kernel

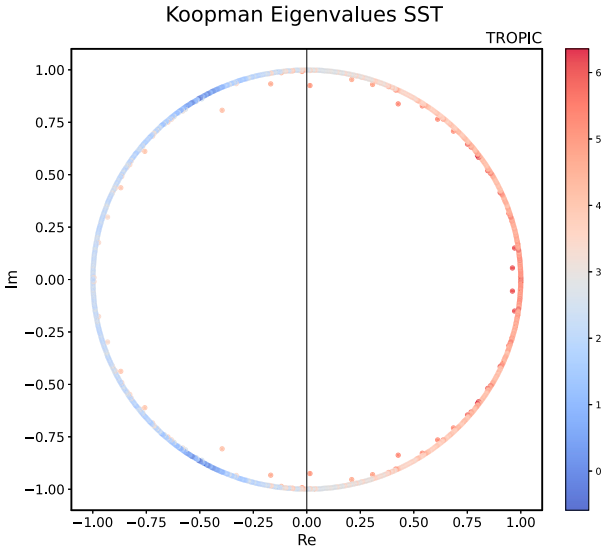

FIG. 1. Koopman spectrum for the tropical SST. A regularization constant of 10^{-5} has been used in the kernel-based algorithm. Only the eigenvalues whose oscillatory time scales are longer than 3 months are shown. The shading is proportional to the norm of the Koopman modes.

uses a larger set of functions yielding a better approximation. In the linear kernel case, i.e., when the kernel is simply the scalar product, the Koopman modes then become the LIM modes (see appendix C).

5. Koopman modes of the tropical ocean

We will consider the tropical Pacific Ocean, between the latitudes of 35°N–35°S. Figure 1 shows the eigenvalues of the Koopman operator. We have used a regularization constant $\epsilon = 10^{-5}$, with little sensitivity with respect to the case with no regularization. The eigenvalues are well distributed around the circle, they are contained within the unit circle, the shading is proportional to the norm of the Koopman modes, and it is possible to see that with few exceptions, the eigenvalues with modulus smaller than 1 have also a small Koopman mode amplitude.

a. *Spectral density*

We can investigate the density of the eigenvalues to discern the character of the estimated spectrum. The state density is shown in Fig. 2, where the eigenvalue density has been estimated using a Gaussian density estimator. Remarkably, several peaks appear elevated over the sea of eigenvalues.

This suggests the existence of quasi-periodic cycles in the neighborhood of the maxima, namely, cycles that are hidden but contribute to the large-scale structure of the phase space (Cvitanović et al. 2020). We can easily identify the peaks around 9.18 and 3.40 years, less evident areas around 2 years and a more definite structure at 1.2 years. The band indicates the interval of one standard deviation of the spectra obtained by scrambling the original data time series and repeating the Koopman decomposition. A more detailed analysis of the

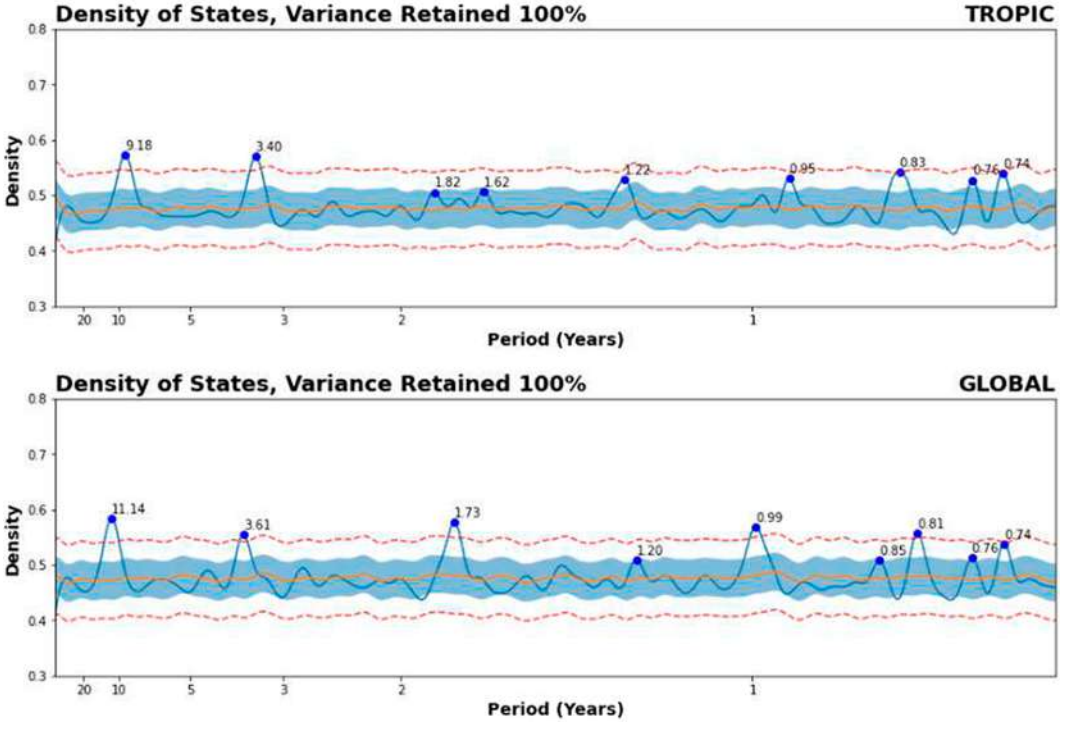

FIG. 2. State density distribution of the eigenvalues. The state density distribution of the eigenvalues for (top) the tropical domain and (bottom) the global domain. It is possible to see several peaks that correspond to quasi-periodic motions. The shaded band shows one standard deviation extent over computing the spectra randomly scrambling the time series 100 times, while the red dotted lines delimit two standard deviations. The period in years is indicated next to the peak.

TABLE 1. The first 10 Koopman modes ordered for norm size.

Mode	Norm	*e*-folding time (yr)	Period (yr)
0	576.24	−11.67	0.82
1	434.61	−2.18	9.13
2	431.40	−10.26	0.83
3	419.99	−4.74	3.41
4	369.94	−14.59	1.22
5	355.72	−192.29	3.58
6	343.90	−1043.24	140.65
7	328.69	285.31	75.83
8	326.15	−8.85	0.74
9	316.36	−10.55	0.95

3.40 peak reveals that in an interval of 6 months around that period, there are 14 modes, whereas in an interval of 12 months around the 9.18 peaks, there are four modes. We are showing in Table 1 the first 10 modes in terms of the norm size. The Koopman decomposition yields here not only quasi-periodic modes with long decay or growth *e*-folding time but also strongly damped modes, like mode 1 that has a damping time scale of about 2 years and an oscillatory period of about 9 years. The evaluation is sensitive to the length of the time series, as the precise frequency and amplitudes of the peaks may vary somewhat, especially the longer periods that are more sensitive to the length of the time series, but it is to be expected in an evaluation that is basically evaluating time behavior. The longer time series will result in peaks that are more identifiable and significant over the uncertainty band visible in Fig. 2.

The time variation in the first eight Koopman eigenfunctions is shown in Fig. 3. It is possible to see the damping scale, indicating that they represent some adjustment process. However, in their neighborhood, there are also modes that have much slower decaying rates almost exactly oscillatory, like mode 5 that with a period of 3.58 years and a decay scale of 192 years is slowly decaying.

The Koopman modes give the value of the Koopman eigenfunction at the grid point, so they can be represented on a geographical map and their evolution through one cycle is shown in the following figures. The time evolution of the *j*th Koopman mode ξ_j is computed by dividing the phase in N steps, ignoring the exponentially decaying term of the eigenvalue as

$$\mathrm{Re}(\mathbf{v}_j e^{i\omega_j t}), \tag{18}$$

for $t_n = 2\pi n\Delta t/T_k$, where $\Delta t = T_k/N$ is the phase increment and T_k is the period corresponding to the kth mode. There are several modes that contribute to the bands, but choosing the modes with the largest norm around the 3.4-yr peak, we can show the first one in Fig. 4. The mode is an equatorially confined mode, with some structure in the midlatitude and mostly limited to the Pacific. The next mode (Fig. 5), which is almost stationary, shows a very similar structure, but with a noticeable phase shift.

The spatial distribution of the first modes at the 9.18 peak can be seen in Fig. 6, and as in the previous pictures, we show the evolution over a cycle without regard for the exponential decay. Also in this case, it is difficult to identify a possible propagation in the strongly damped modes, whereas it is more traceable in stationary mode 5 (Fig. 7). Both modes show a wider distribution with respect to the higher frequency modes suggesting an involvement of off-tropical dynamics.

The picture that results is that even with a continuous spectrum, there are preferred time scales that emerge as the accumulation points of the eigenvalues. We can speculate that, in general, we can have on one extreme the case of a perfectly periodic system that will have, sharp, delta-function eigenvalue distributions and to the other extreme of a totally random system with a flat distribution of the eigenvalues. In between, we will have cases with preferred time scales of variability, showing as the accumulation of eigenvalues at certain frequencies. We can also speculate that it is the interaction among many different Koopman eigenfunction that broadens the peaks, leading to the interesting possibility that in this way, we can identify fundamental periodic motions that are however continuously perturbed by the higher modes.

b. *Stationary Koopman modes*

The spontaneous appearance of quasi-neutral modes suggests that they may play a significant role. They may define

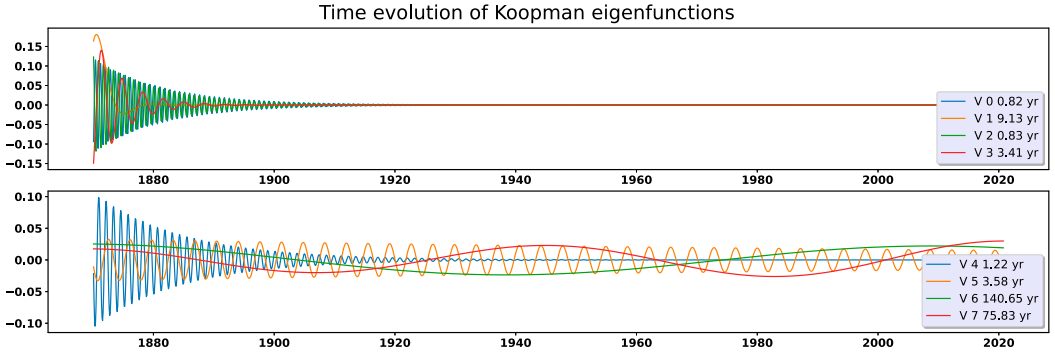

FIG. 3. Time evolution of the values taken by the first eight Koopman eigenfunctions corresponding to the largest norms of the Koopman modes. The eigenfunctions are evaluated on the first time step, and only the real part is shown. Some of the modes are damped with a significant *e*-folding time scale.

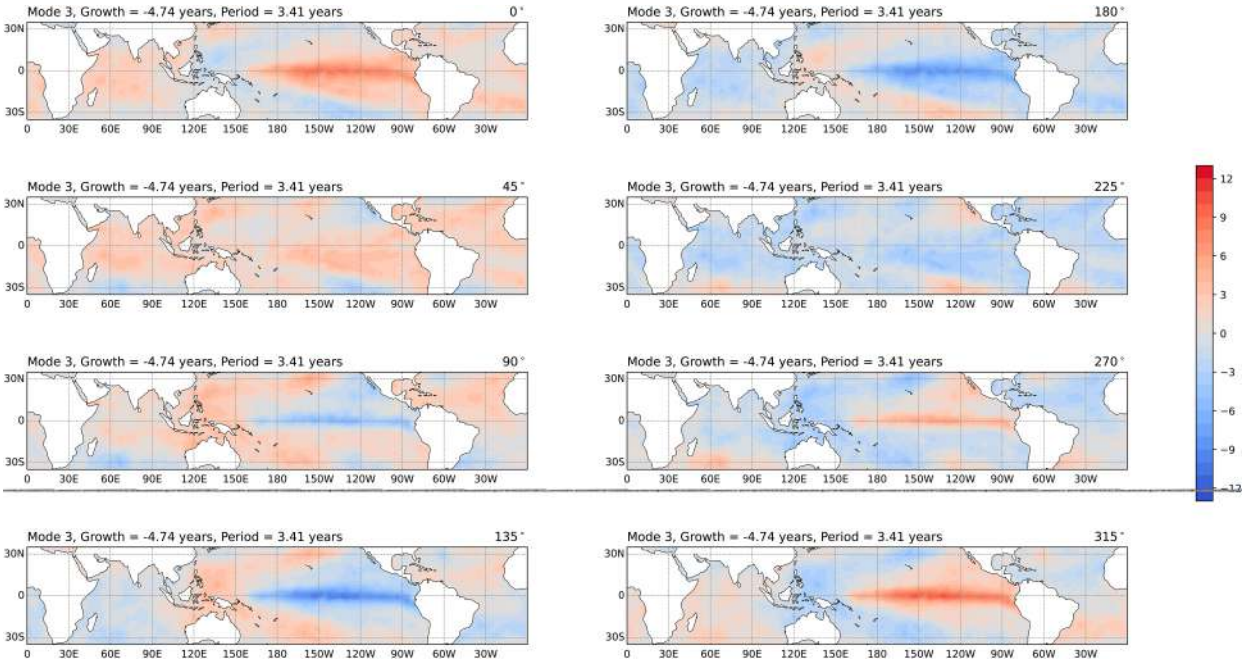

FIG. 4. Time evolution of the oscillatory component of Koopman mode 3 over a single period at different phases. The decaying component is ignored. Contour units are in degrees.

a stationary subspace, which is a portion of the variability that confines the evolution of the system for a long time. The Koopman analysis does not force the stationary character on the modes in the same way that a Fourier analysis does, for instance, and therefore, they identify the "stationary modes" around which short-lived fluctuations appear.

In a numerical estimation, we have to define a level under which we will consider the modes stationary. Table 2 shows the first 10 stationary modes ordered by magnitude. In this case, we

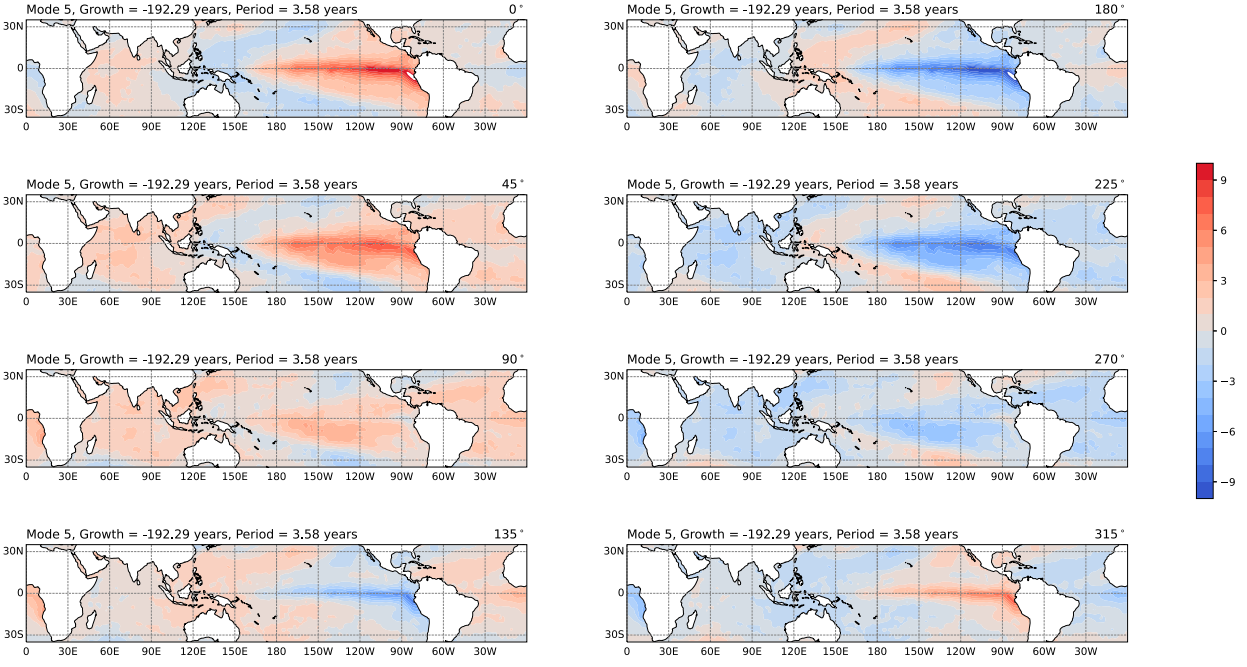

FIG. 5. As in Fig. 4, but for mode 5 with a period of 3.58 years.

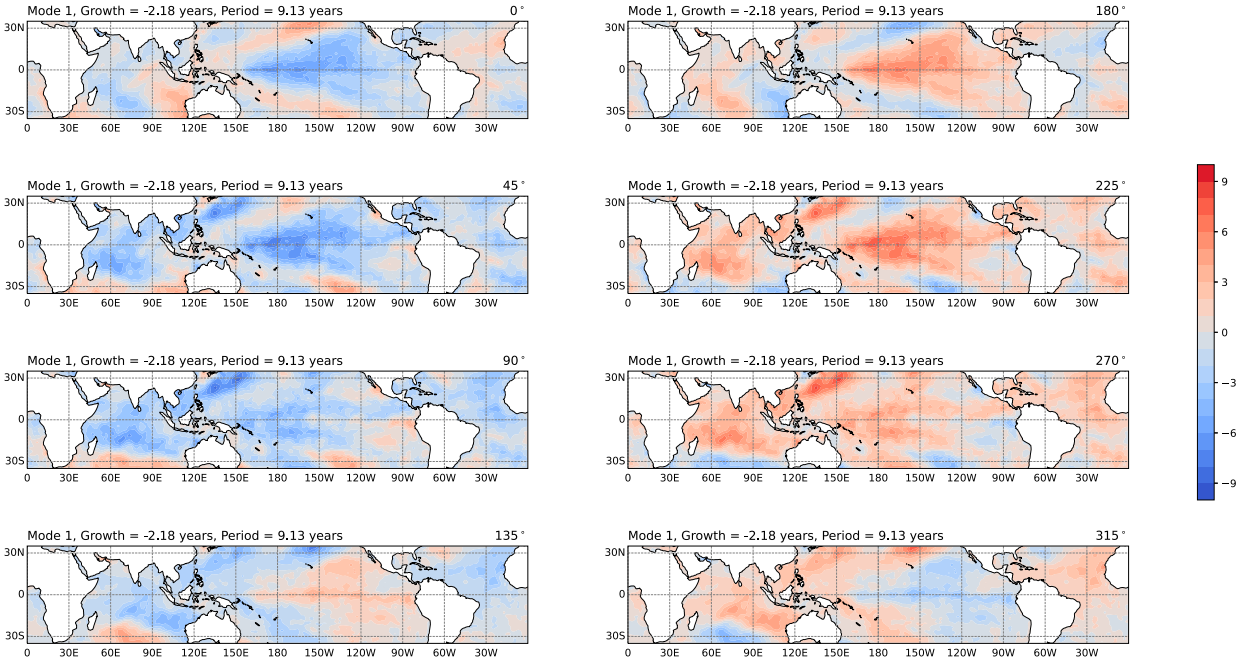

FIG. 6. Time evolution of Koopman mode 1 close to the 9.18-yr peak. Contour units are in degrees.

have chosen the threshold of an *e*-folding time greater than 300 yr^{-1} to retain modes that we call "stationary." The values of the eigenfunctions associated with the first eight stationary modes with the largest norms are shown in Fig. 8.

The interplay between different kinds of modes is shown in Fig. 9; here, the modes are ordered by period and normalized such that the largest mode has unit norm. Using the chosen threshold, we found that 76 modes and their complex conjugates satisfy the stationary condition. The color scale indicates the sign of the real part of the eigenvalues. Eigenvalues with positive real part are in blue, those with negative real part are in light pink, and those whose real part is smaller than the

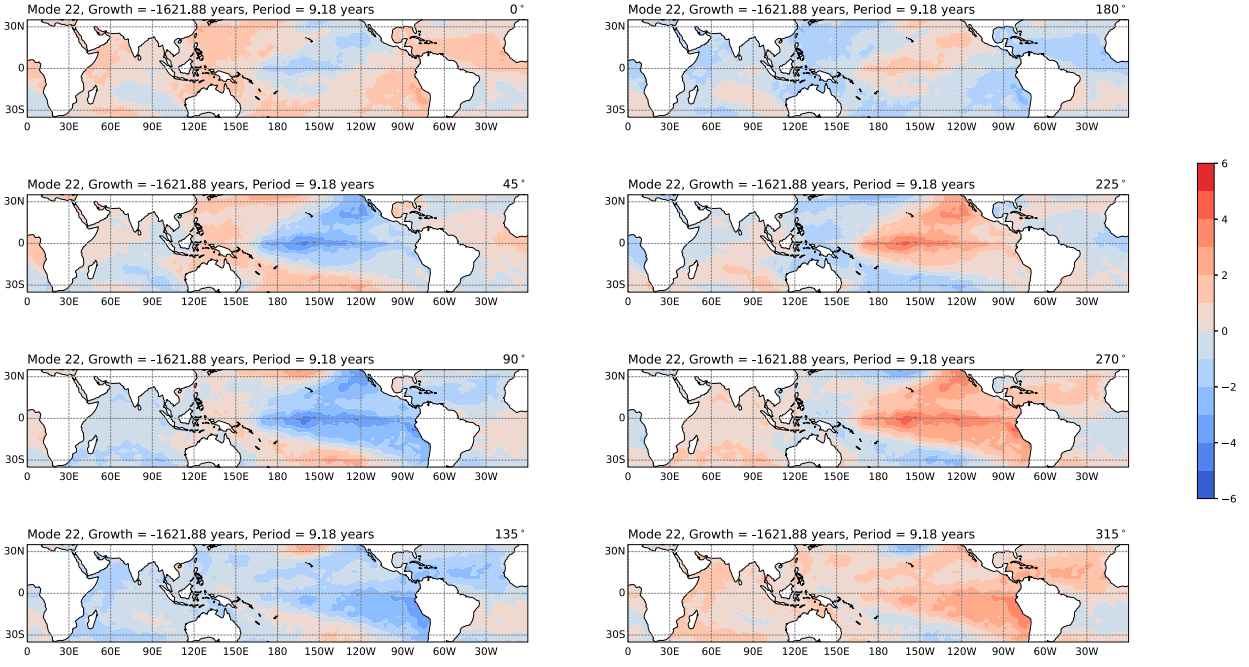

FIG. 7. As in Fig. 6, but for mode 22 corresponding to the peak of period 9.18 years.

TABLE 2. The first 10 stationary Koopman modes ordered by magnitude. The position in the general list of modes is given in parentheses.

Mode	Norm	*e*-folding time (yr)	Period (yr)
0 (6)	343.90	−1043.24	140.65
1 (12)	287.85	−838.18	5.65
2 (20)	247.85	−786.32	48.55
3 (22)	237.98	−1621.88	9.18
4 (24)	222.80	−998.35	13.29
5 (26)	215.15	−373.90	12.15
6 (30)	207.41	−604.13	18.48
7 (32)	202.47	−652.27	6.40
8 (38)	197.97	−724.62	4.20
9 (44)	187.20	−2678.26	2.86

threshold corresponding to a *e*-folding time of 300 years are in coral. The damped modes are the majority, but we can see that modes with small *e*-folding time are also common. There is a rare instance of a growing mode, which is attributed to numerical instabilities, as physically stable systems should not exhibit exponential growth.

We can use the Koopman decomposition to project the data onto some subsets. We can have an idea of what this means by looking at Fig. 10 that shows the Niño-3 index computed on the SST time series reconstructed using in Eq. (16) only Koopman modes whose periods are within a range of 6 months around the 3.4-yr peak (red line) and 12 months (blue line). All modes in the range are used, including decaying ones. However, if in the reconstruction, we use all the stationary modes we obtain the green line. The groups of modes around the 3.4 peaks follow the variation in the Niño-3 index well, but it is only when all the stationary modes are used that we improve also on the amplitude of some of the larger variations. This figure is consistent with Froyland et al. (2021) showing that Koopman decomposition can identify dynamically relevant modes accurately even if we are using here a different method to estimate the Koopman modes and a different, longer dataset.

The stationary periods have emerged spontaneously from the analysis, and they show that the neutral modes can explain a significant part of the Niño-3 index variability. Some mechanism amplifies the stationary variation locally, causing amplification and modulations of the variability and the stationary modes express a rather nontrivial representation of the variability. The stationary subspace therefore reflects a fundamental property of the system, at least as it can be identified from the data and within the limit of the length of time series, but it is tempting to identify it with the internal variability of the SST.

The almost purely oscillatory modes with very weak growth or decay suggest that they will contribute to the stationary portion of the variability of the SST. From a mathematical point of view, these modes will render the Koopman operator unitary. The deviation caused by the blue and pink modes in Fig. 9 represents the deviation from the unitarity of the Koopman operator. If it was not for the presence of some residual and spurious growing modes, which arise from numerical instability, the system would eventually evolve to a subspace spanned only by the stationary modes. It is interesting to note that the modes with the highest norm are strongly damped modes that really exist only for a short time during the evolution of the time series. In this scenario, the damped modes express the distance from equilibrium of the observations in the beginning of the time series. Figure 9 shows that the stationary modes are distributed along the spectrum and so we can find one in the neighborhood of all the peaks we have identified, we can then concentrate only on the stationary part of the spectrum (Fig. 11). Often, these modes are also close to the peaks of the eigenvalue density.

c. *Overlap and interaction among the Koopman modes*

It is of interest to look at the interaction between the various Koopman eigenfunctions. Because of the Hilbert space structure, the interaction is given by the overlapping or interaction integral, which is the scalar product of the eigenfunctions computed numerically over the values of the eigenfunctions at the data points (Fig. 12). The picture shows the overlapping integral of the eigenfunctions as a matrix of interaction integrals. The ordering is such that the first 72 modes correspond to the stationary modes. The diagonal is the norm of the eigenfunctions normalized to 1, and the colors are decreasing values with a logarithmic scale up to values 0.01 as smaller values are omitted.

With this ordering, the bottom left quadrant shows the interactions among the stationary modes. The top right quadrant shows instead the interactions among the transient (damped)

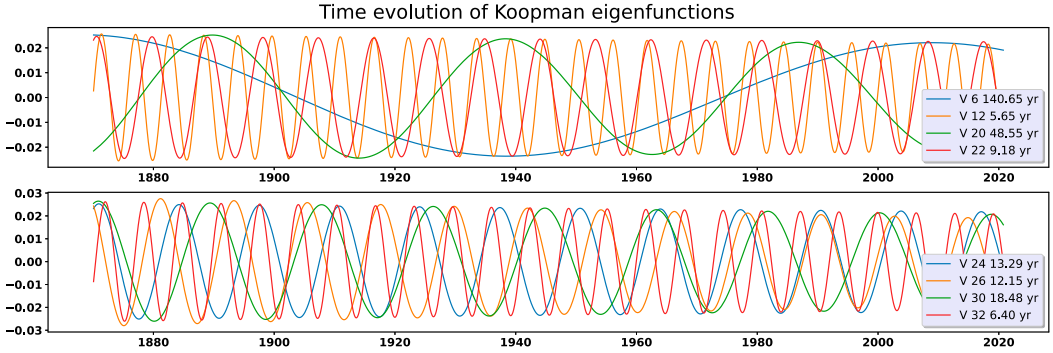

FIG. 8. Time evolution of the values taken by the first eight Koopman stationary eigenfunctions with the largest norms of the Koopman modes on the data values. Only the real part is shown.

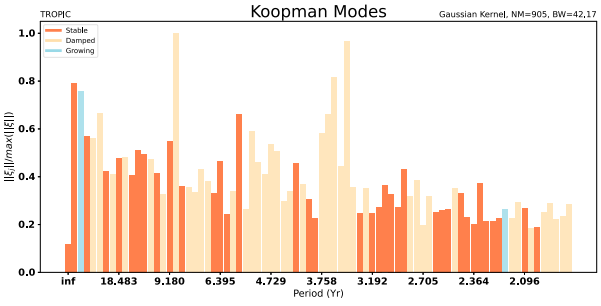

FIG. 9. Koopman modes for the tropical SST normalized so that the norm of the largest mode is 1. In the case of complex conjugate pairs, only one mode is shown. We represent here decreasing time scales up to 1 year. The light pink modes are the damped mode, the blue ones are growing modes, and the coral ones are stationary modes.

modes, and the two remaining quadrants show the interaction between the stationary and damped modes. The interactions among the stationary modes are concentrated on the diagonal. The off-diagonal values are small, with some exceptions. It is also possible to observe the existence of interaction "segments" that show up as blocks on the diagram. There are rather isolated modes around 1.82 periods and a block between periods 1.82 and 1.23 that seems to involve modes that only interact among themselves and very little with the others. Also the block between 3.27 and 1.82 defines another such block, even if more complex. It is interesting to note a series of short period modes in the very top right corner that seems to be almost completely orthogonal.

This must be contrasted with the top right quadrant of the transient modes that instead shows less structured interactions and more long-range interactions among the modes with different periods. The remarkable case of the modes that seem to interact with every other mode corresponds to strongly damped transient modes that result in having interactions with most of the oscillatory modes because of their limited extent. Furthermore, the off-diagonal values are larger and decrease more slowly moving away from the diagonal.

The symmetrical top left and bottom right quadrant show in general smaller values. The maximum interaction value is of the order of 0.1. The plots confirm somewhat a local interaction between the stationary modes and these modes. We can see this from the limited diagonal extent of the nonzero values of the interactions, indicating a tendency for modes within certain period bands to interact with stationary modes in the same band.

d. *Sensitivity analysis*

As previously mentioned, the state distribution is sensitive to the length of the analyzed time series, and the frequencies and amplitudes of the peaks may vary to some extent, especially for longer periods that are more sensitive to the length of the time series. Figure 2 shows several peaks that rise above the shaded bands that represent one standard deviation over the spectrum obtained by randomly scrambling the time series 100 times. This analysis gives an idea of the robustness of those peaks, which can be clearly distinguished from the sea of eigenvalues obtained from the random time series.

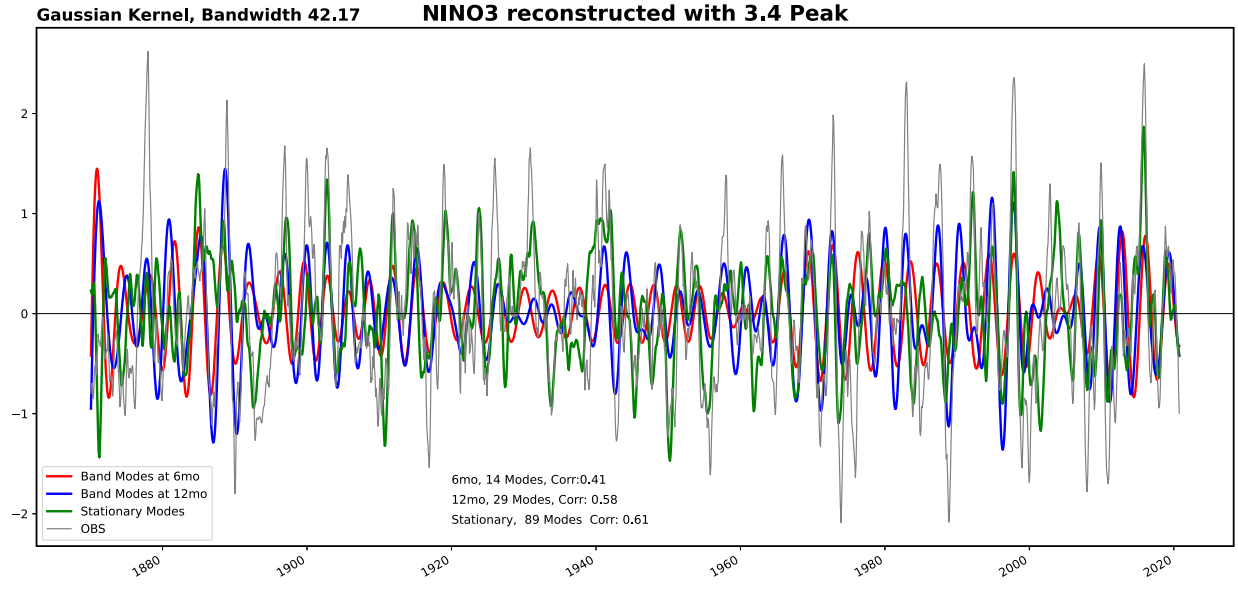

FIG. 10. Niño-3 index reconstructed from the Koopman stationary modes (in color) and the original data (black line). The red line corresponds to a Niño-3 index computed over the SST data series reconstructed using only the modes around a 6-month interval around the 3.4-yr peak (14 modes plus complex conjugates). The blue line corresponds to expanding the interval to 12 months (29 modes) and the green line is obtained by reconstructing the data using all the stationary modes (78 modes). The stationary modes improve marginally the correlation (0.61), but they increase the amplitude of the larger events. The average correlation achieved over 50 extractions of random combinations of the same number of modes (among all of them, not just stationary) yields a correlation of 0.30 ± 0.05.

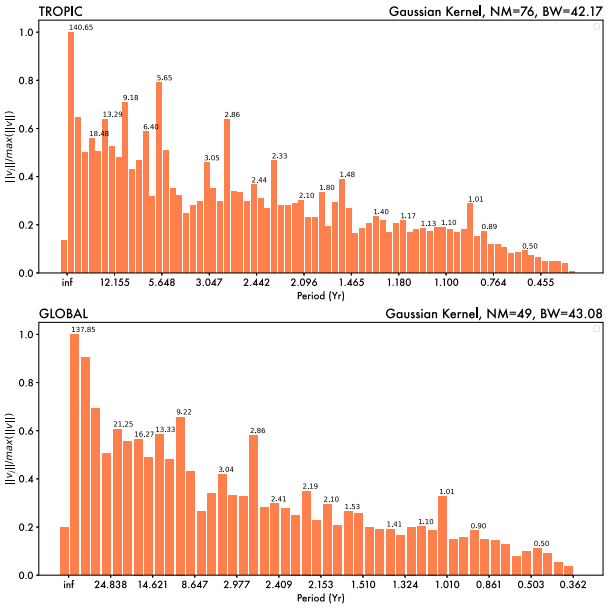

FIG. 11. The stationary Koopman modes norm for (top) the tropical domain and (bottom) the global domain. The amplitude is normalized to the maximum value.

To further investigate this issue, Fig. 13 shows the state density distribution for the whole period (upper panel on the left, equivalent to Fig. 2) plus 15 80-yr-long periods of the temporal domain (with a sliding window of 5 years). The colored circles identify the appearance of the peaks identified in Fig. 2 in each subperiod, as well as two other peaks that recurrently appear in the 80-yr-long subsamples (viz., around 6.16 and 1.69). Peaks have been considered to be the same if they lie within a 15% distance of the value of the total period peak. The peaks below 0.95 have not been considered, as they lie too close to each other to properly identify them. It is possible to see that several peaks correspond among most of the domains, as well as with the whole period distribution shown in Fig. 2. In particular, the peak around 0.95 (marked in red) can be found in 13 out of the 15 subsamples, while the one around 1.22 appears in 14 out of the 15. Regarding the peak around 3.40, it can be found in 7 out of the 15 subsamples if we keep the 15% band around the original peak. However, it can be identified in 12 out of the 15 samples if we stretch it to 30%. In fact, it is easy to see that a distinct peak appears around 3.40 in almost all the subsamples, but it seems to showcase a higher variability in its period. It should be noted, too, that some peaks split into two peaks. This happens more markedly at higher frequencies, where peaks are closer among them and, therefore, more difficult to identify. In addition, two new frequencies emerge consistently in the 80-yr-long samples: one around 6.16 and another around 1.69 (marked in blue and red respectively), while the peak around 9.18 cannot be found in any of them, which demonstrates that longer training periods are needed for the proper identification of lower frequency periods. The exact frequencies of these two new peaks (6.16 and 1.69) have been extracted from the peaks arising when using all the 15-period eigenvalues for the density computation (not shown).

A comparison between Figs. 2 and 13 shows that most part of the main peaks can be identified in both the total and subsampled temporal domains. A further analysis with shorter subsets (50 and 37.5 years long, not shown) shows similar results, but the peaks become increasingly more difficult to identify, with further shifts and an increasing number of splitting peaks. In particular, the 9.18 peak (and 11.14 peak for the

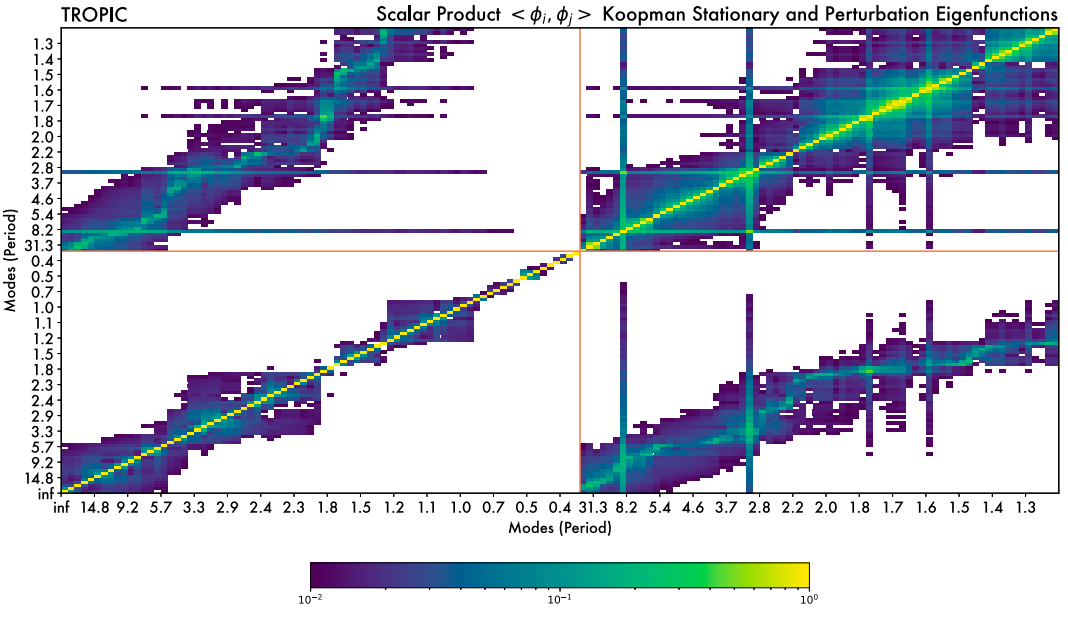

FIG. 12. Overlap integrals for the stationary and transient Koopman eigenfunctions for the tropical case. The bottom left quadrant shows the stationary mode overlaps and the top right quadrant the transient Koopman mode overlaps. The other quadrants show the interactions.

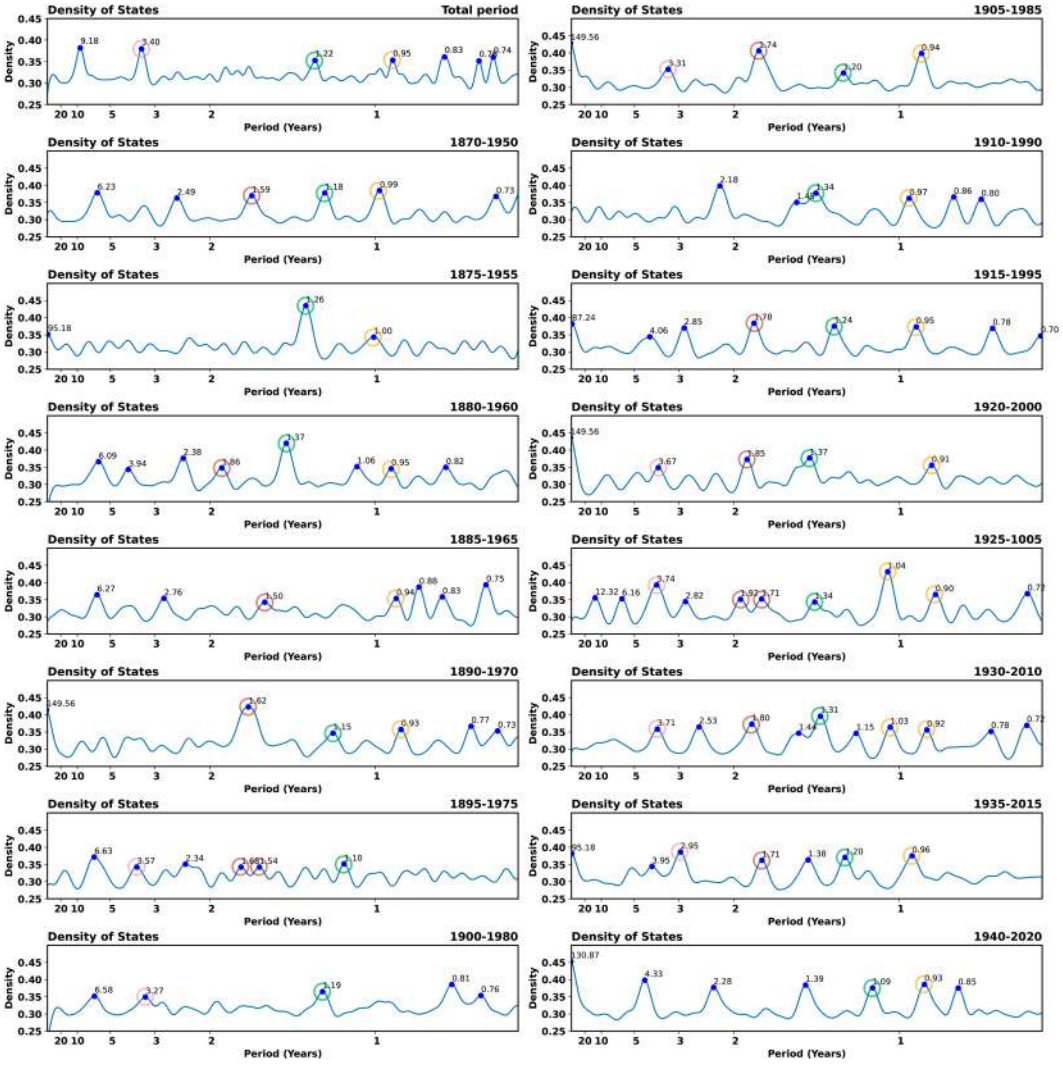

FIG. 13. State density distribution for the whole period (first panel on the left, equivalent to Fig. 2), plus 15 80-yr-long periods of the temporal domain (with a sliding window of 5 years). The colored circles identify the appearance of the peaks identified in Fig. 2 in each subperiod. A peak has been considered to be the same if it lies within a 15% interval from the value of the original peak. The peaks below 0.95 have not been considered, as they lie too close to each other to properly identify them. It is possible to see that several peaks show in many subperiods, as well as within the whole period distribution shown in Fig. 2.

global domain) has proved to be harder to reproduce. While they still emerge for periods slightly shorter than the total time series, they seem to shift to shorter frequencies or to completely disappear for periods significantly shorter than the total time domain. However, this is to be expected, as longer time series result in peaks that are more significant and identifiable.

Regarding the sensitivity to the EOF dimensional reduction, Fig. 14 shows the state density distribution for four different retained variances of the tropical domain. The variances retained, from top to bottom, are equal to 0.99 (number of EOFs = 224), 0.95 (number of EOFs = 56), 0.90 (number of EOFs = 31), and 0.80 (number of EOFs = 15). It can be observed that the peaks and the overall distributions featured in the three first panels show a high correspondence among them, as well as with the full-rank distribution in Fig. 2. On the other hand, the bottom panel (which retains only the first 15 EOFs) is considerably different. Although some of the main peaks still can be identified, it also shows an increasing number of less identifiable peaks. This feature becomes more evident if the retained variance is even lower (0.75 and 0.6, not shown), demonstrating that the method remains stable and robust as long as the retained variance is close enough to one (even if the number of EOFs varies greatly, i.e., from 224 at 0.99 to 31 at 0.9).

A more detailed analysis shows that the number of growing modes increases as the number of EOFs becomes considerably smaller than the full rank (under 0.8), while the

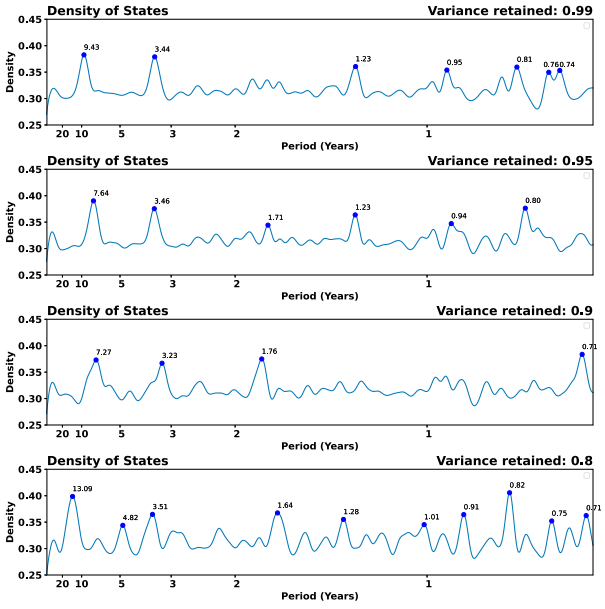

FIG. 14. State density distribution of the eigenvalues for the tropical domain and four different EOF dimensional reductions. The variances retained, from top to bottom, equal to 0.99 (number of EOFs = 224), 0.95 (number of EOFs = 56), 0.90 (number of EOFs = 31), and 0.80 (number of EOFs = 15). It is possible to see that the peaks and the overall distributions featured in the three first panels show a high correspondence among them, as well as with the full-rank distribution shown in Fig. 2. Although some of the main peaks can be identified also in the last panel (which retains only the first 15 EOFs), it shows an increased number of less identifiable peaks.

number of stationary modes decreases (i.e., no stationary modes arise for a retained variance of 0.6, number of EOFs = 6). This suggests that nonlinear dynamics and the lower tails of variability are relevant for the identification of isolated peaks, but further research is needed to fully investigate this issue.

6. Koopman modes of the global SST

The previous analysis has been focused on the entire tropical belt between 35°N and 35°S, and it is now interesting to see the results of the Koopman decomposition for a different domain. In this section, we discuss the Koopman decomposition for an enlarged global domain between 60°N and 60°S. Expanding the domain to higher latitudes will involve longer time-scale processes, so we can expect to see the signature of the different dynamics in the modal structure.

The Koopman decomposition for the global case is shown in the bottom panel of Fig. 2, where we are comparing the density analysis of the eigenvalues for the two cases. The spectrum is rather consistent, and we can see that the same modes, identified by their frequency/period, emerge in both cases. The same mode around 3.6 years is visible and another one at 1.7 years. The low-frequency peak shifts toward longer periods in the global case.

The shift toward longer periods in the global domain including the entire Pacific Ocean can probably be connected to the long-term variability of the Pacific, as the Pacific decadal oscillation (PDO). After the identification of the PDO as a variability mode of the Pacific (Mantua et al. 1997), the early interpretation favored a simple mode of variability, but further research (Mantua and Hare 2002) revealed that the PDO can be really a mixture of several processes and time scales (Newman et al. 2016). This interpretation is supported by the Koopman mode decomposition that shows a number of modes in this spectral region. Figure 15 shows the PDO index computed as the average over the region (20°, 60°N) and (150°, 220°E) (Mantua et al. 1997) for the observations and for the data reconstructed by the stationary Koopman modes. It is interesting to note that in this case, it is difficult to capture

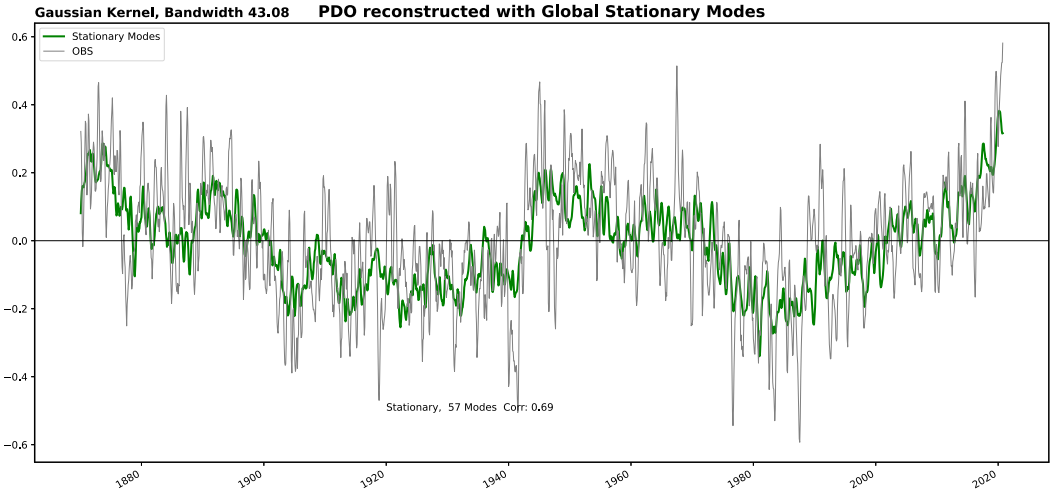

FIG. 15. PDO index reconstructed from the Koopman stationary modes (green) and the original data (black line). The index reconstructed from the stationary modes yield a correlation of 0.69 with the observational index. The average correlation achieved computing an index over 50 extractions of random combinations of the same number of Koopman modes yields a correlation of 0.09 ± 0.07.

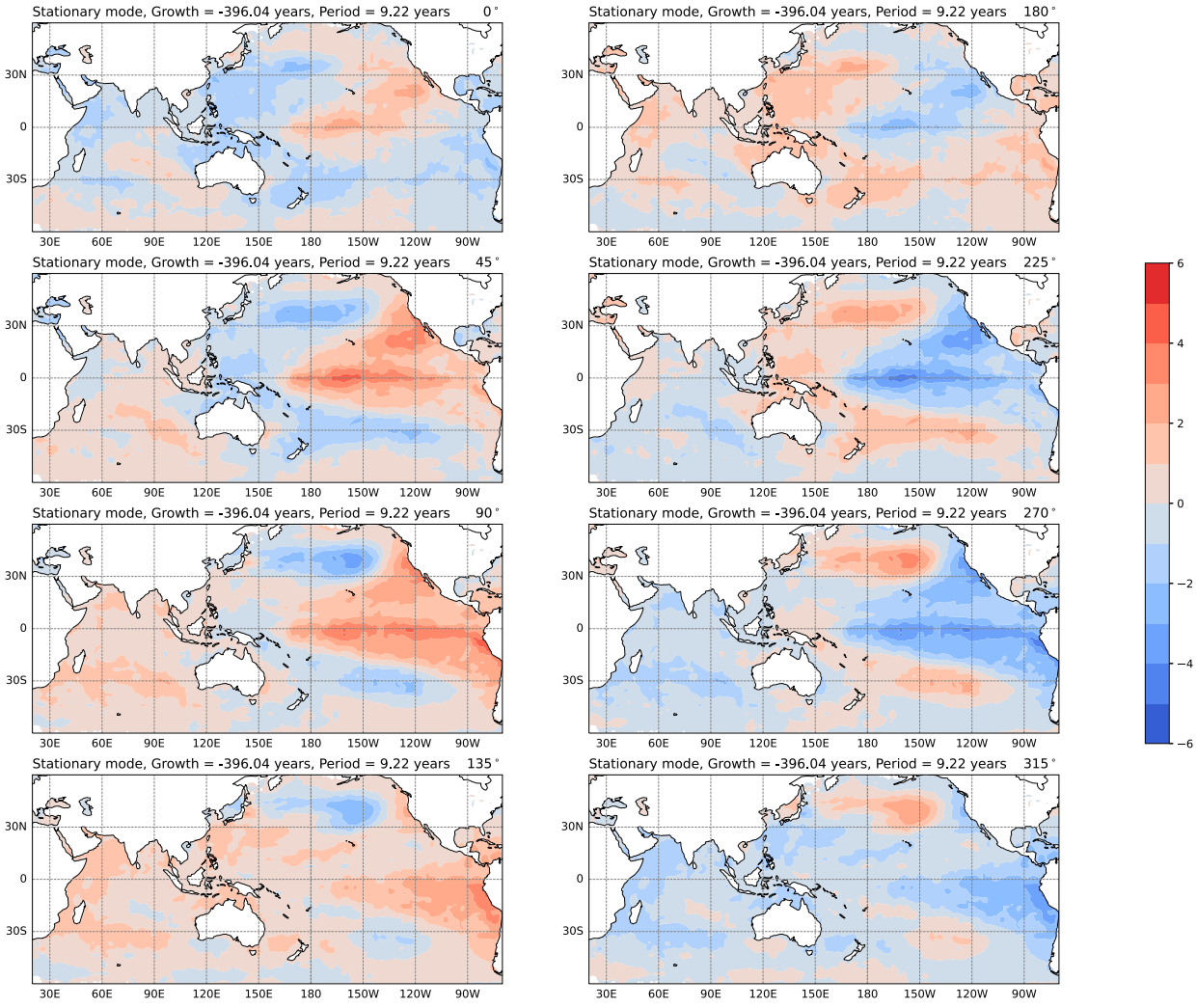

FIG. 16. The global stationary Koopman mode with period 9.22 years.

the behavior of the index with few modes and some correspondence can be obtained only using all the available stationary modes. This is to be compared with the behavior of the Niño-3 index that is showing a tighter spectral concentration of modes with respect to the PDO.

Considering now just the stationary components, selected using also in this case with a threshold value of 300 years^{-1} for the stationarity (i.e., a growth rate smaller than the threshold value, we can compare the Koopman modes in Fig. 11, where we are comparing the stationary Koopman modes for the tropical and global case. The modes have been normalized with respect to the maximum norm. The comparison shows how the modes at 2.86 and 3.04 years appear to be mostly tropical, whereas the mode at 5.65 appearing in the tropical analysis is probably due to the residual influence from the midlatitudes. In the larger domain, this mode is absorbed in the general shift toward longer time scales.

The structure of the global Koopman modes is shown in Fig. 16 for the mode with a period of about 9 years. The pattern is clearly indicating the PDO, and it is basinwide. The second largest mode (Fig. 17) is instead for the period at 2.86. This mode is more confined to the equatorial region. The comparison between these two modes shows how in fact they are not completely separable and how the slow mode, in its evolution around the basin, interacts with the faster equatorial modes. Also the broader shape of the slow mode may indicate the role of slower off-tropical disturbances in the propagation from the North American West Coast to Asia.

Each monthly mean can be expanded over the eigenfunctions, so we can get a different view by looking at the contribution from each eigenfunction for each month, including in this case all modes both exponential and oscillatory. There is considerable variability from 1 month to the next, but it is possible to get some insight by constructing appropriate composites. Figure 19 shows the Koopman spectra for two cases of composites. On the top, the data have been stratified using the NINO34 index, compositing the cases with a value of the index larger than 1.5°C or less than −1.5°C. The

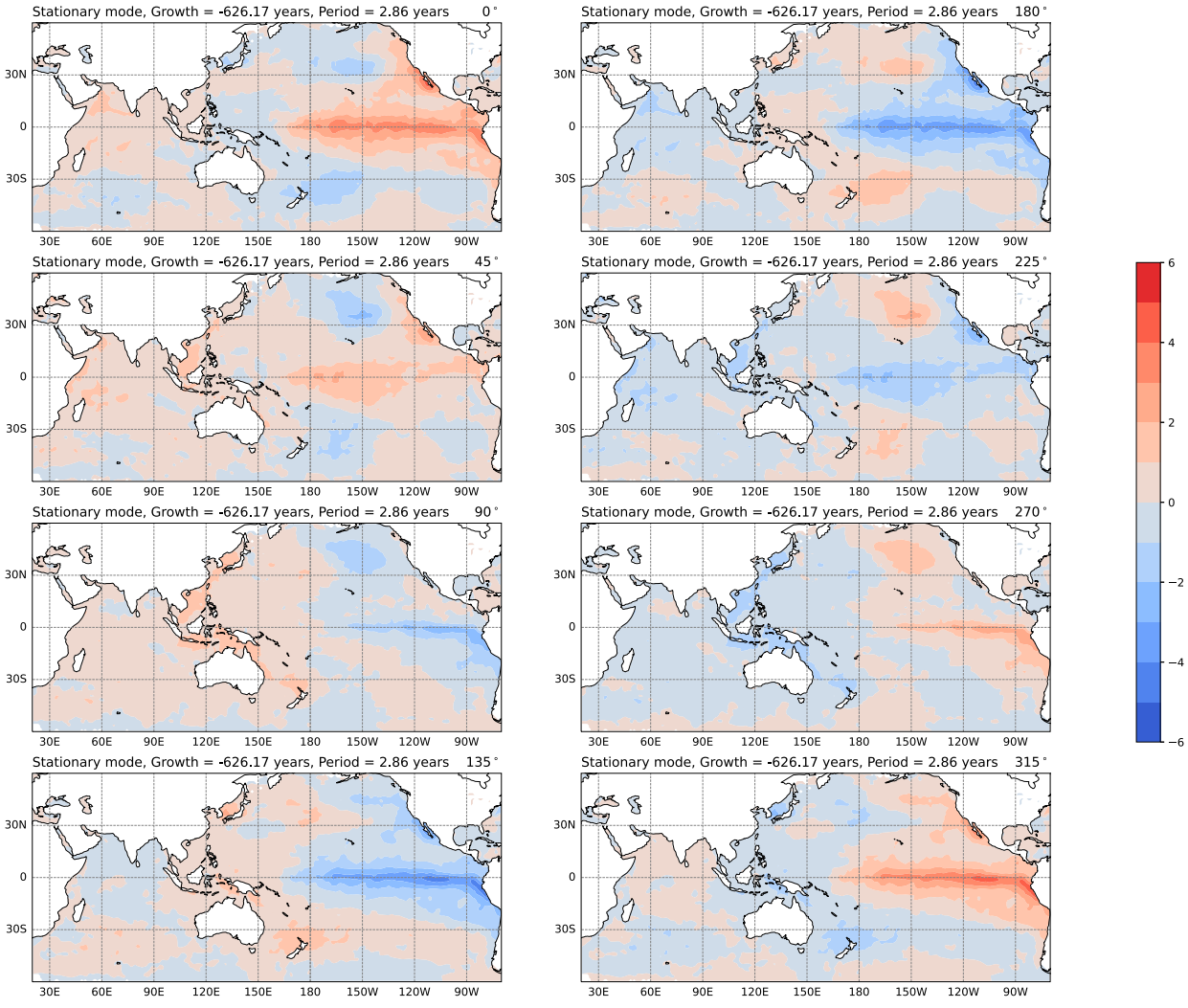

FIG. 17. As in Fig. 16, but for the stationary mode of period 2.86 years.

second composite is for months with a value of the index close to 0.

We can see a basic exponential decline as the frequency increases, but with notable peaks at modes with periods of 2–5 years for the cases with a large Niño/A index. The absolute value of the eigenfunction has no phase information and so it cannot separate positive and negative cases. The neutral case shows no peak at these frequencies.

The bottom panel shows similar composites but for the PDO index, defined as the mean anomaly over the area between (20°, 60°N) and (150°, 220°E) (Mantua et al. 1997). In this case, it is possible to note that the month with a high absolute value (greater than one standard deviation) of the index, either positive or negative, shows a large contribution from the slower modes, with a relatively minor component of the interannual variability. The months with smaller values of the index result in a rather flat distribution.

The interactions among the modes in the global case are shown in Fig. 18. The figure is organized as in Fig. 12 so that the meaning of the quadrants is similar. It is interesting to note that the stationary modes are more clearly orthogonal than in the tropical case and they are organized in well-defined blocks that define the sections of modes that have residual interaction among themselves but much weaker interactions with others. The time scales that emerge are longer periods up to 10–12 years

7. Discussion

The Koopman modes provide a new way of analyzing the variance of a climate field without enforcing a preorganized structure on the modes themselves. They do not force a mathematical structure, as the EOF or even Fourier analyses, but let the dynamically significant modes emerge naturally from the dataset. Koopman-based methods provide a global rather than local linearization and therefore allow the usage of the well-known linear operator theory concepts for the physical interpretation. Used in combination with

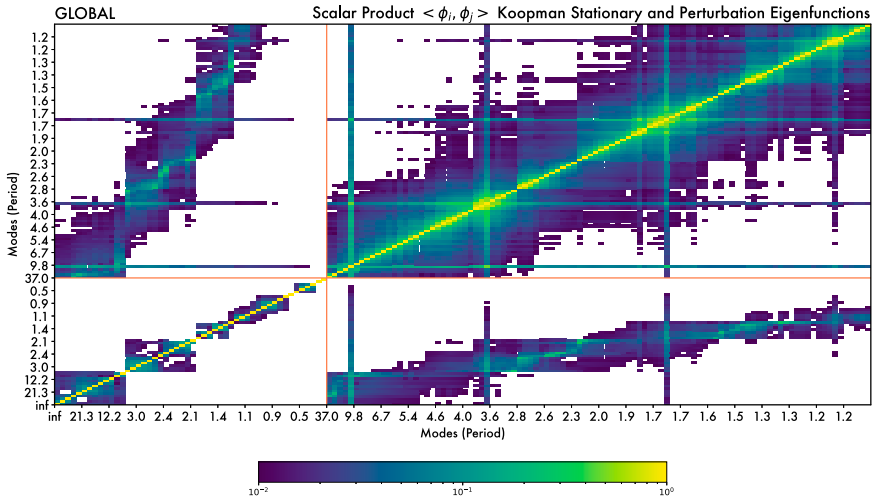

FIG. 18. Overlap integrals for the stationary and transient Koopman eigenfunctions for the global domain. The bottom left quadrant are the stationary mode overlaps, and the top right quadrant are the transient Koopman modes overlaps. The other quadrant denotes the interactions. Note the increased separation between time scales in the stationary sector.

kernel methods, they can provide a very powerful tool, which can be the basis for forecasting, but these methods can also be used to gain insight into the behavior of the system.

The vision that emerges for the tropical Pacific is consistent with a physical scenario where the variability is made up of several different modes, represented by the Koopman eigenfunctions and their projection on the state space, the Koopman modes. We propose here that, at least in the limited view offered only by the SST, the interannual variability is composed by a spectrum of patterns almost regularly oscillatory, which are however perturbed and modified continuously by the interaction with other eigenfunctions, often with larger *e*-folding times that may amplify the amplitude of the oscillation over a relatively short period of time and then decay. This scenario then offers a resolution to the discussion on the oscillatory versus damped oscillatory nature of the variability. The scenario is supported by the interaction coefficients among the eigenfunctions (Figs. 12 and 18), which shows not only how the stationary eigenfunctions are more orthogonal than the rest but also how they interact with the strongly damped eigenfunctions in a continuous interplay. The system that is based on a number of stationary or quasi-stationary modes that collectively define a subspace of regular oscillations and variations. The stationary subspace contributes to the internal variability directly, but it also provides the texture around which the nonorthogonal modes can act to amplify the variability. It is reasonable to expect that the amplification is generated partially by linear instabilities, but also by the repelling and attracting properties of the stationary subspace itself. These ideas will need to be further investigated and verified by extending the analysis to other parameters, from the atmosphere and subsurface in the ocean.

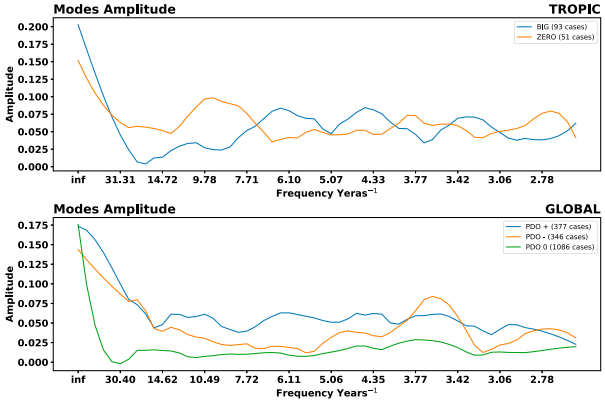

FIG. 19. The distribution of the absolute value of the values for each Koopman eigenfunction for (top) the Niño cases and (bottom) PDO. The Niño-3 + 4 have been stratified as the months with Niño/A cases (BIG: >1.5 or <−1.5) and neutral cases with an absolute value of the index smaller than 0.025. The PDO has been stratified as PDO > 0.15 and PDO < −0.15. The Niño/A cases show peaks, whereas in the neutral case, it is rather a monotonous distribution. The PDO is concentrated in a smooth quantity over the longer time scales.

The Koopman modes are estimated from data, and a theoretical calculation even for simple toy problems is still elusive, so they are limited by the characteristics of the datasets, especially in the choice of stationarity criteria (the "threshold") and in the details of the estimating algorithm (kernel choices, bandwidths, etc.). Some sensitivity with respect to the choice of parameters is to be expected. Although this is somewhat unavoidable for data-driven methods, it must be taken into account. The presence of the continuous spectrum typical of mixing systems, like the turbulent systems we encounter in climate and meteorological applications is a large challenge. However, we have shown that the stationary modes provide a guide to the phase space and also that the distribution of the states is not trivial, revealing an underlying structure.

8. Conclusions

This paper shows that the Koopman analysis can be used to analyze the long time series of large climatic datasets easily. The continuous part of the spectrum can be handled using density estimation techniques that identified the main frequencies. A purely multiperiodic system would be represented by a series of sharp lines in the spectrum, but the interaction with the continuous part causes line broadening and the appearance of wider peaks in the eigenvalue spectrum. A possible interpretation of ENSO variability, for instance, assumes that ENSO is the result of damped modes stochastically driven by high-frequency (weather) noise. The speculation here is that Koopman modes, and in particular the most prominent ones, may play a role as such modes.

The Koopman modes provide a theoretical framework that allow the discovery of the underlying dynamics implied by the (unknown) dynamical system that drives the evolution of the dataset, in this case the SST. The analysis has separated the variability into quasi-stationary modes and decaying modes that are in a continuous interaction. However, the stationary modes collectively are the scaffolding upon which the variability is built, as is shown by their capacity to represent a large part of the variability of Niño and PDO indices. The interactions with decaying modes provide the amplification over relatively short time scales. These ideas open a new avenue of investigation and will need to be further developed and investigated, especially in their relation with other parameters and fields of the coupled tropical system.

Acknowledgments. We thank three anonymous reviewers for their comments, which helped to improve the manuscript.

Data availability statement. Data and codes are available from the authors.

APPENDIX A

Notations

An overview of the notation used in the paper (see Table A1).

APPENDIX B

Spectral Decomposition of the Koopman Operator

The decomposition of the Koopman operator into eigenvalues, eigenfunctions, and modes has been computed using the algorithms proposed in Klus et al. (2020), described also in Navarra et al. (2021). The usage of kernels and reproducing kernel Hilbert spaces (Scholköpf and Smola 2002) leads to an improved approximation by using a richer set of features to estimate the Koopman operator, allowing similarity measures that are more general than the standard inner product. We define a function, the so-called *kernel*,

$$k(\mathbf{x}, \mathbf{y}) = \langle \phi(\mathbf{x}), \phi(\mathbf{y}) \rangle,$$

that generalizes the notion of similarity between two states. Such methods have been extensively used for classification, and other machine learning problems and the properties and the conditions to be satisfied by the kernel have been investigated (see Steinwart and Christmann (2008) for more details). Muandet et al. (2017) classified several possible kernels in terms of their conservation of properties of the basic system. Popular choices include radial basis kernels such as the Gaussian kernel $k(\mathbf{x}, \mathbf{y}) = \exp(-\|\mathbf{x} - \mathbf{y}\|^2/2s^2)$ or the Laplacian kernel $k(\mathbf{x}, \mathbf{y}) = \exp(-\|\mathbf{x} - \mathbf{y}\|/s)$. Also polynomial kernels $k(\mathbf{x}, \mathbf{y}) = (c + \langle \mathbf{x}, \mathbf{y} \rangle)^d$ have been used. The hyperparameters $c \geq 0$, $s > 0$, and $d \in \mathbb{N}$ are arbitrary, and they have to be chosen for each problem. The polynomial kernel reduces to the usual scalar product for $c = 0$ and $d = 1$.

The mathematical derivations described in Klus et al. (2016, 2020), and Navarra et al. (2021), and the corresponding algorithms will yield the eigenvalues, eigenfunctions, and modes. Given time series data of the form,

$$\mathbf{X} = [\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_N], \tag{B1}$$

where $\mathbf{x}_i$ represents snapshots of gridpoint values of physical fields (reshaped as vectors).

TABLE A1. Notations used in the paper.

$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots]$	Data matrix
$\mathbf{Y} = [\mathbf{x}_2, \mathbf{x}_3, \ldots]$	Shifted data matrix
k, f	Kernel and associated feature map
$\mathbf{F} = [f_1, f_2, \ldots]$	Features matrix
K	Koopman operator
$\mathbf{G}_{xx}$	Gram matrices for data matrices $\mathbf{X}$
$\mathbf{G}_{xy}$	Gram matrices for data matrices $\mathbf{X}$ and $\mathbf{Y}$
$\mathbf{\Phi} = [\phi_1, \phi_2, \ldots]$	Matrix of the eigenfunction of the Koopman operator
$\mathbf{\Psi} = [\psi_1, \psi_2, \ldots]$	Matrix of the eigenfunction of the Perron–Frobenius operator
w_{ik}	Coefficient of the feature expansion of the Koopman eigenfunctions
u_{ik}	Coefficient of the feature expansion of the Perron eigenfunctions
$\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \ldots]$	Koopman modes
μ	Eigenvalue of the Koopman operator
λ	Eigenvalue of the generators of the operators

It can be shown that the in the kernel formulation we can use the Gramian matrices $\mathbf{G}_{xx} = [k(\mathbf{x}_i, \mathbf{x}_j)]$ and $\mathbf{G}_{yx} = [k(\mathbf{y}_i, \mathbf{x}_j)]$, to set up the eigenvalues problem (Klus et al. 2020),

$$(\mathbf{G}_{xx} + n \in \mathbf{I})^{-1}\mathbf{G}_{yx}\mathbf{w} = \mu\mathbf{w}, \tag{B2}$$

then the expression

$$\mathbf{\Phi}(\mathbf{X}) = \mathbf{G}_{xx}\mathbf{W}, \tag{B3}$$

where the columns of $\mathbf{W}$ are the eigenvectors $\mathbf{w}$, which will give an estimation of the eigenfunctions of the Koopman operator in terms of their values on the data vectors. The Tikhonov regularization with identity matrix $\mathbf{I}$ can be added in the case of singularity of $\mathbf{G}_{xx}$, and alternatively, a pseudoinverse calculation can be used. The generator eigenvalues λ can then be obtained as $\lambda = \log(\mu)/\Delta t$, where Δt is the lag time, i.e., the difference between t_i and t_{i+1}.

The algorithms will then produce the matrix

$$\mathbf{\Phi}(\mathbf{X}) = \begin{bmatrix} \varphi_1(\mathbf{x}_1) & \varphi_2(\mathbf{x}_1) & \varphi_3(\mathbf{x}_1) \\ \varphi_1(\mathbf{x}_2) & \varphi_2(\mathbf{x}_2) & \dots \\ \varphi_1(\mathbf{x}_3) & \dots & \dots \end{bmatrix} \tag{B4}$$

of the values taken by the eigenfunctions $\varphi_i(\mathbf{x}_j)$ at the sample points. It will also provide coefficients w_{ik} and u_{ik} for the expansion of the eigenfunctions in the feature functions derived from the kernel $k(\cdot, \mathbf{x})$.

Having a vector-valued observable $\mathbf{g} = [g_1, g_2, g_3, \dots]^{\mathrm{T}}$ that takes values on the same sample points,

$$\mathbf{G}(\mathbf{X}) = \begin{bmatrix} g_1(\mathbf{x}_1) & g_1(\mathbf{x}_2) & g_1(\mathbf{x}_3) \\ g_2(\mathbf{x}_1) & g_2(\mathbf{x}_2) & \dots \\ g_3(\mathbf{x}_1) & \dots & \dots \end{bmatrix}, \tag{B5}$$

the Koopman modes can be obtained from the matrix of the value of the observable on the eigenfunctions, via the pseudoinverse $\Phi(X)^+$ (Mezic and Surana 2016),

$$\mathbf{V} = \mathbf{\Phi}(\mathbf{X})^+\mathbf{G}(\mathbf{X}). \tag{B6}$$

Then, the Koopman modes of the state vectors are

$$\mathbf{V} = \mathbf{\Phi}(\mathbf{X})^+\mathbf{X}. \tag{B7}$$

APPENDIX C

Koopman Decomposition and LIM

The computation of the Koopman modes will reduce to DMD modes if a linear kernel is used (Klus et al. 2016). In that case, the Koopman modes are equivalent to a LIM as shown by Tu et al. (2014). We include a slightly different derivation for the sake of completeness.

A numerical calculation of the Koopman modes will produce only discrete eigenvalues, so we can use the expansion in (4) using only the discrete eigenvalues,

$$\mathbf{x}(t)^K = \sum_k \mathbf{v}_k e^{\lambda_k t}\varphi_k[\mathbf{x}(0)]. \tag{C1}$$

In a LIM approximation, spatial covariance matrix $\mathbf{B}$ is used to compute Green's function to construct the predictor,

$$\mathbf{x}(t + \Delta t) = \exp(\mathbf{B}\Delta t)\mathbf{x}(t), \tag{C2}$$

where matrix $\mathbf{B}$ is obtained from the covariance matrix and time-lagged covariance matrix as,

$$\mathbf{B} = \frac{1}{\Delta t}\log[\mathbf{Y}\mathbf{X}^{\mathrm{T}}(\mathbf{X}\mathbf{X}^{\mathrm{T}})^{-1}], \tag{C3}$$

and $\mathbf{X}$ is the data matrix and $\mathbf{Y}$ is the shifted data matrix (Penland and Sardeshmukh 1995). We can see from this expression that LIM constructs an approximation of the generator of the Koopman operator using linear basis functions and that

$$\mathbf{C} = \mathbf{Y}\mathbf{X}^{\mathrm{T}}(\mathbf{X}\mathbf{X}^{\mathrm{T}})^{-1} = \mathbf{Y}\mathbf{X}^{+}$$

is thus an approximation of the Koopman operator with lag time Δt. Matrix $\mathbf{C}$ can be regarded as an EDMD estimate of the operator using linear basis functions, which is equivalent to DMD. Note that $\mathbf{C}$ is the solution of the regression problem

$$\min_{\mathbf{A}\in\mathbb{R}^{d\times d}}\|\mathbf{Y} - \mathbf{A}\mathbf{X}\|_F. \tag{C4},$$

Choosing a linear kernel in (B2) allows us to write the Gramians in terms of the data matrices as

$$\mathbf{G}_{xx} = \mathbf{X}^{\mathrm{T}}\mathbf{X}, \tag{C5}$$

$$\mathbf{G}_{yx} = \mathbf{Y}^{\mathrm{T}}\mathbf{X}, \tag{C6}$$

and the Koopman eigenfunctions are obtained from the eigenvalue problem,

$$(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{Y}^{\mathrm{T}}\mathbf{X}\mathbf{W} = \mathbf{W}\mathbf{\Lambda}_K, \tag{C7}$$

where $\mathbf{W}$ is the eigenvector matrix and $\mathbf{\Lambda}_K$ is a diagonal matrix containing the eigenvalues of the Koopman operator. The corresponding eigenfunctions are then given by $\varphi(x) = (\mathbf{X}w)^{\mathrm{T}}x$. Note that $\mathbf{X}w$ is a right eigenvector of $\mathbf{C}^{\mathrm{T}}$ since, using properties of the pseudoinverse,

$$(\mathbf{X}\mathbf{X}^{\mathrm{T}})^{+}(\mathbf{X}\mathbf{Y}^{\mathrm{T}})\mathbf{X}w = \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{+}(\mathbf{Y}^{\mathrm{T}}\mathbf{X})w = \lambda_K\mathbf{X}w$$

and thus a left eigenvector of $\mathbf{C}$. Note that depending on the size of $\mathbf{X}$, $\mathbf{X}\mathbf{X}^{\mathrm{T}}$, and/or $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ might not be invertible, and we thus have to use pseudoinverses here. The LIM eigenfunctions, on the other hand, are obtained from the right eigenvectors of matrix $\mathbf{B}$, which are also the solutions of the corresponding eigenvalue problem,

$$\mathbf{CU} = \mathbf{Y}\mathbf{X}^{\mathrm{T}}(\mathbf{X}\mathbf{X}^{\mathrm{T}})^{-1}\mathbf{U} = \mathbf{U}\boldsymbol{\Lambda}_L.$$

The difference between LIM and EDMD is thus that LIM computes the right eigenvectors and EDMD computes the left eigenvectors of $\mathbf{C}$.

The Koopman modes $\mathbf{V}$ are obtained by representing the full-state observable $g(x) = x$ in terms of the Koopman eigenfunctions. Since in our case, the dictionary is given by $\phi(x) = x$, this yields

$$\mathbf{V} = (\mathbf{X}\mathbf{W})^{-\mathrm{T}}$$

[see Klus et al. (2016) for more details]. The Koopman modes are then the right eigenvectors of $\mathbf{C}$ and for this special case equivalent to LIM. However, this equivalence holds only if a trivial linear kernel is used. More powerful nonlinear kernels lead to more accurate approximations of the eigenfunctions since they map the data to higher-dimensional or infinite-dimensional feature spaces. The resulting function spaces are, under certain conditions, dense in the space of continuous functions.

REFERENCES

Berry, T., and J. Harlim, 2016: Variable bandwidth diffusion kernels. *Appl. Comput. Harmonic Anal.*, **40**, 68–96, https://doi.org/10.1016/j.acha.2015.01.001.

Budišić, M., R. Mohr, and I. Mezić, 2012: Applied koopmanism. *Chaos*, **22**, 047510, https://doi.org/10.1063/1.4772195.

Cvitanović, P., R. Artuso, R. Mainieri, G. Tanner, and G. Vattay, 2020: *Chaos: Classical and Quantum*. Niels Bohr Institute, 1089 pp., http://ChaosBook.org/.

Das, S., and D. Giannakis, 2020: Koopman spectra in reproducing kernel Hilbert spaces. *Appl. Comput. Harmonic Anal.*, **49**, 573–607, https://doi.org/10.1016/j.acha.2020.05.008.

Dommenget, D., and M. Latif, 2002: A cautionary note on the interpretation of EOFs. *J. Climate*, **15**, 216–225, https://doi.org/10.1175/1520-0442(2002)015<0216:ACNOTI>2.0.CO;2.

Farrell, B., 1988: Optimal excitation of neutral Rossby waves. *J. Atmos. Sci.*, **45**, 163–172, https://doi.org/10.1175/1520-0469(1988)045<0163:OEONRW>2.0.CO;2.

Farrell, B. F., 1989: Optimal excitation of baroclinic waves. *J. Atmos. Sci.*, **46**, 1193–1206, https://doi.org/10.1175/1520-0469(1989)046<1193:OEOBW>2.0.CO;2.

Frederiksen, J. S., 2000: Singular vectors, finite-time normal modes, and error growth during blocking. *J. Atmos. Sci.*, **57**, 312–333, https://doi.org/10.1175/1520-0469(2000)057<0312:SVFTNM>2.0.CO;2.

Froyland, G., D. Giannakis, B. R. Lintner, M. Pike, and J. Slawinska, 2021: Spectral analysis of climate dynamics with operator-theoretic approaches. *Nat. Commun.*, **12**, 6570, https://doi.org/10.1038/s41467-021-26357-x.

Ghil, M., and Coauthors, 2002: Advanced spectral methods for climatic time series. *Rev. Geophys.*, **40**, 1003, https://doi.org/10.1029/2000RG000092.

Giannakis, D., 2019: Data-driven spectral decomposition and forecasting of ergodic dynamical systems. *Appl. Comput. Harmonic Anal.*, **47**, 338–396, https://doi.org/10.1016/j.acha.2017.09.001.

——, and A. J. Majda, 2012a: Comparing low-frequency and intermittent variability in comprehensive climate models through nonlinear Laplacian spectral analysis. *Geophys. Res. Lett.*, **39**, L10710, https://doi.org/10.1029/2012GL051575.

——, and ——, 2012b: Nonlinear Laplacian spectral analysis for time series with intermittency and low-frequency variability. *Proc. Natl. Acad. Sci. USA*, **109**, 2222–2227, https://doi.org/10.1073/pnas.1118984109.

——, J. Slawinska, and Z. Zhao, 2015: Spatiotemporal feature extraction with data-driven Koopman operators. *Proc. First Int. Workshop on Feature Extraction: Modern Questions and Challenges at NIPS 2015*, Montreal, QC, Canada, PLMR, 103–115, https://proceedings.mlr.press/v44/giannakis15.html.

Hannachi, A., I. T. Jolliffe, and D. B. Stephenson, 2007: Empirical orthogonal functions and related techniques in atmospheric science: A review. *Int. J. Climatol.*, **27**, 1119–1152, https://doi.org/10.1002/joc.1499.

Hogg, J., M. Fonoberova, and I. Mezić, 2020: Exponentially decaying modes and long-term prediction of sea ice concentration using Koopman mode decomposition. *Sci. Rep.*, **10**, 16313, https://doi.org/10.1038/s41598-020-73211-z.

Klus, S., P. Koltai, and C. Schütte, 2016: On the numerical approximation of the Perron–Frobenius and Koopman operator. *J. Comput. Dyn.*, **3**, 51–79, https://doi.org/10.3934/jcd.2016003.

——, F. Nüske, P. Koltai, H. Wu, I. Kevrekidis, C. Schütte, and F. Noé, 2018: Data-driven model reduction and transfer operator approximation. *J. Nonlinear Sci.*, **28**, 985–1010, https://doi.org/10.1007/s00332-017-9437-7.

——, I. Schuster, and K. Muandet, 2020: Eigendecompositions of transfer operators in reproducing kernel Hilbert spaces. *J. Nonlinear Sci.*, **30**, 283–315, https://doi.org/10.1007/s00332-019-09574-z.

Koopman, B. O., 1931: Hamiltonian systems and transformation in Hilbert space. *Proc. Natl. Acad. Sci. USA*, **17**, 315–318, https://doi.org/10.1073/pnas.17.5.315.

——, and J. Neumann, 1932: Dynamical systems of continuous spectra. *Proc. Natl. Acad. Sci. USA*, **18**, 255–263, https://doi.org/10.1073/pnas.18.3.255.

Kutzbach, J. E., 1970: Large-scale features of monthly mean northern hemisphere anomaly maps of sea-level pressure. *Mon. Wea. Rev.*, **98**, 708–716, https://doi.org/10.1175/1520-0493(1970)098<0708:LSFOMM>2.3.CO;2.

Lau, K.-M., and P. H. Chan, 1985: Aspects of the 40–50 day oscillation during the northern winter as inferred from outgoing longwave radiation. *Mon. Wea. Rev.*, **113**, 1889–1909, https://doi.org/10.1175/1520-0493(1985)113<1889:AOTDOD>2.0.CO;2.

Lau, N.-C., 1981: A diagnostic study of recurrent meteorological anomalies appearing in a 15-year simulation with a GFDL general circulation model. *Mon. Wea. Rev.*, **109**, 2287–2311, https://doi.org/10.1175/1520-0493(1981)109<2287:ADSORM>2.0.CO;2.

Lorenz, E. N., 1956: Empirical orthogonal functions and statistical weather prediction. Statistical Forecast Project Tech. Rep. 1, 52 pp.

Mantua, N. J., and S. R. Hare, 2002: The Pacific decadal oscillation. *J. Oceanogr.*, **58**, 35–44, https://doi.org/10.1023/A:1015820616384.

——, ——, Y. Zhang, J. M. Wallace, and R. C. Francis, 1997: A Pacific interdecadal climate oscillation with impacts on salmon production. *Bull. Amer. Meteor. Soc.*, **78**, 1069–1080, https://doi.org/10.1175/1520-0477(1997)078<1069:APICOW>2.0.CO;2.

McGibbon, R. T., and V. S. Pande, 2015: Variational cross-validation of slow dynamical modes in molecular kinetics. *J. Chem. Phys.*, **142**, 124105, https://doi.org/10.1063/1.4916292.

Mezić, I., 2005: Spectral properties of dynamical systems, model reduction and decompositions. *Nonlinear Dyn.*, **41**, 309–325, https://doi.org/10.1007/s11071-005-2824-x.

——, 2013: Analysis of fluid flows via spectral properties of the Koopman operator. *Annu. Rev. Fluid Mech.*, **45**, 357–378, https://doi.org/10.1146/annurev-fluid-011212-140652.

Mezic, I., and A. Surana, 2016: Koopman mode decomposition for periodic/quasi-periodic time dependence. *IFAC-PapersOnLine*, **49**, 690–697, https://doi.org/10.1016/j.ifacol.2016.10.246.

Muandet, K., K. Fukumizu, B. Sriperumbudur, and B. Schölkopf, 2017: Kernel mean embedding of distributions: A review and beyond. *Found. Trends Mach. Learn.*, **10** (1–2), 1–141, https://doi.org/10.1561/2200000060.

Navarra, A., J. Tribbia, and S. Klus, 2021: Estimation of Koopman transfer operators for the equatorial Pacific SST. *J. Atmos. Sci.*, **78**, 1227–1244, https://doi.org/10.1175/JAS-D-20-0136.1.

Newman, M., and Coauthors, 2016: The Pacific decadal oscillation, revisited. *J. Climate*, **29**, 4399–4427, https://doi.org/10.1175/JCLI-D-15-0508.1.

Noé, F., and F. Nüske, 2013: A variational approach to modeling slow processes in stochastic dynamical systems. *Multiscale Model. Simul.*, **11**, 635–655, https://doi.org/10.1137/110858616.

Nueske, F., B. G. Keller, G. Perez-Hernandez, A. S. J. S. Mey, and F. Noe, 2014: Variational approach to molecular kinetics. *J. Chem. Theory Comput.*, **10**, 1739–1752, https://doi.org/10.1021/ct4009156.

Penland, C., 1996: A stochastic model of IndoPacific sea surface temperature anomalies. *Physica D*, **98**, 534–558, https://doi.org/10.1016/0167-2789(96)00124-8.

——, and P. D. Sardeshmukh, 1995: The optimal growth of tropical sea surface temperature anomalies. *J. Climate*, **8**, 1999–2024, https://doi.org/10.1175/1520-0442(1995)008<1999:TOGOTS>2.0.CO;2.

Quadrelli, R., and J. M. Wallace, 2004: A simplified linear framework for interpreting patterns of Northern Hemisphere wintertime climate variability. *J. Climate*, **17**, 3728–3744, https://doi.org/10.1175/1520-0442(2004)017<3728:ASLFFI>2.0.CO;2.

Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.*, **108**, 4407, https://doi.org/10.1029/2002JD002670.

Rowley, C. W., I. Mezić, S. Bagheri, P. Schlatter, and D. S. Henningson, 2009: Spectral analysis of nonlinear flows. *J. Fluid Mech.*, **641**, 115–127, https://doi.org/10.1017/S0022112009992059.

Scholköpf, B., and A. J. Smola, 2002: *Learning with Kernels*. MIT Press, 626 pp.

Simmons, A. J., J. M. Wallace, and G. W. Branstator, 1983: Barotropic wave propagation and instability, and atmospheric teleconnection patterns. *J. Atmos. Sci.*, **40**, 1363–1392, https://doi.org/10.1175/1520-0469(1983)040<1363:BWPAIA>2.0.CO;2.

Steinwart, I., and A. Christmann, 2008: *Support Vector Machines*. 1st ed. Springer, 603 pp.

Trenberth, K. E., and D. A. Paolino Jr., 1981: Characteristic patterns of variability of sea level pressure in the Northern Hemisphere. *Mon. Wea. Rev.*, **109**, 1169–1189, https://doi.org/10.1175/1520-0493(1981)109<1169:CPOVOS>2.0.CO;2.

Tu, J. H., 2013: Dynamic mode decomposition: Theory and applications. Ph.D. thesis, Princeton University, 175 pp.

——, C. W. Rowley, D. M. Luchtenburg, S. L. Brunton, and J. N. Kutz, 2014: On dynamic mode decomposition: Theory and applications. *J. Comput. Dyn.*, **1**, 391–421, https://doi.org/10.3934/jcd.2014.1.391.

Tung, W.-W., D. Giannakis, and A. J. Majda, 2014: Symmetric and antisymmetric convection signals in the Madden–Julian Oscillation. Part I: Basic modes in infrared brightness temperature. *J. Atmos. Sci.*, **71**, 3302–3326, https://doi.org/10.1175/JAS-D-13-0122.1.

Ulam, S. M., 1960: *A Collection of Mathematical Problems*. Interscience Publishers, 150 pp.

Wang, X., J. Slawinska, and D. Giannakis, 2020: Extended-range statistical ENSO prediction through operator-theoretic techniques for nonlinear dynamics. *Sci. Rep.*, **10**, 2636, https://doi.org/10.1038/s41598-020-59128-7.

Williams, M. O., I. G. Kevrekidis, and C. W. Rowley, 2015a: A data–driven approximation of the Koopman operator: Extending dynamic mode decomposition. *J. Nonlinear Sci.*, **25**, 1307–1346, https://doi.org/10.1007/s00332-015-9258-5.

——, C. W. Rowley, and I. G. Kevrekidis, 2015b: A kernel-based method for data-driven Koopman spectral analysis. *J. Comput. Dyn.*, **2**, 247–265, https://doi.org/10.3934/jcd.2015005.