

## Estimation of Koopman Transfer Operators for the Equatorial Pacific SST

ANTONIO NAVARRA,<sup>a,b</sup> JOE TRIBBIA,<sup>c</sup> AND STEFAN KLUS<sup>d</sup>

<sup>a</sup> *Centro Euromediterraneo sui Cambiamenti Climatici, Bologna, Italy*

<sup>b</sup> *Department of Biological, Geological, and Environmental Sciences (BIGEA), University of Bologna, Bologna, Italy*

<sup>c</sup> *National Center for Atmospheric Research, Boulder, Colorado*

<sup>d</sup> *Department of Mathematics, University of Surrey, Guildford, United Kingdom*

(Manuscript received 7 May 2020, in final form 5 January 2021)

**ABSTRACT:** In the last years, ensemble methods have been widely popular in atmospheric, climate, and ocean dynamics investigations and forecasts as convenient methods to obtain statistical information on these systems. In many cases, ensembles have been used as an approximation to the probability distribution that has acquired more and more a central role, as the importance of a single trajectory, or member, was recognized as less informative. This paper shows that using results from the dynamical systems and more recent results from the machine learning and AI communities, we can arrive at a direct estimation of the probability distribution evolution and also at the formulation of predictor systems based on a nonlinear formulation. The paper introduces the theory and demonstrates its application to two examples. The first is a one-dimensional system based on the Niño-3 index; the second is a multidimensional case based on time series of monthly mean SST in the Pacific. We show that we can construct the probability distribution and set up a system to forecast its evolution and derive various quantities from it. The objective of the paper is not strict realism, but the introduction of these methods and the demonstration that they can be used also in the complex, multidimensional environment typical of atmosphere and ocean applications.

**KEYWORDS:** Atmosphere; Ocean; Statistical techniques; Superensembles

### 1. Introduction


In the last 20 years the evolution of the methodological approach to atmosphere and ocean forecasting and in general to investigations of the dynamics of weather and climate has progressed toward a massive usage of ensemble techniques. Starting with the pioneering work of Molteni et al. (1996) and Toth and Kalnay (1993), the generation of ensembles has become the dominant approach in forecasting (Barkmeijer et al. 2013; Schwartz et al. 2019; Bell and Kirtman, 2019) and in climate projections and scenarios (Kay et al. 2015; Tebaldi and Knutti 2007; Maher et al. 2019).

The ensemble members are usually generated by perturbing initial conditions, boundary conditions, or involving different models and/or resolutions. The objective was to sample as much as possible the phase space, implicitly recognizing that the essential information cannot be considered contained in the single forecasts, but rather in their distribution and variance. In practice, we are shifting the forecasting problem from an individual forecast to forecasting the probability distribution of the variables of interest.

This shift has been empirically motivated and it has yielded important successes, but it has also significant fundamental consequences, since we are moving away from the “trajectory picture” and adopting instead a “probability picture.” The difference is that we have extensive information on the dynamics and properties of the trajectories, i.e., the forecast

or single integration, but in reality we know much less about the properties of the ensemble of trajectories, i.e., the probability distribution. Recently, some attempts have been made to apply statistical methods, like revisiting the analog method (Ding et al. 2019), or using convolutional neural networks to design a predictive system (Ham et al. 2019). Very recently Wang et al. (2020) extended the analog method by combining it with a kernel approach that allowed them to design a prediction system for the Niño indices in the Pacific. They showed that using a kernel as a similarity measure generalizes the analog concept and includes other measures of similarity than the linear measure based on the spatial inner product that is used in linear inverse model (LIM) methods (Penland and Sardeshmukh 1995).

Of course, there are very good reasons for this. For the atmosphere and the ocean, the probability evolution can probably be formulated theoretically, but with very little chance of being treated in practice to get feasible calculations and estimates. Furthermore, we do not know much about the properties of the probability for the atmosphere and its evolution. In principle, the problem is difficult since it is really the probability of a field configuration (the temperature, the wind, . . .), so it needs to be treated with the tools of functional analysis, a tough problem for nonlinear fluid systems like the atmosphere and the ocean. Discretized systems are equivalent to systems of ordinary differential equations and therefore are simpler to deal with, but the dimensionality quickly becomes a problem. This problem is made explicit by the attempts to use stochastic models leading to the Fokker–Planck equation (Navarra et al. 2013; Majda and Qi 2020) that described the evolution of the probability as a function of the degrees of freedom of the problem, basically confined it to highly idealized models. Berry et al. (2015) used nonparametric methods to estimate forecasting models for

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

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

low-dimension systems, showing that it is indeed possible to estimate the evolution equation for the probability distribution without having to identify the equation itself in closed form. Their approach, however, was applied to low-order systems and used some assumption of stochasticity for the system.

The trajectory picture is not the only one possible for a system. In a couple of exceptional papers, [Koopman \(1931\)](#) and [Koopman and Neumann \(1932\)](#) proposed an alternative approach showing that a system can be equivalently described by an operator acting on a function space. The Koopman operator picture ([Rowley et al. 2009](#); [Budišić et al. 2012](#)) shows that for every dynamical system there is a linear operator acting on a function space whose spectral properties, namely, eigenvalues, eigenfunctions and modes, completely characterize the dynamical system. Chaotic dynamical systems may have a partially or entirely continuous spectrum that in practice is approximated numerically. The link between the Koopman operator and the dynamical system is provided by the fact that the function space on which it operates is the space of the functions of the state variables of the dynamical system itself. The function space can be made into a Hilbert space with a suitable measure.

The explicit expression of the Koopman operator in closed form, however, was only possible for simple systems amenable to analytical treatment, until recently a number of results have improved upon the numerical algorithm by [Ulam \(1960\)](#), introducing the *extended dynamic mode decomposition* (EDMD) ([Williams et al. 2015a,b](#); [Klus et al. 2016](#)) and the *variational approach of conformation dynamics* (VAC) ([Noé and Nüske 2013](#); [Nüske et al. 2014](#)). A review of these methods can be found in [Klus et al. \(2018\)](#), 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 observation and simulation data.

It was further recognized that the adjoint of the Koopman operator is the Perron–Frobenius operator ([Lasota and Mackey 1994](#); [Beck and Schlögl 1995](#)). The Perron–Frobenius operator is very interesting because it acts on the space of densities of the state space of the system. So whereas the Koopman operator provides information on the evolution of functions of the state (sometimes referred to as *observables*), its adjoint, the Perron–Frobenius operator evolves densities of trajectories of the state space. Both operators can be estimated using databased techniques. In this paper we will describe the connection between the Koopman and Perron–Frobenius operator (collectively known as *transfer operators*) and then we will examine some examples using the algorithms of [Klus et al. \(2019\)](#).

## 2. Transfer operators

A deterministic dynamical system can be defined by an ordinary differential equation of the form

$$\frac{d\mathbf{x}}{dt} = F(\mathbf{x}), \quad (1)$$

where  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  is a vector of dimension  $n$ ; we can identify “states,” the  $\mathbf{x}$  vectors, and “observables,”  $g(\mathbf{x})$ ,

basically any function of the states. The Koopman operator  $\mathcal{K}$  evolves the functions of the states in time,

$$\mathcal{K}g(\mathbf{x}) = g[U_\tau(\mathbf{x})],$$

where  $U_\tau$  is the operator associated with  $F$  that evolves the state for a fixed lag time  $\tau$ . Note that the Koopman operator (and consequently also its eigenvalues) implicitly depends on the chosen lag time. For the sake of simplicity, we will, however, omit this dependency and simply write  $\mathcal{K}$ . This operator is linear, even if the underlying dynamics is nonlinear. The operator contains all the information of the dynamical system and it constitutes an alternative formulation to the “trajectory picture.” In this “operator picture” we can use all the machinery developed for the analysis of linear operators, with the complication that this operator is infinite-dimensional even for finite-dimensional dynamical systems.

The adjoint of the Koopman operator is acting on the dual space of the observables. It can be shown ([Mezić 2005](#)) that it acts on measures (densities) that can be interpreted as probabilities,

$$\rho_\tau(\mathbf{x}) = \mathcal{P}\rho_0(\mathbf{x}) = \rho[U_\tau(\mathbf{x})].$$

The probability of finding the system in phase space volume  $\Delta\mathbf{x}$  at time  $t$  is then given by  $\rho_t(\mathbf{x})\Delta\mathbf{x}$ . If the initial probability distribution is highly localized around a certain state  $\mathbf{x}_0$ , then it can be interpreted as the conditional probability of finding the system in the state  $\mathbf{x}$  given that it was at  $\mathbf{x}_0$  at  $t = 0$ . In general it describes the absolute probability for the states.

These two operators describe completely the underlying dynamical system, so that every dynamical system induces a natural measure or density that describes the behavior of sets of trajectories under the transformation of the dynamical system. They are linked by a duality relation

$$\langle \mathcal{K}g_1, g_2 \rangle = \langle g_1, \mathcal{P}g_2 \rangle,$$

where  $\langle \cdot, \cdot \rangle$  is an appropriate inner product.

An alternative formulation is given in terms of the infinitesimal generator of  $\mathcal{K}$ , with

$$\mathcal{K}g(\mathbf{x}) = e^{\mathcal{L}\tau}g(\mathbf{x}),$$

where the operator  $\mathcal{L}$  is given by

$$\mathcal{L}g = \sum_{i=1}^n F_i \frac{\partial g}{\partial x_i}$$

and for the Perron–Frobenius operator

$$\mathcal{P}\rho(\mathbf{x}) = e^{\mathcal{L}^\dagger\tau}\rho(\mathbf{x}),$$

where  $\mathcal{L}^\dagger$  is the adjoint of  $\mathcal{L}$ . The sum of  $\mathcal{L}$  and its adjoint is given by

$$(\mathcal{L} + \mathcal{L}^\dagger)g = -\sum_{i=1}^n \frac{\partial F_i}{\partial x_i} g.$$

So if a dynamical system is volume preserving in the phase space, i.e.,

$$\sum_{i=1}^n \frac{\partial F_i}{\partial x_i} = 0,$$

then the generator is skew adjoint and the Koopman and Perron–Frobenius operators can be characterized by the same generator. In the case that the dynamical system does not preserve volume, for instance for dissipative systems, there is a difference and the two operators will have different generators. This property is depending on the measure that is being used, for instance the Lorenz attractor is dissipative with respect to the standard Lebesgue measure, but it is ergodic and measure preserving with respect to the measure defined by the probability distribution on the attractor.

A unitary operator has the property that the eigenvalues  $\mu_i$  are all of modulus 1, distributed on the unit circle. For eigenvalues inside the unit circle it holds that  $|\mu_i| \leq 1$  and, if the system is ergodic, the eigenfunction of the Koopman operator corresponding to the eigenvalue 1 is a constant function and the associated eigenfunction of the Perron–Frobenius operator is the steady-state probability distribution. Transfer operators associated with complex dynamical systems might have continuous spectra. The analysis of such problems, however, is more challenging and beyond the scope of this paper. Numerically, we are computing transfer operators projected onto finite-dimensional spaces. See [Giannakis \(2019\)](#) for a more detailed discussion.

It is interesting to remark that in the case of a stochastic system, the infinitesimal generator of the Perron–Frobenius operator is the Fokker–Planck equation, whereas the infinitesimal generator of the Koopman operator is the Kolmogorov backward equation.

The definitions can be easily extended to the temporally discrete case,

$$\mathbf{x}(t + \tau) = \mathbf{x}(t) + \tau F(\mathbf{x}).$$

The eigenfunctions are the same and the eigenvalues transform from  $\mu$  for the discrete case to  $\lambda = \log(\mu)/\tau$  for the continuous case.

A comprehensive analysis of the mathematical properties of transfer operators can be found in [Giannakis \(2019\)](#), where the relations between the ergodic properties of the dynamical system and the spectrum of transfer operators are discussed in detail.

In meteorology and/or climate in the past 20 years the usage of ensemble methods has steadily increased and they are now a standard procedure. The underlying assumption is that we can sample the phase space of the system and estimate the probability distribution of the variable of interest. There is an implied realization that the focus of the forecast is shifted from the target of obtaining the “right” trajectory that more closely follows the real evolution of the real atmosphere/ocean system to that of obtaining the distribution of probabilities of various outcomes.

The single integration from a specific initial condition is now less important, because we recognize the fact that the sensitivity to initial conditions and in general to the multiple nonlinear processes present in the system is causing the information to be shifted from a single trajectory to their

collective behavior and properties. The collective behavior is completely described by the probability distribution.

The problem can now be formulated considering that every law governing the evolution of the atmosphere and ocean can be reduced to the abstract form (1), where  $\mathbf{x}$  represents the variable after discretization and  $F(\mathbf{x})$  represents the interactions and processes that regulate the time evolution. A single trajectory is evolved by the time evolution operator  $U$  so that

$$\mathbf{x}(\tau) = U_\tau \mathbf{x}_0$$

and its inverse

$$\mathbf{x}(-\tau) = U_{-\tau} \mathbf{x}_0.$$

If we have an ensemble of initial conditions distributed according to a probability distribution  $\rho_0(\mathbf{x})$ , then the probability density at time  $\tau$  is given by ([Gaspard et al. 1995](#); [Gaspard and Tasaki 2001](#))

$$\begin{aligned} \rho_\tau(\mathbf{x}) &= \int \delta(\mathbf{x} - U_\tau \mathbf{y}) \rho_0(\mathbf{y}) \, d\mathbf{y} \\ &= \left| \frac{\partial(U_{-\tau} \mathbf{x})}{\partial \mathbf{x}} \right| \rho_0(U_{-\tau} \mathbf{x}) = (\mathcal{P} \rho_0)(\mathbf{x}), \end{aligned}$$

and this relation defines  $\mathcal{P}$ , the Perron–Frobenius operator. There are no other assumptions on the probability, but if we require that the trajectories are conserved locally, i.e., no trajectory disappears suddenly in any volume of the phase space, then the probability obeys the Liouville equation ([Poincaré 1906](#); [Gaspard 2007](#)).

### 3. Transfer operators and data

#### a. Estimating transfer operators from data

Transfer operators are a well-defined concept, but their abstract definition has made them elusive to analytical investigation, except in very simple cases. However, it was shown in [Williams et al. \(2015a,b\)](#) and [Klus et al. \(2018, 2019\)](#) that transfer operators can be estimated from data, observations, or simulations, using the EDMD. Assuming such a dataset is available describing a time-evolving system from an initial time  $t = 0$  to a final time  $t = T$ , we can organize the data as follows:

$$\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_m]. \tag{2}$$

Here, the vectors  $\mathbf{z}_i$  of length  $n$  describe the system at different times, from  $t = 0$  for  $i = 1$  to  $t = T$  for  $i = m$ , with a discrete time interval of  $\tau$ . The conceptual picture here is that we take the data to represent the sampled evolution of the system according to an unknown dynamical equations.

This organization of the data is similar to the basic data organization used in many other methods, from the linear inverse methods, LIM, ([Penland and Sardeshmukh 1995](#); [Penland 1996](#)) to simple EOF calculation or more advanced methods such as singular spectra analysis ([Vautard and Ghil 1989](#)). These methods rely on various forms of regression

and assume a stochastic component for the dynamics. The time covariance matrix constructed from the data, often with some lags included, is the central component and it can be defined from the data matrix (6) as

$$\mathbf{C}_{zz} = \mathbf{Z}\mathbf{Z}^T.$$

The dimension of the covariance matrix is given by the spatial length of the data and so usually space reduction is needed before implementing these methods. However, the Gram matrix

$$\mathbf{G}_{zz} = \mathbf{Z}^T\mathbf{Z}$$

has the dimension of the length of the time series, *irrespective* of the length of the vectors, usually grid points that can easily be in the thousands. The covariance matrix contains the covariance information for each spatial degree of freedom, usually grid points, whereas the Gram matrix elements are basically overlap integrals of the spatial field at different time levels, measuring the degree of similarity between them. In what follows we will show that using the Gram matrices we can get useful information from large datasets.

The vectors can be two- or even three-dimensional fields, like temperature or geopotential, expressed as vectors of grid points. The vectors are elements of a finite dimensional space that we will define the “state space” and the vectors themselves are called “states.”

Then we can define the evolution data matrix (defined here for lag one, but other lags can be considered) as the shifted matrix,

$$\mathbf{Y} = [\mathbf{z}_2, \mathbf{z}_3, \dots, \mathbf{z}_{m+1}].$$

It can be shown that the auxiliary matrix eigenvalue problems

$$\begin{aligned} (\mathbf{G}_{zz} + n\epsilon I)^{-1} \mathbf{G}_{zy} \mathbf{v}^k &= \mu \mathbf{v}^k, \\ (\mathbf{G}_{zz} + n\epsilon I)^{-1} \mathbf{G}_{yz} \mathbf{v}^p &= \mu \mathbf{v}^p \end{aligned} \tag{3}$$

can be used to determine eigenvalues and eigenfunctions of the Perron–Frobenius and Koopman operators, respectively; see Klus et al. (2019). Here,  $\epsilon$  is a (Tikhonov) regularization parameter which can be added to ensure that the inverse exists. Alternatively, a pseudoinverse calculation can be used. The continuous eigenvalues can be obtained as  $\lambda = \log(\mu)/\tau$ , where  $\tau$  is the time interval between two successive states of the time series.

If the standard inner product is used, then  $\mathbf{G}_{zz} = \mathbf{Z}^T\mathbf{Z}$  is the Gram matrix of the data matrix  $\mathbf{Z}$ , whereas  $\mathbf{G}_{zy} = \mathbf{Z}^T\mathbf{Y}$  is the Gram matrix of the data matrix  $\mathbf{Z}$  and the shifted data matrix  $\mathbf{Y}$ . The elements of the Gram matrix are estimating the similarity or correlation between the state vectors at different times,

$$\mathbf{G}_{zz} = \begin{bmatrix} \langle \mathbf{z}_1, \mathbf{z}_1 \rangle & \dots & \langle \mathbf{z}_1, \mathbf{z}_m \rangle \\ \vdots & \ddots & \vdots \\ \langle \mathbf{z}_m, \mathbf{z}_1 \rangle & \dots & \langle \mathbf{z}_m, \mathbf{z}_m \rangle \end{bmatrix} \tag{4a}$$

and

$$\mathbf{G}_{zy} = \begin{bmatrix} \langle \mathbf{z}_1, \mathbf{z}_2 \rangle & \dots & \langle \mathbf{z}_1, \mathbf{z}_{m+1} \rangle \\ \vdots & \ddots & \vdots \\ \langle \mathbf{z}_m, \mathbf{z}_2 \rangle & \dots & \langle \mathbf{z}_m, \mathbf{z}_{m+1} \rangle \end{bmatrix}. \tag{4b}$$

The choice of the function that gives the similarity measure determines the class of functions that we use to approximate the transfer operators. Using the standard inner product for a measure of similarity in (9) corresponds to restricting the function space over which we approximate the transfer operator to just linear functions. In this case the Gram matrix contains the same information as the covariance matrix that is used in estimation methods based on autocorrelation like LIM. In fact it can be shown that in this case this approach is equivalent to LIM (Tu 2013; Tu et al. 2014).

### b. A generalization of similarity using kernels

Limiting the approximation of the transfer operators to linear functions is a strong constraint on the approximation. To improve the approximation we have to use a richer set of functions to allow for a better estimation of the transfer operators. This can be achieved by considering a more general similarity measure than the linear inner product, defining a function, usually called *kernel*, such that the similarity between vector  $\mathbf{x}$  and  $\mathbf{y}$  is given by a two-argument function  $k(\mathbf{x}, \mathbf{y})$ . The case in which this function can be expressed as an appropriate inner product of one-argument functions,

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

is particularly important.

Such methods have been extensively investigated in classification and machine learning problems and the properties of several classes of kernels have been identified. The major mathematical result is that the kernels generate a Hilbert function space (of finite or infinite dimension) that enjoys nice properties, the so-called *reproducing kernel Hilbert space* (RKHS).

A full description of kernels and their properties can be found in Schölkopf and Smola (2001) and Steinwart and Christmann (2008). Many possible kernels have been investigated. The linear kernel used in standard correlation turns out to be a special case of the *polynomial* kernel,

$$k(\mathbf{x}, \mathbf{y}) = (\langle \mathbf{x}, \mathbf{y} \rangle + c)^p,$$

where  $p$  is an integer and  $c$  a nonnegative constant. A popular choice is the Gaussian kernel

$$k(\mathbf{x}, \mathbf{y}) = [\exp(-\|\mathbf{x} - \mathbf{y}\|^2/2\sigma^2)] / (\sqrt{2\pi}\sigma),$$

but other kernels are based on exponentials, sigmoid, transcendental, or trigonometric functions. Kernels have been classified as *characteristic kernels* and *universal kernels*; see, e.g., Muandet et al. (2017). A characteristic kernel will preserve the probability distribution of the state space in the RKHS and a universal kernel will generate RKHSs that will be capable to approximate all continuous functions of the state variables. The polynomial kernels are neither of those,

whereas the Gaussian kernel is both universal and characteristic and therefore it is the choice made in this paper.

The choice of the kernel functions is of course of great importance and there is certainly room for further analysis, but for this paper we would like to select a kernel that gives us a good fidelity of the probability aspect of the data and it is amenable also to some analytical calculation. Furthermore, the Gaussian kernel is positive definite and limited between zero and one, whereas other kernels, like the polynomial kernel, require some sort of normalization since they can have very large values.

Similarly important is the choice of the parameters of the kernel. There is no guiding principle for this selection, but in the case of the Gaussian kernel very small or very large values of  $\sigma$  will make the entry in the Gram matrix uniform, destroying the data information. It is clear then that  $\sigma$  must be of the same order as the distances between the data snapshot ( $\|\mathbf{x}-\mathbf{y}\|^2$ ). An empirical choice that is often made is the median of the distribution of the distances (Flaxman et al. 2016), that has been shown to give good results heuristically. We have selected a somewhat similar choice and we have selected  $\sigma$  to be such that the standard deviation of the distribution of the distances is 1. The choice could be refined in specific applications by cross-validation methods to arrive at an optimal, but problem dependent, choice. For the introduction of the method in this paper we have refrained from doing so.

*c. Kernels and the Gram matrix*

Using a kernel the Gram matrix is then given by

$$\mathbf{G}_{zz} = \begin{bmatrix} k(\mathbf{z}_1, \mathbf{z}_1) & \dots & k(\mathbf{z}_1, \mathbf{z}_m) \\ \vdots & \ddots & \vdots \\ k(\mathbf{z}_m, \mathbf{z}_1) & \dots & k(\mathbf{z}_m, \mathbf{z}_m) \end{bmatrix},$$

and  $\mathbf{G}_{yz}$  is defined analogously. The main property of this space is the reproducing property: every function in the RKHS can be calculated using the kernel via the RKHS inner product

$$f(\mathbf{x}) = \langle f(\cdot), k(\cdot, \mathbf{x}) \rangle,$$

where  $\langle \cdot, \cdot \rangle$  is the inner product of the RKHS. We can thus define the canonical feature map  $\phi(\mathbf{x}) = k(\cdot, \mathbf{x})$  and obtain

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

Furthermore, we can define the so-called feature matrix  $\Phi$  by

$$\begin{aligned} \Phi &= [\phi(\mathbf{z}_1), \phi(\mathbf{z}_2), \dots, \phi(\mathbf{z}_m)] \\ &= [k(\cdot, \mathbf{z}_1), k(\cdot, \mathbf{z}_2), \dots, k(\cdot, \mathbf{z}_m)]. \end{aligned}$$

In what follows, we will sometimes use the notation  $\phi_f(\mathbf{x}) = [\phi(\mathbf{z}_j)](\mathbf{x}) = k(\mathbf{x}, \mathbf{z}_j)$ . Every state  $\mathbf{x}$  can be represented in the RKHS by  $\phi(\mathbf{x})$  (Fig. 1). The reproducing property looks like a generalization of the distribution function like the Dirac delta, linking local properties to global integrals.

The matrix eigenvalue problem (8) approximates the corresponding operator eigenvalue problem. The eigenfunctions are then given by

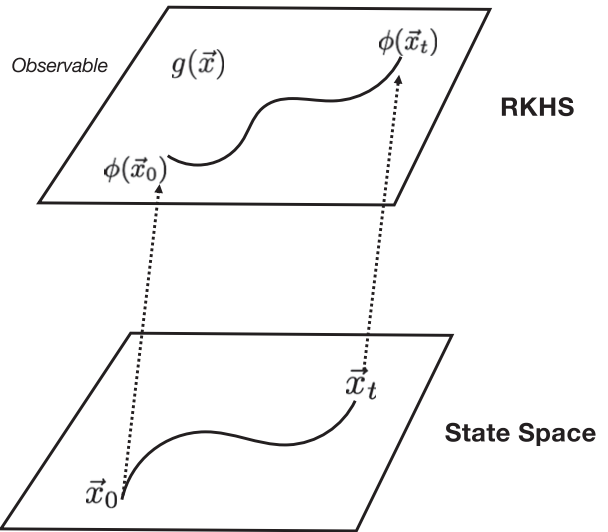


FIG. 1. Schematic of the relations among the state space, the feature mapping, the observables, and the RKHS.

$$\varphi^P = \Phi \mathbf{G}_{zz}^{-1} \mathbf{v}^P = \Phi \mathbf{u}, \quad (\text{Perron-Frobenius})$$

$$\varphi^K = \Phi \mathbf{v}^K, \quad (\text{Koopman})$$

where  $\mathbf{v}$  are the eigenvectors of the auxiliary eigenvalue problem (3). The values at the training data points can then be obtained by evaluating the eigenfunctions in  $\mathbf{z}_1, \dots, \mathbf{z}_m$ . The feature matrix  $\Phi$  evaluated in  $\mathbf{Z}$  becomes the Gram matrix  $\mathbf{G}_{zz}$ .

**4. The spectrum of transfer operators**

The linear nature of the Koopman operator leads to the possibility of its analysis using spectral methods (Rowley et al. 2009; Mezić 2013). We have some freedom in the selection of the function space on which the transfer operators are acting and in what follows we will choose that both the Koopman operator and its adjoint, the Perron-Frobenius operator, will be defined on the Hilbert space of square-integrable functions  $L^2$ . There is a freedom to select the measure for the Hilbert space, for an ergodic and measure-preserving system the invariant measure may be a choice, but others are possible.

The relationship between the eigenvalues  $\mu$  of the Koopman operator for a fixed lag time  $\tau$  and the eigenvalues  $\lambda$  of the generator is given by  $\mu = e^{\lambda\tau}$ . That is, as described above,  $\lambda = \log(\mu)/\tau$ . We can thus make predictions for all possible times  $t$ , not just multiples of  $\tau$ .

As a consequence, every function  $g$  (observable) of the state vector  $\mathbf{x}$  can be written as

$$g(\mathbf{x}, t) = \sum_i \alpha_i e^{\lambda_i t} \varphi_i(\mathbf{x}),$$

yielding the time evolution of the observable, where the  $\alpha_i$  are the coefficients that express the initial value of the observables on the span of the eigenfunctions

$$g(\mathbf{x}_0) = g(\mathbf{x}, 0) = \sum_i \alpha_i \varphi_i(\mathbf{x}).$$

The evolution of probability densities can be obtained from the Perron–Frobenius operator and its eigenfunctions in a similar way:

$$\rho(\mathbf{x}, t) = \sum_i \beta_i e^{\lambda_i t} \varphi_i^P(\mathbf{x}).$$

Once again the  $\beta_i$  are the coefficients of the initial density expanded in the Perron–Frobenius eigenfunctions, obtained from the projection of the initial probability distribution

$$\rho(\mathbf{x}_0) = \rho(\mathbf{x}, 0) = \sum_i \beta_i \varphi_i^P(\mathbf{x}).$$

*a. Observables*

The vector-valued observable defined by  $\mathbf{g}(\mathbf{x}, t) = [g_1(\mathbf{x}), g_2(\mathbf{x}), \dots, g_k(\mathbf{x})]^T$  can be expressed in terms of the first  $N$  Koopman eigenfunction  $\varphi_N$  as

$$\begin{aligned} g_k(\mathbf{x}, t) &= \sum_{i=0}^N \alpha_{ki} e^{\lambda_i t} \varphi_i(\mathbf{x}) \\ &= \sum_i \alpha_{ki} e^{\lambda_i t} \sum_j v_{ij} \phi_j(\mathbf{x}) \\ &= \sum_{ij} \alpha_{ki} e^{\lambda_i t} v_{ij} \phi_j(\mathbf{x}) \end{aligned}$$

because the eigenfunctions can be written in terms of the features as

$$\varphi_l(\mathbf{x}) = \sum_{j=0}^M v_{lj} \phi_j(\mathbf{x}),$$

where  $v_{lj}$  is the  $j$ th component of the  $l$ th eigenvector of the empirical estimates.

Introducing the matrix  $\mathbf{A} = \alpha_{ij}$ , the diagonal matrix  $\mathbf{D}(t)$  having as diagonal elements  $[e^{\lambda_0 t}, e^{\lambda_1 t}, \dots, e^{\lambda_N t}]$ , and letting  $\varphi_N$  be the row vector of the first  $N$  Koopman eigenfunctions, we can write the evolution of the coefficients of the observables in matrix form as

$$\mathbf{g}(\mathbf{x}, t) = \mathbf{A} \mathbf{D}(t) \varphi_N^T(\mathbf{x}).$$

On the other hand the eigenfunctions can be expressed in terms of the features, where  $\mathbf{V}_N$  is the matrix of the  $v_{ij}$  coefficients restricted to the retained eigenfunctions

$$\tilde{\varphi} = [\varphi_1, \varphi_2, \dots, \varphi_N] = \Phi \mathbf{V}_N \tag{5}$$

to obtain

$$\mathbf{g}(\cdot, t) = \mathbf{A} \mathbf{D}(t) \mathbf{V}_N^T \Phi^T. \tag{6}$$

The coefficient matrix  $\mathbf{A}$  can then be obtained from (6) at  $t = 0$ . Evaluating  $\mathbf{g}$  in all training data points  $\mathbf{Z}$ , we obtain

$$\mathbf{g}(\mathbf{Z}) = \mathbf{A} \mathbf{V}_N^T \mathbf{G}_{zz},$$

so we have

$$\mathbf{A} = \mathbf{g}(\mathbf{Z}) (\mathbf{V}_N^T \mathbf{G}_{zz})^{-1},$$

where the matrix  $\mathbf{g}(\mathbf{Z})$  is defined as the values of the observable at every data point,  $\mathbf{g}(\mathbf{Z}) = [\mathbf{g}(z_1), \mathbf{g}(z_2), \dots, \mathbf{g}(z_m)]$ .

*b. Probability*

The probability density of the states can be expanded in the Perron–Frobenius eigenfunctions  $\varphi^P$  in a similar way, taking now into account the fact that it is an scalar function and not necessarily a vector observable as before. The probability  $\rho(\mathbf{x}, t)$  can then be expressed in terms of the first  $N$  Perron–Frobenius eigenfunctions. The approximation cannot guarantee a positive definite distribution at all truncations, but it is achieved in the limit  $N \rightarrow \infty$ . We organize the row vector of functions  $\tilde{\varphi}$  as in the previous case (15),

$$\tilde{\varphi}^P = [\varphi_1, \varphi_2, \dots, \varphi_N] = \Phi \mathbf{V}_N^P,$$

then we can set

$$\rho(\mathbf{x}, t) = \sum_{i=0}^N \beta_i e^{\lambda_i t} \varphi_i^P(\mathbf{x}), \tag{7}$$

and using the column vector  $\mathbf{B}$  that contains as elements the coefficients  $\beta_i$ , the diagonal matrix of the eigenvalues  $\mathbf{D}(t)$  as above, and  $\varphi_N^P$  the row vector of the first  $N$  Perron–Frobenius eigenfunctions, we can write the evolution of the coefficients of the probability distribution in matrix form as

$$\rho(\mathbf{x}, t) = \mathbf{B}^T \mathbf{D}(t) (\tilde{\varphi}^P)^T = \mathbf{B}^T \mathbf{D}(t) (\mathbf{V}_N^P)^T \Phi^T.$$

If we project (20) at  $t = 0$  on the Perron–Frobenius eigenfunctions  $\varphi_i^P(\mathbf{x})$  one at the time we get a set of equations that can be used to obtain the expansion coefficients  $\beta_i$  of the initial condition,

$$\rho(\mathbf{x}, t) \varphi_k^P(\mathbf{x}) = \sum_{i=0}^N \beta_i e^{\lambda_i t} \varphi_k^P(\mathbf{x}) \varphi_i^P(\mathbf{x}), \quad k = 1, 2, \dots, N. \tag{8}$$

Let  $\rho(x, 0)$  denote the probability at time  $t = 0$ . The vector  $\mathbf{B}$  can then be obtained from (8), calculating the values of the initial probability distribution on the data points  $\mathbf{Z}$ , i.e.

$$\rho(\mathbf{Z}, 0) \mathbf{G}_{zz} \mathbf{V}_N^P = \mathbf{B}^T (\mathbf{V}_N^P)^T \mathbf{G}_{zz} \mathbf{V}_N^P,$$

where  $\mathbf{G}_{zz}$  is the Gram matrix of the data. So we have

$$\mathbf{B}^T = \rho(\mathbf{Z}, 0) \mathbf{G}_{zz} \mathbf{V}_N^P [(\mathbf{V}_N^P)^T \mathbf{G}_{zz} \mathbf{V}_N^P]^{-1}.$$

The expectation value with respect to the probability density can then be written as

$$E[f(\mathbf{x})] = \frac{1}{R(t)} \int f(\mathbf{x}) \rho(\mathbf{x}, t) d\mathbf{x},$$

where the integration extends over all state variables,  $d\mathbf{x} = dx_1, dx_2, \dots$  and  $R(t)$  is a normalization factor for the (unnormalized) probability density that will be time dependent for systems converging to a stationary state, given by

$$R(t) = \int \rho(\mathbf{x}, t) d\mathbf{x}.$$

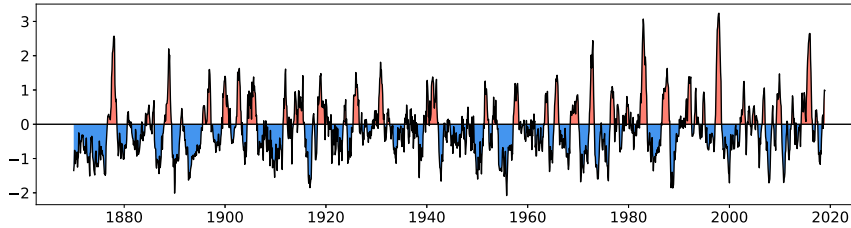


FIG. 2. Niño-3 time series.

The  $\mathbf{B}$  vector we can now be used together with (7) to write the expectation values of any function at any time

$$E[f(\mathbf{x})] = \frac{1}{R(t)} \int \mathbf{B}^T \mathbf{D}(t) (\mathbf{V}_N^p)^T f(\mathbf{x}) \Phi(\mathbf{x})^T d\mathbf{x}.$$

Because  $f(\mathbf{x})$  is a scalar, we can transform it into

$$\begin{aligned} E[f(\mathbf{x})] &= \frac{1}{R(t)} \mathbf{Q}(t) \int f(\mathbf{x}) \Phi(\mathbf{x})^T d\mathbf{x} \\ &= \frac{1}{R(t)} \mathbf{Q}(t) \mathbf{S}, \end{aligned} \tag{9}$$

where the vector  $\mathbf{Q}(t)$  is defined as

$$\mathbf{Q}(t) = \mathbf{B}^T \mathbf{D}(t) (\mathbf{V}_N^p)^T$$

and we have introduced a column vector  $\mathbf{S}$ , the *structure vector*, as

$$\mathbf{S}_i[f] = \int f(\mathbf{x}) \phi_i(\mathbf{x}) d\mathbf{x}. \tag{10}$$

The vector  $\mathbf{Q}$  contains the information coming from the dynamics as it is represented from the data. The normalization factor can then be obtained from

$$R(t) = \mathbf{Q}(t) \int \Phi(\mathbf{x}) d\mathbf{x}.$$

*c. The structure matrix for the Gaussian kernel*

For the Gaussian kernel the structure vector can be computed explicitly. In this case the features are

$$\phi_i(\mathbf{x}) = \exp[-(\mathbf{x} - \mathbf{z}_i)^2 / (2\sigma^2)] / (\sqrt{2\pi} \sigma),$$

so for the expected value of the  $k$ th component of the state vector  $\mathbf{x}^k$ ,  $\mathbf{S}$  is given by

$$\mathbf{S}_i[x^k] = \int x^k \exp[-(\mathbf{x} - \mathbf{z}_i)^2 / (2\sigma^2)] / (\sqrt{2\pi} \sigma) d\mathbf{x} = \mathbf{z}_i^k$$

and similarly for the variance  $\langle (x^k)^2 \rangle$  then

$$\mathbf{S}_i[(x^k)^2] = (\mathbf{z}_i^k)^2 + 2\sigma^2$$

and the cross-variances,  $\langle x^k x^l \rangle$

$$\mathbf{S}_i[x^k x^l] = \mathbf{z}_i^k \mathbf{z}_i^l.$$

From the matrix  $\mathbf{S}$  we can derive the evolution of expected values of the state vector.

**5. Application to the one-dimensional Niño-3 time series**

We start with the simplest example that consists of a one-dimensional monthly means time series, in this case the Niño-3 index time series<sup>1</sup> based on data from Rayner et al. (2003); see Fig. 2. The data are anomaly monthly means values from January 1870 to December 2018 for a total of 1788 data points. In this case the vector entries in the data matrix are just numbers so the data matrix  $\mathbf{Z}$  itself is a one-dimensional vector,

$$\mathbf{Z}_{\text{NINO3}} = [z_1, z_2, \dots, z_m].$$

In this case the zero lag and one-lag time covariance that are the basic blocks of a LIM approach are numbers, but on the other hand the spatial Gram matrices will be  $1788 \times 1788$  and its elements are the similarity between every monthly anomaly with all the others. Now, if we choose as measure of similarity in the Gram matrices (4a) and (4b) the inner product in state space (in this case it is just the product of every monthly anomaly with every other monthly anomaly) the approximated Koopman operator eigenvalue problem (3) will have only one eigenvalue whose value will be proportional to the one-lag covariance.

On the other hand, if we use a nonlinear measure for the similarity, using one of the kernels described in section 4b then the eigenvalue problem for the Koopman or Perron–Frobenius operators will have in general many different eigenvalues and eigenfunctions, in fact for the Gaussian kernel it is usually of full order. The consequence is that now we have many more functions available for the approximation of the operators opening up the possibility of improving the approximation itself.

The interpretation is that in the linear similarity case the approximation obtained to the Koopman operator is essentially the same as in a LIM approach, providing here another interpretation of the LIM procedure as an attempt to get an approximation to the Koopman operator using only linear functions. Note that the Koopman framework does not need to assume the presence of a stochastic component, even if stochastic system can also be treated within the Koopman approach.

<sup>1</sup> [https://www.esrl.noaa.gov/psd/gcos\\_wgsp/Timeseries/Nino3](https://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/Nino3).

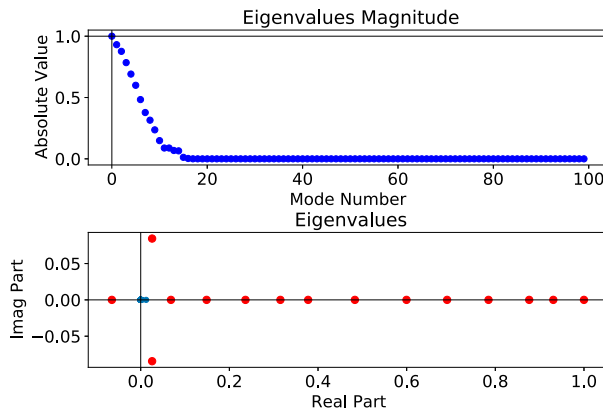


FIG. 3. (top) Magnitude and (bottom) real and imaginary part of the eigenvalues of the Perron–Frobenius operator for the Niño-3 monthly time series.

Then we can see that solving the eigenvalue problem (3) is a similar step to the determination of the empirical normal modes of Penland and Sardeshmukh (1995) from the time lagged covariance matrix appropriately scaled. When we use a Gaussian kernel with bandwidth  $\sigma = 0.5$ , the eigenvalue problems for Koopman and Perron–Frobenius operators become nontrivial and we get many eigenvalues shown in Fig. 3. There is an eigenvalue of magnitude one corresponding to the invariant density and all other eigenvalues are smaller than one, this implies that there is a stationary state corresponding to eigenvalue of unitary size and the remaining eigenvalues are describing decaying eigenfunctions toward that state. We can also see that there are about only 20 eigenvalues large enough to contribute to the time evolution, as most of the others are really numerically zero. The figure also shows the position of the eigenvalues in the complex plane. Most of eigenvalues are real, but there a few that have nonzero imaginary part, indicating an oscillatory component. The  $\mathbf{G}$  matrices are real, so complex eigenvalues and their relative eigenfunctions come in conjugate pairs.

In this one-dimensional case we can compute explicitly the eigenfunctions of the Koopman and Perron–Frobenius operators. The eigenfunctions are going to be weighted by the empirical probability density of the data (Fig. 4). Figure 5 shows the eigenfunctions. The probability distribution relaxes to the lowest eigenfunctions as time progresses, indicating that this eigenfunction is a *stationary state* for this system.

The positive values (Fig. 6) show a similar behavior. In general the stationary state is reached between 6 and 12 months depending on the initial condition, at that point every memory of the initial condition is lost and the probability distribution cannot be distinguished from the average value over the history of the time series. We can see then the transfer operators, in this case the Perron–Frobenius operator, provide another estimation of the predictability limit for the equatorial sea surface temperature (SST) as expressed by the Niño-3 index. The value, between 6 and 12 months, is consistent with estimates from seasonal forecasting systems and other empirical estimates.

The other transfer operator, the Koopman operator, on the other hand, can be used to predict the evolution of observables

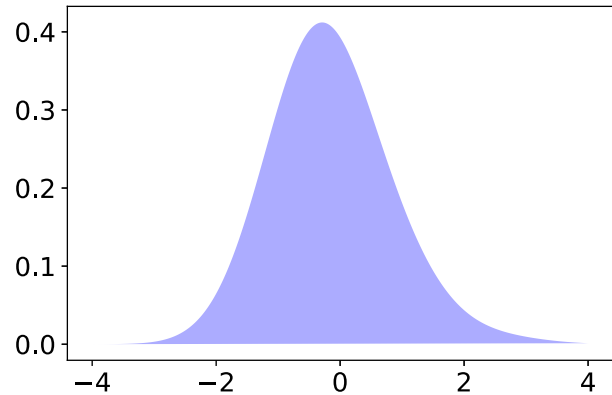


FIG. 4. Empirical probability density for Niño-3. It has been estimated using a kernel density estimator from Scikit-Learn (Pedregosa et al. 2011).

and, of course, of the simplest observable, the state vector itself, in this case the value of the Niño-3 index. We show in Fig. 7 the forecast of the Niño-3 index from various starting points using the Koopman eigenfunctions. For comparison, it is shown the autoregression forecast that can be considered as the simplest approximation of the Koopman operator, namely, approximating it with just linear functions. Improving the approximation with a larger class of functions by use of the kernel yields a richer behavior.

## 6. The Koopman operator for the Pacific SST

We describe as an example in this section the application of transfer operator theory to the evolution of the equatorial Pacific SST. The SST data are obtained from ERA5 (Copernicus Climate Change Service 2017).<sup>2</sup>

The dataset is composed of anomaly monthly means fields from January 1979 to December 2018 for a total of 468 snapshots, normalized by the total standard deviation. The anomalies have been computed with respect to the month by month climatology obtained from the entire time series 1979–2018 and no other preprocessing has been applied, i.e., no detrending has been performed. The resolution of  $0.25^\circ$  translates into 67 796 ocean grid points, taking into account the land-sea mask. Therefore in principle every grid point represents a degree of freedom, showing a typical high-dimensional problem. In more precise terms the issue here is that we are observing the variability of a field that has an infinite number of degrees of freedom. Another way of saying this is that we need an infinite set of numbers to exactly specify a configuration of the SST field, in a real application we realize a discretization that approximates the field with a finite number of points. We can organize the data according to (2), obtaining an array of size  $67\,796 \times 468$  and carry out the calculation as described in

<sup>2</sup>They cover the equatorial Pacific zone with a resolution of  $0.25^\circ$ . The selected region extends from  $15^\circ\text{N}$  to  $15^\circ\text{S}$  and from  $40^\circ\text{E}$  to  $110^\circ\text{W}$ .



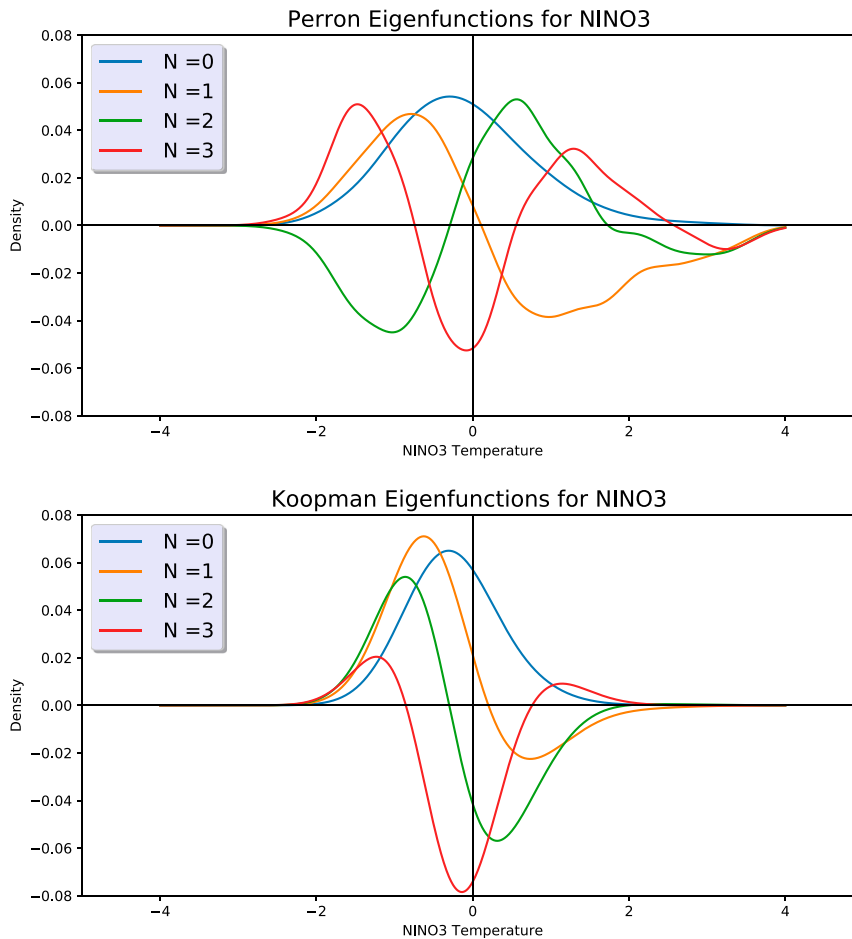


FIG. 5. First four (top) Perron–Frobenius eigenfunctions and (bottom) Koopman eigenfunctions for the Niño-3 time series. The Koopman eigenfunctions have been reweighted by  $\sqrt{\rho_E}$  for convenience, where  $\rho_E$  denotes the density of the time series data (Fig. 2) computed by a kernel density estimation; see Fig. 4. The asymptotic state ( $N = 0$ ) is the lowest one with eigenvalue 1; the other modes are progressively decaying at higher rates.

the previous sections. Note that because the algorithm is using Gram matrices rather than covariance matrices the calculation is feasible.

For computational convenience we have done a preliminary EOF analysis on the anomaly data, keeping a smaller number of EOF modes. The calculations below have been performed with 31 modes, retaining 92% of the variance. After this transformation the state vectors  $\mathbf{x}$  consist of the coefficients of the EOF for every monthly mean anomaly. There is little change in keeping more modes. This reduction is not essential to the calculation and the algorithms work fine even using the full original data in the gridpoint representation or using the entire spectrum of 468 EOFs.

The spectrum of the transfer operator is shown in Fig. 8. We have used here a Gaussian kernel (left panel) with a bandwidth determined from the standard deviation of the distribution of squared norms of the data vectors' Euclidean distances  $\|\mathbf{z}_i - \mathbf{z}_j\|^2$ . The distribution of the eigenvalues of the spectrum indicates that the approximated transfer operator

is almost unitary, except for a few eigenvalues with norm smaller than one inside the unit circle. In contrast, the usage of a polynomial kernel of order one (on the left), that corresponds to using the ordinary covariance matrix, shows a much poorer approximation. It is interesting to note that in the polynomial case only 31 eigenvalues are different from zero, that corresponds to the maximum number of degrees of freedom of the covariance matrix, i.e., the number of EOF modes retained.

It must be clarified that we cannot plot the eigenfunctions themselves, as they are high-dimensional functions and in effect what we are plotting here are the values that the respective eigenfunctions take on the data points. The eigenvalues are located on the unit circle, but we can order them by their respective period, from the slowest to the fastest. Most of the eigenvalues have small growth/decay rates and it is not possible to identify a definite ground state as in the case of the Niño-3 index. There are eigenfunctions both growing and decaying and so in general there is no asymptotic state to which

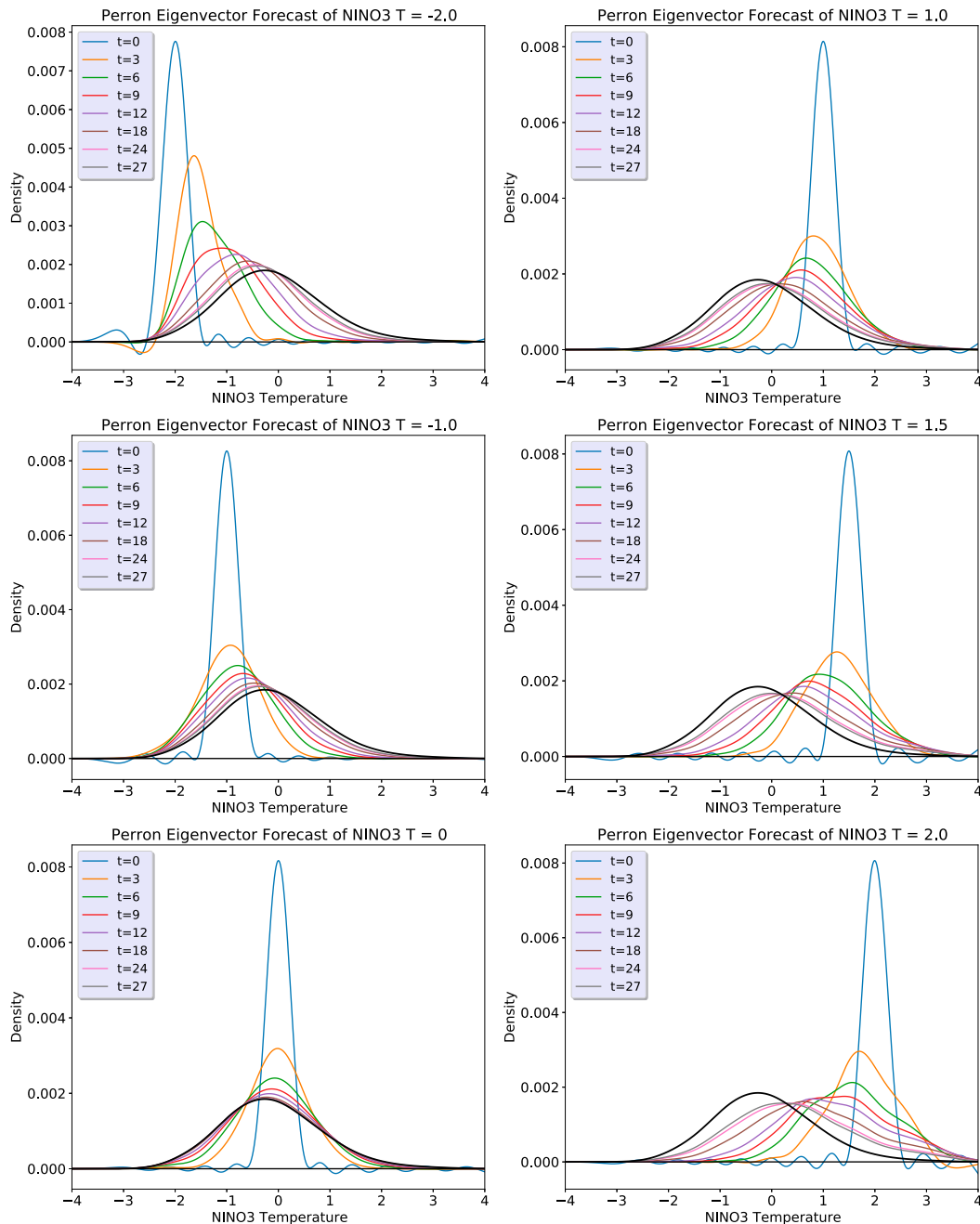


FIG. 6. Evolution of the probability for the Niño-3 index time series for various initial conditions of negative anomalies. The panels show the evolution starting at values of the index equal to (bottom left) 0.0, (middle left)  $-1$ , (top left)  $-2.0$ , (bottom right)  $2.0$ , (middle right)  $1.5$ , (top right)  $1.0$ . The time units in the legend are months.

they relax with time. Figure 9 shows some examples of eigenfunctions. We have plotted here the values of the eigenfunctions at the data points, because of their complex nature we are plotting the amplitude. We can interpret the stable ground state at zero growth rate ( $N = 0$ ) as a normal state and we can notice large deviations corresponding to years of large anomalies. The corresponding eigenfunctions

are a very slow, trend-like evolution of the system, until we reach the sixth or seventh eigenfunctions, with periods almost decadal, where we can notice stable fluctuations of decreasing time scale. These are nonlinear fluctuations, with sharp transitions between states, very different from simple oscillations. Higher eigenfunctions ( $N > 20$ ) result in fast time scales at annual or biannual scale.

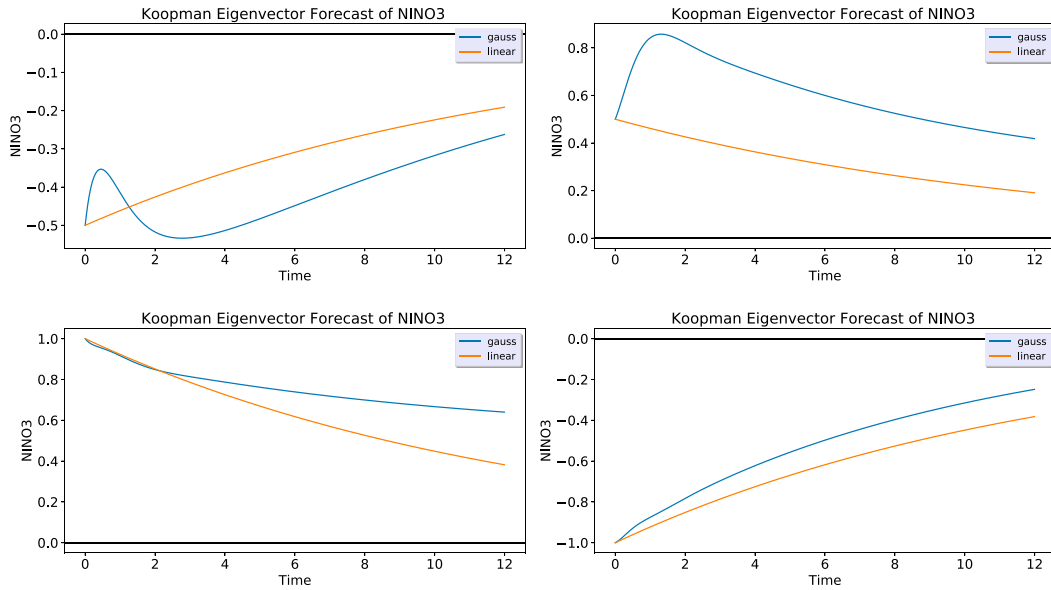


FIG. 7. Forecast of the Niño-3 index time series for various initial conditions. The Koopman forecast and the forecast based on the autocorrelation of the time series are shown. Time units are months.

According to the analysis of [Giannakis \(2019\)](#), the absence of a realization to the ground state seems to indicate that the system is ergodic. It is interesting to compare with the case of the Niño-3 index that was clearly showing instead a dissipative nature. It is reasonable to assume that the full multidimensional field is the main physical system we are considering, so we have to be careful to conclude from low-dimensional slices of the field properties of the entire field.

A linear predictor can be constructed from the Koopman operator eigenfunctions ([Korda and Mezić 2019](#)), by considering the state itself to be an observable expressed as a linear combination of the Koopman eigenfunctions. For a prediction starting in January of a given year, the Koopman eigenfunctions have been calculated using the data up to the preceding December. The verification data have also been projected on the different EOF sets for each starting date. The EOFs have been also calculated only in this selected

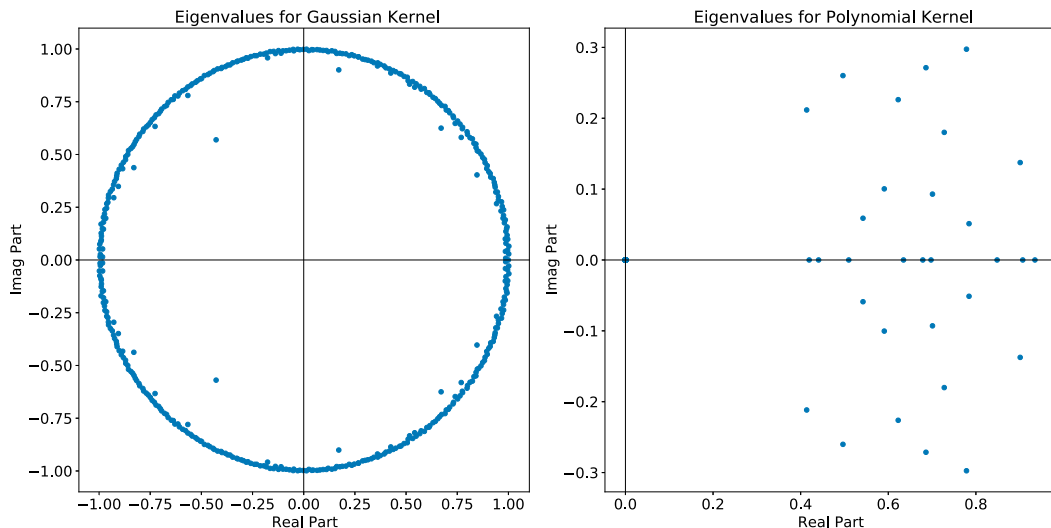


FIG. 8. Eigenvalues of the Koopman operator for (left) a Gaussian kernel and (right) a polynomial kernel of order one with no shift. The eigenvalues within the circle have norm smaller than unity. In the polynomial kernel case there are only 31 eigenvalues different from zero, as most of them are concentrated in the origin.

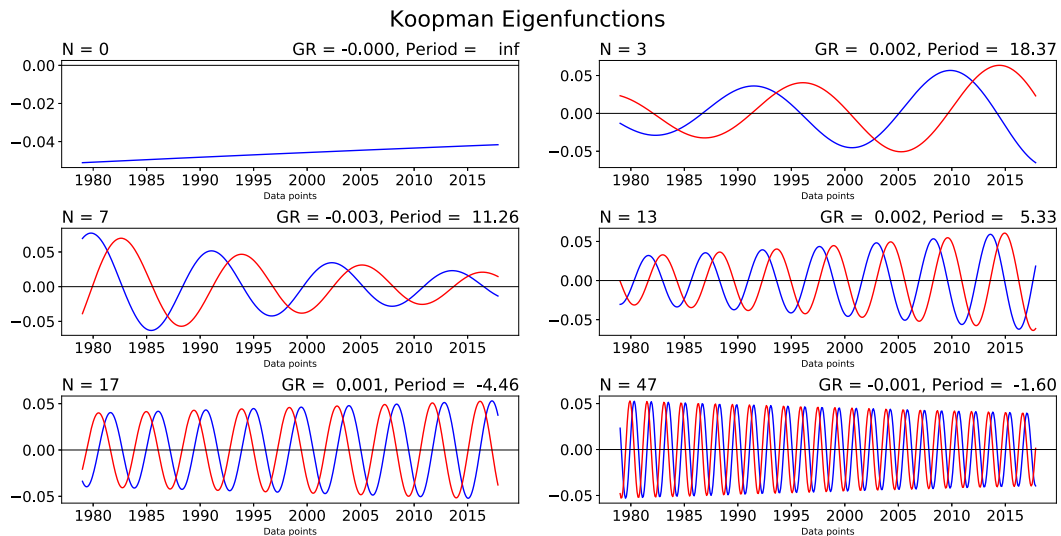


FIG. 9. Real and imaginary part of the Koopman eigenfunctions at the data points.

period. We have retained 31 EOF modes, corresponding to about 92% of the variance. The initial conditions are then obtained by expanding the initial state on the resulting Koopman eigenfunctions using the expansion described in section 4a. We use the spatial correlation coefficient over the entire area as a verification measure. The verification data are projected on the same EOFs obtained from the training period.

The results are shown in Figs. 10–13 for a number of selected cases. There is some freedom in selecting the truncation limit for the expansion in the nonlinear Gaussian kernel case where we get a large number of eigenfunctions. Because we can order the eigenfunctions according to their time scale we have selected the truncation based on the last time period retained in the eigenfunctions. For the case of

**January Cases**

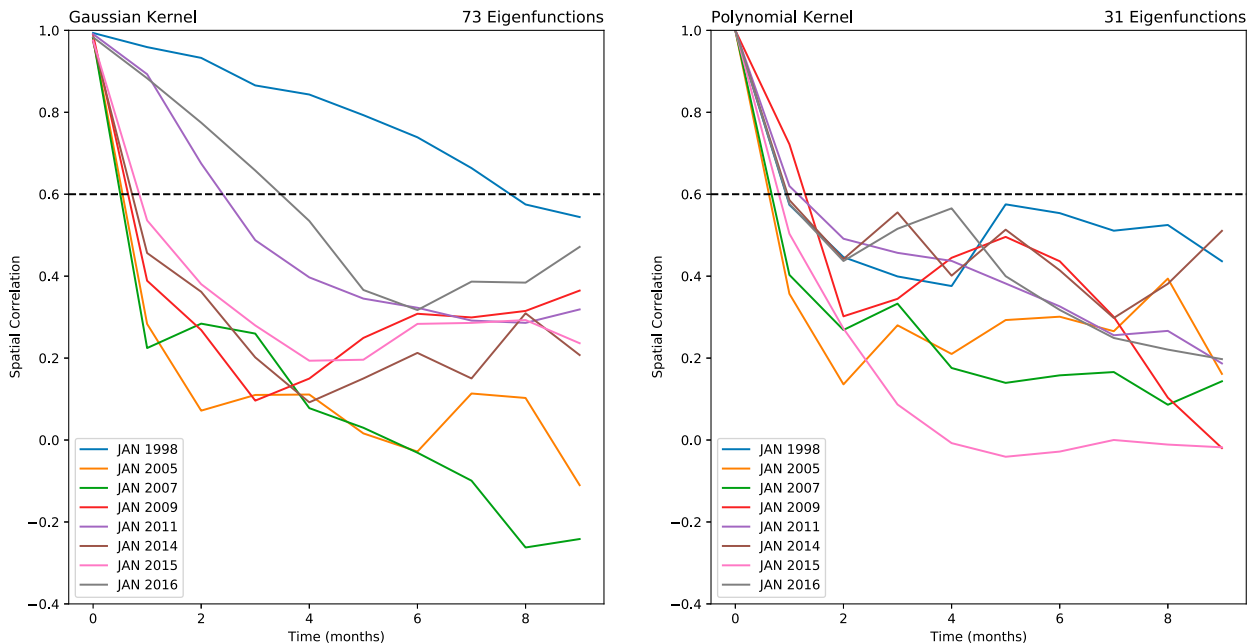


FIG. 10. Anomaly correlation coefficients between the observed SST and the predicted SST reproduced via a linear expansion on the Koopman eigenfunctions. Here they are shown for January initial-condition cases. (left) The Gaussian kernel; (right) a polynomial kernel equivalent to a standard covariance matrix.

April Cases

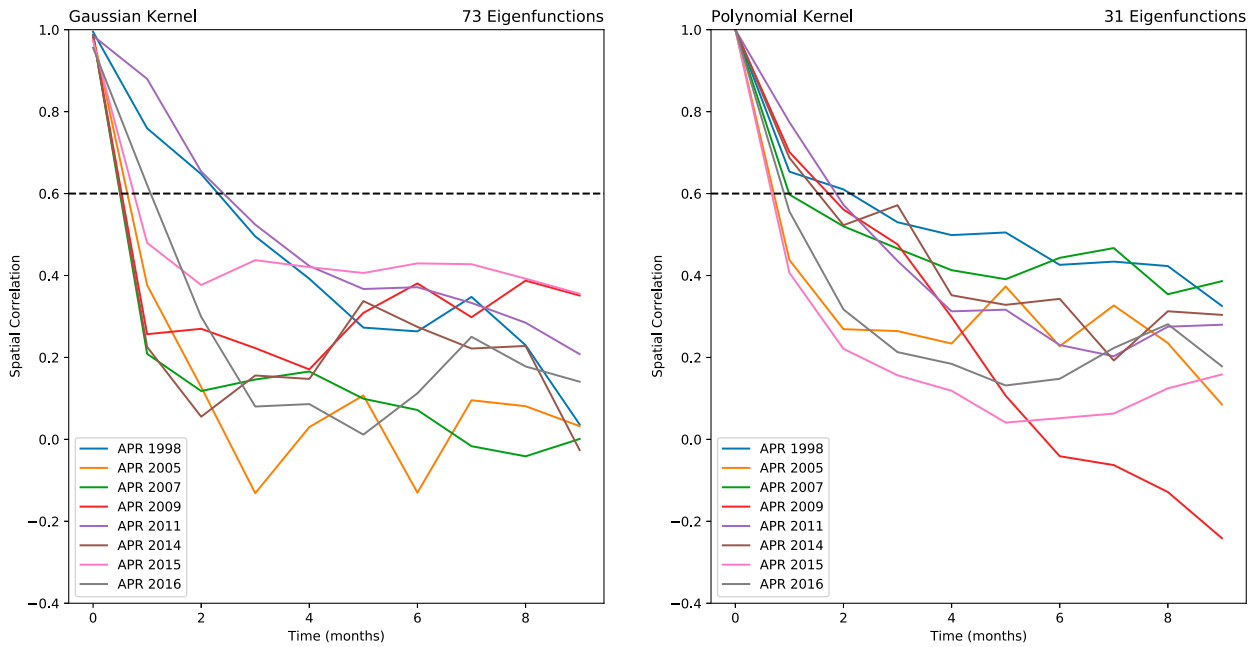


FIG. 11. As in Fig. 10, but for April starting dates.

polynomial linear kernel, essentially similar to a linear inverse model, we have kept all the available eigenfunctions that correspond to the number of SVD modes retained. A number of selected cases have been chosen from the late

1990s onward. The dashed lines indicate the skill for a persistence forecast. For the sake of clarity, we are showing only two cases, but they are fairly representative of the behavior of other cases. In general there is a gain of predictability when the

January Cases

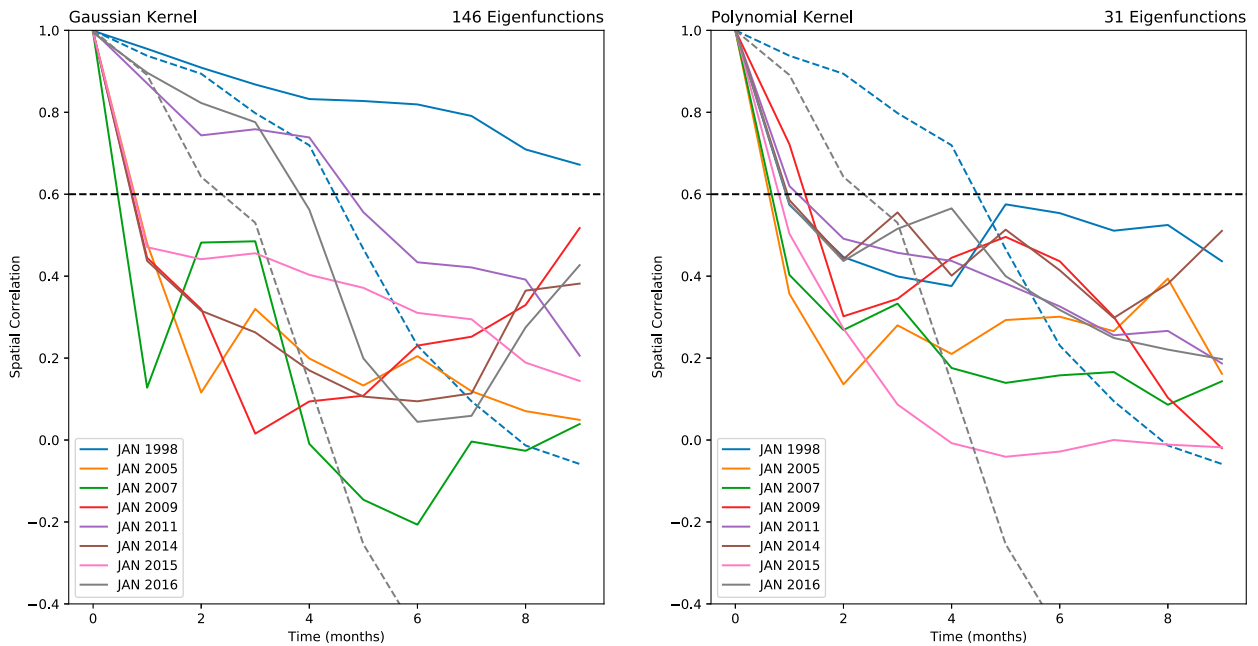


FIG. 12. As in Fig. 10, but extending the truncation in the nonlinear kernel case to 146 eigenfunctions, corresponding to periods up to 3 months. The linear panel is the same as in the preceding figure, reproduced for convenience. The dashed lines indicate the persistence skill score for selected cases.

April Cases

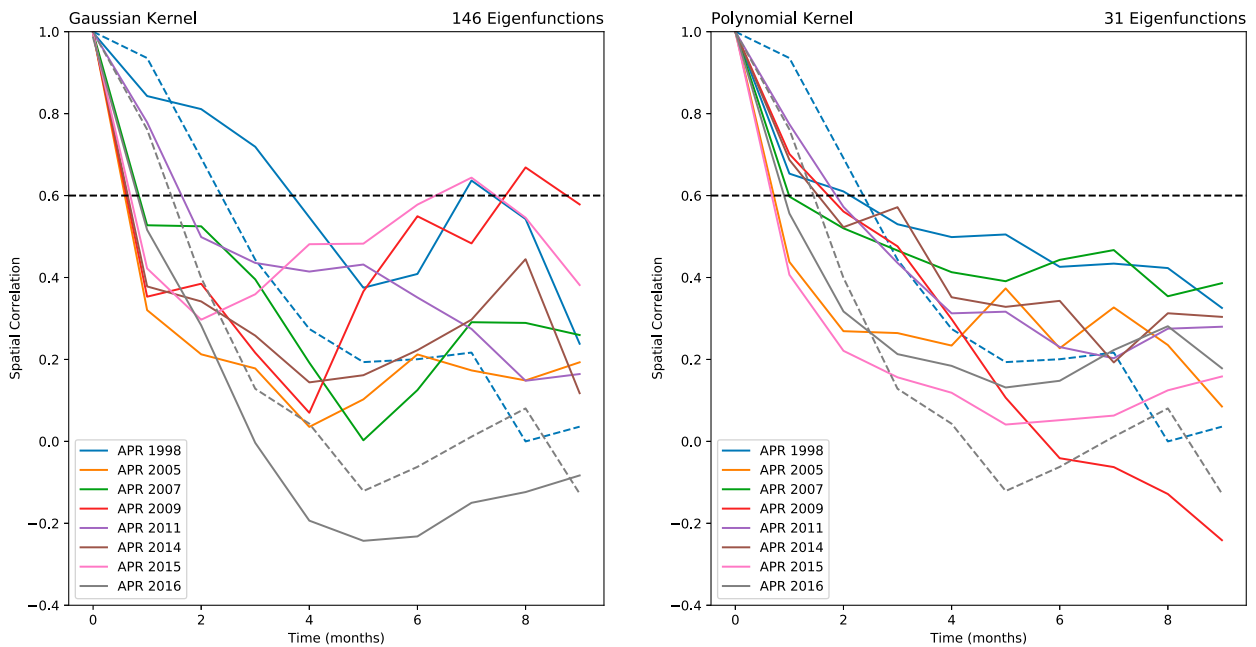


FIG. 13. As in Fig. 12, but for April starting dates.

forecast has skill. This is just a preliminary illustration of the capability of the method, since a full analysis of the performance and design of a predictive system based on the Koopman eigenfunctions is beyond the scope of the present paper, but we think it is useful to give an initial flavor of the potential of the method.

Figure 10 shows the results for the January start dates. The case for a linear polynomial kernel (right panel in Fig. 10) corresponds to a LIM where we have retained all the 31

eigenfunctions that correspond to nonzero eigenvalues. We can see that in general the Gaussian kernel is yielding a better reproducibility of the evolution of the state relative to the linear case. In a few cases the reproducibility is quite significant up to 6–8 months. Some hints of the seasonal dependence can be obtained from Fig. 11 that shows similar results for the April start dates. This is a more difficult case compared to the January and we are losing some predictability.

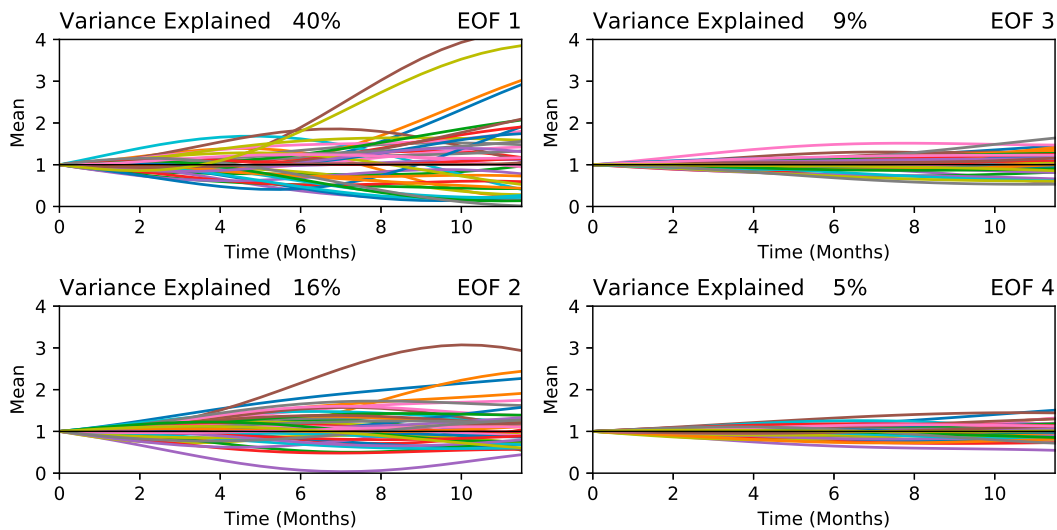


FIG. 14. The evolution of the variance of the some of the EOFs. The variance is normalized with respect to the initial variance.

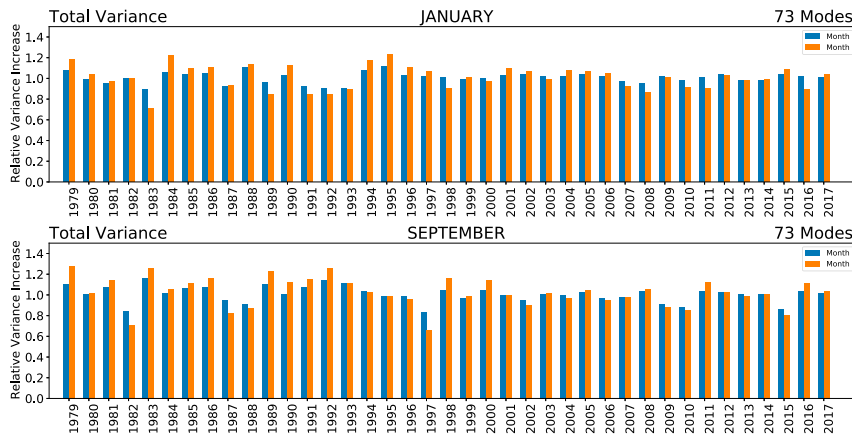


FIG. 15. Total variance for each starting date in (top) January and (bottom) September after 3 months (blue bars) and 6 months (orange bars).

Using a nonlinear kernel produces a larger space for the estimation of Koopman operator, than in the linear case. In the linear case we are limited by the number of SVD modes retained and we have already used the maximum number possible in the Figs. 10 and 11, but in the case of the nonlinear kernel we can use a larger number of eigenfunctions for our predictions. Figure 12 shows the case where we have retained 146 eigenfunctions in the nonlinear case, corresponding to periods up to 2–3 months. We note some improvement for the prediction, both for January and April dates. The improvements are not uniform and they are not present in some cases, interestingly in the April case there is a sort of improvement at month 6–7 of the forecasts.

**7. The Perron–Frobenius operator and the probability of the Pacific SST**

The evolution of the probability can be obtained from the Perron–Frobenius operator eigenfunctions. The initial probability is assumed to be a Gaussian centered at the monthly value of the SST:

$$\rho_i(\mathbf{x}, 0) = \frac{1}{\sqrt{2\pi}\delta} \exp \left[ -\frac{(\mathbf{x} - \mathbf{z}_i)^2}{2\delta^2} \right],$$

where  $\mathbf{z}_i$  is the  $i$ th initial condition, in this case assumed to be the monthly mean for that month, and  $\delta$  is a measure of the uncertainty of the initial condition, assumed here constant for all EOF components for simplicity.

The expected values of observables can then be obtained easily from the results of section 4b. Using (28) we can compute the variance for a single component using the structure matrix in (30) during the evolution for each initial condition. Figure 14 shows such an evolution for four EOF components separately for each January initial condition. We can see how the variance for each component grows with time, in some cases the amplification of the variance is very large, indicating an initial condition that tends to generate a large amplification of perturbations.

We can examine now the total variance, defined as the sum of the variances of each component EOF, defined as expectation values

$$\text{Var}_{\text{SST}} = \sum_k E[(x^k)^2]$$

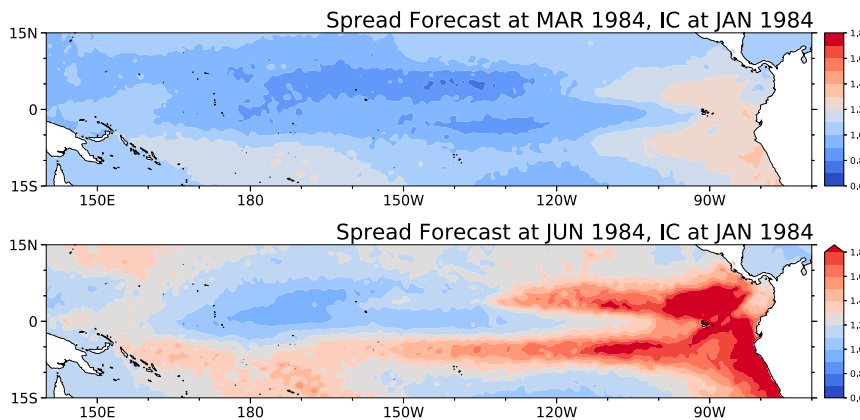


FIG. 16. Ratio of total variance to initial variance starting from January 1984 after (top) 3 and (bottom) 6 months.

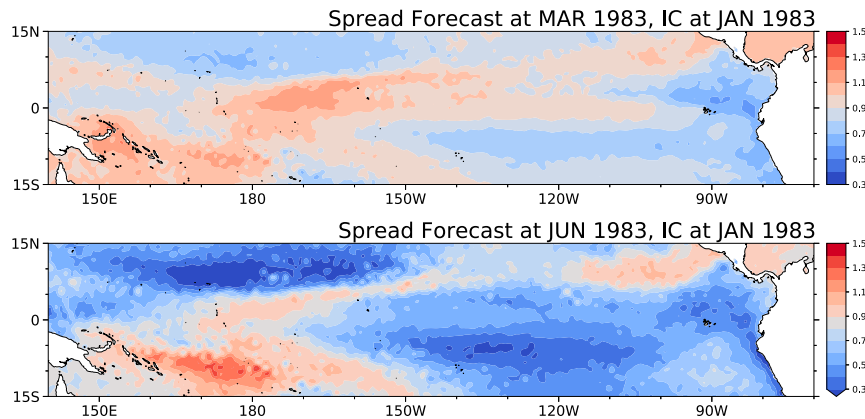


FIG. 17. Ratio of total variance to initial variance starting from January 1983 after (top) 3 and (bottom) 6 months.

for each initial condition. Figure 15 shows the variance after 3 and 6 months for each starting date in detail. There is a seasonal difference between January and September. The inter-annual variability in the evolution of the total variance is greater in September than in January. Certain years show a visible reduction of the total variance at month 6, after an initial increase at month 3, such as 1997 and 1982.

In this analysis we have shown the variance using the EOF coefficients, but we can also use the Perron–Frobenius operator to obtain the expectation value of the variance point by point. The expected value on the grid can be obtained from the representation of the data in terms of the EOFs as

$$\text{SST}(\text{lon}, \text{lat}, t) = \sum_i x_i(t) \phi_i(\text{lon}, \text{lat}),$$

where  $(\text{lon}, \text{lat})$  are longitude and latitude, the  $x_i(t)$  the EOF coefficients and  $\phi_i$  are the EOF patterns. Then the expectation value of the variance at the point  $\text{SST}(\text{lon}, \text{lat}, t)$  is given by

$$E[\text{SST}(\text{lon}, \text{lat}, t)] \\ E \left[ \sum_i x_i(t) \phi_i(\text{lon}, \text{lat}) \sum_j x_j(t) \phi_j(\text{lon}, \text{lat}) \right] = \\ \sum_i E[x_i(t)x_j(t)] \phi_i(\text{lon}, \text{lat}) \phi_j(\text{lon}, \text{lat}),$$

where the  $E[x_i(t)x_j(t)]$  are the expectation values of the covariances of the  $(i, j)$ th EOF coefficients at time  $t$ .

We can therefore look at the geographical distribution of the total variance in time starting from different initial conditions. Figure 16 shows such a ratio for January 1984. This is one of the states that has one of the largest growth of the total variance in general, we can see here that after 6 months the integrated variance has increased 20%. Regional differences are stronger. In the east Pacific, north and south of the equator the local variance at later months is almost double the initial variance. After 6 months a clear bipolar pattern has emerged with the maximum amplification in the east and indeed a smaller variance in the west Pacific.

The situation is different for January 1983 (Fig. 17). In this case we have a relatively weak initial amplification of the variance in the west Pacific, but already at month 3 a tendency for a decreasing variance in the east appears. It becomes very pronounced at month 6 with a drastic decrease of the variance in the central Pacific south of the equator.

## 8. Conclusions

We have presented here some examples of the application of Koopman methods to atmospheric and climate data that show very interesting potential. A complete physical interpretation of these results requires further investigation, but it is possible to say at this point that the Perron–Frobenius modes are sensitive enough to identify regimes and/or states of active dynamics. The Koopman approach also makes it possible to estimate the transfer operators for complex systems extending Hilbert space methods to this area. In particular, it is possible to estimate empirically the evolution equation for the probability distribution even for complex systems. There is also a potential for using these techniques for establishing the dissipative and/or conservative character of a physical system based on data, either from observation or numerical simulations, offering a new approach to the classification of such systems. These methods are based on the Gram matrices rather than on covariance matrices as in many other prior studies, and as such are applicable also to large high-dimensional datasets. The potential that these methods give to estimate the evolution of the covariance matrix without the recourse to the calculation of ensembles of trajectories is particularly interesting. We are currently looking in another paper to the application of these ideas to data assimilation problems.

*Acknowledgments.* We gratefully recognize the partial support of the EU project EUCP 776613. The National Center for Atmospheric Research is supported by the U.S. National Science Foundation.

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



## APPENDIX

## Notation and Definitions Used

$x \in \mathbb{R}^n$	State of the system
$F$	Right-hand side of the ODE
$k, \phi$	Kernel and associated feature map
$\Phi$	Feature matrix
$U_t$	Evolution operator to time $t$ associated with $F$
$\mathcal{K}$	Koopman operator
$\mathcal{P}$	Perron–Frobenius operator
$\mathcal{L}, \mathcal{L}^\dagger$	Koopman generator and its adjoint
$\mathbf{G}_{xx}$	Gram matrices for data matrices $\mathbf{X}$
$\mathbf{G}_{xy}$	Gram matrices for data matrices $\mathbf{X}$ and $\mathbf{Y}$
$\varphi$	Eigenfunction of the Koopman operator
$\varphi^{\mathcal{P}}$	Eigenfunction of the Perron–Frobenius operator
$\mu$	Eigenvalue of the Koopman or Perron–Frobenius operator
$\lambda$	Eigenvalue of the generators of the operators
$\rho_E(\mathbf{x})$	Empirical probability distribution of Niño-3 anomalies
$\mathbf{S}_i[f]$	Structure vector for the expected value of the function $f$

## REFERENCES

- Barkmeijer, J., R. Buizza, E. Källén, F. Molteni, R. Mureau, T. Palmer, S. Tibaldi, and J. Tribbia, 2013: 20 years of ensemble prediction at ECMWF. *ECMWF Newsletter*, No. 134, ECMWF, Reading, United Kingdom, 16–32, <https://www.ecmwf.int/node/17373>.
- Beck, C., and F. Schlögl, 1995: Transfer operator methods. *Thermodynamics of Chaotic Systems*, Cambridge University Press, 190–203.
- Bell, R., and B. Kirtman, 2019: Seasonal Forecasting of wind and waves in the North Atlantic using a grand multimodel ensemble. *Wea. Forecasting*, **34**, 31–59, <https://doi.org/10.1175/WAF-D-18-0099.1>.
- Berry, T., D. Giannakis, and J. Harlim, 2015: Nonparametric forecasting of low-dimensional dynamical systems. *Phys. Rev. E*, **91**, 032915, <https://doi.org/10.1103/PhysRevE.91.032915>.
- Budišić, M., R. Mohr, and I. Mezić, 2012: Applied Koopmanism. *Chaos*, **22**, 047510, <https://doi.org/10.1063/1.4772195>.
- Copernicus Climate Change Service, 2017: ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate. Copernicus Climate Change Service Climate Data Store, accessed 29 September 2019, <https://cds.climate.copernicus.eu/cdsapp#!home>.
- Ding, H., M. Newman, M. A. Alexander, and A. T. Wittenberg, 2019: Diagnosing secular variations in retrospective ENSO seasonal forecast skill using CMIP5 model-analogs. *Geophys. Res. Lett.*, **46**, 1721–1730, <https://doi.org/10.1029/2018GL080598>.
- Flaxman, S., D. Sejdinovic, and J. Cunningham, and S. J. Filippis, 2016: Bayesian learning of kernel embeddings. *Proc. 32nd Conf. on Uncertainty in Artificial Intelligence*, New York, NY, AUAI.
- Gaspard, P., 2007: From dynamical systems theory to nonequilibrium thermodynamics. *Symp. Henri Poincaré*, Brussels, Belgium, International Solvay Institutes for Physics and Chemistry, 97–119.
- , and S. Tasaki, 2001: Liouvillian dynamics of the Hopf bifurcation. *Phys. Rev. E*, **64**, 056232, <https://doi.org/10.1103/PhysRevE.64.056232>.
- , G. Nicolis, A. Provata, and S. Tasaki, 1995: Spectral signature of the pitchfork bifurcation: Liouville equation approach. *Phys. Rev. E*, **51**, 74–94, <https://doi.org/10.1103/PhysRevE.51.74>.
- Giannakis, D., 2019: Data-driven spectral decomposition and forecasting of ergodic dynamical systems. *Appl. Comput. Harmon. Anal.*, **47**, 338–396, <https://doi.org/10.1016/j.acha.2017.09.001>.
- Ham, Y.-G., J.-H. Kim, and J.-J. Luo, 2019: Deep learning for multi-year ENSO forecasts. *Nature*, **573**, 568–572, <https://doi.org/10.1038/s41586-019-1559-7>.
- Kay, J. E., and Coauthors, 2015: The Community Earth System Model (CESM) Large Ensemble project: A community resource for studying climate change in the presence of internal climate variability. *Bull. Amer. Meteor. Soc.*, **96**, 1333–1349, <https://doi.org/10.1175/BAMS-D-13-00255.1>.
- 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, 2019: Eigendecompositions of transfer operators in reproducing kernel Hilbert spaces. *J. Nonlinear Sci.*, <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>.
- Korda, M., and I. Mezić, 2019: Optimal construction of Koopman eigenfunctions for prediction and control. arXiv, <https://arxiv.org/abs/1810.08733>.
- Lasota, A., and M. C. Mackey, 1994: *Chaos, Fractals, and Noise: Stochastic Aspects of Dynamics*. 2nd ed. Applied Mathematical Sciences, Vol. 97, Springer, 472 pp.
- Maher, N., and Coauthors, 2019: The Max Planck Institute Grand Ensemble: Enabling the exploration of climate system variability. *J. Adv. Model. Earth Syst.*, **11**, 2050–2069, <https://doi.org/10.1029/2019MS001639>.
- Majda, A. J., and D. Qi, 2020: Statistical phase transitions and extreme events in shallow water waves with an abrupt depth change. *J. Stat. Phys.*, **179**, 1718–1741, <https://doi.org/10.1007/s10955-019-02465-3>.
- 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>.
- Molteni, F., R. Buizza, T. Palmer, and T. Petroliagis, 1996: The ECMWF Ensemble Prediction System: Methodology and validation. *Quart. J. Roy. Meteor. Soc.*, **122**, 73–119, <https://doi.org/10.1002/qj.49712252905>.
- Muandet, K., K. Fukumizu, B. Sriperumbudur, and B. Schölkopf, 2017: Kernel mean embedding of distributions: A review and beyond. *Found. Trends Mach. Learn.*, **10**, 1–141, <https://doi.org/10.1561/22000000060>.
- Navarra, A., J. Tribbia, and G. Conti, 2013: Atmosphere–ocean interactions at strong couplings in a simple model of El Niño. *J. Climate*, **26**, 9633–9654, <https://doi.org/10.1175/JCLI-D-12-00763.1>.

- 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>.
- Nüske, 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>.
- Pedregosa, F., and Coauthors, 2011: Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.*, **12**, 2825–2830.
- 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](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](https://doi.org/10.1175/1520-0442(1995)008<1999:TOGOTS>2.0.CO;2).
- Poincaré, H., 1906: Reflexions sur la theorie cinetique des gaz. *J. Phys. Theor. Appl.*, **5**, 369–403.
- 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>.
- Schölkopf, B., and A. J. Smola, 2001: *Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond*. MIT Press, 626 pp.
- Schwartz, C. S., G. S. Romine, R. A. Sobash, K. R. Fossell, and M. L. Weisman, 2019: NCAR's real-time convection-allowing ensemble project. *Bull. Amer. Meteor. Soc.*, **100**, 321–343, <https://doi.org/10.1175/BAMS-D-17-0297.1>.
- Steinwart, I., and A. Christmann, 2008: *Support Vector Machines*. 1st ed. Springer, 601 pp.
- Tibaldi, C., and R. Knutti, 2007: The use of the multi-model ensemble in probabilistic climate projections. *Philos. Trans. Roy. Soc.*, **365A**, 2053–2075, <https://doi.org/10.1098/rsta.2007.2076>.
- Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.*, **74**, 2317–2330, [https://doi.org/10.1175/1520-0477\(1993\)074<2317:EFANTG>2.0.CO;2](https://doi.org/10.1175/1520-0477(1993)074<2317:EFANTG>2.0.CO;2).
- Tu, J. H., 2013: Dynamic mode decomposition: Theory and applications. Ph.D. thesis, Mechanical and Aerospace Engineering Dept., Princeton University, 123 pp.
- , C. W. Rowley, D. M. Luchtenburg, S. B. Brunton, and J. N. Kutz, 2014: On dynamic mode decomposition: Theory and applications. *J. Comput. Dyn.*, **1**, 391, <https://doi.org/10.3934/jcd.2014.1.391>.
- Ulam, S. M., 1960: *A Collection Of Mathematical Problems*. Interscience Publishers, 150 pp.
- Vautard, R., and M. Ghil, 1989: Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series. *Physica D*, **35**, 395–424, [https://doi.org/10.1016/0167-2789\(89\)90077-8](https://doi.org/10.1016/0167-2789(89)90077-8).
- 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>.