A Koopman-based framework for forecasting the spatiotemporal evolution of chaotic dynamics with nonlinearities modeled as exogenous forcings

M. A. Khodkar[1],[*] Pedram Hassanzadeh[1,2],[†] and Athanasios Antoulas[3,4,5]

[1]*Department of Mechanical Engineering,*

Rice University, Houston, TX 77005, USA

[2]*Department of Earth, Environmental and Planetary Sciences,*

Rice University, Houston, TX 77005, USA

[3]*Department of Electrical and Computer Engineering,*

Rice University, Houston, TX 77005, USA

[4]*Baylor College of Medicine, 1 Baylor Plaza, Houston, TX 77030, USA*

[5]*Data-Driven System Reduction and Identification (DRI) Group,*

Max Planck Institute for Dynamics of Complex Technical Systems,

Sandtorstrasse 1, 39106 Magdeburg, Germany

(Dated: September 4, 2019)

Abstract

We introduce a data-driven method and shows its skills for spatiotemporal prediction of high-dimensional chaotic dynamics and turbulence. The method is based on a finite-dimensional approximation of the Koopman operator where the observables are vector-valued and delay-embedded, and the nonlinearities are treated as external forcings. The predictive capabilities of the method are demonstrated for well-known prototypes of chaos such as the Kuramoto-Sivashinsky equation and Lorenz-96 system, for which the data-driven predictions are accurate for several Lyapunov timescales. Similar performance is seen for two-dimensional lid-driven cavity flows at high Reynolds numbers.

---

[*] mkhodkar@rice.edu

[†] pedram@rice.edu

1

I. INTRODUCTION

Predicting the spatiotemporal evolution of high-dimensional and nonlinear dynamical systems, such as turbulent flows, has been of long-standing interest in science and engineering [1, 2]. For example, forecasting turbulent flows plays a key role in controlling and optimizing various engineering systems (e.g. wind farms) and predicting the state of the atmosphere and/or ocean (e.g. day-to-day weather) [3–5]. For many of these problems, an objective of particular interest and wide-ranging applications is predicting extreme events at some useful lead time [6–8].

Data-driven prediction of chaotic dynamics and turbulent flows has received significant attention in recent years, in particular, for problems in which the high-dimensional, non-linear governing equations cannot be solved fast enough to be useful (e.g. for online control/optimization), or in which some of the physical processes (and thus the governing equations) are not fully understood but observational data from the past are available (e.g. the weather/climate systems) [9–14]. Rapid advances have been recently made in this area based on using techniques from machine learning or approximating the Koopman operator. These approaches involve using the past data to build/train a model that can produce accurate and fast predictions about the future spatiotemporal evolution of the flow. Promising results for prototypes of chaotic dynamics, e.g. Lorenz-63/96 and Kuramoto-Sivashinsky (K-S) equations, or classical fluid examples such as vortex shedding past a cylinder and homogenous isotropic turbulence, have been reported using machine learning methods such as long short-term memory (LSTM) networks, physics-informed neural networks, an reservoir computing [15–22].

The Koopman operator [23], which is an infinite-dimensional linear operator that shifts observables forward in time, offers a powerful framework for analyzing and tackling nonlinear systems such as fluid flows [24, 25]. Data-driven, finite-dimensional approximations of the Koopman operator, using methods such as dynamic mode decomposition (DMD) and its variants [26–32], have been extensively used to analyze various flows in recent years [25, 33]. Coupling the Exact DMD of Tu *et al.* [28] with the delay-embedding theorem of Takens [34], Arbabi and Mezić [30] introduced the Hankel-DMD method and proved its convergence to the Koopman operator. Korda and Mezić [35] have utilized this method for the short-term forecasting of a forced Van der Pol oscillator as well as for the feedback control of a

bilinear DC motor. This concept was further employed by Arbabi *et al.* [36] to develop a predictive-control framework for the purpose of stabilizing a two-dimensional (2D) lid-driven cavity flow that has bifurcated to a limit cycle, with emphasis on the delay-embedding of measurements to reduce the number of placed sensors.

Furthermore, Giannakis and collaborators have shown that projecting the delay-embedded data onto the eigenfunctions of Laplace-Beltrami [37] or Koopman operator [12] leads to the frameworks which are skillful in model reduction, mode decomposition and forecasting of time series for ergodic dynamical systems with strongly nonlinear phenomena such as intermittitencies or energy bursts. Brunton *et al.* [38] have built another successful data-driven model, skilled in detecting the low-probability events of chaotic dynamics, based on the sparse identification of nonlinear dynamics (SINDy) described in Brunton *et al.* [39], and the representation of chaos as an intermittent forcing in the form of the principal component (PC) of the last retained mode given by the singular value decomposition (SVD) of the delay-embedded data. They notice that the statistics of this intermittent forcing is non-Gaussian, and it is connected to the PCs of other retained modes present in the state vector in a nonlinear fashion. The model, referred to as Hankel alternative view of Koopman (HAVOK) by the authors, then accurately predicts the lobe-switching events of various chaotic attractors such as Lorenz-63, Rössler and double-pendulum, one time unit ahead of their occurrences.

The corresponence between the aforementioned Koopman-based methods which benefit from delay-embedded measurements and the system identification method matrix pencil (which is called Loewner method when used in frequency-domain) in time-domain is mathematically established by Ionita and Antoulas [40]. This approach has been recently used for data-driven model reduction of nonlinear systems such as Burgers' equation [41–43]. The underlying connections between the two methods have been further investigated in Pogorelyuk and Rowley [44].

Despite the success of previous models, data-driven spatiotemporal prediction of high-dimensional and highly chaotic systems for reasonably long times is still the subject of ongoing research. Towards this end, we develop a data-driven Koopman-based method which models the nonlinearities as external forcings (actuations), while the observables forming the state vector are vector-valued, delay-embedded and linear. The unknown maps (matrices) appearing in the method are found using the method of Proctor *et al.* [45], known as

dynamic mode decomposition with control (DMDc). The paper is organized as follows. The mathematical derivation of the method and its implementation for some well-known prototypes of chaos are discussed in Sec. II. Further results regarding the predictive capabilities of the method for chaotic dynamical systems (Lorenz-63, K-S and Lorenz-96) as well as a fluid test case (a 2D lid-driven cavity flow at high Reynolds numbers) are presented in Sec. III. Section IV summarizes the methodology and the findings, and outlines the prospects of the proposed method.

II. METHODOLOGY AND APPLICATIONS TO SOME CHAOTIC TEST CASES

In the following, we present a Koopman-based data-driven method, which enables the spatiotemporal prediction of chaotic dynamics such as the K-S equation and Lorenz-96 system. We also shed some insights into the accurate representation of nonlinearity. General guides regarding the proper selection of the method's parameters are also provided in Appendix A.

A. A data-driven predictive framework for the Kuramoto-Sivashinsky equation

We use the K-S equation, a widely-used prototype for spatiotemporal chaotic systems, as an example to formulate our proposed data-driven method. The K-S equation is described by

$$\frac{\partial u}{\partial t} = -u\frac{\partial u}{\partial x} - \frac{\partial^2 u}{\partial x^2} - \frac{\partial^4 u}{\partial x^4} + \eta \cos(2\pi x/\delta) , \tag{1}$$

where u denotes the K-S variable. The last term in Eq. (1) is a periodic forcing which causes spatial inhomogeneity [18]. Here, we take $\delta = L/2$, where L is the domain length. The choice for the number of collocation points n depends on the system's chaoticity, which is increased by the domain length L, so n will be reported individually for each case. A periodic boundary condition is enforced, and a pseudospectral solver with the classic fourth-order Runge-Kutta (RK4) is used for integrating (1), to construct training sets with $N = 70000$ data points and sampling interval $\tau = 0.02\tau_d$, where τ_d is the decorrelation timescale of the principal component of the leading POD mode (PC1). The training set refers to the part of the dataset used for building the Koopman-based predictive method, which is fully separate

4

from the dataset used for examining the performance of the model, called the testing set. The length of our training sets and sampling intervals are identical to those of Pathak *et al.* [18]. We have also constructed 20 independent testing sets, each with 30000 samples, to evaluate the performance of the proposed data-driven methods.

Relative errors are calculated as

$$E(t) = \|\boldsymbol{u}_{pred} - \boldsymbol{u}_{num}\|_2 / \|\boldsymbol{u}_{num}\|_2 , \tag{2}$$

and are averaged over all these testing sets. Note that $\|\ \|_2$ denotes the Euclidean norm, and $\boldsymbol{u}_{pred}$ and $\boldsymbol{u}_{num}$ represent the values given by the data-driven predictive methods and the pseudospecral numerical solver, respectively, while the latter is taken as ground truth. Another measure of error can also be defined as $E_{ave} = 1/t_l \int_0^{t_l} E(t)\mathrm{d}t$, which is simply the temporal mean of $E(t)$ from the starting point of prediction ($t = 0$) to the divergence time t_l, where t_l corresponds to the time at which E first exceeds 0.3.

Suppose N samples of vector-valued observables $\boldsymbol{u}^i$ obtained from running the numerical solver are arranged in the Hankel matrix $\mathcal{H}$

$$\mathcal{H} = \begin{bmatrix} \boldsymbol{u}^1 & \boldsymbol{u}^2 & \dots & \boldsymbol{u}^{N-q+1} \\ \boldsymbol{u}^2 & \boldsymbol{u}^3 & \dots & \boldsymbol{u}^{N-q+2} \\ \vdots & \vdots & \dots & \vdots \\ \boldsymbol{u}^q & \boldsymbol{u}^{q+1} & \dots & \boldsymbol{u}^N \end{bmatrix} , \tag{3}$$

where $\boldsymbol{u}^i \in \mathbb{R}^n$ is sampled at $t = i\tau$, and q is the delay-embedding dimension. The size of $\mathcal{H}$ is then $(n \times q) \times (N - q + 1)$. Following the Hankel-DMD formulation of Arbabi and Mezić [30], we construct

$$\mathrm{X} = \mathcal{H}(:, 1:N-q) , \quad \mathrm{Y} = \mathcal{H}(:, 2:N-q+1) , \tag{4}$$

and conduct reduced SVD to obtain $\mathrm{X} = \mathrm{USV}^*$, in the subspace of leading r singular vectors ($*$ indicates conjugate transpose). The data-driven approximation of Koopman operator using Hankel-DMD is computed as [30]

$$\mathrm{A}_{\mathrm{HDMD}} = \mathrm{U}^*\mathrm{YVS}^{-1}, \tag{5}$$

which has the size $r \times r$.

Once $\mathrm{A}_{\mathrm{HDMD}}$ is calculated, a future vector-valued observable $\boldsymbol{u}^{m+1}$, that was not part of the training set, can be predicted from

$$\boldsymbol{\mathcal{U}}^{m+1,r} = \mathrm{A}_{\mathrm{HDMD}}\, \boldsymbol{\mathcal{U}}^{m,r} , \tag{6}$$

where $\boldsymbol{\mathcal{U}}^{m+1,r}$, a vector of length r, is $\boldsymbol{\mathcal{U}}^{m+1} = [\boldsymbol{u}^{m-q+2}\ \boldsymbol{u}^{m-q+3}\ \cdots\ \boldsymbol{u}^{m+1}]^T$ (T indicates transpose), a vector of length $n \times q$, projected onto the subspace of first r singular vectors. The first block of $\boldsymbol{\mathcal{U}}^{m+1}$ is then taken as the prediction of the new state. All values in $\boldsymbol{\mathcal{U}}^m = [\boldsymbol{u}^{m-q+1}\ \boldsymbol{u}^{m-q+2}\ \cdots\ \boldsymbol{u}^m]^T$ and its projection onto the subspace of retained singular vectors, $\boldsymbol{\mathcal{U}}^{m,r}$, are either known from the initial condition or already predicted. We notice that the first $q-1$ predictions are the reconstruction of initial condition, and only the qth prediction is the forecast value of the future state. Hereafter, we refer to this method, which is thoroughly based on the Hankel-DMD method introduced by Arbabi and Mezić [30], as M1.

One may suspect that the linear combination of a finite number of DMD modes, specifically those given by a fairly short dataset of linear observables, cannot accurately reproduce the nonlinear characteristics of a chaotic dynamics for a reasonably long period of time. Prior to investigating this hypothesis and the performance of M1, we attempt to develop a modified method (M2), which is specialized to tackle the issue of the nonlinearity. Inspired by HAVOK model of Brunton *et al.* [38], in which adding a forcing term to the linear model is seen to approximate the nonlinear dynamics more accurately and yield better predictions, we incorporate the nonlinear effects in the form of external forcings, so that any dynamical system can be modeled as

$$\boldsymbol{u}^{m+1} = \mathrm{A}\boldsymbol{u}^m + \mathrm{B}\boldsymbol{f}^m. \tag{7}$$

Unlike [38] which uses the last retained singular vector as forcing, we choose the forcing term in a physics-driven fashion. For instance, when some knowledge of the governing equations is available, or one can intuitively speculate the form of nonlinearity, the forcing term can be chosen according to that knowledge or intuition. Consequently, for the K-S equation, forcing vector $\boldsymbol{f}^i$ includes the square of u at the same snapshot and every grid point, i.e. $\boldsymbol{f}^i = \left[(u_1^i)^2\ (u_2^i)^2\ \cdots\ (u_n^i)^2\right]^T$. The delay-embedded form of Eq. (7) reads

$$\boldsymbol{\mathcal{U}}^{m+1} = \mathrm{A}\boldsymbol{\mathcal{U}}^m + \mathrm{B}\boldsymbol{\mathcal{F}}^m, \tag{8}$$

where the definition of $\boldsymbol{\mathcal{U}}^m$ is the same as before. Now the forcing vectors should also be

sampled at each snapshot and sorted in the following Hankel matrix $\mathcal{F}$

$$\mathcal{F} = \begin{bmatrix} \boldsymbol{f}^1 & \boldsymbol{f}^2 & \cdots & \boldsymbol{f}^{N-q} \\ \boldsymbol{f}^2 & \boldsymbol{f}^3 & \cdots & \boldsymbol{f}^{N-q+1} \\ \vdots & \vdots & \cdots & \vdots \\ \boldsymbol{f}^q & \boldsymbol{f}^{q+1} & \cdots & \boldsymbol{f}^{N-1} \end{bmatrix}. \tag{9}$$

Note that however for the K-S system the state vector $\boldsymbol{u}$ and the forcing term $\boldsymbol{f}$ are of the same size, but in general the forcing vector could be much larger and of the length $n' \gg n$, depending on the form of nonlinearities and the number of nonlinear processes in the dynamical system. The size of $\mathcal{F}$ is thus $(n' \times q) \times (N - q)$.

The unknown maps A and B are then found using DMDc method presented in Proctor *et al.* [45], which simply minimizes the Frobenius norm $\|Y - \mathrm{A}X - \mathrm{B}\mathcal{F}\|_F$ to achieve

$$\mathrm{A} = \hat{\mathrm{U}}^* Y \tilde{\mathrm{V}} \tilde{\mathrm{S}}^{-1} \tilde{\mathrm{U}}_1^* \hat{\mathrm{U}},$$
$$\mathrm{B} = \hat{\mathrm{U}}^* Y \tilde{\mathrm{V}} \tilde{\mathrm{S}}^{-1} \tilde{\mathrm{U}}_2^*. \tag{10}$$

Here, $Y = \hat{\mathrm{U}} \hat{\mathrm{S}} \hat{\mathrm{V}}^*$, while the truncation value is taken as r, i.e. $\hat{\mathrm{U}} \in \mathbb{R}^{nq \times r}$, $\hat{\mathrm{S}} \in \mathbb{R}^{r \times r}$ and $\hat{\mathrm{V}} \in \mathbb{R}^{(N-q) \times r}$, and $[X\ \mathcal{F}]^T = \tilde{\mathrm{U}} \tilde{\mathrm{S}} \tilde{\mathrm{V}}^*$ with the truncation value selected as p, so that $\tilde{\mathrm{U}} \in \mathbb{R}^{(n+n')q \times p}$, $\tilde{\mathrm{S}} \in \mathbb{R}^{p \times p}$ and $\tilde{\mathrm{V}} \in \mathbb{R}^{(N-q) \times p}$. $\tilde{\mathrm{U}}_1$ and $\tilde{\mathrm{U}}_2$ are made up of the first nq and the remaining $n'q$ rows of $\tilde{\mathrm{U}}$, respectively. Note that A and B are calculated in a reduced-dimension subspace, and they have the respective sizes $r \times r$ and $r \times nq$. Similar to M1, once training is done and the unknown maps are calculated, the first q data points in the testing set will be used to initialize the state vector, i.e. these points are not predicted. *Nonetheless all results shown after $t = 0$ are newly predicted values by the data-driven methods, and were not part of the initial condition.* Figure 1 summarizes the training and forecasting steps of this method.

Figure 2(a) displays the temporal evolution of $u(x = 8,\ t)$ predicted by M1 (dashed magenta) and M2 (dashed red), and compares them to the actual data (truth) obtained via the numerical integration of Eq. (1). *We reiterate that all results shown for M1 and M2 after $t = 0$ are new forecast values, and were not used during training and building the model or as a part of the initial condition.* Time is scaled by Lyapunov timescale $1/\Lambda_{max}$, where the leading Lyapunov exponent Λ_{max} is calculated following Wolf *et al.* [46]. As shown in this figure, predictions rendered by the conventional Hankel-DMD method (M1)
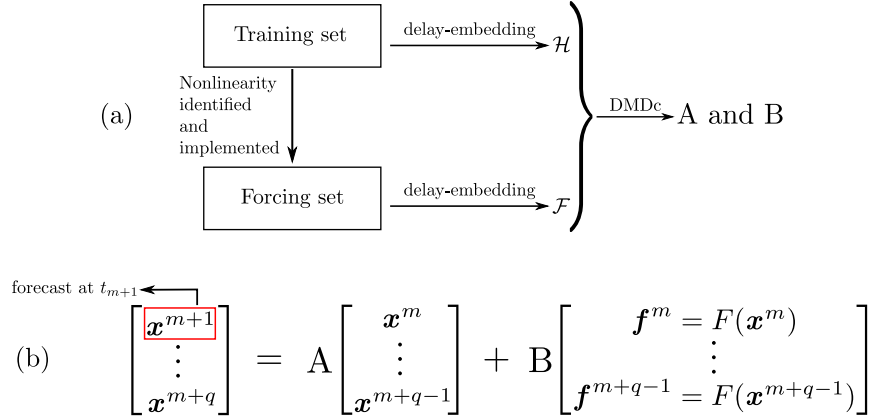
7

FIG. 1. Schematic of data-driven method M2: (a) Training on vector-valued and time-delay-embedded observables yields unknown matrices A and B, while nonlinearites are modeled as external forcings and DMDc of Proctor *et al.* [45] is utilized. (b) Matrices A and B can then be employed for the spatiotemporal forecasting of the dynamical system. All vectors on the right-hand side are either known from initial condition or already predicted. Note that $\boldsymbol{f}^i$ is a function of $\boldsymbol{x}^i$.

diverge from the testing data fairly rapidly in less than a Lyapunov timescale, and after that, predictions gradually decay to zero. This can be attributed to the fact all eigenvalues of A_{HDMD} fall inside the unit circle, with most of them located in its vicinity, meaning all the modes corresponding to these eigenvalues are decaying (Fig. 3(a) and (c)). To distinguish between the eigenvalues more clearly, Fig. 3 shows the eigenvalues λ of $\exp(\tau A)$. Therefore, the eigenvalues of A inside/outside the unit circle correspond to the eigenvalues of $\exp(\tau A)$ to the left/right of the imaginary axis. It is noteworthy that the special case of $q = 1$, leading to the conventional Exact DMD, does not reveal any predictive skill, so that its predictions become inaccurate in less than ten iterations, or $0.1/\Lambda_{max}$. On the other hand, the inclusion of nonlinearity in the form of external forcings has substantially improved the performance of the predictive framework so that, compared to the best results of M1, prediction horizon t_l is increased by a factor larger than 10. However, usually a few of eigenvalues of A given by M2 (six in the case with $L = 22$ and eight in the case with $L = 100$) fall outside the unit circle (Figs. 3(b) and (d)), indicating that the dynamical system has some growing modes, the nonlinear part of the predictive method ($B\boldsymbol{f}$) suppresses the unbounded growth of these modes. This is fully compatible with the underlying physics of nonlinear dynamical systems, in which the unbounded growth of the unstable modes is suppressed by the energy-conserving
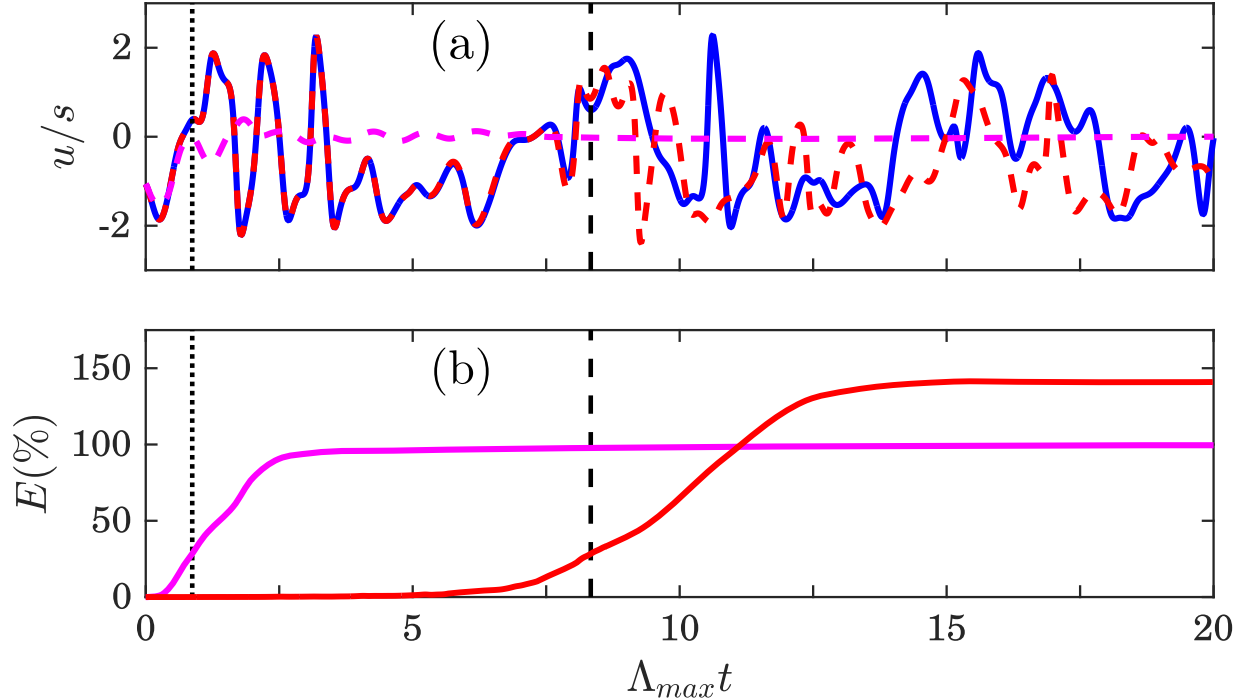
FIG. 2. (a) Predictions given by M1 (dashed magenta) and M2 (dashed red) for the time series of u at $x = 8$, compared to the ground truth obtained via the numerical integration of Eq. (1) (solid blue). Note that y-axis is normalized by the standard deviation s of testing data. (b) Variation of relative error E rendered by each method with time scaled by Lyapunov timescale $1/\Lambda_{max}$. Again, magenta and red lines represent the results of M1 and M2, respectively, while the vertical dotted and dashed lines mark the time at which the prediction of each method diverges from the actual data. The studied K-S system is unforced with the domain length $L = 22$, the number of collocation points $n = 64$, and the attractor dimension $D_{KY} = 5.20$, computed based on the Kaplan-Yorke formulation [50].

nonlinear interactions that transfer energy to the stable modes where the dissipation occurs [47–49]. The close match between the blue circles and red crosses in the right panels of Fig. 3 also shows that halving the length of training set does not change the eigenvalues identified by M2, confirming that these eigenvalues are captured robustly, and the growing modes are integral to the dynamical system. We also highlight that training in time-delay coordinate is a crucial part of M2 as choosing $q = 1$ for this method yields predictions that diverge from the actual data in less than $0.4/\Lambda_{max}$.
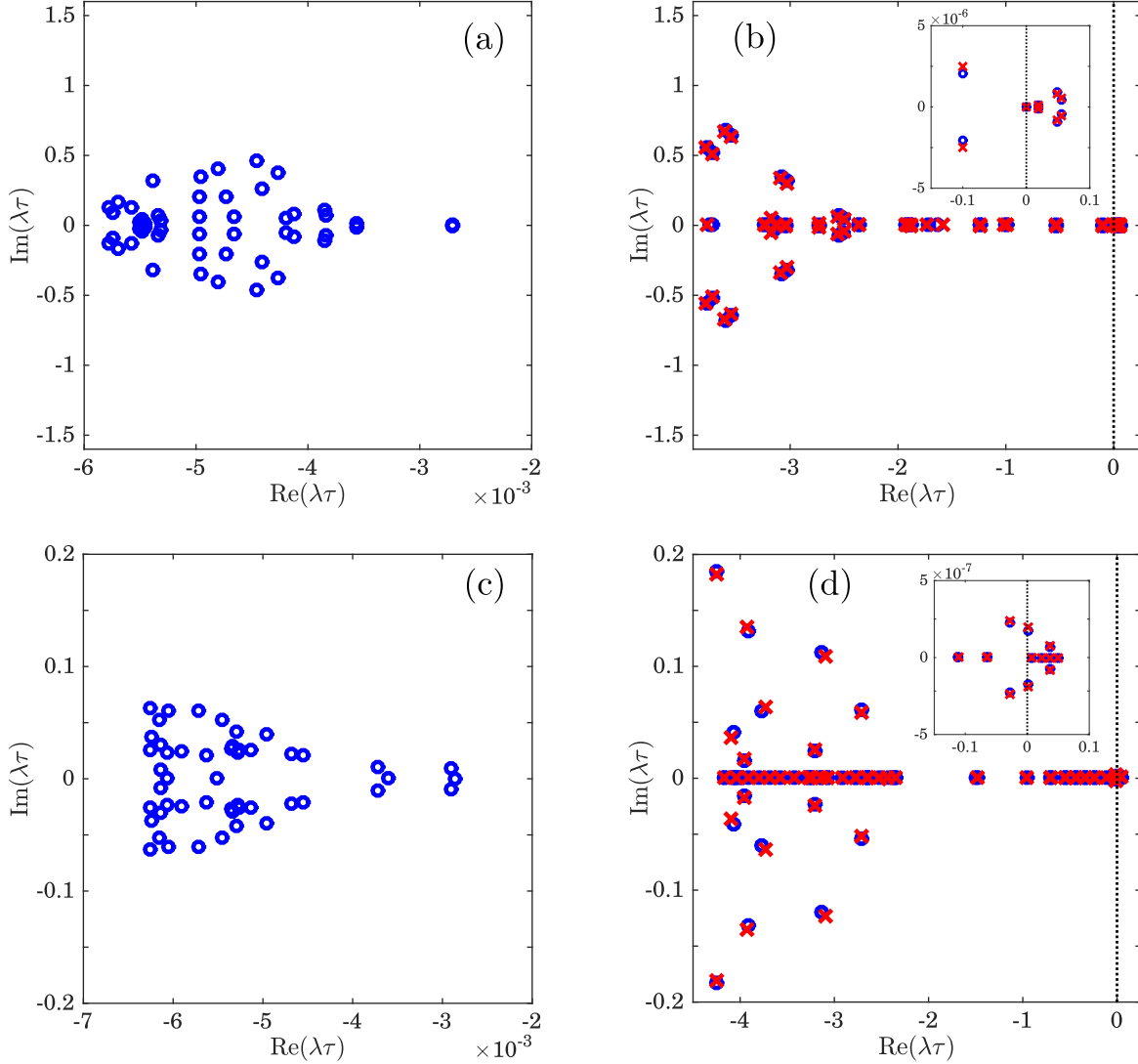
FIG. 3. Eigenvalues λ of $\exp(\tau A)$ for K-S equations with $L = 22$ (top) and $L = 100$ (bottom), normalized by sampling frequency $1/\tau$, and calculated via (a) M1 with $(r, q) = (2500, 50)$, (b) M2 with $(r, p, q) = (400, 400, 20)$, (c) M1 with $(r, q) = (2200, 40)$, and (d) M2 with $(r, p, q) = (800, 800, 10)$. The blue circles show the eigenvalues provided by the entire training set with 70000 data points, while the red crosses in panels (b) and (d) indicate the eigenvalues identified when half the training set is used. The close agreement between the two suggests that the eigenvalues of the systems are captured robustly. The vertical dotted lines in these panels mark the imaginary axis. For clarity, only the first 50 eigenvalues are depicted, and the smaller panels within panels (b) and (d) focusing on the eigenvalues of M2 with growing modes are also included. We notice that except for the few leading eigenvalues, modes given by M2 decay at much faster rates than those given by M1. The results of the case with $L = 100$ will be further discussed in Sec. III B.

B. Accurate representation of nonlinearity beyond squared terms

Here, we consider a different commonly-explored prototype for chaotic dynamics, Lorenz-96 system [51], whose governing ordinary differential equation (ODE) is given by

$$\dot{X}_j = (X_{j+1} - X_{j-2})X_{j-1} - X_j + F \,, \tag{11}$$

where X and overdot indicate the Lorenz variable and time derivative, respectively, and j varies from 1 to $n = 40$. The external forcing term F determines the level of chaoticity. Here, we take $F = 16$. Again, RK4 along with periodic boundary condition is used to numerically integrate the system, and to build training sets with $N = 100000$ data points uniformly sampled at every $\tau = 0.02\tau_d$.

For this test case, although the nonlinearity still has a quadratic form, constructing the forcing vector $\boldsymbol{f}$ using only X_j^2 yields short prediction horizons (see the point corresponding to $J = 0$ in Fig. 4(a)). Seemingly, this is owing to the different nature of nonlinearities appearing in ODE (11) of Lorenz-96 compared to those in the K-S equation; where the former involves terms which are the product of X at some distinct grid points, e.g. $X_{j-1}X_{j+1}$. Motivated by the underlying dynamics of Lorenz-96, we construct the forcing vector so that in addition to X_j^2, it involves terms in the form of $[X_j X_{j+1} \; X_j X_{j+2} \; \cdots \; X_j X_{j+J}]$, where J indicates the number of 'neighboring' points incorporated for building the forcing vector. The vector hence finds the following form

$$\boldsymbol{f}^T = \left[\; X_1^2 \; X_1 X_2 \; \cdots \; X_1 X_{J+1} \; \cdots \; X_j^2 \; X_j X_{j+1} \; \cdots \; X_j X_{j+J} \; \cdots \; X_n X_{n+J} \; \right] \,. \tag{12}$$

Note that when the neighboring points fall outside the domain, boundary conditions are invoked.

Figure 4(a) demonstrates how the prediction horizon t_l changes as the number of neighboring points in the forcing vector grows. An abrupt jump in t_l is observed when J increases from 1 to 2, which is consient with the underlying ODE of Lorenz-96 in which nonlinear terms in the form of $X_{j-1}X_{j+1}$ are present. Further increase in J results in a gradual decline in t_l, as the size of forcing vector, and subsequently, its coefficient matrix B, unnecessarily grow, which leads to less accurate approximations of the components of this matrix.

In the problems for which no knowledge of the underlying dynamics is available, a fully data-driven alternative approach for detecting the nonlinearities can be sought by calculating
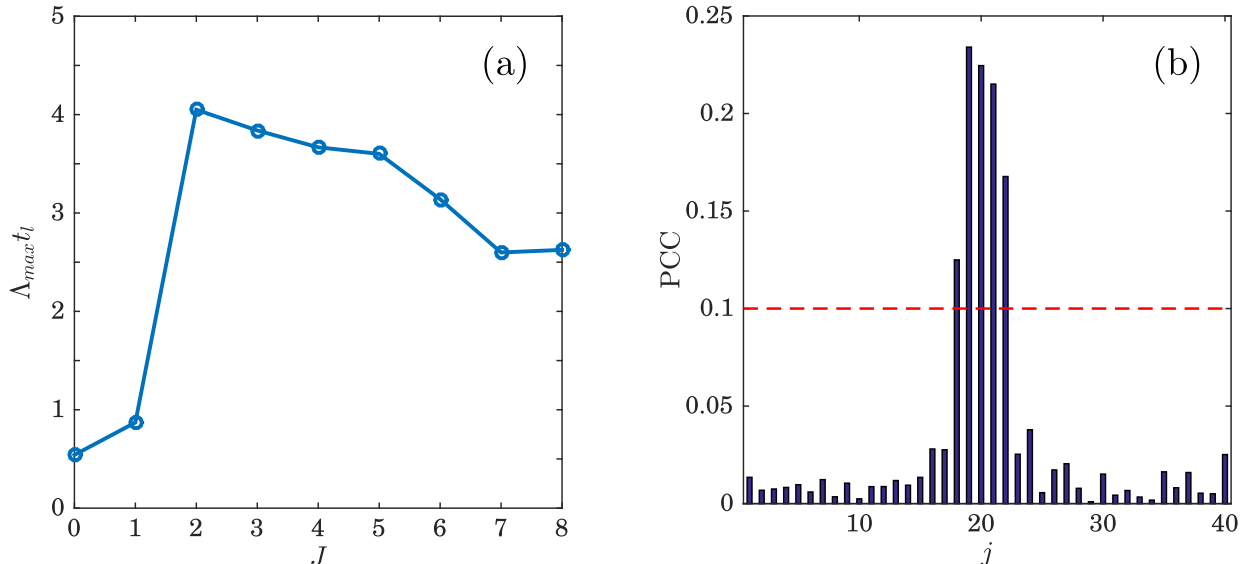
FIG. 4. (a) Variation of prediction horizon t_l with the number of neighboring points J used for constructing quadratic forcing terms. (b) Pearson correlation coefficient (PCC) between the time series of the temporal derivative at a certain point $(\dot{X}_{20})$ and the time series of X at every grid point. The horizontal dashed line corresponds to the threshold 0.1 for choosing the neighboring points. For the Lorenz-96 system under consideration, the number of grid points, external forcing term and attractor dimension respectively equal $n = 40$, $F = 16$ and $D_{KY} = 32$.

the Pearson correlation coefficients (PCC) between the time series of the temporal derivative at a certain point $\dot{X}_I$ and the time series of X_j at all points as

$$\text{PCC} = \frac{\boldsymbol{E}\big[(X_j - \mu_{X_j})(\dot{X}_I - \mu_{\dot{X}_I})\big]}{s_{X_j} s_{\dot{X}_I}}, \tag{13}$$

where μ and s denote the mean and standard deviation of each time series, respectively, and $\boldsymbol{E}$ represents expecation operator. However, due to the chaoticity of the system, the temporal evolution of Lorenz variable at all grid points are interconnected, the dependence is anticipated to be stronger, when according to ODE (11), $\dot{X}_I$ is directly a fucntion of X_j. Consequently, as can be seen in Fig. 4(b), PCC is substantially larger for grid point I or the points in one- or two-grid-point distance from I, since terms involving these points explicitly appear in the underlying ODE of Lorenz-96. The points above the threshold (dashed red line in Fig. 4(b)) can then be selected as the neighboring points while constructing the forcing vector. We remark that this data-driven approach solely identifies the neighboring points, and does not provide any information with regard to the order of nonlinearity. Once the

neighboring points are detected, one can obtain low- to high-order monomials by multiplying the Lorenz variables at the neighboring points with each other, and stack up these terms to build the forcing vector. The highest-order term required for constructing the forcing vector can be determined by the investigator's speculation or intuition.

III. TESTING THE PERFORMANCE OF METHOD M2 FOR VARIOUS CHAOTIC DYNAMICAL SYSTEMS

In the following section, we present the detailed results of the Koopman-based method that treats the nonlinearities as external actuations (M2), when it is used to predict the spatiotemporal evolution of a variety of chaotic systems, from the simple and commonly-used Lorenz-63 to high-dimensional and highly chaotic K-S and Lorenz-96 systems. The section is then concluded by a more complex and larger-scale fluid example (a 2D lid-driven cavity flow at Re = 20000 and 30000). For all examples, prediction horizon t_l and averaged error E_{ave} are calculated based on the definitions and procedures introduced in Sec. II.

A. The Lorenz-63 system

Lorenz-63, one of the most well-known prototypes of chaotic dynamics, was originally developed by Lorenz [1] as a simplified mathematical model for the atmospheric convection from the relatively complicated equations of motion and heat transport for an incompressible Boussinesq flow. The dynamics of the Lorenz-63 system is characterized by the following ODEs

$$
\begin{aligned}
\dot{x} &= \sigma(y - x)\,, \\
\dot{y} &= x(\rho - z) - y\,, \\
\dot{z} &= xy - \beta z\,,
\end{aligned}
\tag{14}
$$

where σ and ρ represent the Prandtl and scaled Rayleigh numbers, respectively, and β is related to the dimensions of the atmospheric layer. Following Lorenz [1], we take $\sigma = 10$, $\rho = 28$ and $\beta = 8/3$, for which the system reveals a chaotic beahaviour with a strange, butterfly-like attractor. The same RK4 scheme as Sec. II is used to numerically integrate ODEs (14), and to construct training sets with 100000 data points, which are uniformly
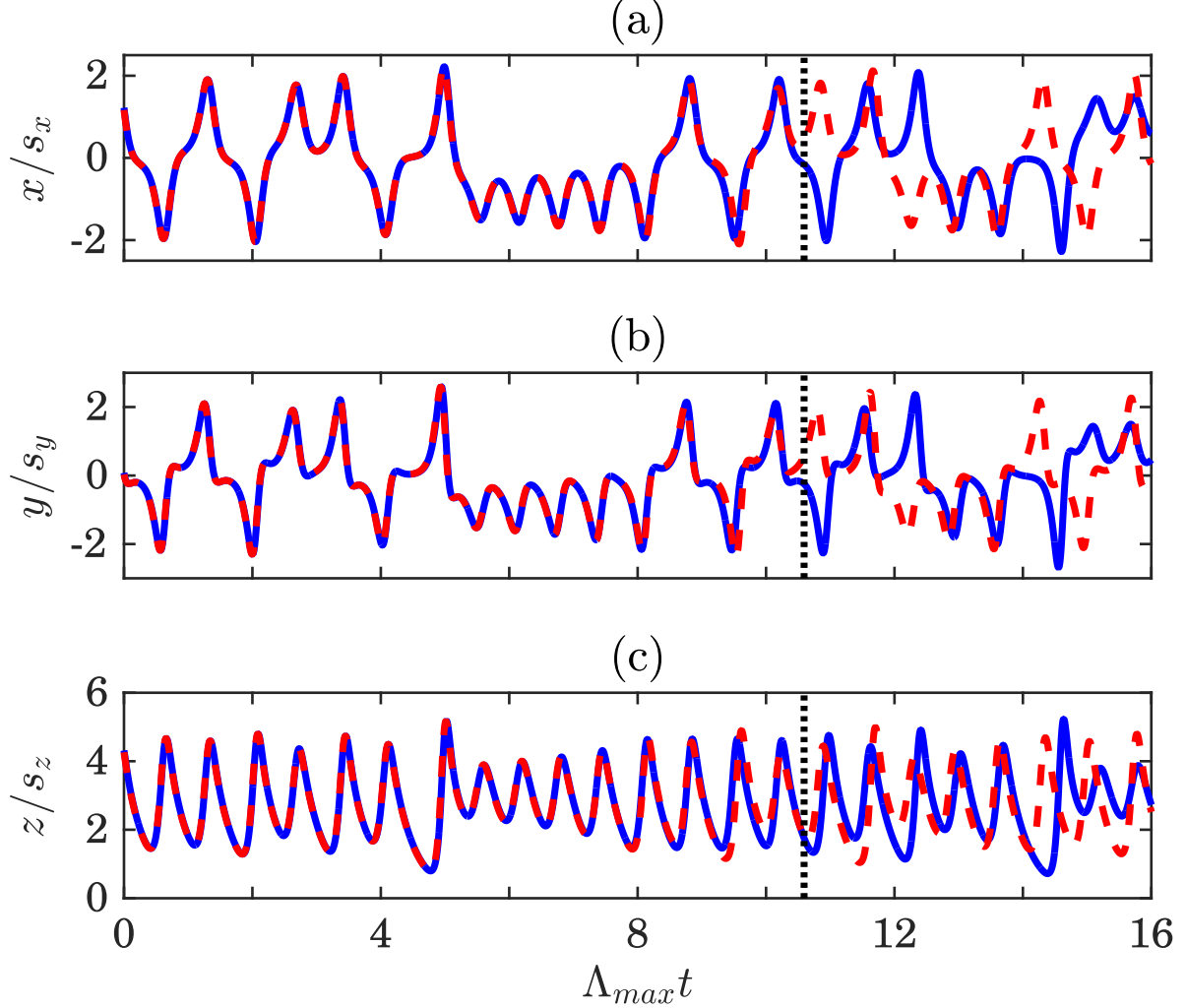
FIG. 5. Predictions of the present method M2 for the temporal evolution of the variables in a Lorenz-63 system with $\sigma = 10$, $\rho = 28$ and $\beta = 8/3$, where the method parameters are selected as $(r, p, q) = (60, 120, 50)$. For this Lorenz-63 system, the Kaplan-Yorke-based attractor dimension and the leading Lyapunov exponent of the system are found to be $D_{KY} = 2.06$ and $\Lambda_{max} = 0.91$. Note that vertical axes are normalized by the standard deviation of the corresponding variable from testing data, denoted by s_x, s_y or s_z.

sampled at every $\tau = 0.02\tau_d$. Note that while developing a data-driven model via M2, all possible quadratic combinations of x, y and z are included in the forcing vector of Eq. (7), i.e. $\boldsymbol{f}^T = [x^2 \ y^2 \ z^2 \ xy \ xz \ yz]$.

As shown in Fig. 5, the predictions of M2 follow the true trajectory for more than 10 Lyapunov timescales with $E_{ave} = 2.10\%$, while the extreme events, i.e. events at which $|x| > 2s_x$ and $|y| > 2s_y$, are also accurately captured. We reiterate that using M1 or M2
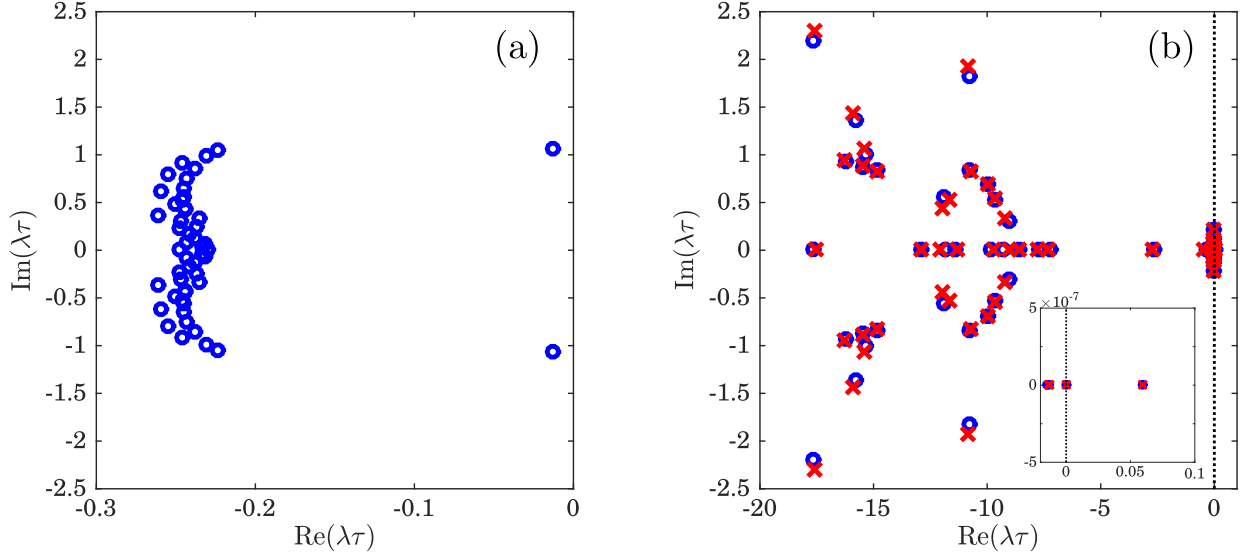
14

FIG. 6. Eigenvalues λ of $\exp(\tau A)$ for the under consideration Lorenz-63 system, scaled by sampling frequency $1/\tau$, and obtained using (a) M1 with $(r, q) = (100, 50)$, (b) M2 with $(r, p, q) = (60, 120, 50)$. The blue circles show the eigenvalues given by the entire training set with 100000 samples, while the red crosses in panel (b) indicate the eigenvalues identified using half the training set. The close agreement between the two suggests that the eigenvalues of the system are captured robustly. The vertical dotted line in panel (b) corresponds to the imaginary axis. For clarity, only the first 50 eigenvalues are depicted. Smaller panel within panel (b) magnifies the eigenvalue of M2 with positive real part. Note that except for the few leading eigenvalues, modes given by M2 decay at much faster rates than those given by M1.

with $q = 1$ results in rapidly diverging predictions, so that for both $\Lambda_{max} t_l < 0.1$, suggesting that the delay-embedding of the measurements along with the incorporation of physics-driven frocings are vital for reasonably long-time accurate predictions. As expected, again, all the eigenvalues of the system identified by M1 fall inside the unit circle (Fig. 6a), which is consistent with the quickly vanishing predictions of this method for Lorenz-63.

B.   Kuramoto-Sivashinsky equation

The spatiotemporal prediction of M2 for a K-S equation with domain length $L = 100$ is displayed in Fig. 7(b), and is compared against the testing set provided by the numerical solver of Sec. II A in Fig. 7(a). The difference of the two is shown in panel (c). Futhermore,
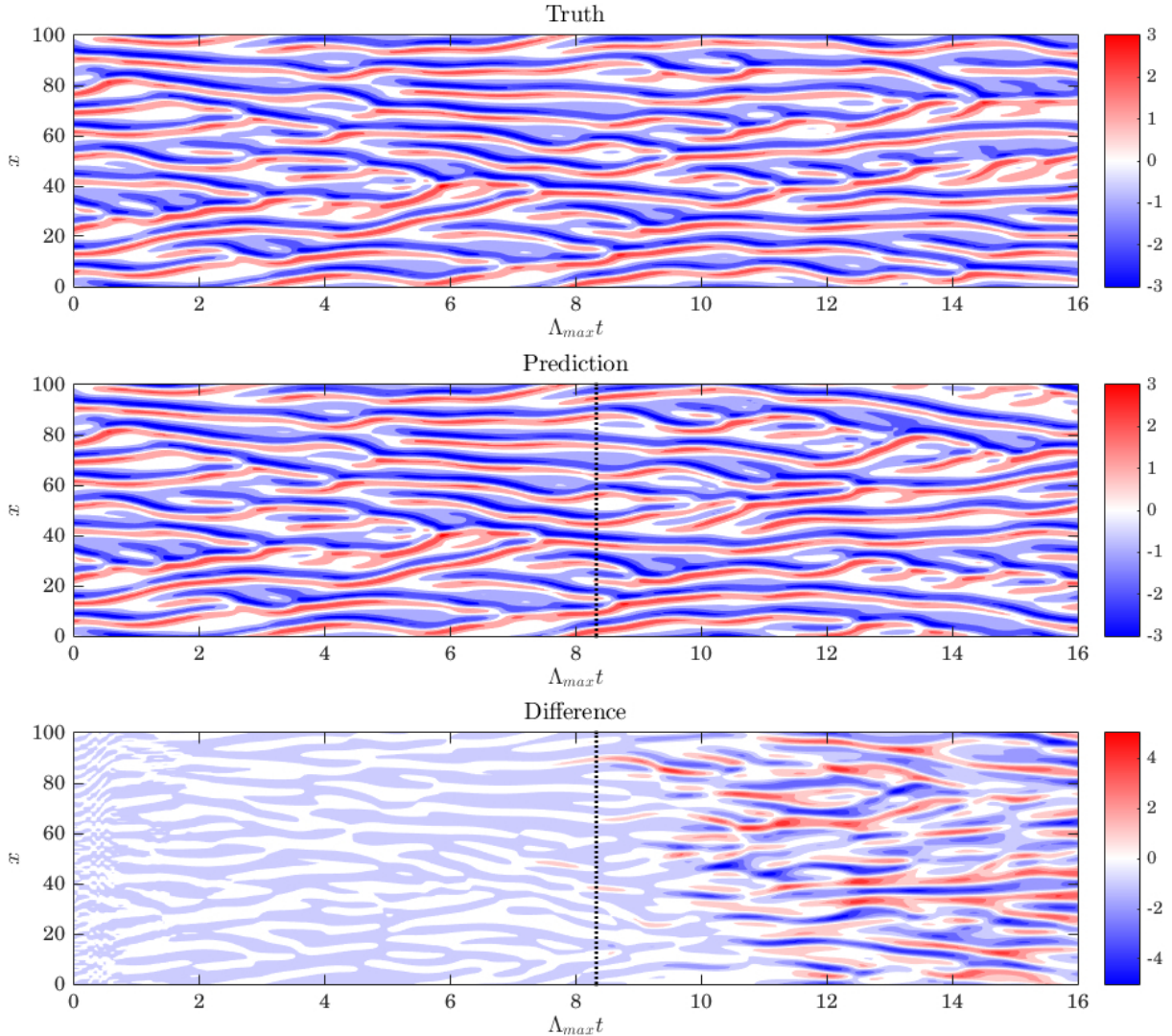
15

FIG. 7. Spatiotemporal evolution of an unforced K-S system with the domain length $L = 100$ and the Kaplan-Yorke dimension $D_{KY} = 23.2$. Shading shows (a) $u(x,t)$ from integration of Eq. (1); (b) $u(x,t)$ predicted by M2 when a training set with 70000 data points is used; (c) Difference of (a) and (b). The vertical dotted lines in panels (b) and (c) mark the divergence time t_l of the data-driven predictions. Λ_{max} is the largest positive Lyapunov exponent.

the number of grid points n needed for stably advancing the K-S equation (1) in time, the properties of the attractor (D_{KY} and Λ_{max}), the method parameters (r, p, q) leading to the best results, and the assessment of the method performance (t_l and E_{ave}) are detailed in Table I, for this case and several other K-S systems with different domain lengths. All cases in Table I are identical to those examined in Pathak *et al.* [18] in terms of domain length,

16

TABLE I. Prediction horizon t_l and averaged error E_{ave} for K-S equation (1) at different levels of chaoticity controlled by domain length L. n is the number of grid points or the physical dimension, and D_{KY} denotes the attractor dimension calculated based on the Kaplan-Yorke formulation. Method parameters (r, p, q) giving the best results are also reported. All cases are unforced, except those with asterisks for which $\eta = 0.01$.

L	n	D_{KY}	(r, p, q)	$\Lambda_{max} t_l$	E_{ave}
22	64	5.20	$(400, 400, 20)$	8.35	3.36
100	128	23.2	$(800, 800, 10)$	8.30	5.73
100*	128	24.1	$(800, 800, 10)$	8.03	3.81
200	256	43.3	$(1200, 1200, 10)$	8.17	4.82
200*	256	44.8	$(1200, 1200, 10)$	8.17	6.95
400	256	85.0	$(3000, 3000, 15)$	7.51	6.12
800	512	167	$(6000, 6000, 15)$	7.12	7.32
1600	1024	338	$(16000, 16000, 20)$	6.43	10.9

potential forcing, attractor properties, and the length and sampling interval of the training set. As can be seen in this table, for moderately chaotic systems ($L \leq 200$), the data-driven predictions remain accurate for more than $8/\Lambda_{max}$, while E_{ave} is below 7%. As L and choticity further grow, t_l slowly declines and E_{ave} gradually increases. Notwithstanding, for all considered cases, M2 provides skillfull forecasts for relatively long times, and it modestly outperforms the reservoir computing approach of Pathak *et al.* [18] for which, t_l was found to be around $6/\Lambda_{max}$ for all cases, when an adequeate number of parallel reservoirs had been used.

C. The Lorenz-96 system

Figure 8 depicts the spatiotemporal evolution of a Lorenz-96 system with the external forcing $F = 8$, given by RK4 integration of Eq. (11) (panel (a)) and the data-driven prediction of M2 (panel (b)). The last panel shows the difference between the two results. In addition to this case, two other systems with $F = 4$ (lower chaoticity) and $F = 16$ (higher chaoticity) are explored as well. For all cases, the number of grid points is fixed at $n = 40$.
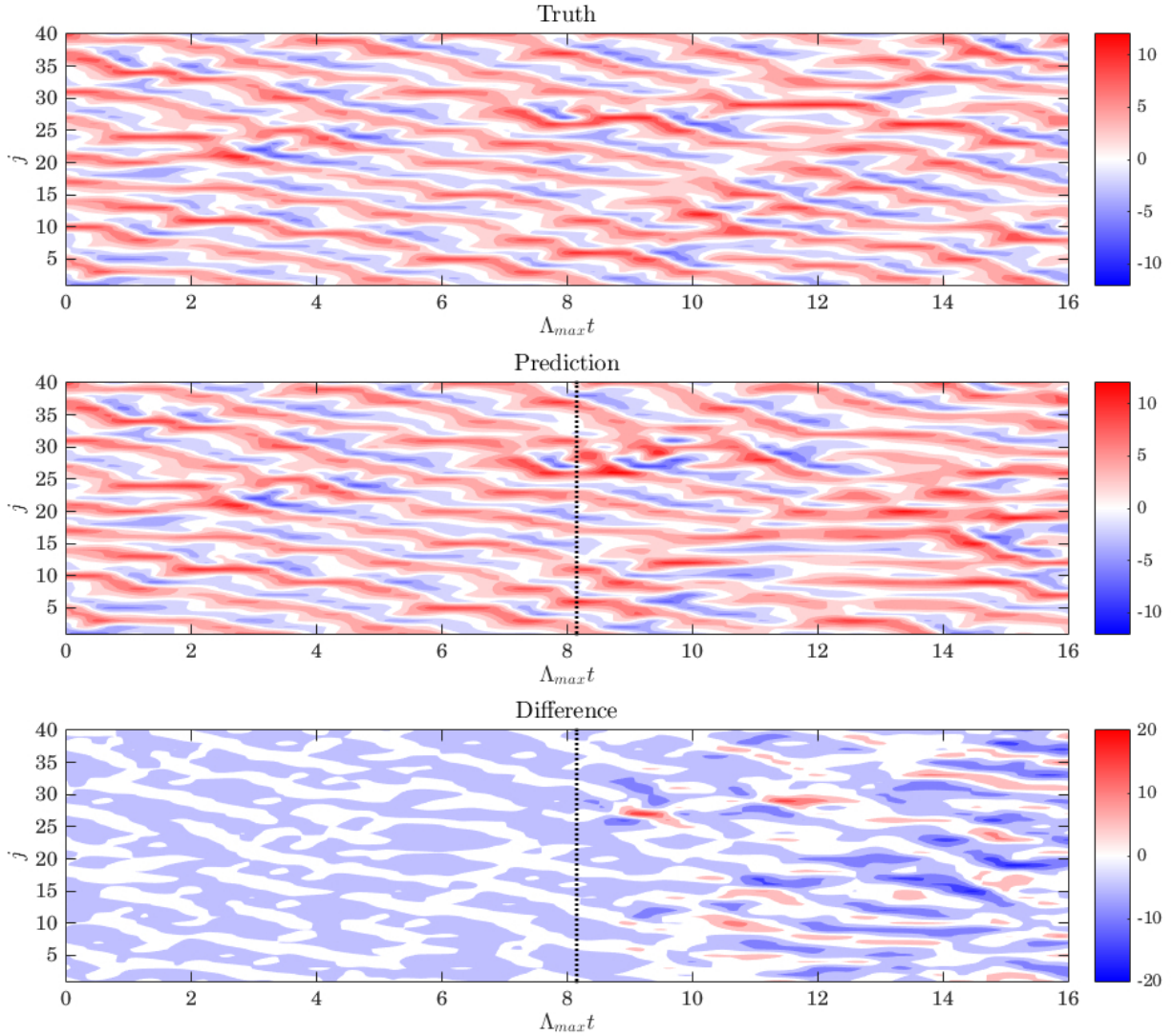
FIG. 8. Spatiotemporal evolution of a Lorenz-96 system with the external forcing $F = 8$ and the Kaplan-Yorke dimension $D_{KY} = 28.4$. (a) Testing set obtained by numerically intergrating Eq. (11); (b) M2 predictions while the model is built using a training set with 100000 samples; (c) Difference of (a) and (b). The vertical dotted lines in panels (b) and (c) correspond to the time t_l at which the predictions of the data-driven method diverge from the actual data. Λ_{max} is the largest positive Lyapunov exponent.

Further details of each case, viz., attractor properties (D_{KY} and Λ_{max}), the trio of optimal method parameters (r, p, q), the prediction horizon t_l and the averaged error E_{ave} are reported in Table II. It is observed that when the system exhibits a quasiperiodic behavior and has a power spectrum with some local maxima (e.g. the test case with $F = 4$), contingent
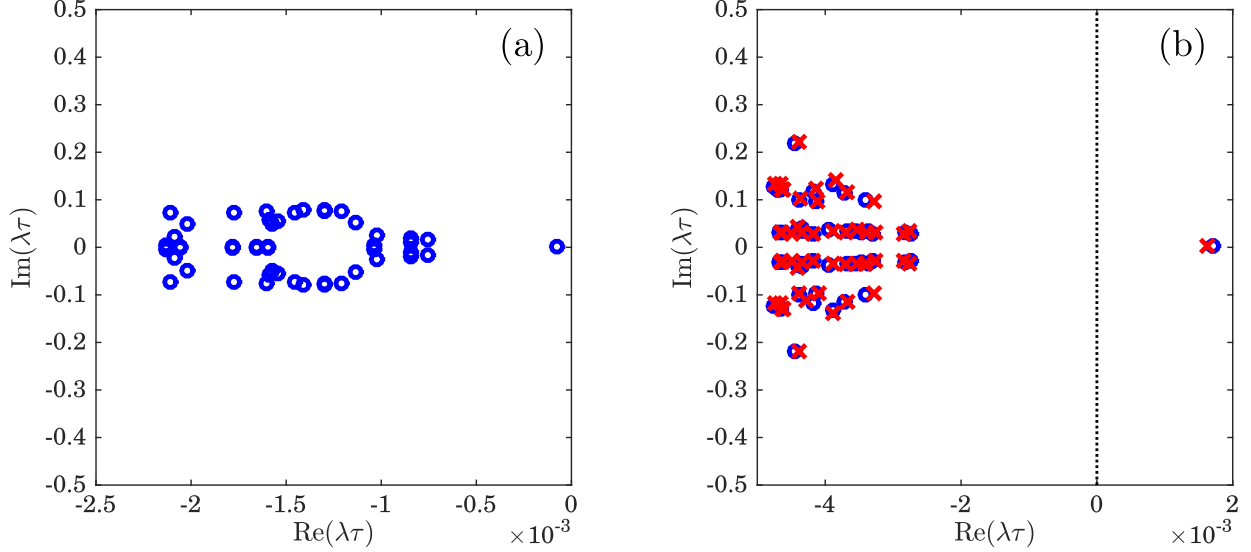
18

FIG. 9. Eigenvalues λ of $\exp(\tau A)$ for the studied Lorenz-96 system with $F = 8$, scaled by sampling frequency $1/\tau$, and provided by (a) M1 with $(r, q) = (200, 20)$, (b) M2 with $(r, p, q) = (700, 2100, 20)$. The blue circles and red crosses show the eigenvalues detected by the entire and half the training set, respectively. The close agreement between the two suggests that the eigenvalues of the system are captured robustly. Vertical dotted lines in panel (b) indicates the imaginary axis. For clarity, only the first 50 eigenvalues are depicted. Smaller panel within panel (b) magnifies the eigenvalues of M2 corresponding to the growing modes. Note that except for the few leading eigenvalues, modes given by M2 decay at much faster rates than those given by M1.

upon the availability of enough snapshots for training, M2 predictions can be accurate for very long times, and occassionally they may never diverge, meaning the underlying dynamics can be fully discovered. As the chaoticity of the Lorenz-96 system is increased by doubling F, t_l becomes finite $(8.16/\Lambda_{max})$ and E_{ave} grows by almost 2.5%. Further doubling of F makes the system highly chaotic so that forecasting its spatiotemporal evolution becomes challenging. Nevertheless, the present Koopman-based method still yields predictions which are accurate for more than $4/\Lambda_{max}$ with $E_{ave} = 7.67\%$, when $F = 8$.

Figure 9 displays the eigenvalues of the Lorenz-96 system with $F = 8$ given by M1 (left) and M2 (right). As demonstrated by this figure, not surprisingly, all modes of A_{HDMD} are decaying again, leading to the predictions that approach zero fairly rapidly, and lose the true trajectory in less than $0.1/\Lambda_{max}$. Furthermore, delay-embedding of the vector-valued observables was found to be essential in M2, so that choosing $q = 1$ using this method

19

TABLE II. Prediction horizon t_l and averaged error E_{ave} for Lorenz-96 system (11) with different external forcings F. n and D_{KY} indicate the number of grid points and Kaplan-Yorke-based attractor dimension, respectively. Method parameters (r, p, q) leading to the most accurate predictions are also presented. Despite using very long testing sets ($\sim 30/\Lambda_{max}$), predictions for the case with $F = 4$ were seen to agree very closely with the actual data for the entire length of testing sets so that the prediction error always remained below the divergence threshold.

F	n	D_{KY}	(r, p, q)	$\Lambda_{max} t_l$	E_{ave}
4	40	15.3	$(400, 1200, 20)$	No divergence observed	4.31
8	40	28.4	$(700, 2100, 20)$	8.16	6.82
16	40	32.1	$(1200, 1200, 40)$	4.05	7.67

resulted in $t_l < 0.2/\Lambda_{max}$.

D. 2D lid-driven cavity flow

The 2D lid-driven cavity flow has been employed for decades as a benchmark for validation of new numerical models and computational schemes (See e.g. [52–54]). Here, we choose this problem as a gateway to the implementation of our Koopman-based method to large-scale fluid flows at high Reynolds numbers. The schematic of the 2D cavity flow is sketched in Fig. 10. The constant-density fluid is confined by a square box whose walls are stationary, except for the top wall (lid), which moves to the right with the velocity $U(x)$. This produces a shear-driven flow mixing the entire fluid via the clockwise primary vortex at the center, as well as some smaller-scale vortices at the corners, if Reynolds number is sufficiently large. The nondimensional equations of motion for this unsteady and incompressible flow are in the following form

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0\,,$$
$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} = -\frac{\partial p}{\partial x} + \frac{1}{\mathrm{Re}}\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right),$$
$$\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} = -\frac{\partial p}{\partial y} + \frac{1}{\mathrm{Re}}\left(\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2}\right),$$

(15)

where u, v and p denote the horizontal velocity, the vertical velocity and the pressure fields, respectively. Characteristic length and velocity are selected as the domain length L
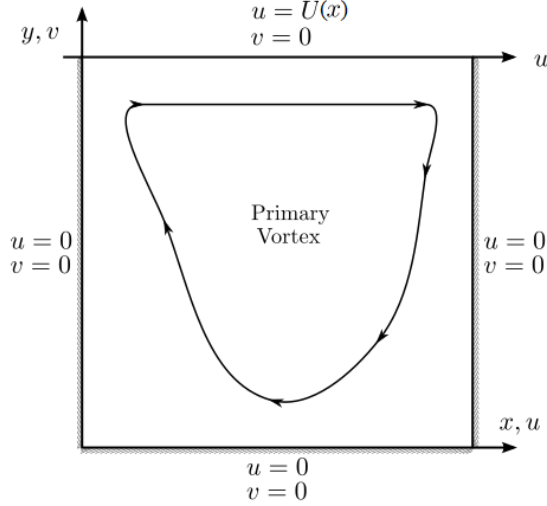
FIG. 10. Schematic of a 2D lid-driven cavity flow. Here, u and v are the horizontal and vertical velocities, respectively, and $U(x) = 16x^2(1-x)^2$.

and maximum lid velocity U_{max}, so that $\text{Re} = U_{max}L/\nu$, with ν indicating the kinematic viscosity of the fluid. This also means that the time is nondimensionalized by the advective timescale L/U_{max}. No-slip boundary conditions are enforced along all walls, except for the lid at which

$$U(x) = 16x^2(1-x)^2 \, , \, v = 0 \, . \tag{16}$$

This boundary condition allows for smooth transitions in top corners, while satisfying the continuity and incompressibility.

   To construct the required training and testing sets, direct numerical simulations (DNS) of the flow based on primitive variables are conducted using a Chebyshev-Chebyshev pseu-dospectral solver with 70 grid points in each direction, and dimensionless time-step $\Delta t = 0.005$. The sufficiency of the number of grid points was examined via mesh refinement. The simulations are then carried out to build the model for the fluid system at two relatively high Reynolds numbers $\text{Re} = 20000$ and $\text{Re} = 30000$. We then sample horizontal and vertical velocities at every other grid point (in total 35 grid points in each direction). The sampling interval and the length of training set for both cases are taken as $\tau \approx 0.025\tau_d$ and $T_{train} = 2000\tau_{adv}$. We then remove their corresponding long time-mean values, $\bar{u}$ and $\bar{v}$, to obtain anomalous velocities $u' = u - \bar{u}$ and $v' = v - \bar{v}$, and finally arrange these mean-removed velocities in 1D vectors $\boldsymbol{u}'$ and $\boldsymbol{v}'$. The state vector $\boldsymbol{V}$ at each time consists

21
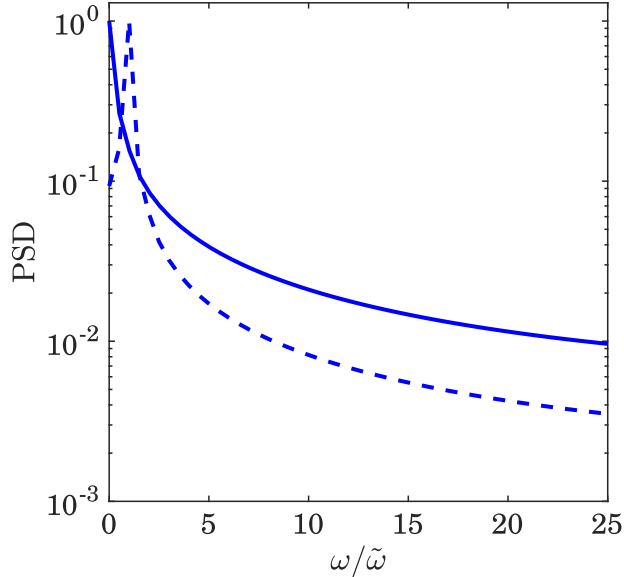
FIG. 11. The Power Spectral Densities (PSDs) of the time series of the PC1 calculated for the velocity field at Re = 20000 (dashed) and Re = 30000 (solid). The PSDs are evaluated by dividing the entire data consisting of 100000 data points ($\sim 2000\tau_{adv}$) into 200 windows with the same length, and carrying out fast Fourier transform (FFT) coupled with Hann filter for each window. The results of all windows are then averaged to obtain the plotted PSDs. Frequency ω is normalized by the frequency of the advective timescale $\tilde{\omega} = 2\pi/\tau_{adv}$.

of these two vectors stacked on top of each other, i.e. $\boldsymbol{V} = [\boldsymbol{u}\ \boldsymbol{v}]^T$. Motivated by the underlying physics of the flow, we choose the Reynolds stress terms at each sampled grid point (u'^2, $u'v'$ and v'^2) to form the nonlinear forcing vector $\boldsymbol{f}$. The results are reported for the normal-to-the-plane vorticity field $\omega = \partial v/\partial x - \partial u/\partial y$, and the PC1 timeseries of the horizontal velocity.

As discussed in Arbabi and Mezić [30], for Re $\leq$ 10000, the cavity flow converges to a steady-state laminar solution whose corresponding attractor is in the form of a fixed point. Upon slight increase in Re above 10000, a periodic flow with a single oscillation frequency emerges. This behavior persists until at Re $\geq$ 15000 another bifurcation occurs and a flow with quasiperiodic behavior (multiple basic frequencies) forms. The third bifurcation occurs around Re = 18000 leading to a rapid rise in the level of kinetic energy. The kinetic energy then continually increases so that at Re $\gtrsim$ 22000, the fluid system becomes fully chaotic and no quasiperiodic compoenents can be further detected. Figure 11 is consistent
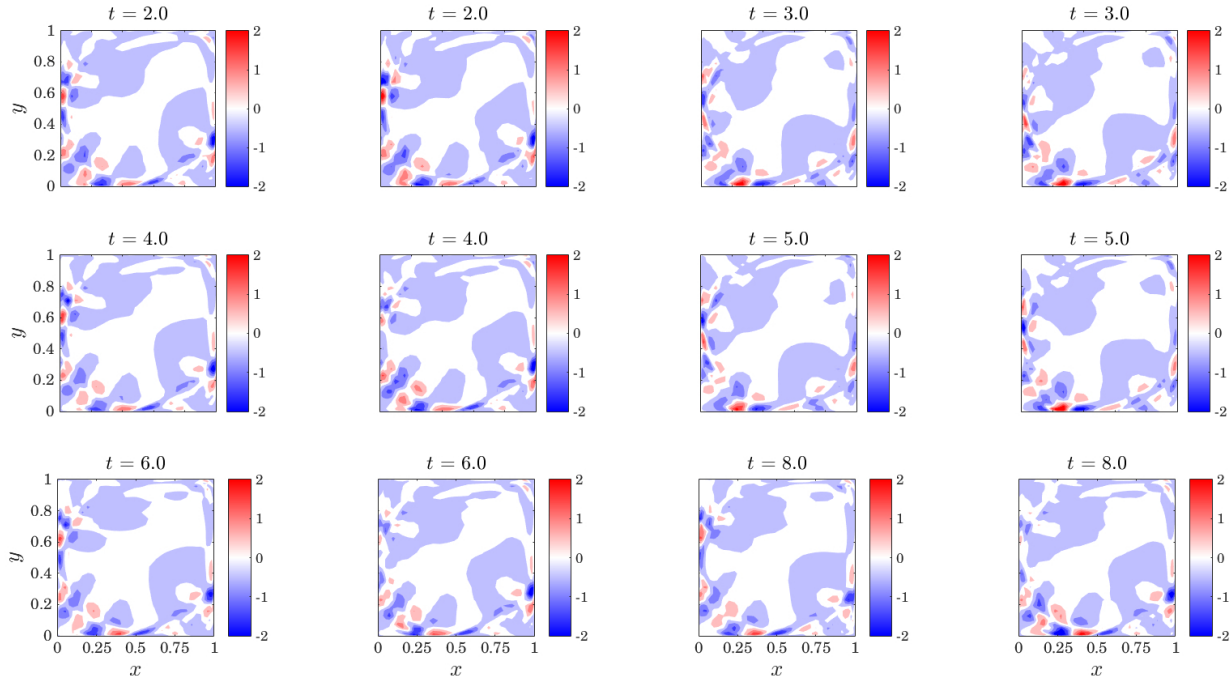
22

FIG. 12. Snapshots of vorticity field for a lid-driven cavity flow at Re = 20000 at six distinct dimensionless times given by DNS solver (first and third columns) and the best predictions of M2 (second and fourth columns). These results were obtained by choosing $(r, p, q) = (400, 600, 3)$, leading to $t_l = 5.2\tau_{adv} = 1040\Delta t$ and $E_{ave} = 16.6\%$. Note that for the bottom panels, data-driven predictions have already deviated from DNS data.

with the findings of Arbabi and Mezić [30] so that for the cavity flow at Re = 20000, the power spectrum calculated for the timeseries of PC1 of the velocity field has a maximum at $\omega/\tilde{\omega} \approx 1$, manifesting that the underlying dynamics is not fully chaotic yet. In fact, according to [30], such flow has a mixed spectrum, i.e. it contains both discrete and continuous components. In contrast, flow at Re = 30000 exhibits a monotonically decaying spectrum, which is indicative of chaotic behaviour.

As demonstrated by Figs. 12 and 13, the Koopman-based method M2 renders accurate predictions for several advective timescales, equivalent to hundreds of DNS time-steps (see the captions of the figures for the exact values), during which the vorticity fields unergo substantial changes. The presence of quasiperiodic components in the flow at Re = 20000 is distinctly illustrated by Fig. 12, e.g. compare the snapshots of the vorticty field at $t = 2$ and $t = 4$ (or at $t = 3$ and $t = 5$), or see the PC1 timeseries shown in Fig. 14(a). The performance of M2 somewhat degrades from Re = 20000 to Re = 30000 as the underlying dynamics
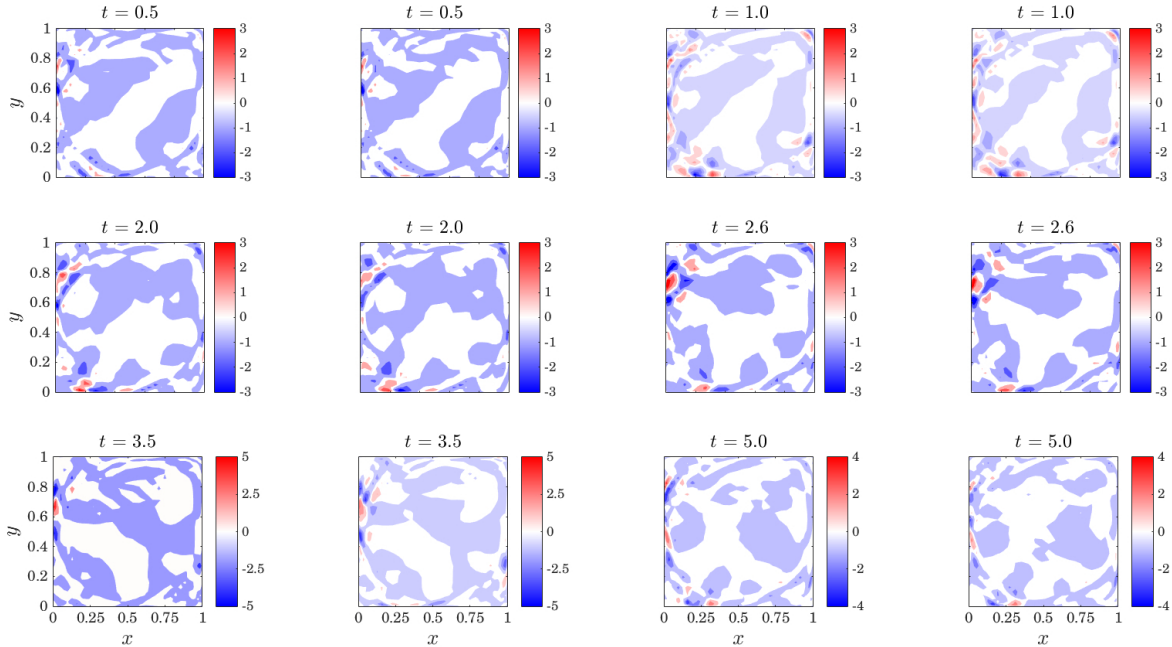
FIG. 13. Similar to Fig. 13, but for a flow at Re = 30000. The optimal method parameters are found to be $(r, p, q) = (1200, 1800, 3)$, for whcih $t_l = 2.7\tau_{adv} = 540\Delta t$ and $E_{ave} = 16.8\%$.

becomes more complex, so that t_l is nearly halved, while E_{ave} varies insignificantly. Exact values are provided in the captions of corresponding figures. These observations are further supported by Fig. 14 in which the timeseries for the PC1 of horizontal velocity obtained from DNS data (solid blue) and M2 predictions (dashed red) are displayed. Nonetheless, the results of M2 for the studied fluid example seem promising, and can lead to a new avenue for predictive modeling of high-dimensional and highly turbulent flows, specifically if some modifications are considered. This prospect will be discussed in detail in Sec. IV.

Similar to chaotic dynamical systems such as K-S equation and Lorenz-96, M2 discovers a few growing modes for the 2D cavity flow, irrespective of its Reynolds number (Figs. 15(b) and (d)), whose growths are suppressed by the existing nonlinearites as discussed in Sec. II A. Moreover, M1 yields rapidly vanishing predictions which diverge from the actual flowfield in less than $40\Delta t$ when Re = 20000 and $20\Delta t$ when Re = 30000, as all eigenvalues detected by this method lie to the left of imaginary axis (Figs. 15(a) and (b)). Note also that the delay-embedding dimension chosen for this problem may seem small, but in fact $q\tau \approx 0.2\tau_d$, which is in the same range as what was found for the dynamical systems discussed earlier in this section.
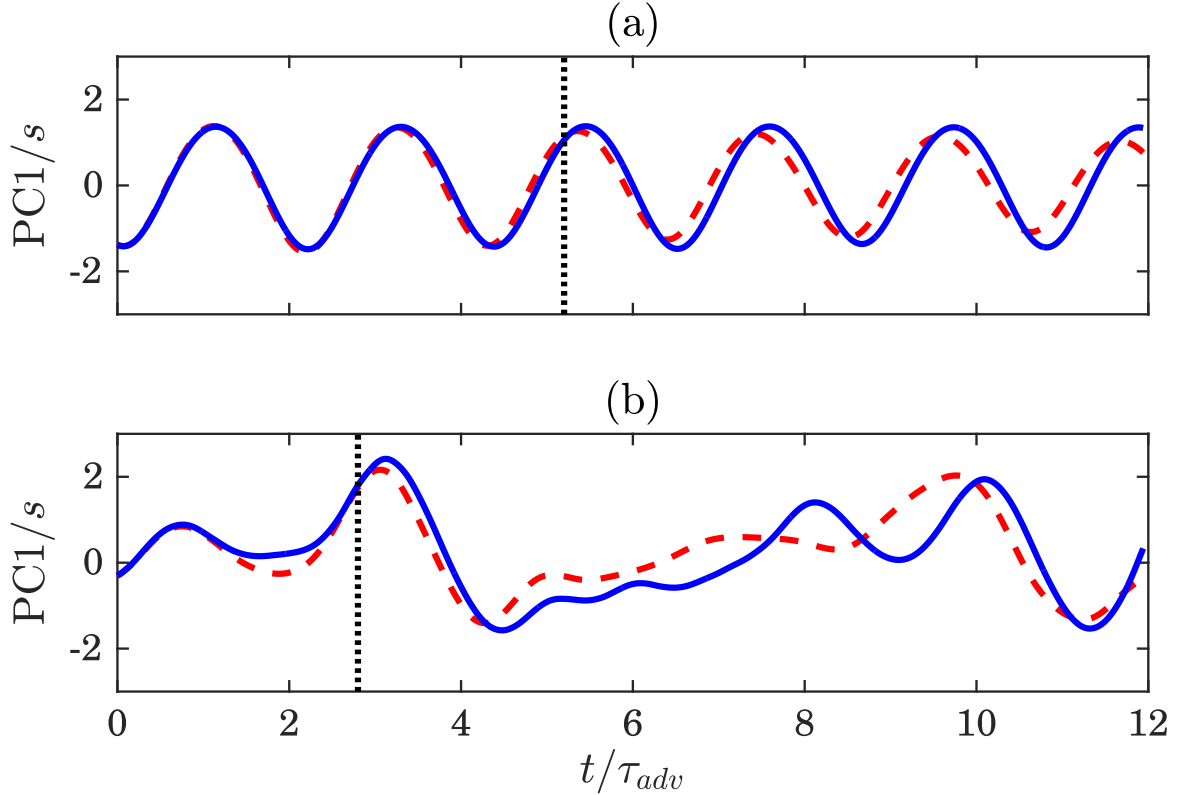
FIG. 14. Timeseries of the PC1 of horizontal velocity, calculated from DNS (solid blue) and predicted by M2 (dashed red), at Re = 20000 (top) and Re = 30000 (bottom). The results are scaled by the standard deviation s of the PC1 timeseries from testing data. The vertical dotted lines correspond to the time at which the predictions of M2 for the entire flowfield (not merely the PC1 of u) diverge from DNS data.

IV. CONCLUSIONS

Within the present investigation, we have proposed a data-driven Koopman-based method, referred to as M2 in the text, which treats the nonlinearities of the system as external actuations, whereas the observables are linear, vector-valued, and time-delay-embedded. Hence, a linear framework (7) is built whose unknown maps are found via the DMDc technique of Proctor *et al.* [45]. This data-driven predictive framework is shown to accurately forecast the spatiotemporal evolution of common examples of chaos such as Lorenz-63, K-S and Lorenz-96 systems as well as a high-Reynolds-number fluid flow, namely, a 2D lid-driven cavity flow, for several Lyapunov or advective timescales, which, in the case of cavity flow, is equivalent to hundreds of numerical solver time-steps.
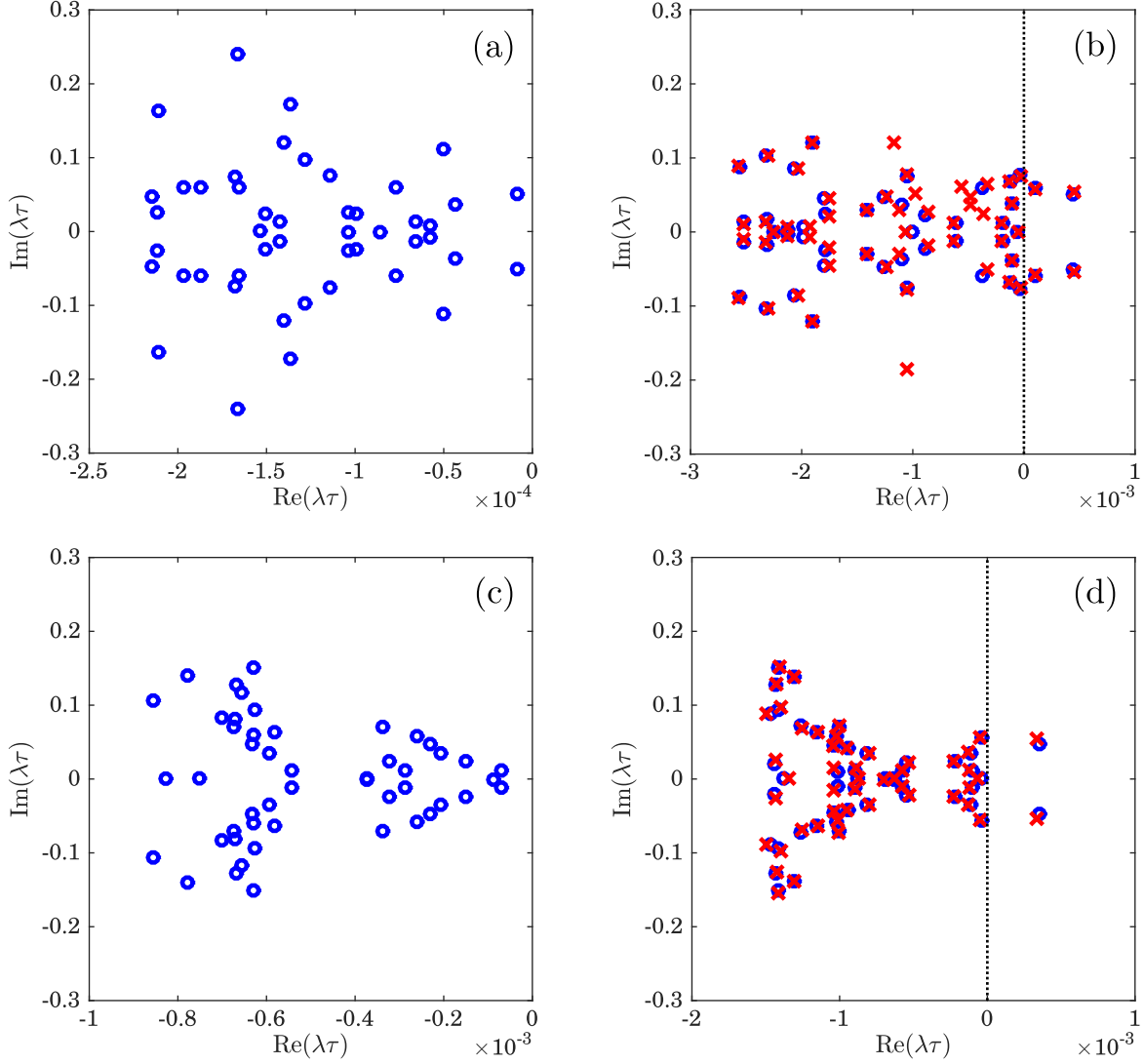
25

FIG. 15. Eigenvalues λ of $\exp(\tau A)$ for the described 2D cavity flow at Re = 20000 (top) and Re = 30000 (bottom), normalized by the sampling frequency $1/\tau$, and calculated via (a) M1 with $(r, q) = (1000, 10)$, (b) M2 with $(r, p, q) = (400, 600, 3)$, (c) M1 with $(r, q) = (1000, 10)$, and (d) M2 with $(r, p, q) = (1200, 1800, 3)$. The blue circles show the eigenvalues provided by the entire training set with the length $T_{train} = 2000\tau_{adv}$, whereas the red crosses in panels (b) and (d) exhibit the eigenvalues obtained using half the training set. The close agreement between the two suggests that the eigenvalues of the systems are captured robustly. Vertical dotted lines in these panels mark the imaginary axis. For clarity, only the first 50 eigenvalues are displayed.

As shown in Secs. II and III, the strong performance of M2 hinges on the simultaneous use of vector-valued, delay-embedded observables and physics-inspired forcings. The resulting linear model is built using the DMDc algorithm of Proctor *et al.* [45]. The advantages of delay-embedding had been shown in previous studies [12, 28, 30, 35–38]. The novelty of M2 is in the last feature, i.e., the accurate representation of the underlying nonlinear processes using a linear model that treats the nonlinearities as exogenous forcings, which builds on the work of Brunton *et al.* [38]. Note that in M2, the forcing terms are updated as the predictions of the new (future) state become available. Such representation of nonlinearities enables us to capture the potentially present unstable modes, whose unbounded growth is suppressed by the energy-conserving nonlinear interplay between the unstable and stable modes [47–49]. It is worth noting that attempts on including such unstable modes in a linear model such as M1 (6) leads to predictions that grow exponentially unboundedly, while attempts on excluding them, which are integral to the spatiotemporal evolution of the system, leads to inaccurate predictions. M2 provides a linear framework for accurately accounting for these unstable modes.

Similar to most data-driven methods, and specifically for very large-scale and high-dimensional systems, the success of the present method to some degree depends on the availability of sufficiently long training sets. This is at least computationally very demanding, if not prohibitive, for three-dimensional and highly turbulent flows. Motivated by the success of Mohan *et al.* [20] in accurately reproducing the long-term statistics of isotropic turbulence, we speculate that this issue might be rectified by initially compressing the 3D turbulence data via methods such as autoencoders from machine learning. The compressed data can then be used for training by M2, whose predictions can later be decoded and brought back to the physical space. Otto and Rowley [55] have also shown that a neural network combining an autoencoder with linear recurrent dynamics can be employed to provide a low-dimensional dictionary of linear and nonlinear observales for the approximation of Koopman operator. The method of [55] has proven to be skillful in identifying the salient dynamical modes, and making short-term predictions for some well-known chaotic dynamics. Successful implementation of these ideas can result in data-assisted surrogate models for computational fluid dynamics (CFD) solvers, that significantly reduce the computational time of these solvers by accelerating the advancement of the flow in time. We aim to pursue these lines of research in the subsequent sudies.

ACKNOWLEDGMENT

Appendix A: Methodical selection of Koopman-based method's parameters

For each case, the optimal choices of delay-embedding dimension q and the size of the reduced subspace r are obtained by a comprehensive search over a broad range of these parameters on validation sets, i.e. datasets which are fully independent from training or testing sets, and are specifically built for finding the optimal parameters. For all studied systems, it was typically seen that choosing $q\tau = O(\tau_d)$ leads to accurate results; the optimal q however was not found to be necessarily equal to τ_d/τ, and in some cases it could be as low as one-fifth of this value. It should be noted that the decorrelation timescale τ_d provides a measure of dynamical system's memory. We also observed that the optimal hard threshold presented in Brunton and Kutz [56] yields a good criterion for the truncation value r. We again highlight that the optimal value of r might be somewhat lower than what is given by this hard thresholding, which might be associated with the shortage of data. Nonetheless, the results are not too sensitive to the choice of r, so long as the selected r is not very far from the threshold. Choosing r significantly larger than this criterion leads to a rapidly diverging model, since many eigenvalues of A fall outside the unit circle in this case. Finally, it was always observed that taking $p = \alpha r$ results in the most accurate predictions, where

α is the ratio of the length of forcing vector to the length of state vector.

[1] E. N. Lorenz, "Deterministic nonperiodic flow," J. Atoms. Sci. **20**, 130–141 (1963).

[2] G. E. P. Box, Jenkins G. M., G. C. Reinsel, and G. M. Ljung, *Time series analysis: forecasting and control* (John Wiley & Sons, 2015).

[3] G. Van Kuik *et al.*, "Long-term research challenges in wind energy - A research agenda by the European academy of wind energy," Wind Energ. Sci. **1**, 1–39 (2016).

[4] T. Duriez, S. L. Brunton, and B. R. Noack, *Machine Learning Control-Taming Nonlinear Dynamics and Turbulence* (Springer, 2017).

[5] A. J. Majda, "Challenges in climate science and contemporary applied mathematics," Commun. Pure Appl. Math. **65**, 920–948 (2012).

[6] P. Bauer, A. Thorpe, and G. Brunet, "The quiet revolution of numerical weather prediction," Nature **525**, 47 (2015).

[7] A. Majda and N. Chen, "Model error, information barriers, state estimation and prediction in complex multiscale systems," Entropy **20**, 644 (2018).

[8] M. Farazmand and T. P. Sapsis, "A variational approach to probing extreme events in turbulent dynamical systems," Sci. Adv. **3** (2017).

[9] C. Wunsch, "The interpretation of short climate records, with comments on the North Atlantic and Southern Oscillations," Bull. Am. Meteorol. Soc. **80**, 245–256 (1999).

[10] H. Van den Dool, *Empirical methods in short-term climate prediction* (Oxford University Press, 2007).

[11] N. R. Cavanaugh, T. Allen, A. Subramanian, B. Mapes, H. Seo, and A. J. Miller, "The skill of atmospheric linear inverse models in hindcasting the Madden–Julian Oscillation," Clim. Dyn. **44**, 897–906 (2015).

[12] D. Giannakis, "Data-driven spectral decomposition and forecasting of ergodic dynamical systems," Appl. Comput. Harmon. Anal. (2017).

[13] D. Comeau, Z. Zhao, D. Giannakis, and A. J. Majda, "Data-driven prediction strategies for low-frequency patterns of North Pacific climate variability," Clima. Dyn. **48**, 1855–1872 (2017).

[14] M. A. Khodkar and P. Hassanzadeh, "Data-driven reduced modelling of turbulent Rayleigh-Bénard convection using dmd-enhanced fluctuation-dissipation theorem," J. Fluid Mech. **852** (2018), 10.1017/jfm.2018.586.

[15] P. L. McDermott and C. K. Wikle, "An ensemble quadratic echo state network for non-linear spatio-temporal forecasting," Stat **6**, 315–330 (2017).

[16] R. Yu, S. Zheng, A. Anandkumar, and Y. Yue, "Long-term forecasting using tensor-train RNNs," arXiv:1711.00073 (2017).

[17] P. R. Vlachas, W. Byeon, Z. Y. Wan, T. Sapsis, and P. Koumoutsakos, "Data-driven forecasting of high-dimensional chaotic systems with long short-term memory networks," Proc. R. Soc. A **474** (2018).

[18] U. Pathak, B. Hunt, M. Girvan, Z. Lu, and E. Ott, "Model-free prediction of large spatiotemporally chaotic systems from data: A reservoir computing approach," Phys. Rev. Lett. **120** (2018).

[19] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for learning forward and inverse problems involving nonlinear partial differential equations," J. Comp. Phys. **378**, 686–707 (2019).

[20] A. T. Mohan, D. Daniel, M. Chertkov, and D. Livescu, "Compressed convolutional LSTM: An efficient deep learning framework to model high fidelity 3D turbulence," arXiv:1903.00033 (2019).

[21] P. L. McDermott and C. K. Wikle, "Deep state networks with uncertainty quantification for spatio-temporal forecasting," Environmetrics **30**, e2553 (2019).

[22] A. Chattopadhyay, P. Hassanzadeh, K. Palem, and D. Subramanian, "Data-driven prediction of a multi-scale lorenz96 chaotic system using a hierarchy of deep learning methods: Reservoir computing, ANN, and RNN-LSTM," arXiv:1906.08829 (2019).

[23] B. O. Koopman, "Hamiltonian systems and transformation in Hilbert space," Proc. Natl. Acad. Sci. **17** (1931).

[24] I. Mezić, "Spectral properties of dynamical systems, model reduction and decompositions," Nonlin. Dyn. **41**, 309–325 (2005).

[25] I. Mezić, "Analysis of fluid flows via spectral properties of the Koopman operator," Annu. Rev. Fluid Mech. **45** (2013).

[26] P. J. Schmid, "Dynamic mode decomposition of numerical and experimental data," J. Fluid Mech. **656**, 5–28 (2010).

[27] C. W. Rowley, I. Mezić, S. Bagheri, P. Schlatter, and D. S. Henningson, "Spectral analysis of nonlinear flows," J. Fluid Mech. **641**, 115–127 (2009).

[28] J. H. Tu, C. W. Rowley, Luchtenburg D. M., S. L. Brunton, and J. N. Kutz, "On dynamic mode decomposition: Theory and applications," J. Comp. Dyn. **1**, 391–421 (2014).

[29] M. O. Williams, I. G. Kevrekidis, and C. W. Rowley, "A data–driven approximation of the Koopman operator: Extending dynamic mode decomposition," J. Nonlin. Sci. **25** (2015).

[30] H. Arbabi and I. Mezić, "Ergodic theory, dynamic mode decomposition, and computation of spectral properties of the Koopman operator," SIAM J. Appl. Dyn. Syst. **16**, 2096–2126 (2017).

[31] H. Arbabi and I. Mezić, "Study of dynamics in post-transient flows using Koopman mode decomposition," Phys. Rev. Fluids **2**, 124402 (2017).

[32] M. Korda and I. Mezić, "On convergence of extended dynamic mode decomposition to the Koopman operator," J. Nonlin. Sci. **28**, 687–710 (2018).

[33] C. W. Rowley and S. T. M. Dawson, "Model reduction for flow analysis and control," Annu. Rev. Fluid Mech. **49**, 387–417 (2017).

[34] F. Takens, "Detecting strange attractors in turbulence," Lect. Notes Math. **898**, 366–381 (1981).

[35] M. Korda and I. Mezić, "Linear predictors for nonlinear dynamical systems: Koopman operator meets model predictive control," Automatica **93**, 149–160 (2018).

[36] H. Arbabi, M. Korda, and I. Mezić, "A data-driven Koopman model predictive control for nonlinear flows," arXiv:1804.05291 (2018).

[37] D. Giannakis and A. J. Majda, "Nonlinear laplacian spectral analysis for time series with intermittency and low-frequency variability," Proc. Natl Acad. Sci. **113**, 3932–3937 (2012).

[38] S. L. Brunton, B. W. Brunton, J. L. Proctor, E. Kaiser, and J. N. Kutz, "Chaos as an intermittently forced linear system," Nat. Commun. **8**, 19 (2017).

[39] S. L. Brunton, J. L. Proctor, E. Kaiser, and J. N. Kutz, "Discovering governing equations from data by sparse identification of nonlinear dynamical systems," Proc. Natl Acad. Sci. **113**, 3932–3937 (2016).

[40] A. C. Ionita and A. C. Antoulas, "Matrix pencils in time and frequency domain system identification," in *Developments in Control Theory: Towards Glocal Control*, Vol. 76, edited by L. Qiu, J. Chen, T. Iwasaki, and H. Fujioka (IET Control Engineering Series, 2012) pp. 79–88.

[41] A. C. Antoulas, I. V. Gosea, and A. C. Ionita, "Model reduction of bilinear systems in the Loewner framework," SIAM J. Sci. Comput. **38**, B889–B916 (2016).

[42] A. C. Antoulas, I. V. Gosea, and M. Heinkenschloss, "On the Loewner framework for model reduction of Burgers' equation," in *Active Flow and Combustion Control 2018* (Springer, 2019) pp. 255–270.

[43] S. Gugercin, C. A. Beattie, and A. C. Antoulas, *Data-driven and interpolatory model reduction* (Society for Industrial and Applied Mathematics, 2019).

[44] L. Pogorelyuk and C. W. Rowley, "Clustering of series via dynamic mode decomposition and the matrix pencil method," arXiv:1802.09878 (2018).

[45] J. L Proctor, S. L. Brunton, and J. N. Kutz, "Dynamic mode decomposition with control," SIAM J. Appl. Dynam. Syst. **15**, 142–161 (2016).

[46] A. Wolf, J. B. Swift, H. L. Swinney, and J. A. Vastano, "Determining Lyapunov exponents from a time series," Physica D **16**, 285–317 (1985).

[47] T. Sapsis and A. J. Majda, "Statistically accurate low-order models for uncertainty quantification in turbulent dynamical systems," Proc. Natl. Acad. Sci. **110** (2013).

[48] A. J. Majda, "Introduction to turbulent dynamical systems for complex systems," in *Frontiers in Applied Dynamical Systems: Reviews and Tutorials* (Springer, 2016).

[49] D. Qi and A. J. Majda, "Low-dimensional reduced-order models for statistical response and uncertainty quantification: Two-layer baroclinic turbulence," J. Atoms. Sci. **73**, 4609–4639 (2016).

[50] J. L. Kaplan and J. A. Yorke, *Functional Differential Equations and the Approximation of Fixed Points*, Lecture Notes in Mathem, Vol. 730 (Springer, 1978) pp. 204–227.

[51] E. N. Lorenz, "Predictability - a problem partly solved," in *Predictability of Weather and Climate*, edited by T. Palmer and R. Hagedorn (Cambridge University Press, 2006) pp. 40–58.

[52] U. Ghia, N. K. Ghia, and C. T. Shin, "High-Re solutions for incompressible flow using the Navier-Stokes equations and a multigrid method," J. Comput. Phys. **48**, 387–411 (1982).

[53] H. B. Schreiber and H. B. Keller, "Driven cavity flows by efficient numerical techniques," J. Comput. Phys. **49**, 310–333 (1983).

[54] M. Sahin and R. G. Owens, "A novel fully implicit finite volume method applied to the lid-driven cavity problempart I: High Reynolds number fow calculations," J. Numer. Meth. Fluids **42**, 57–77 (2003).

[55] S. E. Otto and C. W Rowley, "Linearly recurrent autoencoder networks for learning dynamics," SIAM J. Sci. Comput. **18**, 558–593 (2019).

[56] S. L. Brunton and J. N. Kutz, "Singular Value Decomposition (svd) and Principal Component Analysis (PCA)," in *Data-driven Science and Engineering: Machine Learning, Dynamical Systems, and Control* (Cambridge University Press, 2019) Chap. 1.