Shock trace prediction by reduced models for a viscous stochastic Burgers equation

Cite as: Chaos **32**, 043109 (2022); doi: 10.1063/5.0084955
Submitted: 11 January 2022 · Accepted: 15 March 2022 ·
Published Online: 4 April 2022

Nan Chen,[1,a)] Honghu Liu,[2,b)] and Fei Lu[3,c)]

AFFILIATIONS

[1]Department of Mathematics, University of Wisconsin-Madison, Madison, Wisconsin 53705, USA
[2]Department of Mathematics, Virginia Tech, Blacksburg, Virginia 24061, USA
[3]Department of Mathematics, Johns Hopkins University, Baltimore, Maryland 21218, USA

[a)]chennan@math.wisc.edu
[b)]hhliu@vt.edu
[c)]**Author to whom correspondence should be addressed:** feilu@math.jhu.edu

ABSTRACT

Viscous shocks are a particular type of extreme event in nonlinear multiscale systems, and their representation requires small scales. Model reduction can thus play an essential role in reducing the computational cost for the prediction of shocks. Yet, reduced models typically aim to approximate large-scale dominating dynamics, which do not resolve the small scales by design. To resolve this representation barrier, we introduce a new qualitative characterization of the space–time locations of shocks, named the "shock trace," via a space–time indicator function based on an empirical resolution-adaptive threshold. Unlike exact shocks, the shock traces can be captured within the representation capacity of the large scales, thus facilitating the forecast of the timing and locations of the shocks utilizing reduced models. Within the context of a viscous stochastic Burgers equation, we show that a data-driven reduced model, in the form of nonlinear autoregression (NAR) time series models, can accurately predict the random shock traces, with relatively low rates of false predictions. Furthermore, the NAR model, which includes nonlinear closure terms to approximate the feedback from the small scales, significantly outperforms the corresponding Galerkin truncated model in the scenario of either noiseless or noisy observations. The results illustrate the importance of the data-driven closure terms in the NAR model, which account for the effects of the unresolved dynamics brought by nonlinear interactions.

Viscous shocks are waves with large spatial gradients. They are commonly encountered in nonlinear multiscale systems arising in science and engineering applications. However, an accurate representation of viscous shocks requires resolving a broad range of spatial scales in the dynamics. To alleviate this representation barrier, we introduce a new effective and qualitative characterization of the space–time locations of shocks, named the "shock trace," via a space–time indicator function based on an empirical resolution-adaptive threshold. Unlike exact shocks, shock traces can be captured using only a small number of dominant large-scale modes. Thus, they facilitate the forecast of the timing and locations of the shocks utilizing reduced models. Within the context of a viscous stochastic Burgers equation, we show that a data-driven reduced model, in the form of nonlinear autoregression (NAR) time series models, can accurately predict the shock traces, with relatively low rates of false predictions. The key elements for such a success are data-driven closure terms in the NAR model, which account for the effects of the unresolved dynamics brought by nonlinear interactions.

I. INTRODUCTION

Extreme events occur in many high-dimensional multiscale systems in geophysics, engineering, neural science, and material science.[1–8] Their prediction with uncertainty quantification has significant scientific and societal impacts. However, it can be computationally prohibitive to run these systems for the ensemble prediction that aims to quantify the uncertainty. Reduced models can bring down the computational cost by orders of magnitude compared with that for the original system, providing surrogate models that make the ensemble prediction feasible. In particular, stochastic reduced

TABLE I. Parameter settings of the full and reduced models.

Full model	$\nu = 0.02$	Viscosity constant
	$N = 128$	Number of modes
	$\Delta t = 0.001$	Time step size
	$K_0 = 4$	Number of modes in the stochastic force
	$\sigma = 0.2 \text{or } 1$	Strength of stochastic force
Reduced models	$K = 8$	Number of modes for the reduced models
	$\delta = \Delta t \times 10$	Time step size of reduced models

models have been built to quantify the uncertainty[9–13] and have been applied to forecast extreme events.[2,14–17] Many of these reduced models primarily aim at the statistical forecast of the extreme values, for example, predicting the probability density functions (PDFs). Other reduced models are designed for the short-term prediction of intermittent time series from nature that contain extreme events.[18,19] In addition, statistical and machine learning tools have been developed to predict the onset of extreme events in complex systems; see, e.g., Refs. 20 and 23.

Shock (or shock wave) is a special type of extreme event that is observed in many complex nonlinear systems.[24–26] Shocks have a unique feature that makes them extremely difficult to predict: a shock is characterized by an abrupt change of the state, with the spatial derivative rather than the state itself reaching an extreme value. Thus, the exact representation of a shock requires the information from small or fast scales (see Fig. 1 and discussions in Sec. II C), which are beyond the reach of a reduced model that models the large scales. Yet, the large scales contain rich information about the shocks because they dominate the small scales through the nonlinear interactions. Therefore, it is of particular interest to uncover the relation between the shocks and the large scales and to investigate the possibility of predicting partial information of shocks by the reduced models. As far as we know, despite the importance of predicting the shocks, this issue has not been addressed yet.

This paper shows that reduced models can predict the timing and location of random shocks of viscous stochastic Burgers equations.[27,28] The Burgers equation is a prototype for nonlinear conservation equations that can develop shocks. Here, on top of the nonlinear deterministic dynamics, external stochastic forces are added into the Burgers equation. These stochastic forces are smooth in space and white in time. They play an important role in randomly triggering shocks. An additional small viscous term is further incorporated into the equation such that the random shocks will be dissipated before the appearance of a discontinuity. Therefore, these shocks are known as viscous shocks.[29] In addition to the fundamental mathematical topics such as the existence of the solution and the invariant measure,[27,28,30] the statistics of shocks has been studied for the Burgers equation in various contexts, either with viscosity or in the inviscid limit, with or without stochastic forcing, and possibly subject to additional random initial conditions.[29,31–34] Burgers equation has also frequently served as a testbed for illustrating various model reduction techniques, although usually not in the context of shock tracking; see, e.g., Refs. 35, 38.

To effectively predict the spatiotemporal structures of the shocks, we introduce a new qualitative characterization via a simple indicator function that depends on both time and space. It is named as "shock trace" (see Sec. II C). By prescribing an empirical threshold value, the spatial derivative of the reconstructed spatiotemporal solution from a reduced model is mapped to this indicator function. If the spatial derivative of the forecast solution is more negative than the threshold, then it is mapped to 1, indicating the occurrence of a shock. Otherwise, it is mapped to 0, standing for no shock occurrence at that location. This new characterization has several desirable features in facilitating the prediction of the shocks. First, it mitigates the complexity of the shock forecast as it avoids the forecast of the precise values of the solution. Nevertheless, the simplified representation of the solution via the indicator function preserves the key spatiotemporal structure of the shocks, providing an effectual characterization of the abrupt and sharp changes in the solution. Second, the threshold value that maps the model solution to the indicator function is adaptive to spatial resolutions (i.e., the number of Fourier modes being adopted). Therefore, the same mapping criterion is applicable to the solutions of both the full model and reduced models. Third, the indicator function provides a simple but effective way to quantitatively count the false predictions, allowing a quantification of the forecast uncertainty.

We consider a data-driven reduced model in the form of a nonlinear autoregression (NAR) model.[39,40] The NAR model has three attractive features as a prototype reduced model. First, it includes only the time evolution of the large-scale Fourier coefficients with low wave-numbers, and as a result, it allows a large time step size since there is no stiffness in these large scales. Therefore, it significantly decreases the computational cost, often several orders of magnitude from an accurate full model. Second, as a closure model, the NAR model effectively parameterizes the nonlinear feedback from the unresolved small scales, and its parametric form is derived from a numerical integrator of the system. Third, its parameters can be efficiently estimated by least squares regression from the data (see Sec. III).

Numerical results (see Sec. IV) show that the NAR model can predict the shock traces accurately with low rates of false predictions, almost as good as the corresponding projection of the full model solution. The NAR model significantly outperforms the Galerkin truncated system in predicting the shock traces, in situations with both the true initial condition and noisy observations. In the latter case, an efficient data assimilation scheme is utilized with the NAR model. These results highlight the importance and effectiveness of the data-driven closure that captures the effects of unresolved scales. The key element, the parametric form of these closure terms, is inspired by a Picard iteration scheme within the context of approximating the discrete-time flow map of the resolved dynamics (see Sec. III).

The rest of the paper is organized as follows. Following a brief review of the basic properties of the stochastic Burgers equation and its numerical integration scheme, we introduce the concept of shock trace based on the indicator function with a resolution-adaptive threshold in Sec. II. Section III presents the NAR modeling

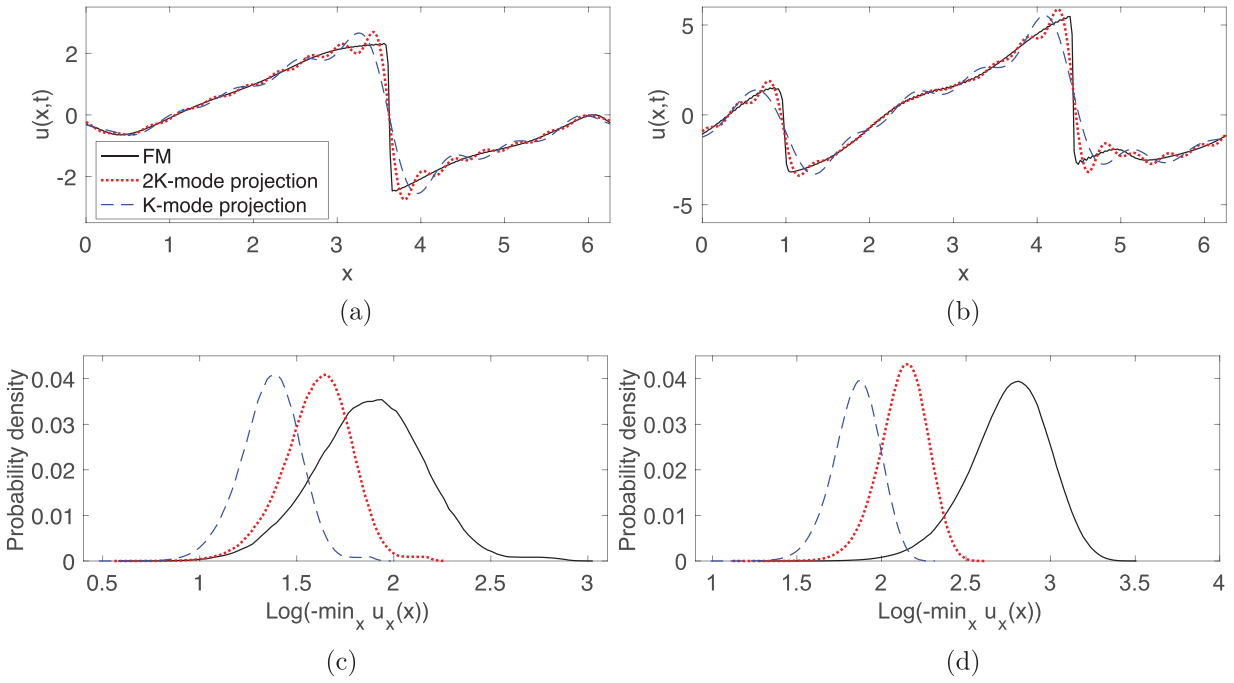


FIG. 1. Typical shocks [panels (a) and (b)] and the empirical densities of most negative derivatives [panels (c) and (d)] of the full model solution (denoted by FM), and its K-mode and $2K$-mode projections with $K = 8$. The full model is simulated with $N = 128$ Fourier modes and the viscosity is $\nu = 0.02$; see Table I for the value of other parameters. Note that neither the K-mode nor the $2K$-mode projection have sharp gradients as the viscous shocks, but their most negative derivatives occur at the same locations as those of the full model solution, i.e., capturing the locations of shocks. We applied logarithm (with base 10) to the absolute value of the most negative derivatives when computing the PDFs for visualization purpose. (a) A typical shock, $\sigma = 0.2$. (b) A typical shock, $\sigma = 1$. (c) PDF of most negative derivatives, $\sigma = 0.2$. (d) PDF of most negative derivatives, $\sigma = 1$.

framework and its parameter inference. The forecast skill of the reduced model is studied in Sec. IV for cases with both weak and strong stochastic forces. Conclusion and some final remarks are then presented in Sec. V.

II. THE VISCOUS STOCHASTIC BURGERS EQUATION AND SHOCK TRACE

A. The viscous stochastic Burgers equation

The model considered in this article is the following viscous stochastic Burgers equation posed on $(0, 2\pi)$ supplemented with periodic boundary conditions and suitable initial condition:[27,28]

$$\begin{aligned} &\partial_t u = \nu \partial_{xx} u - u\partial_x u + f(x,t), \quad 0 < x < 2\pi, \ t > 0, \\ &u(0,t) = u(2\pi,t), \quad \partial_x u(0,t) = \partial_x u(2\pi,t), \quad t \geq 0, \\ &u(x,0) = u_0(x), \quad 0 < x < 2\pi. \end{aligned} \tag{2.1}$$

Here, $\nu > 0$ is the viscosity constant and u_0 is a given square-integrable function. The stochastic force $f(x,t)$ is smooth in space and white in time,[41] acting on a few low-frequency (i.e., large-scale) Fourier modes, which is chosen to be of the form

$$f(x,t) = \sigma \sum_{m=1}^{K_0} \Big(\sin(mx)\dot{W}_m(t) + \cos(mx)\dot{W}'_m(t) \Big), \tag{2.2}$$

where $\sigma > 0$ represents the strength of the stochastic force, $\{W_m, W'_m\}$ are independent Brown motions with $\{\dot{W}_m, \dot{W}'_m\}$ denoting the white noises, and K_0 is a fixed positive integer. With the chosen boundary conditions and the above form of the stochastic force f, the quantity $\int_0^{2\pi} u(x,t)\,dx$ is conserved for all $t \geq 0$. Without loss of generality, we assume that the initial condition has mean zero, which leads to

$$\int_0^{2\pi} u(x,t)\,dx = 0, \quad t \geq 0. \tag{2.3}$$

Equation. (2.1) is interpreted as an infinite-dimensional stochastic differential equation (SDE) for the corresponding Fourier modes,

$$\frac{d}{dt}\widehat{u}_k = -\nu k^2 \widehat{u}_k - \frac{ik}{2} \sum_{l=-\infty}^{\infty} \widehat{u}_l \widehat{u}_{k-l} + \widehat{f}_k(t), \tag{2.4}$$

where $\widehat{u}_k$ are the Fourier coefficients,

$$\widehat{u}_k(t) = \mathcal{F}[u]_k = \frac{1}{2\pi}\int_0^{2\pi} u(x,t)e^{-ikx}dx,$$

$$u(x,t) = \mathcal{F}^{-1}[\widehat{u}] = \sum_{k=-\infty}^{\infty} \widehat{u}_k(t)e^{ikx},$$

with $\mathcal{F}[u]$ being the Fourier transform of u. Here, $\widehat{f}_k(t)$ are white noises consisting of linear combinations of $\{\dot{W}_k, \dot{W}'_k\}$ in (2.2) for $|k| \le K_0$, and $\widehat{f}_k(t) = 0$ for $k > K_0$. Note also that $\widehat{u}_0(t) \equiv 0$ thanks to (2.3).

The above viscous stochastic Burgers equation driven by the stochastic force, which is smooth in space and white in time, has been proven to have an invariant measure in Refs. 27 and 28. We refer the readers to Ref. 30 for the Burgers equation driven by spatiotemporal white noises. Note that the "one force, one solution" principle holds: for each realization of the stochastic force, there exists a unique solution globally and the random attractor consists of a single trajectory, almost surely. For each realization of the force, the shocks form randomly but dissipate before reaching a discontinuity due to the viscosity. Thus, they are called viscous shocks.[29]

B. Numerical integration scheme of the full system and parameter regimes

A high-resolution numerical integration scheme is adopted for the full system (2.1) to generate the true solution. The numerical scheme utilized here is a Galerkin spectral method.[42] More specifically, the function $u(x,t)$ is represented at grid points $x_j = j\Delta x$ with $j = 0, \dots, 2N-1$ and $\Delta x = \frac{2\pi}{2N}$. The Fourier transform $\mathcal{F}$ is replaced by discrete Fourier transform,

$$\widehat{u}_k(t) = \mathcal{F}_{2N}[u]_k = \sum_{j=0}^{2N-1} u(x_j,t)e^{-ikx_j}, \quad k = -N+1, \\ -N+2, \dots, N,$$
$$u(x_j,t) = \mathcal{F}_{2N}^{-1}[\widehat{u}]_j = \frac{1}{2N}\sum_{k=-N+1}^{N} \widehat{u}_k e^{ikx_j}, \quad j = 0,1,\dots,2N-1. \tag{2.5}$$

Since u is real, we have $\widehat{u}_{-k} = \widehat{u}_k^*$, with the superscript * denoting the complex conjugate. Recall also that $\widehat{u}_0$ is identically zero thanks to the mean-zero assumption on the initial data [cf. (2.3)]. Then, by setting $\widehat{u}_{-N} = 0$ to simplify the notations, we obtain from (2.4) the following truncated system:

$$\frac{\mathrm{d}}{\mathrm{d}t}\widehat{u}_k = -\nu k^2 \widehat{u}_k - \frac{ik}{2}\sum_{|k-l|\le N, |l|\le N} \widehat{u}_l \widehat{u}_{k-l} + \widehat{f}_k, \quad \text{with} \\ |k| = 1, \dots, N-1, \text{ or } k = N. \tag{2.6}$$

The system (2.6) is solved using the exponential time differencing fourth-order Runge–Kutta method (ETDRK4) (see Refs. 43 and 44) with the standard 3/2 zero-padding for dealiasing (see e.g., Ref. 45), where the force term $\widehat{f}_k$ is treated as a constant in each time step. Such a hybrid scheme is of strong order 1, but it has an advantage of preserving both the numerical stability of ETDRK4 and the simplicity of the Euler–Maruyama scheme.

Hereafter, for a given $K < N$, we call $u_K(x,t) = \sum_{|j|\le K} \widehat{u}_j(t)e^{-ijx}$ the K-mode projection of the full model solution $u_N(x,t) = \sum_{|j|\le N} \widehat{u}_j(t)e^{-ijx}$, where $\widehat{u}_j(t)$ is solved from the above full model (2.6) with N-pairs of Fourier modes. The reduced model to be introduced later in Sec. III aims to approximate this K-mode projection for a suitable K.

TABLE II. Threshold values in defining shock traces.

Forcing strength	τ_N	τ_{2K}	τ_K
$\sigma = 0.2$	−21.26	−26.02	−15.77
$\sigma = 1$	−320.41	−98.16	−52.22

The threshold values are computed using (2.7) for the full model solution (τ_N), and its $2K$-mode (τ_{2K}) and K-mode projections (τ_K), where $N = 128$ and $K = 8$. The system parameters are those listed in Table I.

The parameters used in the numerical experiments are chosen as follows. The viscosity is set to be $\nu = 0.02$, which is small enough to allow shocks to emerge, when subject to the stochastic force that will be further specified below. Of course, a smaller viscosity constant usually demands a higher spatiotemporal resolution in order to obtain numerically accurate solutions of (2.1) and hence an accurate description of the emergence of the shocks. For the chosen ν, we have set $N = 128$ for the high-dimensional Galerkin system (2.6) and we used $\Delta t = 0.001$ as the time step size. Only the first four pairs of Fourier modes are forced with $K_0 = 4$ in (2.2), and we consider two values for the strength of the stochastic force, $\sigma = 0.2$ and $\sigma = 1$, which correspond to two dynamical regimes exhibiting moderate and strong "turbulent" behavior, respectively. For the regime with larger σ, shocks appear more frequently both in time and space, and the (spatial) gradients presented in the viscous shock profiles become sharper as well; cf. Figs. 1 and 2. It is also worth mentioning that, for both the forcing scenarios, we checked that the mean Courant–Friedrichs–Lewy (CFL) number (computed over a sufficiently long solution trajectory) is well below one: it is 0.045 for the case $\sigma = 0.2$ and 0.139 for the case $\sigma = 1$. For latter reference, we summarize the parameters in Table I, in which we also listed the dimension and time step size used for the reduced models to be presented in Sec. III below. Here, we set $K_0 < K$, i.e., the stochastic force acts only on the first K_0 large-scale resolved modes. In this setting, the closure model as will be constructed in Sec. III mainly accounts for the model error due to nonlinear interactions with the unresolved modes. Thus, it can make meaningful trajectory-wise predictions. In general, the dimension of the closure model can be set to be below the stochastic forcing scales when the goal is only to predict statistical quantities, and it was shown in Ref. 40 that for the same parameter regimes utilized here, when $K_0 = 4$, a closure model with $K = 2$ can already reproduce the associated energy spectrum, marginal probability densities, and temporal correlations for the resolved modes.

C. Trace of the random viscous shocks

A new simple criterion is developed here to qualitatively characterize the occurrence of the random shocks. This criterion is adaptive to spatial resolutions and is, thus, applicable to reduced models as well. The procedure of applying this new criterion to trace the shocks is as follows:

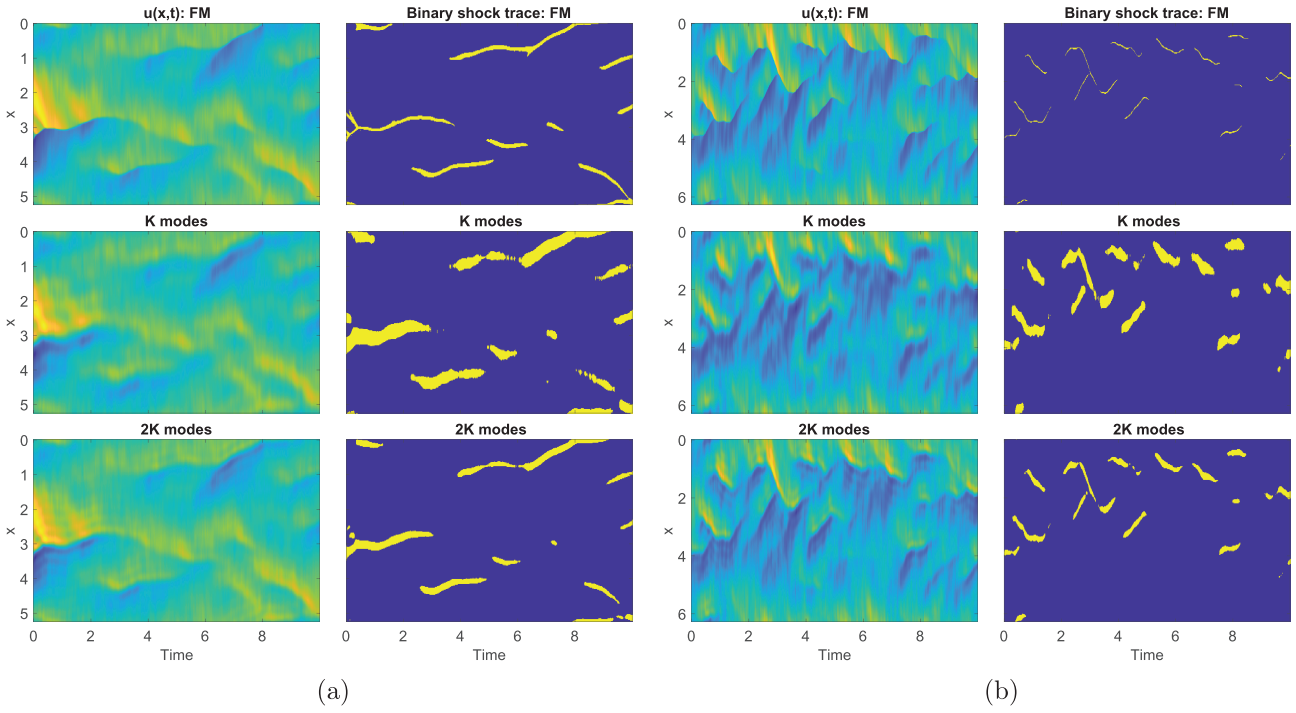


FIG. 2. The full model solution, and its K-mode and $2K$-mode projections, in the space–time plot [left column of panels (a) and (b)], and their binary shock traces (right column). The shock trace indicator function is defined by (2.8), where the involved threshold parameter is given by Table II. The K-model projection already encodes the shock trace information, providing a basis for using reduced models to predict the shock traces. Here, $K = 8$ and the full model has $N = 128$ Fourier modes.

1. Prescribe a statistical threshold value for the spatial derivative of the process to determine the occurrence of shocks: a shock occurs at (x, t) if the spatial derivative is more negative than the threshold value. The threshold value is computed empirically from a large ensemble of sample trajectories of the process.
2. Define an indicator function that identifies the shock trace: it has value 1 at (x, t) if a shock occurs according to the criterion in step 1 above and is 0 otherwise.

This new criterion mitigates the complexity of the shock forecast as it avoids the forecast of the precise values of the solution. Nevertheless, the simplified representation of the solution via the indicator function preserves the key spatiotemporal structure of the shocks, providing an effectual indicator of the abrupt and sharp changes in the solution. Note that the threshold value is adaptive to spatial resolutions (i.e., the number of Fourier modes being adopted). Therefore, the same mapping criterion is applicable to the solutions of both the full and reduced models. In addition, the indicator function provides a simple but effective way to quantitatively count the false predictions, allowing a quantification of the forecast uncertainty, as will be seen in Sec. IV.

We set the threshold value to be 1 standard deviation above the mean of the most negative spatial derivatives, which are empirically computed from a large ensemble of trajectories. This criterion applies to processes with different spatial resolutions, e.g., the k-mode projection $u_k(x,t) = \sum_{|j|\le k} \widehat{u}_j(t) e^{ijx}$ of the full model solution, for any $k \ge 1$. More specifically, we compute the empirical threshold value τ as follows, using this k-mode projection u_k as an example. First, generate a large ensemble of M representative trajectories of the full model whose initial conditions are sampled from a long trajectory of the full model. Denote the k-mode projection of these trajectories by $\{u_k^{(m)}(x,t_l), 0 \le l \le L]\}_{m=1}^{M}$, where $t_l = l\delta$ and $L = \frac{T}{\delta}$ with δ corresponding to the sampling frequency and T the length of the time interval over which the solution trajectories are computed. Second, compute the threshold value τ_k for the k-mode projection as

$$\tau_k = \overline{D}_k + \eta_k, \quad \text{with } \overline{D}_k = \frac{1}{LM}\sum_{l=1}^{L}\sum_{m=1}^{M} D_k^{(m)}(t_l),$$
$$\eta_k^2 = \frac{1}{LM}\sum_{l=1}^{L}\sum_{m=1}^{M} |D_k^{(m)}(t_l) - \overline{D}_k|^2, \quad (2.7)$$

where $D_k^{(m)}(t_l) = \min_x \partial_x u_k^{(m)}(x,t_l)$ denotes the most negative spatial derivative at $t = t_l$ for the mth trajectory of u_k. Note that $\overline{D}_k$ and η_k are, respectively, the mean and standard deviation of $D_k^{(m)}(t_l)$ among all times and all trajectories.

Figures 1(a) and 1(b) show typical shocks presented in the full model solution u_N along with their K-mode and $2K$-mode projections, u_K and u_{2K}, with $K = 8$. Note that neither the K-mode projection nor the $2K$-mode projection capture well the sharp gradients presented in the viscous shocks. Figures 1(c) and 1(d) show the empirical probability distributions of the most negative derivatives $D_k^{(m)}(t_l)$ for u_N, u_K, and u_{2K}. They are computed from $M = 200$ trajectories with length $T = 100$. The full model solution's most negative derivatives are about a magnitude larger than the other two. The shocks have derivatives at the scale -10^3, on the right end of the distribution of the full model solution, which are out of the reach of the low-frequency Fourier modes. Thus, the low-frequency Fourier modes alone cannot represent the shocks. However, the three density functions have similar tails. Together with observation that the K-mode and $2K$-mode projections have large spatial derivatives at the locations of shocks in (a) and (b), they suggest a connection between the most negative derivatives of the full model solution and those computed from its lower-dimensional projections. Our empirical criterion above exploits this connection to detect the trace of shocks.

The empirical threshold values τ_k are shown in Table II. The threshold value increases as the resolution increases in the case of a strong stochastic force. However, in the regime of weak stochastic forcing, due to the large variation of the most negative derivative of the full model [see Fig. 1(c)], τ_{2K} turns out to be more negative than τ_N.

With the threshold value τ_k for a k-mode projection u_k as above, we define a *shock trace indicator* function

$$\mathbf{1}_{S_{u_k}}(x,t) = \begin{cases} 1, & (x,t) \in S_{u_k}, \\ 0, & (x,t) \notin S_{u_k}, \end{cases} \quad \text{with}$$

$$S_{u_k} = \{(y,s) \in [0,2\pi] \times [0,T] : \partial_x u_k(y,s) < \tau_k\}. \tag{2.8}$$

The binary shock trace indicator function has value 1 in the set S_{u_k}, a neighborhood where shock occurs, and is 0 otherwise. It is applicable to both the full model solution and its low-dimensional projections (see Fig. 2, bottom row), as well as the solutions of reduced models to be studied in later sections.

Figure 2 shows that our thresholds defined in (2.7) can detect the trace of the viscous shocks from either the full model solution or its K-mode and $2K$-mode projections. These thresholds, set to be 1 standard deviation above the corresponding mean, are based on an empirical balance between accuracy of detection and robustness of tolerating the random perturbations from the stochastic force. Due to the choice of these thresholds as well as the fact that the projected shock profiles become less steep and span wider as the number of modes used in the projection decreases [see again panels (a) and (b) of Fig. 1], the binary shock traces become wider as the number of Fourier modes of the truncated solution decreases from $2K$ to K (with $K = 8$ here). A more negative threshold (e.g., one standard deviation below the mean) can lead to narrower shock traces, but for the parameter regimes considered, the corresponding shock traces identified by the K-mode and $2K$-mode projections match less good than those identified from the full model solution.

In our numerical tests in Sec. IV, we will consider reduced models with $K = 8$ to show that the proposed closure model is able to faithfully reproduce the shock traces as revealed by the projection of the true solution on the first K modes while a standard K-mode Galerkin truncation fails.

Remark 2.1. *Our empirical threshold provides a preliminary criterion for the detection of shocks for different resolutions. The threshold values are resolution dependent. Here, we set the threshold value to be 1 standard deviation above the mean of the most negative spatial derivatives as in (2.7) for simplicity. One can further refine the definition by multiplying for instance a resolution-dependent factor λ_k to the standard deviation, leading to $\tau_k = \overline{D}_k + \lambda_k \eta_k$. One could also set the threshold to be simply the maximal of the corresponding density function such as shown in panels (c) and (d) of Fig. 1. In general, this empirical threshold should be chosen such that the shock traces of the k-mode projection of the full model solution can accurately capture the space–time locations of the shocks. Thus, this threshold is independent of reduced models. Once calibrated, it is used in computing the shock traces for reduced models of these k modes.*

III. DATA-DRIVEN REDUCED MODEL WITH CLOSURE MODELING

To reduce the computational cost and facilitate an efficient ensemble forecast of the shock traces, a data-driven nonlinear reduced model is developed here that involves the dynamics of only the leading K Fourier modes. The development of such a reduced model is based on a nonlinear autoregression (NAR) modeling framework, which was introduced in Refs. 9, 39, and 40. The key idea in the construction of the NAR model is a parametric approximation of the discrete-time flow map of the leading K modes. In addition to the reduction of the spatial dimension, a much larger numerical integration time step is used in simulating the NAR model, which enhances the overall computational efficiency. It is important to note that although the discrete-time flow map is a time-varying infinite-dimensional functional that depends on the K modes and the trajectory of the stochastic forces, our inference of the NAR model does not suffer from the curse of dimensionality because we make use of the nonlinear structure in the equation to derive an informative parametric approximation of the flow map, as will be seen in Sec. III B.

A. Discrete-time flow map and its approximation

We start with the general framework of the model reduction in terms of flow map approximation, which can be used to construct the reduced models with suitable closure terms.

To simplify notion, we write the stochastic Burgers Eq. (2.1) in an operator form as

$$\frac{du}{dt} + Au = B(u) + f, \; u(0) = u_0, \tag{3.1}$$

where the linear operator $A : H_p^1 \to L^2$ and the nonlinear operator $B : H_p^1 \to L^2$ are defined by

$$A = -\nu\Delta, \quad B(u) = -(u^2)_x/2,$$

and f is the stochastic force. Here, H_p^1 denotes the standard Sobolev space of periodic functions on $[0, 2\pi]$ that are square integrable and whose first-order weak derivatives are also square integrable.

We first decompose the system into the resolved and the unresolved scales (e.g., the low-frequency and high-frequency Fourier modes). That is, we write the solution in the form

$$u = Pu + Qu = v + w,$$

where P and $Q = I - P$ denote the projection operators to the resolved and unresolved scales, respectively. For example, in terms of Fourier expansion $u = \sum_{|k|=1}^{\infty} \widehat{u}_k(t)e^{ikx}$, we set P and Q be the projections from H_p^1 to the low and high wavenumber Fourier modes, with

$$v = \sum_{|k|=1}^{K} \widehat{u}_k(t)e^{ikx}, \quad w = \sum_{|k|=K+1}^{\infty} \widehat{u}_k(t)e^{ikx}. \tag{3.2}$$

With these notations, we can write the system (3.1) as

$$\frac{dv}{dt} = -PAv + PB(v) + Pf + [PB(v+w) - PB(v)], \tag{3.3a}$$

$$\frac{dw}{dt} = -QAw + QB(v+w) + Qf. \tag{3.3b}$$

Equation (3.3a) for the resolved variables v is not closed since it depends on w due to the nonlinear coupling brought by the nonlinear operator B. Thus, a closure model for the resolved-scale variables aims to approximate $[PB(v+w) - PB(v)]$ (see, e.g., Refs. 9, 37, and 46, 48) by a function of v and possible additional noise terms. The Mori–Zwanzig formalism for deterministic systems[39,49–52] shows that an exact closure model involves memory effects (i.e., the dependence on the history of v) and the uncertainties from the unknown initial condition of w. Both the memory effects and uncertainties are difficult to model from physical principles. Therefore, data-driven approaches, combined with physical insights, have led to various constructions of closure models by statistical learning of the discrete-time flow map, either by parametric models[9,40,53] or by nonparametric machine learning representations.[54–56] These closure models can describe the statistical and dynamical properties of the resolved scales, leading to accurate prediction of v with uncertainty being quantified.

The NAR reduced model we adopted here aims to approximate the flow map of v. More precisely, let $t_n = n\delta$ for $n = 1, 2, \ldots$, with δ being one time step. The flow map of v, according to (3.3), takes the form

$$v(t_n) = \mathcal{G}(v(t_{n-1}), w(t_{n-1}), f(s)_{s\in[t_{n-1},t_n)}), \tag{3.4}$$

where $\mathcal{G}$ is a functional forwarding the flow of the resolved variable v from time t_{n-1} to t_n, which depends on the current state $[v(t_{n-1}), w(t_{n-1})]$ and the trajectory of the stochastic forcing on the time interval $[t_{n-1}, t_n)$. We seek a parametric function F to approximate the flow map $\mathcal{G}$ that is independent of the unresolved variable w but, instead, is dependent on the history of the resolved variable v and the stochastic force. For this purpose, we denote by $v^{1:n-1}$ the discrete-time trajectory of v at $t_1, \ldots, t_{n-1}$, and by $f^{1:n}$ the discrete-time trajectory of f at $t_1, \ldots, t_n$. Denote also $v^n = v(t_n)$. We aim to construct F such that

$$v^n \approx F(\theta, v^{1:n-1}, f^{1:n}) + g^n, \tag{3.5}$$

where θ is a multivariate parameter to be estimated from data and the precise form of F is inspired by a Picard approximation of the unresolved variable w as detailed in Sec. III B. The additional process $\{g^n\}$ aims to represent the residual $\mathcal{G}(v(t_{n-1}), w(t_{n-1}), f(s)_{s\in[t_{n-1},t_n)}) - F(\theta, v^{1:n-1}, f^{1:n})$. In this paper, g^n is set to be Gaussian for simplicity, but other forms are also possible, taking the moving average[9,57] or stochastic Stuart–Landau oscillators[58] as examples.

B. The parametric reduced model with closure modeling terms

The construction of a closure model for v of the form (3.5) is through a statistical learning from data, which was originally introduced in Ref. 40. The construction consists of three steps:

1. Deriving a family of parametric functions from the Picard approximation of the high-frequency modes,
2. Estimating the parameters by maximizing the likelihood of the data, and
3. Selecting the model that best fits the data.

1. Derivation of the parametric model

We first derive the parametric function $F(\theta, v^{1:n-1}, f^{1:n})$ in (3.5) from numerical integrators of the full model. In view of Eq. (3.3a), a closure model has to represent the residual $PB(v+w) - PB(v)$ in terms of v and its history. Note that $w(t)$ is a functional of the history of v by integrating (3.3b),

$$w(t) = e^{-QAr}w(t-r) + \int_{t-r}^{t} e^{-QA(t-s)}[QB(v(s)+w(s)) + Qf(s)]ds, \tag{3.6}$$

where $r \in [0, t]$. Given $w(t-r)$ and a trajectory $[v(s), Qf(s), s \in [t-r, t]]$, the Picard iteration provides us an explicit approximation of $w(t)$ as a functional of the trajectory of v. That is, the sequence of functions $\{w^{(l)}\}$, defined by

$$w^{(l+1)}(t) = e^{-QAr}w^{(l)}(t-r) + \int_{t-r}^{t} e^{-QA(t-s)}[QB(v(s)+w^{(l)}(s)) + Qf(s)]ds, \tag{3.7}$$

with $w^{(0)}(s) = 0$ for $s \in [t-r, t]$, converges under suitable conditions to the function $[w(s), s \in [t-r, t]]$ as $l \to \infty$. In particular, the first Picard iteration provides a closed representation

$$w^{(1)}(t) = \int_{t-r}^{t} e^{-QA(t-s)}[QB(v(s)) + Qf(s)]ds. \tag{3.8}$$

While a rigorous convergence analysis of the above Picard iteration is beyond the scope of the current paper, we will show below that the NAR model built from the first iteration (3.8) can already provide good performances. Note that integrals of the form (3.8) also arise naturally in the approximation of center manifolds as mentioned in Remark 3.1.

Now, substituting w in (3.3a) by $w^{(1)}$, we obtain an approximate closed integro-differential equation for v. We then discretize this integro-differential equation in a parametric way to obtain a parametric function that aims to approximate the discrete-time flow map of v. For this purpose, let $\delta = t_n - t_{n-1}$ denote the time step size and let $r = p\delta$ in (3.8). By parameterizing a Riemann sum

approximation of the above integral (3.8) for $w^{(1)}(t)$, we get $w^{(1)}(t_n) \approx \sum_{j=0}^{p} c_j e^{-QAj\delta}[QB(v(t_{n-j})) + Qf(t_{n-j})]$. We denote

$$w^{(1),n} = \sum_{j=0}^{p} c_j e^{-QAj\delta}[QB(v^{n-j}) + Qf^{n-j}], \tag{3.9}$$

where $c_j \in \mathbb{R}$ are parameters to be estimated from data. Furthermore, to keep linear dependence on the parameters, we approximate the term $PB(v+w) - PB(v)$ in (3.3a) as follows by ignoring the nonlinear self-interaction term $Pw\partial_x w$:

$$PB(v + w^{(1)}) - PB(v) \approx -P(\partial_x v w^{(1)} + v\partial_x w^{(1)}).$$

Then, by parameterizing a numerical integrator of (3.3a), we obtain

$$v^n = v^{n-1} + \delta[R^\delta(v^{n-1}) + Pf^{n-1} + \Phi^{n-1}] + g^n, \tag{3.10a}$$

$$\Phi^{n-1} = \partial_x v^{n-1} w^{(1),n-1} + v^{n-1}\partial_x w^{(1),n-1} + \sum_{j=1}^{p}[c_j^v v^{n-j} + c_j^R R^\delta(v^{n-j}) + c_j^f P f^{n-j}], \tag{3.10b}$$

where the nonlinear function $R^\delta(\cdot)$ comes from a numerical integration of the deterministic truncated Galerkin equation $\frac{dv}{dt} \approx -PAv + PB(v)$ at time t_{n-1} and with time step size δ. Here, the extra parameterized terms $[c_j^R R^\delta(v^{n-j}) + c_j^f P f^{n-j}]$ aim to further account for the memory. Note that the high frequency component Qf of the noise does not enter the closure of the low modes because $PQf = 0$ in the linear approximation.

Next, we rewrite (3.10a) and (3.10b) in terms of Fourier modes, and we further parameterize the modes component-wise to account for the different dynamics between modes. Denote $\widehat{v}^n = (\widehat{v}_k^n, |k| \le K) \in \mathbb{C}^{2K}$, the low modes in the reduced model that approximates the original low modes $(\widehat{u}_k(t_n), |k| \le K)$. The reduced model is in the form of a nonlinear autoregression (NAR) model, which reads

$$\widehat{v}_k^n = \widehat{v}_k^{n-1} + \delta[\widehat{R}_k^\delta(\widehat{v}^{n-1}) + \widehat{f}_k^{n-1} + \Phi_k^{n-1}] + \widehat{g}_k^n, \quad 1 \le k \le K, \tag{3.11a}$$

$$\Phi_k^{n-1} = \sum_{j=1}^{p}\left[c_{k,j}^v \widehat{v}_k^{n-j} + c_{k,j}^R \widehat{R}_k^\delta(\widehat{v}^{n-j}) + c_{k,j}^f \widehat{f}_k^{n-j} + c_{k,j}^w \sum_{\substack{|k-l|\le K, K<|l|\le 2K \\ \text{or } |l|\le K, K<|k-l|\le 2K}} \widetilde{v}_l^{n-1}\widetilde{v}_{k-l}^{n-j}\right], \tag{3.11b}$$

where $\widehat{R}^\delta$ comes from the ETDRK4 integrator of the K-mode truncated system, $\widehat{f}_k^n$ and $\widehat{g}_k^n$ denote the kth mode coefficient in the Fourier transform of, respectively, f^n and g^n, and the notation $\widetilde{v}_l^{n-j}$ is defined by

$$\widetilde{v}_k^{n-j} = \begin{cases} \widehat{v}_k^{n-j}, & 1 \le k \le K; \\ \frac{ik}{2} e^{-\nu k^2 j\delta} \sum_{|l|\le K, |k-l|\le K} \widehat{v}_{k-l}^{n-j}\widehat{v}_l^{n-j}, & K < k \le 2K. \end{cases} \tag{3.12}$$

We set $\widehat{v}_{-k}^n = (\widehat{v}_k^n)^*$. Here, the coefficients vary component-wise to allow more flexibility in fitting data.

Equation (3.11) defines a nonlinear autoregression type model when the residual term $\{g^n\}$ is modeled by independent Gaussian noise. One may further improve the model in two directions: (1) include more nonlinear terms by parameterizing higher order Picard iteration than the first iteration; (2) consider spatial correlation between the components of g or by using moving average models.[9,57] We assume for simplicity that the g has independent components, so that the coefficients can be estimated by the least squares fitting.

Remark 3.1. *The parameterization of the unresolved variable w given by (3.8) is closely related to the Lyapunov–Perron integrals arising from the approximation of center manifolds,[59–61] in which the parameter r is pushed to infinite and the true dynamics of v involved in the integral (3.8) is replaced by solutions from a linearized equation; see e.g., Theorem 1 in Ref. 46 and Theorem 6.1 in 59. Treating r in (3.8) as a free parameter to be optimized using solution data leads to deformations of the manifold that aims to approximate the unresolved dynamics in an optimal way. Such a data-driven optimization is the key of the parameterizing manifold approach proposed in Refs. 46 and 59 for dimension reduction of stochastic and chaotic systems. The difference here is that we only utilize the functional form of the parameterization but optimize all of the associated coefficients using the data.*

2. Parameter estimation

The coefficients in (3.11) are estimated by maximizing the likelihood of data. The data can be either a long trajectory or many independent short trajectories. We denote the data consisting of M independent trajectories by

$$\text{Data:} \quad \{\widehat{v}^{1:N_t,m}, \widehat{f}^{1:N_t,m}\}_{m=1}^M = \{(\widehat{u}_k^{(m)}(t_{1:N_t})), (\widehat{f}_k^{(m)}(t_{1:N_t})), k = 1, \dots, K\}_{m=1}^M, \tag{3.13}$$

where m indexes the trajectories and N_t denotes the number of steps for each trajectory.

The maximal likelihood estimator (MLE) of the coefficients is computed by least squares since the model depends linearly on them.[40] To be precise, let us write (3.11b) as

$$\Phi_k^n(\boldsymbol{\theta}_k) = \sum_{j=1}^{4p} \boldsymbol{\theta}_{k,j} \boldsymbol{\Phi}_{k,j}^n,$$

with $\boldsymbol{\theta}_k = (c^v_{k,j}, c^R_{k,j}, c^f_{k,j}, c^w_{k,j}, j = 1, \ldots, p) \in \mathbb{R}^{4p}$ denoting the involved parameters and $\boldsymbol{\Phi}^n_k = (\widehat{v}^{n-j}_k, \widehat{R}^\delta_k(\widehat{v}^{n-j}), \widehat{f}^{n-j}_k, \sum_l \widetilde{v}^{n-1}_l \widetilde{v}^{n-j}_{k-l}, j = 1, \ldots, p) \in \mathbb{C}^{4p}$ denoting the parametric terms. Note that the summation over l in $\boldsymbol{\Phi}^n_k$ is subject to the same constraints as in (3.11b). We compute the MLE as

$$\widehat{\boldsymbol{\theta}}_k = (\mathbf{A}_k)^{-1}\mathbf{b}_k, \quad 1 \le k \le K,$$
$$\widehat{\sigma}^g_k = \frac{1}{M(N_t-1)} \sum_{n,m=1}^{N_t,M} \|\widehat{v}^{n,m}_k - \big(\widehat{v}^{n-1,m}_k + \delta \widehat{R}^\delta_k(\widehat{v}^{n-1,m}_k) + \delta \widehat{f}^{n-1,m}_k + \delta \boldsymbol{\Phi}^{n-1,m}_k(\widehat{\boldsymbol{\theta}})\big)\|^2, \tag{3.14}$$

where the normal matrix $\mathbf{A}_k$ and vector $\mathbf{b}_k$ are defined by

$$\mathbf{A}_k(j', j) = \frac{\delta}{M(N_t-1)} \sum_{n,m=1}^{N_t,M} \langle \boldsymbol{\Phi}^{n-1,m}_{k,j'}, \boldsymbol{\Phi}^{n-1,m}_{k,j} \rangle, \quad 1 \le j', j \le 4p,$$
$$\mathbf{b}_k(j) = \frac{1}{M(N_t-1)} \sum_{n,m=1}^{N_t,M} \langle \widehat{v}^{n,m}_k - \big(\widehat{v}^{n-1,m}_k + \delta \widehat{R}^\delta_k(\widehat{v}^{n-1,m}_k) + \delta \widehat{f}^{n-1,m}_k\big), \boldsymbol{\Phi}^{n-1,m}_{k,j} \rangle. \tag{3.15}$$

In practice, we use pseudo-inverse or regularization to solve the least squares problem in (3.14) when $\mathbf{A}_k$ is ill-conditioned. By fitting parameters to the data of true solution, we obtain the optimal function in the parametric family, correcting the numerical error and model error.

3. Model selection

The model selection step aims to determine the time lag p and remove the redundant terms in the model in (3.11b). We select the simplest model that fits the data the best in the sense that (i) it reproduces the statistics such as the energy spectrum, the marginal invariant densities, and temporal correlations; and (ii) the estimator converges as the data size increases.

For the settings in Table I, we take $p = 1$ to yield the simplest models, i.e., our reduced models are Markovian. Non-Markovian models can improve the results and we refer to Ref. 40 for further discussions on the memory length and parameter convergence. With $p = 1$, the model selection step suggests that we estimate the parameters $(c^v_{k,1}, c^R_{k,1}, c^f_{k,1}, c^w_{k,1})$ for $1 \le k \le K$, along with the variance of the residuals of the K modes. The estimators converge fast as the data size increases, either in the number of trajectories or in the length of a trajectory.

The estimators are shown in Tables V and VI in Appendix B, utilizing data consisting of 512 trajectories, each with length 160 time units and with an initial condition downsampled from a long trajectory. All the estimators depend on the strength of the stochastic force, but they have the following common features. The estimators $(\widehat{c}^v_{k,1}, k = 1, \ldots, K)$ are all negative and are more negative as k increases, so that the corresponding linear terms all dissipate energy. The estimator $(\widehat{c}^R_{k,1}, k = 1, \ldots, K)$ are around 1. These two estimators, together with the estimators $(\widehat{c}^f_{k,1}, k = 1, \ldots, K)$, which are small, act as a calibration of the Euler–Maruyama type approximation in (3.11). At last, the estimator $(\widehat{c}^w_{k,1}, k = 1, \ldots, K)$ are at a smaller magnitude, but they are important for the NAR model to account for the effects of the unresolved scales.

These estimated parameters specify our NAR model to predict the shock trace in Sec. IV.

IV. SHOCK TRACE PREDICTION BY THE REDUCED MODEL

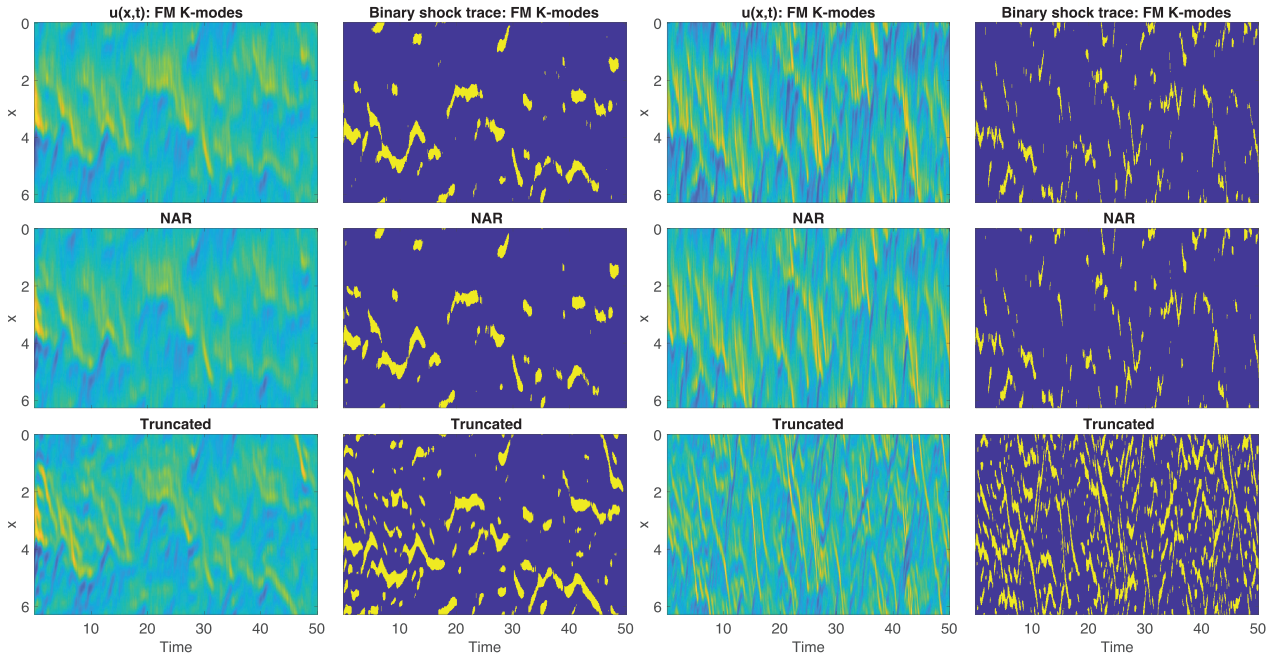
In this section, we demonstrate the skill of the constructed NAR reduced model (3.11) for predicting the binary shock trace in the spatiotemporal solution fields of the stochastic Burgers Eq. (2.1). The accuracy is quantified by computing the rate of false positive (FP) and false negative (FN) events as detailed in Sec. IV A. We consider two prediction scenarios, one with noise-free initial data (Sec. IV B) and the other with noisy data, for which we use data assimilation techniques to estimate the state and make ensemble prediction to quantify the uncertainty (Sec. IV C).

Our choice of the reference shock traces is based on the assessments carried out in Sec. II C. It was shown there that the binary shock traces computed from the eight-mode projection of the true solution already provide decent indication of the timing and locations for the occurrence of shocks in the true solution for both of the two forcing regimes $\sigma = 0.2$ and $\sigma = 1$ within the parameter setup given by Table I; see the second row in Fig. 2. For this reason, we take the binary shock traces computed from the eight-mode projection of the true solution as the ground truth; and we set the first eight Fourier modes as the resolved modes for the NAR reduced model. A truncated Galerkin system of (2.6) for the same resolved modes and the same time step size, referred to as the truncated system below, is used as a comparison. A truncated system with more modes would require a smaller time step size to be stable and a different threshold for shock trace. Thus, in this study, both the NAR model and the truncated system are in exactly the same setting, and the only difference between them is the closure terms. Hence, the NAR's superior performance demonstrates the importance of the closure terms. As will be shown in Secs. IV B and IV C, the NAR reduced model consistently outperforms the truncated system for both the weak noise forcing regime ($\sigma = 0.2$) and the strong noise forcing regime ($\sigma = 1$).

Besides visual comparison of the binary shock trace plots, we provide below a more quantitative way of assessing the performance, which offers statistics that reveals the robustness of the good performance achieved by the proposed NAR reduction framework.

A. Quantification of shock trace prediction skills by false positive/negative rate

We define in this section the false positive and false negative rates in predicting the binary shock traces for a reduced model. For this purpose, we take the binary shock traces computed from a K-mode projection of the true solution as the ground truth (with $K = 8$ here). We get a *false negative* from a reduced model at a space–time location (x, t) when its predicted value is 0 while the truth is 1, and similarly a *false positive* occurs when its predicted value is 1 while the truth is 0. The *false negative rate* (or *false positive rate*) is the ratio between the number of false negatives (or false positives) and the number of true ones.



(a) Stochastic force $\sigma = 0.2$ (b) Stochastic force $\sigma = 1$

FIG. 3. Spatiotemporal fields and the corresponding binary traces of viscous shocks from the NAR reduced model (middle row) and the truncated system (bottom row), in comparison with those from the true K-mode projection of the full model solution with $K = 8$ (top row). The results are from a typical simulation, with a realization of the stochastic force when either $\sigma = 0.2$ or $\sigma = 1$. The rates of false predictions of the binary shock traces are shown in Table III.

More precisely, denote by $u_K(x,t) = \sum_{|k|\le K} \widehat{u}_k(t) e^{ikx}$ the K-mode projection of the true solution, and by v the prediction from a reduced model subject to the same realization of the stochastic force used in the full model. Let τ_K be the binary shock trace threshold associated with u_K defined by (2.7). Following the definition of the binary shock trace in (2.8), we denote the sets $S_{u_K} = \{(x,t) \in [0,2\pi] \times [0,T] : \partial_x u_K(x,t) < \tau_K\}$ and $S_v = \{(x,t) \in [0,2\pi] \times [0,T] : \partial_x v(x,t) < \tau_K\}$. Then, the false negative rate and false positive rate for the prediction v are defined to be

$$R_{\mathrm{FN}} = \frac{|S_{u_K} \setminus (S_v \cap S_{u_K})|}{|S_{u_K}|}, \quad R_{\mathrm{FP}} = \frac{|S_v \setminus (S_v \cap S_{u_K})|}{|S_{u_K}|}, \tag{4.1}$$

where $|\cdot|$ denotes the area of a given set. When these rates are computed numerically, the involved sets and the related areas are approximated through the discrete space–time mesh used for the simulation of the reduced model.

The false negative (FN) rate R_{FN} simply quantifies the percentage of shock traces in u_K that is missed by the reduced model; and the false positive (FP) rate R_{FP} quantifies the relative number of fake shock traces predicted by the reduced model with respect to the total number of shock traces in u_K. Apparently, an accurate prediction would lead to low rates for both false positives and false negatives. Note that these rates depend on the realization of the stochastic force f used as well as the additional stochastic closure terms (if any) in the reduced model. To account for the uncertainty from the randomness, we will compute these rates from many simulations and present them using box charts in Secs. IV B and IV C.

TABLE III. Rates of false predictions.

	Weak forcing $\sigma = 0.2$		Strong forcing $\sigma = 1$	
Rate	False positive	False negative	False positive	False negative
NAR model	0.08	0.17	0.11	0.24
Truncated system	1.06	0.31	2.71	0.55

The reported rates are computed according to (4.1) for the parameter regimes listed in Table I in a typical simulation. See Fig. 3 for the corresponding binary shock trace plots.

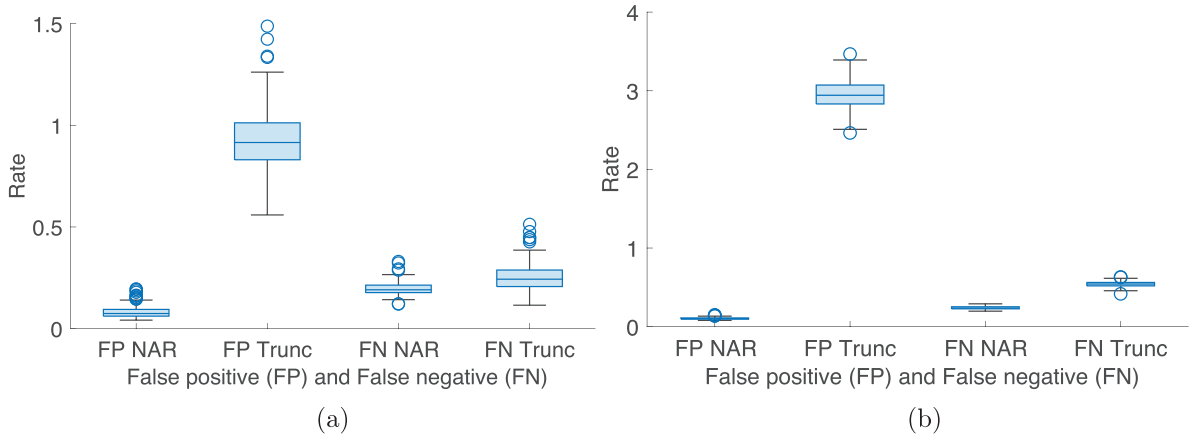


FIG. 4. Rate of false positive and false negative predictions of shocks in box plots. We compare the NAR model and the truncated model with the eight-mode projection of the true solution. The rates are out of 200 simulations. In each box, the central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually in circle marker symbol.

B. Prediction from noiseless observations

In this section, we examine the performance of the NAR reduced model in predicting the shock traces when noiseless observations are used as initial data for the prediction. The case with noisy observation data will be dealt with in Sec. IV C.

The numerical setup is as follows. The parameter regimes are those given by Table I. Both the NAR model and the Galerkin truncated system aim to model the dynamics of the first eight Fourier modes, and the time step δ is set to be ten times that for the full model. These reduced models' initial conditions are projections of the full model's initial conditions, and their stochastic forces are the same as those used for the full model after coarsening to the larger time step δ. Thus, each reduced model produces an approximation to the flow map (3.4) for the first eight Fourier modes, with time step δ.

The NAR model is of the form (3.11), with the involved parameters trained according to the procedure described in Sec. III B 2. We have also set $p = 1$ in (3.11). Thus, no memory terms are involved here (see Sec. III B 3 for further discussions). The additional noise term g^n in (3.11) is modeled simply by a Gaussian fitting to the residual. In terms of computational cost, aside from the offline training stage for inferring the involved parameters, the online simulation time used by the NAR model is only slightly longer than that of the truncated model due to the additional closure terms that are absent in the truncated system, but the NAR model's computational cost remains orders of magnitude smaller than the full model due to its lower space dimension and larger time step size.

TABLE IV. Fraction of energy in the unresolved modes. These numbers are computed based on a typical realization of the full model solution over a 400 time-unit window. The fraction of energy in the unresolved modes at each time t is computed via $\|u(\cdot,t) - u_K(\cdot,t)\|^2/\|u(\cdot,t)\|^2 \times 100\%$, where $\|\cdot\|$ denotes the (spatial) L^2-norm and $K = 8$.

Forcing strength	Mean	Standard deviation	Outliers
$\sigma = 0.2$	3.26%	2.04%	Above 10%
$\sigma = 1$	6.09%	3.00%	Above 20%

Figure 3 shows the prediction results in a typical simulation. We present the spatiotemporal fields obtained from the two reduced models and compare them with the eight-mode projection of the true solution field. Also shown are their binary shock traces computed according to (2.8) with u_k therein taken to be the spatiotemporal field from either the eight-mode projection of u (top row) or the NAR reduced model (middle row) or the truncated model (bottom row). The threshold τ_k in (2.8) is taken here to be the τ_K given by Table II. We also report in Table III at a more quantitative level the visual results given in Fig. 3, using the false negative/positive rates defined in Sec. IV A.

For both the weak force regime ($\sigma = 0.2$) and the strong force regime ($\sigma = 1$), the spatiotemporal field and the binary shock traces of the projected true dynamics are very well reproduced by the NAR model. In contrast, the truncated system already performs visibly less good for the weak force regime, and its predictive skill dramatically decreases in the strong forcing regime. The poor performance of the truncated system is because it does not include any closure term to account for the impact from the unresolved modes. Such closure terms become increasingly important as the forcing strengthens. Indeed, for both forcing regimes considered, the unresolved modes still retain a significant amount of energy, especially for the strong force regime; see Table IV.

The robust superior performance of the NAR model is further confirmed by the statistics of the false negative and false positive rates in multiple simulations. For this purpose, we run the above simulation of the full model and the reduced models for 200 different realizations of the solution of the full model, where the initial conditions are sampled from a long trajectory of the full model. Figure 4 presents the box plots of the false positive and false negative

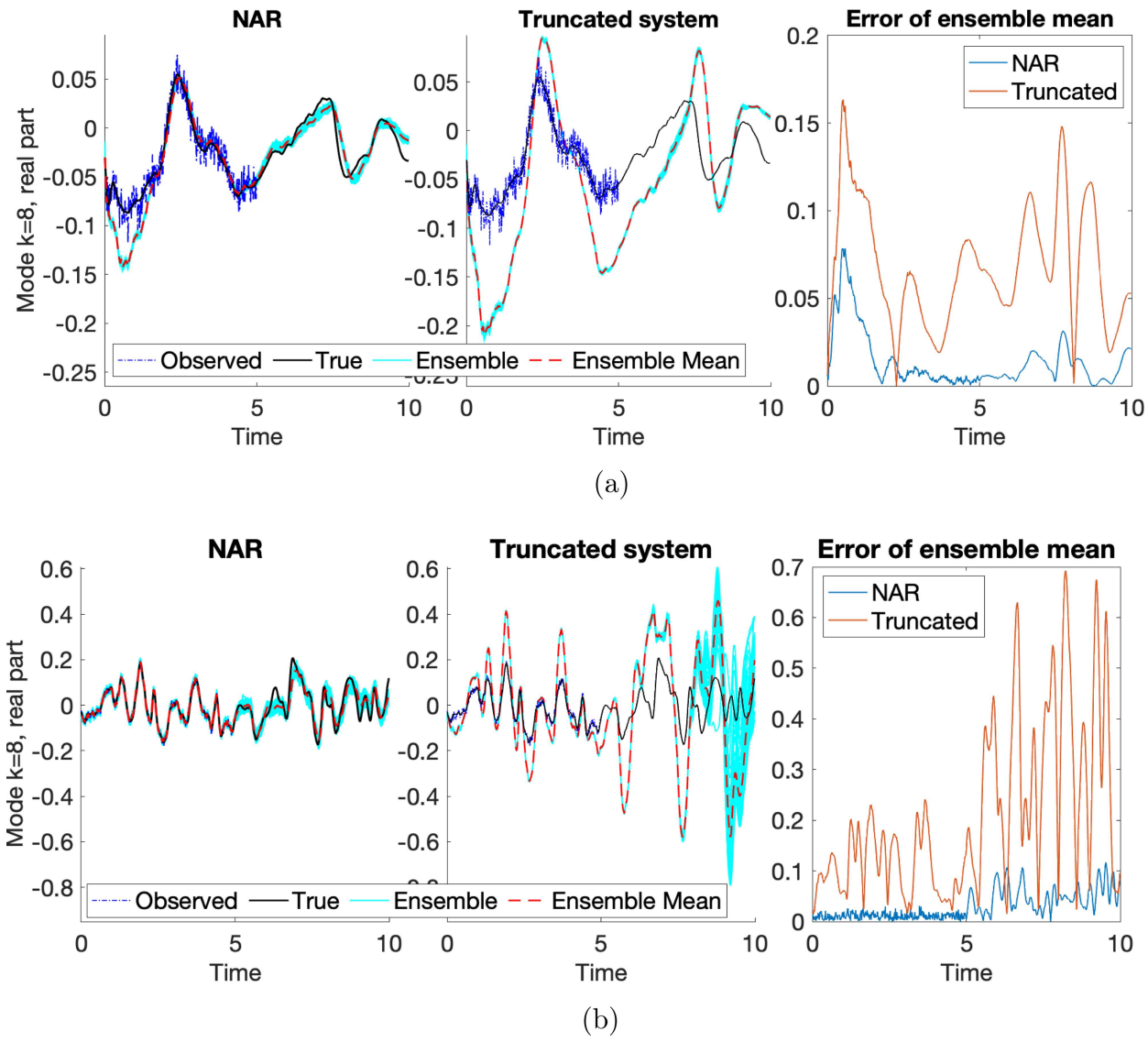


FIG. 5. Ensemble trajectories of the real part of the eighth Fourier mode in the assimilation stage ($t \in [0, 5]$) and the prediction stage ($t \in (5, 10]$) compared with the corresponding true dynamics. In both stages and in both cases of weak ($\sigma = 0.2$) and strong ($\sigma = 1$) stochastic forces, the NAR model has ensemble trajectories closer to the true signal (left column) than the truncated system (middle column), leading to smaller errors in the ensemble mean estimator shown in the right column.

rates associated with these predictions. The NAR reduced model has significantly smaller false rates than the truncated system for both forcing regimes, and the variation of these rates is also significantly smaller; see the caption of Fig. 4 for more details.

C. Data assimilation and prediction from noisy observations

Noisy observations are commonly seen in practice due to either measurement or sampling errors presented in the data collection process. When the observations are noisy, the shock trace prediction consists of two stages: a data assimilation stage to estimate the initial conditions by filtering and a prediction stage that advances forward in time from the estimated initial conditions. In the data assimilation stage, we filter out the noise in the observations by the ensemble Kalman filter (EnKF)[62,63] (see Appendix A for a brief review), with the NAR reduced model and the truncated system as the forecast model, respectively.

1. Test settings

For each realization of the true solution, the corresponding shock trace prediction experiment consists of first performing data

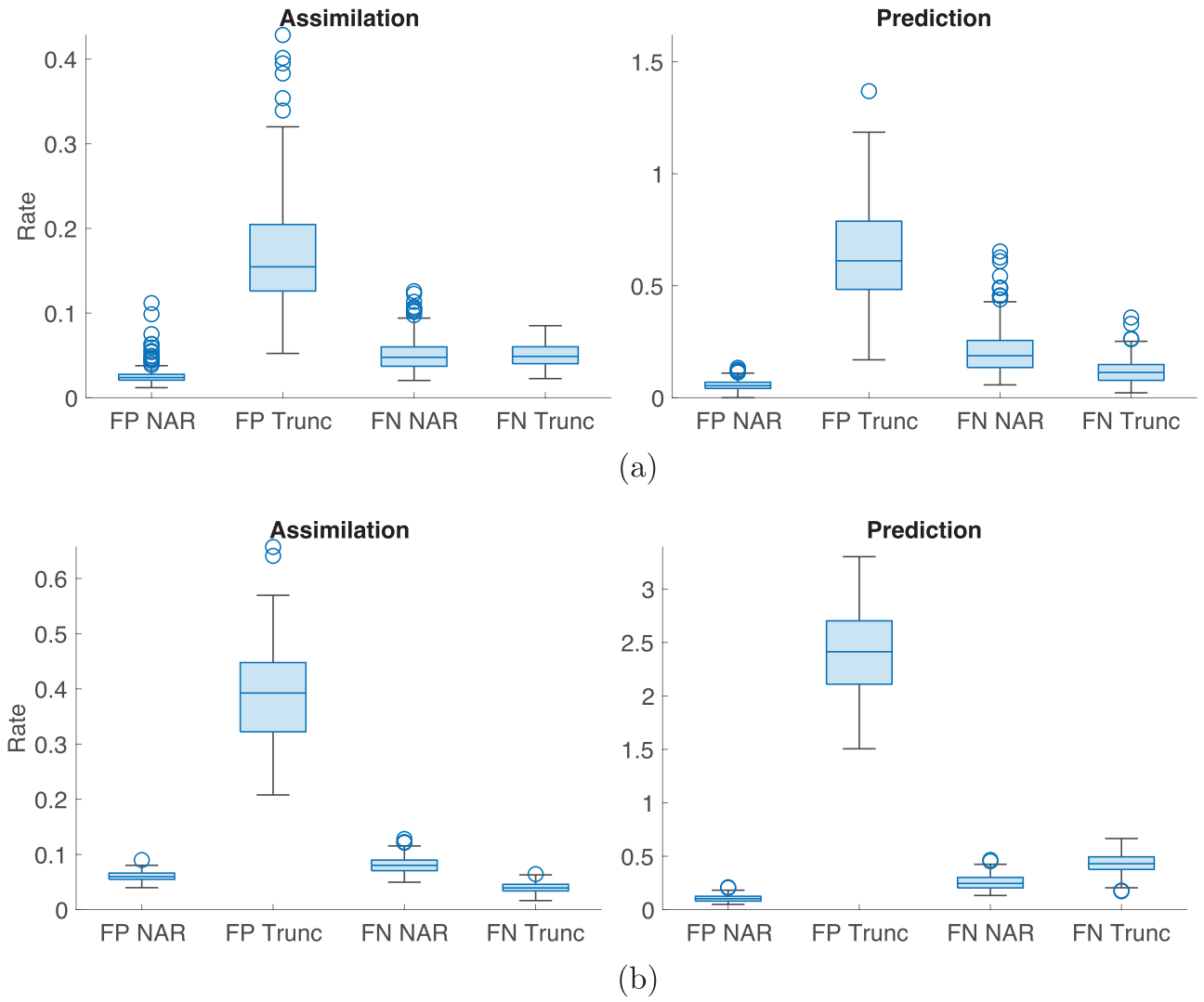


FIG. 6. Rate of false positive (FP) and false negative (FN) in the prediction of the binary shock traces obtained from the NAR model and the truncated system. These rates are computed from the prediction for 200 realizations of the true solution. In each prediction experiment, the spatiotemporal fields for the NAR model and the truncated system used in computing the false rates are taken to be the corresponding ensemble mean of the EnKF with 100 particles. See also the caption of Fig. 4 for the meaning of each element in the box plots.

assimilation using a reduced model (either the NAR model or the truncated system) in the time interval $[0, 5]$, and then carrying out prediction using the same reduced model over the time interval $(5, 10]$. Thus, we have 500 time steps in both stages since the time step is set to be $\delta = 0.01$ for the reduced model.

To generate noisy observations for the data assimilation, we add independent Gaussian noise to the first eight Fourier modes' real and imaginary parts with a standard deviation of 0.01 and mean of 0 (while preserving that $\widehat{u}_{-k} = \widehat{u}_k^*$). These noisy perturbations are relatively large, with the ratio between this standard deviation 0.01 and the mean absolute values of the first to the eighth mode of the true dynamics ranging from about 7% to 40% for the weak stochastic force regime ($\sigma = 0.2$) and from 2% to 14% for the strong stochastic force regime ($\sigma = 1$). This ratio increases as the wave number k increases because the mean absolute value of the Fourier mode decreases as k increases.

In the data assimilation stage, the forecast model of the EnKF (either the NAR model or the truncated system) is initialized from an ensemble of 100 initial conditions randomly sampled from a Gaussian distribution centered at the noisy observation at time $t = 0$ with standard deviation 0.025, which is slightly larger than the standard deviation of the observation noise. In the prediction stage, we

simply simulate the forecast model with the end point from each of the assimilation ensemble as the initial condition. Throughout this section, the NAR model is the same as the one used in Sec. IV B, i.e., the NAR model is trained offline from noiseless data.

2. Performance comparison for a typical prediction

As in Sec. IV B, we first compare the performance for a typical realization of the true solution. For this purpose, we show the ensemble trajectories in the data assimilation and the prediction stages. The results are presented in Fig. 5 for the real part of the mode with wave number $k = 8$, in the settings of a weak stochastic force (top row, $\sigma = 0.2$) and a strong stochastic force (bottom row, $\sigma = 1$).

At the data assimilation stage ($t \in [0, 5]$), the EnKF ensembles of the NAR model are much closer to the true trajectory than the truncated system's. Also, the NAR model's ensemble mean estimators have errors less than the observation noise's standard deviation after a short time period, whereas the truncated system's ensembles deviate far away from the truth; see the right column of Fig. 5. The NAR model's prediction ensembles continue to be spread around the true trajectory, whereas those of the truncated system struggle to make a reasonable prediction for both $\sigma = 0.2$ and $\sigma = 1$.

Similar superior performances are observed for the NAR model in the ensemble prediction of other modes with wave number k less than 8 at both the data assimilation stage and the prediction stage. The truncated system's prediction skill gradually improves as k decreases, because the signal to noise ratio improves as k decreases, but the skill remains inferior to that of the NAR model.

Finally, it is worth pointing out that we set the prediction interval to be (5, 10] here just for illustration purpose. Since the full model is not chaotic, and the same realization of the stochastic force f is used in both the full model and the reduce systems, one can expect that the good prediction skill of the NAR model shown here to still hold for even longer prediction intervals.

3. Performance comparison in multiple simulations

To assess the robustness of the NAR model, we repeat the above simulation for 200 different realizations of the true solution and present the performance using again box plots for the rates of false predictions in Fig. 6. The procedure of computing the false positive and false negative rates is the same as before but now carried out separately for the assimilation stage and the prediction stage. For each realization of the true solution, the spatiotemporal fields for the NAR model and the truncated system used in computing these rates are taken to be the corresponding ensemble mean from simulations with 100 different initial conditions, which are sampled in the same way adopted to produce Fig. 5.

The NAR model has significantly smaller rates of false predictions than the truncated system in both stages and for both regimes of the stochastic forces (see Fig. 6). In particular, the improvement is significant in the prediction stage, reducing the median rates of false positive from 0.61 to 0.05 in the weak forcing case and from about 2.41 to less than 0.10 in the strong forcing case. In the assimilation stage, the improvement is less significant, reducing the median rates of false positive from 0.15 to 0.02 in the weak forcing case and from 0.39 to 0.06 in the strong forcing case. The assimilation stage has smaller rates than the prediction stage because of the information supplied by the observation data.

We have, thus, illustrated that the NAR model is able to predict accurately the pathwise behavior, i.e., the timing and location, of the shocks with uncertainty quantification. This contrasts to predicting only the long-term statistics of the fat-tailed PDFs, which is often the focus of forecasting many other extreme events using reduced models. While the same realizations (after coarsening) of the stochastic force as the full model were applied to the reduced models for diagnostic purpose, the accurate pathwise performance offered by the NAR model presented above shows that even when the stochastic force is unknown in practice, the prediction provided by the K-mode NAR model would be comparable with those obtained from the K-mode projection of the prediction computed from the full model itself.

V. DISCUSSION AND CONCLUSIONS

To summarize, this work shows that a reduced model with systematic closure terms, namely, a nonlinear autoregression (NAR) model, can accurately predict the timing and locations of random shocks for the viscous stochastic Burgers equation, even though the reduced model cannot represent the precise shocks. A key element is a new characterization of shocks, called *shock trace*, which is a binary indicator that reveals the timing and location of the occurrence of a shock. The shock trace is defined via empirical resolution-adaptive thresholds.

The NAR model describes the evolution of only the leading K Fourier modes with K being a friction of the full model and with a time step size ten times the full model. Thus, it reduces the computational cost by orders of magnitude. The NAR model predicts the shock trace almost as good as those computed from the K-mode projection of the full model's solution. The NAR model consistently outperforms the corresponding Galerkin truncated system in forecasting the shock trace from both noiseless and noisy observations. Thus, the data-driven closure in the NAR model plays an essential role in its success of shock trace prediction.

We conclude by reiterating some key elements that contributed to the success of the closure model utilized here and also mention potential future works.

Data representation and importance of closure terms. The reduction and prediction approach presented here is independent of the orthogonal basis adopted to decompose the solution field. Instead of the Fourier basis, one could use for instance empirically computed modes obtained from, e.g., the proper orthogonal decomposition method[64] and its variants[65] or the dynamic mode decomposition.[66,67]

The superior performance of the NAR model over the truncated system demonstrates the importance of incorporating suitable closure terms into the reduced model in order to effectively represent the nonlinear feedback from the unresolved-scale variables to the resolved ones. Such nonlinear feedback effects are important to be properly approximated here since a significant amount of energy in the shocks can be contained in the high-frequency variables, which are not resolved by a low-dimensional reduced system (cf. Table IV). Of course, besides the NAR model, many other

closure modeling techniques can potentially be used for this purpose; see, e.g., Refs. 39, 46, and 68, 71.

Predicting key information of extreme events by reduced models. By relaxing the goal from predicting the full shock profiles to predicting only their space–time locations, we arrive at a task within the representation capacity of a low-dimensional reduced subspace, which is thus much more accessible by a reduced-order modeling approach. The indicator function (2.8) introduced here for identifying binary shock traces is based on a simple resolution-adaptive threshold defined by (2.7), which is computed empirically using an ensemble of the true solution trajectories or their projections onto a given number of Fourier modes. These binary shock traces encode the shock locations in space–time, providing thus important partial information about the shocks.

While we illustrated the approach on the stochastic Burgers equation, it is expected that the developed strategy will be of interest for data-driven predictive modeling of a broad range of extreme events, particularly those beyond the representation capacity of a reduced-order modeling framework. Note that the concerned extreme events do not have to be the extreme values[1,2,21,23] of the states of the model; the shocks correspond to extreme values in the spatial gradient field instead. It is reasonable to expect that the representation capacity offered by a fixed subspace will be violated more often when higher derivatives are involved in defining the concerned extreme events.

Prediction of the exact shocks. Going beyond shock trace prediction, if one would wish for instance to approximate the sharp gradients presented in the shock profiles, it would require either to resolve more Fourier modes, or switch to a data-adaptive empirical basis such as those mentioned above, or adopt a combination of good low-dimensional reduced models with additional techniques to recover the unresolved high-frequency modes. A few possibilities are available for the latter option depending on the setup.

For instance, when observations are only available for the low-frequency modes, one can design computationally efficient data assimilation strategies within the conditional Gaussian framework[72,73] to approximate the dynamics of the high-frequency modes with quantified uncertainties by a suitable dynamical model for the unresolved modes. On the other hand, one can parameterize the high-frequency modes using suitable random functions of the resolved low-frequency dynamics. These random functions can be designed for instance through the dynamics-based parameterizing manifold approach[46,74] or other data-driven and machine learning approaches. When good parameterizations of the unresolved modes are available, they can be used to both build reduced models for the resolved modes and provide approximations of the unresolved modes.

ACKNOWLEDGMENTS

The authors are grateful to the anonymous reviewers for their valuable feedback that helped significantly improve the manuscript. The research of N.C. was partially funded by the Office of VCRGE at UW-Madison and the ONR (No. N00014-21-1-2904). The work of H.L. was partially funded by the NSF (Award No. DMS-2108856). The work of F.L. was funded by the NSF (Award Nos. DMS-1821211 and DMS-1913243).

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

APPENDIX A: THE ENSEMBLE KALMAN FILTER

The Ensemble Kalman Filter (EnKF) is a Monte Carlo implementation of Bayesian filtering with the Kalman filter update.[62,75–77] It uses an ensemble of random samples, also called particles, to approximate the forecast and analysis distributions by Gaussian distributions whose means and covariances are given by ensemble means and covariances. Among various EnKF algorithms, we consider the version with perturbed observations, introduced in Refs. 62 and 77, and we refer to Ref. 78 for a comparison of different versions of EnKF algorithms.

Suppose the filter uses a forecast model

$$\mathbf{x}_n = \mathbf{F}_n(\mathbf{x}_{n-l:n-1}), \tag{A1}$$

where $\mathbf{x}_n \in \mathbb{R}^{d_x}$ is the state variable, $\mathbf{x}_{n-l:n-1} = (\mathbf{x}_{n-l}, \ldots, \mathbf{x}_{n-1})$, and $\mathbf{F}_n$ is a forecast operator at time n which maps $\mathbb{R}^{l\times d_x}$ to $\mathbb{R}^{d_x}$ with $1 \le l \le n-1$. The forecast model can be either stochastic or deterministic, and either Markovian (e.g., $l = 1$) or non-Markovian (e.g., $l > 1$). The state variable is observed through a linear observation operator with Gaussian noise,

$$\mathbf{z}_n = H\mathbf{x}_n + \epsilon_n,$$

TABLE V. Parameters in the NAR model: weak stochastic force $\sigma = 0.2$.

k	1	2	3	4	5	6	7	8
c^v	−0.01	−0.05	−0.12	−0.26	−0.73	−1.28	−2.02	−2.68
c^R	1.08	1.04	1.00	0.93	0.93	0.83	0.65	0.26
c^f ($\times 10^{-3}$)	−0.12	−0.47	−1.07	−1.90	−0.00	−0.00	−0.00	−0.00
c^w ($\times 10^{-5}$)	−1.38	4.59	−3.41	−4.47	−5.48	−5.37	−0.00	−5.91
$\widehat{\sigma}_g$	0.04	0.13	0.23	0.32	0.44	0.55	0.70	0.93

TABLE VI. Parameters in the NAR model: strong stochastic force $\sigma = 1$.

k	1	2	3	4	5	6	7	8
c^v	−0.05	−0.23	−0.54	−1.07	−2.71	−4.49	−6.84	−9.20
c^R	1.08	1.02	0.97	0.87	0.88	0.78	0.64	0.35
c^f ($\times 10^{-3}$)	−0.26	−1.05	−2.39	−4.25	0.00	−0.00	0.00	−0.00
c^w ($\times 10^{-6}$)	−0.70	−6.88	−1.87	−4.57	−1.84	−2.27	−1.43	−7.74
$\widehat{\sigma}_g$	0.92	2.06	3.19	4.08	5.12	5.96	7.28	9.74

where $H \in \mathbb{R}^{d_z \times d_x}$ is the observation matrix, and the $\epsilon_n \sim N(0, R)$ are independent Gaussian noises. In this study, we assume that the observation matrix R is known.

The EnKF iterates the following two steps, with an initial ensemble of particles $\{\mathbf{x}_0^{a,(i)}, i = 1, \ldots, M\}$ sampled from the forecast distribution of the state variable $\mathbf{x}$ (e.g., the stationary distribution of the forecast model).

1. Forecast step: from the ensemble $\{\mathbf{x}_{1:n-1}^{a,(i)}\}$ at time $n-1$, generate a forecast ensemble $\{\mathbf{x}_n^{f,(i)}\}$ using the forecast model in (A1), i.e., $\mathbf{x}_n^{f,(i)} = F_n(\mathbf{x}_{n-l:n-1}^{a,(i)})$. Here, the superscript in $\mathbf{x}_n^f$ denotes the ensemble from the forecast model, and the superscript in $\mathbf{x}_n^a$ denotes the ensemble of the posterior distribution after assimilating data in the following analysis step. If the forecast model is stochastic, independent realizations should be used at different times.
2. Analysis step: Given new observation $\mathbf{z}_n$, update the forecast ensemble to get a posterior ensemble of $\mathbf{x}_n$,

$$\mathbf{x}_n^{a,(i)} = \mathbf{x}_n^{f,(i)} + K_n(\mathbf{z}_n^{(i)} - H\mathbf{x}_n^{f,(i)}), \tag{A2}$$

for $i = 1, \ldots, M$, where the Kalman gain matrix is

$$K_n = C_n^f H^T (HC_n^f H^T + R)^{-1}, \tag{A3}$$

where the matrix C_n^f is the sample covariance of the forecast ensemble,

$$C_n^f = \frac{1}{M-1} \sum_{i=1}^{M} \left(\mathbf{x}_n^{f,(i)} - \overline{\mathbf{x}}_n^f\right)\left(\mathbf{x}_n^{f,(i)} - \overline{\mathbf{x}}_n^f\right)^T,$$

where $\overline{\mathbf{x}}_n^f = \frac{1}{M}\sum_{i=1}^{M} \mathbf{x}_n^{f,(i)}$ and the $\mathbf{z}_n^{(i)}$ are obtained by adding random perturbations $\epsilon_n^{(i)} \sim N(0, R)$ to $\mathbf{z}_n$,

$$\mathbf{z}_n^{(i)} = \mathbf{z}_n + \epsilon_n^{(i)}.$$

APPENDIX B: PARAMETERS IN THE NAR MODELS

The parameters for the NAR models in the regimes of weak and strong stochastic forcing are shown in Tables V and VI (see Sec. III B 3 for the specifications).

REFERENCES

[1] M. Ghil, P. Yiou, S. Hallegatte, B. Malamud, P. Naveau, A. Soloviev, P. Friederichs, V. Keilis-Borok, D. Kondrashov, V. Kossobokov *et al.*, "Extreme events: Dynamics, statistics and prediction," Nonlinear Processes Geophys. **18**(3), 295–350 (2011).

[2] F. Kwasniok, "Data-based stochastic subgrid-scale parametrization: An approach using cluster-weighted modelling," Philos. Trans. R. Soc. A **370**(1962), 1061–1086 (2012).

[3] A. J. Majda, *Introduction to Turbulent Dynamical Systems in Complex Systems* (Springer, 2016).

[4] A. J. Majda and N. Chen, "Model error, information barriers, state estimation and prediction in complex multiscale systems," Entropy **20**(9), 644 (2018).

[5] A. Saha and U. Feudel, "Extreme events in FitzHugh-Nagumo oscillators coupled with two time delays," Phys. Rev. E **95**(6), 062219 (2017).

[6] S. A. Sheard and A. Mostashari, "Principles of complex systems for systems engineering," Syst. Eng. **12**(4), 295–311 (2009).

[7] W.-K. Tao, J.-D. Chern, R. Atlas, D. Randall, M. Khairoutdinov, J.-L. Li, D. E. Waliser, A. Hou, X. Lin, C. Peters-Lidard, W. Lau, J. Jiang, and J. Simpson, "A multiscale modeling system: Developments, applications, and critical issues," Bull. Am. Meteorol. Soc. **90**(4), 515–534 (2009).

[8] D. C. Wilcox, "Multiscale model for turbulent flows," AIAA J. **26**(11), 1311–1320 (1988).

[9] A. J. Chorin and F. Lu, "Discrete approach to stochastic parametrization and dimension reduction in nonlinear dynamics," Proc. Natl. Acad. Sci. U.S.A. **112**(32), 9804–9809 (2015).

[10] F. Lu, X. Tu, and A. J. Chorin, "Accounting for model error from unresolved scales in ensemble Kalman filters by stochastic parameterization," Mon. Weather Rev. **145**(9), 3709–3723 (2017).

[11] A. J. Majda and M. Branicki, "Lessons in uncertainty quantification for turbulent dynamical systems," Discrete Contin. Dyn. Syst. **32**(9), 3133 (2012).

[12] T. Palmer, P. Williams, and P. Williams, *Stochastic Physics and Climate Modelling* (Cambridge University Press Cambridge, 2010), Vol. 480.

[13] R. C. Smith, *Uncertainty Quantification: Theory, Implementation, and Applications* (Siam, 2013), Vol. 12.

[14] W. Cousins and T. P. Sapsis, "Quantification and prediction of extreme events in a one-dimensional nonlinear dispersive wave model," Physica D **280**, 48–58 (2014).

[15] C. Franzke, "Predictability of extreme events in a nonlinear stochastic-dynamical model," Phys. Rev. E **85**(3), 031134 (2012).

[16] A. J. Majda and D. Qi, "Strategies for reduced-order models for predicting the statistical responses and uncertainty quantification in complex turbulent dynamical systems," SIAM Rev. **60**(3), 491–549 (2018).

[17] D. Qi and A. J. Majda, "Predicting extreme events for passive scalar turbulence in two-layer baroclinic flows through reduced-order stochastic models," Commun. Math. Sci. **16**(1), 17–51 (2018).

[18] N. Chen and A. J. Majda, "Predicting observed and hidden extreme events in complex nonlinear dynamical systems with partial observations and short training time series," Chaos **30**(3), 033101 (2020).

[19] W. Cousins and T. P. Sapsis, "Reduced-order precursors of rare events in unidirectional nonlinear water waves," J. Fluid Mech. **790**, 368–388 (2016).

[20] M. Farazmand and T. P. Sapsis, "Extreme events: Mechanisms and prediction," Appl. Mech. Rev. **71**(5), 050801 (2019).

[21] D. Qi and A. J. Majda, "Using machine learning to predict extreme events in complex systems," Proc. Natl. Acad. Sci. U.S.A. **117**(1), 52–59 (2020).

[22] T. P. Sapsis, "Statistics of extreme events in fluid flows and waves," Annu. Rev. Fluid Mech. **53**, 85–111 (2021).

[23] Z. Y. Wan, P. Vlachas, P. Koumoutsakos, and T. Sapsis, "Data-assisted reduced-order modeling of extreme events in complex dynamical systems," PLoS ONE **13**(5), e0197704 (2018).

[24]J. D. Anderson, Jr., *Fundamentals of Aerodynamics* (Tata McGraw-Hill Education, 2010).
[25]R. Courant and K. O. Friedrichs, *Supersonic Flow and Shock Waves* (Springer Science & Business Media, 1999), Vol. 21.
[26]A. H. Shapiro, *The Dynamics and Thermodynamics of Compressible Fluid Flow* (Ronald Press, New York, 1953).
[27]W. E. K. Khanin, A. Mazel, and Y. G. Sinai, "Invariant measures for Burgers equation with stochastic forcing," Ann. Math. **151**(3), 877–960 (2000).
[28]W. E. K. Khanin, A. Mazel, and Y. Sinai, "Probability distribution functions for the random forced Burgers equation," Phys. Rev. Lett. **78**(10), 1904–1907 (1997).
[29]A. Dunlap and L. Ryzhik, "Viscous shock solutions to the stochastic Burgers equation," Arch. Ration. Mech. Anal. **242**(2), 1–35 (2021).
[30]G. Da Prato, A. Debussche, and R. Temam, "Stochastic Burgers' equation," Nonlinear Differential Equations Appl. **1**(4), 389–402 (1994).
[31]J. Bec and K. Khanin, "Burgers turbulence," Phys. Rep. **447**(1–2), 1–66 (2007).
[32]H. Cho, D. Venturi, and G. E. Karniadakis, "Statistical analysis and simulation of random shocks in stochastic Burgers equation," Proc. R. Soc. A. **470**(2171), 20140080 (2014).
[33]A. J. Chorin and O. H. Hald, "Viscosity-dependent inertial spectra of the Burgers and Korteweg-deVries-Burgers equations," Proc. Natl. Acad. Sci. U.S.A. **102**(11), 3921–3923 (2005).
[34]Y. G. Sinai, "Statistics of shocks in solutions of inviscid Burgers equation," Commun. Math. Phys. **148**(3), 601–621 (1992).
[35]M. Beck and C. E. Wayne, "Using global invariant manifolds to understand metastability in the Burgers equation with small viscosity," SIAM J. Appl. Dyn. Syst. **8**(3), 1043–1065 (2009).
[36]S. Dolaptchiev, U. Achatz, and I. Timofeyev, "Stochastic closure for local averages in the finite-difference discretization of the forced Burgers equation," Theor. Comput. Fluid Dyn. **27**(3–4), 297–317 (2013).
[37]C. Mou, B. Koc, O. San, and T. Iliescu, "Data-driven variational multiscale reduced order models," Comput. Methods Appl. Mech. Eng. **373**, 113470 (2020).
[38]Z. Wang, I. Akhtar, J. Borggaard, and T. Iliescu, "Two-level discretizations of nonlinear closure models for proper orthogonal decomposition," J. Comput. Phys. **230**(1), 126–146 (2011).
[39]K. K. Lin and F. Lu, "Data-driven model reduction, Wiener projections, and the Koopman-Mori-Zwanzig formalism," J. Comput. Phys. **424**, 109864 (2021).
[40]F. Lu, "Data-driven model reduction for stochastic Burgers equations," Entropy **22**(12), 1360 (2020).
[41]C. W. Gardiner, *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*, Springer Series in Synergetics Vol. 13 (Springer-Verlag, Berlin, 2004).
[42]B. Guo, *Spectral Methods and Their Applications* (World Scientific, 1998).
[43]S. M. Cox and P. C. Matthews, "Exponential time differencing for stiff systems," J. Comput. Phys. **176**(2), 430–455 (2002).
[44]A. K. Kassam and L. N. Trefethen, "Fourth-order time stepping for stiff PDEs," SIAM J. Sci. Comput. **26**(4), 1214–1233 (2005).
[45]D. Gottlieb and S. Orszag, *Numerical Analysis of Spectral Methods: Theory and Applications* (SIAM, Philadelphia, PA, 1977).
[46]M. D. Chekroun, H. Liu, and J. C. McWilliams, "Variational approach to closure of nonlinear dynamical systems: Autonomous case," J. Stat. Phys. **179**(5), 1073–1160 (2020).
[47]B. Gershgorin, J. Harlim, and A. J. Majda, "Test models for improving filtering with model errors through stochastic parameter estimation," J. Comput. Phys. **229**(1), 1–31 (2010).
[48]J. Harlim, "Model error in data assimilation," in *Nonlinear and Stochastic Climate Dynamics*, edited by C. Franzke and T. O'Kane (Cambridge University Press, Oxford, 2016).
[49]A. J. Chorin and O. H. Hald, *Stochastic Tools in Mathematics and Science*, 3rd ed. (Springer, New York, 2013).
[50]A. J. Chorin, O. H. Hald, and R. Kupferman, "Optimal prediction with memory," Physica D **166**(3), 239–257 (2002).
[51]A. J. Chorin, "Conditional expectations and renormalization," Multiscale Model. Simul. **1**(1), 105–118 (2003).
[52]M. Santos Gutiérrez, V. Lucarini, M. D. Chekroun, and M. Ghil, "Reduced-order models for coupled dynamical systems: Data-driven methods and the Koopman operator," Chaos **31**(5), 053116 (2021).
[53]F. Lu, K. K. Lin, and A. J. Chori, "Data-based stochastic model reduction for the Kuramoto–Sivashinsky equation," Physica D **340**, 46–57 (2017).
[54]A. Gouasmi, E. J. Parish, and K. Duraisamy, "*A priori* estimation of memory effects in reduced-order models of nonlinear systems using the Mori–Zwanzig formalism," Proc. R. Soc. A **473**(2205), 20170385 (2017).
[55]J. Harlim, S. W. Jiang, S. Liang, and H. Yang, "Machine learning for prediction with missing dynamics," J. Comput. Phys. **428**, 109922 (2020).
[56]C. Ma, J. Wang, and W. E, "Model reduction with memory and the machine learning of dynamical systems," Commun. Comput. Phys. **25**(4), 947–962 (2018).
[57]F. Lu, K. K. Lin, and A. J. Chorin, "Comparison of continuous and discrete-time data-based modeling for hypoelliptic systems," Commun. Appl. Math. Comput. Sci. **11**(2), 187–216 (2016).
[58]M. D. Chekroun, H. Liu, and J. C. McWilliams, "Stochastic rectification of fast oscillations on slow manifold closures," Proc. Natl. Acad. Sci. U.S.A. **118**(48), e2113650118 (2021).
[59]M. D. Chekroun, H. Liu, and S. Wang, *Approximation of Stochastic Invariant Manifolds: Stochastic Manifolds for Nonlinear SPDEs I*, Springer Briefs in Mathematics (Springer, 2015).
[60]D. Henry, *Geometric Theory of Semilinear Parabolic Equations*, Lecture Notes in Mathematics Vol. 840 (Springer-Verlag, Berlin, 1981).
[61]T. Ma and S. Wang, *Bifurcation Theory and Applications*, World Scientific Series on Nonlinear Science, Series A: Monographs and Treatises Vol. 53 (World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2005).
[62]G. Burgers, P. J. van Leeuwen, and G. Evensen, "Analysis scheme in the ensemble Kalman filter," Mon. Weather Rev. **126**, 1719–1724 (1998).
[63]G. Evensen, *Data Assimilation: The Ensemble Kalman Filter*, 2nd ed. (Springer, Berlin, 2007).
[64]P. Holmes, J. L. Lumley, and G. Berkooz, *Turbulence, Coherent Structures, Dynamical Systems and Symmetry* (Cambridge, 1996).
[65]K. Taira, S. L. Brunton, S. T. Dawson, C. W. Rowley, T. Colonius, B. J. McKeon, O. T. Schmidt, S. Gordeyev, V. Theofilis, and L. S. Ukeiley, "Modal analysis of fluid flows: An overview," AIAA J. **55**, 4013–4041 (2017).
[66]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).
[67]P. J. Schmid, "Dynamic mode decomposition of numerical and experimental data," J. Fluid Mech. **656**, 5–28 (2010).
[68]S. E. Ahmed, S. Pawar, O. San, A. Rasheed, T. Iliescu, and B. R. Noack, "On closures for reduced order models—A spectrum of first-principle to machine-learned avenues," Phys. Fluids **33**(9), 091301 (2021).
[69]A. J. Majda, I. Timofeyev, and E. Vanden Eijnden, "A mathematical framework for stochastic climate models," Commun. Pure Appl. Math. **54**(8), 891–974 (2001).
[70]J. Wouters and V. Lucarini, "Multi-level dynamical systems: Connecting the Ruelle response theory and the Mori-Zwanzig approach," J. Stat. Phys. **151**, 850–860 (2013).
[71]X. Xie, M. Mohebujjaman, L. G. Rebholz, and T. Iliescu, "Data-driven filtered reduced order modeling of fluid flows," SIAM J. Sci. Comput. **40**(3), B834–B857 (2018).
[72]N. Chen, Y. Li, and H. Liu, arXiv:2112.05226 (2021).
[73]N. Chen and A. J. Majda, "Conditional Gaussian systems for multiscale nonlinear stochastic systems: Prediction, state estimation and uncertainty quantification," Entropy **20**(7), 509 (2018).
[74]M. D. Chekroun, H. Liu, and S. Wang, *Stochastic Parameterizing Manifolds and Non-Markovian Reduced Equations: Stochastic Manifolds for Nonlinear SPDEs II*, Springer Briefs in Mathematics (Springer, 2015).
[75]G. Evensen, "Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics," J. Geophys. Res. Oceans **99**(C5), 10143–10162, https://doi.org/10.1029/94JC00572 (1994).
[76]G. Evensen and P. J. Van Leeuwen, "Assimilation of Geosat altimeter data for the Agulhas current using the ensemble Kalman filter with a quasigeostrophic model," Mon. Weather Rev. **124**(1), 85–96 (1996).
[77]P. L. Houtekamer and H. L. Mitchell, "Data assimilation using an ensemble Kalman filter technique," Mon. Weather Rev. **126**(3), 796–811 (1998).
[78]J. Lei, P. Bickel, and C. Snyder, "Comparison of ensemble Kalman filters under non-Gaussianity," Mon. Weather Rev. **138**(4), 1293–1306 (2010).