Physics of Fluids

# On closures for reduced order models—A spectrum of first-principle to machine-learned avenues

Cite as: Phys. Fluids **33**, 091301 (2021); https://doi.org/10.1063/5.0061577
Submitted: 28 June 2021 . Accepted: 01 September 2021 . Published Online: 22 September 2021

Shady E. Ahmed, Suraj Pawar, Omer San, Adil Rasheed, Traian Iliescu, Bernd R. Noack, et al.

Physics of Fluids REVIEW scitation.org/journal/phf

# On closures for reduced order models—A spectrum of first-principle to machine-learned avenues

Cite as: Phys. Fluids **33**, 091301 (2021); doi: 10.1063/5.0061577
Submitted: 28 June 2021 · Accepted: 1 September 2021 ·
Published Online: 22 September 2021

View Online Export Citation CrossMark

Shady E. Ahmed,[1] Suraj Pawar,[1] Omer San,[1,a)] Adil Rasheed,[2] Traian Iliescu,[3] and Bernd R. Noack[4,5]

AFFILIATIONS

[1]School of Mechanical and Aerospace Engineering, Oklahoma State University, Stillwater, Oklahoma 74078, USA
[2]Department of Engineering Cybernetics, Norwegian University of Science and Technology, N-7465 Trondheim, Norway
[3]Department of Mathematics, Virginia Tech, Blacksburg, Virginia 24061, USA
[4]School of Mechanical Engineering and Automation, Harbin Institute of Technology, Shenzhen 518058, China
[5]Hermann-Föttinger-Institut für Strömungsmechanik, Technische Universität Berlin, D-10623 Berlin, Germany

[a)]Author to whom correspondence should be addressed: osan@okstate.edu

## ABSTRACT

For over a century, reduced order models (ROMs) have been a fundamental discipline of theoretical fluid mechanics. Early examples include Galerkin models inspired by the Orr–Sommerfeld stability equation and numerous vortex models, of which the von Kármán vortex street is one of the most prominent. Subsequent ROMs typically relied on first principles, like mathematical Galerkin models, weakly nonlinear stability theory, and two- and three-dimensional vortex models. Aubry *et al.* [J. Fluid Mech. **192**, 115–173 (1988)] pioneered the data-driven proper orthogonal decomposition (POD) modeling. In early POD modeling, available data were used to build an optimal basis, which was then utilized in a classical Galerkin procedure to construct the ROM, but data have made a profound impact on ROMs beyond the Galerkin expansion. In this paper, we take a modest step and illustrate the impact of data-driven modeling on one significant ROM area. Specifically, we focus on ROM closures, which are correction terms that are added to the classical ROMs in order to model the effect of the discarded ROM modes in under-resolved simulations. Through simple examples, we illustrate the main modeling principles used to construct the classical ROMs, motivate and introduce modern ROM closures, and show how data-driven modeling, artificial intelligence, and machine learning have changed the standard ROM methodology over the last two decades. Finally, we outline our vision on how the state-of-the-art data-driven modeling can continue to reshape the field of reduced order modeling.



## I. INTRODUCTION

One of the very first exciting experiences that kids go through is playing with water; they might throw a stone in a lake, float a rubber duck in a bathtub, or even stir a straw while enjoying a tasty cup of juice! They like doing this over and over again because of the magnificent patterns that keep forming every time. These patterns or *coherent structures* are ubiquitous in the world, in general, and in fluid flows, in particular. They attracted Leonardo da Vinci more than five centuries ago, resulting in some of his outstanding artwork.[1] Fluid dynamicists are especially lucky to enjoy the beauty of these formations on a daily basis, but other than their esthetic value, these patterns come with a practical benefit. In particular, these coherent structures are the cornerstone in the development of *reduced order models (ROMs)* for fluid flows. ROMs are built by using available *data* to identify and rank these structures, choosing the most effective few of them, and tracking their dynamical behavior in order to approximate the evolution of the underlying flow. The computational cost of the relatively low-dimensional ROMs is dramatically lower than the computational cost of a direct numerical simulation, which aims at capturing all the flow scales. Since their introduction to the field of fluid dynamics more than fifty years ago,[2] ROMs have witnessed tremendous changes. Arguably, *data-driven modeling* has been the main driving force behind these changes. Over the last two decades, the state-of-the-art methods from *machine learning (ML)* have reshaped the field of reduced order modeling.

The main objective of this study is to provide an overview of data-driven reduced order modeling strategies relevant to the fluid dynamics applications. The topic spans a wide spectrum, and there are many review articles on the pertinent discussions, methodologies, and applications in fluids[3–19] as well as closely related fields.[20–43] Therefore, it is not our intention to include a detailed discussion, but rather to survey one important ROM research area, closure modeling, and provide our *subjective* perspectives on how data-driven modeling has made an impact in this area. In particular, given the recent interest in ML applications in fluid dynamics, our survey is intended to encourage the cross-disciplinary efforts between the practitioners, physicists, mathematicians, and data scientists. We hope that our paper will shed light on the new ideas of integrating both physics in ML models and ML-enabled capabilities in principled models, a rapidly emerging field that came to be known as physics-guided ML (PGML). These developments are born to conform with the scientific foundations that are moving rapidly to the industry to enable the next generation of digital twin technologies.[44]

This paper first aims at identifying the imminent practical and mathematical needs in designing closure approaches for ROMs of nonlinear parameterized fluid dynamics systems, i.e., complex natural or engineered systems comprising coupled partial differential equations (PDEs) with variable parameters, and initial and boundary conditions.[45] Such models usually serve as the inner-workhorse for outer-workflow loops, such as optimal design,[46] control,[47,48] estimation, and discovery.[28,49–52] In particular, there has recently been an increasing interest in ROMs from the fluid dynamics community, where the emerging data-driven methods prevail. This is primarily due to the fact that data and centralized powerful open-source machine learning and optimization libraries have become widespread, as indicated in Fig. 1. Although we mainly focus on the incompressible flows, we emphasize that there have been inspiring works done in the compressible case.[53–67]

To begin with, the reduced order modeling can be viewed as the art of converting existing prior information and collected data into a dramatically more efficient, yet relatively accurate, surrogate model to be used on demand. For example, a conceivable strategy of flow control is to put most of the demanding calculations offline and to keep only the low-rank updates for fluid flow evolutions online.[68,69] Emerging digital twin infrastructures are one of the main beneficiaries and driving forces behind the efficient surrogate model development efforts.[70,71] Although the ROM concept is not new in fluid dynamics, there are still many new fronts and opportunities, mainly due to the recent advances in ML algorithms and easy-to-use open-source packages that can be utilized in many control and optimization processes. We also note that, in many fluid dynamics applications, the typical data sparsity (due to the number of resolved degrees of freedom being orders of magnitude larger than the number of available sensors) and corruption (e.g., due to signal noise, interference, and sensor malfunctioning) motivate the physics-informed data-driven modeling.

Among fluid dynamicists, the projection-based linear methods have become popular. Both proper orthogonal decomposition (POD)[72] and dynamic mode decomposition (DMD)[73] enabled approaches have been exploited. In our work, we mostly focus on POD-relevant literature and refer the reader to Kutz *et al.*[73] for the DMD principles. The mathematical foundations behind the POD-based linear subspace approaches go back to the principal component analysis (PCA), pioneered by Pierson[74] in 1901 and later demonstrated graphically by Hotelling[75] in 1933. This powerful statistical approach (also known as Kosambi–Karhunen–Loève expansion[76,77] or empirical orthogonal functions[78,79]) was first introduced in the fluid dynamics community by Lumley[2,80,81] and came to be known as POD. In practice, the *method of snapshots*, established by Sirovich,[82] was a key enabler to efficiently determine the POD modes for large-scale problems, often encountered in fluid dynamics. Of particular interest when characterizing the dynamics of coherent structures in wall bounded flows, the beauty of the POD modeling approach was

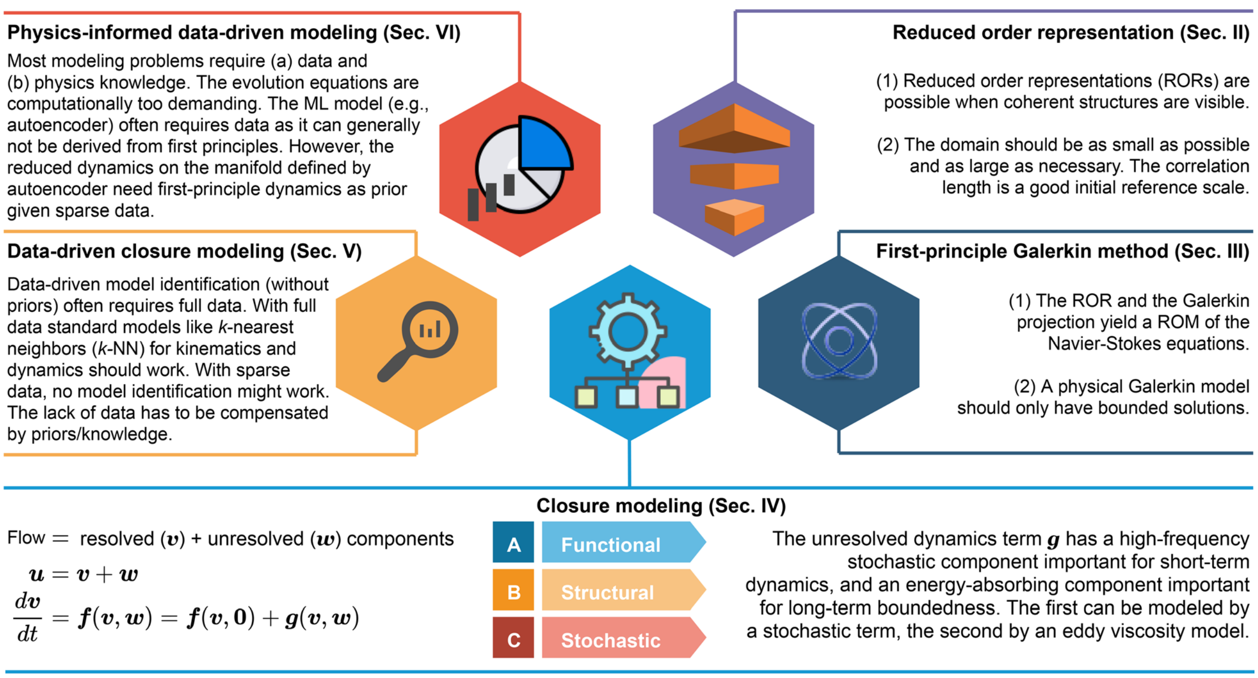


**FIG. 1.** An overview of data-driven reduced order modeling in fluid dynamics.

demonstrated in the seminal works by Aubry *et al.*[83,84] An admittedly incomplete chronological evolution of projection-based ROMs is given in Table I.

A key advantage of POD is the guaranteed minimal representation error for the employed snapshots with respect to all other Galerkin expansions with the same number of modes. Another advantage is the orthogonality of the constructed modes that naturally leverages the use of a Galerkin type projection onto the governing PDEs to obtain a system of ordinary differential equations, defining a dynamical system for the amplitudes of the POD modes. These models are often *intrusive* in the sense that both the governing equations and the prerecorded snapshot data are required to build a ROM. Postulating an alternative approach, DMD-based approaches and many other *nonintrusive* models bypass this equation dependency in order to construct the ROM solely based on the prerecorded snapshots.[13,14,28,40,85–94]

If we adopt the training and testing terminologies from ML, both POD- and DMD-based models require the *training* snapshot data to build the ROM (i.e., data-driven modeling). The fundamental question in practice is how well these ROMs will perform in testing conditions (i.e., the conditions that are not included in the training data). The trade-off performance between training and testing constitutes one of the crucial questions about the credibility of the proposed ROMs and motivates more efforts, ideas, and collaborations to push the frontiers of existing ROM frameworks. Physics-informed ML has also made an

**TABLE I.** An incomplete chronological list of key contributions to projection-based ROMs in fluid dynamics.

| Year | Study | Key contribution |
|---|---|---|
| 1915 | Galerkin[95] | Galerkin method for solving (initial) boundary value problems |
| 1962 | Saltzman[96] | Low-dimensional modeling (with 7 modes, see also Ref. 97 for a revisit) |
| 1963 | Lorenz[98] | Low-dimensional modeling (with 3 modes) |
| 1967 | Lumley[2] | Proper orthogonal decomposition (POD) |
| 1987 | Sirovich[82] | Method of snapshots |
| 1988 | Aubry *et al.*[83] | First POD model: dynamics of coherent structures and global eddy viscosity modeling |
| 1994 | Rempfer and Fasel[99] | Linear modal eddy viscosity closure |
| 1995 | Everson and Sirovich[100] | Gappy POD |
| 2000 | Ravindran[101] | Galerkin ROM for optimal flow control problems |
| 2001 | Kunisch and Volkwein[102] | First numerical analysis of Galerkin ROM for parabolic problems |
| 2002 | Willcox and Peraire[85] | Balanced truncation with POD |
| 2003 | Couplet *et al.*[103] | Guidelines for modeling unresolved modes in POD-Galerkin models |
| 2004 | Sirisup and Karniadakis[104] | Spectral viscosity closure for POD models |
| 2004 | Barrault *et al*[86] | Empirical interpolation method (EIM) |
| 2005 | Mezić[87] | Spectral decomposition of the Koopman operator |
| 2007 | Rozza *et al.*[105] | Reduced basis approximation |
| 2007 | Cao *et al.*[106] | Galerkin ROM for four-dimensional variational data assimilation |
| 2008 | Amsallem and Farhat[107] | Interpolation method based on the Grassmann manifold approach |
| 2008 | Astrid *et al.*[108] | Missing point estimation |
| 2009 | Rowley *et al.*[109] | Spectral analysis of nonlinear flows |
| 2009 | Sapsis and Lermusiaux[110] | Dynamically orthogonal field equations |
| 2010 | Schmid[91] | A purely nonintrusive perspective: dynamic mode decomposition (DMD) |
| 2010 | Chaturantabut and Sorensen[90] | Discrete empirical interpolation method (DEIM) |
| 2013 | Carlberg *et al.*[111] | The Gauss–Newton with approximated tensors (GNAT) method |
| 2013 | Cordier *et al.*[112] | Proof of global boundedness of nonlinear eddy viscosity closures |
| 2014 | Östh *et al.*[113] | $\sqrt{K}$-scaled eddy viscosity concept |
| 2015 | Ballarin *et al.*[114] | Stabilization of POD Galerkin approximations |
| 2015 | Schlegel and Noack[115] | On bounded solutions of Galerkin models |
| 2016 | Peherstorfer and Willcox[116] | Data-driven operator inference nonintrusive ROMs |
| 2016 | Brunton *et al.*[117] | Sparse identification of nonlinear dynamics (SINDy) |
| 2016 | Sieber *et al.*[118] | Spectral POD |
| 2018 | Towne *et al.*[119] | On the relationship between spectral POD, DMD, and resolvent analysis |
| 2018 | Reiss *et al.*[120] | Shifted/transported snapshot POD |
| 2018 | Loiseau *et al.*[121] | Feature-based manifold modeling |
| 2019 | Mendez *et al.*[122] | Multiscale proper orthogonal decomposition |
| 2021 | Li *et al.*[123] and Fernex *et al.*[124] | Cluster-based network models |

impact in reduced order modeling. In this hybrid approach, while the training data provide a set of global basis functions, the underlying governing equations (i.e., physics) constrain the evolution within the linear subspace defined by these POD basis functions. Deep discussions on hybrid approaches that combine deterministic and statistical modeling can be found elsewhere.[20,125–131]

Projection-based ROMs have been explored for decades, and these explorations have been paying off in many applications. They have great promise for flow control of industrial processes, for enlarging the ensemble size for flow problems with uncertain data, and even for providing accurate forecasts of fluid behavior. The apparent success of the low-rank ensemble nonlinear filtering methods utilized in weather forecasting centers also suggests that there is a prospect of using a system whose dimension is substantially lower than the dimension of the state space. Yet, the ROMs' potential has been only realized for a small collection of canonical flows. One of the main roadblocks for ROMs of realistic flows is that they are not accurate models for the dominant modes. In practice, a closure or correction term is generally added.[112,113,132–139] In many cases, there are complementary physical, statistical, and computational challenges that arise in the development of ROMs and ROM closures, topics that we will systematically survey in this work toward establishing foundations to close the gap between what ROMs can do and where they are needed.

## II. REDUCED ORDER REPRESENTATION

A *reduced order representation* (ROR) can be viewed as a generalization of the latent space or manifold, i.e., a simplifying kinematic approximation. For example, the POD procedure introduced in Sec. II B constitutes a best-fit linear manifold to establish a ROR. The ROR facilitates a data compression for an ensemble of snapshot data. Physically interpretable RORs are possible when dominant coherent structures are present. This is clearly ubiquitous in the fluid flows that we encounter in our daily life, as introduced in Sec. I, as well as in large-scale and industrial settings. For spatiotemporal dynamical systems, the rank-r ROR of the state $u(\boldsymbol{x}, t)$ can be simply written as

$$u(\boldsymbol{x}, t) = \sum_{i=1}^{r} a_i(t)\psi_i(\boldsymbol{x}), \tag{1}$$

where $\boldsymbol{x}$ refers to the spatial coordinates, $t$ is the time, $\psi_i$ denotes the $i$th mode in the ROR, and $a_i$ is the corresponding amplitude or coefficient. Although it is generally assumed that the basis functions $\psi_i$ are time-independent and the dynamical evolution is encapsulated in the coefficients $a_i$, there have been studies that admit time-evolving basis functions as well.[110,140–142]

### A. Eigenfunction expansion

Any set of $n$ linearly independent vectors can serve as a basis for an $n$-dimensional vector space. Any vector in this space can be expressed as a linear combination of these linearly independent basis vectors. In an infinite dimensional vector space of functions, there exists an infinite set of linearly independent basis functions $\{\psi_i(\boldsymbol{x})\}_{i=1,2,\ldots}$ such that a given function $u$ in this space can be written as a linear combination of these functions. It is straightforward to show that any periodic, piecewise continuous function can be written as an infinite sum of sines and cosines (e.g., Fourier series[143]). The eigenfunction expansion can be viewed as a generalization of the Fourier series expansion for arbitrary boundary conditions, where the Sturm–Liouville theory provides an infinite sequence of eigenvalue–eigenfunction pairs. ROR also seeks an expansion of an arbitrary function in terms of a given set of basis functions. However, in contrast to the methods mentioned in this section, ROR aims at finding a low-dimensional basis instead of an infinite dimensional one.

To illustrate these concepts, let us consider a linear advection problem in a spatial domain $[0, L]$,

$$\frac{\partial u}{\partial t} + c\frac{\partial u}{\partial x} = 0, \tag{2}$$

where $c$ is the wave speed. To select the appropriate basis functions, we consider the boundary conditions. For example, if we have homogeneous Dirichlet boundary conditions, i.e.,

$$u(x = 0, t) = 0, \quad u(x = L, t) = 0 \quad \text{for} \quad t \in [0, T], \tag{3}$$

we might choose a set of *orthonormal* basis functions $\psi_i(x)$ defined as

$$\psi_i(x) = \sqrt{\frac{2}{L}}\sin\left(\frac{i\pi}{L}x\right), \tag{4}$$

in order to approximate $u$ as follows:

$$u(x, t) = \sum_{i=1}^{r} a_i(t)\psi_i(x). \tag{5}$$

In a more sophisticated scenario with the homogeneous Neumann condition on the left boundary ($x = 0$) and the homogeneous Dirichlet condition on the right boundary ($x = L = 1$), i.e.,

$$\left.\frac{\partial u}{\partial x}\right|_{x=0} = 0, \quad u|_{x=1} = 0 \quad \text{for} \quad t \in [0, T], \tag{6}$$

we can define a set of *non-orthogonal* functions that satisfy Eq. (6) as follows:

$$\phi_i(x) = \cos(i\pi x) - (-1)^i, \tag{7}$$

and apply the Gram–Schmidt *orthonormalization* process to obtain the following set of basis functions:

$$\psi_i(x)=\sqrt{\frac{4i-2}{2i+1}}\left[\frac{(-1)^{i+1}}{2i-1}+\cos(i\pi x)+\frac{2}{2i-1}\sum_{j=1}^{i-1}(-1)^{i+j+1}\cos(j\pi x)\right]. \tag{8}$$

Here, we note that these basis functions are orthonormal, i.e.,

$$\int_0^1 \psi_i(x)\psi_k(x)dx = \delta_{ik}, \tag{9}$$

and they are derived from the Fourier harmonics that satisfy the boundary conditions.

Next, we focus on an illustrative example with periodic boundary conditions, with the domain length $L = 2\pi$, the maximum time $T = 2\pi$, and the wave speed $c = 1$. For a given initial condition $u(x, t = 0) = \cos(x)$, Eq. (2) admits an analytical solution in the form of a right traveling wave $u(x, t) = \cos(x - t)$. Let us approximate $u$ using a modal expansion with only two Fourier harmonics defined by

$$u(x,t) = a_1(t)\cos(x) + a_2(t)\sin(x). \tag{10}$$

Substituting Eq. (10) into Eq. (2), we get

$$\frac{\partial}{\partial t}(a_1(t)\cos(x) + a_2(t)\sin(x)) + \frac{\partial}{\partial x}(a_1(t)\cos(x) + a_2(t)\sin(x)) = 0. \tag{11}$$

Once we multiply Eq. (11) with $\cos(x)$ and integrate over the domain, we obtain an equation for $a_1(t)$,

$$\frac{da_1}{dt} = -a_2, \tag{12}$$

and similarly, multiplying Eq. (11) with $\sin(x)$, the evolution equation for $a_2$ becomes

$$\frac{da_2}{dt} = a_1. \tag{13}$$

Equations (12) and (13) constitute the well-known Galerkin system. Using the initial condition given at $t=0$,

$$a_1(0) = 1, \quad a_2(0) = 0, \tag{14}$$

we can obtain an analytical solution of the Galerkin system given by Eqs. (12) and (13) as

$$a_1(t) = \cos(t), \quad a_2(t) = \sin(t). \tag{15}$$

Therefore, the two-mode Galerkin model approximation given by Eq. (10) yields a solution

$$u(x,t) = \cos(t)\cos(x) + \sin(t)\sin(x), \tag{16}$$

which can be further written as

$$u(x,t) = \cos(x - t). \tag{17}$$

As we illustrated in this example, the two-mode approximation retrieves the exact solution. One of the key aspects in such a modal scheme is, therefore, related to the characteristics of the selected basis functions, which ultimately provided the best possible expansion in this example. A central question is how we would know *a priori* the appropriate $\cos(x)$ and $\sin(x)$ basis functions to approximate $u$.

We also note that the multimodal method utilizes similar arguments to represent the solution as a superposition of an infinite set of generalized Fourier basis functions and time-dependent coefficients. The multimodal method has been extensively exploited to study the sloshing problem,[144–146] where a set of natural harmonic functions are defined to satisfy the boundary conditions. This definition is challenging since each individual tank shape requires a dedicated applied mathematical and physical study. Moreover, Faltinsen and Timokha[147] reported that the simple truncation of the infinite sum to a finite sum $r$ can yield either inaccurate or expensive computations. Thus, the selection of the dominant modes is a non-trivial task. A truncation based on employing special asymptotic relationships, postulated following mathematical or physical arguments, has been shown to produce good results for the sloshing problem.[148–153] Nevertheless, these relations are valid under reasonable assumptions for specific tank geometries, which poses a fundamental challenge in the multimodal method's application in arbitrary settings.[147] It is, therefore, tempting to explore the emerging data-driven tools to mitigate such problems. For example, as discussed in Sec. II B, one could consider the POD procedure, which provides a systematic framework that yields a set of basis functions (accompanied by a sorting mechanism) from a set of snapshots.

### B. Proper orthogonal decomposition: Linear best-fit basis functions

In addition to the boundary conditions, we might have archival data (i.e., snapshot fields) to help us construct the basis functions. PCA[74] can be used to construct the basis functions that optimally represent the data. In 1933, a geometric representation of PCA has been proposed by Hotelling,[75] and this concept has later become popular as empirical orthogonal functions (EOF)[154] in environmental science, and POD[84] in the fluid dynamics community. The *method of snapshots*[82] has been instrumental in the development of POD-based approaches.[19] To compute the POD basis functions, let us assume that we have access to $m$ snapshots $\boldsymbol{u}(\boldsymbol{x}, t_i)$ for $i = 1, 2, \ldots, m$. A Reynolds decomposition-like expansion can be written as

$$\boldsymbol{u}(\boldsymbol{x}, t_i) = \bar{\boldsymbol{u}}(\boldsymbol{x}) + \boldsymbol{v}(\boldsymbol{x}, t_i), \tag{18}$$

where $\bar{\boldsymbol{u}}(\boldsymbol{x})$ is a reference (e.g., ensemble mean) field, which can be obtained as

$$\bar{\boldsymbol{u}}(\boldsymbol{x}) = \frac{1}{m}\sum_{i=1}^{m}\boldsymbol{u}(\boldsymbol{x}, t_i), \tag{19}$$

and the set of anomaly snapshots $\boldsymbol{v}(\boldsymbol{x}, t_i)$ for $i = 1, 2, \ldots, m$ can be defined as

$$\boldsymbol{v}(\boldsymbol{x}, t_i) = \boldsymbol{u}(\boldsymbol{x}, t_i) - \bar{\boldsymbol{u}}(\boldsymbol{x}). \tag{20}$$

For clarity, we introduce the POD procedure for a scalar field (POD is normally applied to the velocity vector field). Let us denote $\upsilon(\boldsymbol{x}, t_j)$ a component of the anomaly velocity vector field (e.g., the $x$-component). A temporal correlation matrix $\mathbf{A} = [\alpha_{ij}]$ can be constructed from these anomaly snapshots:

$$\alpha_{ij} = \int_{\Omega} \upsilon(\boldsymbol{x}, t_i)\upsilon(\boldsymbol{x}, t_j)d\boldsymbol{x}, \quad \boldsymbol{x} \in \Omega, \tag{21}$$

where $\Omega$ is the spatial domain, and $i$ and $j$ refer to the snapshot indices. We define the $L^2$ inner product of two functions $f$ and $g$ as

$$(f(\cdot), g(\cdot)) = \int_{\Omega} f(\boldsymbol{x})g(\boldsymbol{x})d\boldsymbol{x}, \tag{22}$$

which yields $\alpha_{ij} = (\upsilon(\boldsymbol{x}, t_i), \upsilon(\boldsymbol{x}, t_j))$ from Eq. (21). The data correlation matrix $\mathbf{A} = [\alpha_{ij}]$ is a non-negative, symmetric $m \times m$ matrix, also known as the Gramian matrix of $\upsilon(\boldsymbol{x}, t_1), \upsilon(\boldsymbol{x}, t_2), \ldots, \upsilon(\boldsymbol{x}, t_m)$. If we define the diagonal eigenvalue matrix $\Lambda = \text{diag}[\lambda_1, \ldots., \lambda_m]$ and a right eigenvector matrix $\Gamma = [\gamma_1, \ldots., \gamma_m]$ whose columns are the corresponding eigenvectors of $\mathbf{A}$, we can solve the following eigenvalue problem to obtain the optimal POD basis functions:[101]

$$\mathbf{A}\Gamma = \Gamma\Lambda. \tag{23}$$

In general, most of the subroutines for solving Eq. (23) give $\Gamma$ with all of the eigenvectors normalized to unity. The orthonormal POD basis functions for the anomaly field, $\upsilon$, can be thus calculated as follows:

$$\psi_i(\boldsymbol{x}) = \frac{1}{\sqrt{\lambda_i}} \sum_{k=1}^{m} \gamma_i^k \upsilon(\boldsymbol{x}, t_k), \tag{24}$$

where $\lambda_i$ is the $i$th eigenvalue, $\gamma_i^k$ is the $k$th component of the $i$th eigenvector, and $\psi_i(\boldsymbol{x})$ is the $i$th POD mode.

The eigenvalues are often stored in descending order for practical purposes, i.e., $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m \geq 0$, and the eigenvectors are normalized in such a way that the basis functions satisfy the following orthonormality condition:

$$(\psi_i, \psi_j) = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases} \tag{25}$$

Now, we can linearly represent the anomaly field variable $\upsilon(\boldsymbol{x}, t)$ using the POD modes as follows:

$$\upsilon(\boldsymbol{x}, t) = \sum_{i=1}^{r} a_i(t)\psi_i(\boldsymbol{x}), \tag{26}$$

where $a_i$ are the time-dependent (pseudo) modal coefficients, and $r$ is the total number of retained modes after the truncation, with $r \ll m$. These $r$ modes with the largest energy content correspond to the largest eigenvalues $(\lambda_1, \lambda_2, \ldots, \lambda_r)$. In general, adding more POD modes reduces the POD-ROM error. We note, however, that this is not always true. For example, adding POD modes that are polluted by numerical noise can actually decrease the POD-ROM error (see, e.g., the numerical investigation in Ref. 155). Often, the value of $r$ is determined by using the relative information content (RIC) index,[156] which is defined as

$$\text{RIC} = \frac{\sum_{i=1}^{r} \lambda_i}{\sum_{i=1}^{m} \lambda_i}, \tag{27}$$

where $\text{RIC} = 1$ refers to a complete representation of the data snapshots. For example, if one records a set of snapshots from the field given by Eq. (17) [i.e., the solution of the advection problem in Eq. (2)], let us say, $m = 100$ or more equally distributed snapshots between $t = 0$ and $t = T$, the POD analysis could offer a perfect representation with $\text{RIC} = 1$ using only two retained modes ($r = 2$) since the underlying dynamics can be constructed by a linear superposition of two harmonics. Of course, that is not always the case, and RIC becomes smaller than unity even if we retain a substantial number of modes, especially for turbulent flows.

This need for a large number of modes is one of the chief motivating factors for developing closure models to compensate the effects of the truncated modes in ROMs. However, it is believed that there is no separation of scales in turbulence, and therefore, most turbulent flow problems cannot be characterized by a high RIC index. Specifically, if there is no significant pattern in the evolution dynamics, there is a slow decay rate for the eigenvalues $\lambda_k$, and retaining only a few modes cannot capture the essential dynamics of turbulence. Thus, it is not surprising that many ROM practitioners have often demonstrated their proposed methodologies for problems that show somehow an underlying pattern (e.g., a shedding pattern in simulating the von Karman street).

This picture can be linked to the Kolmogorov barrier,[157] where the linear reducibility (i.e., representing the underlying fluctuation field as a linear superposition/span of a finite/limited number of basis functions) is hindered. *Modal expansions have an elliptic nature by construction, and using such tools for convection-dominated flows with higher degrees of hyperbolicity might often add another level of complexity when designing projection ROMs.* Then, a central question might arise about these data-driven procedure: *Do we really get any benefit using the POD basis functions generated from prerecorded snapshots?* Although there might be a trade-off between storage, accuracy, and efficiency, the answer probably depends on the problem at hand. It might be a big *yes* if there is an underlying pattern (e.g., limit cycles or quasi-periodic oscillations), and might be a *no* if the flow is highly turbulent in a statistically non-equilibrium and chaotic state. In the latter case, one might consider a standard local discretization (e.g., finite difference/element/volume) method or a pseudo-spectral method (supported by the harmonics that satisfy the boundary conditions) without attempting to perform the POD procedure to compute a set of data-driven global basis functions.

For instance, the fast Fourier transform (FFT) provides an extremely efficient computational framework for models with such global basis functions without requiring any additional storage for precomputed or measured snapshots to generate a set of data-driven basis functions. The trade-off between the accuracy and computational efficiency should always be considered carefully in generating data-driven models like Galerkin ROMs. The complexity of a typical right-hand side (RHS) computation of a pseudo-spectral solver becomes slightly bigger than $\mathcal{O}(n)$, where $n$ refers to the number of grid points. In contrast, the complexity of a typical Galerkin ROM is $\mathcal{O}(r^3)$ (there are also additional costs associated with, e.g., collecting and processing snapshots or solving an eigenvalue problem to generate a set of basis functions). Therefore, the Galerkin ROM becomes a computationally feasible approach *if and only if* a few number of retained modes are utilized. As a rule of thumb, $r$ should be significantly less than the number of grid points in each direction for a canonical 3D problem (e.g., $r \ll 256$ for a $256^3$ problem). Otherwise, it would be hard to justify that the model is indeed *reduced order*, since instead we could simply use the FFT algorithm to integrate the dynamical system equations in the harmonic space. Same arguments hold true for using a more flexible and convenient localized model, especially for problems with more complicated geometries (e.g., with the finite element, finite difference, or finite volume method, where the RHS can be obtained in $\mathcal{O}(n)$ computations).

## C. Leveraging Reduced Order Representation

One of the major reasons for the inaccuracy of current ROMs in the numerical simulation of complex flows is the quality of ROR, which is the ability or inability of the ROM framework to represent the underlying complex dynamics. Specifically, in order to determine whether there is a valid ROR of the given system, we need to answer the following questions: (i) Is the ROM basis able to accurately approximate the dynamics? (ii) Is the Galerkin projection yielding an accurate ROM?

Furthermore, the issue of the selection of a convenient domain often comes into play. If the domain of influence is too small, there might be no good dynamical prediction. On the other hand, when it is too large, there could be too many uncorrelated events that have to be lumped in global modes. These uncorrelated events might work against the modeling accuracy since they often increase the deformation of modal expansion or degradation of the model representation. In other words, the domain should be as small as possible and as large as necessary. The correlation length might be a good initial reference scale to define the domain of interest. Dynamic mode adaptation, parameter-space-time domain partitioning as well as smart clustering ideas have been explored, although we believe this topic is still in its infancy.

In practice, the construction of *a good* low-order space is a cornerstone in projection-based ROM. That said, the representability of POD basis functions becomes questionable for non-stationary, strongly evolving, and convection-dominated flows. Being a linear-based approach, POD might not be sufficient to describe the nonlinear processes. More importantly, using a Galerkin projection based on elliptical ansatz for a hyperbolic problem could generate numerical oscillations. Moreover, the POD is optimal *globally* in the sense that it minimizes the *averaged* $L^2$ error across all the snapshot data. This raises the issue of modal deformation by the rapidly varying flow field state in such a way that the resulting modes are not representative of any of the system's states. Furthermore, since the POD modes are ranked based on their energy content, excursions in state spaces that contain a small amount of energy can be overlooked by POD even if these excursions might have significant impact on the dynamical evolution (see Cazemier *et al.*,[158] for example). Similar scenarios arise for parameterized systems spanning a large parameter space when the system's behavior highly depends on the parameter value. Therefore, we devote the rest of this section to the efforts aimed at enhancing the basis representability by either improving the offline construction stages or efficiently updating the ROM during online deployment.

One of the simplest approaches to improve the quality of the POD basis functions is to enrich the snapshot data matrix with *extra* information. For example, in addition to the exact flow field data, the scaled difference between the consecutive snapshots (i.e., the difference quotients) can be utilized such that the time derivative information is better represented in the resulting modes.[102,159] Moreover, instead of collecting snapshot data at arbitrary time intervals and/or parameter values, more effective sampling techniques should be pursued. In Ref. 160, a ROM is integrated into a Markov chain Monte Carlo (MCMC) framework, where the posterior distribution estimated by the MCMC algorithm is utilized to adaptively select the parameter values at which snapshots are evaluated.

In an effort toward the accurate identification of coherent structures from experimental or numerical data, a spectral POD approach has been developed,[118] and its relationship to DMD and resolvent analysis has been established.[119] These studies use a technique that performs several PODs on individual frequencies obtained from Fourier-transformed windows of snapshot data. Thus, the modes we get for each frequency correspond to a coherent frequency domain structure. If an analysis of the eigenvalue spectrum for each frequency reveals coherent structures, it can indicate that there is a physical process which is occurring at that frequency. Hence, this becomes a useful data analysis tool on top of providing orthogonal modes for ROMs. In addition, transported snapshots POD approaches[120,161] have been introduced for convection-dominated transport systems. In particular, these studies use a shifting operator on the snapshots (requiring interpolation on unstructured grids and some knowledge of the transport speed) to allow POD or DMD to (more) efficiently approximate advective systems. In their recent works, Mendible *et al.*[162] employed an unsupervised traveling wave identification with shifting and truncation (UnTWIST) algorithm[161] to discover moving coordinate frames into which the data are shifted, thus overcoming the limitations imposed by the underlying translational invariance and allowing for the application of traditional dimensionality reduction techniques. Etter and Carlberg[163] proposed a novel online adaptive basis refinement mechanism for efficiently enriching the trial basis in a manner that ensures convergence of the ROM to the FOM.

Localization methods have been successfully pursued to mitigate the modal deformation of the POD basis by partitioning the state space,[164–167] time domain,[157,168–175] physical domain,[176–179] or parameter space[180–182] using multiple local, piecewise affine subspace approximations instead of a single global approximation. These partitioning or time-varying approaches work by parsing the available snapshot data into a few overlapping or non-overlapping groups (e.g., based on solution value, time, parameter, geometry, or component) and applying standard modal decomposition techniques (e.g., POD) for each region separately. This eventually yields a library of compact ROMs, each suitable for a specific region and/or dynamics, and interpolation methods can be utilized when the region of interest does not exist in the available library.

In this context, clustering techniques can be also utilized to effectively perform such partitioning. Indeed, cluster-based reduced order models (CROMs) have been proposed to tackle some of the potential pitfalls of classical GROMs (e.g., the mismatch between the modal expansion approach and the underlying dynamics, see Noack[183]); CROMs start by sorting the snapshot data into a small number of clusters (e.g., using k-means approach) with centroids being the representative states in each cluster. Conceptually, this is similar to coarse-graining the state-space (or generally the feature-space) into centroidal Voronoi tessellation (CVT) generators.[184] The transition dynamics between these centroids can be modeled as a probabilistic Markov model[185–187] or a deterministic-stochastic network model.[123,124,188]

As highlighted in Sec. II B, POD provides an efficient way to compress the data and explain the variance of the data better than any other linear combination.[189] Indeed, from a linear algebra perspective, it can often be formulated as a singular value decomposition, providing an optimal low-rank matrix approximation. This can leverage highly performant and scalable algorithms to handle extremely large datasets, benefiting from the rich legacy of linear algebra investigations. From a statistical point of view, this orthogonal projection provides *linearly uncorrelated* features. However, it cannot reveal nonlinear correlations in the data. In contrast, *manifold learning* (or representation learning) techniques aim at accounting for such nonlinear correlations to further reduce the dimensionality of the problem. The generalizations of PCA to nonlinear settings often define a curve in the latent space which minimizes the mean squared error of all variables. Yet, the smoothness of the curve can be varied by the method. For example, an autoassociative or autoencoding neural network model[190–192] and a kernel PCA[193] are two successful approaches of such a nonlinear PCA (NLPCA) framework. We refer the reader to recent works[194–196] for excellent discussions on autoencoder technology in fluid dynamics (see Fig. 2). We also note that other nonlinear dimensionality reduction techniques, such as principal curves,[197] locally linear embedding,[198] isomap,[199] and self-organized map[200] approaches, can also be regarded as a discrete version of NLPCA.

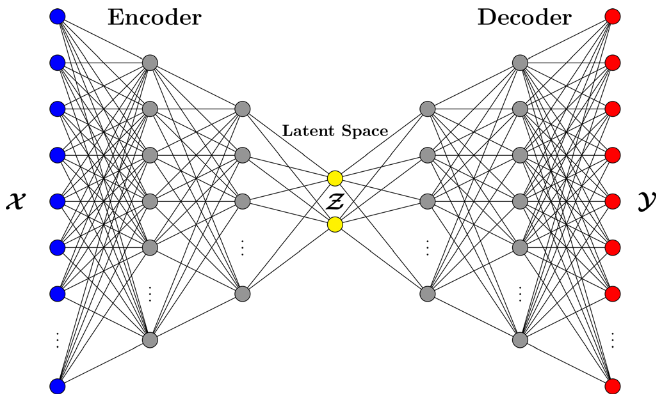


**FIG. 2.** A schematic diagram for an autoencoder (AE) for latent space construction, where the input $\mathcal{X}$ is the full field, the output $\mathcal{Y}$ designates its reconstruction, and $\mathcal{Z}$ represents the latent space (compressed) variables.

It is worth noting that during the modal truncation step (reducing the dimensionality of the system), dependencies among the retained and discarded modes generally yield inaccurate results if the closure problem is not addressed. In this regard, persistent homology (PH)[201–205] provides a delicate balance between data simplification and intrinsic structure extraction. PH is a tool in topological data analysis that aims at studying and extracting the features that persist across multiple scales by casting the multiscale organization into a mathematical formalism. In particular, PH measures the lifetime of intrinsic topological features using a filtration process to distinguish between the long-lived features and the short-lived ones (which are considered topological noise).[206]

However, the application of PH has been largely dedicated to qualitative data classification and analysis, and its utility for quantitative modeling and prediction (including ROM) is scarce.[207] PH has been utilized to characterize the time series from dynamical systems based on topological features that appear in the solution manifold or attractor.[208–210] Rieck and Leitte[211] used PH as an evaluation tool to compare the performance of different dimensionality reduction algorithms (e.g., PCA and isomap mentioned earlier). One of the major challenges of employing PH is its prohibitive computational cost (for the worst-case scenario). To increase the PH efficiency, Moitra *et al.*[212] utilized a clustering technique to represent similar groups of data points with their cluster centroid and applied PH onto these clusters.

## III. FIRST-PRINCIPLE GALERKIN METHOD

Finite dimensional low-order models routinely arise when we apply Galerkin type projection techniques to infinite dimensional PDE models.[213–215] We formalize the model reduction problem for fluid flow systems, considering a generic prognostic equation as follows:

$$\frac{\partial \mathbf{u}}{\partial t} = \mathcal{F}(\mathbf{u}; \mathbf{x}, t), \tag{28}$$

where $\mathbf{u}$ denotes the discrete approximation of a three-dimensional (3D) dependent variable (e.g., density, velocity, temperature, moisture); $\mathbf{x}$ denotes the independent spatial variables (e.g., latitude, longitude, and height); and $\mathcal{F}$ defines the model's dynamical core (e.g., semi-discretized PDEs representing mass, momentum, and energy conservation), all written in vector form. Specifically, we explore the autonomous dynamics case for a quantity of interest $\mathbf{u}$, with $\mathcal{F}$ being decomposed into linear $\mathcal{L}$ and nonlinear $\mathcal{N}$ operators as

$$\frac{\mathrm{d}\mathbf{u}}{\mathrm{d}t} = \mathcal{L}\mathbf{u} + \mathcal{N}(\mathbf{u}), \tag{29}$$

where $\mathbf{u} \in \mathbb{R}^n$ and $n$ is the number of degrees of freedom in the spatial discretization. Considering the Navier–Stokes equations as a typical mathematical framework for fluid flow modeling, we highlight that the linear and nonlinear operators often represent the diffusive and convective effects, respectively.

In order to build the projection-based ROM, the solution $\mathbf{u}$ is approximated in a low-dimensional affine subspace of dimension $r$ via the Galerkin ansatz as follows:

$$\mathbf{u}(t) \approx \bar{\mathbf{u}} + \boldsymbol{\Psi}\boldsymbol{a}(t), \tag{30}$$

where $\bar{\mathbf{u}} \in \mathbb{R}^n$ is a reference solution representing the affine offset, $\boldsymbol{\Psi} \in \mathbb{R}^{n\times r}$ denotes the trial basis, and $\boldsymbol{a}(t) \in \mathbb{R}^r$ is the vector of reduced (generalized) coordinates, also called modal weights or coefficients. The reference solution $\bar{\mathbf{u}}$ as well as the basis $\boldsymbol{\Psi}$ are constructed during an offline stage from a collection of FOM evaluations (called snapshots). Without loss of generality, we suppose that the time-averaged field defines the reference solution, $\bar{\mathbf{u}}$, and the basis $\boldsymbol{\Psi}$ is constructed using the POD technique. Then, the low-rank approximation given by Eq. (30) is substituted into Eq. (29) and an inner product with a test basis is performed to yield a system of ODEs for the unknown modal coefficients, $\boldsymbol{a}(t)$. In Galerkin projection-based ROM (GROM), the test basis is chosen to be the same as the trial basis. We make use of the orthonormality property (i.e., $\boldsymbol{\Psi}^\top\boldsymbol{\Psi} = \boldsymbol{I}_r$, where $\boldsymbol{I}_r$ is the $r \times r$ identity matrix and the superscript $\top$ denotes the matrix transpose, assuming a Euclidean state space) as follows:

$$\boldsymbol{\Psi}^\top \frac{\mathrm{d}}{\mathrm{d}t}(\bar{\mathbf{u}} + \boldsymbol{\Psi}\boldsymbol{a}) = \boldsymbol{\Psi}^\top\mathcal{L}(\bar{\mathbf{u}} + \boldsymbol{\Psi}\boldsymbol{a}) + \boldsymbol{\Psi}^\top\mathcal{N}(\bar{\mathbf{u}} + \boldsymbol{\Psi}\boldsymbol{a}). \tag{31}$$

Since both $\bar{\mathbf{u}}$ and $\boldsymbol{\Psi}$ are considered time-independent, Eq. (31) reduces to

$$\boldsymbol{\Psi}^\top\boldsymbol{\Psi}\frac{\mathrm{d}\boldsymbol{a}}{\mathrm{d}t} = \boldsymbol{\Psi}^\top\mathcal{L}\bar{\mathbf{u}} + \boldsymbol{\Psi}^\top\mathcal{L}\boldsymbol{\Psi}\boldsymbol{a} + \boldsymbol{\Psi}^\top\mathcal{N}(\bar{\mathbf{u}} + \boldsymbol{\Psi}\boldsymbol{a}). \tag{32}$$

Note that $\boldsymbol{\Psi}^\top\mathcal{L}\bar{\mathbf{u}}$ and $\boldsymbol{\Psi}^\top\mathcal{L}\boldsymbol{\Psi}$ can be precomputed during the offline construction stage, reducing the online computational cost of evaluating the first two terms on the right-hand side to $O(r)$, which is independent of the FOM dimension, $n$. However, generally speaking, computing the third term representing the system's nonlinearity depends on $n$, limiting the computational benefit of ROM. In order to mitigate this limitation, hyperreduction approaches have been developed to alleviate this dependency on $n$, by *approximating*, rather than *evaluating*, the nonlinear term in a reduced order subspace.[63,216,217] Examples of hyperreduction include the empirical interpolation method (EIM),[86] its discrete version (DEIM),[90,218] the gappy POD,[100,111,219] and the missing point estimation (MPE),[108,220] where the approximation is performed using sampling techniques. On the other hand, tensorial ROM can benefit from the quadratic (or generally polynomial) nonlinearity, which is ubiquitous in fluid flow systems, to rewrite Eq. (32) as follows:

$$\frac{d\boldsymbol{a}}{dt} = \mathfrak{B} + \mathfrak{L}\boldsymbol{a} + \boldsymbol{a}^\top \mathfrak{N}\boldsymbol{a}, \tag{33}$$

where the vector $\mathfrak{B}$, the matrix $\mathfrak{L}$, and the tensor $\mathfrak{N}$ are precomputed during the offline stage, reducing the computational cost of solving the GROM defined in Eq. (33) to $O(r^3)$ in the case of quadratic nonlinearity (which is the case for the Navier–Stokes equations). When the true underlying dynamics of the system are non-polynomial, *lifting* transformations can be exploited to yield a finite-dimensional coordinate representation in which the system dynamics have quadratic structure.[221–224] Although such transformation is not universally guaranteed, a large class of smooth nonlinear systems that appear in engineering applications (e.g., elementary functions like exponential and trigonometric functions or polynomials) can be equivalently lifted to quadratic form.

## A. POD Galerkin projection: Burgers equation

To illustrate the POD Galerkin approach for flow systems with quadratic nonlinearity, let us consider the Burgers equation,

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = \nu \frac{\partial^2 u}{\partial x^2}, \tag{34}$$

which is often used as a simplified prototype by fluid dynamicists. Using the POD procedure outlined in Sec. II B, we can define the $u$ field as a linear superposition of the mean field and the POD basis functions,

$$u(x,t) = \bar{u}(x) + \sum_{i=1}^{r} a_i(t)\psi_i(x), \tag{35}$$

and substitute this approximation of our field variable into Eq. (34). Once we perform an orthonormal Galerkin projection, the resulting dynamical system for $a_k(t)$ can be written as

$$\frac{da_k}{dt} = \mathfrak{B}_k + \sum_{i=1}^{r} \mathfrak{L}_k^i a_i + \sum_{i=1}^{r}\sum_{j=1}^{r} \mathfrak{N}_k^{ij} a_i a_j, \tag{36}$$

where

$$\begin{aligned} \mathfrak{B}_k &= \left(\nu \frac{\partial^2 \bar{u}}{\partial x^2} - \bar{u}\frac{\partial \bar{u}}{\partial x}, \psi_k\right), \\ \mathfrak{L}_k^i &= \left(\nu \frac{\partial^2 \psi_i}{\partial x^2} - \bar{u}\frac{\partial \psi_i}{\partial x} - \psi_i \frac{\partial \bar{u}}{\partial x}, \psi_k\right), \\ \mathfrak{N}_k^{ij} &= \left(-\psi_i \frac{\partial \psi_j}{\partial x}, \psi_k\right). \end{aligned} \tag{37}$$

This tensorial system consists of $r$ coupled ODEs and it is often written as Eq. (33), where $\boldsymbol{a}$ is the vector of unknown coefficients $a_k(t),\ k = 1, 2, \ldots, r$, $\mathfrak{B}$ is a scaling vector coming from the reference mean field with entries $\mathfrak{B}_k$, $\mathfrak{L}$ is an $r \times r$ matrix with entries $\mathfrak{L}_k^i$ for the contribution stemming from the linear viscous term, and $\mathfrak{N}$ is an $r \times r \times r$ tensor with entries $\mathfrak{N}_k^{ij}$ arising from the nonlinear advection term, $1 \le i, j, k \le r$. In this tensorial form, the corresponding model coefficients $\mathfrak{B}_k$, $\mathfrak{L}$, and $\mathfrak{N}$ are *precomputed* from the available snapshots. Alternatively, there are a number of *online* approaches where we can compute the nonlinear part using hyperreduction or principled sampling strategies to approximate the full nonlinear state from a small number of measurement or collocation points.[86,90,100,225–227] Ştefănescu *et al.*[216] performed a comparative study between the direct (online) and tensorial (precomputed) methods. Moreover, Karasözen *et al.*[228] recently discussed the structure preserving ROMs and compared the direct and tensorial POD approaches.

## B. Projection-based ROMs

Computational models for the Navier–Stokes equations could make a tremendous impact in critical applications, such as the biomedical and engineering applications that we describe next. Despite their enormous potential, FOMs have not fully transitioned to engineering practice. The main roadblock is the extraordinary computational cost incurred by computational models in many applications. For example, although preliminary studies of aortic dissections showed that uncertainties in the geometry and inflow conditions have a fundamental role, performing an *uncertainty quantification* study requires a huge number of computational model runs. Similarly, performing a *shape optimization* study to determine the optimal vascular configuration for the total cavopulmonary connection surgery requires again many computational model runs. Also, in renewable energy applications, performing *data assimilation* to incorporate the available observations in the control of wind-power production requires numerous model runs.

Since running current computational models hundreds and thousands of times can take days and weeks on high performance computing (HPC) platforms, a brute-force computational approach for these biomedical and engineering applications is simply not possible. Therefore, what is needed is a modeling strategy that allows model runs that take minutes to hours on a laptop.

For structure-dominated systems, ROMs can decrease the FOM computational cost by *orders of magnitude*. ROMs are (extremely) low-dimensional models that are trained (constructed) from available data. As explained in Sec. III, in an offline phase, the FOM is run for a few parameters values to construct a low-dimensional (e.g., 10-dimensional) ROM basis $\{\psi_1, \ldots, \psi_{10}\}$, which is used to build the ROM:

$$\frac{d\boldsymbol{a}}{dt} = \boldsymbol{f}(\boldsymbol{a}), \tag{38}$$

where $\boldsymbol{a}$ is the vector of coefficients in the ROM approximation $\sum_{i=1}^{10} a_i(t)\psi_i(\boldsymbol{x})$ of the variable of interest and $\boldsymbol{f}$ comprises the ROM operators (e.g., vectors, matrices, and tensors) that can be preassembled from the ROM basis in the offline phase. In the online phase, the low-dimensional ROM given by Eq. (38) is then used for parameters values that are *different* from those used in the training stage. Since ROM is low-dimensional (10-dimensional), its computational cost is orders of magnitude lower than the FOM cost. Thus, for the biomedical and engineering applications described above, ROMs appear as a natural alternative to the prohibitively expensive FOMs.

Unfortunately, current ROMs cannot be used in clinical and engineering practice, since they would require too many modes (degrees of freedom). For example, to capture all the geometric scales in aortic dissection, one might need hundreds or even thousands of ROM modes (e.g., see Table II). Similarly, to cope with the high Reynolds number in the wind farm optimization, a large number of ROM modes are necessary. Thus, although ROMs decrease the FOM computational cost by orders of magnitude, they are still too expensive: current ROMs cannot be run in minutes or hours on a laptop and thus cannot be used easily in clinical and engineering practice.

TABLE II. A non-exhaustive list illustrating the energy characteristics and range of the number of retained modes $r$ for $m$ given snapshots.

| Study | Problem | $r$ | $m$ | Comment |
|---|---|---|---|---|
| Östh *et al.*[113] | Ahmed body | 10–100 | 2000 | The first 500 modes resolve 60% of the kinetic energy. |
| San and Borggaard[171] | Marsigli flows | 6–30 | 400 | The first 30 modes resolve 90% of the kinetic energy. |
| Rahman *et al.*[229] | Quasigeostrophic flows | 10–80 | 400 | The first 50 modes resolve 80% of the kinetic energy. |
| Ballarin *et al.*[230] | Hemodynamics | 50 | 400 | O(10–100) modes are required to obtain a reliable approximation. |
| VerHulst and Meneveau[231] | Wind farm | N/A | 7200 | 430 POD modes are required to capture 80% of the total energy. |
| Shah and Bou-Zeid[232] | Atmospheric boundary layer | N/A | 2500 | 500 POD modes are required to capture 80% of the total energy. |
| Zhang and Stevens[233] | Atmospheric boundary layer | N/A | 5000 | 2000 POD modes are required to capture 80% of the total energy. |

TABLE III. A chronological list of key contributions to ROM closure modeling.

| Year | Study | Key contribution |
|---|---|---|
| 1988 | Aubry *et al.*[83] | First closure model: global eddy viscosity modeling |
| 1994 | Rempfer and Fasel[99] | Linear modal eddy viscosity closure |
| 1995 | Selten[252] | Time-averaging closure modeling |
| 1997 | Selten[253] | A statistical closure of a barotropic model |
| 1998 | Cazemier *et al.*[158] | Penalty term closure model based on the energy conservation principles |
| 2003 | Couplet *et al.*[103] | Guidelines for modeling unresolved modes in POD-Galerkin models |
| 2004 | Sirisup and Karniadakis[104] | Spectral viscosity closure for POD models |
| 2008 | Noack *et al.*[254] | Finite time thermodynamics and ensemble averaging closure models |
| 2009 | Bergmann *et al.*[255,256] | Residual-based variational multiscale POD |
| 2011 | Borggaard *et al.*[257] | First numerical analysis of closure models: artificial viscosity model |
| 2011 | Akhtar *et al.*[239] | Nonlinear eddy viscosity model based on the Frobenius norm of the Jacobian |
| 2011 | Wang *et al.*[258] | Two-level discretization model |
| 2012 | Wang *et al.*[135] | Eddy viscosity variational multiscale and dynamic Smagorinsky closures |
| 2013 | Balajewicz *et al.*[259] | Subspace calibration using the Navier–Stokes equations |
| 2013 | Cordier *et al.*[112] | Proof of global boundedness of nonlinear eddy viscosity closures |
| 2014 | Östh *et al.*[113] | $\sqrt{K}$-scaled eddy viscosity concept |
| 2014 | Iliescu and Wang[260] | Projection-based eddy viscosity variational multiscale POD |
| 2014 | San and Iliescu[261] | Smagorinsky and Chollet-Lesieur spectral vanishing eddy viscosity models |
| 2015 | Stinis,[262] Chorin and Lu,[263] and Li *et al.*[264] | Mori–Zwanzig (MZ) formalism |
| 2017 | Gouasmi *et al.*[265] | MZ ROM closures |
| 2017 | Rebollo *et al.*[266] | Reduced basis methods for the Smagorinsky closure model |
| 2017 | Xie *et al.*[267] | Approximate deconvolution reduced order modeling |
| 2017 | Benosman *et al.*[137] | Lyapunov control theory to design learning-based closure models |
| 2018 | San and Maulik[268] | Extreme learning machine closure model |
| 2018 | San and Maulik[269] | Neural network closures for ROM |
| 2018 | Pan and Duraisamy[270] | Sparse polynomial regression and neural network for closure model |
| 2019 | Rahman *et al.*[229] | Dynamic closure model based on a test (secondary) truncation approach |
| 2019 | Stabile *et al.*[138] | Reduced order variational multiscale approach for turbulent flows |
| 2020 | Imtiaz and Akhtar[271] | Nonlinear closure model based on the Jacobian of the Galerkin model |
| 2020 | Reyes and Codina[139] | Variational multiscale ROMs |
| 2020 | Xie *et al.*[272] | Residual neural network closures |
| 2020 | Wang *et al.*[273] | Recurrent neural network closures |
| 2021 | Mou *et al.*[274] | Data-driven variational multiscale ROMs |
| 2021 | Gupta and Lermusiaux[275] | Neural closure models |

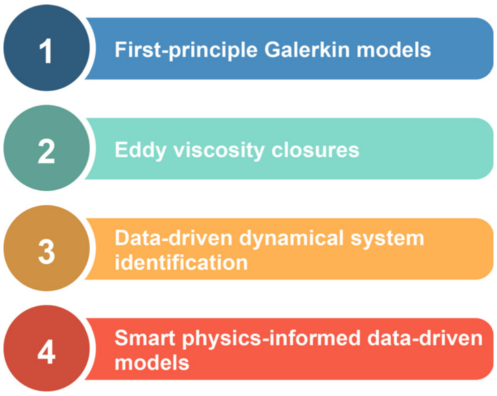


**FIG. 3.** Evolution of the Galerkin ROM approaches.

With the evolution of ROM approaches outlined in Fig. 3 in mind, the ROM community is at a *crossroads*. On the one hand, current ROMs can be used for academic test problems for which a handful of ROM modes can model simple dynamics with substantial success. On the other hand, realistic, complex flows require high-dimensional ROMs that cannot be used in clinical and engineering practice. What is needed is low-dimensional, *efficient* ROMs that are *accurate* so that they can be utilized in such vital applications.

One of the main reasons for the notorious inaccuracy of current ROMs in complex clinical and engineering settings is the *drastic ROM truncation*: instead of using many (e.g., 100) ROM modes $\{\psi_1, \ldots, \psi_{100}\}$, current ROMs use only a handful of ROM modes $\{\psi_1, \ldots, \psi_{10}\}$ to ensure a low computational cost. This drastic truncation yields acceptable results in simple, academic test problems, but produces inaccurate results in practical clinical and engineering settings.[113] Thus, for accurate results, the *ROM closure problem* needs to be solved: one needs to model the effect of the discarded ROM modes $\{\psi_{11}, \ldots, \psi_{100}\}$ on the ROM dynamics, i.e., on the time evolution of resolved ROM modes $\{\psi_1, \ldots, \psi_{10}\}$,

$$\frac{d\boldsymbol{a}}{dt} = \boldsymbol{f}(\boldsymbol{a}) + \mathfrak{C}, \tag{39}$$

where $\mathfrak{C}$ is a low-dimensional term that models the effect of the discarded ROM modes $\{\psi_{11}, \ldots, \psi_{100}\}$ on $\{\psi_1, \ldots, \psi_{10}\}$. The closure term is also known as *unresolved tendency* or *model error* in different disciplines.

The closure problem is prevalent in numerical simulation of complex systems. For example, the classical numerical discretization of turbulent flows (e.g., finite element or finite volume methods) inevitably takes place in the *under-resolved regime* (e.g., on coarse meshes) and requires closure modeling (i.e., modeling the subgrid scale effects). In computational fluid dynamics (CFD), e.g., large eddy simulation (LES), there are hundreds (if not thousands) of closure models.[234] This is in stark contrast to reduced order modeling, where only relatively few ROM closure models have been investigated. The reason for the discrepancy between ROM closure and LES closure is that the latter has been mostly built around physical insight stemming from Kolmogorov's statistical theory of turbulence (e.g., the concept of eddy viscosity), which is generally posed in the Fourier setting.[234,235] Much of this physical insight is not generally available in a ROM setting. Thus, current ROM closure models have been deprived of this powerful methodology that represents the core of most LES closure models. To construct low-dimensional and efficient ROMs that are *accurate*, a set of principled, mathematical, and/or data-driven ROM closure modeling strategies need to be utilized (Table III). In Sec. IV, we survey the main types of closure models developed in the reduced order modeling community.

## IV. CLOSURE MODELING

Although the solution of Eq. (33) becomes independent of the FOM dimension $n$, the cubic scaling with respect to $r$ hurts the turnaround of such ROMs. This is especially true for fluid flows of practical interest (e.g., turbulent and convection-dominated flows), where the FOM solution manifold is characterized by a large and slowly decaying Kolmogorov $n$-width.[236,237] Thus, a large number of modes are required to maintain the solution accuracy, resulting in excessive computational overhead, which may even exceed the FOM computational cost. Therefore, in these complex settings, ROM will always incur a degree of under-resolution by sacrificing some degree of accuracy for the sake of computational efficiency. This under-resolution has direct and indirect consequences. The direct outcome is the projection error affecting the Galerkin ansatz [Eq. (30)], where some of the underlying flow features are lost. The indirect ramifications are related to the nonlinearity of the system, implying that the discarded modes indeed interact with the retained ones. By performing severe modal truncation (remember, computational efficiency is a priority!), we suppress these interactions and Eq. (33) no longer captures the projected trajectory, decreasing the solution accuracy.[238]

To illustrate the above discussion, consider a state variable $\mathbf{u}(t) \in \mathbb{R}^n$, which can be *exactly* written as a superposition of $n$ basis functions as $\mathbf{u}(t) = \boldsymbol{\Psi}\boldsymbol{a}(t) + \Phi\boldsymbol{b}(t)$, where $\boldsymbol{\Psi} \in \mathbb{R}^{n\times r}$ and $\Phi \in \mathbb{R}^{n\times(n-r)}$ represent the modes to be retained and truncated, respectively, and $\boldsymbol{a}(t) \in \mathbb{R}^r$ and $\boldsymbol{b}(t) \in \mathbb{R}^{(n-r)}$ are the corresponding time-dependent coefficients. A Galerkin projection of the governing equations onto $\boldsymbol{\Psi}$ and $\Phi$ yields the following:

$$\frac{d}{dt}\begin{bmatrix}\boldsymbol{a}\\ \boldsymbol{b}\end{bmatrix} = \begin{bmatrix}\boldsymbol{f}_a(\boldsymbol{a},\boldsymbol{b})\\ \boldsymbol{f}_b(\boldsymbol{a},\boldsymbol{b})\end{bmatrix}. \tag{40}$$

We note that Eq. (40) is an exact representation of the system's dynamics. In reduced order modeling, we are only interested in the resolved part of the dynamics, which can be written as

$$\frac{d\boldsymbol{a}}{dt} = \boldsymbol{f}_a(\boldsymbol{a},\boldsymbol{b}). \tag{41}$$

Nevertheless, Eq. (41) is not practical because its solution requires the knowledge of the unresolved variable, $\boldsymbol{b}$. In a classic truncated ROM, it is often assumed that $\boldsymbol{f}_a(\boldsymbol{a},\boldsymbol{b}) = \boldsymbol{f}_a(\boldsymbol{a},\boldsymbol{0}) = \boldsymbol{f}(\boldsymbol{a})$. However, for nonlinear cases, this relation does not hold [i.e., $\boldsymbol{f}_a(\boldsymbol{a},\boldsymbol{b}) \neq \boldsymbol{f}_a(\boldsymbol{a},\boldsymbol{0})$].

Following the Kolmogorov hypotheses[242,243] from turbulence modeling and assuming an analogy between POD and Fourier modes (see Fig. 4), it is commonly agreed in the ROM community that the first POD modes resolve the large energy-containing flow scales, while the last modes correspond to the low-energy dissipative scales. Indeed, this analogy has been demonstrated theoretically and numerically for different flow scenarios (e.g., flow over a cylinder[244] and a turbulent flow past a backward-facing step[103]) Thus, truncating the low-energy

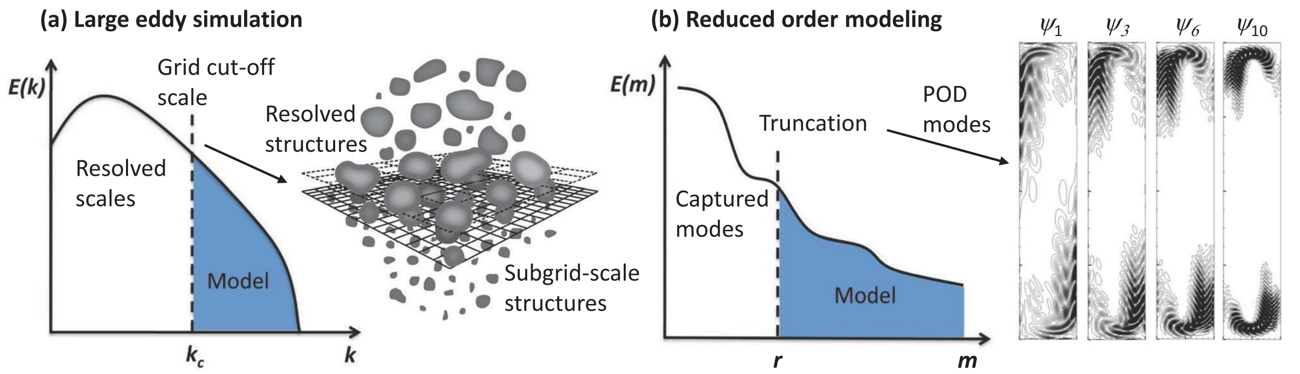


FIG. 4. Closure modeling analogy between LES and ROM, where the higher values of $k$ and $m$ refer to the smaller scales (adapted from Refs. 239–241). Shaded areas in energy spectra represent the discarded scales that must be modeled.

scales is believed to result in a pileup of energy levels, leading to the solution instability. We also highlight that this argument has been recently the focus of scientific revisits. For instance, Grimberg *et al.*[245] state, using mathematical arguments and analogies from finite element analysis, that the solution instabilitiy observed in most studies dealing with GROM is a by-product of the Galerkin projection step. Moreover, they show that a ROM based on Petrov–Galerkin projection, where the test basis differs from the trial basis, yields a more accurate and stable solution than standard Galerkin projection. However, the test basis needs to be updated at each iteration and time step, increasing the computational complexity of the resulting ROM. As a highly promising approach, an adjoint Petrov–Galerkin method for nonlinear model reduction has been recently put forth by Parish *et al.*[246] Rather than constructing a low-dimensional subspace for the entire state space in a monolithic fashion, Hoang *et al.*[247] recently proposed a dynamic methodology to construct separate subspaces for the different subdomains. Although the Petrov–Galerkin projection[248] could mitigate some of the challenges the Galerkin ROMs have to face in the under-resolved simulation of turbulent flows, we limit ourselves to Galerkin projection-based ROMs in the current review, where closure models have been mainly developed to improve the solution accuracy and stability properties. The major aim of closure models is to make up for the effects of discarded modes onto the dynamics of resolved modes. Specifically, the objective is to modify Eq. (33) to correctly resolve the time dynamics of $\boldsymbol{\Psi}$,

$$\frac{d\boldsymbol{a}}{dt} = \mathfrak{B} + \mathfrak{L}\boldsymbol{a} + \boldsymbol{a}^{\top}\mathfrak{N}\boldsymbol{a} + \mathfrak{C}, \tag{42}$$

where $\mathfrak{C}$ represents the closure model that needs to be determined.

The closure problem has historical roots in CFD, in particular in turbulence modeling, including the Reynolds averaged Navier–Stokes (RANS) and LES. There are relatively few closure models that have been investigated in a ROM context. In general, the ROM closure modeling approaches can be classified into (1) functional (phenomenological), which use physical insights to postulate a model form for the closure term (e.g., a dissipative term) and (2) structural (mathematical), which often rely on filtering techniques to reveal the closure term without using any physical assumptions or additional phenomenological arguments. We emphasize that placing a given ROM closure model in one of these categories is not always straightforward, as these categories sometimes overlap. We also refer the readers to the unified exposition of several mean field modeling ideas[249] as well as other closure techniques for probability density function (PDF)[250] or moment closures for kinetic theories.[251]

## A. Functional closure models

The functional, or phenomenological, closure modeling investigations have been largely focused on the concept of eddy viscosity, which is added to the physical viscosity of the system to drain the excessive energy. This modeling concept is inspired by Kolmogorov's ideas[242,243] about the energy spectrum and energy cascade. In this section, we outline several functional (phenomenological) ROM closure modeling strategies centered around the concept of eddy viscosity.

### *1. Mixing-length ROM closure*

The first mixing-length ROM closure model was proposed by Aubry *et al.*,[83] who studied the wall region of the turbulent boundary layer and used a simple generalization of the Heisenberg spectral model in homogeneous turbulence to provide the eddy viscosity closure term. Specifically, the authors assumed that the deviatoric component of the Reynolds stress $\hat{\boldsymbol{\tau}}$ of the unresolved field (represented by truncated modes), acting on the resolved field (i.e., retained modes), is proportional to the strain rate $\mathbf{S}$ of the resolved field,

$$\hat{\boldsymbol{\tau}} = -2\alpha\nu_e\mathbf{S}, \tag{43}$$

where $\nu_e$ is the eddy viscosity parameter and $\alpha$ is a dimensionless adjustable parameter. Moreover, they expressed the eddy viscosity term as a function of the eigenvalues and eigenfunctions of the first neglected modes based on the assumption that the energy decreases rapidly with increasing mode index. The adequacy of this model was quantitatively validated using numerical simulations in Ref. 276, and further investigations have been performed by Podvin and Lumley (e.g., for minimal flow unit[277] and channel flow[278]) and also in Refs. 135 and 279. The two main drawbacks of the global eddy viscosity modeling approach are as follows:

(i) This formulation is equivalent to using Navier–Stokes equations at a lower Reynolds number.
(ii) In this formulation, a linear closure term models the nonlinear turbulence dynamics.

#### 2. Smagorinsky ROM closure

As an improvement of the mixing-length model in Ref. 83, the celebrated Smagorinsky model[280] developed for LES has been utilized for the ROM closure problem. The eddy viscosity coefficient adapts in time in the Smagorinsky ROM closure model, but not in the mixing-length ROM closure model. Thus, the former is expected to be more accurate than the latter. To our knowledge, the first Smagorinsky ROM closure model was proposed in Noack *et al.*[281] (see also Ref. 282). Ullmann and Lang[283] used the same Smagorinsky closure model of the original FOM simulations for ROMs based on LES snapshots of the turbulent flow around a circular cylinder. However, the eddy viscosity term does not appear explicitly in their ROM equation. Instead, the reconstructed velocity field is utilized to update the (spatially varying) Smagorinsky eddy viscosity term in the FOM space, which in turn updates the corresponding model coefficients at each time step of the time integration of the ROM. Borggaard *et al.*[257] proposed the inclusion of an artificial viscosity term in the ROM equation that resembles the one used in the Smagorinsky model even if the original FOM (for data generation) does not involve such a term. Rebollo *et al.*[266] investigated the Smagorisnky ROM closure model in a reduced basis method (RBM) setting. A rigorous numerical analysis of the Smagorinsky ROM closure model was performed in Ref. 257, where error estimates for the time discretization were proven. To our knowledge, this represents the first numerical analysis of ROM closures. Error estimates for the time and space discretizations of the Smagorinsky ROM closure model were later proven in Ref. 266 in RBM context.

#### 3. Dynamic SGS ROM closure

The dynamic SGS model[284] is the state-of-the-art closure model in LES. The main improvement in the dynamic SGS model over the standard Smagorinsky model is that it uses an eddy viscosity coefficient that is updated in time by using a secondary filtering operation. The dynamic SGS closure model was extended for the first time to a ROM setting by Borggaard *et al.*[282] and was later investigated by Wang *et al.*[135] in the numerical simulation of a 3D flow past a cylinder, where it yielded significantly more accurate results than both the mixing-length and the Smagorinsky ROM closure models. A more efficient numerical discretization of the dynamic SGS ROM closure model was proposed by Rahman *et al.*[229]

#### 4. Mode-dependent eddy viscosity closure

Rather than adopting a single global eddy viscosity value $\nu_e$ for all the modes (as in the mixing-length ROM closure model), a mode-dependent eddy viscosity was proposed by Rempfer[285] and Rempfer and Fasel[286] to use a different amount of dissipation for each scale. In Refs. 285 and 286, the effective viscosity is calculated by requiring the energy variation of different modes in ROM to match the energy variation of the coherent structures in FOM. A modification to the mixing-length model can be incorporated by introducing a mode-dependent kernel. The importance of such a mode dependent kernel was first stressed by Rempfer.[99,285,286] Sirisup and Karniadakis[104] applied a vanishing viscosity kernel, which adds a small amount of mode-dependent dissipation that satisfies the entropy condition, yet retains spectral accuracy. The intrinsic stabilization scheme proposed in Ref. 132 utilizes information from available snapshots and POD modes to define a mode-dependent stabilization. In Ref. 261, linear, quadratic, and squareroot kernels were investigated for the 1D Burgers problem.

#### 5. Variational multiscale eddy viscosity ROM closure

The variational multiscale (VMS) methods developed by Hughes and his collaborators[287–289] have made a significant impact in classical CFD. The VMS methods center around the principle of locality of energy transfer, which states that energy is transferred mainly between neighboring scales or modes. Since ROMs use hierarchical bases in which the large and small structures are clearly displayed, the VMS framework was naturally extended to the ROM setting. Next, we present some of the eddy viscosity ROM closure models developed in a VMS framework. Borggaard *et al.*[282] proposed the first VMS ROM closure model, which was later investigated in Wang *et al.*[135] in the numerical simulation of a 3D flow past a circular cylinder. The VMS ROM in Refs. 135 and 282 used a *three-scale* decomposition of the flow field into resolved large, resolved small, and unresolved scales, and employed the Smagorinsky model to dissipate energy only from the resolved small scales. A *two-scale* decomposition of the flow field into resolved and unresolved scales was used by Bergmann *et al.*[256] to develop a VMS ROM closure model with a residual based eddy viscosity component. Iliescu and Wang[260,290] put forth a three-scale VMS ROM closure model, in which the ROM projection was used to construct an eddy viscosity term that acts only on the small resolved scales.

A similar three-scale VMS ROM was proposed by Roop[291] for a generalized Oseen problem. Eroglu *et al.*[292] developed a different three-scale VMS ROM that uses the ROM projection to add an eddy viscosity term acting only on the small resolved scales in a modular fashion. This VMS ROM was successfully tested in the numerical simulation of a turbulent channel flow at $Re_\tau = 395$[292] and was extended by Eroglu *et al.*[293] to the Darcy–Brinkman equations with the double diffusive convection. Stabile *et al.*[138] proposed a two-scale residual-based VMS ROM closure model in which the VMS strategy is used at both the FOM and the ROM levels to ensure model consistency. Reyes and Codina[139] (see also Ref. 294) developed a two-scale VMS ROM which is equipped with time-dependent orthogonal subgrid scales that leverage the orthonormal nature of the POD basis. Two-scale VMS-ROMs based on the orthogonal subgrid scales were used by Reyes *et al.*[295] for thermally coupled low Mach flows and by Tello *et al.*[296] for a fluid structure interaction problem. The first numerical analysis of VMS ROMs was performed in Refs. 260 and 290, where the stability and convergence were rigorously proven. A numerical analysis of VMS ROMs was also performed by Roop[291] and Eroglu *et al.*[292] We also refer to the recent studies by Rubino and his co-workers[297,298] for multistage ROM stabilization approaches in advection-dominated problems.

### 6. Finite-time thermodynamics ROM closure

In the majority of the aforementioned studies, the closure term eventually appears as a linear term in the GROM (i.e., $\mathfrak{C} = \tilde{\mathfrak{B}} + \tilde{\mathfrak{L}}\boldsymbol{a}$). (The Smagorinsky and dynamic SGS ROM closure models are notable exceptions.) Noack *et al.*[299] highlighted that the energy transfer is actually caused by nonlinear mechanisms. Thus, they introduced a nonlinear eddy viscosity term $\nu_e(\boldsymbol{a})$ that is state dependent. A finite-time thermodynamics (FTT)[254] approach was utilized to quantify the nonlinear eddy viscosity by matching the modal energy transfer effect as follows:

$$\nu_e(\boldsymbol{a}) = \nu_0\sqrt{\frac{K(t)}{\bar{K}}}, \tag{44}$$

where $K(t) = \sum_{i=1}^{r}\frac{1}{2}a_i(t)^2$ represents the total turbulence kinetic energy resolved by the Galerkin expansion and $\bar{K}$ denotes its time-averaged value. This led to damping levels more consistent with energy fluctuations than those defined by a linear eddy viscosity model. The FTT-based nonlinear eddy viscosity with an energy-based scaling model was successfully applied to a 3D turbulent jet[300] and a 2D mixing layer.[112] It was further extended to a mode-dependent nonlinear eddy viscosity for a high Reynolds number flow over a square-back Ahmed body.[113]

### 7. Efficient numerical discretization of ROM closures

Although the eddy viscosity closure models discussed in this section can significantly improve the ROM accuracy, their brute-force numerical discretization can be extremely inefficient. For example, the Smagorinsky ROM closure model depends on the Frobenius norm of the deformation tensor, which is a non-polynomial nonlinearity that cannot be preassembled in the offline stage. Thus, alternative, efficient numerical discretizations have been proposed over the last decade, which we outline next. Wang *et al.*[258] proposed a two-level method to avoid the brute-force discretization of the closure term onto the FOM fine mesh. Specifically, the POD bases constructed from the original fine grid resolution snapshot data were interpolated onto a coarse grid, and then they were used to efficiently compute the ROM closure term. To avoid the assembly of the FOM strain rate tensor at each time step, Akhtar *et al.*[239] used the Jacobian of the GROM right-hand side as an eddy viscosity coefficient. A precomputed eddy viscosity approach was adopted by San and Iliescu[301] by simplifying the nonlinear interaction in the original Smagorinsky model. San and Iliescu[261] also explored various closure approaches including constant, polynomial, and spectral vanishing viscosity models. Rebollo *et al.*[266] were the first to use an efficient hyper-reduction method[63] (i.e., EIM[86]) to discretize the Smagorinsky ROM closure in an RBM setting.

## B. Structural closure models

Structural closure models are generally derived through mathematical rather than phenomenological arguments. This often includes a filtering procedure, where the filtered field is assumed to have larger spatial structures than those in the original one. Therefore, the filtered flow variables require fewer modes in the ROM approximation. In other words, for the same number of modes, ROM is capable of approximating the filtered flow field more accurately than the unfiltered field. This approach is similar to LES, where the filtered flow variables can be approximated on the given coarse mesh more accurately than the original unfiltered flow variables. In this section, we survey ROM closure models developed by using different types of ROM filtering.

### 1. Spatial filtering: Projection

Given the hierarchical nature of the ROM basis, not surprisingly, the most popular type of ROM filtering has been the ROM projection, i.e., the projection of various (nonlinear) terms living in the $r$-dimensional ROM space spanned by the first $r$ ROM basis functions onto a smaller, $s$-dimensional ROM space spanned by the first $s$ ROM basis functions, where $s < r$. A classical example of ROM closure models constructed by using the ROM projection is the VMS-ROMs,[138,139,294,302] which are discussed in Sec. IV A. The ROM projection, however, has been used to develop other types of ROM closures. For example, the ROM projection has been utilized to construct parametrized manifolds ROM closures,[303,304] which are based on dynamical systems approaches, e.g., approximate inertial manifolds. The ROM projection has also been used to build ROM closures based on stochastic dynamical systems ideas.[305–307]

### 2. Spatial filtering: Differential filter and approximate deconvolution

Using the analogy between LES and ROM, we mention that a lot of ideas and techniques in image and signal processing are also applicable in ROM, and vice versa! In LES, the approximate deconvolution (AD) represents one of the most popular techniques in this class. It is based on the deconvolution approaches developed in the image processing and inverse problems communities to recover the original signal from a blurred filtered signal.

In stark contrast to the abundance of functional closure studies (beginning in the 1980s), there are only a few structural closure models in the ROM literature. The AD-ROM was proposed by Xie *et al.*[267] for the three-dimensional flow past a circular cylinder. To construct the AD-ROM, a ROM differential filter is applied to the Navier–Stokes equations, followed by a Galerkin projection of the *filtered* equations. It is usually assumed that the filtering and differentiation operators commute, and the repercussions of this assumption are investigated in Ref. 308. It is well known that, in general, nonlinearity and ROM spatial filtering do not commute. Therefore, the resulting equations include a *filtered* nonlinear term of the *unfiltered* variables [i.e., $\widehat{\mathcal{N}(\mathbf{u})}$, where the hat operator denotes the filtering process], rather than a nonlinear operator of the *filtered* variables [i.e., $\mathcal{N}(\widehat{\mathbf{u}})$]. A regularized deconvolution is adopted to provide the ROM approximation of the unfiltered flow variables in order to compute the nonlinear term. Thus, the filtering process increases the accuracy of the ROM in the sense that the filtered field contains larger spatial structures, and thus can be accurately captured by the ROM approximation. In addition, the AD technique solves the ROM closure problem by providing an estimate of the unfiltered flow variables.

Remark. We note that ROM spatial filtering has also been used to develop *regularized ROMs (Reg-ROMs)*, i.e., ROMs in which spatial filtering is used to smoothen (regularize) various terms in the Navier–Stokes equations and increase the numerical stability of the ROM. We emphasize that, while related, regularization and closure are different: the latter adds a closure term, whereas the former usually does not. ROM spatial filtering has been used to develop various types of Reg-ROMs: Wells *et al.*[309] proposed, for the first time, an evolve-then-filter approach in which the GROM [Eq. (33)] is integrated (evolved) for one time step, after which a ROM spatial filter is applied to filter the intermediate solution obtained in the evolve step. This filtering reduces the numerical oscillations of the flow variables (i.e., adds numerical stabilization to the ROM). Gunzburger *et al.*[310] proposed an evolve-filter-relax approach that considers the additional step of relaxation, which averages the unfiltered and filtered flow variables to control the amount of numerical dissipation introduced by the filter. Recently, Girfoglio *et al.*[311,312] have investigated the evolve-filter approach in a finite volume setting. The ROM differential filter has also been used to develop the Leray Reg-ROM in Refs. 313–315.

#### 3. PDF filtering: Mori–Zwanzig formalism and memory effects

A different type of filtering, based on filtering with respect to the PDF of the initial conditions, has been instrumental in adding memory effects to ROM closures, with a rationale based on the Mori–Zwanzig (MZ) formalism.[316–318] Recently, the MZ formalism has been intensely used to define closures for both LES and ROM settings. Next, we outline some of these developmemnts. Stinis[262] introduced a generalized MZ framework for the construction of ROMs for systems without scale separation. For example, assume that Eq. (40) defines the following linear system:[265,270,273,319]

$$\frac{d}{dt}\begin{bmatrix}\boldsymbol{a}\\ \boldsymbol{b}\end{bmatrix} = \begin{bmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{bmatrix}\begin{bmatrix}\boldsymbol{a}\\ \boldsymbol{b}\end{bmatrix}. \tag{45}$$

The evolution of the unresolved state $\boldsymbol{b}$ can be evaluated as follows (assuming that $\boldsymbol{a}$ is known):[265,273,320]

$$\boldsymbol{b}(t) = \int_0^t e^{A_{22}(t-s)} A_{21}\boldsymbol{a}(s)ds + e^{A_{22}t}\boldsymbol{b}(0). \tag{46}$$

Therefore, Eq. (41) for the dynamics of $\boldsymbol{a}$ can be written as

$$\begin{aligned}\frac{d\boldsymbol{a}}{dt} &= A_{11}\boldsymbol{a} + A_{12}\boldsymbol{b}\\ &= A_{11}\boldsymbol{a} + A_{12}\int_0^t e^{A_{22}(t-s)}A_{21}\boldsymbol{a}(s)ds + A_{12}e^{A_{22}t}\boldsymbol{b}(0).\end{aligned} \tag{47}$$

Equation (47) expresses the dynamics of the resolved scales using a Markovian term (i.e., $A_{11}\boldsymbol{a}$, which depends only on the current value of $\boldsymbol{a}$), a memory integral term depending on the history of the resolved scales $\boldsymbol{a}$, and a term describing the contribution of the initial conditions. This derivation can be extended to nonlinear settings where, for instance, the nonlinear ordinary differential equation can be written as a linear partial differential equation using the Liouville operator. The exact evolution equations for the reduced state can be written as

$$\frac{d\boldsymbol{a}}{dt} = \boldsymbol{f}_{\boldsymbol{a}}(\boldsymbol{a},0) + \int_0^t \mathcal{K}(\boldsymbol{a}(s), t-s)ds + \mathcal{O}(\boldsymbol{a}(0), \boldsymbol{b}(0)), \tag{48}$$

$$\approx \boldsymbol{f}(\boldsymbol{a}) + \mathfrak{C}, \tag{49}$$

which is the closure term in Eq. (42). In Eq. (48), $\mathcal{K}$ is called the memory kernel and $\mathcal{O}$ designates the contribution from the initial conditions. The memory integral term implies that the accurate resolution of $\boldsymbol{a}$ comprises a non-Markovian contribution. However, the direct computation of Eq. (48) is generally prohibitive, and the estimation of the memory effect is often sought. Li *et al.*[264] included a great discussion on the incorporation of memory effects in coarse-grained modeling via the MZ formalism. A discrete approach to stochastic parametrization, dimension reduction, and their connections to the MZ formalism of statistical physics has been proposed by Chorin and Lu.[263] In an LES setting, Parish and Duraisamy[320] framed the MZ closure modeling approach by exploiting similarities between two levels of coarse-graining via the Germano identity of fluid mechanics and by assuming that memory effects have a finite temporal support. The concept has been also generalized to provide a mathematically consistent framework for the construction of ROMs of dynamical systems.[265] Moreover, Parish and Duraisamy[321] established an analogy between MZ and VMS approaches.

#### 4. Ensemble averaging

Noack *et al.*[299] were the first to use the ensemble averaging to construct a finite-time thermodynamics (FTT)[254] framework. Gunzburger *et al.*[322] built ensemble-based POD ROMs, where the nonlinear advection term in the Navier–Stokes equations is replaced by a linear term in the equations for the resolved scales. This *linearization* is performed by using an ensemble of solution trajectories by propagating an ensemble of ROMs with varying parameters and/or initial conditions and updating the ensemble average at each time step. Later on, this ensemble-based approach was equipped with Leray regularization to develop regularized ROMs for convection-dominated flows.[315]

#### 5. Time averaging

Selten[252,253] used time averaging to develop ROM closures. In particular, by estimating the rate at which the ROM trajectory drifts away from the projection of the FOM solution on the ROM subspace, Selten[252] added a linear damping to expand the doubling-time of the error resulting from the modal truncation. Berselli *et al.*[323] developed mathematical support for eddy viscosity modeling of time-averaged ROM closures. While being interested in a statistical equilibrium problem exploring possible forward and backward average transfer of energy among ROM basis functions, they proved that the time-averaged energy exchange from low index POD modes to high index POD modes is positive for long enough time intervals. This study provides, for the first time, mathematical support for the ROM eddy viscosity methodology, where the energy transfer to the truncated modes is modeled by employing extra viscous dissipation.

#### 6. Calibrating the POD space with a Navier–Stokes based side constraint

The last modeling approach that we discuss in this section is that proposed by Balajewicz *et al.*[259] Although this approach does not add a ROM closure model, it does leverage mathematical arguments to model the effect of the truncated modes. In this approach, the POD subspace is subjected to a Navier–Stokes based side constraint. Specifically, the power balance for the fluctuation energy is required to be satisfied by the attractor data after the Galerkin projection on the adjusted POD space. This procedure can be conceptualized as rotating the POD subspace into a more dissipative regime, in which the extra dissipation is now performed by more dissipative POD modes.

### C. Stochastic closure models

Although we are mainly focusing on deterministic closure modeling in this review, we emphasize that the need for stochastic modeling was already formulated in Aubry *et al.*[83] to avoid statistically nonstationary behavior for some homoclinic orbits. The dynamics of the unresolved scales and hence their interactions with the resolved scales are unknown. Thus, we can only form an approximate idea of how the truncated modes behave and affect the ROM solution. Even with the best closure model, we can never be certain about its accuracy in practical settings. Thus, it is natural to model the dynamics of the unresolved modes using a random or *stochastic* process, from which we can infer the unresolved modes' contribution to the evolution of large scales in a statistical sense. We refer to Leith,[324] Chorin and Hald,[325] Majda *et al.*,[305] Majda and Harlim,[306] Harlim *et al.*,[307] Majda,[326] Resseguier *et al.*,[327] Chorin and Lu,[263] Lu *et al.*,[328] Lu,[329] Sieber *et al.*,[330] and Chekroun *et al.*,[303] for detailed discussions on the probabilistic modeling of such random/chaotic systems as well as the development of statistically accurate ROMs and stochastic closure models.[110,331–338] We also note that nonparametric stochastic modeling approaches have been proposed for representative stochastic Itô drift diffusion forecast models.[339] Next, we briefly outline a few of these strategies.

Stochastic closure approaches seek to account for the effects of the unresolved scales on the long-term statistics of the resolved scales. In particular, the closure term is modeled by a stochastic process, usually represented by Markovian and/or non-Markovian dynamics with a random forcing (e.g., random noise). For a truncated ROM of the Kuramoto–Sivashinsky system, Lu *et al.*[328] defined a discrete-time closure term $z^n$ at time $t_n$ as follows:

$$z^n = \Phi^n + \xi^n, \tag{50}$$

where $\xi$ is a sequence of independent identically distributed random variables, which are sampled from Gaussian distributions and characterize the stochastic component of the closure, while $\Phi$ is a function of current and past values of the resolved scales $\boldsymbol{a}$ and the forcing $\xi$. The authors used the nonlinear autoregression moving average with exogenous input (NARMAX) approach to parameterize $\Phi$. A similar approach was adopted in Ref. 263 for the Lorenz 96 model and Ref. 329 for the stochastic Burgers equation. The multiscale Lorenz 96 model,[340] which has been considered as a nontrivial test problem for stochastic parametrization in geophysical fluid dynamics studies, can be written as

$$\frac{dX_i}{dt} = -X_{i-1}(X_{i-2} - X_{i+1}) - X_i - \frac{hc}{b}\sum_{j=1}^{J} Y_{j,i} + F, \tag{51}$$

$$\frac{dY_{j,i}}{dt} = -cbY_{j+1,i}(Y_{j+2,i} - Y_{j-1,i}) - cY_{j,i} + \frac{hc}{b}X_i, \tag{52}$$

where Eq. (51) represents the evolution of slow, high-amplitude variables $X_i$ $(i = 1, \ldots, I)$, and Eq. (52) describes the evolution of coupled fast, low-amplitude variables $Y_{j,i}$ $(j = 1, \ldots, J)$. In order to investigate different closure approaches, $X$ can be considered as the resolved scales, while $Y$ can be considered as the unresolved ones. Therefore, Eq. (52) is assumed to be unknown and is only used for generating *true* data. Furthermore, the form of the term $-\frac{hc}{b}\sum_{j=1}^{J} Y_{j,i}$, representing the contribution of $Y$ to the dynamics of $X$, is also assumed to be unknown. A closure term is parameterized as a function of the resolved scales. Although the Lorenz 96 equations are deterministic, Wilks[331] showed the existence of multiple closure values that are consistent with any given large-scale variable, i.e., that different values of the closure term yield *statistically* similar results. Therefore, Wilks[331] defined the closure term using both deterministic and stochastic components. Specifically, the author used a fourth-order polynomial fitting for the deterministic part that represents the average trend, and a first-order auto-regression model for the stochastic part that defines the deviation of different realizations from the fitted curve. Arnold *et al.*[335] explored several parametrization schemes for the stochastic component, including the additive and multiplicative noise.

Although the distinction between the resolved and unresolved variables in the multiscale Lorenz 96 system is not driven by a Galerkin truncation as is the case for most projection-based ROMs (which is the focus of the current review), the same arguments apply in both scenarios. For example, Mémin[341] assumed that the flow field is decomposed into a deterministic resolved component and a generalized random field that models the unresolved flow component and all the uncertainties in the flow. In other words, the projection of the flow field onto the truncated space is treated as a *realization* or sample of the stochastic component of the flow and is modeled using Brownian motion. Resseguier *et al.*[327] provided numerical investigations of this methodology using POD-Galerkin projection for flow past a cylinder. Nonetheless, we believe that this is an open research area that needs fresh ideas to translate statistical closure strategies[342] from turbulence modeling to the ROM arena.

## V. DATA-DRIVEN CLOSURE MODELING

With the abundant supply of big data, open-source cutting edge and easy-to-use machine learning libraries, cheap computational infrastructure, and high quality, readily available training resources, data-driven closure modeling has become very popular. Since projection-based ROMs are usually constructed from snapshot data (either collected experimentally or computationally), it is natural to further exploit this set of data to estimate the closure term efficiently.

In this section, we survey data-driven closure modeling approaches in which the closure problem is cast into a regression task. The majority of the recent data-driven closure studies can be viewed as a regression task, where the closure model is defined partially or completely as a function of available information (e.g., resolved scales). We first discuss the early investigations based on classical least squares approaches. Then, we introduce ML tools that perform this regression task using neural networks and Gaussian processes regression. Finally, we explore legacy data assimilation, parameter estimation, and system identification tools that can efficiently be used to solve the closure problem. We note that some subsections are entirely devoted to closure modeling (Secs. V A and V B), some only partially address closure modeling (Secs. V C and V E), and some do not address closure modeling at all (Secs. V D, V F, and V G). Although the approaches in the latter subsections have not yet been used for closure modeling, we believe that they will soon make an impact in this dynamic research field.

### A. Trajectory regression vs model regression

There are two main schools of thought in data-driven ROM closure modeling: (i) trajectory regression and (ii) model regression.

The *trajectory regression* approach aims at finding the ROM closure model $\mathfrak{C}$ in the closed ROM [i.e., the dynamical system in Eq. (42)] that yields the best ROM trajectory. In this approach, the following constrained regression problem is solved:

$$\begin{aligned} &\underset{\mathfrak{C}\text{ parameters}}{\text{minimize}} \, ||\boldsymbol{a}^{\text{ROM}} - \boldsymbol{a}^{\text{FOM}}||^2, \\ &\text{subject to } \boldsymbol{a}^{\text{ROM}} \text{ solves closed ROM [Eq. (42)]} \end{aligned} \tag{53}$$

where $\boldsymbol{a}^{\text{FOM}}$ is the vector of ROM coefficients computed with the FOM data and $\boldsymbol{a}^{\text{ROM}}$ is the vector of ROM coefficients yielded by the closed ROM. [i.e., Eq. (42)]. We note that the trajectory regression approach is reminiscent of the a posteriori testing used in LES.[234]

The *model regression* approach aims at finding the ROM closure model $\mathfrak{C}$ that is the best approximation of the "true" closure model. In this approach, the following unconstrained regression problem is solved:

$$\underset{\mathbf{c}}{\text{minimize}} \, ||\mathfrak{C}^{\text{ROM}} - \mathfrak{C}^{\text{FOM}}||^2, \tag{54}$$

where $\mathfrak{C}^{\text{FOM}}$ is the ROM closure model computed with the FOM data, $\mathfrak{C}^{\text{ROM}}$ is the postulated ROM closure model form, and $\mathbf{c}$ is the vector of parameters used to define the ROM closure model form. We note that the model regression approach is similar in spirit to the *a priori* testing used in LES.[234]

We emphasize that the two approaches are fundamentally different. The trajectory regression is a black-box approach in which the precise formula for the ROM closure term, $\mathfrak{C}$, is not actually known. Instead, the trajectory regression first postulates a model form for $\mathfrak{C}$ (e.g., by using one of the functional models in Sec. IV A) and then solves the constrained optimization problem Eq. (53) to find the closure model parameters that yield the most accurate ROM trajectory (i.e., the trajectory $\boldsymbol{a}^{\text{ROM}}$ that is closest to the projection of the FOM data on the ROM basis, $\boldsymbol{a}^{\text{FOM}}$). In contrast, the model regression first employs one of the filters described in Sec. IV B to determine a precise formula for the ROM closure term, $\mathfrak{C}^{\text{FOM}}$. Then, it postulates a model form for $\mathfrak{C}^{\text{ROM}}$. Finally, it solves the unconstrained optimization problem Eq. (54) to find the closure model parameters that yield the most accurate ROM closure model (i.e., the closure model $\mathfrak{C}^{\text{ROM}}$ that is closest to the "true" closure model, $\mathfrak{C}^{\text{FOM}}$, computed from FOM data).

There are pros and cons for both approaches. The trajectory regression is conceptually simpler than the model regression since it does not need to determine the "true" form of the ROM closure term. The trajectory regression is also more flexible than the model regression since it can model not only the ROM closure term, but also other sources of ROM uncertainty, such as the numerical discretization error and the missing data. Finally, according to Noack's conjecture,[343] the model regression is more accurate in the predictive regime (i.e., outside the training interval), whereas the trajectory regression is more accurate in the reconstructive regime (i.e., inside the training interval). Noack motivated his conjecture by noting that using data to match models appears more robust to perturbations than using data to match trajectories. To our knowledge, Noack's conjecture has not been investigated numerically.

## B. Least squares regression

In this section, we survey the data-driven ROM closure models that use a least squares formulation in the optimization problems given by Eqs. (53) and (54).

### 1. Trajectory regression

Given its conceptual simplicity, the trajectory regression has been used from the earliest days of reduced order modeling to develop closures. The general idea used to develop these closure models is simple: (i) postulate a ROM closure model form, either functional (such as the eddy viscosity models surveyed in Sec. IV A) or structural (such as the models surveyed in Sec. IV B) and (ii) use a least squares problem in Eq. (53) to determine the various parameters in the postulated model form.

*a. Functional models.* Probably the first functional trajectory regression closure is the mixing-length model proposed in the pioneering work by Aubry *et al.*,[83] in which trajectory regression is used to determine the mixing-length constant (see Ref. 135 for related work). A least squares trajectory regression was also used by Wang *et al.*[135] to determine the eddy viscosity constants in the Smagorinsky and VMS closure models. Further improvements to the least squares trajectory regression of eddy viscosity ROM closure models were proposed by Östh *et al.*[113] and Protas *et al.*[136] The eddy viscosity trajectory regression has also been used by Stabile *et al.*,[138] Reyes and Codina,[139] and Bergmann *et al.*[256]

*b. Structural models.* Probably the first structural trajectory regression closures are those constructed with *calibration* methods, which have been introduced to directly modify (or calibrate) the GROM polynomial coefficients (i.e., $\mathfrak{B}$, $\mathfrak{L}$, and $\mathfrak{N}$), rather than introducing an additional closure term. Buffoni *et al.*[344] calibrated the constant and linear terms (i.e., $\mathfrak{B}$ and $\mathfrak{L}$), while leaving the nonlinear term, $\mathfrak{N}$, as derived from the Galerkin projection step. The modified coefficients are then found using a pseudo-spectral method such that the model prediction is as close as possible to the actual reference solution.[345] An extension to calibrate *all* the polynomial coefficients (linear and quadratic) was employed by Couplet *et al.*,[346] where the cost function is defined to penalize the deviation of the calibrated ROM behavior with respect to the projection of true snapshot data. Moreover, Perret *et al.*[347] considered a cubic polynomial to represent the dynamics of the POD modal amplitudes for supersonic jet-mixing layer data, and adopted a least squares regression to define the coefficients of the polynomial. Baiges *et al.*[348] used a calibration method in a VMS framework to build a structural trajectory regression closure. Specifically, the authors postulated a linear model for the unresolved sub-scale term as a function of the resolved field (i.e., $\mathfrak{C} = \tilde{\mathfrak{B}} + \tilde{\mathfrak{L}}\boldsymbol{a}$) and then solved the constrained least-square problem Eq. (53) for the components of $\tilde{\mathfrak{B}}$ and $\tilde{\mathfrak{L}}$.

An assessment of various calibration techniques using the two-dimensional flow around a cylinder can be found in Ref. 349. A vital merit of the calibration methods is that the computational costs of these methods are reasonable since they employ the temporal part of the POD information for the regression task, while the Galerkin projection method exploits the much more voluminous spatial POD information to construct the ROM polynomial.

### 2. Model regression

These closure models are constructed as follows: (i) use the filters surveyed in Sec. IV B to determine a mathematical formula for the true closure model, $\mathfrak{C}^{\text{FOM}}$, computed from FOM data; (ii) postulate a

ROM closure model form for the ROM closure model, $\mathfrak{C}^{\text{ROM}}$; and (iii) use the least squares problem in Eq. (54) to determine the various parameters in the postulated model form. To our knowledge, the vast majority of ROM closure model forms that have been proposed in this direction are of structural type. A notable exception is the model proposed by Hijazi *et al.*,[350] which uses the finite volume RANS data in conjunction with a model regression approach to determine the eddy viscosity component of the ROM.

A model regression closure that uses the ROM projection as a spatial filter is the data-driven VMS-ROM model proposed by Mou *et al.*[302] A two-scale version was investigated in Refs. 351 and 353, and a three-scale version was proposed in Ref. 302. The behavior of the linear,[352] quadratic,[302,351] and even cubic[353] terms were studied. A data-driven VMS-ROM to increase the pressure accuracy was proposed in Ref. 354. The verifiability of the data-driven VMS-ROM was rigorously proven by Koc *et al.*[355] Other model regression closures that use the ROM projection as a spatial filter were employed to build parameterized manifold closures by Liu and his collaborators,[303] and by Lu and his co-workers[328,329,356] to construct stochastic ROM closures. A model regression closure that uses the ROM differential filter was proposed by Koc *et al.*[308] The resulting data-driven LES-ROM uses the ROM differential filter to determine a mathematical formula for the true closure model, $\mathfrak{C}^{\text{FOM}}$, computed from FOM data. To our knowledge, the data-driven LES-ROM in Koc *et al.*[308] is the only model regression approach that utilizes a *spatial filter* (i.e., the differential filter) instead of the commonly used ROM projection.

A model regression closure that uses the PDF filtering in an MZ setting was proposed by Duraisamy *e tal.*[246,265,357] In the MZ framework, PDF filtering is used to express the true closure model, $\mathfrak{C}^{\text{FOM}}$, as a memory term, which is then approximated by using the FOM data and solving the least squares problem in Eq. (54). Moreover, a model regression closure that uses time filtering was proposed by Selten[253] and utilized in the numerical simulation of a barotropic model. Mohebujjaman *et al.*[358] used physical constraints in the model regression closure to improve the stability and accuracy of the data-driven VMS-ROM model.[302] Specifically, they equipped the least squares problem in Eq. (54) with physical constraints to enforce the regressed matrix and tensor to have similar characteristics as the GROM operators (e.g., $\tilde{\mathfrak{L}}$ being negative semi-definite and $\tilde{\mathfrak{N}}$ being energy conserving). A similar quadratic formula was adopted in Ref. 359 to recover the hidden physical processes (e.g., source terms) for a system with incomplete governing equations.

## C. Neural network regression

The introduction of neural network regression into ROM was highly motivated by the desire to construct purely data-driven nonintrusive ROM (NIROM) frameworks,[8,360] which solely rely on data to learn the dynamics of the relevant solution manifold without the need to access the governing equations. Nonintrusive approaches are attractive due to their portability since they do not necessarily require the exact form of the equations or the methods used to generate the data. In addition, nonintrusive models offer a unique advantage in multidisciplinary collaborative environments, where only data can be shared without revealing the proprietary or sensitive information. Nonintrusive approaches are also useful when the detailed governing equations of the problem are unknown. This modeling approach can benefit from the enormous amount of data collected from experiments, sensor measurements, and large-scale simulations to build NIROMs based on the assumption that data are a manifestation of *all* the underlying dynamics and processes.

Machine learning tools, in particular the artificial neural networks (ANNs) equipped with the universal approximation theorem,[361] have been widely exploited in this regard. A typical feed-forward neural network is depicted in Fig. 5, where a mapping $\boldsymbol{\mathcal{M}}$ from the input $\boldsymbol{\mathcal{X}}$ to the output $\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{M}}(\boldsymbol{\mathcal{X}})$ is inferred through a learning algorithm. For transient flows, a single-layer feed-forward neural network was proposed by San *et al.*[362] to provide accurate predictions of the ROM coefficients with varying control parameter values, using sequential and residual approaches. In the sequential approach, a mapping from the current values of $\boldsymbol{a}$ to their future values is approximated. Moreover, the input layer is augmented with the acting Reynolds number and the time. That is, $\boldsymbol{\mathcal{X}} = \{\text{Re}, t_n, \boldsymbol{a}(t_n)\}$, while $\boldsymbol{\mathcal{Y}} = \{\boldsymbol{a}(t_{n+1})\}$. On the other hand, the residual implementation relies on learning the deviation of the future state from its current value (i.e., $\boldsymbol{\mathcal{Y}} = \{\boldsymbol{a}(t_{n+1}) - \boldsymbol{a}(t_n)\}$). Pawar *et al.*[363] employed deep neural networks (DNN) to bypass the Galerkin projection step and build a fully NIROM for the two-dimensional Boussinesq equations with a differentially heated cavity flow setup at various Rayleigh numbers. In particular, the evolution of the POD modal amplitudes $\boldsymbol{a}(t_{n+1})$ was predicted from their past values using residual and backward difference scheme formulas. The application of variants of ANNs as regression models for the dynamics of low-order states (e.g., POD amplitudes) has gained substantial popularity,[364–366] owing to the availability of open-source and user-friendly ML libraries. This is a hot topic and dozens of new papers appear every week in different journals and conferences all over the world, dealing with different aspects of NIROM based on ANNs (e.g., different architectures, test bed problem, and error bounds).

Recurrent neural networks (RNNs) are very effective for sequence predictions in numerous applications, e.g., speech

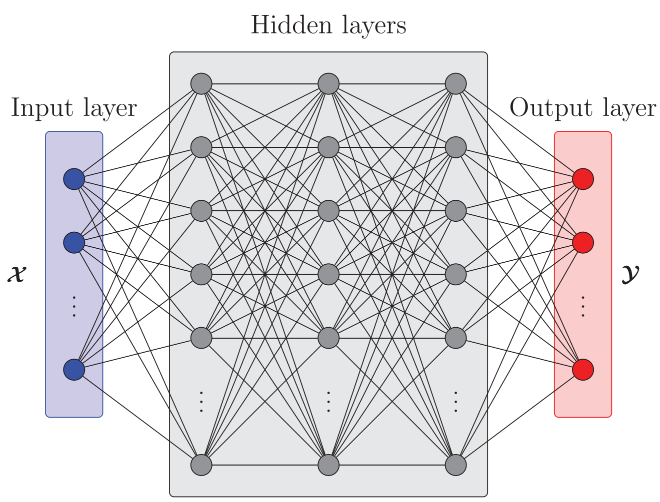


**FIG. 5.** A schematic diagram of a typical feed-forward neural network with an input layer, hidden layers, and an output layer. We note that these general-purpose dense deep network architectures have been evolving to more specific neural network designs.[367] We discuss some of them briefly in Sec. VI.

recognition and translation. RNNs contain loops that allow them to retain information from one time step to another so as to enforce the temporal dependencies. A deep residual RNN was utilized by Kani and Elsheikh[368] for the model reduction of nonlinear dynamical systems. However, one of the limitations of RNNs is vanishing (or exploding) gradient to capture the long-term dependencies, stemming from the repetitive multiplication of gradient with potentially ill-conditioned weight matrices during the backpropagation learning algorithm. The long short-term memory (LSTM) neural networks mitigate this issue by employing a gating mechanism that allows information to be forgotten. Vlachas *et al.*[369] trained an LSTM to predict the derivative of $\boldsymbol{a}$ with respect to time from a short history of $\boldsymbol{a}$ values, where a first-order forward difference scheme was adopted to represent the temporal derivative. They also combined the LSTM with a mean stochastic model to cope with attractor regions that are not captured in the training set. Mohan and Gaitonde[370] explored the bidirectional variant of LSTM, employing two LSTM networks: one in the forward and the other in the reverse direction, for NIROM of forced isotropic turbulence and magneto-hydrodynamic turbulence using the Johns Hopkins turbulence database.[371] Rahman *et al.*[372] utilized LSTMs for the NIROM implementation of the two-dimensional single-layer quasi-geostrophic ocean circulation model. A sliding window approach was adopted to predict the evolution of the POD amplitudes. In another interesting article, Wang *et al.*[273] utilized a conditioned LSTM for the memory term in the GROM equations, representing the closure model, for parametric systems.

Instead of using ML to entirely replace the GROM with NIROM, data-driven ML can be utilized along with the physics-based GROM to construct the closure model. This hybrid approach was adopted in Ref. 268, which used a dissipative term employing an eddy viscosity coefficient and utilized a single layer extreme learning machine (ELM) to estimate a modal $\nu_e$ as a function of the mode index, GROM right-hand side (RHS), and modal amplitudes. Furthermore, a clipping procedure was carried out by discarding the negative values of $\nu_e$. ML tools can be exploited to provide the numerical value of the closure term, without constraints on the functional form of the closure model. San and Maulik[269,373] utilized an ELM[374] to learn the value of the closure term as a function of the GROM RHS, i.e., $\mathfrak{C} = f(\mathfrak{B} + \mathfrak{L}\boldsymbol{a} + \boldsymbol{a}^\top \mathfrak{N}\boldsymbol{a})$. For training purposes, the true closure term is computed from the projection of the pure evolution of the PDE onto the reduced space. In other words, Eq. (29) is first evaluated in the FOM space, then projected onto the basis functions, while the GROM RHS [Eq. (33)] is computed directly in the reduced space.

Wan *et al.*[375] utilized LSTM architectures to learn the mismatch resulting from the imperfect GROM RHS as a function of the sequence of past values of the resolved state. The ML model is exploited to assist the imperfect model whenever data are available, while for locations with sparse data, the GROM still provides an acceptable baseline for the prediction of the system state. The estimation of the closure term as a function of the time history of the resolved scales has roots in the Mori–Zwanzig formalism,[325,376] and the memory embedding of LSTM implementation for closure modeling is supported by the Takens embedding theorem.[377] Gupta and Lermusiaux[275] employed RNNs to learn the non-Markovian term appearing in Eq. (48) using the neural delay differential equations with discrete delays as follows:

$$\frac{d\boldsymbol{a}}{dt} \approx \boldsymbol{f}(\boldsymbol{a}) + g_{RNN}(\boldsymbol{a}(t), \boldsymbol{a}(t-\tau_1), \ldots, \boldsymbol{a}(t-\tau_K); \theta), \tag{55}$$

where $K$ is the number of discrete-delays and $\theta$ represents the neural network weights. For distributed delays, the ROM evolution is written as

$$\frac{d\boldsymbol{a}}{dt} \approx \boldsymbol{f}(\boldsymbol{a}) + g\left(\boldsymbol{a}(t), \int_{t-\tau_2}^{t-\tau_1} h(\boldsymbol{a}(\tau), \tau) d\tau, t\right), \tag{56}$$

where the delay is distributed over past time periods $t-\tau_2$ and $t-\tau_1$. Gupta and Lermusiaux[275] approximated the $g$ and $h$ functions defining the delay term using two different coupled neural networks as follows:

$$\frac{d\boldsymbol{a}}{dt} \approx \boldsymbol{f}(\boldsymbol{a}) + g_{NN}(\boldsymbol{a}(t), y(t), t; \theta_g), \tag{57}$$

$$\frac{dy}{dt} \approx h_{NN}(\boldsymbol{a}(t-\tau_1), t-\tau_1; \theta_h) - h_{NN}(\boldsymbol{a}(t-\tau_2), t-\tau_2; \theta_h), \tag{58}$$

where the memory effect can be embedded without the use of any specific recurrent neural network. The authors showed that the non-Marovian closure outperforms its Marovian counterpart (with no time delays).

## D. Kernel regression

Gaussian process regression (GPR)[378] has the advantage of the simultaneous prediction of both the system's dynamics and the associated uncertainty. Although GPR provides a powerful tool for probabilistic inference that enables modelers to strike a balance between model complexity and data fitting,[379] its use is often limited to relatively small training datasets due to the well-known cubic scaling characteristics of the Gaussian processes. However, fast algorithms using approximate matrix-vector products can be utilized for large datasets.[380,381] GPR has gained prominence providing surrogate models for complex and multidimensional systems.[382–384] In ROM applications, these confidence measures can be particularly informative when the ROM dimension is lower than the intrinsic dimension of the system. Wan and Sapsis[385] formulated a stochastic model based on GPR dynamics and utilized a Monte Carlo framework for the forecast of the system's state and corresponding uncertainty. Maulik *et al.*[386] utilized a Gaussian process emulator for the dynamical evolution of the latent space state variables, obtained from POD and autoencoder compression, for the shallow water equations. We also note that Raissi and Karniadakis[379] developed a Gaussian process framework to learn PDEs from relatively small quantities of data. Xiao *et al.*[387] applied a second-order Taylor series scheme and a Smolyak sparse grid collocation method to calculate the POD modal coefficients at each time step from their values at earlier time steps. A radial basis function (RBF) multidimensional interpolation was used in Ref. 388 for similar purposes. RBF interpolations have been also utilized for parameterized problems as mappings from the parameter space to the ROM space.[389]

## E. Data assimilation and error correction

An excellent discussion of the similarities between ML and data assimilation (DA) tools has been recently provided by Alan Geer.[390] The synergistic integration of ML and DA is essential for developing improved approaches.[391–402] Looking at the ROM framework from a data assimilation point of view opens up innovative avenues to tackle the closure problem. Ideas from optimal control theory, data assimilation, and parameter estimation were proven to be valuable in this regard. DA is a generic framework combining the available observations with the underlying dynamical principles governing the system to estimate the physical quantities of interest. This is usually accomplished by starting from a background solution and computing an optimal estimation of the true state of the system that minimizes the discrepancy between the model predictions and collected observations. This minimization problem is solved while taking into account the respective statistical confidence of different observations, background solution, and model uncertainty.[401] In order to emphasize the model's error (due to truncation), D'Adamo *et al.*[402] added a Gaussian variable to the ROM equation. This is similar to the weak variational data assimilation framework implemented in Ref. 403 for ROMs using real experimental conditions with noisy particle image velocimetry data. The GROM is augmented with an additive stochastic control variable, representing the model's uncertainty that reflects the effect of unresolved scales on the resolved dynamics. Estimating such an uncertainty function can be incorporated to improve the ROM predictions. Zerfas *et al.*[404] utilized the nudging algorithm to improve predictions by adding a feedback control term that nudges the ROM approximation toward the reference solution corresponding to the observed data. The authors also presented a strategy to dynamically adjust the nudging parameter by controlling the dissipation arising from the nudging term as well as a numerical analysis of the proposed DA-ROM. A combination of the nudging methodology and LSTM framework was adopted by Ahmed *et al.*[405,406] to correct the ROM trajectory, considering the initial condition mismatch and GROM model deficiency.

DA has been also utilized to *calibrate* the ROM coefficients so that the ROM predictions agree with the available observations. Cordier *et al.*[112] adopted a four-dimensional variational (4DVAR) formulation to tune the computed ROM coefficients, where the background values were obtained from standard Galerkin projection. The projection of snapshot data onto the POD basis was treated as synthetic observations of the reduced system's state. Although a good match was observed *within* the assimilation window, the model's stability on the forecast window was not ensured. This is similar to the strong variational formulation in Ref. 402 aiming at directly correcting the model's coefficients assuming a deterministic dynamical model.

DA tools can also be exploited to provide a good estimate of the free parameters in classical closure models. For example, Cordier *et al.*[112] adopted the nonlinear eddy viscosity model from Refs. 254 and 300 and applied the 4DVAR framework to estimate the eddy viscosity parameter to increase the physical reliability of the model *beyond* the assimilation window. More recently, Ahmed *et al.*[407] used a linear eddy viscosity model and exploited the forward sensitivity method (FSM) to compute and update the mode-dependent eddy viscosity parameters. Given the plummeting costs of sensors and the potential of ROM in the real-time monitoring and control, we emphasize that DA appears to be a good candidate for future developments leveraging the increasingly available heterogeneous measurement data to build more robust ROM closures.

## F. Operator inference approaches

An *operator inference* (OI) approach was proposed by Peherstorfer and Willcox[116] to infer the ROM operators from data. Next, we briefly outline the OI approach. First, we note that the quadratic term in Eq. (33) can be written as $[\boldsymbol{a}^\top \mathfrak{N} \boldsymbol{a}]_k = \sum_{i=1}^r \sum_{j=1}^r \mathfrak{N}_{ijk} a_i a_j$. In Ref. 116, this is rewritten as a matrix-vector product to exploit the commutative property of multiplication and avoid redundancy (i.e., we consider a single term $a_i a_j$ as a representative of both $a_i a_j$ and $a_j a_i$) as follows:

$$\frac{\mathrm{d}\boldsymbol{a}}{\mathrm{d}t} = \mathfrak{B} + \mathfrak{L}\boldsymbol{a} + \mathfrak{D}\boldsymbol{a}^2, \tag{59}$$

where $\mathfrak{D} \in \mathbb{R}^{r \times r(r+1)/2}$ is the quadratic operator and $\boldsymbol{a}^2 = [\boldsymbol{a}^{(1)^\top}, \boldsymbol{a}^{(2)^\top}, \ldots, \boldsymbol{a}^{(r)^\top}]^\top \in \mathbb{R}^{r(r+1)/2}$, with $\boldsymbol{a}^{(i)} \in \mathbb{R}^i$ defined as

$$\boldsymbol{a}^{(i)} = a_i \begin{bmatrix} a_1 \\ \vdots \\ a_i \end{bmatrix}. \tag{60}$$

Then, the components of $\mathfrak{B}$, $\mathfrak{L}$, and $\mathfrak{D}$ are computed by solving $r$ least squares problems, corresponding to each mode dynamics (i.e., $\frac{\mathrm{d}a_k}{\mathrm{d}t}$). The OI algorithm in Ref. 116 can be extended to any arbitrary polynomial nonlinear terms in the state. However, the computational cost grows exponentially with the order of the polynomial nonlinear term rendering it feasible only for low-order polynomials. We highlight that even if the Galerkin projection step is not often required in the previous calibration (and OI) studies, it is generally assumed that the true governing equation has a quadratic (or polynomial) structure. This limitation was addressed in Ref. 408, which introduced a lift & learn approach, where the ROM polynomial coefficients are efficiently calibrated even if the high-dimensional dynamics are not quadratic using lifting transformations. Recent OI developments are discussed in Refs. 409–412.

## G. System identification approaches

In many fluid dynamics applications, system identification approaches become viable tools to identify nonlinear low-order models.[413] However, robust identification of realistic dynamical systems constitutes a grand challenge. For example, let us decompose the flow into two parts:

$$\text{Flow} = \text{resolved} + \text{unresolved dynamics} \tag{61}$$

or

$$\boldsymbol{u} = \boldsymbol{v} + \boldsymbol{w}. \tag{62}$$

Then, the evolution equation for $\boldsymbol{v}$ can be abstracted as follows:

$$\frac{d\boldsymbol{v}}{dt} = \boldsymbol{f}(\boldsymbol{v}, \boldsymbol{w}) \doteq \boldsymbol{f}(\boldsymbol{v}, 0) + \boldsymbol{g}(\boldsymbol{v}, \boldsymbol{w}). \tag{63}$$

The unresolved dynamics term, $\boldsymbol{g}$, has a high-frequency stochastic component important for short-term dynamics and an energy-absorbing component important for long-term boundedness. The first component can be modeled by a stochastic term, and the second by an eddy viscosity model, as we discussed earlier.

In general, data-driven model identification (without priors) requires full data. With full data a $k$-nearest neighbor ($k$-NN) model for kinematics and dynamics should work. With sparse data, no model identification might work. The lack of data has to be compensated by priors/knowledge. There are two issues here:

*Warning 1.* Lack of resolution (unknown $\boldsymbol{w}$) leads to one closure problem. Lack of data (under-resolved $\boldsymbol{v}$) leads to another closure problem.
*Warning 2.* The model complexity and data richness are strongly interwoven.[414]

That being said, there are many methods for model identification, starting with brute-force data interpolation.[121] Although many of the methodologies presented in Sec. V E can be viewed as system identification approaches, we dedicate the following discussion to the efforts and potential opportunities that aim at revealing the mathematical representation of the closure term. This strategy is fundamentally different from assuming a specific form for model closure and fitting it to data to compute the unknown parameters and/or coefficients. In this regard, symbolic regression (SR) techniques have been recently exploited to identify interpretable closed form approximations of the governing equations, by observing the dynamical behavior and response of the system of interest.[415,416] SR techniques can be largely classified into two categories: (1) approaches that utilize compressed sensing and sparsity-promoting techniques to choose a few functions from a large feature library of potential basis functions that have the expressive power to define the dynamics; (2) evolutionary algorithms that search for functional abstractions with a preselected set of *basic* mathematical operators and operands. Examples that belong to the first category include the sparse identification of nonlinear dynamics (SINDy) framework,[117,417] the sequential threshold ridge regression (STRidge) algorithm,[418] and the PDE-functional identification of nonlinear dynamics (PDE-FIND) technique,[418] while genetic programing (GP)[419–421] and gene expression programing (GEP)[422] represent the major drivers for evolutionary SR explorations. Loiseau *et al.*[423] utilized SINDy to identify a system of nonlinearly coupled ODEs governing the evolution of the first pair of POD modes' amplitudes (i.e., $a_1$ and $a_2$) for the 2D flow over a cylinder. Considering monomials of $a_1$ and $a_2$, a library of candidate functions is constructed as follows:

$$\boldsymbol{\Theta} = \begin{bmatrix} 1 & a_1 & a_2 & a_1^2 & a_1 a_2 & a_2^2 & a_1^3 & a_1^2 a_2 & a_1 a_2^2 & a_2^3 \end{bmatrix}. \quad (64)$$

The identified ROM equations thus take the form

$$\frac{\mathrm{d}\boldsymbol{a}}{\mathrm{d}t} = \boldsymbol{\Theta}\boldsymbol{\zeta}, \quad (65)$$

where $\boldsymbol{\zeta}$ encapsulates the coefficient of each candidate function, computed using a sparsity-promoting regression problem. Note that the library $\boldsymbol{\Theta}$ can be enriched with any arbitrary functions that potentially describe the system's dynamics. Since the system given in Eq. (65) is distilled from data, the effect of truncated modes on $a_1$ and $a_2$ (i.e., the closure model) is inherited in the identified model. Indeed, the results of integrating the model derived by SINDy outperformed those from standard GROM. A recent innovation,[424] trapping SINDy, identifies dynamics with a trapping region, i.e., guarantees boundedness of the solution.[115] This innovation is particularly important for higher-dimensional dynamics, where sparse identification is prone to give rise to unbounded solutions otherwise.

SR approaches have been also pursued to discover *discrepancy models* and reveal the hidden physics and dynamical processes that are not represented in the available governing equations. Vaddireddy *et al.*[425] applied the GEP and STRidge algorithms to recover the hidden physics (e.g., source or forcing terms) in the 2D Navier–Stokes equations using Eulerian sensor measurements. Likewise, Kaheman *et al.*[426] utilized SINDy to model the mismatch between simplified models and measurement data. With particular relevance to closure modeling, Vaddireddy *et al.*[425] also demonstrated the application of GEP and STRidge to identify the truncation error due to numerical discretization, and recover the eddy viscosity kernels, manifested as source terms in the LES equations. With the ongoing advancement and maturity of SR tools, great leaps in closure modeling are expected, especially given the lack of physical intuition in a ROM context.

## VI. PHYSICS-INFORMED DATA-DRIVEN MODELING

In most of the works related to NIROMs, the general idea is to employ the ML model to learn the temporal evolution of the field variable in the reduced order subspace and thereby bypass the intrusive Galerkin projection. Gao *et al.*[427] proposed a framework that utilizes a fully connected neural network to approximate the nonlinear velocity function in the ROM equations by leveraging the same data used for the reduced basis construction. They illustrated the stable performance of the framework in the parametric viscous Burgers equation and two-dimensional premixed $H_2$-air flame model. Xu and Duraisamy[428] proposed a data-driven framework composed of a convolutional encoder to identify nonlinear basis functions, a temporal convolutional encoder to learn the temporal dynamics of latent variables, and a fully connected neural network to learn the mapping between the parameters of the system and the latent variables. They demonstrated the predictive performance of this framework for problems involving discontinuities, wave propagation, and coherent structures. Although NIROMs have been successful for many complex nonlinear problems, the typical sparsity of data motivates the physics-informed data-driven modeling.

Remark. We note that most computational modeling approaches require (a) data and (b) physics knowledge. The evolution equations are computationally too demanding and numerical discretization methods are not feasible for multi-query tasks. The data-driven methods are computationally efficient, but data-hungry in nature. For example, the autoencoders require a lot of data to build a generalizable ROM that cannot be derived from first principles. However, the reduced dynamics on the autoencoder need to follow first-principles dynamics, which can be imposed as a prior given sparse data.

The neural networks are one of the most popular algorithms for learning the reduced order dynamics of nonlinear problems. However, as the complexity of the system increases, the depth of the neural network also grows to learn the complicated nonlinear dynamics, and the number of trainable parameters explodes quickly. In the presence of sparse data, deep neural networks exhibit high epistemic uncertainty, and this adversely affects the trustworthiness of NIROMs. In these situations, it is important to inject physical relationships explicitly or implicitly in the training process or the model architecture. For example, partial differential equations, conservation laws, and symmetries can be exploited toward building physics-informed data-driven models. This will allow the training of deep neural networks with limited training data, speed up the training process, and also ensure physically consistent prediction. To this end, different approaches have been proposed in scientific machine learning, e.g., cost function modification to accommodate the model Jacobian,[429] grow-when-required network,[430] physics-informed neural networks,[431–434] embedding *hard* physical constraints in a neural network,[435,436] physics-based feature extraction,[437] leveraging uncertainty information,[438] developing visualization tool of the network analysis,[439] physics-guided machine learning,[440,441] and hybrid modeling.[20,238,406,442] Readers are referred to the recent review article by Karniadakis *et al.*[434] for a detailed discussion of embedding physics into machine learning for tackling scientific problems. In recent years, innovative applications of tailored neural network architectures have become increasingly common among fluid dynamicists. A central question in many studies is often how to exploit prior knowledge about the problem at hand to build more trustworthy models.[443] Various hybrid modeling principles that aim at combining the machine learning and data-driven models with the physics-based models have been recently discussed by San *et al.*[20]

The physics-informed data-driven modeling has been successful in many applications, such as turbulence closure modeling,[444] super-resolution of turbulent flows,[445,446] and generative modeling of dynamical systems,[447–449] and it holds a great potential for ROMs. Next, we present several recent developments in physics-informed data-driven modeling. Chen *et al.*[450] built the physics-reinforced neural network (PRNN) trained by minimizing the mean squared residual error of the reduced order equations, and the mean squared error between the neural network prediction and the projection of high-fidelity data on the reduced basis. The incorporation of reduced order equations in the loss function can be considered as a physics-based regularization. The PRNN was demonstrated to be more accurate than a purely data-driven neural network for complex nonlinear problems. Mohan *et al.*[435] proposed a physics-embedded convolutional autoencoder (PhyCAE) in which the divergence-free condition is imposed as a hard constraint through non-trainable layers after the decoder. The PhyCAE combined with a recurrent neural network for modal coefficients prediction can provide a physics-constrained NIROM that satisfies the conservation laws. Lee and Carlberg[451] developed a framework that computes the lower-dimensional embedding using a convolutional autoencoder and enforces physical conservation laws by modeling the latent-dynamics as a solution to a constrained optimization problem. The objective function of the optimization problem is defined as the sum of squares of conservation-law violations over a control volume of the finite volume discretization.

Kaptanoglu *et al.*[452] developed a physics-constrained low-dimensional model for magnetohydrodynamics by enforcing symmetries derived from global conservation laws into data-driven models. Sawant *et al.*[453] proposed a physics-informed regularizer and structure preserving (such as symmetry and definiteness in linear terms) OI formulation and demonstrated their framework's performance in terms of improved stability and accuracy for nonlinear systems. The OI framework with physics-based regularization has also been applied for building predictive ROMs for rocket engine combustion dynamics.[454] A constrained sparse Galerkin approach has been introduced by Loiseau and Brunton.[413] Finally, Mohebujjaman *et al.*[358] used physical constraints to increase the stability and accuracy of their data-driven variational multiscale ROM closure framework.

There are a number of limitations of data-driven methods, including extrapolation beyond the training dataset, the curse of dimensionality, stability issues, and boundedness of the model. Many of these potential challenges can be mitigated using the decades of progress in physics-based modeling. For example, one way to address the curse of dimensionality, i.e., when the optimization problem can become highly non-convex and computationally expensive for very high-dimensional systems, is to tailor the feature space based on prior knowledge about the system.[455] Furthermore, the extrapolation beyond training dataset, which is a central challenge in many data-driven methods, can be effectively tackled by enforcing conservation laws or by using a custom neural network architecture that incorporates prior information about the system at hand. Finally, as the research in physics-informed data-driven modeling is rapidly progressing, reproducible research through open-source benchmark datasets and test cases should also be developed.

## VII. CONCLUSIONS AND OUTLOOK

The ever-increasing need for computational efficiency and improved accuracy of many applications leads to very large-scale dynamical systems whose simulations and analyses make unmanageable demands on computational resources. Significantly simplifying the computational complexity of the underlying mathematical model, ROMs offer promise in many applications, like shape optimization, uncertainty quantification, and control. Over the past decades, for example, DNS, LES, and RANS have made a tremendous impact in the numerical simulation of turbulent flows. However, these FOMs cannot be generally used in many query applications because of their prohibitively high computational cost.

ROMs are efficient, relatively low-dimensional models often created from available data. In fluid dynamics, ROMs have been successfully used as surrogate models for structure-dominated problems, mainly in simple, academic test problems. However, traditional ROMs generally fail for more realistic flows because a low-dimensional ROM basis cannot accurately represent the complex dynamics, a topic that vastly waits for new explorations. Our paper provides a glimpse into various approaches to generating accurate ROMs, focusing on phenomenological, mathematical, and data-centric closure modeling approaches. We primarily focus on subspace projection-based methods and discuss their feasibility for various nonlinear problems in fluid dynamics. A chief emphasis is given to closure models and, in particular, to the analogy between LES and ROMs. Various methodologies are leveraged and examples are included to provide a broader overview of forward-looking reduced order modeling practices in the age of data. Of course, our paper is only a first step in an exhaustive discussion of ROM closure modeling. Although we tried to include as many relevant contributions as possible, we left out important developments (e.g., stochastic closure modeling, compressible flows, and ROM pressure approximations). We hope, however, that our paper will serve as a stepping stone toward a more comprehensive discussion of the exciting research areas of data-driven modeling and ROM closures, where these and many other new developments will be carefully presented.

We envision that ROMs will be a key enabler in the development of a big data cybernetics infrastructure, an approach to controlling an asset or process using real-time big data. These tools and concepts offer many new perspectives to our rapidly digitized society and its seamless interactions with many different fields. With the recent wave of digitalization, the latest trend in every industry is to build systems and approaches that will help not only during the conceptualization, prototyping, testing, and design optimization phase but also during the operation phase with the ultimate aim to use them throughout the whole product life cycle. While the numerical simulation tools and lab-scale experiments are clearly important in the first phase, the potential of real-time availability of data in the operational phase is opening up new avenues for monitoring and improving operations throughout the life cycle of a product. We believe that ROMs will be crucial to the improvement of emerging digital twin technologies.

Currently, the development of robust monolithic ROMs for a single operating condition with adequate data is a well-established art. We remark that "sufficient data" for post-transient dynamics implies that all snapshots will be approximately revisited multiple times. This is already quite a challenge for turbulent flows given the stochasticity of the dynamics.

A much more common task is modeling transient, controlled, or multiparametric dynamics. For instance, an airplane needs to be designed to prevent flutter under a large range of operating conditions, e.g., velocity, angle of ascent or descent, position of flaps, and maneuvers. There will never be enough data for these cases and the large terra incognitae (oceans of missing data) have to be replaced by prior knowledge or clever guesses. Hence, physics-informed data-driven ROMs will become a necessity for most applications.

Most likely, such a range of dynamics will not be facilitated by a single "monolithic" reduced order representation, but a large set of representations, leading to a large set of ROMs with overlapping domains of validity. This trend is already foreshaddowed for the transient cylinder wake. An accurate reduced order representation requires about 50 POD modes[121] or, staying in the Galerkin framework, a set of adjustable mean-field and adjustable Galerkin expansion modes.[456]

Another requirement is human interpretability of the kinematics and dynamics. In an ideal case, this leads to a low-dimensional manifold for the data and a sparse representation for the dynamics. Most likely, only cartographic visualization of the state space and the dynamics will be achievable. One example is the cluster-based network with many centroids for many operating conditions and a transition network for the dynamics.

As the treasures of high-quality experimental and numerical data and the spectrum of increasingly powerful methods are evolving, an open-source distribution of data and methods becomes of increasing importance. Hitherto, sharing of data and methods is still in its infancy and requires dedicated efforts. A noteworthy positive development is ever-increasing number of journals that encourage the publication of data and methods.

We conclude with two guidelines which are independent of the field but may easily be overlooked. First, there is the need for a well chosen set of guiding benchmark problems serving as lighthouses for model development. The second advice is best formulated by Harrington Emerson (1853–1931, efficiency engineer):

> "As to methods there may be a million and then some, but principles are few. The man who grasps principles can successfully select his own methods. The man who tries methods, ignoring principles, is sure to have trouble."

## ACKNOWLEDGMENTS

O.S. gratefully acknowledges the U.S. DOE Early Career Research Program support No. DE-SC0019290 and the National Science Foundation support No. DMS-2012255. A.R. acknowledges the financial support received by the Research Council of Norway and the industrial partners of the following projects: EXAIGON–*Explainable AI systems for gradual industry adoption* (Grant No. 304843), *Hole cleaning monitoring in drilling with distributed sensors and hybrid methods* (Grant No. 308823), and RaPiD–*Reciprocal Physics and Data-driven models* (Grant No. 313909). T.I. acknowledges the support through the National Science Foundation Grant No. DMS-2012253. B.R.N. acknowledges the funding from the Harbin Institute of Technology, Shenzhen (Starting grant), the Peacock Talent A Plan from Shenzhen Government, and the German Science Foundation (DFG) via Grant No. SE 2504/3-1.

## DATA AVAILABILITY

The data that support the findings of this study are available within the article.

## REFERENCES

[1] I. Marusic and S. Broomhall, "Leonardo da Vinci and fluid mechanics," Annu. Rev. Fluid Mech. **53**, 1–25 (2021).
[2] J. L. Lumley, "The structure of inhomogeneous turbulent flows," *Atmospheric Turbulence and Radio Wave Propagation* (Nauka, Moscow, 1967).
[3] J. Kou and W. Zhang, "Data-driven modeling for unsteady aerodynamics and aeroelasticity," Prog. Aerosp. Sci. **125**, 100725 (2021).
[4] A. Beck and M. Kurz, "A perspective on machine learning methods in turbulence modeling," GAMM-Mitt. **44**, e202100002 (2021).
[5] S. L. Brunton, B. R. Noack, and P. Koumoutsakos, "Machine learning for fluid mechanics," Annu. Rev. Fluid Mech. **52**, 477–508 (2020).
[6] K. Taira, M. S. Hemati, S. L. Brunton, Y. Sun, K. Duraisamy, S. Bagheri, S. T. Dawson, and C.-A. Yeh, "Modal analysis of fluid flows: Applications and outlook," AIAA J. **58**, 998–1022 (2020).
[7] G. Mendonça, F. Afonso, and F. Lau, "Model order reduction in aerodynamics: Review and applications," Proc. Inst. Mech. Eng., Part G **233**, 5816–5836 (2019).
[8] J. Yu, C. Yan, and M. Guo, "Non-intrusive reduced-order modeling for fluid problems: A brief review," Proc. Inst. Mech. Eng., Part G **233**, 5896–5912 (2019).
[9] M. P. Brenner, J. D. Eldredge, and J. B. Freund, "Perspective on machine learning for advancing fluid mechanics," Phys. Rev. Fluids **4**, 100501 (2019).
[10] R. Yondo, K. Bobrowski, E. Andrés, and E. Valero, "A review of surrogate modeling techniques for aerodynamic analysis and optimization: Current limitations and future challenges in industry," in *Advances in Evolutionary and Deterministic Methods for Design, Optimization and Control in Engineering and Sciences* (Springer, Berlin, 2019), pp. 19–33.
[11] K. Taira, S. L. Brunton, S. 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).
[12] C. W. Rowley and S. T. M. Dawson, "Model reduction for flow analysis and control," Ann. Rev. Fluid Mech. **49**, 387–417 (2017).
[13] S. L. Brunton and B. R. Noack, "Closed-loop turbulence control: Progress and challenges," Appl. Mech. Rev. **67**, 050801 (2015).
[14] T. Lassila, A. Manzoni, A. Quarteroni, and G. Rozza, "Model order reduction in fluid dynamics: Challenges and perspectives," in *Reduced Order Methods for Modeling and Computational Reduction* (Springer, Berlin, 2014), pp. 235–273.
[15] I. Mezić, "Analysis of fluid flows via spectral properties of the Koopman operator," Annu. Rev. Fluid Mech. **45**, 357–378 (2013).
[16] F. Chinesta, P. Ladeveze, and E. Cueto, "A short review on model order reduction based on proper generalized decomposition," Arch. Comput. Methods Eng. **18**, 395 (2011).
[17] N. Massarotti, P. Nithiarasu, E. Samadiani, and Y. Joshi, "Reduced order thermal modeling of data centers via proper orthogonal decomposition: A review," Int. J. Numer. Methods Heat Fluid Flow **20**, 529–550 (2010).
[18] D. J. Lucia, P. S. Beran, and W. A. Silva, "Reduced-order modeling: New approaches for computational physics," Prog. Aerosp. Sci. **40**, 51–117 (2004).

[19]G. Berkooz, P. Holmes, and J. L. Lumley, "The proper orthogonal decomposition in the analysis of turbulent flows," Annu. Rev. Fluid Mech. **25**, 539–575 (1993).
[20]O. San, A. Rasheed, and T. Kvamsdal, "Hybrid analysis and modeling, eclecticism, and multifidelity computing toward digital twin revolution," GAMM-Mitt. **44**, e202100007 (2021).
[21]A. Heinlein, A. Klawonn, M. Lanser, and J. Weber, "Combining machine learning and domain decomposition methods for the solution of partial differential equations-a review," GAMM-Mitt. **44**, e202100001 (2021).
[22]J. Blechschmidt and O. G. Ernst, "Three ways to solve partial differential equations with neural networks-a review," GAMM-Mitt. **44**, e202100006 (2021).
[23]K. Kashinath, M. Mustafa, A. Albert *et al.*, "Physics-informed machine learning: Case studies for weather and climate modelling," Proc. R. Soc. A **379**, 20200093 (2021).
[24]P. Bauer, B. Stevens, and W. Hazeleger, "A digital twin of earth for the green transition," Nat. Clim. Change **11**, 80–83 (2021).
[25]M. Frank, D. Drikakis, and V. Charissis, "Machine-learning methods for computational science and engineering," Computation **8**, 15 (2020).
[26]S. Zhu and A. P. Piotrowski, "River/stream water temperature forecasting using artificial intelligence models: A systematic review," Acta Geophys. **68**, 1433–1442 (2020).
[27]P. Tahmasebi, S. Kamrava, T. Bai, and M. Sahimi, "Machine learning in geo- and environmental sciences: From small to large scale," Adv. Water Resour. **142**, 103619 (2020).
[28]P. Benner, S. Gugercin, and K. Willcox, "A survey of projection-based model reduction methods for parametric dynamical systems," SIAM Rev. **57**, 483–531 (2015).
[29]M. J. Asher, B. F. Croke, A. J. Jakeman, and L. J. Peeters, "A review of surrogate models and their application to groundwater modeling," Water Resour. Res. **51**, 5957–5973, https://doi.org/10.1002/2015WR016967 (2015).
[30]V. Machairas, A. Tsangrassoulis, and K. Axarli, "Algorithms for optimization of building design: A review," Renewable Sustainable Energy Rev. **31**, 101–112 (2014).
[31]M. P. Mignolet, A. Przekop, S. A. Rizzi, and S. M. Spottswood, "A review of indirect/non-intrusive reduced order modeling of nonlinear geometric structures," J. Sound Vib. **332**, 2437–2460 (2013).
[32]M. A. Bazaz and S. Janardhanan, "A review of parametric model order reduction techniques," in *IEEE International Conference on Signal Processing, Computing and Control* (IEEE, 2012), pp. 1–6.
[33]S. Razavi, B. A. Tolson, and D. H. Burn, "Review of surrogate modeling in water resources," Water Resour. Res. **48**, W07401, https://doi.org/10.1029/2011WR011527 (2012).
[34]C. Theodoropoulos, "Optimisation and linear control of large scale nonlinear systems: A review and a suite of model reduction-based techniques," in *Coping with Complexity: Model Reduction and Data Analysis* (Springer, Berlin, 2011), pp. 37–61.
[35]M. B. Wagner, A. Younan, P. Allaire, and R. Cogill, "Model reduction methods for rotor dynamic analysis: A survey and review," Int. J. Rotating Mach. **2010**, 273716 (2010).
[36]J. P. Kleijnen, "Kriging metamodeling in simulation: A review," Eur. J. Oper. Res. **192**, 707–716 (2009).
[37]R. Pinnau, "Model reduction via proper orthogonal decomposition," in *Model Order Reduction: Theory, Research Aspects and Applications* (Springer, Berlin, 2008), pp. 95–109.
[38]Y. S. Ong, P. B. Nair, and A. J. Keane, "Evolutionary optimization of computationally expensive problems via surrogate modeling," AIAA J. **41**, 687–696 (2003).
[39]R. W. Freund, "Model reduction methods based on Krylov subspaces," Acta Numer. **12**, 267–319 (2003).
[40]Z. Bai, "Krylov subspace techniques for reduced-order modeling of large-scale dynamical systems," Appl. Numer. Math. **43**, 9–44 (2002).
[41]A. Chatterjee, "An introduction to the proper orthogonal decomposition," Curr. Sci. **78**, 808–817 (2000), see https://www.jstor.org/stable/24103957.
[42]D. Bonvin and D. Mellichamp, "A unified derivation and critical review of modal approaches to model reduction," Int. J. Control **35**, 829–848 (1982).
[43]Z. Elrazaz and N. K. Sinha, "A review of some model reduction techniques," Can. Electr. Eng. J. **6**, 34–40 (1981).
[44]A. Rasheed, O. San, and T. Kvamsdal, "Digital twin: Values, challenges and enablers from a modeling perspective," IEEE Access **8**, 21980–22012 (2020).
[45]E. N. Karatzas, F. Ballarin, and G. Rozza, "Projection-based reduced order models for a cut finite element method in parametrized domains," Comput. Math. Appl. **79**, 833–851 (2020).
[46]J. V. Aguado, D. Borzacchiello, C. Ghnatios, F. Lebel, R. Upadhyay, C. Binetruy, and F. Chinesta, "A simulation app based on reduced order modeling for manufacturing optimization of composite outlet guide vanes," Adv. Model. Simul. Eng. Sci. **4**(1), 1 (2017).
[47]J. V. Aguado, A. Huerta, F. Chinesta, and E. Cueto, "Real-time monitoring of thermal processes by reduced-order modeling," Int. J. Numer. Methods Eng. **102**, 991–1017 (2015).
[48]Z. Zainib, F. Ballarin, S. Fremes, P. Triverio, L. Jiménez-Juan, and G. Rozza, "Reduced order methods for parametric optimal flow control in coronary bypass grafts, towards patient-specific data assimilation," Int. J. Numer. Methods Biomed. Eng. (published online) (2020).
[49]A. Quarteroni and G. Rozza, *Reduced Order Methods for Modeling and Computational Reduction* (Springer, Berlin, 2014), Vol. 9.
[50]Y. Wang, S. W. Cheung, E. T. Chung, Y. Efendiev, and M. Wang, "Deep multiscale model learning," J. Comput. Phys. **406**, 109071–109071 (2020).
[51]R. Swischuk, L. Mainini, B. Peherstorfer, and K. Willcox, "Projection-based model reduction: Formulations for physics-based machine learning," Comput. Fluids **179**, 704–717 (2019).
[52]K. Lee and K. T. Carlberg, "Model reduction of dynamical systems on nonlinear manifolds using deep convolutional autoencoders," J. Comput. Phys. **404**, 108973 (2020).
[53]A. Iollo, S. Lanteri, and J.-A. Désidéri, "Stability properties of POD–Galerkin approximations for the compressible Navier–Stokes equations," Theor. Comput. Fluid Dyn. **13**, 377–396 (2000).
[54]C. W. Rowley, T. Colonius, and R. M. Murray, "Model reduction for compressible flows using POD and Galerkin projection," Physica D **189**, 115–129 (2004).
[55]R. Bourguet, M. Braza, and A. Dervieux, "Reduced-order modeling for unsteady transonic flows around an airfoil," Phys. Fluids **19**, 111701 (2007).
[56]X. Gloerfelt, "Compressible proper orthogonal decomposition/Galerkin reduced-order model of self-sustained oscillations in a cavity," Phys. Fluids **20**, 115105 (2008).
[57]R. Bourguet, M. Braza, A. Sévrain, and A. Bouhadji, "Capturing transition features around a wing by reduced-order modeling based on compressible Navier–Stokes equations," Phys. Fluids **21**, 094104 (2009).
[58]M. F. Barone, I. Kalashnikova, D. J. Segalman, and H. K. Thornquist, "Stable Galerkin reduced order models for linearized compressible flow," J. Comput. Phys. **228**, 1932–1946 (2009).
[59]I. Kalashnikova and M. F. Barone, "On the stability and convergence of a Galerkin reduced order model (ROM) of compressible flow with solid wall and far-field boundary treatment," Int. J. Numer. Methods Eng. **83**, 1345–1375 (2010).
[60]A. Placzek, D.-M. Tran, and R. Ohayon, "A nonlinear POD-Galerkin reduced-order model for compressible flows taking into account rigid body motions," Comput. Methods Appl. Mech. Eng. **200**, 3497–3514 (2011).
[61]M. Balajewicz, I. Tezaur, and E. Dowell, "Minimal subspace rotation on the Stiefel manifold for stabilization and enhancement of projection-based reduced order models for the compressible Navier–Stokes equations," J. Comput. Phys. **321**, 224–241 (2016).
[62]A. Ferrero, A. Iollo, and F. Larocca, "Global and local POD models for the prediction of compressible flows with DG methods," Int. J. Numer. Methods Eng. **116**, 332–357 (2018).
[63]M. Yano, "Discontinuous Galerkin reduced basis empirical quadrature procedure for model reduction of parametrized nonlinear conservation laws," Adv. Comput. Math. **45**, 2287–2320 (2019).
[64]M. Yu, H. Wei-Xi, and X. Chun-Xiao, "Data-driven construction of a reduced-order model for supersonic boundary layer transition," J. Fluid Mech. **874**, 1096–1114 (2019).
[65]E. Rezaian and M. Wei, "A hybrid stabilization approach for reduced-order models of compressible flows with shock-vortex interaction," Int. J. Numer. Methods Eng. **121**, 1629–1646 (2020).

[66]V. Zucatti, W. Wolf, and M. Bergmann, "Calibration of projection-based reduced-order models for unsteady compressible flows," J. Comput. Phys. **433**, 110196 (2021).

[67]E. H. Krath, F. L. Carpenter, P. G. Cizmas, and D. A. Johnston, "An efficient proper orthogonal decomposition based reduced-order model for compressible flows," J. Comput. Phys. **426**, 109959 (2021).

[68]S. Hovland, J. T. Gravdahl, and K. E. Willcox, "Explicit model predictive control for large-scale systems via model reduction," J. Guidance, Control, Dyn. **31**, 918–926 (2008).

[69]F. Ballarin, G. Rozza, and M. Strazzullo, "Reduced order methods for parametric flow control problems and applications," preprint arXiv:2011.12101 (2020).

[70]M. G. Kapteyn, D. J. Knezevic, D. Huynh, M. Tran, and K. E. Willcox, "Data-driven physics-based digital twins via a library of component-based reduced-order models," Int. J. Numer. Methods Eng. (published online) (2020).

[71]M. G. Kapteyn, J. V. Pretorius, and K. E. Willcox, "A probabilistic graphical model foundation for enabling predictive digital twins at scale," Nat. Comput. Sci. **1**, 337–347 (2021).

[72]P. Holmes, J. L. Lumley, and G. Berkooz, *Turbulence, Coherent Structures, Dynamical Systems and Symmetry* (Cambridge University Press, New York, 1998).

[73]J. N. Kutz, S. L. Brunton, B. W. Brunton, and J. L. Proctor, *Dynamic Mode Decomposition: Data-Driven Modeling of Complex Systems* (SIAM, 2016).

[74]K. Pearson, "Principal components analysis," London, Edinburgh, Dublin Philos. Mag. J. Sci. **2**, 559 (1901).

[75]H. Hotelling, "Analysis of a complex of statistical variables into principal components," J. Educ. Psychol. **24**, 417 (1933).

[76]D. Kosambi, *Statistics in Function Space* (Springer, Berlin, 2016), pp. 115–123.

[77]M. Loeve, *Probability Theory: Foundations, Random Sequences* (Van Nostrand Company, New York, 1955).

[78]E. N. Lorenz, *Empirical Orthogonal Functions and Statistical Weather Prediction* (Massachusetts Institute of Technology, Department of Meteorology, Cambridge, 1956).

[79]A. H. Monahan, J. C. Fyfe, M. H. Ambaum, D. B. Stephenson, and G. R. North, "Empirical orthogonal functions: The medium is the message," J. Clim. **22**, 6501–6514 (2009).

[80]J. L. Lumley, *Stochastic Tools in Turbulence* (Dover Publications, New York, 2007).

[81]J. L. Lumley, "Coherent structures in turbulence," in *Transition and Turbulence* (Academic Press, New York, 1981), pp. 215–242.

[82]L. Sirovich, "Turbulence and the dynamics of coherent structures. I-Coherent structures. II-Symmetries and transformations. III-Dynamics and scaling," Q. Appl. Math. **45**, 561–571 (1987).

[83]N. Aubry, P. Holmes, J. L. Lumley, and E. Stone, "The dynamics of coherent structures in the wall region of a turbulent boundary layer," J. Fluid Mech. **192**, 115–173 (1988).

[84]N. Aubry, "On the hidden beauty of the proper orthogonal decomposition," Theor. Comput. Fluid Dyn. **2**, 339–352 (1991).

[85]K. Willcox and J. Peraire, "Balanced model reduction via the proper orthogonal decomposition," AIAA J. **40**, 2323–2330 (2002).

[86]M. Barrault, Y. Maday, N. C. Nguyen, and A. T. Patera, "An 'empirical interpolation' method: Application to efficient reduced-basis discretization of partial differential equations," C. R. Math. **339**, 667–672 (2004).

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

[88]M. A. Grepl, Y. Maday, N. C. Nguyen, and A. T. Patera, "Efficient reduced-basis treatment of nonaffine and nonlinear partial differential equations," ESAIM: Math. Modell. Numer. Anal. **41**, 575–605 (2007).

[89]S. Gugercin, A. C. Antoulas, and C. Beattie, "$H_2$ model reduction for large-scale linear dynamical systems," SIAM J. Matrix Anal. Appl. **30**, 609–638 (2008).

[90]S. Chaturantabut and D. C. Sorensen, "Nonlinear model reduction via discrete empirical interpolation," SIAM J. Sci. Comput. **32**, 2737–2764 (2010).

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

[92]M. O. Williams, I. G. Kevrekidis, and C. W. Rowley, "A data–driven approximation of the Koopman operator: Extending dynamic mode decomposition," J. Nonlinear Sci. **25**, 1307–1346 (2015).

[93]K. Carlberg, C. Bou-Mosleh, and C. Farhat, "Efficient non-linear model reduction via a least-squares Petrov–Galerkin projection and compressive tensor approximations," Int. J. Numer. Methods Eng. **86**, 155–181 (2011).

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

[95]B. G. Galerkin, "Series-solutions of some cases of equilibrium of elastic beams and plates," Vestn. Inshenernov **1**, 897–908 (1915).

[96]B. Saltzman, "Finite amplitude free convection as an initial value problem-I," J. Atmos. Sci. **19**, 329–341 (1962).

[97]S. Lakshmivarahan, J. M. Lewis, and J. Hu, "Saltzman's model. Part I: Complete characterization of solution properties," J. Atmos. Sci. **76**, 1587–1608 (2019).

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

[99]D. Rempfer and H. F. Fasel, "Dynamics of three-dimensional coherent structures in a flat-plate boundary layer," J. Fluid Mech. **275**, 257–283 (1994).

[100]R. Everson and L. Sirovich, "Karhunen–Loeve procedure for gappy data," J. Opt. Soc. Am. A **12**, 1657–1664 (1995).

[101]S. S. Ravindran, "A reduced-order approach for optimal control of fluids using proper orthogonal decomposition," Int. J. Numer. Methods Fluids **34**, 425–448 (2000).

[102]K. Kunisch and S. Volkwein, "Galerkin proper orthogonal decomposition methods for parabolic problems," Numer. Math. **90**, 117–148 (2001).

[103]M. Couplet, P. Sagaut, and C. Basdevant, "Intermodal energy transfers in a proper orthogonal decomposition-Galerkin representation of a turbulent separated flow," J. Fluid Mech. **491**, 275 (2003).

[104]S. Sirisup and G. E. Karniadakis, "A spectral viscosity method for correcting the long-term behavior of POD models," J. Comput. Phys. **194**, 92–116 (2004).

[105]G. Rozza, D. B. P. Huynh, and A. T. Patera, "Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations," Arch. Comput. Methods Eng. **15**, 229 (2007).

[106]Y. Cao, J. Zhu, I. M. Navon, and Z. Luo, "A reduced-order approach to four-dimensional variational data assimilation using proper orthogonal decomposition," Int. J. Numer. Methods Fluids **53**, 1571–1583 (2007).

[107]D. Amsallem and C. Farhat, "Interpolation method for adapting reduced-order models and application to aeroelasticity," AIAA J. **46**, 1803–1813 (2008).

[108]P. Astrid, S. Weiland, K. Willcox, and T. Backx, "Missing point estimation in models described by proper orthogonal decomposition," IEEE Trans. Automat. Control **53**, 2237–2251 (2008).

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

[110]T. P. Sapsis and P. F. Lermusiaux, "Dynamically orthogonal field equations for continuous stochastic dynamical systems," Physica D **238**, 2347–2360 (2009).

[111]K. Carlberg, C. Farhat, J. Cortial, and D. Amsallem, "The GNAT method for nonlinear model reduction: Effective implementation and application to computational fluid dynamics and turbulent flows," J. Comput. Phys. **242**, 623–647 (2013).

[112]L. Cordier, B. R. Noack, G. Tissot, G. Lehnasch, J. Delville, M. Balajewicz, G. Daviller, and R. K. Niven, "Identification strategies for model-based control," Exp. Fluids **54**, 1580 (2013).

[113]J. Östh, B. R. Noack, S. Krajnović, D. Barros, and J. Borée, "On the need for a nonlinear subscale turbulence term in POD models as exemplified for a high-Reynolds-number flow over an Ahmed body," J. Fluid Mech. **747**, 518–544 (2014).

[114]F. Ballarin, A. Manzoni, A. Quarteroni, and G. Rozza, "Supremizer stabilization of POD–Galerkin approximation of parametrized steady incompressible Navier–Stokes equations," Int. J. Numer. Methods Eng. **102**, 1136–1161 (2015).

[115]M. Schlegel and B. R. Noack, "On long-term boundedness of Galerkin models," J. Fluid Mech. **765**, 325–352 (2015).

[116]B. Peherstorfer and K. Willcox, "Data-driven operator inference for nonintrusive projection-based model reduction," Comput. Methods Appl. Mech. Eng. **306**, 196–215 (2016).
[117]S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Discovering governing equations from data by sparse identification of nonlinear dynamical systems," Proc. Natl. Acad. Sci. U. S. A. **113**, 3932–3937 (2016).
[118]M. Sieber, C. O. Paschereit, and K. Oberleithner, "Spectral proper orthogonal decomposition," J. Fluid Mech. **792**, 798–828 (2016).
[119]A. Towne, O. T. Schmidt, and T. Colonius, "Spectral proper orthogonal decomposition and its relationship to dynamic mode decomposition and resolvent analysis," J. Fluid Mech. **847**, 821–867 (2018).
[120]J. Reiss, P. Schulze, J. Sesterhenn, and V. Mehrmann, "The shifted proper orthogonal decomposition: A mode decomposition for multiple transport phenomena," SIAM J. Sci. Comput. **40**, A1322–A1344 (2018).
[121]J.-C. Loiseau, B. R. Noack, and S. L. Brunton, "Sparse reduced-order modelling: Sensor-based dynamics to full-state estimation," J. Fluid Mech. **844**, 459–490 (2018).
[122]M. Mendez, M. Balabane, and J.-M. Buchlin, "Multi-scale proper orthogonal decomposition of complex fluid flows," J. Fluid Mech. **870**, 988–1036 (2019).
[123]H. Li, D. Fernex, R. Semaan, J. Tan, M. Morzyński, and B. R. Noack, "Cluster-based network model," J. Fluid Mech. **906**, A21 (2021).
[124]D. Fernex, B. R. Noack, and R. Semaan, "Cluster-based network modeling-From snapshots to complex dynamical systems," Sci. Adv. **7**, eabf5006 (2021).
[125]V. M. Krasnopolsky and M. S. Fox-Rabinovitz, "A new synergetic paradigm in environmental numerical modeling: Hybrid models combining deterministic and machine learning components," Ecol. Modell. **191**, 5–18 (2006).
[126]V. M. Krasnopolsky and M. S. Fox-Rabinovitz, "Complex hybrid models combining deterministic and machine learning components for numerical climate modeling and weather prediction," Neural Networks **19**, 122–134 (2006).
[127]M. Reichstein, G. Camps-Valls, B. Stevens, M. Jung, J. Denzler, N. Carvalhais *et al.*, "Deep learning and process understanding for data-driven Earth system science," Nature **566**, 195–204 (2019).
[128]E. de Bezenac, A. Pajot, and P. Gallinari, "Deep learning for physical processes: Incorporating prior scientific knowledge," J. Stat. Mech. **2019**, 124009.
[129]A. Karpatne, G. Atluri, J. H. Faghmous, M. Steinbach, A. Banerjee, A. Ganguly, S. Shekhar, N. Samatova, and V. Kumar, "Theory-guided data science: A new paradigm for scientific discovery from data," IEEE Trans. Knowl. Data Eng. **29**, 2318–2331 (2017).
[130]C. M. Childs and N. R. Washburn, "Embedding domain knowledge for machine learning of complex material systems," MRS Commun. **9**, 806–820 (2019).
[131]J. Willard, X. Jia, S. Xu, M. Steinbach, and V. Kumar, "Integrating physics-based modeling with machine learning: A survey," preprint arXiv:2003.04919 (2020).
[132]V. L. Kalb and A. E. Deane, "An intrinsic stabilization scheme for proper orthogonal decomposition based low-dimensional models," Phys. Fluids **19**, 054106 (2007).
[133]D. Amsallem and C. Farhat, "Stabilization of projection-based reduced-order models," Int. J. Numer. Methods Eng. **91**, 358–377 (2012).
[134]M. Balajewicz and E. H. Dowell, "Stabilization of projection-based reduced order models of the Navier–Stokes," Nonlinear Dyn. **70**, 1619–1632 (2012).
[135]Z. Wang, I. Akhtar, J. Borggaard, and T. Iliescu, "Proper orthogonal decomposition closure models for turbulent flows: A numerical comparison," Comput. Meth. Appl. Mech. Eng. **237–240**, 10–26 (2012).
[136]B. Protas, B. R. Noack, and J. Östh, "Optimal nonlinear eddy viscosity in Galerkin models of turbulent flows," J. Fluid Mech. **766**, 337–367 (2015).
[137]M. Benosman, J. Borggaard, O. San, and B. Kramer, "Learning-based robust stabilization for reduced-order models of 2D and 3D Boussinesq equations," Appl. Math. Modell. **49**, 162–181 (2017).
[138]G. Stabile, F. Ballarin, G. Zuccarino, and G. Rozza, "A reduced order variational multiscale approach for turbulent flows," Adv. Comput. Math. **4**(5), 2349–2368 (2019).
[139]R. Reyes and R. Codina, "Projection-based reduced order models for flow problems: A variational multiscale approach," Comput. Methods Appl. Mech. Eng. **363**, 112844 (2020).
[140]M. Cheng, T. Y. Hou, and Z. Zhang, "A dynamically bi-orthogonal method for time-dependent stochastic partial differential equations I: Derivation and algorithms," J. Comput. Phys. **242**, 843–868 (2013).
[141]D. Ramezanian, A. G. Nouri, and H. Babaee, "On-the-fly reduced order modeling of passive and reactive species via time-dependent manifolds," Comput. Methods Appl. Mech. Eng. **382**, 113882 (2021).
[142]P. Patil and H. Babaee, "Real-time reduced-order modeling of stochastic partial differential equations via time-dependent subspaces," J. Comput. Phys. **415**, 109511 (2020).
[143]R. Haberman, *Applied Partial Differential Equations with Fourier Series and Boundary Value Problems* (Pearson, New York, 2012).
[144]O. M. Faltinsen and A. N. Timokha, *Sloshing* (Cambridge University Press, Cambridge, 2009), Vol. 577.
[145]I. Lukovsky and A. Timokha, "Multimodal method in sloshing," J. Math. Sci. **220**, 239–253 (2017).
[146]I. Gavrilyuk, I. Lukovsky, and A. Timokha, "A multimodal approach to nonlinear sloshing in a circular cylindrical tank," Hybrid Methods Eng. **2**, 22 (2000).
[147]O. M. Faltinsen and A. N. Timokha, "An adaptive multimodal approach to nonlinear sloshing in a rectangular tank," J. Fluid Mech. **432**, 167–200 (2001).
[148]O. M. Faltinsen, O. F. Rognebakke, I. A. Lukovsky, and A. N. Timokha, "Multidimensional modal analysis of nonlinear sloshing in a rectangular tank with finite water depth," J. Fluid Mech. **407**, 201–234 (2000).
[149]O. M. Faltinsen and A. N. Timokha, "Asymptotic modal approximation of nonlinear resonant sloshing in a rectangular tank with small fluid depth," J. Fluid Mech. **470**, 319–357 (2002).
[150]O. M. Faltinsen and A. N. Timokha, "A multimodal method for liquid sloshing in a two-dimensional circular tank," J. Fluid Mech. **665**, 457–479 (2010).
[151]I. Lukovsky, D. Ovchynnykov, and A. Timokha, "Asymptotic nonlinear multimodal modeling of liquid sloshing in an upright circular cylindrical tank. I. Modal equations," Nonlinear Oscillations **14**, 512–525 (2012).
[152]M. Ansari, R. Firouz-Abadi, and M. Ghasemi, "Two phase modal analysis of nonlinear sloshing in a rectangular container," Ocean Eng. **38**, 1277–1282 (2011).
[153]J. Gómez-Goñi, C. A. Garrido-Mendoza, J. L. Cercós, and L. González, "Two phase analysis of sloshing in a rectangular container with volume of fluid (VOF) methods," Ocean Eng. **73**, 208–212 (2013).
[154]A. Obukhov, "Statistically homogeneous fields on a sphere," Usp. Mat. Nauk **2**, 196–198 (1947).
[155]S. Giere, T. Iliescu, V. John, and D. Wells, "SUPG reduced order models for convection-dominated convection-diffusion-reaction equations," Comput. Methods Appl. Mech. Eng. **289**, 454–474 (2015).
[156]M. D. Gunzburger, *Perspectives in Flow Control and Optimization* (SIAM, Philadelphia, 2002).
[157]S. E. Ahmed and O. San, "Breaking the Kolmogorov barrier in model reduction of fluid flows," Fluids **5**, 26 (2020).
[158]W. Cazemier, R. Verstappen, and A. Veldman, "Proper orthogonal decomposition and low-dimensional models for driven cavity flows," Phys. Fluids **10**, 1685 (1998).
[159]B. Koc, S. Rubino, M. Schneier, J. R. Singler, and T. Iliescu, "On optimal pointwise in time error bounds and difference quotients for the proper orthogonal decomposition," SIAM J. Numer. Anal. **59**, 2163–2196 (2021).
[160]T. Cui, Y. M. Marzouk, and K. E. Willcox, "Data-driven model reduction for the Bayesian solution of inverse problems," Int. J. Numer. Methods Eng. **102**, 966–990 (2015).
[161]A. Mendible, S. L. Brunton, A. Y. Aravkin, W. Lowrie, and J. N. Kutz, "Dimensionality reduction and reduced-order modeling for traveling wave physics," Theor. Comput. Fluid Dyn. **34**, 385–400 (2020).
[162]A. Mendible, J. Koch, H. Lange, S. L. Brunton, and J. N. Kutz, "Data-driven modeling of rotating detonation waves," Phys. Rev. Fluids **6**, 050507 (2021).
[163]P. A. Etter and K. T. Carlberg, "Online adaptive basis refinement and compression for reduced-order models via vector-space sieving," Comput. Methods Appl. Mech. Eng. **364**, 112931 (2020).
[164]D. Amsallem, M. J. Zahr, and C. Farhat, "Nonlinear model order reduction based on local reduced-order bases," Int. J. Numer. Methods Eng. **92**, 891–916 (2012).
[165]K. Washabaugh, D. Amsallem, M. Zahr, and C. Farhat, "Nonlinear model reduction for CFD problems using local reduced-order bases," in *42nd AIAA Fluid Dynamics Conference and Exhibit* (AIAA, 2012), p. 2686.

[166]B. Peherstorfer, D. Butnaru, K. Willcox, and H.-J. Bungartz, "Localized discrete empirical interpolation method," SIAM J. Sci. Comput. **36**, A168–A192 (2014).
[167]T. Taddei, S. Perotto, and A. Quarteroni, "Reduced basis techniques for nonlinear conservation laws," ESAIM **49**, 787–814 (2015).
[168]W. IJzerman, "Signal representation and modeling of spatial structures in fluids," Ph.D. thesis (University of Twente, 2000).
[169]M. Dihlmann, M. Drohmann, and B. Haasdonk, "Model reduction of parametrized evolution problems using the reduced basis method with adaptive time-partitioning," in *Proceedings of the International Conference on Adaptive Modeling and Simulation* (ADMOS, 2011), p. 64.
[170]J. Borggaard, A. Hay, and D. Pelletier, "Interval-based reduced order models for unsteady fluid flow," Int. J. Numer. Anal. Model. **4**, 353–367 (2007).
[171]O. San and J. Borggaard, "Principal interval decomposition framework for POD reduced-order modeling of convective Boussinesq flows," Int. J. Numer. Methods Fluids **78**, 37–62 (2015).
[172]H. Babaee and T. P. Sapsis, "A variational principle for the description of time-dependent modes associated with transient instabilities," Philos. Trans. R. Soc. London **472**, 20150779 (2016).
[173]S. Chaturantabut, "Temporal localized nonlinear model reduction with a priori error estimate," Appl. Numer. Math. **119**, 225–238 (2017).
[174]M. Ahmed and O. San, "Stabilized principal interval decomposition method for model reduction of nonlinear convective systems with moving shocks," Comput. Appl. Math. **37**, 6870–6902 (2018).
[175]S. E. Ahmed, S. M. Rahman, O. San, A. Rasheed, and I. M. Navon, "Memory embedded non-intrusive reduced order modeling of non-ergodic flows," Phys. Fluids **31**, 126602 (2019).
[176]Y. Maday and E. M. Rønquist, "A reduced-basis element method," J. Sci. Comput. **17**, 447–459 (2002).
[177]A. E. Løvgren, Y. Maday, and E. M. Rønquist, "A reduced basis element method for the steady Stokes problem," ESAIM **40**, 529–552 (2006).
[178]L. Iapichino, A. Quarteroni, and G. Rozza, "A reduced basis hybrid method for the coupling of parametrized domains represented by fluidic networks," Comput. Methods Appl. Mech. Eng. **221–222**, 63–82 (2012).
[179]J. L. Eftang and A. T. Patera, "Port reduction in parametrized component static condensation: Approximation and *a posteriori* error estimation," Int. J. Numer. Methods Eng. **96**, 269–302 (2013).
[180]J. L. Eftang and B. Stamm, "Parameter multi-domain '*hp*' empirical interpolation," Int. J. Numer. Methods Eng. **90**, 412–428 (2012).
[181]A. Moosavi, R. Stefanescu, and A. Sandu, "Efficient construction of local parametric reduced order models using machine learning techniques," preprint arXiv:1511.02909 (2015).
[182]M. Hess, A. Alla, A. Quaini, G. Rozza, and M. Gunzburger, "A localized reduced-order modeling approach for PDEs with bifurcating solutions," Comput. Methods Appl. Mech. Eng. **351**, 379–403 (2019).
[183]B. R. Noack, "From snapshots to modal expansions–bridging low residuals and pure frequencies," J. Fluid Mech. **802**, 1–4 (2016).
[184]J. Burkardt, M. Gunzburger, and H.-C. Lee, "Centroidal voronoi tessellation-based reduced-order modeling of complex systems," SIAM J. Sci. Comput. **28**, 459–484 (2006).
[185]E. Kaiser, B. R. Noack, L. Cordier, A. Spohn, M. Segond, M. Abel, G. Daviller, J. Östh, S. Krajnović, and R. K. Niven, "Cluster-based reduced-order modelling of a mixing layer," J. Fluid Mech. **754**, 365–414 (2014).
[186]E. Kaiser, B. R. Noack, A. Spohn, L. N. Cattafesta, and M. Morzyński, "Cluster-based control of a separating flow over a smoothly contoured ramp," Theor. Comput. Fluid Dyn. **31**, 579–593 (2017).
[187]H. Li and J. Tan, "Cluster-based Markov model to understand the transition dynamics of a supersonic mixing layer," Phys. Fluids **32**, 56104 (2020).
[188]D. Fernex, R. Semaan, M. Albers, P. S. Meysonnat, W. Schröder, R. Ishar, E. Kaiser, and B. R. Noack, "Cluster-based network model for drag reduction mechanisms of an actuated turbulent boundary layer," Proc. Appl. Math. Mech. **19**, e201900219 (2019).
[189]I. T. Jolliffe, *Principal Component Analysis* (Springer, Berlin, 2002).
[190]M. A. Kramer, "Nonlinear principal component analysis using autoassociative neural networks," AIChE J. **37**, 233–243 (1991).
[191]W. W. Hsieh, "Nonlinear principal component analysis by neural networks," Tellus A **53**, 599–615 (2001).
[192]W. W. Hsieh, *Machine Learning Methods in the Environmental Sciences: Neural Networks and Kernels* (Cambridge University Press, New York, 2009).
[193]B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," Neural Comput. **10**, 1299–1319 (1998).
[194]S. E. Otto and C. W. Rowley, "Linearly recurrent autoencoder networks for learning dynamics," SIAM J. Appl. Dyn. Syst. **18**, 558–593 (2019).
[195]K. Fukami, T. Nakamura, and K. Fukagata, "Convolutional neural network based hierarchical autoencoder for nonlinear mode decomposition of fluid field data," Phys. Fluids **32**, 095110 (2020).
[196]L. Agostini, "Exploration and prediction of fluid dynamical systems using auto-encoder technology," Phys. Fluids **32**, 067103 (2020).
[197]T. Hastie and W. Stuetzle, "Principal curves," J. Am. Stat. Assoc. **84**, 502–516 (1989).
[198]S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," Science **290**, 2323–2326 (2000).
[199]J. B. Tenenbaum, V. De Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," Science **290**, 2319–2323 (2000).
[200]T. Kohonen, "Self-organized formation of topologically correct feature maps," Biol. Cybern. **43**, 59–69 (1982).
[201]H. Edelsbrunner, D. Letscher, and A. Zomorodian, "Topological persistence and simplification," in *Proceedings 41st Annual Symposium on Foundations of Computer Science* (IEEE, 2000), pp. 454–463.
[202]H. Edelsbrunner, J. Harer *et al.*, "Persistent homology-a survey," Contemp. Math. **453**, 257–282 (2008).
[203]H. Edelsbrunner and D. Morozov, "Persistent homology: Theory and practice," in *European Congress of Mathematics, Kraków, 2–7 July, 2012* (European Mathematical Society, 2014), pp. 31–50.
[204]A. Zomorodian and G. Carlsson, "Localized homology," Comput. Geom. **41**, 126–148 (2008).
[205]A. Zomorodian and G. Carlsson, "Computing persistent homology," Discrete Comput. Geom. **33**, 249–274 (2005).
[206]N. Otter, M. A. Porter, U. Tillmann, P. Grindrod, and H. A. Harrington, "A roadmap for the computation of persistent homology," EPJ Data Sci. **6**, 17–38 (2017).
[207]B. Wang and G.-W. Wei, "Object-oriented persistent homology," J. Comput. Phys. **305**, 276–299 (2016).
[208]C. M. Pereira and R. F. de Mello, "Persistent homology for time series and spatial data clustering," Expert Syst. Appl. **42**, 6026–6038 (2015).
[209]J. Garland, E. Bradley, and J. D. Meiss, "Exploring the topology of dynamical reconstructions," Physica D **334**, 49–59 (2016).
[210]S. Maletić, Y. Zhao, and M. Rajković, "Persistent topological features of dynamical systems," Chaos **26**, 053105 (2016).
[211]B. Rieck and H. Leitte, "Persistent homology for the evaluation of dimensionality reduction schemes," Comput. Graph. Forum **34**, 431–440 (2015).
[212]A. Moitra, N. O. Malott, and P. A. Wilsey, "Cluster-based data reduction for persistent homology," in *IEEE International Conference on Big Data (Big Data)* (IEEE, Seattle, WA, 2018), pp. 327–334.
[213]S. Lakshmivarahan and Y. Wang, "On the relation between energy-conserving low-order models and a system of coupled generalized Volterra gyrostats with nonlinear feedback," J. Nonlinear Sci. **18**, 75–97 (2008).
[214]S. Lakshmivarahan and Y. Wang, "On the structure of the energy conserving low-order models and their relation to Volterra gyrostat," Nonlinear Anal. **9**, 1573–1589 (2008).
[215]Y. Wang and S. Lakshmivarahan, "On the relation between energy conserving low-order models and Hamiltonian systems," Nonlinear Anal. **71**, e351–e358 (2009).
[216]R. Ştefănescu, A. Sandu, and I. M. Navon, "Comparison of POD reduced order strategies for the nonlinear 2D shallow water equations," Int. J. Numer. Methods Fluids **76**, 497–521 (2014).
[217]G. Dimitriu, R. Ştefănescu, and I. M. Navon, "Comparative numerical analysis using reduced-order modeling strategies for nonlinear large-scale systems," J. Comput. Appl. Math. **310**, 32–43 (2017).
[218]S. Chaturantabut and D. C. Sorensen, "Discrete empirical interpolation for nonlinear model reduction," in *Proceedings of the 48th IEEE Conference on Decision and Control* (IEEE, Shanghai, China, 2009), pp. 4316–4321.
[219]T. Bui-Thanh, M. Damodaran, and K. Willcox, "Aerodynamic data reconstruction and inverse design using proper orthogonal decomposition," AIAA J. **42**, 1505–1516 (2004).

[220]R. Zimmermann and K. Willcox, "An accelerated greedy missing point estimation procedure," SIAM J. Sci. Comput. **38**, A2827–A2850 (2016).
[221]C. Gu, "QLMOR: A projection-based nonlinear model order reduction approach using quadratic-linear representation of nonlinear systems," IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst. **30**, 1307–1320 (2011).
[222]P. Benner and T. Breiten, "Two-sided projection methods for nonlinear model order reduction," SIAM J. Sci. Comput. **37**, B239–B260 (2015).
[223]P. Benner, P. Goyal, and S. Gugercin, "$H_2$-quasi-optimal model order reduction for quadratic-bilinear control systems," SIAM J. Matrix Anal. Appl. **39**, 983–1032 (2018).
[224]B. Kramer and K. E. Willcox, "Nonlinear model order reduction via lifting transformations and proper orthogonal decomposition," AIAA J. **57**, 2297–2307 (2019).
[225]K. Willcox, "Unsteady flow sensing and estimation via the gappy proper orthogonal decomposition," Comput. Fluids **35**, 208–226 (2006).
[226]B. Yildirim, C. Chryssostomidis, and G. E. Karniadakis, "Efficient sensor placement for ocean measurements using low-dimensional concepts," Ocean Modell. **27**, 160–173 (2009).
[227]Z. Drmac and S. Gugercin, "A new selection operator for the discrete empirical interpolation method—Improved a priori error bound and extensions," SIAM J. Sci. Comput. **38**, A631–A648 (2016).
[228]B. Karasözen, S. Yildiz, and M. Uzunca, "Structure preserving model order reduction of shallow water equations," Math. Methods Appl. Sci. **44**, 476–492 (2021).
[229]S. M. Rahman, S. E. Ahmed, and O. San, "A dynamic closure modeling framework for model order reduction of geophysical flows," Phys. Fluids **31**, 046602 (2019).
[230]F. Ballarin, E. Faggiano, S. Ippolito, A. Manzoni, A. Quarteroni, G. Rozza, and R. Scrofani, "Fast simulations of patient-specific haemodynamics of coronary artery bypass grafts based on a POD–Galerkin method and a vascular shape parametrization," J. Comput. Phys. **315**, 609–628 (2016).
[231]C. VerHulst and C. Meneveau, "Large eddy simulation study of the kinetic energy entrainment by energetic turbulent flow structures in large wind farms," Phys. Fluids **26**, 025113 (2014).
[232]S. Shah and E. Bou-Zeid, "Very-large-scale motions in the atmospheric boundary layer educed by snapshot proper orthogonal decomposition," Boundary-Layer Meteorol. **153**, 355–387 (2014).
[233]M. Zhang and R. J. A. M. Stevens, "Characterizing the coherent structures within and above large wind farms," Boundary-Layer Meteorol. **174**, 61–80 (2020).
[234]P. Sagaut, "Large eddy simulation for incompressible flow," in *Scientific Computation*, 3rd ed. (Springer-Verlag, Berlin, 2006), p. xxx+556.
[235]S. Pope, *Turbulent Flows* (Cambridge University Press, Cambridge, 2000), p. xxxiv+771.
[236]A. Kolmogoroff, "Über die beste Annäherung von Funktionen einer gegebenen Funktionenklasse," Ann. Math. **37**, 107–110 (1936).
[237]A. Pinkus, *N-Widths in Approximation Theory* (Springer-Verlag, Berlin, 1985), Vol. 7.
[238]S. E. Ahmed, O. San, A. Rasheed, and T. Iliescu, "A long short-term memory embedding for hybrid uplifted reduced order models," Physica D **409**, 132471 (2020).
[239]I. Akhtar, Z. Wang, J. Borggaard, and T. Iliescu, "A new closure strategy for proper orthogonal decomposition reduced-order models," J. Comput. Nonlinear Dyn. **7**, 034503 (2012).
[240]O. San and J. Borggaard, "Basis selection and closure for POD models of convection dominated Boussinesq flows," in 21st International Symposium on Mathematical Theory of Networks and Systems (2014), Vol. 5.
[241]L. N. Azadani and A. E. Staples, "Large-eddy simulation of turbulent barotropic flows in spectral space on a sphere," J. Atmos. Sci. **72**, 1727–1742 (2015).
[242]A. N. Kolmogorov, "The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers," Cr Acad. Sci. URSS **30**, 301–305 (1941).
[243]A. N. Kolmogorov, "The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers," Proc. R. Soc. London, Ser. A **434**, 9–13 (1991).
[244]B. R. Noack, K. Afanasiev, M. Moryński, G. Tadmor, and F. Thiele, "A hierarchy of low-dimensional models for the transient and post-transient cylinder wake," J. Fluid Mech. **497**, 335–363 (2003).
[245]S. Grimberg, C. Farhat, and N. Youkilis, "On the stability of projection-based model order reduction for convection-dominated laminar and turbulent flows," J. Comput. Phys. **419**, 109681 (2020).
[246]E. J. Parish, C. R. Wentland, and K. Duraisamy, "The Adjoint Petrov–Galerkin method for non-linear model reduction," Comput. Methods Appl. Mech. Eng. **365**, 112991 (2020).
[247]C. Hoang, Y. Choi, and K. Carlberg, "Domain-decomposition least-squares Petrov-Galerkin (DD-LSPG) nonlinear model reduction," Comput. Methods Appl. Mech. Eng. **384**, 113997 (2021).
[248]D. Xiao, F. Fang, J. Du, C. C. Pain, I. M. Navon, A. G. Buchan, A. H. Elsheikh, and G. Hu, "Non-linear Petrov–Galerkin methods for reduced order modelling of the Navier–Stokes equations using a mixed finite element pair," Comput. Methods Appl. Mech. Eng. **255**, 147–157 (2013).
[249]S. T. Johnston, M. J. Simpson, and R. E. Baker, "Mean-field descriptions of collective migration with strong adhesion," Phys. Rev. E **85**, 051922 (2012).
[250]C. Brennan and D. Venturi, "Data-driven closures for stochastic dynamical systems," J. Comput. Phys. **372**, 281–298 (2018).
[251]C. D. Levermore, "Moment closure hierarchies for kinetic theories," J. Stat. Phys. **83**, 1021–1065 (1996).
[252]F. M. Selten, "An efficient description of the dynamics of barotropic flow," J. Atmos. Sci. **52**, 915–936 (1995).
[253]F. M. Selten, "A statistical closure of a low-order barotropic model," J. Atmos. Sci. **54**, 1085–1093 (1997).
[254]B. R. Noack, M. Schlegel, B. Ahlborn, G. Mutschke, M. Morzynski, P. Comte, and G. Tadmor, "A finite-time thermodynamics of unsteady fluid flows," J. Non-Equilib. Thermodyn. **33**, 103–148 (2008).
[255]M. Bergmann, C.-H. Bruneau, and A. Iollo, "Improvement of reduced order modeling based on POD," in *Computational Fluid Dynamics* (Springer, Berlin, 2009), pp. 779–784.
[256]M. Bergmann, C.-H. Bruneau, and A. Iollo, "Enablers for robust POD models," J. Comput. Phys. **228**, 516–538 (2009).
[257]J. Borggaard, T. Iliescu, and Z. Wang, "Artificial viscosity proper orthogonal decomposition," Math. Comput. Modell. **53**, 269–279 (2011).
[258]Z. Wang, I. Akhtar, J. Borggaard, and T. Iliescu, "Two-level discretizations of nonlinear closure models for proper orthogonal decomposition," J. Comput. Phys. **230**, 126–146 (2011).
[259]M. J. Balajewicz, E. H. Dowell, and B. R. Noack, "Low-dimensional modelling of high-Reynolds-number shear flows incorporating constraints from the Navier-Stokes equation," J. Fluid Mech. **729**, 285–308 (2013).
[260]T. Iliescu and Z. Wang, "Variational multiscale proper orthogonal decomposition: Navier-Stokes equations," Numer. Methods Partial Differ. Equations **30**, 641–663 (2014).
[261]O. San and T. Iliescu, "Proper orthogonal decomposition closure models for fluid flows: Burgers equation," Int. J. Numer. Anal. Mod., Ser. B **5**, 285–305 (2014).
[262]P. Stinis, "Renormalized Mori–Zwanzig-reduced models for systems without scale separation," Proc. R. Soc. A **471**, 20140446 (2015).
[263]A. J. Chorin and F. Lu, "Discrete approach to stochastic parametrization and dimension reduction in nonlinear dynamics," Proc. Nat. Acad. Sci. U. S. A. **112**, 9804–9809 (2015).
[264]Z. Li, X. Bian, X. Li, and G. E. Karniadakis, "Incorporation of memory effects in coarse-grained modeling via the Mori-Zwanzig formalism," J. Chem. Phys. **143**, 243128 (2015).
[265]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**, 20170385 (2017).
[266]T. C. Rebollo, E. D. Ávila, M. G. Mármol, F. Ballarin, and G. Rozza, "On a certified Smagorinsky reduced basis turbulence model," SIAM J. Numer. Anal. **55**, 3047–3067 (2017).
[267]X. Xie, D. Wells, Z. Wang, and T. Iliescu, "Approximate deconvolution reduced order modeling," Comput. Methods Appl. Mech. Eng. **313**, 512–534 (2017).
[268]O. San and R. Maulik, "Extreme learning machine for reduced order modeling of turbulent geophysical flows," Phys. Rev. E **97**, 042322 (2018).

[269]O. San and R. Maulik, "Neural network closures for nonlinear model order reduction," Adv. Comput. Math. **44**, 1717–1750 (2018).
[270]S. Pan and K. Duraisamy, "Data-driven discovery of closure models," SIAM J. Appl. Dyn. Syst. **17**, 2381–2413 (2018).
[271]H. Imtiaz and I. Akhtar, "Nonlinear closure modeling in reduced order models for turbulent flows: A dynamical system approach," Nonlinear Dyn. **99**, 479–494 (2020).
[272]X. Xie, C. Webster, and T. Iliescu, "Closure learning for nonlinear model reduction using deep residual neural network," Fluids **5**, 39 (2020).
[273]Q. Wang, N. Ripamonti, and J. S. Hesthaven, "Recurrent neural network closure of parametric POD-Galerkin reduced-order models based on the Mori-Zwanzig formalism," J. Comput. Phys. **410**, 109402 (2020).
[274]C. Mou, Z. Wang, D. R. Wells, X. Xie, and T. Iliescu, "Reduced order models for the quasi-geostrophic equations: A brief survey," Fluids **6**, 16 (2020).
[275]A. Gupta and P. F. Lermusiaux, "Neural closure models for dynamical systems," preprint arXiv:2012.13869 (2020).
[276]B. Podvin, "On the adequacy of the ten-dimensional model for the wall layer," Phys. Fluids **13**, 210–224 (2001).
[277]B. Podvin and J. Lumley, "A low-dimensional approach for the minimal flow unit," J. Fluid Mech. **362**, 121–155 (1998).
[278]B. Podvin, "A proper-orthogonal-decomposition–based model for the wall layer of a turbulent channel flow," Phys. Fluids **21**, 015111 (2009).
[279]C. Mou, E. Merzari, O. San, and T. Iliescu, "A numerical investigation of the lengthscale in the mixing-length reduced order model of the turbulent channel flow," preprint arXiv:2108.02254 (2021).
[280]J. Smagorinsky, "General circulation experiments with the primitive equations: I. the basic experiment," Mon. Weather Rev. **91**, 99–164 (1963).
[281]B. R. Noack, P. Papas, and P. A. Monkewitz, "Low-dimensional Galerkin model of a laminar shear-layer," Technical Report No. 2002-01 (École Polytechnique Fédérale de Lausanne,2002).
[282]J. Borggaard, A. Duggleby, A. Hay, T. Iliescu, and Z. Wang, "Reduced-order modeling of turbulent flows," in Proceedings of the MTNS (2008).
[283]S. Ullmann and J. Lang, "A POD-Galerkin reduced model with updated coefficients for smagorinsky LES," in Proceedings of V European Conference on Computational Fluid Dynamics, ECCOMAS (CFD 2010), Lisbon, Portugal, edited by J. C. F. Pereira and A. Sequeira (2010).
[284]M. Germano, U. Piomelli, P. Moin, and W. Cabot, "A dynamic subgrid-scale eddy viscosity model," Phys. Fluids A **3**, 1760–1765 (1991).
[285]D. Rempfer, "Koharente strukturen und chaos beim laminar-turbulenten grenzschichtumschlag," Ph.D. thesis (University Stuttgart, 1991).
[286]D. Rempfer and H. Fasel, "The dynamics of coherent structures in a flat-plate boundary layer," in *Advances in Turbulence IV* (Springer, Berlin, 1993), pp. 73–77.
[287]T. J. R. Hughes, G. R. Feijóo, L. Mazzei, and J.-B. Quincy, "The variational multiscale method—A paradigm for computational mechanics," Comput. Methods Appl. Mech. Eng. **166**, 3–24 (1998).
[288]T. J. Hughes, L. Mazzei, and K. E. Jansen, "Large eddy simulation and the variational multiscale method," Comput. Visualization Sci. **3**, 47–59 (2000).
[289]T. J. Hughes, A. A. Oberai, and L. Mazzei, "Large eddy simulation of turbulent channel flows by the variational multiscale method," Phys. Fluids **13**, 1784–1799 (2001).
[290]T. Iliescu and Z. Wang, "Variational multiscale proper orthogonal decomposition: Convection-dominated convection-diffusion-reaction equations," Math. Comput. **82**, 1357–1378 (2013).
[291]J. P. Roop, "A proper-orthogonal decomposition variational multiscale approximation method for a generalized Oseen problem," Adv. Numer. Anal. **2013**, 974284 (2013).
[292]F. G. Eroglu, S. Kaya, and L. G. Rebholz, "A modular regularized variational multiscale proper orthogonal decomposition for incompressible flows," Comput. Methods Appl. Mech. Eng. **325**, 350–368 (2017).
[293]F. G. Eroglu, S. Kaya, and L. G. Rebholz, "Decoupled modular regularized VMS-POD for Darcy-Brinkman equations," IAENG Int. J. Appl. Math. **49**(2), 1–11 (2019).
[294]R. Reyes, "Stabilized reduced order models for low speed flows," Ph.D. thesis (Universitat Politècnica de Catalunya, 2020).
[295]R. Reyes, R. Codina, J. Baiges, and S. Idelsohn, "Reduced order models for thermally coupled low mach flows," Adv. Model. Simul. Eng. Sci. **5**, 28 (2018).
[296]A. Tello, R. Codina, and J. Baiges, "Fluid structure interaction by means of variational multiscale reduced order models," Int. J. Numer. Methods Eng. **121**, 2601–2625 (2020).
[297]S. Rubino, "Numerical analysis of a projection-based stabilized POD-ROM for incompressible flows," SIAM J. Numer. Anal. **58**, 2019–2058 (2020).
[298]M. Azaïez, T. C. Rebollo, and S. Rubino, "A cure for instabilities due to advection-dominance in POD solution to advection-diffusion-reaction equations," J. Comput. Phys. **425**, 109916 (2021).
[299]B. R. Noack, M. Morzynski, and G. Tadmor, *Reduced-Order Modelling for Flow Control* (Springer-Verlag, Berlin, 2011), Vol. 528.
[300]M. Schlegel, B. R. Noack, P. Comte, D. Kolomenskiy, K. Schneider, M. Farge, D. M. Luchtenburg, J. E. Scouten, and G. Tadmor, "Reduced-order modelling of turbulent jets for noise control," in *Numerical Simulation of Turbulent Flows and Noise Generation* (Springer, Berlin, 2009), pp. 3–27.
[301]O. San and T. Iliescu, "A stabilized proper orthogonal decomposition reduced-order model for large scale quasigeostrophic ocean circulation," Adv. Comput. Math. **41**, 1289–1319 (2015).
[302]C. Mou, B. Koc, O. San, L. G. Rebholz, and T. Iliescu, "Data-driven variational multiscale reduced order models," Comput. Methods Appl. Mech. Eng. **373**, 113470 (2021).
[303]M. D. Chekroun, H. Liu, and S. Wang, *Stochastic Parameterizing Manifolds and non-Markovian Reduced Equations: Stochastic Manifolds for Nonlinear SPDEs II* (Springer, Berlin, 2014).
[304]M. D. Chekroun, H. Liu, and S. Wang, *Approximation of Stochastic Invariant Manifolds: Stochastic Manifolds for Nonlinear SPDEs I* (Springer, Berlin, 2015), Vol. 1.
[305]A. J. Majda, J. Harlim, and B. Gershgorin, "Mathematical strategies for filtering turbulent dynamical systems," Discrete Contin. Dyn. Syst. **27**, 441–486 (2010).
[306]A. J. Majda and J. Harlim, "Physics constrained nonlinear regression models for time series," Nonlinearity **26**, 201 (2013).
[307]J. Harlim, A. Mahdi, and A. J. Majda, "An ensemble kalman filter for statistical estimation of physics constrained nonlinear regression models," J. Comput. Phys. **257**, 782–812 (2014).
[308]B. Koc, M. Mohebujjaman, C. Mou, and T. Iliescu, "Commutation error in reduced order modeling of fluid flows," Adv. Comput. Math. **45**, 2587–2621 (2019).
[309]D. Wells, Z. Wang, X. Xie, and T. Iliescu, "An evolve-then-filter regularized reduced order model for convection-dominated flows," Int. J. Numer. Methods Fluids **84**, 598–615 (2017).
[310]M. Gunzburger, T. Iliescu, M. Mohebujjaman, and M. Schneier, "An evolve-filter-relax stabilized reduced order stochastic collocation method for the time-dependent Navier–Stokes equations," SIAM/ASA J. Uncertainty Quantif. **7**, 1162–1184 (2019).
[311]M. Girfoglio, A. Quaini, and G. Rozza, "A POD-Galerkin reduced order model for a LES filtering approach," J. Comput. Phys. **436**, 110260 (2021).
[312]M. Girfoglio, A. Quaini, and G. Rozza, "Pressure stabilization strategies for a LES filtering reduced order model," preprint arXiv:2106.15887 (2021).
[313]F. Sabetghadam and A. Jafarpour, "$\alpha$ regularization of the POD-Galerkin dynamical systems of the Kuramoto–Sivashinsky equation," Appl. Math. Comput. **218**, 6012–6026 (2012).
[314]X. Xie, D. Wells, Z. Wang, and T. Iliescu, "Numerical analysis of the Leray reduced order model," J. Comput. Appl. Math. **328**, 12–29 (2018).
[315]M. Gunzburger, T. Iliescu, and M. Schneier, "A Leray regularized ensemble-proper orthogonal decomposition method for parameterized convection-dominated flows," IMA J. Numer. Anal. **40**, 886–913 (2020).
[316]H. Mori, "Transport, collective motion, and Brownian motion," Prog. Theor. Phys. **33**, 423–455 (1965).
[317]R. Zwanzig, "Problems in nonlinear transport theory," in *Systems Far from Equilibrium* (Springer, Berlin, 1980), pp. 198–225.
[318]A. J. Chorin, O. H. Hald, and R. Kupferman, "Optimal prediction and the Mori–Zwanzig representation of irreversible processes," Proc. Nat. Acad. Sci. U. S. A. **97**, 2968–2973 (2000).
[319]R. Zwanzig, *Nonequilibrium Statistical Mechanics* (Oxford University Press, Oxford, 2001).
[320]E. J. Parish and K. Duraisamy, "A dynamic subgrid scale model for large eddy simulations based on the Mori–Zwanzig formalism," J. Comput. Phys. **349**, 154–175 (2017).

[321]E. J. Parish and K. Duraisamy, "A unified framework for multiscale modeling using the Mori-Zwanzig formalism and the variational multiscale method," preprint arXiv:1712.09669 (2017).
[322]M. Gunzburger, N. Jiang, and M. Schneier, "An ensemble-proper orthogonal decomposition method for the nonstationary Navier-Stokes equations," SIAM J. Numer. Anal. **55**, 286–304 (2017).
[323]L. C. Berselli, T. Iliescu, B. Koc, and R. Lewandowski, "Long-time Reynolds averaging of reduced order models for fluid flows: Preliminary results," Math. Eng. **2**, 1–25 (2020).
[324]C. E. Leith, "Stochastic models of chaotic systems," Physica D **98**, 481–491 (1996).
[325]A. J. Chorin and O. H. Hald, *Stochastic Tools in Mathematics and Science* (Springer-Verlag, New York, 2009), Vol. 1.
[326]A. J. Majda, "Statistical energy conservation principle for inhomogeneous turbulent dynamical systems," Proc. Nat. Acad. Sci. U. S. A. **112**, 8937–8941 (2015).
[327]V. Resseguier, E. Mémin, and B. Chapron, "Stochastic fluid dynamic model and dimensional reduction," in *Ninth International Symposium on Turbulence and Shear Flow Phenomena* (Begel House Inc., 2015).
[328]F. Lu, K. K. Lin, and A. J. Chorin, "Data-based stochastic model reduction for the Kuramoto–Sivashinsky equation," Physica D **340**, 46–57 (2017).
[329]F. Lu, "Data-driven model reduction for stochastic Burgers equations," Entropy **22**, 1360 (2020).
[330]M. Sieber, C. O. Paschereit, and K. Oberleithner, "Stochastic modelling of a noise-driven global instability in a turbulent swirling jet," J. Fluid Mech. **916**, A7 (2021).
[331]D. S. Wilks, "Effects of stochastic parametrizations in the Lorenz'96 system," Q. J. R. Meteorol. Soc. **131**, 389–407 (2005).
[332]T. P. Sapsis, "Attractor local dimensionality, nonlinear energy transfers and finite-time instabilities in unstable dynamical systems with applications to two-dimensional fluid flows," Proc. R. Soc. A **469**, 20120550 (2013).
[333]T. P. Sapsis and A. J. Majda, "A statistically accurate modified quasilinear Gaussian closure for uncertainty quantification in turbulent dynamical systems," Physica D **252**, 34–45 (2013).
[334]T. P. Sapsis and A. J. Majda, "Statistically accurate low-order models for uncertainty quantification in turbulent dynamical systems," Proc. Natl. Acad. Sci. U. S. A. **110**, 13705–13710 (2013).
[335]H. Arnold, I. Moroz, and T. Palmer, "Stochastic parametrizations and model uncertainty in the Lorenz'96 system," Philos. Trans. R. Soc. A **371**, 20110479 (2013).
[336]M. Pulido, P. Tandeo, M. Bocquet, A. Carrassi, and M. Lucini, "Stochastic parameterization identification using ensemble Kalman filtering combined with maximum likelihood methods," Tellus A. **70**, 1–17 (2018).
[337]A. Chattopadhyay, A. Subel, and P. Hassanzadeh, "Data-driven super-parameterization using deep learning: Experimentation with multiscale Lorenz 96 systems and transfer learning," J. Adv. Model. Earth Syst. **12**, e2020MS002084 (2020).
[338]D. J. Gagne, H. M. Christensen, A. C. Subramanian, and A. H. Monahan, "Machine learning for stochastic parameterization: Generative adversarial networks in the Lorenz'96 model," J. Adv. Model. Earth Syst. **12**, e2019MS001896 (2020).
[339]T. Berry, D. Giannakis, and J. Harlim, "Nonparametric forecasting of low-dimensional dynamical systems," Phys. Rev. E **91**, 032915 (2015).
[340]E. N. Lorenz, "Predictability: A problem partly solved," in Proceedings of Seminar on Predictability (1996), Vol. 1.
[341]E. Mémin, "Fluid flow dynamics under location uncertainty," Geophys. Astrophys. Fluid Dyn. **108**, 119–146 (2014).
[342]Y. Zhou, "Turbulence theories and statistical closure approaches," Phys. Rep. (published online) (2021).
[343]B. R. Noack, P. Papas, and P. A. Monkewitz, "The need for a pressure-term representation in empirical Galerkin models of incompressible shear flows," J. Fluid Mech. **523**, 339–365 (2005).
[344]M. Buffoni, S. Camarri, A. Iollo, and M. V. Salvetti, "Low-dimensional modelling of a confined three-dimensional wake flow," J. Fluid Mech. **569**, 141–150 (2006).
[345]B. Galletti, A. Bottaro, C.-H. Bruneau, and A. Iollo, "Accurate model reduction of transient and forced wakes," Eur. J. Mech.-B/Fluids **26**, 354–366 (2007).
[346]M. Couplet, C. Basdevant, and P. Sagaut, "Calibrated reduced-order POD-Galerkin system for fluid flow modelling," J. Comput. Phys. **207**, 192–220 (2005).
[347]L. Perret, E. Collin, and J. Delville, "Polynomial identification of POD based low-order dynamical system," J. Turbul. **7**, N17 (2006).
[348]J. Baiges, R. Codina, and S. Idelsohn, "Reduced-order subscales for POD models," Comput. Methods Appl. Mech. Eng. **291**, 173–196 (2015).
[349]L. Cordier, B. Abou El Majd, and J. Favier, "Calibration of POD reduced-order models using Tikhonov regularization," Int. J. Num. Methods Fluids **63**, 269–296 (2010).
[350]S. Hijazi, G. Stabile, A. Mola, and G. Rozza, "Data-driven POD-Galerkin reduced order model for turbulent flows," J. Comput. Phys **416**, 109513 (2020).
[351]X. Xie, M. Mohebujjaman, L. G. Rebholz, and T. Iliescu, "Data-driven filtered reduced order modeling of fluid flows," SIAM J. Sci. Comput. **40**, B834–B857 (2018).
[352]C. Mou, H. Liu, D. R. Wells, and T. Iliescu, "Data-driven correction reduced order models for the quasi-geostrophic equations: A numerical investigation," Int. J. Comput. Fluid Dyn. **34**, 147–159 (2020).
[353]C. Mou, "Cross-validation of the data-driven correction reduced order model," M.S. thesis (Virginia Tech, 2018).
[354]A. Ivagnes, "Data enhanced reduced order methods for turbulent flows," Ph.D. thesis (Politecnico di Torino, 2021).
[355]B. Koc, C. Mou, H. Liu, Z. Wang, G. Rozza, and T. Iliescu, "Verifiability of the data-driven variational multiscale reduced order model," preprint arXiv:2108.04982 (2021).
[356]K. K. Lin and F. Lu, "Data-driven model reduction, Wiener projections, and the Koopman-Mori-Zwanzig formalism," J. Comput. Phys. **424**, 109864 (2021).
[357]E. J. Parish, C. Wentland, and K. Duraisamy, "A residual-based Petrov-Galerkin reduced-order model with memory effects," preprint arXiv:1810.03455 (2018).
[358]M. Mohebujjaman, L. G. Rebholz, and T. Iliescu, "Physically-constrained data-driven correction for reduced order modeling of fluid flows," Int. J. Numer. Methods Fluids **89**, 103–122 (2019).
[359]S. Pawar, S. E. Ahmed, O. San, and A. Rasheed, "An evolve-then-correct reduced order model for hidden fluid dynamics," Mathematics **8**, 570 (2020).
[360]S. Fresca, L. Dede, and A. Manzoni, "A comprehensive deep learning-based approach to reduced order modeling of nonlinear time-dependent parametrized pdes," J. Sci. Comput. **87**, 1–36 (2021).
[361]B. C. Csáji, "Approximation with artificial neural networks," M.Sc. thesis (Faculty of Sciences, Eötvös Loránd University, Hungary, 2001).
[362]O. San, R. Maulik, and M. Ahmed, "An artificial neural network framework for reduced order modeling of transient flows," Commun. Nonlinear Sci. Numer. Simul. **77**, 271–287 (2019).
[363]S. Pawar, S. Rahman, H. Vaddireddy, O. San, A. Rasheed, and P. Vedula, "A deep learning enabler for nonintrusive reduced order modeling of fluid flows," Phys. Fluids **31**, 085101 (2019).
[364]J. S. Hesthaven and S. Ubbiali, "Non-intrusive reduced order modeling of nonlinear problems using neural networks," J. Comput. Phys. **363**, 55–78 (2018).
[365]Q. Wang, J. S. Hesthaven, and D. Ray, "Non-intrusive reduced order modeling of unsteady flows using artificial neural networks with application to a combustion problem," J. Comput. Phys. **384**, 289–307 (2019).
[366]P. Wu, J. Sun, X. Chang, W. Zhang, R. Arcucci, Y. Guo, and C. C. Pain, "Data-driven reduced order model with temporal convolutional neural network," Comput. Methods Appl. Mech. Eng. **360**, 112766 (2020).
[367]I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, Cambridge, Massachusetts, 2016).
[368]J. N. Kani and A. H. Elsheikh, "DR-RNN: A deep residual recurrent neural network for model reduction," preprint arXiv:1709.00939 (2017).
[369]P. R. Vlachas, W. Byeon, Z. Y. Wan, T. P. Sapsis, and, and P. Koumoutsakos, "Data-driven forecasting of high-dimensional chaotic systems with long short-term memory networks," Proc. R. Soc. A **474**, 20170844 (2018).
[370]A. T. Mohan and D. V. Gaitonde, "A deep learning based approach to reduced order modeling for turbulent flow control using LSTM neural networks," preprint arXiv:1804.09269 (2018).

[371]J. Graham, K. Kanov, X. Yang, M. Lee, N. Malaya, C. Lalescu, R. Burns, G. Eyink, A. Szalay, R. Moser *et al.*, "A web services accessible database of turbulent channel flow and its use for testing a new integral wall model for LES," J. Turbul. **17**, 181–215 (2016).
[372]S. M. Rahman, S. Pawar, O. San, A. Rasheed, and T. Iliescu, "A nonintrusive reduced order modeling framework for quasigeostrophic turbulence," Phys. Rev. E **100**, 053306 (2019).
[373]O. San and R. Maulik, "Machine learning closures for model order reduction of thermal fluids," Appl. Math. Modell. **60**, 681–710 (2018).
[374]G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," Neurocomputing **70**, 489–501 (2006).
[375]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**, e0197704 (2018).
[376]A. J. Chorin, O. H. Hald, and R. Kupferman, "Optimal prediction with memory," Physica D **166**, 239–257 (2002).
[377]F. Takens, "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence, Warwick* 1980 (Springer, 1981), pp. 366–381.
[378]C. E. Rasmussen, "Gaussian processes in machine learning," in *Summer School on Machine Learning* (Springer, New York, 2003), pp. 63–71.
[379]M. Raissi and G. E. Karniadakis, "Hidden physics models: Machine learning of nonlinear partial differential equations," J. Comput. Phys. **357**, 125–141 (2018).
[380]V. C. Raykar, R. Duraiswami, and L. H. Zhao, "Fast computation of kernel estimators," J. Comput. Graphical Stat. **19**, 205–220 (2010).
[381]K. Chalupka, C. K. Williams, and I. Murray, "A framework for evaluating approximation methods for Gaussian process regression," J. Mach. Learn. Res. **14**, 333–350 (2013).
[382]A. I. Forrester, A. Sóbester, and A. J. Keane, "Multi-fidelity optimization via surrogate modelling," Proc. R. Soc. A **463**, 3251–3269 (2007).
[383]P. Perdikaris, D. Venturi, J. O. Royset, and G. E. Karniadakis, "Multi-fidelity modelling via recursive co-kriging and Gaussian–Markov random fields," Proc. R. Soc. A **471**, 20150018 (2015).
[384]A. Feldstein, D. Lazzara, N. Princen, and K. Willcox, "Multifidelity data fusion: Application to blended-wing-body multidisciplinary analysis under uncertainty," AIAA J. **58**, 889–906 (2020).
[385]Z. Y. Wan and T. P. Sapsis, "Reduced-space gaussian process regression for data-driven probabilistic forecast of chaotic dynamical systems," Physica D **345**, 40–55 (2017).
[386]R. Maulik, T. Botsas, N. Ramachandra, L. R. Mason, and I. Pan, "Latent-space time evolution of non-intrusive reduced-order models using Gaussian process emulation," Physica D **416**, 132797 (2021).
[387]D. Xiao, F. Fang, A. G. Buchan, C. C. Pain, I. M. Navon, and A. Muggeridge, "Non-intrusive reduced order modelling of the Navier–Stokes equations," Comput. Methods Appl. Mech. Eng. **293**, 522–541 (2015).
[388]D. Xiao, F. Fang, C. Pain, and G. Hu, "Non-intrusive reduced-order modelling of the Navier–Stokes equations based on RBF interpolation," Int. J. Numer. Methods Fluids **79**, 580–595 (2015).
[389]D. Rajaram, C. Perron, T. G. Puranik, and D. N. Mavris, "Randomized algorithms for non-intrusive parametric reduced order modeling," AIAA J. **58**(12), 5389–5407 (2020).
[390]A. J. Geer, "Learning earth system models from observations: Machine learning or data assimilation?," Philos. Trans. R. Soc. A **379**, 20200089 (2021).
[391]Y. Tang and W. W. Hsieh, "Coupling neural networks to incomplete dynamical systems via variational data assimilation," Mon. Weather Rev. **129**, 818–834 (2001).
[392]A. Liaqat, M. Fukuhara, and T. Takeda, "Applying a neural network collocation method to an incompletely known dynamical system via weak constraint data assimilation," Mon. Weather Rev. **131**, 1696–1714 (2003).
[393]J. Brajard, A. Carrassi, M. Bocquet, and L. Bertino, "Combining data assimilation and machine learning to infer unresolved scale parametrization," Philos. Trans. R. Soc. A **379**, 20200086 (2021).
[394]J. Brajard, A. Carrassi, M. Bocquet, and L. Bertino, "Combining data assimilation and machine learning to emulate a dynamical model from sparse and noisy observations: A case study with the Lorenz 96 model," J. Comput. Sci. **44**, 101171 (2020).
[395]M. Bonavita and P. Laloyaux, "Machine learning for model error inference and correction," J. Adv. Model. Earth Syst. **12**, e2020MS002232 (2020).
[396]A. Farchi, P. Laloyaux, M. Bonavita, and M. Bocquet, "Using machine learning to correct model error in data assimilation and forecast applications," preprint arXiv:2010.12605 (2020).
[397]M. Bocquet, J. Brajard, A. Carrassi, and L. Bertino, "Bayesian inference of chaotic dynamics by merging data assimilation, machine learning and expectation-maximization," Found. Data Sci. **2**, 55–80 (2020).
[398]A. Farchi, M. Bocquet, P. Laloyaux, M. Bonavita, and Q. Malartic, "A comparison of combined data assimilation and machine learning methods for offline and online model error correction," preprint arXiv:2107.11114 (2021).
[399]S. Pawar and O. San, "Data assimilation empowered neural network parametrizations for subgrid processes in geophysical flows," Phys. Rev. Fluids **6**, 050501 (2021).
[400]S. Pawar, O. San, A. Rasheed, and I. M. Navon, "A nonintrusive hybrid neural-physics modeling of incomplete dynamical systems: Lorenz equations," preprint arXiv:2104.00114 (2021).
[401]S. E. Ahmed, S. Pawar, and O. San, "PyDA: A hands-on introduction to dynamical data assimilation with python," Fluids **5**, 225 (2020).
[402]J. D'Adamo, N. Papadakis, E. Memin, and G. Artana, "Variational assimilation of POD low-order dynamical systems," J. Turbul. **8**, N9 (2007).
[403]G. Artana, A. Cammilleri, J. Carlier, and E. Mémin, "Strong and weak constraint variational assimilations for reduced order fluid flow modeling," J. Comput. Phys. **231**, 3264–3288 (2012).
[404]C. Zerfas, L. G. Rebholz, M. Schneier, and T. Iliescu, "Continuous data assimilation reduced order models of fluid flow," Comput. Methods Appl. Mech. Eng. **357**, 112596 (2019).
[405]S. E. Ahmed, S. Pawar, O. San, and A. Rasheed, "Reduced order modeling of fluid flows: Machine learning, Kolmogorov barrier, closure modeling, and partitioning," in *AIAA AVIATION 2020 FORUM* (AIAA, 2020), p. 2946.
[406]S. E. Ahmed, S. Pawar, O. San, A. Rasheed, and M. Tabib, "A nudged hybrid analysis and modeling approach for realtime wake-vortex transport and decay prediction," Comput. Fluids **221**, 104895 (2021).
[407]S. E. Ahmed, K. Bhar, O. San, and A. Rasheed, "Forward sensitivity approach for estimating eddy viscosity closures in nonlinear model reduction," Phys. Rev. E **102**, 043302 (2020).
[408]E. Qian, B. Kramer, B. Peherstorfer, and K. Willcox, "Lift & learn: Physics-informed machine learning for large-scale nonlinear dynamical systems," Physica D **406**, 132401 (2020).
[409]P. Benner, P. Goyal, B. Kramer, B. Peherstorfer, and K. Willcox, "Operator inference for non-intrusive model reduction of systems with non-polynomial nonlinear terms," Comput. Methods Appl. Mech. Eng. **372**, 113433 (2020).
[410]S. A. McQuarrie, C. Huang, and K. E. Willcox, "Data-driven reduced-order models via regularised operator inference for a single-injector combustion process," J. R. Soc. New Zealand **51**, 194–211 (2021).
[411]B. Peherstorfer, "Sampling low-dimensional Markovian dynamics for preasymptotically recovering reduced models from data with operator inference," SIAM J. Sci. Comput. **42**, A3489–A3515 (2020).
[412]S. Yildiz, P. Goyal, P. Benner, and B. Karasözen, "Learning reduced-order dynamics for parametrized shallow water equations from data," Int. J. Num. Methods Fluids **93**(8), 2803–2821 (2021).
[413]J.-C. Loiseau and S. L. Brunton, "Constrained sparse Galerkin regression," J. Fluid Mech. **838**, 42–67 (2018).
[414]Y. S. Abu-Mostafa, M. Magdon-Ismail, and H.-T. Lin, *Learning from Data* (AMLBook New York, 2012).
[415]M. Quade, M. Abel, K. Shafi, R. K. Niven, and B. R. Noack, "Prediction of dynamical systems by symbolic regression," Phys. Rev. E **94**, 012214 (2016).
[416]C. Luo and S.-L. Zhang, "Parse-matrix evolution for symbolic regression," Eng. Appl. Artif. Intell. **25**, 1182–1193 (2012).
[417]B. de Silva, K. Champion, M. Quade, J.-C. Loiseau, J. Kutz, and S. Brunton, "PySINDy: A Python package for the sparse identification of nonlinear dynamical systems from data," J. Open Source Software **5**, 2104–2104 (2020).
[418]S. H. Rudy, S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Data-driven discovery of partial differential equations," Sci. Adv. **3**, e1602614 (2017).
[419]J. R. Koza, *Genetic Programming: On the programming of Computers by Means of Natural Selection* (MIT Press, Cambridge, MA, 1992), Vol. 1.

[420]J. Bongard and H. Lipson, "Automated reverse engineering of nonlinear dynamical systems," Proc. Natl. Acad. Sci. U. S. A. **104**, 9943–9948 (2007).
[421]M. Schmidt and H. Lipson, "Distilling free-form natural laws from experimental data," Science **324**, 81–85 (2009).
[422]C. Ferreira, "Gene expression programming: A new adaptive algorithm for solving problems," preprint arXiv:cs/0102027 (2001).
[423]J.-C. Loiseau, S. L. Brunton, and B. R. Noack, "From the POD-Galerkin method to sparse manifold models," Handb. Model-Order Reduct. **2**, 1–47 (2020).
[424]A. A. Kaptanoglu, J. L. Callaham, C. J. Hansen, A. Aravkin, and S. L. Brunton, "Promoting global stability in data-driven models of quadratic nonlinear dynamics," arXiv:2105.01843 (2021).
[425]H. Vaddireddy, A. Rasheed, A. E. Staples, and O. San, "Feature engineering and symbolic regression methods for detecting hidden physics from sparse sensor observation data," Phys. Fluids **32**, 015113 (2020).
[426]K. Kaheman, E. Kaiser, B. Strom, J. N. Kutz, and S. L. Brunton, "Learning discrepancy models from experimental data," preprint arXiv:1909.08574 (2019).
[427]H. Gao, J.-X. Wang, and M. J. Zahr, "Non-intrusive model reduction of large-scale, nonlinear dynamical systems using deep learning," Physica D **412**, 132614 (2020).
[428]J. Xu and K. Duraisamy, "Multi-level convolutional autoencoder networks for parametric prediction of spatio-temporal dynamics," Comput. Methods Appl. Mech. Eng. **372**, 113379 (2020).
[429]J.-W. Lee and J.-H. Oh, "Hybrid learning of mapping and its Jacobian in multilayer neural networks," Neural Comput. **9**, 937–958 (1997).
[430]S. Marsland, J. Shapiro, and U. Nehmzow, "A self-organising network that grows when required," Neural Networks **15**, 1041–1058 (2002).
[431]M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," J. Comput. Phys. **378**, 686–707 (2019).
[432]Y. Zhu, N. Zabaras, P.-S. Koutsourelakis, and P. Perdikaris, "Physics-constrained deep learning for high-dimensional surrogate modeling and uncertainty quantification without labeled data," J. Comput. Phys. **394**, 56–81 (2019).
[433]S. Pan and K. Duraisamy, "Physics-informed probabilistic learning of linear embeddings of nonlinear dynamics with guaranteed stability," SIAM J. Appl. Dyn. Syst. **19**, 480–509 (2020).
[434]G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, "Physics-informed machine learning," Nat. Rev. Phys. **3**, 422–440 (2021).
[435]A. T. Mohan, N. Lubbers, D. Livescu, and M. Chertkov, "Embedding hard physical constraints in neural network coarse-graining of 3D turbulence," preprint arXiv:2002.00021 (2020).
[436]T. Beucler, S. Rasp, M. Pritchard, and P. Gentine, "Achieving conservation of energy in neural network emulators for climate modeling," preprint arXiv:1906.06622 (2019).
[437]K. Meidani and A. B. Farimani, "Data-driven identification of 2D partial differential equations using extracted physical features," Comput. Methods Appl. Mech. Eng. **381**, 113831 (2021).
[438]C. Leibig, V. Allken, M. S. Ayhan, P. Berens, and S. Wahl, "Leveraging uncertainty information from deep neural networks for disease detection," Sci. Rep. **7**, 1–14 (2017).
[439]D. Sacha, M. Sedlmair, L. Zhang, J. A. Lee, J. Peltonen, D. Weiskopf, S. C. North, and D. A. Keim, "What you see is what you can change: Human-centered machine learning by interactive visualization," Neurocomputing **268**, 164–175 (2017).
[440]S. Pawar, O. San, B. Aksoylu, A. Rasheed, and T. Kvamsdal, "Physics guided machine learning using simplified theories," Phys. Fluids **33**, 011701 (2021).
[441]S. Pawar, O. San, A. Nair, A. Rasheed, and T. Kvamsdal, "Model fusion with physics-guided machine learning: Projection-based reduced-order modeling," Phys. Fluids **33**, 067123 (2021).
[442]S. Pawar, S. E. Ahmed, O. San, and A. Rasheed, "Data-driven recovery of hidden physics in reduced order modeling of fluid flows," Phys. Fluids **32**, 036602 (2020).
[443]M. Bonavita, R. Arcucci, A. Carrassi, P. Dueben, A. J. Geer, B. Le Saux, N. Longépé, P.-P. Mathieu, and L. Raynaud, "Machine learning for earth system observation and prediction," Bull. Am. Meteorol. Soc. **102**, E710–E716 (2021).
[444]J. Ling, A. Kurzawski, and J. Templeton, "Reynolds averaged turbulence modelling using deep neural networks with embedded invariance," J. Fluid Mech. **807**, 155–166 (2016).
[445]A. Subramaniam, M. L. Wong, R. D. Borker, S. Nimmagadda, and S. K. Lele, "Turbulence enrichment using physics-informed generative adversarial networks," arXiv–2003 (2020).
[446]M. Bode, M. Gauding, Z. Lian, D. Denker, M. Davidovic, K. Kleinheinz, J. Jitsev, and H. Pitsch, "Using physics-informed enhanced super-resolution generative adversarial networks for subfilter modeling in turbulent reactive flows," Proc. Combust. Inst. **38**, 2617 (2021).
[447]N. B. Erichson, M. Muehlebach, and M. W. Mahoney, "Physics-informed autoencoders for Lyapunov-stable fluid flow prediction," preprint arXiv:1905.10866 (2019).
[448]J.-L. Wu, K. Kashinath, A. Albert, D. Chirila, H. Xiao *et al.*, "Enforcing statistical constraints in generative adversarial networks for modeling chaotic dynamical systems," J. Comput. Phys. **406**, 109209 (2020).
[449]N. Geneva and N. Zabaras, "Modeling the dynamics of PDE systems with physics-constrained deep auto-regressive networks," J. Comput. Phys. **403**, 109056 (2020).
[450]W. Chen, Q. Wang, J. S. Hesthaven, and C. Zhang, "Physics-informed machine learning for reduced-order modeling of nonlinear problems," preprint (2020).
[451]K. Lee and K. Carlberg, "Deep conservation: A latent-dynamics model for exact satisfaction of physical conservation laws," preprint arXiv:1909.09754 (2019).
[452]A. A. Kaptanoglu, K. D. Morgan, C. J. Hansen, and S. L. Brunton, "Physics-constrained, low-dimensional models for magnetohydrodynamics: First-principles and data-driven approaches," Phys. Rev. E **104**, 015206 (2021).
[453]N. Sawant, B. Kramer, and B. Peherstorfer, "Physics-informed regularization and structure preservation for learning stable reduced models from data with operator inference," preprint arXiv:2107.02597 (2021).
[454]R. Swischuk, B. Kramer, C. Huang, and K. Willcox, "Learning physics-based reduced-order models for a single-injector combustion process," AIAA J. **58**, 2658–2672 (2020).
[455]S. Beetham and J. Capecelatro, "Formulating turbulence closures using sparse regression with embedded form invariance," Phys. Rev. Fluids **5**, 084611 (2020).
[456]S. Siegel, K. Cohen, J. Seidel, D. Luchtenburg, and T. McLaughlin, "Low dimensional modelling of a transient cylinder wake using double proper orthogonal decomposition," J. Fluid Mech. **610**, 1–42 (2008).