An Efficient Data-Driven Multiscale Stochastic Reduced Order Modeling Framework for Complex Systems

Changhong Mou

Department of Mathematics, University of Wisconsin-Madison, 480 Lincoln Drive, Madison, WI 53706, USA.

Nan Chen

Department of Mathematics, University of Wisconsin-Madison, 480 Lincoln Drive, Madison, WI 53706, USA.

Traian Iliescu

Department of Mathematics, Virginia Tech, Virginia Tech, Blacksburg, VA 24061, USA.

Abstract

Suitable reduced order models (ROMs) are computationally efficient tools in characterizing key dynamical and statistical features of nature. In this paper, a systematic multiscale stochastic ROM framework is developed for complex systems with strong chaotic or turbulent behavior. The new ROMs are fundamentally different from the traditional Galerkin ROM (G-ROM) or those deterministic ROMs that aim at minimizing the path-wise errors and applying mainly to laminar systems. Here, the new ROM focuses on recovering the large-scale dynamics to the maximum extent while it also exploits cheap but effective conditional linear functions as the closure terms to capture the statistical features of the medium-scale variables and its feedback to the large scales. In addition, physics constraints are incorporated into the new ROM. One unique feature of the resulting ROM is that it facilitates an efficient and accurate scheme for nonlinear data assimilation, the solution of which is provided by closed analytic formulae. Such an analytic solvable data assimilation solution significantly accelerates the computational efficiency and allows the new ROM to avoid many potential numerical

Preprint submitted to arXiv.org

March 23, 2022
and sampling issues in recovering the unobserved states from partial observations. The overall model calibration is efficient and systematic via explicit mathematical formulae. The new ROM framework is applied to complex nonlinear systems, in which the intrinsic turbulent behavior is either triggered by external random forcing or deterministic nonlinearity. It is shown that the new ROM significantly outperforms the G-ROM in both scenarios in terms of reproducing the dynamical and statistical features as well as recovering unobserved states via the associated efficient data assimilation scheme.

**Keywords:** Reduced Order Models, Data Assimilation, Physics Constraints, Analytically Solvable Conditional Statistics, Turbulent Systems

**2010 MSC:** 37N10, 65C20, 86A04, 62M10

1. **Introduction**

Modeling complex nonlinear turbulent systems is of practical importance. These complex nonlinear system appear in many areas [1, 2, 3, 4, 5], such as geophysics, climate science, engineering, atmosphere and ocean science, neural science, and material science. They often involve strong nonlinearities and multiscale structures [3, 6, 7, 8]. Energy is transferred across different scales via the intrinsic nonlinearity that triggers regime switching, intermittent instabilities, extreme events and many other chaotic and turbulent features [9, 10, 11, 12, 13, 14, 15, 16]. As a result, the associated probability density functions (PDFs) are usually non-Gaussian with fat tails. Building appropriate mathematical models for complex systems that succeed in characterizing these key features not only advances the understanding of nature, but facilitates effective state estimation and skillful forecast as well. However, the complexity of nature makes it extremely challenging to develop a full order model (FOM) and apply it to forecast many quantities of interest. In fact, due to the lack of physical understanding or the inadequate resolution in computations, a perfect FOM is never available in practice. Even if a nearly perfect FOM is accessible, the model is often of high dimensionality and contains a complicated model structure [17, 18]. As a consequence, there exist many mathematical and computational difficulties in analyzing and simulating such a FOM.
Suitable reduced order models (ROMs) \cite{19, 20, 21, 22, 23} are desirable surrogates for the FOMs to reproduce certain key dynamical and statistical features of nature with a much lower computational cost. Linear regression models are possibly the simplest class of ROMs \cite{24, 25} that can provide certain skill for short-term forecasts but they usually struggle in characterizing the underlying nonlinear physics. Beyond the regression strategies, one systematic and commonly used approach to developing ROMs is to project the starting complex nonlinear system to the leading a few energetic modes in light of the Galerkin proper orthogonal decomposition methods \cite{20} or other empirical basis functions such as the principal interaction patterns \cite{26, 27} and the dynamic mode decomposition \cite{28, 29}. With a careful design of the closure terms (see \cite{30} for a survey) to compensate the truncation error, these ROMs are skillful in resolving certain problems in many areas, including fluids and turbulence \cite{31, 32, 21, 23, 33}. While many traditional ROM strategies focus on seeking for the optimal approximate models with deterministic model structures, stochastic effects have been incorporated into many recent developments of ROMs for complex turbulent systems. The stochasticity is essential in assessing the uncertainty and model error in ROMs. It also plays a pivotal role in facilitating data assimilation and advancing probabilistic forecast, which is vital for predicting non-Gaussian extreme events. The MTV (named after Majda, Timofeyev and Vanden Eijnden) strategy is one of the most straightforward approaches to incorporate stochastic effects in the ROMs \cite{34, 35}, the idea of which is to replace the self-interactions among the high frequencies by random noise. The recently proposed nonlinear autoregression (NAR) model emphasizes the random effect in the development of regression-based ROMs and its significance in advancing the statistical forecast \cite{36, 37}. On the other hand, physics-constrained regression models are a set of nonlinear stochastic ROMs \cite{38, 39, 40}, which take into account the energy conserving nonlinear interactions in the model development that guarantees the well-posedness of long-term behavior of the system. In addition, stochastic parameterizations are powerful ways to reduce the complexity of the unresolved part of the dynamics that nevertheless provides effective statistical feedbacks to the resolved-scale dynamics \cite{41, 42, 43, 44, 45}.

The goal of this paper is to build a systematic multiscale stochastic ROM frame-
work for complex systems with strong chaotic or turbulent behavior. The new ROM framework has several unique features. First, a new strategy of building effective closure terms is incorporated into the framework that includes both deterministic and stochastic components. Fundamentally different from the traditional closure methods, the development of the closures in this new ROM framework exploits the multiscale characteristics of the underlying system to reach a systematic formulation that is mathematically more concise and tractable than most of the traditional nonlinear closures. Nevertheless, appropriate stochastic closure terms are combined with the deterministic ones in this framework that succeeds in providing effective statistical feedbacks between different scales. These closure terms significantly advance the ROM in reproducing both the dynamics and statistics of nature. Second, data assimilation is a necessary procedure in understanding and forecasting turbulent systems [46, 47, 48, 49]. It aims at recovering the unobserved state variables using Bayesian inference, and is the prerequisite of the ensemble forecast for turbulent flows. Different from most of the traditional ROMs that have to rely on ensemble based data assimilation methods for state estimation and the initialization of the forecast, the ROM with the new concise closures facilitates an efficient and accurate scheme for nonlinear data assimilation, the solution of which is provided by closed analytic formulae. Such an analytic solvable data assimilation solution increases the computational efficiency and allows the new ROM to avoid many potential numerical and sampling issues that usually require careful ad hoc tunings. Third, physics constraints can be naturally incorporated into this new ROM framework, which is essential for preserving the basic principles in physics and preventing finite-time blow-up solutions in the statistical forecast [38, 39, 50]. Another desirable feature is that the ROM and the associated closures can be systematically determined via a computationally efficient optimization procedure with no ad hoc tuning. Note that the development of the ROM in this framework does not require to know the exact FOM, which is often not accessible in practice. The framework is in fact amenable to data-driven scenarios. Specifically, the partially available information from the FOM can be used to determine the large-scale structure of the ROM, while the remaining part of the ROM including the closures can be inferred from data. Finally, it is worthwhile to highlight that many well-known complex nonlinear systems
in geophysics, neural science, and engineering turbulence have the same structure as those belonging to the new ROM framework [51], which is a further justification of the framework.

The new ROM framework will be applied to complex nonlinear systems that belong to two categories, distinguished by the triggering mechanisms of the turbulent behavior. The first test model is a viscous stochastic Burgers equation [52, 53]. In such a complex system, external random forcing interacts with the deterministic nonlinearity to induce intermittent signals in the form of viscous shocks, which will be dissipated via the viscous term before the appearance of discontinuities. Here, the stochastic forcing triggers turbulent features. The second test example is the quasi-geostrophic (QG) equation [1, 9]. The QG equation is a deterministic system and the turbulent features are induced completely by the nonlinear interactions between the state variables at different spatial scales. In both cases, only a small number of the state variables are involved in the ROM that allows a much lower computational cost to reproduce the key dynamical and statistical features. It is also significant to see the skill of the stochastic closure terms in effectively characterizing the turbulent features triggered completely by the deterministic nonlinearity.

The rest of the paper is organized as follows. Section 2 includes the new ROM framework, the associated data assimilation scheme, and the model calibration. Section 3 demonstrates the skill of the new ROMs in reproducing the key dynamical and statistical features of both the viscous stochastic Burgers equation and the QG system. The comparison with the state-of-the-art ROMs in terms of both the model simulation and the data assimilation skill is presented in this section as well. The paper is concluded in Section 4, together with a discussion of potential extensions of the current framework. All the technical details are included in Appendix.
2. The New Reduced Order Modeling (ROM) Framework

2.1. The mathematical framework of the ROM with a conditional linear closure

Many complex turbulent nonlinear systems in nature have the following abstract mathematical structure [6, 1, 9, 47],

\[
\frac{du}{dt} = (L + D)u + B(u, u) + F(t) + \sigma(t) \dot{W}(t),
\]

(1)

where \( u \in \mathbb{C}^N \) is the multidimensional state variable and \( t \) is time. In (1), \( Lu \) and \( Du \) represent linear dispersion and dissipation effects, respectively, where \( L \) is skew-symmetric and \( D \) is negative-definite. The nonlinear effect is introduced through an energy-conserving quadratic form with \( u \cdot B(u, u) = 0 \). The system is also subject to external forcing effects that contain a deterministic component, \( F(t) \), and possibly a stochastic disturbance as well, which is assumed to be a Gaussian random noise, \( \sigma(t) \dot{W}(t) \), for simplicity. The model in (1) can be understood as discretizing a (stochastic) partial differential equation by projecting it onto a given set of basis functions \( \{ \varphi_1, \varphi_2, \ldots, \varphi_N \} \), resulting in a high-dimensional system, where the stochastic forcing accounts for the truncation error. Due to the intrinsic turbulent behavior, the system in (1) is highly nonlinear and the state variable \( u \) can possess strongly non-Gaussian statistics in both the marginal and joint PDFs.

Decompose the state variable \( u \) into \( u = (v^T, w^T, z^T)^T \) with \( :^T \) being the vector transpose, where \( v \in \mathbb{C}^{N_1}, w \in \mathbb{C}^{N_2}, z \in \mathbb{C}^{N_3} \), and \( N_1 + N_2 + N_3 = N \). A typical criterion for such a decomposition is based on the rank of the scales for the state variables. In other words, \( v \) includes the leading few modes corresponding to the large-scale components of the state variables, \( w \) contains the next a few modes describing the medium-scale features, while \( z \) involves the remaining modes that are treated as the small-scale variables. Note that different from many traditional ROM development, where the state variables are decomposed into only the large- and small-scale components, a three-scale decomposition is utilized here. Such a new decomposition facilitates a smoother transition from large- to small-scale variables by means of a set of intermediate-scale variables. It also allows the feedback from medium- to large-scale variables to be explicitly incorporated into the ROM, which is often crucial in enhancing the model accuracy in characterizing energy transfer and intermittency. Finally, the
choice of $N_1, N_2,$ and $N_3$ is case dependent but in a typical scenario $z$ accounts for the majority number of the modes.

A natural way to build ROMs with a reduced dimension of the coupled system in (1) is to project $u$ onto its large- and medium-scale components $v$ and $w$,

$$
\frac{dv}{dt} = A_{(v)}^{(u)} v + v^{*}\tilde{B}_{v,v}^{(u)} v + v^{*}\tilde{B}_{v,w}^{(u)} w + w^{*}\tilde{B}_{w,v}^{(u)} w + \text{residual}_1, \quad (2a)
$$

$$
\frac{dw}{dt} = A_{(w)}^{(u)} w + v^{*}\tilde{B}_{v,v}^{(u)} v + v^{*}\tilde{B}_{v,w}^{(u)} w + w^{*}\tilde{B}_{w,w}^{(u)} w + \text{residual}_2, \quad (2b)
$$

where $\cdot^{*}$ is the conjugate transpose. It is seen that the linear and the quadratic nonlinear terms in (1) have been explicitly expressed in (2). Therefore, $A^{(\cdot)}$ and $\tilde{B}^{(\cdot)}$ in (2) are all constant matrices and tensors while the two residual terms include the contribution from the variable $z$ that has been truncated here. The notation in (2) is slightly abused in the sense that the quadratic nonlinear terms between $v$ and $w$ are simply written as $v^{*}\tilde{B}_{v,v}^{(u)} w$ and $v^{*}\tilde{B}_{v,w}^{(u)} w$ while the other parts of the associated nonlinear terms, say $w^{*}\tilde{B}_{v,v}^{(u)} v$ and $w^{*}\tilde{B}_{v,w}^{(u)} v$, are omitted in (2) for notation simplicity. Note that if the state variables are all real-valued, then $v^{*}\tilde{B}_{v,v}^{(u)} w$ and $v^{*}\tilde{B}_{v,w}^{(u)} w$ are sufficient to characterize the quadratic nonlinear interactions between $v$ and $w$. Since each complex-valued variable can always be rewritten as two real-valued variables, representing its real and imaginary parts, all the state variables in the following will be regarded as real-valued. Similarly, the notation $\cdot^{*}$ hereafter simply denotes the transpose of a real-valued vector. In the development of ROMs, various methods have been proposed to deal with the residual terms. If these residual terms are simply ignored, then the resulting ROM becomes the so-called Galerkin ROM (G-ROM) [26, 27]. However, it has been shown that the G-ROM often introduces inaccuracy in reproducing even the large-scale dynamics of the FOM. Such a finding implies the importance of building approximate functions to compensate the contribution from the residuals. If these functions depend only on the state variables of the ROM, then they are known as the closures since they allow the ROM to be closed [30, 31, 21, 23, 33]. In most of the existing methods, the closure terms are assumed to be a linear or quadratic function of the entire ROM state variables, which are $(v, w)$ in the setup of (2).

The new ROM to be developed here has several fundamental differences compared with the existing ROMs. First, only the first three terms on the right hand side of (2)

7
are retained in the new ROM framework, which represent the basic dynamics. In other words, the self nonlinear interactions of $w$ are omitted. One motivation for such a simplification is that $w$ is a smaller scale variable compared with $v$ and therefore it is often a faster variable and contains less energy. As a consequence, $w^* \tilde{B}_{v,v}^{(v)} w$ and $w^* \tilde{B}_{v,w}^{(w)} w$ are expected to have a weaker role than the other nonlinear terms in (2).

Another important reason is related to the development of an efficient data assimilation scheme for the ROM, which will be discussed in Section 2.2. Next, a hybrid closure configuration is introduced to approximate the residual part as well as the contribution of the self nonlinear interactions of $w$. Here, the closure includes both a set of deterministic functions and a stochastic component. The building block of the deterministic part of the closure in this new ROM framework is very distinct from that in the existing methods, which is either linear or quadratic nonlinear in terms of both $v$ and $w$. Here, a special type of nonlinear functions are adopted, which are fully nonlinear with respect to $v$ but are conditional linear with respect to $w$. The motivation is again that, due to the relatively faster nature of $w$, a conditional linear function and stochastic terms can effectively approximate the statistical behavior of $w$ and its feedback to the large-scale component, $v$. With these new ingredients, the resulting ROM with the closure reads:

$$\frac{dv}{dt} = A_v^{(v)} v + v^* B_{v,v}^{(v)} v + v^* B_{v,w}^{(v)} w + (\tau_{w}^{(v)} (v) w + \tau_{w}^{(v)} (v)) + \sigma_v \dot{W}_v, \quad (3a)$$

$$\frac{dw}{dt} = A_w^{(w)} w + v^* B_{v,v}^{(w)} v + v^* B_{v,w}^{(w)} w + (\tau_{w}^{(w)} (v) w + \tau_{w}^{(w)} (v)) + \sigma_w \dot{W}_w, \quad (3b)$$

where $\dot{W}_v$ and $\dot{W}_w$ are the standard Gaussian white noises. The noise coefficients $\sigma_v$ and $\sigma_w$ are constant matrices and for simplicity they are diagonal. The four matrix or vector functions in the closure terms $\tau_{w}^{(v)} (v)$, $\tau_{w}^{(v)} (v)$, $\tau_{w}^{(w)} (v)$, and $\tau_{w}^{(w)} (v)$ can be arbitrarily nonlinear functions of $v$. It is worthwhile to highlight that the conditional linear functions of $w$ are not linear functions of $w$. The conditional linearity means that $w$ is linear with $v$ being fixed. Therefore, the closure in (3) includes the effects from the nonlinear interactions between $v$ and $w$. Here, for the consistency with the quadratic nonlinearity in the starting model (2), the closure terms $\tau_{w}^{(v)} (v) w + \tau_{w}^{(v)} (v)$
and \( \tau^{(w)}_w(v) w + \tau^{(w)}_v(v) \) are assumed to be quadratic, and therefore

\[
\tau^{(v)}_w(v) = v^* C^{(v)}_{w\tau} + D^{(v)}_{w\tau}, \\
\tau^{(v)}_v(v) = v^* B^{(v)}_{v\tau} v + C^{(v)}_{v\tau} v + D^{(v)}_{v\tau}, \\
\tau^{(w)}_w(v) = v^* C^{(w)}_{w\tau} + D^{(w)}_{w\tau}, \\
\tau^{(w)}_v(v) = v^* B^{(w)}_{v\tau} v + C^{(w)}_{v\tau} v + D^{(w)}_{v\tau},
\]

(4)

where \( D^{(v)}_{v\tau} \) and \( D^{(w)}_{w\tau} \) are constant vectors, while \( B^{(v)} \), \( C^{(v)} \), and \( D^{(v)} \) are constant matrices.

It is worthwhile to highlight that many complex nonlinear systems have already fitted into the framework of (3). Some well-known classes of the models are [51, 54]:

- Physics-constrained nonlinear stochastic models. Examples include the noisy versions of Lorenz models, low-order models of Charney-DeVore flows, and a paradigm model for topographic mean flow interaction.

- Stochastically coupled reaction-diffusion models in neuroscience and ecology. Examples include stochastically coupled FitzHugh-Nagumo models and stochastically coupled SIR epidemic models.

- Multi-scale models in turbulence, fluids and geophysical flows. Example include the Boussinesq equations with noise and stochastically forced rotating shallow water equation.

Other applications that share similar motivations or ideas as (3) include predicting the intermittent time-series of the monsoon intraseasonal variabilities [55, 56], recovering the turbulent ocean flows with noisy observations from Lagrangian tracers [57, 58], cheap solvable forecast models in dynamic stochastic superresolution [59, 60], and stochastic superparameterization for geophysical turbulence [8]. These examples further justify that the new ROM framework (3) is appropriate to characterize or approximate many nonlinear and non-Gaussian systems in various disciplines.

2.2. Efficient data assimilation utilizing the new ROM

One of the major goals of building a ROM is to use it as a cheap forecast model. For turbulent systems, the forecast skill relies heavily on the accuracy of the initial values. However, it is often the case that only partial observations are available for complex
systems. These partial observations usually correspond to the large-scale modes as they are the easiest to identify, while the variables that are below certain spatial scales are often hard to approximate. Therefore, it is important to combine the ROM with the partial observations for recovering the unobserved state variables in an accurate fashion, which then can be used as initializations for the forecast [61, 62, 63, 64, 65, 66, 67]. This is known as data assimilation (or filtering) [47, 68, 49, 46, 48].

Ensemble based data assimilation schemes [46, 69], such as the ensemble Kalman filter, have been widely utilized in practice to lead to many successes. However, there are a few intrinsic difficulties in applying these classical data assimilation methods. The main challenge comes from the sampling error as the state estimation is completely determined by the ensembles. To prevent the filter divergence, localization, noise inflation, and many other ad hoc tuning procedures are usually inevitable as effective remedies when using the ensemble based methods [70, 71]. In addition, if a large number of ensembles are used in the system to guarantee the accuracy of the data assimilation results, then the computational cost can significantly increase.

Different from the traditional data assimilation methods, the new ROM in (3) allows an efficient and accurate data assimilation scheme. The solution of the data assimilation result is given by closed analytic formulae and therefore it avoids all the ad hoc tuning procedures. In fact, the ROM in (3) can be written into a more abstract form,

\[
\frac{dv(t)}{dt} = \left[A_0(v, t) + A_1(v, t)w(t)\right] + B_1(v, t)\dot{W}_1(t), \quad (5a)
\]

\[
\frac{dw(t)}{dt} = \left[a_0(v, t) + a_1(v, t)w(t)\right] + b_2(v, t)\dot{W}_2(t), \quad (5b)
\]

where \(A_0, a_0, A_1, a_1, B_1, \) and \(b_2\) are matrices and vectors that can depend nonlinearly on the state variables \(v\). Despite being highly nonlinear and non-Gaussian, the process of \(w\) in (5) is linear by design if a realization of the time series \(v(t)\) for \(t \in [0, t]\) is given. Due to such a conditional linear property, the conditional distribution of \(w(t)\), given a realization of \(v(s)\) for \(s \in [0, t]\), is Gaussian,

\[
p(w(t)|v(s), s \leq t) = \mathcal{N}(\mu(t), \mathbf{R}(t)). \quad (6)
\]

One desirable feature of the ROM framework (5) is that the conditional mean \(\mu\) and
the conditional covariance $R$ can be solved via the following closed analytic formulae

\[ \begin{align*}
\frac{d\mu}{dt} &= (a_0 + a_1\mu) + (RA_1^*)(B_1B_1^*)^{-1}\left(\frac{dv}{dt} - (A_0 + A_1\mu)\right), \\
\frac{dR}{dt} &= a_1R + Ra_1^* + b_2b_2^* - (RA_1^*)(B_1B_1^*)^{-1}(A_1R).
\end{align*} \]

(7a) (7b)

The explicit formulae in (6)–(7) correspond to the optimal nonlinear filter (namely the online data assimilation) solution of the state variable $w(t)$ given a realization of the observed time series $v(s)$ for $s \in [0, t]$. Thus, $\mu$ and $R$ are also known as the filter posterior mean and the filter posterior covariance, respectively. The closed analytic formulae for the filtering solution allow an efficient and accurate online state estimation of the unobserved state $w$ from the observed time series of $v$ and the model structure of the ROM. The real-time recovery of the unobserved state provides the initial condition of $w$, which together with the observed initial value of $v$, advances the forecast utilizing the ROM. Note that the classical Kalman-Bucy filter [73] is the simplest special example of (7).

Due to the solvable conditional Gaussian statistics in (6)–(7), the new ROM is named as the conditional Gaussian ROM (CG-ROM).

2.3. Data-driven model calibration

What remains is to determine the matrices and vectors in the ROM (3), which correspond to the model parameters. As was discussed above, partial observation means only a time series of the observed large-scale variable $v$ is available in practice. Yet, it is assumed that a time series of both $v$ and $w$ is accessible in the model calibration stage to simplify the procedure of determining the model parameters. This can be explained as using the reanalysis data in many practical applications, where the historical trajectories of the unobserved data has been inferred via certain Bayesian approaches [74]. Otherwise, an iteration process involving an alternation between state estimation and parameter estimation can be adopted for the model calibration. The basic algorithm of parameter estimation remains the same but the entire process is more complicated, which will diverge from the ROM development here, and is therefore treated as future work (see the discussion in Section 4). Nevertheless, only the value of the observed
variable $v$ is available in the forecast stage such that data assimilation becomes essential in recovering the initial value of $w$.

The parameters in the ROM (3) can be divided into three sets:

(A). The parameters in the basic dynamics: $A_v^{(v)}$, $B_{vv}^{(v)}$, $B_{vw}^{(v)}$, $A_w^{(w)}$, $B_{vw}^{(w)}$, $B_{ww}^{(w)}$ in (3).

(B). The parameters in the deterministic part of the closure: $B_v^{(v)}$, $C_v^{(v)}$ and $D_v^{(v)}$ in (4).

(C). The parameters in the stochastic part of the closure: $\sigma_v$ and $\sigma_w$ in (3).

The parameters in Set (A) are inherited from the FOM if the FOM is available. Otherwise, they can be handled in the same way as those in Sets (B) and (C), which are computed using the following optimization algorithm. For the simplicity of discussion, consider the discrete approximation of the original continuous version of the ROM (3) by applying the Euler-Maruyama scheme:

$$v^{j+1} = v^j + \left[ A_v^{(v)} v^j + (v^j)^* B_{vv}^{(v)} v^j + (v^j)^* B_{vw}^{(w)} w^j \right] \Delta t + \left[ \tau_v^{(v)} (v^j) w^j + \tau_v^{(v)} (v^j) \right] \Delta t + \left( \sigma_v \sqrt{\Delta t} \right) \varepsilon_v,$$

(8a)

$$w^{j+1} = w^j + \left[ A_w^{(w)} v^j + (v^j)^* B_{vw}^{(v)} v^j + (v^j)^* B_{ww}^{(w)} w^j \right] \Delta t + \left[ \tau_w^{(w)} (v^j) w^j + \tau_w^{(w)} (v^j) \right] \Delta t + \left( \sigma_w \sqrt{\Delta t} \right) \varepsilon_w,$$

(8b)

where $\varepsilon_v$ and $\varepsilon_w$ are standard independent and identically distributed Gaussian random variables at time $t_j$. Assume the noise coefficients are constants, namely $\sigma_v := \sigma_v$ and $\sigma_w := \sigma_w$ for all $j$. Denote $\Sigma$ a block diagonal matrix with entries $(\sigma_v)^2 \Delta t$ and $(\sigma_w)^2 \Delta t$. Note that in many practical situations $\Sigma$ is a diagonal matrix, from which $\sigma_v$ and $\sigma_w$ can be inferred. Further denote

$$\theta = \begin{bmatrix} \theta_L \\ \theta_S \end{bmatrix}, \quad C = \begin{bmatrix} C_L \\ C_S \end{bmatrix}, \quad M^T \theta = \begin{bmatrix} M_L \theta_L \\ 0 \\ M_S \theta_S \end{bmatrix},$$

(9)
where $\theta_L$ contains all the parameters in $\tau_w^{(v)}, \tau_v^{(v)}$ and $\theta_S$ contains all those in $\tau_w^{(w)}, \tau_v^{(w)}$. Here, the parameters are assumed to appear as multiplicative pre-factors of functions consisting of state variables $M^j$ such that the closure terms can be written in an abstract form $M^j\theta$. The vector $C^j$ in (9) includes all those terms on the right hand side of (3) that do not contain parameters. Then the optimization problem based on a linear regression can be formulated as [75]:

$$
\tilde{L} = \min_{\theta, \Sigma} \left( \sum_{j=1}^{J} \left( J \log |\Sigma| + \frac{1}{2} \sum_j \left( z_{j+1} - M^j\theta - C^j \right)^* R^{-1} \left( z_{j+1} - M^j\theta - C^j \right) \right) \right),
$$

(10)

where $z_{j+1} = [v_{j+1}; w_{j+1}]$ and $J$ is the number of training snapshots. To solve the optimization problem (10), we use the iterative method to compute the following:

$$
\Sigma = \frac{1}{J} \sum_{j=1}^{J} \left( z_{j+1} - M^j\theta - C^j \right) \left( z_{j+1} - M^j\theta - C^j \right)^*,
$$

(11)

$$
\theta = \left( \sum_{j} \left( M^j \right)^* \Sigma^{-1} M^j \right)^{-1} \left( \sum_{j} \left( M^j \right)^* \Sigma^{-1} \left( z_{j+1} - C^j \right) \right).
$$

(12)

2.4. Incorporating physics constraints into the ROM

Physics constraints correspond to the conservation of the energy in the quadratic nonlinear terms in (1), namely, $u \cdot B(u, u) = 0$. Incorporating the physics constraint into the ROM is central in preventing finite-time blow-up of the solutions and facilitates a skillful medium- to long-range forecast [38, 39, 50].

The physics constraints (or in general any constraint) for the model can be written as the following matrix equality:

$$
H\theta = 0,
$$

(13)

where $H$ contains those functions that are involved in the constraints. Then the optimization problem in (10) is reformulated as a constrained optimization problem,

$$
\tilde{L} = \min_{\theta, \Sigma} \left( -\frac{1}{2} \log |\Sigma^{-1}| + \frac{1}{2} \sum_j \left( z_{j+1} - M^j\theta - C^j \right)^* \Sigma^{-1} \left( z_{j+1} - M^j\theta - C^j \right) \right),
$$

(14)

s.t. $H\theta = 0$.
The constrained optimization problem can be solved by the Lagrangian multiplier method \cite{76}, where the Lagrangian function yields

\[
f(\theta, \Sigma; \lambda) = \left( -\frac{J}{2} \log |\Sigma^{-1}| + \frac{1}{2} \sum_{j} (z^{j+1} - M^j \theta - C^j)^\ast \Sigma^{-1} \cdot (z^{j+1} - M^j \theta - C^j) + \lambda^\ast \mathbf{H} \theta. \right) \tag{16}
\]

The solution can be obtained using the iterative method:

\[
\Sigma = \frac{1}{J} \sum_{j=1}^{J} (z^{j+1} - M^j \theta - C^j)^\ast (z^{j+1} - M^j \theta - C^j), \tag{17}
\]

\[
\lambda = (\mathbf{H} \mathcal{K}^{-1} \mathbf{H}^\ast)^{-1} (\mathbf{H} \mathcal{K}^{-1} \mathbf{b}), \tag{18}
\]

\[
\theta = \mathcal{K}^{-1} (\mathbf{b} - \mathbf{H}^\ast \lambda), \tag{19}
\]

where

\[
\mathcal{K} = \sum_{j} (M^j)^\ast \Sigma^{-1} M^j, \tag{20}
\]

\[
b = \sum_{j} (M^j)^\ast \Sigma^{-1} (z^{j+1} - C^j). \tag{21}
\]
In Figure 1, we summarize all the models and corresponding equations used in this paper.

3. Applications to Complex Turbulent Flows

3.1. The viscous stochastic Burgers equation

The first application of the CG-ROM framework is the following viscous stochastic Burgers equation

\[
\frac{\partial u}{\partial t} = \nu u_{xx} + \lambda u - \gamma uu_x + \sigma \dot{W}_t, \tag{22}
\]

which is posed on the spatial domain \(0, L\) supplemented with the Dirichlet boundary conditions \(u(0, t) = u(L, t) = 0\) for \(t \geq 0\) and a suitable initial condition \(u(x, 0) = \ldots\)
$u_0(x)$ for $x \in (0, L)$. In (22), $\nu > 0$ is the viscosity constant, $\lambda$ is the linear drag coefficient, $\gamma$ is the advection strength, $\sigma$ stands for the level of the stochastic noise, and $\dot{W}_t$ is a spatiotemporal white noise source.

The viscous stochastic Burgers equation (22) is very different from the standard deterministic inviscid Burgers equation, which is well-known for the development of discontinuities, namely shock waves. In (22), the external random forcing interacts with the deterministic nonlinearity to induce intermittent bursts of the signals in the form of viscous shocks, which will be dissipated via the viscous term before the appearance of discontinuities. Therefore, the viscous stochastic Burgers equation is a suitable test model that mimics turbulent behavior in nature with intermittency. It is also noticeable that the stochastic forcing is triggering turbulent features.

3.1.1. Numerical solver of the FOM

The viscous stochastic Burgers equation (22) is solved by a semi-implicit Euler-Maruyama scheme. At each time step the nonlinearity $uu_x = \left( u^2 \right)_x / 2$ and the noise term $\sigma \dot{W}_t$ are treated explicitly, where the random noises are drawn independently from a normal distribution $\mathcal{N}(0, 1)$, while the other terms are treated implicitly. The Laplacian operator is discretized using the standard second-order central difference approximation.

The spatial domain considered in the following tests is $x \in (0, 2\pi)$. In the numerical discretization, the total number of grid points is $N_x = 512$, and therefore $\Delta x \approx 1.2 \times 10^{-2}$. The simulations of the viscous stochastic Burgers equation (22) are performed with a numerical integration time step $\Delta t = 10^{-3}$ and are integrated up to $T = 10^4$. The FOM solutions are recorded every $5 \times 10^{-2}$ simulation time units, and the initial condition is given by

$$u(x, 0) = 0.1 \sqrt{\frac{1}{\pi}} \sin\left(\frac{x}{2}\right) + 0.1 \sqrt{\frac{1}{\pi}} \sin(2x).$$

In particular, the stochastic noise is added as

$$\sigma = \sum_{k_x = 1}^{N_k} \hat{\sigma}_i \sqrt{\frac{2}{L}} \sin\left(\frac{k_x \pi x}{L}\right),$$

where $N_k$ is chosen to be 4 and $\hat{\sigma}_i$ is a constant for $i = 1, 2, 3, 4$. 

![](image1)
3.1.2. ROM formulation

To build the standard Galerkin ROM, the simplest way is to adopt the Fourier sine modes as the basis functions due to the Dirichlet boundary conditions:

\[ \varphi_k = \sqrt{\frac{2}{L}} \sin \left( \frac{k\pi x}{L} \right), \quad k = 1, 2, \cdots, R. \]  

(25)

Then, the G-ROM yields the following:

\[ (du, \varphi_k) + \left[ (\nu \partial_x u, \partial_x \varphi_k) - (\lambda u, \varphi_k) + (\gamma uu_x, \varphi_k) \right] dt = (\sigma dW_t, \varphi_k), \]  

(26)

\[ k = 1, 2, \cdots, R. \]

With the sought solution, 

\[ u_r = \sum_{l=1}^{r} a_l(t) \varphi_l(x), \]  

where \( r \) is the dimension of reduced space and \( \{a_l\}_{l=1}^{r} \) are the sought ROM coefficients varying with time, one can replace \( u \) with \( u_r \) in the equation (26) and hence obtain the Galerkin ROM (G-ROM):

\[ da_k = \sum_{l=1}^{r} A_{kl} a_l dt + \sum_{l=1}^{r} \sum_{m=1}^{r} B_{lmk} a_l a_m dt + \sigma_k dW_t, \]  

\[ k = 1, \cdots, r, \]  

(27)

where \( A \) is a matrix and \( B \) is a tensor which are defined as

\[ A_{kl} = -\nu \left( \frac{\partial \varphi_l}{\partial x}, \frac{\partial \varphi_k}{\partial x} \right) + \lambda (\varphi_l, \varphi_k) \]  

(28)

\[ B_{lmk} = -\gamma (\varphi_l, \frac{\partial \varphi_m}{\partial x}, \varphi_k). \]  

(29)

Thanks to the explicit formulation of ROM basis functions (i.e., they are sine functions (25)), we can explicitly calculate the formulas in (28) and (29):

\[ (\varphi_l, \varphi_k) = \delta_{lk}, \quad \left( \frac{\partial \varphi_l}{\partial x}, \frac{\partial \varphi_k}{\partial x} \right) = \frac{k^2 \pi^2}{L^2} \delta_{lk}, \quad l, k = 1, \cdots, r, \]  

(30)

\[ \left( \varphi_l, \frac{\partial \varphi_m}{\partial x}, \varphi_k \right) = \begin{cases} \left( \frac{m \pi}{4}, \frac{2}{L} \right)^{3/2} & \text{if } l + m - k = 0 \text{ or } l - m - k = 0, \\ -\left( \frac{m \pi}{4}, \frac{2}{L} \right)^{3/2} & \text{if } l - m + k = 0, \\ 0 & \text{otherwise}. \end{cases} \]  

(31)

The CG-ROM starts from (27). It drops the quadratic nonlinear interaction between the medium-scale variables but adds closure terms as was described in Section 2.1. Physics constraints are imposed in the development of the CG-ROM, while they are automatically included in the G-ROM.
3.1.3. Dynamical regimes

Two different regimes are considered in the following tests, with the associated parameters listed in Table 1. The two dynamical regimes use a different linear drag coefficient, $\lambda$. A larger (i.e., more positive value of) $\lambda$ is more prone to induce linear instability and, thus, yields more turbulent features. Linear analysis shows that only one eigenvalue is positive in Regime I, while three positive eigenvalues are observed in Regime II. Therefore, Regime I is a weakly chaotic regime, while Regime II is a much more challenging regime with strong instabilities.

| Parameters | $\nu$  | $\lambda$ | $\gamma$ | $\hat{\sigma}$ |
|------------|--------|-----------|----------|----------------|
| Regime I   | 0.005  | 0.00375   | 1        | 0.003          |
| Regime II  | 0.005  | 0.01125   | 1        | 0.003          |

Table 1: Parameters of the two dynamical regimes of the viscous stochastic Burgers equation (22).

3.1.4. Results: Comparison of the FOM and ROMs in model simulations and statistics

To present numerical results of ROMs, we mainly consider the ROMs with $r = 5$ modes, which account for 96.05% of the FOM energy for Regime I, and 93.81% of the FOM energy for Regime II, where the FOM system has in total 512 modes. In the CG-ROM, the dimension of the observed variable $v$ is $r_1 = 2$ and therefore the dimension of the unobserved medium-scale variable $w$ is $r_2 = 3$, such that the total number of the modes in the CG-ROM is the same as that in the G-ROM. For the simplicity of the notation, these leading five modes are denoted by $a_1, \ldots, a_5$.

Figure 2 compares the trajectories, PDFs, and autocorrelation functions (ACFs) of the FOM (black), G-ROM (red), and CG-ROM (blue) for Regime I. In order to contrast the path-wise behavior, the same random noise source is utilized in the three models. Yet, it is important to notice that in practice the noise is unknown and therefore the statistics, namely the PDFs and the ACFs, are more suitable metrics for model comparison. The G-ROM is nearly as accurate as the CG-ROM for the first two modes in terms of both the simulated trajectories and the statistics. However, due to the lack of closure terms, the G-ROM loses its skill in capturing the statistics of modes $a_3$, $a_4$, and $a_5$. The amplitudes of the corresponding trajectories are also overestimated. Table
includes the $L^2$ errors for different $r_1$ and $r$ values in the ROMs. The results in Table 2 consistently show that the CG-ROM outperforms the G-ROM for both the large-scale modes $v$ and the medium-scale modes $w$, indicating the robustness of the CG-ROM performance.

Figure 3 compares the trajectories and the statistics of the FOM (black) and the CG-ROM (blue) in Regime II, which is a much tougher regime for the test. Despite being strongly turbulent with significant non-Gaussian statistics in all the leading $r = 5$ modes, the CG-ROM succeeds in reproducing both the trajectories (by assigning the same random forcing as the FOM) and the overall non-Gaussian PDFs. Note that the solution of the G-ROM suffers from strong instability and blows up in a finite time in such a test case. This further indicates the necessity of incorporating the closure terms in the development of the ROMs. Figure 4 shows the reconstructed spatiotemporal patterns. The CG-ROM is able to reproduce the intermittent spatiotemporal patterns of the viscous shocks, which highly resemble those from the FOM with a pattern correlation that is above 0.95.

| $r_1$, $r$ | ROMs       | $E (v; L^2)$ | $E (w; L^2)$ | $E (u; L^2)$ |
|------------|------------|--------------|--------------|--------------|
| $r_1 = 2$, $r = 4$ | G-ROM ($r$ dim) | 5.818e-2     | 6.476e-2     | 8.705e-2     |
|            | CG-ROM     | 2.051e-2     | 1.924e-2     | 2.812e-2     |
| $r_1 = 2$, $r = 5$ | G-ROM ($r$ dim) | 5.109e-2     | 5.919e-2     | 7.819e-2     |
|            | CG-ROM     | 2.251e-2     | 2.086e-2     | 3.069e-2     |
| $r_1 = 3$, $r = 6$ | G-ROM ($r$ dim) | 2.193e-2     | 3.302e-2     | 3.964e-2     |
|            | CG-ROM     | 2.957e-2     | 1.668e-2     | 3.395e-2     |

Table 2: Regime I of the viscous stochastic Burgers equation. The $L^2$ errors for different $r_1$ and $r$ values in the ROMs. The same random noise source is utilized in the FOM and the two ROMs.

3.1.5. Results: Comparison of the FOM and ROMs in data assimilation

The focus of this subsection is on the data assimilation (i.e., filtering) of the unobserved variable $w = (a_3, a_4, a_5)^T$ using ROMs given one realization of the observed trajectories of the leading two modes $v = (a_1, a_2)^T$ generated from the FOM. As in
Figure 2: Regime I of the viscous stochastic Burgers equation. Comparison of the trajectories, PDFs, and ACFs of the FOM (black), G-ROM (red), and CG-ROM (blue) for $r = 5$ and $r_1 = 2$. The same random noise source is utilized in the three models. The PDFs and the ACFs are computed based on a finite trajectory length of $5 \times 10^4$ units.

In the previous subsection, the two ROMs utilized here are the G-ROM and the CG-ROM. Recall that the closed analytic formulae (7) associated with the CG-ROM provide an efficient and accurate solution of the filter posterior distribution. However, no such a closed formula can be applied to the G-ROM. Therefore, the ensemble Kalman-Bucy filter (EnKBF) [77], which is a widely used ensemble based data assimilation scheme for continuous-in-time observations, is utilized for the G-ROM. As will be seen below, the solutions using G-ROM with the EnKBF are much less accurate than those using the CG-ROM with closed analytic filter formulae. To understand if the data assimilation scheme or the ROM itself is the main source accounting for the poor behavior of the former, the EnKBF is also applied to the CG-ROM. In other words, three filters will be tested in our numerical investigation:

- **CG-ROM Closed Form**: using the closed analytic formulae (7) as the data assimilation scheme for the CG-ROM,
Figure 3: Regime II of the viscous stochastic Burgers equation. Comparison of the trajectories, PDFs, and ACFs of the FOM (black) and the CG-ROM (blue) for $r = 5$ and $r_1 = 2$. The same random noise source is utilized in the two models. The PDFs and the ACFs are computed based on a finite trajectory length of $5 \times 10^4$ units. The path-wise solution of the G-ROM blows up around time unit 100 and is therefore omitted here.

- G-ROM EnKBF: using the EnKBF as the data assimilation scheme for the G-ROM, and
- CG-ROM EnKBF: using the EnKBF as the data assimilation scheme for the CG-ROM.

For the EnKBF, the size of ensemble is taken to be 100, which has been tested to be sufficient to reach reasonably accurate and robust results. Note that although the G-ROM or CG-ROM are often relatively low dimensional, running such a moderate ensemble size forward through the system via EnKBF is still much more computationally expensive than the closed form in the CG-ROM.

Figure 5 shows the posterior mean time series of $w = (a_3, a_4, a_5)^T$ from data assimilation using the three different filters in Regime I. Consistent with the model simulation results presented in Figure 2, the G-ROM EnKBF is significantly worse than the
Regime II of the viscous stochastic Burgers equation. Reconstructed spatiotemporal patterns of the FOM and CG-ROM.

Figure 4: Regime II of the viscous stochastic Burgers equation. Reconstructed spatiotemporal patterns of the FOM and CG-ROM.

CG-ROM Closed Form, especially for $a_5$. As the results of CG-ROM EnKBF are only slightly worse than those of CG-ROM Closed Form in such a case, it is conclusive that the main error source for the unskillful behavior in the G-ROM EnKBF is the model error in the G-ROM, not the filtering method EnKBF. In fact, the bare truncation in the G-ROM effectively breaks the coupling between $v$ and $w$ to a large extent. Therefore, it is more difficult to infer the unobserved states from the observations via a model with a poorly described coupling relationship. Similar qualitative conclusions are reached in Figure 6 that illustrates the results of Regime II. However, the G-ROM EnKBF is quantitatively much worse than its counterparts due to the stronger turbulent behavior in this regime. Note that, despite the big error, the data assimilation results using the G-ROM EnKBF do not suffer from the finite-time blow-up issue as in the forward run of the G-ROM. The reason is that in data assimilation observations are an additional input, which constantly helps mitigate the pathological behavior of the model. Finally, Figure 7 displays a spatiotemporal reconstruction of the three unobserved modes $w = (a_3, a_4, a_5)^T$, using the recovered results from data assimilation, which is compared with the truth. It further validates that recovered fields from the CG-ROM Closed
Form and from the CG-ROM EnKBF are comparably accurate, while that from the G-ROM is completely biased.

Figure 5: Regime I of the viscous stochastic Burgers equation. Truth and posterior mean time series of $w = (a_3, a_4, a_5)^T$ from data assimilation using the three different filters.
Figure 6: Regime II of the viscous stochastic Burgers equation. Truth and posterior mean time series of $w = (a_3, a_4, a_5)^T$ from data assimilation using the three different filters.

Figure 7: Regime II of the viscous stochastic Burgers equation. Reconstructed spatiotemporal field based on the three modes $w = (a_3, a_4, a_5)^T$ from data assimilation and the truth, where the truth is computed by projecting the true signal on the basis functions associated with these three modes.
3.2. The QG model

The second application is the quasi-geostrophic (QG) equation:

\[
\frac{\partial \omega}{\partial t} + J(\omega, \psi) - Ro^{-1} \frac{\partial \psi}{\partial x} = Re^{-1} \Delta \omega + Ro^{-1} F,
\]

\[
\omega = -\Delta \psi,
\]

which is a widely used model that describes the large scale ocean circulation \[1, 9\]. In the QG equation (32), \(\omega\) is the vorticity, \(\psi\) is the streamfunction, \(F\) is the external forcing, and the Jacobian is defined as \(J(A, B) = \partial A/\partial x \partial B/\partial y - \partial A/\partial y \partial B/\partial x\).

There are two non-dimensional numbers in (32): \(Re\) is the Reynolds number and \(Ro\) is the Rossby number. The former is the ratio of inertial forces to viscous forces, representing the strength of turbulence, while the latter stands for the ratio of inertial force to Coriolis force. Note that, different from the viscous stochastic Burgers equation, the QG equation is a deterministic system and its turbulent features are induced completely by the nonlinear interactions between the state variables at different spatial scales.

In the following, a symmetric double-gyre wind forcing is imposed to the QG equation \[78, 79, 80, 81\] with \(F = \sin(\pi(y - 1))\). A rectangular domain is considered here with \(\Omega = [0, 1] \times [0, 2]\). Homogeneous Dirichlet boundary conditions are applied for both \(\psi\) and \(\omega\):

\[
\psi(t, x, y) = 0, \quad \omega(t, x, y) = 0 \quad \text{for} \quad (x, y) \in \partial \Omega \quad \text{and} \quad t \geq 0.
\]

The two non-dimensional numbers are set to be \(Re = 450\) and \(Ro = 0.0036\), such that the system has moderate turbulent behavior and a fast rotation.

3.2.1. Numerical solver of the FOM

A spectral method with a 256 × 512 spatial resolution and an explicit Runge-Kutta method are utilized to solve the QG model. The FOM is simulated for 80 time units. The solution displays a transient behavior on the time interval \([0, 10]\), and then converges to a statistically steady state on the time interval \([10, 80]\). The FOM solutions are recorded on the time interval \([10, 80]\) every \(10^{-2}\) simulation time units, which ensures that the snapshots used in the construction of the ROM basis are equally spaced. In Figure 8, the vorticity and streamfunction snapshots at \(t = 20, 30, 40, 50, 60\) are plotted.
3.2.2. ROM formulation

The ROMs start from projecting the vorticity $\omega$ to its POD bases and keeping the leading $r$ modes. Correspondingly, a reduced set of $r$ basis functions are also introduced for $\psi$, which are subordinate to the above POD basis functions for $\omega$ via $\phi_i(x, y) = -\Delta^{-1}\varphi_i(x, y)$, i.e., they solve the following Poisson equation:

$$-\Delta \phi_i(x, y) = \varphi_i(x, y), \quad \text{subject to } \phi_i(x, y) = 0,$$

for $(x, y) \in \partial \Omega, \quad i = 1, 2, \cdots, r. \quad (34)$

Note that while the POD basis $\{\varphi_i\}$ for the vorticity $\omega$ is an orthonormal basis under the $L^2$ inner product, the basis $\{\phi_i\}$ for the streamfunction $\psi$ is not orthogonal. Given the G-ROM approximation $\omega_r = \sum_{i=1}^{r} a_i(t) \phi_i(x, y)$ of $\omega$, the corresponding $\psi$ is approximated by $\psi_r = \sum_{i=1}^{r} a_i(t) \phi_i(x, y)$, which results from the ansatz $\psi_r = -\Delta^{-1}\omega_r$ and
definition of the basis function \( \phi_i \). With the above notations, the \( r \)-dimensional G-ROM for the problem (32)–(33) is given by:

\[
\left( \frac{\partial \omega_r}{\partial t}, \phi_i \right) + (J(\omega_r, \psi_r), \phi_i) - R_o^{-1} \left( \frac{\partial \psi_r}{\partial x}, \phi_i \right) + R_e^{-1} (\nabla \omega_r, \nabla \psi_r) = R_o^{-1} (F, \phi_i),
\]

(35)

where \((\cdot, \cdot)\) denotes the \( L^2 \) inner product over the spatial domain, and \( i = 1, \cdots, r \).

The component-wise form of the G-ROM is as follows: for \( i = 1, 2, \cdots, r \)

\[
\dot{a}_i(t) = b_i + \sum_{m=1}^{r} A_{im} a_m(t) + \sum_{m=1}^{r} \sum_{n=1}^{r} B_{imn} a_m(t) a_n(t),
\]

(36)

where

\[
b_i = R_o^{-1} (F, \phi_i),
\]

\[
A_{im} = R_o^{-1} \left( \frac{\partial \phi_m}{\partial x}, \phi_i \right) - R_e^{-1} (\nabla \phi_m, \nabla \phi_i),
\]

\[
B_{imn} = - (J(\phi_m, \phi_n), \phi_i).
\]

In the above procedure, 3500 equally spaced FOM vorticity snapshots in the time interval \([10, 45]\) are utilized to construct the ROM bases \[79\]. For computational efficiency, the FOM vorticity is interpolated onto a uniform mesh with the resolution \( 257 \times 513 \) over the spatial domain \( \Omega = [0, 1] \times [0, 2] \), i.e., with a mesh size \( \Delta x = \Delta y = 1/256 \).

Based on the interpolated snapshots, the corresponding eigenvalue problem is solved to generate the ROM basis. The fourth-order Runge-Kutta scheme (RK4) is utilized for the temporal discretization. To ensure the numerical stability of the time discretization, a time step size \( \Delta t = 0.001 \) is utilized. The ROM data is stored every ten time steps to match the FOM sampling rate.

On the other hand, the CG-ROM follows the procedure described in Section 2.1 It starts from (36) and then drops the quadratic nonlinear interaction between the medium-scale variables but adds a closure term. Physics constraints are also included in the CG-ROM.

Figure 9 shows the time series and the associated statistics of the leading 6 POD modes of the FOM. Fast oscillations and turbulent behavior can be observed in these time series. In addition, the intermittency introduces extreme events, which lead to non-Gaussian PDFs with fat tails. The shading period up to \( t = 45 \) is utilized to calibrate
the ROMs, while the remaining time series is utilized to test for the data assimilation skill.

In the following, \( r = 20 \) modes are kept in both the G-ROM and the CG-ROM. For notation simplicity, these 20 modes are denoted by identified with their time-dependent coefficients, \( a_1, \ldots, a_{20} \). Assume that the first \( r_1 = 10 \) modes are accurately resolved in the observations. Therefore, \( \mathbf{v} = (a_1, \ldots, a_{10}) \) and \( \mathbf{w} = (a_{11}, \ldots, a_{20}) \).

\[ \text{QG} \]

![Figure 9: The time series and the associated PDFs and ACFs of the leading 6 POD modes of the QG equation. The shading period up to \( t = 45 \) is utilized to calibrate the ROMs, while the remaining time series is utilized to test for the data assimilation skill.}

**3.2.3. Results: Comparison of the FOM and the ROMs in model statistics**

For the viscous stochastic Burgers equation, the random noise source can be artificially assumed to be known, which allows a path-wise comparison between the ROMs and the FOM. In contrast, the turbulent nature of the QG equation makes the trajectories of different models diverge very quickly. Therefore, in this case it is more justified to compare the statistics of these models.

Figure 10 shows the comparison of the PDFs using the FOM and the two ROMs.
for different POD modes. For the first \( r_1 = 10 \) modes, the CG-ROM captures the true statistics quite accurately. Particularly, the non-Gaussian behavior in the PDFs is well characterized by the CG-ROM. In contrast, the resulting PDFs from the G-ROM are more biased. For \( a_2 \) and \( a_3 \), there is a severe mean shift in the result from G-ROM. For \( a_7, a_8, a_9 \) and \( a_{10} \), the variance of the G-ROM is significantly overestimated. For the remaining 10 medium-scale modes, neither the G-ROM nor the CG-ROM is very accurate. Nevertheless, the CG-ROM is still more skillful than the G-ROM in recovering the statistics.

![PDFs using the FOM and the two ROMs for different POD modes.](image)

**Figure 10:** Comparison of the PDFs using the FOM and the two ROMs for different POD modes.

### 3.2.4. Results: Comparison of the FOM and the ROMs in data assimilation

As in Section 3.1.5, the three filters — CG-ROM Closed Form, G-ROM EnKBF and CG-ROM EnKBF — are utilized for assimilating the trajectories of the 10 unobserved state variables \( a_{11}, \ldots, a_{20} \). Figures 11 and 12 include the comparison of the assimilated posterior mean time series and the associated statistics using different filters. Despite that the statistics related to the model free run using the CG-ROM contain

---

29
certain bias for all these variables, the data assimilation solutions using the CG-ROM remain reasonably accurate. This is not surprising since the observations help correct the model error in the ROM simulations. In contrast to the filters based on the CG-ROM, the G-ROM EnKBF is significantly worse.

For the sake of a qualitative comparison between the filtered posterior mean time series and the truth using different filters, three skill scores are utilized here: the root-mean-square error (RMSE), the pattern correlation (Corr), and the relative entropy. They are defined as follows:

\[
\text{Corr} = \frac{\sum_{i=1}^{n} (u^M_i - \bar{u}^M)(u^{ref}_i - \bar{u}^{ref})}{\sqrt{\sum_{i=1}^{n} (u^M_i - \bar{u}^M)^2} \sqrt{\sum_{i=1}^{n} (u^{ref}_i - \bar{u}^{ref})^2}}.
\]

\[
\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (u^M_i - u^{ref}_i)^2}{n}}.
\]

Relative Entropy = \( \mathcal{P}(p^{ref}(u), p^M(u)) = \int p^{ref}(u) \ln \frac{p^{ref}(u)}{p^M(u)} du, \)

where \( u^M_i \) and \( u^{ref}_i \) are the assimilated solution and the truth, respectively, at time \( t = t_i \). The time averages of the assimilated and the true time series are denoted by \( \bar{u}^M \) and \( \bar{u}^{ref} \), respectively. The RMSE and the Corr are widely used metrics in practice to quantify the path-wise error. A smaller RMSE and a larger Corr correspond to a skillful assimilated time series. On the other hand, the relative entropy is an information criterion \[82, 83, 84, 85\], which is adopted to quantify the statistics error between the two distributions formulated by the assimilated time series and the true signal, respectively. The relative entropy has many attractive features. First, \( \mathcal{P}(p, p^M) \geq 0 \) with equality if and only if \( p = p^M \). Second, \( \mathcal{P}(p, p^M) \) is invariant under general nonlinear changes of variables. A smaller relative entropy value corresponds to a smaller statistical error. Figure [13] shows the skill scores using different filters. First, the larger RMSE and the smaller Corr in the G-ROM EnKBF confirm the intuition from Figures [11] and [12] that the path-wise error using the G-ROM EnKBF is much larger than the filters using the CG-ROMs. Second, the relative entropy, which quantifies the error in the PDFs formed by the posterior mean time series and the true signal, indicates that the G-ROM has a bigger statistical error for all the modes as well. It is also worthwhile to mention that the difference between CG-ROM Closed Form and CG-ROM EnKBF implies that
closed analytic formulae in the CG-ROM is advantageous in not only decreasing the computational time but reducing the sampling error as well.

Figure 14 shows the reconstructed spatiotemporal patterns from the posterior mean based on modes $a_{11}, \ldots, a_{20}$. It indicates that the CG-ROM Closed Form can overall recover the streamlines (and thus velocity fields) and the vortices in an accurate fashion. The CG-ROM EnKBF can capture the overall tendency but the sampling error leads to some biases in the amplitude. In contrast, the G-ROM EnKBF leads to a much larger error in reproducing the spatiotemporal patterns of the truth.

![Figure 11: The data assimilation results of the POD modes $a_{11}$ to $a_{15}$ in the QG equation. Comparison of the posterior mean time series and the associated statistics using different filters with the truth.](image-url)
Figure 12: Similar to Figure 11 but for the POD modes $a_{16}$ to $a_{20}$.

Figure 13: The skill scores of the posterior mean time series. The three panels show the root-mean-square error (RMSE), the pattern correlation (Corr), and the relative entropy.
Figure 14: The reconstructed spatiotemporal patterns of $\omega$ (first row) and streamfunction $\psi$ (second row) from the posterior mean based on modes $a_{11}, \ldots, a_{20}$, compared with the truth.
4. Discussion and Conclusions

In this paper, a new multiscale stochastic ROM framework, known as the CG-ROM, is developed. The new ROM focuses on recovering the multiscale dynamical features as well as the statistical feedback between different scales. The CG-ROM exploits cheap but effective conditional linear functions as the closure terms, which facilitates an efficient and accurate data assimilation scheme that is crucial in recovering the unobserved states in turbulent systems. Physics constraints are incorporated into the CG-ROM to prevent the occurrence of pathological behavior of the model.

There are a few issues that remain as future work. First, both the trajectories of $v$ and $w$ are assumed known in the training period for model calibration. Admittedly, the availability of the time series of the unobserved variable $w$ can be explained as the result of the reanalysis. However, it is more natural to assume that only the data of the observed variable $v$ is accessible in practice. Then determining the model parameters is more challenging than the case with full observations in (10). Estimating the model parameters and recovering the unobserved states have to be carried out simultaneously. Fortunately, due to the analytic formulae of the state estimation, such a procedure can still be efficiently using an expectation-maximization algorithm [75]. It is then interesting to see how such an uncertainty affects the skill of the ROM. Second, the ROMs are often preferred to be parsimonious, which prevents the overfitting issue and can yield more physical results. Such a constraint has not been included in this work, as the focus here is on reproducing the statistics and data assimilation for which the current setup is sufficient. Nevertheless, existing techniques, such as the LASSO (least absolute shrinkage and selection operator) regression [80] or information-theoretic based sparse identification methods [87][88], can be incorporated into the current framework. Lastly, as in other ROMs, there is often no rigorous way to determine the dimension of the ROMs. The data assimilation framework may have the potential to provide a different criterion to determine the optimal number of the state variables in the ROMs in addition to using the explained variance as in most of the existing work.
Acknowledgments

The research of N.C. is partially funded by the Office of VCRGE at UW-Madison and ONR N00014-21-1-2904. The research of T.I. is partially funded by NSF through grant DMS-2012253 and CDS&E-MSS-1953113.

Appendix A. Burgers equation

Appendix A.1. Eigenvalue Problems

Consider the linear part of the viscous stochastic Burgers equation (22), which yields the following initial boundary value problem (IBVP):

\[ u_t = \nu u_{xx} + \lambda u, \quad (x,t) \in (0,1) \times \mathbb{R}^+; \quad t \geq 0, \quad (A.1) \]
\[ u(0,t) = u(L,t) = 0, \quad t \geq 0, \quad (A.2) \]
\[ u(x,0) = u_0(x), \quad x \in (0,L). \quad (A.3) \]

This IBVP yields the following eigenvalue problem:

\[ -\nu \frac{\partial^2 w_j}{\partial x^2} - \lambda w_j = \sigma_j w_j, \quad x \in \Omega, \quad (A.4) \]
\[ w_j = 0, \quad x \in \partial \Omega, \quad (A.5) \]

where \( \{w_j\}_{j=1}^\infty \) are eigenfunctions for the operator \( \mathcal{L} = -\nu \frac{\partial^2}{\partial x^2} - \lambda \) on \( \Omega = (0,1) \) with zero boundary conditions. Note that for \( (A.4)-(A.5), (\lambda + \sigma_j)/\nu > 0 \) must be true for all \( j \). Letting \( \alpha_j = \frac{\lambda + \sigma_j}{\nu} \), \( (A.4)-(A.5) \) become

\[ \frac{\partial^2 w_j}{\partial x^2} + \alpha_j w_j = 0, \quad x \in \Omega, \quad (A.6) \]
\[ w_j = 0, \quad x \in \partial \Omega. \quad (A.7) \]

Since \( w_j(x) \) is a nontrivial solution, \( \alpha_j > 0 \) and the boundary condition imply

\[ \alpha_j = \left( \frac{j\pi}{l} \right)^2, \quad w_j(x) = c_j \sin \left( \frac{j\pi x}{l} \right), \quad j \in \mathbb{N}^+, \quad (A.8) \]

where \( c_j \) are arbitrary constants. We choose \( \{w_j\}_{j=1}^\infty \) to be orthonormal in \( L^2(\Omega) \), i.e.,

\[ \int_{\Omega} w_i(x) w_j(x) dx = \delta_{ij}, \quad i, j = 1, 2, \ldots \quad (A.9) \]
Consequently, the eigenfunctions in (A.8) can be written as follows:

\[ w_j(x) = \sqrt{\frac{2}{l}} \sin \left( \frac{j\pi x}{l} \right), \quad j = 1, 2, \ldots \]  

(A.10)

**Appendix A.2. Unstable Eigen-modes**

To apply the Galerkin projection, we define the \( L^2 \) inner product, i.e., \( (f, g)_{L^2} = \int_{\Omega} f(x)g(x)dx \). Then, with a finite-dimensional approximation space spanned by \( \{w_k\}_{k=1}^N \), the Galerkin projection of the viscous stochastic Burgers equation (22) yields

\[ (u_t, w_k)_{L^2} + \nu (\partial_x u, \partial_x w_k)_{L^2} - \lambda (u, w_k)_{L^2} + \gamma (u \partial_x u, w_k)_{L^2} = 0, \quad k = 1, 2, \ldots, N. \]  

(A.11)

Writing \( u(x,t) = \sum_{j=1}^{N} d_j(t)w_j(x) \), we have

\[ \sum_{j=1}^{N} \dot{d}_j(w_j, w_k)_{L^2} + \nu \sum_{j=1}^{N} d_j(\partial_x w_j, \partial_x w_k)_{L^2} - \lambda \sum_{j=1}^{N} d_j(w_j, w_k)_{L^2} + \gamma \sum_{i=1}^{N} \sum_{j=1}^{N} d_i d_j(w_i \partial_x w_j, w_k)_{L^2} = 0, \quad k = 1, 2, \ldots, N. \]  

(A.12)

With the inner products \( (w_i \partial_x w_j, w_k)_{L^2}, (w_j, w_k)_{L^2}, \) and \( (\partial_x w_j, \partial_x w_k)_{L^2} \) evaluated in equations (30) and (31), respectively, the Galerkin projection (A.12) (i.e., the G-ROM (26) in Section 3.1.2) can be reformulated as a set of ODE:

\[ \dot{d}_1 = \left( \lambda - \nu \frac{\pi^2}{l^2} \right) d_1 + h.o.t. \]  

(A.13)

\[ \dot{d}_2 = \left( \lambda - \nu \frac{4\pi^2}{l^2} \right) d_2 + h.o.t. \]  

(A.14)

\[ \dot{d}_3 = \left( \lambda - \nu \frac{9\pi^2}{l^2} \right) d_3 + h.o.t. \]  

(A.15)

\[ \vdots \]

\[ \dot{d}_n = \left( \lambda - \nu \frac{n^2\pi^2}{l^2} \right) d_n + h.o.t., \]  

(A.16)

where \( h.o.t. \) denotes the higher order terms and \( 1 \leq n \leq N \). As a result, the choice of the parameter pair, \( \lambda \) and \( \nu \), can determine the linear stability of the viscous stochastic
Burgers equation. For example, the Regime I in Table 1 yields one unstable eigen-mode, while the Regime II yields three unstable eigen-modes. Thus, the Regime II is more challenging than Regime I.

Appendix B. Optimization Problem

Appendix B.1. Conditional Gaussian ROM with Physical Constraints

In what follows, the details of the multiscale physics constraints are presented. In particular, in Section Appendix B.2, the mathematical formulation of the multiscale physics constraints are derived and the matrix forms for implementation are presented. In Section Appendix B.2.1 a simple example of multiscale physics constraints is given.

Appendix B.2. Multiscale physical constraints

The proposed conditional Gaussian ROMs aim at preserving the multiscale dynamical features of the underlying system. Therefore, multiscale physics constraints, are considered. In particular, the multiscale physics constraints that are imposed for the quadratic terms are expected to satisfy the following:

\[ \left[ v; w \right]^* \left( v^* \left( B_{vv} \oplus C_{ww} \right) [v; w] \right) = 0. \] (B.1)

This constraint enforces the conservation of energy with respect to both the large and the small scales. The discrete formulation of (B.1) yields the following:

\[
\begin{aligned}
\sum_{(i,j,k)=(1,1,1)}^{(r_1,r_1,r_1)} (B_{vv})_{ij} v_i v_j v_k &+ \sum_{(i,j,k)=(1,1,1)}^{(r_1,r_2,r_1)} (C_{ww})_{ij} v_i w_j v_k \\
+ \sum_{(i,j,k)=(1,1,1)}^{(r_1,r_1,r_2)} (B_{wv})_{ij} v_i v_j w_k &+ \sum_{(i,j,k)=(1,1,1)}^{(r_1,r_2,r_2)} (C_{ww})_{ij} v_i w_j w_k = 0.
\end{aligned}
\] (B.2)

Taking into account the symmetries of the nonlinear terms in equation (B.2), the second and third terms on the LHS of (B.2) should share their energy conservation:

\[
\begin{aligned}
\sum_{(i,j,k)=(1,1,1)}^{(r_1,r_2,r_1)} (C_{ww})_{ij} v_i w_j v_k &+ \sum_{(i,j,k)=(1,1,1)}^{(r_1,r_1,r_2)} (B_{wv})_{ij} v_i v_j w_k = 0. \quad (B.3)
\end{aligned}
\]
On the other hand, the other two terms should enforce their own energy conservation:

\[
\sum_{(i,j,k)=(1,1,1)}^{(r_1,r_2)} B_{v\tau}^{(v)}(ijk) v_i v_j v_k = 0 \quad \text{and} \quad \sum_{(i,j,k)=(1,1,1)}^{(r_1,r_2)} C_{w\tau}^{(w)}(ijk) v_i w_j w_k = 0.
\]  
(B.4)

In particular, these terms yield the following discrete relations:

**\(B_{v\tau}^{(v)}\) and \(B_{v\tau}^{(w)}\):**

\[
\left(C_{w\tau}^{(v)} \right)_{(iki)} + \left(B_{v\tau}^{(w)} \right)_{(ikk)} = 0, \quad i = 1, \ldots, r_1, k = 1, \ldots, r_2; \quad (B.5)
\]

\[
\left(C_{w\tau}^{(v)} \right)_{(ikj)} + \left(C_{w\tau}^{(v)} \right)_{(jki)} + \left(B_{v\tau}^{(w)} \right)_{(ijk)} + \left(B_{v\tau}^{(w)} \right)_{(jik)} = 0, \quad i, j = 1, \ldots, r_1, i \neq j; k = 1, \ldots, r_2, \quad (B.6)
\]

**\(B_{v\tau}^{(v)}\):**

\[
\left(B_{v\tau}^{(v)} \right)_{(iii)} = 0, \quad i = 1, \ldots, r_1. \quad (B.7)
\]

\[
\left(B_{v\tau}^{(v)} \right)_{(ijj)} + \left(B_{v\tau}^{(v)} \right)_{(jjj)} + \left(B_{v\tau}^{(v)} \right)_{(iji)} = 0, \quad i, j = 1, \ldots, r_1, i \neq j. \quad (B.8)
\]

\[
\left(B_{v\tau}^{(v)} \right)_{(ikj)} + \left(B_{v\tau}^{(v)} \right)_{(jki)} + \left(B_{v\tau}^{(v)} \right)_{(jik)} = 0, \quad i, j, k = 1, \ldots, r_1, i \neq j \neq k. \quad (B.9)
\]

**\(C_{w\tau}^{(w)}\):**

\[
\left(C_{w\tau}^{(w)} \right)_{(ijj)} = 0, \quad i = 1, \ldots, r_1; j = 1, \ldots, r_2, \quad (B.10)
\]

\[
\left(C_{w\tau}^{(w)} \right)_{(ijk)} + \left(C_{w\tau}^{(w)} \right)_{(ikj)} = 0, \quad i = 1, \ldots, r_1; j, k = 1, \ldots, r_2, j \neq k. \quad (B.11)
\]

To further construct the constraint coefficient matrix \(H\) in equation (13), let \(\theta_L\) contain every element in \(B_{v\tau}^{(v)}, C_{w\tau}^{(v)}, D_{w\tau}^{(v)}, C_{v\tau}^{(v)}\), and let \(\theta_S\) contain every element in \(B_{v\tau}^{(w)}, C_{w\tau}^{(w)}, D_{w\tau}^{(w)}, C_{v\tau}^{(w)}\). Then, the physical constraints (B.5) to (B.11) can be converted into a matrix form:

\[
H \begin{pmatrix} \theta_L \\ \theta_S \end{pmatrix} = 0, \quad (B.12)
\]
where $\theta = [\theta_L; \theta_S]$. In particular, we can formulate $\theta_L$ and $\theta_S$ as follows:

$$
\theta_L = \left[ \begin{array}{c} 
B^{(v)}_{v\tau} \begin{pmatrix} 111 \end{pmatrix} , B^{(v)}_{v\tau} \begin{pmatrix} 121 \end{pmatrix} , \cdots , B^{(v)}_{v\tau} \begin{pmatrix} r_1 r_1 r_1 \end{pmatrix} , C^{(v)}_{w\tau} \begin{pmatrix} 111 \end{pmatrix} , C^{(v)}_{w\tau} \begin{pmatrix} 121 \end{pmatrix} , \cdots , C^{(v)}_{w\tau} \begin{pmatrix} r_1 r_1 r_1 \end{pmatrix} , \\
D^{(w)}_{w\tau} \begin{pmatrix} 111 \end{pmatrix} , \cdots , D^{(w)}_{w\tau} \begin{pmatrix} r_1 r_1 \end{pmatrix} , C^{(w)}_{v\tau} \begin{pmatrix} 111 \end{pmatrix} , C^{(w)}_{v\tau} \begin{pmatrix} 121 \end{pmatrix} , \cdots , C^{(w)}_{v\tau} \begin{pmatrix} r_1 r_1 r_2 \end{pmatrix} \end{array} \right]^T
$$

$$
\theta_S = \left[ \begin{array}{c} 
B^{(w)}_{v\tau} \begin{pmatrix} 111 \end{pmatrix} , B^{(w)}_{v\tau} \begin{pmatrix} 121 \end{pmatrix} , \cdots , B^{(w)}_{v\tau} \begin{pmatrix} r_1 r_1 r_1 \end{pmatrix} , C^{(w)}_{w\tau} \begin{pmatrix} 111 \end{pmatrix} , C^{(w)}_{w\tau} \begin{pmatrix} 121 \end{pmatrix} , \cdots , C^{(w)}_{w\tau} \begin{pmatrix} r_1 r_1 r_1 \end{pmatrix} , \\
D^{(w)}_{w\tau} \begin{pmatrix} 111 \end{pmatrix} , \cdots , D^{(w)}_{w\tau} \begin{pmatrix} r_1 r_2 \end{pmatrix} , C^{(w)}_{v\tau} \begin{pmatrix} 111 \end{pmatrix} , C^{(w)}_{v\tau} \begin{pmatrix} 121 \end{pmatrix} , \cdots , C^{(w)}_{v\tau} \begin{pmatrix} r_1 r_2 r_2 \end{pmatrix} \end{array} \right]^T
$$

Appendix B.2.1. A Simple Example

A simple example, e.g., with $r_1 = 3$, $r = 4$, can be used to illustrate the constraints from (B.7) to (B.11).

$$
B^{(v)}_{v\tau} \in \mathbb{R}^{3 \times 3 \times 3}:
$$

$$
\begin{align*}
(B^{(v)}_{v\tau})_{(111)} &= (B^{(v)}_{v\tau})_{(222)} = (B^{(v)}_{v\tau})_{(333)} = 0, \\
(B^{(v)}_{v\tau})_{(122)} + (B^{(v)}_{v\tau})_{(212)} + (B^{(v)}_{v\tau})_{(221)} &= 0, \\
(B^{(v)}_{v\tau})_{(133)} + (B^{(v)}_{v\tau})_{(313)} + (B^{(v)}_{v\tau})_{(331)} &= 0, \\
(B^{(v)}_{v\tau})_{(233)} + (B^{(v)}_{v\tau})_{(323)} + (B^{(v)}_{v\tau})_{(332)} &= 0, \\
(B^{(v)}_{v\tau})_{(211)} + (B^{(v)}_{v\tau})_{(121)} + (B^{(v)}_{v\tau})_{(112)} &= 0, \\
(B^{(v)}_{v\tau})_{(311)} + (B^{(v)}_{v\tau})_{(131)} + (B^{(v)}_{v\tau})_{(113)} &= 0, \\
(B^{(v)}_{v\tau})_{(322)} + (B^{(v)}_{v\tau})_{(232)} + (B^{(v)}_{v\tau})_{(223)} &= 0, \\
(B^{(v)}_{v\tau})_{(123)} + (B^{(v)}_{v\tau})_{(132)} + (B^{(v)}_{v\tau})_{(213)} + (B^{(v)}_{v\tau})_{(231)} + (B^{(v)}_{v\tau})_{(312)} + (B^{(v)}_{v\tau})_{(321)} &= 0.
\end{align*}
$$
Appendix B.3. Optimization Problem Formulation

In this section, the details of constructing the optimization problem in Section 2.3 and Section 2.4 are presented. In Section Appendix B.3.1, the matrix formulations of $M_j^L$ in the constrained optimization problem (14)-(15) are sketched. In Section Appendix B.3.2, the ill conditioning in the general formulation of the optimization problem is identified and the mathematical explanation is provided. In Section Appendix B.3.3, a special regularization for the constrained optimization problem is proposed, which can overcome the ill conditioning described in Section Appendix B.3.2.
Appendix B.3.1. General Formulation

Following the notations in Section 2.3 and Section 2.4, the $M_j^L$, the coefficient matrix for $\theta^L$, can be expressed as:

$$M_j^L = \Delta T \cdot \begin{pmatrix} (v(j))^T \otimes (u(j))^T & 0 & \ldots & 0 & (u(j))^T & 0 & \ldots & 0 \\ 0 & (v(j))^T \otimes (u(j))^T & \ldots & 0 & 0 & (u(j))^T & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & (v(j))^T \otimes (u(j))^T & 0 & 0 & \ldots & (u(j))^T \end{pmatrix}$$

(B.29)

where $v(j)$ is the snapshot of $v$ at $t_j$, $u(j)$ is the snapshot of $u = [v; w]$ at $t_j$, and $\otimes$ is the Kronecker tensor product. Using the notations from tensor products, $M_j^L$ can also be expressed as $M_j^L = \Delta T \cdot (I_{r_1 \times r_1} \otimes \left((v(j))^T \otimes (u(j))^T\right)) \oplus (I_{r_1 \times r_1} \otimes (u(j))^T)$, where $I_{r_1 \times r_1}$ is the $r_1 \times r_1$ identity matrix.

Appendix B.3.2. Ill Conditioning Problem

When solving the constrained optimization problem, $\sum_j (M_j^L)^* \Sigma^{-1} M_j^L$ is expected to be full rank. However, due to the general formulation presented in Section Appendix B.3.1, $\sum_j (M_j^L)^* \Sigma^{-1} M_j^L$ will inevitably be rank deficient, which leads to the ill conditioning. To track the source of the ill conditioning, it is essential to check the matrix form of $(M^j)^* (\Sigma^{(k)})^{-1} M^j$. In particular, $(M^j)^* (\Sigma^{(k)})^{-1} M^j$ yields the
following:

\[(M^j)^* (\Sigma^{(k)})^{-1} M^j = \begin{bmatrix}
\alpha_j^* & 0 & \ldots & 0 \\
0 & \alpha_j^* & \ldots & 0 \\
\vdots & \ddots & \ddots & \vdots \\
0 & 0 & \alpha_j^* & \ldots \\
\beta_j^* & 0 & \ldots & 0 \\
0 & \beta_j^* & \ldots & 0 \\
\vdots & \ddots & \ddots & \vdots \\
0 & 0 & \ldots & \beta_j^*
\end{bmatrix}
\begin{bmatrix}
\Sigma_{11} & \ldots & 0 \\
\vdots & \ddots & \vdots \\
0 & \ldots & \Sigma_{r_1,r_1}
\end{bmatrix}
\begin{bmatrix}
\alpha_j^* & 0 & \ldots & 0 & \beta_j^* & 0 & \ldots & 0 \\
0 & \alpha_j^* & \ldots & 0 & \beta_j^* & 0 & \ldots & 0 \\
\vdots & \ddots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\
0 & \ldots & \alpha_j^* & 0 & \ldots & \beta_j^*
\end{bmatrix}
\]

\[(B.31)\]
Moreover, the products between \( \alpha_j \) and \( \beta_j \) can be written as

\[
\alpha_j \ast \alpha_j = \left[ v^{(j)} \otimes u^{(j)} \right] \otimes \left[ \left( v^{(j)} \right)^T \otimes \left( u^{(j)} \right)^T \right] \\
\beta_j \ast \alpha_j = \left[ v^{(j)} \otimes u^{(j)} \right] \otimes \left[ \left( u^{(j)} \right)^T \otimes \left( u^{(j)} \right)^T \right]
\]
\[ \alpha_j^* \beta_j = \left[ u^{(j)} \right] \bigotimes \left[ (u^{(j)})^T \bigotimes (u^{(j)})^T \right] \]

\[ = \begin{bmatrix}
    u_{1}^{(j)} v_{1}^{(j)} u_{1} u_{1}^{(j)} v_{1}^{(j)} u_{3} \cdots u_{1}^{(j)} v_{r}^{(j)} u_{r} \\
    u_{2}^{(j)} v_{1}^{(j)} u_{1} u_{2}^{(j)} u_{2} v_{1}^{(j)} u_{3} \cdots u_{2}^{(j)} v_{r}^{(j)} u_{r} \\
    \vdots \\
    u_{r}^{(j)} v_{1}^{(j)} u_{1} u_{r}^{(j)} u_{2}^{(j)} v_{1}^{(j)} u_{3} \cdots u_{r}^{(j)} v_{r}^{(j)} u_{r} \\
\end{bmatrix} \in \mathbb{R}^{r_1 r \times r_1 r}, \tag{B.36} \]

Then

\[ \beta_j^* \beta_j = u^{(j)} \bigotimes (u^{(j)})^T \]

\[ = \begin{bmatrix}
    u_{1}^{(j)} u_{1}^{(j)} u_{1}^{(j)} u_{3} \cdots u_{1}^{(j)} u_{r} \\
    u_{2}^{(j)} u_{2}^{(j)} u_{2}^{(j)} u_{3} \cdots u_{2}^{(j)} u_{r} \\
    \vdots \\
    u_{r}^{(j)} u_{r}^{(j)} u_{r}^{(j)} u_{3} \cdots u_{r}^{(j)} u_{r} \\
\end{bmatrix} \in \mathbb{R}^{r \times r}. \tag{B.37} \]

The rank deficiency occurs from \( \alpha_j^* \alpha_j \) where the repeated entries appear in different rows. The entry-wise expression of \( \alpha_j^* \alpha_j \) yields

\[ \left( \alpha_j^* \alpha_j \right)_{p,q} = (v_{l_1} u_{l_2}) (v_{m_1} u_{m_2}), \quad p = (r(l_1 - 1) + l_2) \]

\[ q = (r(m_1 - 1) + m_2), \tag{B.38} \]

where \( 1 \leq l_1, m_1 \leq r_1, 1 \leq l_2, m_2 \leq r, p \) is the row index, and \( q \) is the column index of \( \alpha_j^* \alpha_j \), which is also the block matrix in \( (M_L)^* (\Sigma_L^{(k)})^{-1} M_L \). Note that, since \( v_{l_1} = u_{l_1} \), \( 1 \leq l_1 \leq r_1 \), the following equality holds:

\[ v_{l_1} u_{l_2} = v_{l_2} u_{l_1}, \quad \text{provided that } 1 \leq l_1, l_2 \leq r_1. \tag{B.39} \]

Therefore, for \( l_1 \neq l_2 \) and \( 1 \leq l_1, l_2 \leq r_1 \) \( (r_1 \geq 2) \), there must exist \( p_1 = (r(l_1 - 1) + l_2) \) and \( p_2 = (r(l_2 - 1) + l_1) \), such that

\[ \left( \alpha_j^* \alpha_j \right)_{p_1,q} = \left( \alpha_j^* \alpha_j \right)_{p_2,q}, \quad p_1 \neq p_2. \tag{B.40} \]

Then \( \frac{r_1 (r_1 - 1)}{2} \) pairs of rows are the same in each sub-block \( (r_1 \text{ sub-blocks in total}) \) of \( (M_L^*)^* (\Sigma_L^{(k)})^{-1} M_L \), and hence the deficient rank of matrix \( K_L \) will be \( \frac{r_1^2 (r_1 - 1)}{2} \).
Remark. To illustrate the ill conditioning, we use \( r_1 = 2 \) and \( r = 4 \) as a simple case.

Let

\[
  u = \begin{pmatrix} v_1 \\ v_2 \\ w_1 \\ w_2 \end{pmatrix}, \quad v = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}, \quad v = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}.
\]

Then

\[
  \mathbf{M}_L' = \begin{pmatrix} \alpha^{(j)} & 0_{1 \times 8} & \beta^{(j)} & 0_{1 \times 4} \\ 0_{1 \times 8} & \alpha^{(j)} & 0_{1 \times 4} & \beta^{(j)} \end{pmatrix},
\]

with

\[
  \alpha^{(j)} = \begin{pmatrix} v_1^{(j)} v_1^{(j)} & v_1^{(j)} v_2^{(j)} & v_2^{(j)} w_1^{(j)} & v_2^{(j)} v_1^{(j)} & v_2^{(j)} v_2^{(j)} & v_1^{(j)} w_1^{(j)} & v_1^{(j)} w_2^{(j)} & v_2^{(j)} w_1^{(j)} & v_2^{(j)} w_2^{(j)} \end{pmatrix},
\]

\[
  \beta^{(j)} = \begin{pmatrix} v_2 \\ v_1 \\ w_1 \\ w_2 \end{pmatrix},
\]

\[
  0_{1 \times 8} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix},
\]

\[
  0_{1 \times 4} = \begin{pmatrix} 0 & 0 & 0 & 0 \end{pmatrix}.
\]

The unknown vector \( \theta_L \) yields the following: I don’t think we’re allowed colored text.

\[
  \theta_L = \begin{bmatrix} 
  \begin{pmatrix} B_{v}^{(v)} \end{pmatrix}_{111} & \begin{pmatrix} B_{v}^{(v)} \end{pmatrix}_{121} & \begin{pmatrix} C_{v}^{(v)} \end{pmatrix}_{(111)} & \begin{pmatrix} C_{w}^{(v)} \end{pmatrix}_{(121)} & \\
  \begin{pmatrix} B_{v}^{(v)} \end{pmatrix}_{211} & \begin{pmatrix} B_{v}^{(v)} \end{pmatrix}_{221} & \begin{pmatrix} C_{v}^{(v)} \end{pmatrix}_{(211)} & \begin{pmatrix} C_{w}^{(v)} \end{pmatrix}_{(221)} & \\
  \begin{pmatrix} B_{v}^{(v)} \end{pmatrix}_{112} & \begin{pmatrix} B_{v}^{(v)} \end{pmatrix}_{122} & \begin{pmatrix} C_{v}^{(v)} \end{pmatrix}_{(112)} & \begin{pmatrix} C_{w}^{(v)} \end{pmatrix}_{(122)} & \\
  \begin{pmatrix} B_{v}^{(v)} \end{pmatrix}_{212} & \begin{pmatrix} B_{v}^{(v)} \end{pmatrix}_{222} & \begin{pmatrix} C_{v}^{(v)} \end{pmatrix}_{(212)} & \begin{pmatrix} C_{w}^{(v)} \end{pmatrix}_{(222)} & \\
  \begin{pmatrix} C_{v}^{(v)} \end{pmatrix}_{(11)} & \begin{pmatrix} C_{v}^{(v)} \end{pmatrix}_{(12)} & \begin{pmatrix} D_{v}^{(v)} \end{pmatrix}_{(11)} & \begin{pmatrix} D_{w}^{(v)} \end{pmatrix}_{(12)} & \\
  \begin{pmatrix} C_{v}^{(v)} \end{pmatrix}_{(21)} & \begin{pmatrix} C_{v}^{(v)} \end{pmatrix}_{(22)} & \begin{pmatrix} D_{v}^{(v)} \end{pmatrix}_{(21)} & \begin{pmatrix} D_{w}^{(v)} \end{pmatrix}_{(22)} &
\end{bmatrix}
\]

where the first four rows in \( \text{(B.47)} \) terms denote the nonlinear parameters in \( \theta_L \), and the last two rows in \( \text{(B.47)} \) terms denote the linear parameters in \( \theta_L \). The \( j \)th row of
product \((M_L^j)^T (\Sigma^{(k)})^{-1} M_L^j\) is related to the \(j\)th row of the residual. In particular, the first row of the product \((M_L^j)^T (\Sigma^{(k)})^{-1} M_L^j\) yields the following:

\[
\left((M_L^j)^* (\Sigma^{(k)})^{-1} M_L^j\right)_1 = \begin{bmatrix}
\Sigma_{11} v_1 v_1 v_1 v_1, & \sum_{11} v_1 v_1 v_1 v_1, & \sum_{11} v_1 v_1 v_2 w_1, & \sum_{11} v_1 v_1 v_1 w_2, \\
0, & 0, & 0, & 0, & 0, & 0, \\
0, & 0, & 0, & 0, & 0, & 0
\end{bmatrix}.
\]

(B.48)

The problem occurs in the second and fifth rows. The second row yields:

\[
\left((M_L^j)^* (\Sigma^{(k)})^{-1} M_L^j\right)_2 = \begin{bmatrix}
\Sigma_{11} v_1 v_2 v_1 v_1, & \sum_{11} v_1 v_2 v_1 v_1, & \sum_{11} v_1 v_2 v_2 w_1, & \sum_{11} v_1 v_2 v_1 w_2, \\
0, & 0, & 0, & 0, & 0, & 0, \\
0, & 0, & 0, & 0, & 0, & 0
\end{bmatrix}.
\]

(B.49)

The fifth row yields:

\[
\left((M_L^j)^* (\Sigma^{(k)})^{-1} M_L^j\right)_5 = \begin{bmatrix}
\Sigma_{11} v_2 v_1 v_1 v_1, & \sum_{11} v_2 v_1 v_1 v_1, & \sum_{11} v_2 v_1 v_1 w_1, & \sum_{11} v_2 v_1 v_1 w_2, \\
0, & 0, & 0, & 0, & 0, & 0, \\
0, & 0, & 0, & 0, & 0, & 0
\end{bmatrix}.
\]

(B.50)

The terms \(v_2 v_1\) and \(v_1 v_2\) are equal. These two terms are corresponding to the second and fifth elements of \(\theta_L\), i.e., \(B_{\nu\nu}^{(\nu)}\) and \(B_{\nu\nu}^{(\nu)}\). In fact, this issue may occur in solving the least square problem if all the nonlinear unknowns are involved, i.e., all entries of the unknown tensors are assumed non-zeros.
Appendix B.3.3. Special Regularization in Optimization Problem

To overcome the rank-deficiency from the formulation of $M_L$, we consider that $B_{B^v \tau}$ and $B_{B^w \tau}$ are upper triangular matrices in each dimension, i.e., $(B_{B^v \tau})_{ijk}$ and $(B_{B^w \tau})_{ijk}$ are upper triangular matrices for a fixed index $k$. In particular, these terms share the following:

\begin{equation}
(B_{B^v \tau})_{ijk} = 0, \quad \text{provided that } i > j; i, j, k = 1, \ldots, r_1 \tag{B.51}
\end{equation}

\begin{equation}
(B_{B^w \tau})_{ijk} = 0, \quad \text{provided that } i > j; i, j = 1, \ldots, r_1; k = 1, \ldots, r_2. \tag{B.52}
\end{equation}

Note that, since the constraints for $(B_{B^w \tau})_{ijk}$ require that $(B_{B^w \tau})_{iii} = 0$ and $(B_{B^w \tau})_{ijk} + (B_{B^w \tau})_{jik} = 0$, $i \neq j$, $B_{B^w \tau}$ vanishes. Hence, the discrete formulation of the new multiscale physics constraints yield the following:

$B_{B^v \tau}$:

\begin{equation}
(B_{B^v \tau})_{(iii)} = 0, \quad i = 1, \ldots, r_1, \tag{B.53}
\end{equation}

\begin{equation}
(B_{B^v \tau})_{(jjj)} = 0, \quad i, j = 1, \ldots, r_1, i < j, \tag{B.54}
\end{equation}

\begin{equation}
(B_{B^v \tau})_{(ijk)} + (B_{B^v \tau})_{(ikj)} + (B_{B^v \tau})_{(jki)} = 0, \quad i, j, k = 1, \ldots, r_1, i < j < k. \tag{B.55}
\end{equation}

$C_{w^v \tau}$ and $B_{w^v \tau}$:

\begin{equation}
(C_{w^v \tau})_{(iki)} + (B_{w^v \tau})_{(iik)} = 0, \quad i = 1, \ldots, r_1, k = 1, \ldots, r_2; \tag{B.56}
\end{equation}

\begin{equation}
(C_{w^v \tau})_{(ikj)} + (C_{w^v \tau})_{(jk)j} + (B_{w^v \tau})_{(jki)} = 0, \quad i < j, i, j = 1, \ldots, r_1, k = 1, \ldots, r_2. \tag{B.57}
\end{equation}

$C_{w^w \tau}$:

\begin{equation}
(C_{w^w \tau})_{(jjj)} = 0, \quad i = 1, \ldots, r_1; j = 1, \ldots, r_2, \tag{B.58}
\end{equation}

\begin{equation}
(C_{w^w \tau})_{(ijk)} + (C_{w^w \tau})_{(ikj)} = 0, \quad i = 1, \ldots, r_1; j, k = 1, \ldots, r_2, j \neq k. \tag{B.59}
\end{equation}
References

[1] G. K. Vallis, Atmospheric and oceanic fluid dynamics, Cambridge University Press, 2017.

[2] S. H. Strogatz, Nonlinear dynamics and chaos with student solutions manual: With applications to physics, biology, chemistry, and engineering, CRC press, 2018.

[3] D. C. Wilcox, Multiscale model for turbulent flows, AIAA journal 26 (11) (1988) 1311–1320.

[4] S. A. Sheard, A. Mostashari, Principles of complex systems for systems engineering, Systems Engineering 12 (4) (2009) 295–311.

[5] M. Ghil, S. Childress, Topics in geophysical fluid dynamics: atmospheric dynamics, dynamo theory, and climate dynamics, Springer Science & Business Media, 2012.

[6] A. J. Majda, Introduction to turbulent dynamical systems in complex systems, Springer, 2016.

[7] W.-K. Tao, J.-D. Chern, R. Atlas, D. Randall, M. Khairoutdinov, J.-L. Li, D. E. Waliser, A. Hou, X. Lin, C. Peters-Lidard, et al., A multiscale modeling system: Developments, applications, and critical issues, Bulletin of the American Meteorological Society 90 (4) (2009) 515–534.

[8] A. J. Majda, I. Grooms, New perspectives on superparameterization for geophysical turbulence, Journal of Computational Physics 271 (2014) 60–77.

[9] R. Salmon, Lectures on geophysical fluid dynamics, Oxford University Press, 1998.

[10] H. A. Dijkstra, Nonlinear climate dynamics, Cambridge University Press, 2013.

[11] T. Palmer, A nonlinear dynamical perspective on climate change, Weather 48 (10) (1993) 314–326.
[12] M. Farazmand, T. P. Sapsis, Extreme events: Mechanisms and prediction, Applied Mechanics Reviews 71 (5).

[13] K. E. Trenberth, J. T. Fasullo, T. G. Shepherd, Attribution of climate extreme events, Nature Climate Change 5 (8) (2015) 725–730.

[14] H. Moffatt, Extreme events in turbulent flow, Journal of Fluid Mechanics 914.

[15] A. Majda, Introduction to PDEs and Waves for the Atmosphere and Ocean, Vol. 9, American Mathematical Soc., 2003.

[16] P. Manneville, Y. Pomeau, Intermittency and the Lorenz model, Physics Letters A 75 (1-2) (1979) 1–2.

[17] A. J. Majda, Challenges in climate science and contemporary applied mathematics, Communications on Pure and Applied Mathematics 65 (7) (2012) 920–948.

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

[19] J. S. Hesthaven, G. Rozza, B. Stamm, Certified Reduced Basis Methods for Parametrized Partial Differential Equations, Springer, 2015.

[20] P. Holmes, J. L. Lumley, G. Berkooz, Turbulence, Coherent Structures, Dynamical Systems and Symmetry, Cambridge, 1996.

[21] B. R. Noack, M. Morzynski, G. Tadmor, Reduced-order modelling for flow control, Vol. 528, Springer Science & Business Media, 2011.

[22] A. Quarteroni, A. Manzoni, F. Negri, Reduced Basis Methods for Partial Differential Equations: An Introduction, Vol. 92, Springer, 2015.

[23] K. Taira, M. S. Hemati, S. L. Brunton, Y. Sun, K. Duraisamy, S. Bagheri, S. T. Dawson, C.-A. Yeh, Modal analysis of fluid flows: Applications and outlook, AIAA journal 58 (3) (2020) 998–1022.

[24] X. Yan, X. Su, Linear regression analysis: theory and computing, World Scientific, 2009.
[25] D. A. Freedman, Statistical models: theory and practice, Cambridge University Press, 2009.

[26] K. Hasselmann, PIPs and POPs: The reduction of complex dynamical systems using principal interaction and oscillation patterns, Journal of Geophysical Research: Atmospheres 93 (D9) (1988) 11015–11021.

[27] F. Kwasniok, The reduction of complex dynamical systems using principal interaction patterns, Physica D: Nonlinear Phenomena 92 (1-2) (1996) 28–60.

[28] C. W. Rowley, I. Mezić, S. Bagheri, P. Schlatter, D. S. Henningson, Spectral analysis of nonlinear flows, Journal of Fluid Mechanics 641 (2009) 115–127.

[29] P. J. Schmid, Dynamic mode decomposition of numerical and experimental data, Journal of Fluid Mechanics 656 (2010) 5–28.

[30] S. E. Ahmed, S. Pawar, O. San, A. Rasheed, T. Iliescu, B. R. Noack, On closures for reduced order models—A spectrum of first-principle to machine-learned avenues, Physics of Fluids 33 (9) (2021) 091301.

[31] K. Carlberg, C. Farhat, J. Cortial, D. Amsallem, The GNAT method for nonlinear model reduction: effective implementation and application to computational fluid dynamics and turbulent flows, Journal of Computational Physics 242 (2013) 623–647.

[32] C. Mou, B. Koc, O. San, L. G. Rebholz, T. Iliescu, Data-driven variational multiscale reduced order models, Computer Methods in Applied Mechanics and Engineering 373 (2021) 113470.

[33] X. Xie, M. Moheddujman, L. G. Rebholz, T. Iliescu, Data-driven filtered reduced order modeling of fluid flows, SIAM Journal on Scientific Computing 40 (3) (2018) B834–B857.

[34] A. J. Majda, I. Timofeyev, E. Vanden Eijnden, A mathematical framework for stochastic climate models, Communications on Pure and Applied Mathematics 54 (8) (2001) 891–974.
[35] A. J. Majda, I. Timofeyev, E. V. Eijnden, Models for stochastic climate prediction, Proceedings of the National Academy of Sciences 96 (26) (1999) 14687–14691.

[36] F. Lu, Data-driven model reduction for stochastic burgers equations, Entropy 22 (12) (2020) 1360.

[37] K. K. Lin, F. Lu, Data-driven model reduction, wiener projections, and the Koopman-Mori-Zwanzig formalism, Journal of Computational Physics 424 (2021) 109864.

[38] A. J. Majda, J. Harlim, Physics constrained nonlinear regression models for time series, Nonlinearity 26 (1) (2012) 201.

[39] J. Harlim, A. Mahdi, A. J. Majda, An ensemble Kalman filter for statistical estimation of physics constrained nonlinear regression models, Journal of Computational Physics 257 (2014) 782–812.

[40] D. Kondrashov, M. D. Chekroun, M. Ghil, Data-driven non-Markovian closure models, Physica D: Nonlinear Phenomena 297 (2015) 33–55.

[41] T. N. Palmer, A nonlinear dynamical perspective on model error: A proposal for non-local stochastic-dynamic parametrization in weather and climate prediction models, Quarterly Journal of the Royal Meteorological Society 127 (572) (2001) 279–304.

[42] A. J. Majda, B. Gershgorin, Improving model fidelity and sensitivity for complex systems through empirical information theory, Proceedings of the National Academy of Sciences 108 (25) (2011) 10044–10049.

[43] D. Crommelin, E. Vanden-Eijnden, Subgrid-scale parameterization with conditional Markov chains, Journal of the Atmospheric Sciences 65 (8) (2008) 2661–2675.

[44] T. J. Phillips, G. L. Potter, D. L. Williamson, R. T. Cederwall, J. S. Boyle, M. Fiorino, J. J. Hnilo, J. G. Olson, S. Xie, J. J. Yio, Evaluating parameterizations in general circulation models: Climate simulation meets weather prediction, Bulletin of the American Meteorological Society 85 (12) (2004) 1903–1916.
[45] M. Branicki, N. Chen, A. J. Majda, Non-Gaussian test models for prediction and state estimation with model errors, Chinese Annals of Mathematics, Series B 34 (1) (2013) 29–64.

[46] G. Evensen, Data assimilation: the ensemble Kalman filter, Springer Science & Business Media, 2009.

[47] E. Kalnay, Atmospheric modeling, data assimilation and predictability, Cambridge University Press, 2003.

[48] K. Law, A. Stuart, K. Zygalakis, Data assimilation, Cham, Switzerland: Springer 214.

[49] A. J. Majda, J. Harlim, Filtering complex turbulent systems, Cambridge University Press, 2012.

[50] A. J. Majda, Y. Yuan, Fundamental limitations of ad hoc linear and quadratic multi-level regression models for physical systems, Discrete and Continuous Dynamical Systems B 17 (4) (2012) 1333–1363.

[51] N. Chen, A. Majda, Conditional Gaussian systems for multiscale nonlinear stochastic systems: Prediction, state estimation and uncertainty quantification, Entropy 20 (7) (2018) 509.

[52] W. E, K. Khanin, A. Mazel, Y. Sinai, Invariant measures for Burgers equation with stochastic forcing, Annals of Mathematics (2000) 877–960.

[53] W. E, K. Khanin, A. Mazel, Y. Sinai, Probability distribution functions for the random forced Burgers equation, Physical Review Letters 78 (10) (1997) 1904.

[54] N. Chen, A. J. Majda, Filtering nonlinear turbulent dynamical systems through conditional Gaussian statistics, Monthly Weather Review 144 (12) (2016) 4885–4917.

[55] N. Chen, A. J. Majda, D. Giannakis, Predicting the cloud patterns of the Madden-Julian oscillation through a low-order nonlinear stochastic model, Geophysical Research Letters 41 (15) (2014) 5612–5619.
[56] N. Chen, A. J. Majda, Predicting the real-time multivariate Madden–Julian oscillation index through a low-order nonlinear stochastic model, Monthly Weather Review 143 (6) (2015) 2148–2169.

[57] N. Chen, A. J. Majda, X. T. Tong, Information barriers for noisy Lagrangian tracers in filtering random incompressible flows, Nonlinearity 27 (9) (2014) 2133.

[58] N. Chen, A. J. Majda, X. T. Tong, Noisy Lagrangian tracers for filtering random rotating compressible flows, Journal of Nonlinear Science 25 (3) (2015) 451–488.

[59] M. Branicki, A. J. Majda, Dynamic stochastic superresolution of sparsely observed turbulent systems, Journal of Computational Physics 241 (2013) 333–363.

[60] S. R. Keating, A. J. Majda, K. S. Smith, New methods for estimating ocean eddy heat transport using satellite altimetry, Monthly Weather Review 140 (5) (2012) 1703–1722.

[61] M. Kaercher, S. Boyaval, M. A. Grepl, K. Veroy, Reduced basis approximation and a posteriori error bounds for 4D-Var data assimilation, Optim. Eng. (2018) 1–33.

[62] Y. Maday, A. T. Patera, J. D. Penn, M. Yano, A parameterized-background data-weak approach to variational data assimilation: formulation, analysis, and application to acoustics, Int. J. Num. Meth. Engng. 102 (5) (2015) 933–965.

[63] S. Pagani, A. Manzoni, A. Quarteroni, Efficient state/parameter estimation in nonlinear unsteady PDEs by a reduced basis ensemble Kalman filter, SIAM-ASA J. Uncertain. 5 (1) (2017) 890–921.

[64] A. A. Popov, C. Mou, T. Iliescu, A. Sandu, A multifidelity ensemble Kalman filter with reduced order control variates, SIAM J. Sci. Comput. 43 (2) (2021) A1134–A1162.

[65] R. Ştefănescu, A. Sandu, I. M. Navon, POD/DEIM reduced-order strategies for efficient four dimensional variational data assimilation, J. Comput. Phys. 295 (2015) 569–595.
[66] D. Xiao, J. Du, F. Fang, C. C. Pain, J. Li, Parameterised non-intrusive reduced order methods for ensemble Kalman filter data assimilation, Comput. & Fluids 177 (2018) 69–77.

[67] C. Zerfas, L. G. Rebholz, M. Schneier, T. Iliescu, Continuous data assimilation reduced order models of fluid flow, Comput. Meth. Appl. Mech. Eng. 357 (2019) 112596.

[68] W. Lahoz, B. Khattatov, R. Ménard, Data assimilation and information, in: Data Assimilation, Springer, 2010, pp. 3–12.

[69] J. S. Whitaker, T. M. Hamill, Ensemble data assimilation without perturbed observations, Monthly weather review 130 (7) (2002) 1913–1924.

[70] B. R. Hunt, E. J. Kostelich, I. Szunyogh, Efficient data assimilation for spatiotemporal chaos: A local ensemble transform kalman filter, Physica D: Nonlinear Phenomena 230 (1-2) (2007) 112–126.

[71] P. L. Houtekamer, F. Zhang, Review of the ensemble kalman filter for atmospheric data assimilation, Monthly Weather Review 144 (12) (2016) 4489–4532.

[72] R. S. Liptser, A. N. Shiryaev, Statistics of random processes II: Applications, Vol. 6, Springer Science & Business Media, 2013.

[73] R. E. Kalman, R. S. Bucy, New results in linear filtering and prediction theory, Journal of Basic Engineering 83 (1) (1961) 95–108.

[74] S. Saha, S. Moorthi, H.-L. Pan, X. Wu, J. Wang, S. Nadiga, P. Tripp, R. Kistler, J. Woollen, D. Behringer, et al., The ncep climate forecast system reanalysis, Bulletin of the American Meteorological Society 91 (8) (2010) 1015–1058.

[75] N. Chen, Learning nonlinear turbulent dynamics from partial observations via analytically solvable conditional statistics, Journal of Computational Physics 418 (2020) 109635.

[76] S. Boyd, L. Vandenberghe, Convex optimization, Cambridge University Press, 2004.
[77] K. Bergemann, S. Reich, An ensemble Kalman-Bucy filter for continuous data assimilation, Meteorologische Zeitschrift 21 (2012) 213–219.

[78] R. J. Greatbatch, B. T. Nadiga, Four-gyre circulation in a barotropic model with double-gyre wind forcing, Journal of Physical Oceanography 30 (6) (2000) 1461–1471.

[79] C. Mou, H. Liu, D. R. Wells, T. Iliescu, Data-driven correction reduced order models for the quasi-geostrophic equations: A numerical investigation, International Journal of Computational Fluid Dynamics 34 (2) (2020) 147–159.

[80] O. San, T. Iliescu, A stabilized proper orthogonal decomposition reduced-order model for large scale quasigeostrophic ocean circulation, Advances in Computational Mathematics (2015) 1289–1319.

[81] O. San, A. E. Staples, Z. Wang, T. Iliescu, Approximate deconvolution large eddy simulation of a barotropic ocean circulation model, Ocean Modelling 40 (2011) 120–132.

[82] S. Kullback, R. A. Leibler, On information and sufficiency, The Annals of Mathematical Statistics 22 (1) (1951) 79–86.

[83] S. Kullback, Letter to the editor: The kullback-leibler distance, American Statistician.

[84] R. Kleeman, Information theory and dynamical system predictability, Entropy 13 (3) (2011) 612–649.

[85] A. Majda, R. V. Abramov, M. J. Grote, Information theory and stochastics for multiscale nonlinear systems, Vol. 25, American Mathematical Soc., 2005.

[86] S. L. Brunton, J. L. Proctor, J. N. Kutz, Discovering governing equations from data by sparse identification of nonlinear dynamical systems, Proceedings of the National Academy of Sciences 113 (15) (2016) 3932–3937.
[87] J. Sun, E. M. Bollt, Causation entropy identifies indirect influences, dominance of neighbors and anticipatory couplings, Physica D: Nonlinear Phenomena 267 (2014) 49–57.

[88] J. Elinger, Information theoretic causality measures for parameter estimation and system identification, Ph.D. thesis, Georgia Institute of Technology (2020).