

# Model and data reduction for data assimilation: Particle filters employing projected forecasts and data with application to a shallow water model <sup>☆</sup>



Aishah Albarakati <sup>a</sup>, Marko Budišić <sup>a</sup>, Rose Crocker <sup>b</sup>, Juniper Glass-Klaiber <sup>c</sup>, Sarah Iams <sup>d</sup>, John Maclean <sup>b</sup>, Noah Marshall <sup>e</sup>, Colin Roberts <sup>f</sup>, Erik S. Van Vleck <sup>g,\*</sup>

<sup>a</sup> Department of Mathematics, Clarkson University, Potsdam, NY 13699, United States of America

<sup>b</sup> School of Mathematical Sciences, University of Adelaide, Adelaide SA 5005, Australia

<sup>c</sup> Department of Physics, Mount Holyoke College, South Hadley, MA 01075, United States of America

<sup>d</sup> Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, United States of America

<sup>e</sup> Department of Mathematics and Statistics, McGill University, Montreal, Quebec H3A 0B9, Canada

<sup>f</sup> Department of Mathematics, Colorado State University, Ft. Collins, CO 80523, United States of America

<sup>g</sup> Department of Mathematics, University of Kansas, Lawrence, KS 66045, United States of America

## ARTICLE INFO

### Keywords:

Data assimilation  
Particle filters  
Order reduction  
Proper orthogonal decomposition  
Dynamic mode decomposition  
Shallow water equation

## ABSTRACT

The understanding of nonlinear, high dimensional flows, e.g., atmospheric and ocean flows, is critical to address the impacts of global climate change. Data Assimilation (DA) techniques combine physical models and observational data, often in a Bayesian framework, to predict the future state of the model and the uncertainty in this prediction. Inherent in these systems are noise (Gaussian and non-Gaussian), nonlinearity, and high dimensionality that pose challenges to making accurate predictions. To address these issues we investigate the use of both model and data dimension reduction based on techniques including Assimilation in the Unstable Subspace (AUS), Proper Orthogonal Decomposition (POD), and Dynamic Mode Decomposition (DMD). Algorithms to take advantage of projected physical and data models may be combined with DA techniques such as Ensemble Kalman Filter (EnKF) and Particle Filter (PF) variants. The projected DA techniques are developed for the optimal proposal particle filter and applied to the Lorenz'96 model (L96) and Shallow Water Equations (SWE) to test the efficacy of our techniques in high dimensional, nonlinear systems.

## Contents

1. Introduction . . . . .	195
2. Background on data assimilation . . . . .	196
2.1. Data assimilation: nonlinearity and non-Gaussian posteriors . . . . .	196
2.2. The Standard Particle Filter (PF) . . . . .	196
2.3. Optimal Proposal Particle Filter (OP-PF) . . . . .	197
3. Particle filters with dimension reduction in state and observation . . . . .	197
3.1. Dimension reduction of state and observation . . . . .	197
3.2. Projected Particle Filter (Proj-PF) . . . . .	198
3.3. Projected Optimal Proposal Particle Filter (Proj-OP-PF) . . . . .	198
4. Techniques for model reduction . . . . .	199
4.1. Proper Orthogonal Decomposition (POD) . . . . .	199
4.2. Dynamic Mode Decomposition (DMD) . . . . .	199

<sup>☆</sup> This research was supported in part by NSF grants DMS-1714195 and DMS-1722578 and originated as part of the AIM 2019 MCRN summer school and academic year engagement program.

\* Corresponding author.

E-mail addresses: albaraaa@clarkson.edu (A. Albarakati), mbudisic@clarkson.edu (M. Budišić), rose.crocker@adelaide.edu.au (R. Crocker), glass22j@mtsholyoke.edu (J. Glass-Klaiber), siams@seas.harvard.edu (S. Iams), John.Maclean@adelaide.edu.au (J. Maclean), noah.marshall2@mail.mcgill.ca (N. Marshall), colin.roberts@rams.colostate.edu (C. Roberts), erikvv@ku.edu (E.S. Van Vleck).

<https://doi.org/10.1016/j.camwa.2021.05.026>

Available online 9 June 2021

0898-1221/© 2021 Elsevier Ltd. All rights reserved.

4.3.	Assimilation in the Unstable Subspace (AUS)	201
5.	Numerical results	201
5.1.	Common experimental setup	201
5.2.	Lorenz '96 equations	201
5.2.1.	Model and parameters	201
5.2.2.	Experiment 1 ( $F = 3, 8$ , $M = 40$ )	203
5.2.3.	Experiment 2 ( $F = 3, 8$ , $M = 400$ )	203
5.2.4.	Experiment 3 ( $F = 3, 4, 6, 8$ , $M = 400$ )	205
5.3.	Shallow Water Equations (SWE)	205
5.3.1.	Model, simulation, and parameters	205
5.3.2.	Experiment 1 (number of particles)	208
5.3.3.	Experiment 2 (observation scenarios)	208
5.3.4.	Experiment 3 (sparse and complete observations)	208
5.3.5.	Experiment 4 (larger observation error covariances $\mathbf{R}$ )	208
6.	Discussion and conclusions	210
	References	210
	Acronyms	211

## 1. Introduction

Several important challenges impede the development of Data Assimilation (DA) techniques: nonlinear physical models, high dimensional models and data, and non-Gaussian posterior distributions. Particle filters and their variants are well suited to handle nonlinearities and non-Gaussian distributions, but due to so-called filter degeneracy still struggle to make accurate predictions for high dimensional problems, especially partial differential equations (PDEs). Variants of particle filters have been developed to ameliorate this difficulty, including implicit particle filters, proposal density methods, the optimal proposal, [1–4]. A related set of works analyze the performance of particle filters: [5,6] show that particle filter degeneracy is induced by a ‘curse of dimensionality’ associated with the data and/or model dimension.

Our contribution in this paper is to develop an approach to particle filtering based on reduced-dimension physical and data models, employing projections of the data and model spaces. The crucial benefit of this approach is that it directly targets the filter degeneracy induced by, for example, simulating some high-dimensional PDEs, while maintaining the Bayesian framework of the particle filter suitable for nonlinear, non-Gaussian DA. We build on the substantial development of Assimilation in the Unstable Subspace (AUS) methods (see, e.g., [7–11]), and recent work on projected data models for particle filtering [12]. The AUS methods are based on state space projections to subspaces spanned by Lyapunov vectors corresponding to the dominant Lyapunov exponents of the system dynamics. The AUS state space projections can greatly improve Kalman Filter (KF) and variational DA schemes: for example, an AUS implementation of 3D-Var works efficiently in high dimensions [13]. However for ensemble-based DA schemes, including the particle filter, the AUS approach has limited promise as the ensemble forecast already tends to align with the dominant Lyapunov exponents [14]. Here we extend the AUS approach by developing projections based on other common model reduction techniques, such as Proper Orthogonal Decomposition (POD) and Dynamic Mode Decomposition (DMD), and by considering combinations of projected physical and data models. Although our focus here is on particle filters (both the standard and optimal proposal forms), the projected models developed here are also applicable to KF and variational techniques. We combine dimension reduction in both the physical model and the data model and compare projections based on AUS, POD, and DMD. Using these projected models, we develop projected particle filter algorithms and apply them to the Lorenz’96 model (L96) and the Shallow Water Equations (SWE).

While the present paper is motivated in large part by techniques developed using AUS projections, see, e.g., [10,12,15], we are also motivated by recently developed techniques for localized particle filters. Recent work has often focused on the issue of localization, such as [16],

and currently two localized particle filtering algorithms [17,18] have been applied in an operational geophysical framework. In the localized particle filter of [18] observations are projected onto the subspace spanned by the ensemble of model forecasts to reduce the dimension of the observations. In [19], the authors apply a localized particle filter which reduces the number of particles needed for effective assimilation. In this scheme, particle weights are updated locally near observations, but are preserved away from observations to mimic the covariance localization in Ensemble Kalman Filter (EnKF). Other techniques such as the *Dynamically Orthogonal* formulation in [20,21] can be interpreted in terms of reduced order physical and data models (see [22–24]).

Although the original contribution of this paper is in combining data-driven Reduced Order Models (ROM) with particle filters, both POD and DMD have been used in concert with other data assimilation techniques. Kalman filter assimilation with a DMD-ROM has been used to predict wind turbine wakes in [25], while in [26] the DMD was used to enhance a Bayesian-optimized Kalman filter to predict events in the upper atmosphere.

For medium- to high-dimensional models, POD can be used to determine the dominant energy modes and, by reducing the model to the corresponding subspace, exploit a possible low dimensional structure of the model space for use in the nonlinear filtering problem [27], with EnKF [28], and with the Four-dimensional variational data assimilation (4D-Var) assimilation scheme [29–32]. Likewise, POD, tensorial POD, and discrete empirical interpolation have been used in the 4D-Var scheme to reduce the state space in application to the shallow water equations [33]. In [34], efficacy of assimilation based on merging DMD, neural networks, and 4D-Var was evaluated on chaotic dynamical systems.

While we primarily focus on using DMD and POD to enhance data assimilation algorithms, it is interesting to point out that [35,36] used data assimilation techniques, specifically the Kalman filter and its variants, to enhance the computation of DMD, in order to reduce the contribution of system noise and lead to constructing better estimates on the eigenvalues corresponding to DMD modes.

In addition, our use of POD and DMD as dimension reduction techniques provides a bridge between techniques developed to assimilate coherent structures and the model reduction literature. Methods such as those developed in [37] and [38] assimilate coherent structures extracted from data, but without an explicit form for the likelihood of the coherent structures. Instead these works use likelihood-free sequential Monte Carlo methods, or an ad hoc *perturbed observations* approach. In this paper we derive an explicit likelihood that corresponds to coherent structures extracted by POD or DMD.

In section 2 we present background on data assimilation techniques including ensemble Kalman filters and particle filters. In section 3 we formulate, using abstract projections, the projected physical and data

models and their use in the context of standard and optimal proposal particle filters. We outline in section 4 the basics of the POD, DMD, and AUS model reduction techniques. Section 5 contains numerical results obtained using the algorithms developed to take advantage of these projected physical and data models. These methods are applied to two nonlinear models: L96 and SWE. Numerical results show the efficacy of the techniques that have been developed. In section 6 we summarize and analyze our results and outline directions for future research.

## 2. Background on data assimilation

### 2.1. Data assimilation: nonlinearity and non-Gaussian posteriors

Data assimilation is a suite of methods commonly used for obtaining accurate estimates of states and/or parameters associated with large-scale geophysical systems such as the climate and atmosphere. DA schemes seek to optimally combine the information contained in an *observation*  $y_t$ , informed by collected data, with that in a *forecast*  $u_t$ , given by a mathematical model of the system. The observation and forecast naturally contain associated error, due to factors such as instrumentation error in the observations and model errors or noise in the forecast. The key aim of any DA scheme is to propitiously balance these different sources of error and the defining characteristic of a particular DA scheme is how it does so [39].

We formulate the data assimilation problem in terms of a physical model and a data model. Consider the discrete time stochastic model with additive noise

$$u_{t+1} = F_t(u_t) + \omega_t, \quad (1)$$

where  $u_t \in \mathbb{R}^M$  is the model state at time index  $t$ ,  $\omega_t \in \mathbb{R}^M$  are Gaussian errors in the model,  $\omega_t \sim \mathcal{N}(0, Q_t)$ , and the function  $F_t : \mathbb{R}^M \rightarrow \mathbb{R}^M$  is the deterministic component of the model. The data model represents the measurement of the physical state as observations  $y_t$  given by

$$y_t = H u_t + \eta_t, \quad (2)$$

where  $\eta_t \in \mathbb{R}^D$  are Gaussian errors in the observation,  $\eta_t \sim \mathcal{N}(0, R_t)$ , while  $H : \mathbb{R}^M \rightarrow \mathbb{R}^D$ ,  $D \leq M$  is the matrix<sup>1</sup> representing the observation operator.

Equation (2) implies, first, that the data  $y_t$  is generated from the true (unknown) state by  $y_t = H u_t^{\text{truth}} + \eta_t$ , and second, that there is a framework to convert estimates of the state  $u_t$  into ‘data space’ via  $H u_t$ . The task for the DA algorithm is to provide the best estimate of the state  $u_t$  using the combined information from eqs. (1) and (2). Complete observation ( $H = I$ ) is useful in analysis, as it reduces the DA problem to removing the influence of observation noise.

We are interested in DA problems in which the model eq. (1) is sufficiently nonlinear to pose challenges for DA algorithms that rely implicitly on assumptions of linearity, chiefly the EnKF. In order to expand on this criterion, we now formalize the DA problem from a Bayesian perspective.

Let us formally write our estimates of the state and data as Probability Density Functions (PDFs). The distribution of the state evolves from time  $t-1$  to time  $t$ , including information from the data at time  $t$ , according to Bayes’ theorem

$$P(u_t | u_{t-1}, y_t) = \frac{P(y_t | u_t) P(u_t | u_{t-1})}{P(y_t)}, \quad \text{or} \quad P(u_t | u_{t-1}, y_t) \propto P(y_t | u_t) P(u_t | u_{t-1}). \quad (3)$$

The left-hand-side of this equation is the distribution we want to approximate with a DA algorithm. It is the *posterior* distribution, of the

state at time  $t$  given (or conditioned on) the past history of the state and also given the data collected at time  $t$ . The right-hand-side of eq. (3) is a blueprint for the action of a DA scheme. The PDF  $P(u_t | u_{t-1})$ , the *prior* distribution, contains our understanding of the state evolution from time  $t-1$  to  $t$ . The second term in the numerator of eq. (3) is the *likelihood*  $P(y_t | u_t)$  and can be explicitly evaluated from eq. (2). The normalizing constant  $P(y_t)$  in the denominator of eq. (3) is typically impossible to calculate directly; however the PF we develop below implicitly calculates  $P(y_t)$  in a discrete setting, by normalizing weights.

The formulation of DA via eq. (3) may appear incomplete (for example, we have employed without proof a formulation in which data from previous observation times is not used again at time  $t$ ), but can be rigorously justified with some weak assumptions: [40, §3] provides a clear exposition for the standard DA formulation, and [41, e.g.] covers formulations of DA in which the same observations may be assimilated at multiple times.

Having set up the DA problem, let us clarify the goal of this paper. We are specifically interested in DA problems in which the nonlinearity of the model eq. (1), together with factors such as sparse observations,  $D \ll M$ , results in a non-Gaussian posterior distribution in eq. (3). Standard methods such as the EnKF are inappropriate for strongly non-Gaussian posteriors, but the PF described in section 2.2 is suitable. The key difficulty, motivated earlier and developed explicitly for PFs in sections 2.2 and 2.3, is applying the PF to high-dimensional models and data—and the remainder of the paper develops a suite of methods to address this difficulty.

### 2.2. The Standard Particle Filter (PF)

Particle Filters, in their many forms, are one of the most common DA methods for nonlinear systems, as they can be proven to reproduce the true target posterior in the limit of large particle populations [42]. Although this property makes PFs very useful, they also suffer from several issues, including particle degeneracy and the curse of dimensionality, which will be discussed in the context of the standard particle filter.

The simplest form of particle filter is the *Sequential Importance Resampling*, or the *Standard Particle Filter* [43,44, e.g.]. In this method, the probability density function for the state is approximated by a weighted set of guesses at the state variable. The probability density function for the state at time  $t-1$  is

$$P(u_{t-1}) = \sum_{\ell=1}^L w_{t-1}^{\ell} \delta(u_{t-1} - u_{t-1}^{\ell}), \quad (4)$$

where each particle is a state vector  $u_{t-1}^{\ell} \in \mathbb{R}^M$ , each weight  $w_{t-1}^{\ell} \geq 0$ , the sum of all weights  $\sum_{\ell=1}^L w_{t-1}^{\ell} = 1$  and  $\delta(\cdot)$  is the Dirac-delta distribution. This density function is quite simple to interpret: it says that we have  $L$  guesses for the state, and that each guess is judged to be good or bad by its weight.

The Standard PF consists of the following two steps:

- **Particle update:**

Each particle  $u_{t-1}^{\ell}$ ,  $\ell$  from 1 to  $L$ , is updated to time  $t$  by running eq. (1). This completes the step, but let us pause to interpret what has happened. In the Bayesian formulation of eq. (3), consider the prior distribution  $P(u_t | u_{t-1})$ : then eq. (1) implies that for any particular  $u_{t-1}^{\ell}$

$$P(u_t^{\ell} | u_{t-1}^{\ell}) \sim \mathcal{N}(F_{t-1}(u_{t-1}^{\ell}), Q).$$

Combining this equation with eq. (4), we see the exact prior distribution is a Gaussian mixture, and updating each particle according to eq. (1) approximates the exact prior distribution by sampling once from each Gaussian distribution in the mixture.

- **Weight update:**

$$w_t^{\ell} = \frac{1}{c} \exp \left[ -\frac{1}{2} (y_t - H u_t^{\ell})^{\top} R^{-1} (y_t - H u_t^{\ell}) \right] w_{t-1}^{\ell},$$

<sup>1</sup> In this work we focus only on linear observation operators  $H$ , in order to obtain closed equations for the OP-PF algorithm in section 2.3 and the reduced data models in section 3.1. In general the observation operator may be nonlinear, and we discuss extensions of our work to this case in section 6.

where  $c$  is a normalization constant such that  $\sum_{l=1}^L w_l^i = 1$ . As before this brief description concludes the step, but we pause to connect the step to eq. (3). The exponential in the weight update is (up to normalization) precisely the term  $P(y_i | u_i^{\ell})$  (compare eq. (2)). The normalization constant  $c$  implicitly calculates the remaining factor from eq. (3), the denominator  $P(y_i)$ . That is, the weight update multiplies each particle by the precise terms needed to complete eq. (3).

Despite the standard PF being suitable for DA in nonlinear systems, the algorithm suffers from several related issues. The method performs badly if one of the particle weights approaches 1. When this occurs, all others approach zero and so the posterior distribution is effectively being approximated by a single particle. This issue is addressed in the Standard PF by monitoring the Effective Sample Size (ESS),  $\sum_{\ell=1}^L 1/(w_{\ell}^i)^2$ , and resampling (for which there are standard algorithms, see for example [45, §3.3.2]) when the ESS drops below some threshold to refresh the particle ensemble.

The PF's tendency to exhibit particle degeneracy is related to what is often referred to as the *curse of dimensionality*, a phenomenon suffered by all importance sampling algorithms in which efficiency decreases rapidly with increasing dimension of the state space [46]. In high-dimensional spaces, resampling is insufficient to prevent degeneracy and  $L$  must be very large to give an appropriate estimate of the posterior, decreasing the computational efficiency of the algorithm. In fact, it has been shown that the theoretically required sample size to avoid particle degeneracy scales exponentially with the state dimension [47]. Several methods have been developed to reduce the sample size needed to counter particle degeneracy in high dimensions, such as the Optimal Proposal Particle Filter (OP-PF).

### 2.3. Optimal Proposal Particle Filter (OP-PF)

The OP-PF ameliorates the issue of degeneracy in PFs by attempting to ensure all posterior particles have similar weights. The idea now is to use the data in both, the particle and weight updates—first to nudge all particles towards the observations, and then to weight them. The ‘proposal’ in OP-PF refers to the distribution used to update particles from one time step to the next. In the standard particle filter the proposal density is  $P(u_i^{\ell} | u_{i-1}^{\ell}) \sim \mathcal{N}(F_{i-1}(u_{i-1}^{\ell}), \mathbf{Q})$ , as the particles are updated using the model. Comparatively, in the OP-PF, the proposal distribution is conditioned on  $y_i$ , as in  $P(u_i^{\ell} | u_{i-1}^{\ell}, y_i)$ . Given the additive noise of the model (1), the optimal proposal update in each particle is Gaussian with  $P(u_i^{\ell} | u_{i-1}^{\ell}, y_i) \sim \mathcal{N}(\mathbf{m}_i^{\ell}, \mathbf{Q}_p)$ , and the particle update is

$$u_i^{\ell} = \mathbf{m}_i^{\ell} + \boldsymbol{\varphi}, \quad \boldsymbol{\varphi} \sim \mathcal{N}(0, \mathbf{Q}_p) \quad (5)$$

where

$$\mathbf{Q}_p^{-1} = \mathbf{Q}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \quad (6)$$

$$\mathbf{m}_i^{\ell} = \mathbf{F}_{i-1}(u_{i-1}^{\ell}) + \mathbf{Q}_p \mathbf{H}^T \mathbf{R}^{-1} (y_i - \mathbf{H} \mathbf{F}_{i-1}(u_{i-1}^{\ell})) . \quad (7)$$

Having employed the data in the particle update, the weight update must now be adjusted so that the overall scheme obeys Bayes’ law. By rearranging eq. (3) it can be shown [3, e.g.] that the  $\ell$ th particle weight drawn from the proposal distribution satisfies  $w_i^{\ell} \propto P(y_i | u_{i-1}^{\ell}) w_{i-1}^{\ell}$  and is also Gaussian, so that

$$w_i^{\ell} \propto \exp \left[ -\frac{1}{2} (\mathbf{I}_i^{\ell})^T (\mathbf{H} \mathbf{Q} \mathbf{H}^T + \mathbf{R})^{-1} (\mathbf{I}_i^{\ell}) \right] w_{i-1}^{\ell} \quad (8)$$

where  $\mathbf{I}_i^{\ell} := y_i - \mathbf{H} \mathbf{F}_{i-1}(u_{i-1}^{\ell})$  is the *innovation vector*. Degeneracy as characterized by a single particle or a few particles with large weight still affects the OP-PF, but less than in many other particle filters. In [48] it is shown that among all PF techniques that obtain  $u_i^{\ell}$  using  $u_{i-1}^{\ell}$  and  $y_i$ , the ‘optimal proposal’ has the minimum variance in the weights and suffers the least from weight degeneracy. This was extended in [49] to any PF scheme that obtains  $u_i^{\ell}$  using the ensemble at the previous

time step  $u_{i-1}^{1:L}$  and  $y_i$ . The distributions used in the OP-PF are not always available, as the additive model error and the form of the linear observation operator are required to obtain closed forms for particle and weight update schemes.

However, in [3] it is shown that the optimal proposal requires an ensemble size  $L$  satisfying  $\log(L) \propto M \times D$  for a linear model, or will suffer from filter degeneracy. Thus, degeneracy is deeply connected to the dimension of model and observations, posing a fundamental stumbling block to the use of PFs in high dimensional problems.

### 3. Particle filters with dimension reduction in state and observation

#### 3.1. Dimension reduction of state and observation

Consider the physical model (1) and the data model (2). There are two routes to dimension reduction using these models, namely *physical model projection* and *data model projection*. Recall that the physical state of the system is given by  $\mathbf{u}_i \in \mathbb{R}^M$  and our observation data is given by  $\mathbf{y}_i \in \mathbb{R}^D$ . As previously discussed, the issue with geophysical models is that the dimensions of the physical and data space,  $M$  and  $D$  respectively, can be extremely large. This poses a problem with data assimilation methods, for example, particle filters, due to the constant need to re-draw and re-weight the particles. In that case, the benefit to assimilation is lost. We cannot move forward in time more than a few steps without re-sampling the set of particles, which overwrites previously gained information about the system. Lowering either the physical model dimension or the data model dimension helps in mitigating this problem. Here, we focus exclusively on linear, projection-based reduction of order, which amount to techniques for selecting the dimension and coordinates for the target subspace on which the models will be projected.

Starting with dimension reduction of the state, consider a matrix  $\mathbf{V}_i \in \mathbb{R}^{M \times M^q}$  whose columns form a time-dependent orthonormal basis ( $\mathbf{V}_i^T \mathbf{V}_i \equiv \mathbf{I}$ ) for the  $M^q$ -dimensional subspace of the state space. *Dimension reduction* or, simply, *reduction* of the state vector  $\mathbf{u}_i$  is given by inner products with columns of  $\mathbf{V}_i$ , which can be interpreted as the matrix multiplication  $\mathbf{v}_i^T : \mathbb{R}^M \rightarrow \mathbb{R}^{M^q}$ :

$$\mathbf{v}_i = \mathbf{V}_i^T \mathbf{u}_i, \quad \mathbf{v}_i \in \mathbb{R}^{M^q}. \quad (9)$$

Since typically  $M^q \ll M$ , this operation is not invertible.

The *reconstruction*  $\mathbf{v}_i : \mathbb{R}^{M^q} \rightarrow \mathbb{R}^M$  generates the state  $\mathbf{v}_i^r$  corresponding to  $\mathbf{v}_i$

$$\mathbf{v}_i^r := \mathbf{V}_i \mathbf{v}_i = \mathbf{V}_i \mathbf{V}_i^T \mathbf{u}_i \quad (10)$$

$\mathbf{v}_i^r$  is an element of the full state space  $\mathbb{R}^M$ , restricted to the subspace spanned by columns of  $\mathbf{V}_i$ , or the span of  $\mathbf{V}_i$ . Due to orthogonality ( $\mathbf{V}_i^T \mathbf{V}_i = \mathbf{I}$ ), reducing the reconstruction recovers the reduced state  $\mathbf{V}_i^T (\mathbf{V}_i \mathbf{v}_i) = \mathbf{v}_i$ , however computing the reconstruction after a reduction does not recover the state  $\mathbf{u}_i$  itself, since  $\mathbf{u}_i \neq \mathbf{V}_i \mathbf{V}_i^T \mathbf{u}_i$ , unless the  $\mathbf{u}_i \in \text{span } \mathbf{V}_i$  to begin with. In general, (10) computes the element of span  $\mathbf{V}_i$  nearest to  $\mathbf{u}_i$

$$\mathbf{v}_i^r = \arg \min_{\mathbf{x} \in \text{span } \mathbf{V}_i} \|\mathbf{x} - \mathbf{u}_i\|_2. \quad (11)$$

The transformation  $\mathbf{\Pi}_i^p : \mathbb{R}^M \rightarrow \mathbb{R}^M$  whose matrix is given by

$$\mathbf{\Pi}_i^p = \mathbf{V}_i \mathbf{V}_i^T, \quad (12)$$

is the orthogonal projection onto the span  $\mathbf{V}_i$ . In certain applications, the projection matrix may be given first, in which case an associated reduction matrix  $\mathbf{V}_i$  can be computed via Singular Value Decomposition (SVD) of  $\mathbf{\Pi}_i^p$ .

To evolve reduced states  $\mathbf{v}_i$  using the physical model, we first reconstruct the state to form  $\mathbf{V}_i \mathbf{v}_i$ , apply (1) to it, and then reduce the output using  $\mathbf{V}_i^T$ :



$$\mathbf{v}_{t+1} = \mathbf{V}_{t+1}^\top (\mathbf{F}_t(\mathbf{V}_t \mathbf{v}_t) + \boldsymbol{\omega}_t) = \mathbf{V}_{t+1}^\top \mathbf{F}_t(\mathbf{V}_t \mathbf{v}_t) + \mathbf{V}_{t+1}^\top \boldsymbol{\omega}_t \equiv \mathbf{F}_t^q(\mathbf{v}_t) + \boldsymbol{\omega}_t^q. \quad (13)$$

Since the orthogonal map of Gaussian random variables results in Gaussian outputs,  $\boldsymbol{\omega}_t^q \sim \mathcal{N}(0, \mathbf{Q}_t^q)$ , where  $\mathbf{Q}_t^q = \mathbf{V}_{t+1}^\top \mathbf{Q}_t \mathbf{V}_{t+1}$ . To this evolution, in the reduced-dimensional space  $\mathbb{R}^{M^q}$ , corresponds the evolution in the full space  $\mathbb{R}^M$ :

$$\boldsymbol{\Pi}_{t+1}^p \mathbf{u}_{t+1} = \boldsymbol{\Pi}_{t+1}^p \mathbf{F}_t(\boldsymbol{\Pi}_t^p \mathbf{u}_t) + \boldsymbol{\Pi}_{t+1}^p \boldsymbol{\omega}_t. \quad (14)$$

The observation space can be similarly reduced using another set of vectors  $\mathbf{U}_t$ . Here, we follow [12] in assuming that  $\mathbf{U}_t$  is a  $M \times D^q$  matrix, that is the reduction of the dimension still acts on the state space, although we will use it to reduce the observation. This allows for the comparison of model-based order reduction methods and data-driven order reduction methods on an equal footing. In this way, it is possible to use the same procedure to derive both  $\mathbf{U}_t$  and  $\mathbf{V}_t$ , although this is not necessary.

We start by defining the reduction of the observation space as

$$\mathbf{z}_t := \mathbf{U}_t^\top (\mathbf{H}^\dagger \mathbf{y}_t), \quad (15)$$

where  $\dagger$  denotes the Moore-Penrose pseudoinverse. Working with  $\mathbf{H}^\dagger \mathbf{y}_t$  instead of simply  $\mathbf{y}_t$  allows for the use of  $\mathbf{U}_t^\top$  whose input space is the state space, as explained above. Since we assume that  $\mathbf{H}$  has a full row-rank, the pseudoinverse  $\mathbf{H}^\dagger$  is an injection, so no information is lost in the process.

Applying this reduction to observations of states constrained to the model-reduced subspace  $\mathbf{y}_t = \mathbf{H}(\mathbf{V}_t \mathbf{v}_t) + \boldsymbol{\eta}_t$  results in the observation equation

$$\mathbf{z}_t = \mathbf{U}_t^\top \mathbf{H}^\dagger [\mathbf{H}(\mathbf{V}_t \mathbf{v}_t) + \boldsymbol{\eta}_t] = \underbrace{\mathbf{U}_t^\top \mathbf{H}^\dagger \mathbf{H} \mathbf{V}_t}_{\mathbf{H}_t^q} \mathbf{v}_t + \underbrace{\mathbf{U}_t^\top \mathbf{H}^\dagger}_{\boldsymbol{\eta}_t^q} \boldsymbol{\eta}_t \quad (16)$$

The composition  $\mathbf{H}^\dagger \mathbf{H} = \boldsymbol{\Pi}_H$  is the projection on the row (input) space of the data operator, that is the observable subspace of the state space. The reduced data operator  $\mathbf{H}_t^q : \mathbb{R}^{M^q} \rightarrow \mathbb{R}^{D^q}$ ,

$$\mathbf{H}_t^q := \mathbf{U}_t^\top \boldsymbol{\Pi}_H \mathbf{V}_t \quad (17)$$

is therefore the composition of the model reconstruction, the projection to the observable subspace space, and the data reduction. The reduced noise is again Gaussian  $\boldsymbol{\eta}_t^q \sim \mathcal{N}(0, \mathbf{R}_t^q)$ , with  $\mathbf{R}_t^q = \mathbf{U}_t^\top \mathbf{H}^\dagger \mathbf{R} (\mathbf{H}^\dagger)^\top \mathbf{U}_t$ .

The alternative to the state-space based reduction of observations is to directly employ a reduction of the observation space, via the matrix  $\mathbf{U}_t \in \mathbb{R}^{D \times D^q}$ , and reduction  $\mathbf{U}_t^\top : \mathbb{R}^D \rightarrow \mathbb{R}^{D^q}$ . This avoids the need to first pull-back observations into the state space by  $\mathbf{H}^\dagger$  as in (15), allowing for the reduction analogous to (9):

$$\mathbf{z}_t := \mathbf{U}_t^\top \mathbf{y}_t, \quad (18)$$

leading directly to the reduced model

$$\mathbf{z}_t = \mathbf{U}_t^\top [\mathbf{H}(\mathbf{V}_t \mathbf{v}_t) + \boldsymbol{\eta}_t] = \underbrace{\mathbf{U}_t^\top \mathbf{H} \mathbf{V}_t}_{\mathbf{H}_t^q} \mathbf{v}_t + \underbrace{\mathbf{U}_t^\top}_{\boldsymbol{\eta}_t^q} \boldsymbol{\eta}_t, \quad (19)$$

with the Gaussian reduced noise  $\boldsymbol{\eta}_t^q \sim \mathcal{N}(0, \mathbf{R}_t^q)$ , with  $\mathbf{R}_t^q = \mathbf{U}_t^\top \mathbf{R} \mathbf{U}_t$ .

In summary, the original state and observation equations eq. (1) and eq. (2) are replaced by reduced order equations through the use of orthogonal system matrices  $\mathbf{V}_t$  and  $\mathbf{U}_t$ ,

$$\mathbf{v}_{t+1} = \mathbf{F}_t^q(\mathbf{v}_t) + \boldsymbol{\omega}_t^q, \quad \mathbf{F}_t^q(\mathbf{v}) = \mathbf{V}_{t+1}^\top \mathbf{F}_t(\mathbf{V}_t \mathbf{v}), \quad (20)$$

$$\boldsymbol{\omega}_t^q \sim \mathcal{N}(0, \mathbf{Q}_t^q) \quad \mathbf{Q}_t^q = \mathbf{V}_{t+1}^\top \mathbf{Q}_t \mathbf{V}_{t+1}, \quad (21)$$

$$\mathbf{z}_t = \mathbf{H}_t^q \mathbf{v}_t + \boldsymbol{\eta}_t^q, \quad \boldsymbol{\eta}_t^q \sim \mathcal{N}(0, \mathbf{R}_t^q),$$

where the two options for the reduced data model are

$$\mathbf{H}_t^q = \mathbf{U}_t^\top \mathbf{H}^\dagger \mathbf{H} \mathbf{V}_t, \quad \mathbf{R}_t^q = \mathbf{U}_t^\top \mathbf{H}^\dagger \mathbf{R} (\mathbf{H}^\dagger)^\top \mathbf{U}_t, \quad \text{when } \mathbf{U}_t^\top : \mathbb{R}^M \rightarrow \mathbb{R}^{D^q}, \quad (22a)$$

or

$$\mathbf{H}_t^q = \mathbf{U}_t^\top \mathbf{H} \mathbf{V}_t, \quad \mathbf{R}_t^q = \mathbf{U}_t^\top \mathbf{R} \mathbf{U}_t, \quad \text{when } \mathbf{U}_t^\top : \mathbb{R}^D \rightarrow \mathbb{R}^{D^q}. \quad (22b)$$

The choice between two options is decided by whether  $\mathbf{U}_t^\top$  is developed as a model-based reduction, or an observation-based reduction. Additionally, the projected optimal proposal particle filter is modified to take the reduced state variable  $\mathbf{V}_t$  as the input by lifting it to the full space and then applying the observation model to it

$$\mathbf{y}_t = \mathbf{H} \mathbf{V}_t \mathbf{v}_t + \boldsymbol{\eta}_t. \quad (23)$$

The orthonormal bases  $\mathbf{V}_t$  and  $\mathbf{U}_t$  can be obtained from different dimension reduction techniques, in particular, POD, DMD, and AUS, described in more detail in section 4. Since AUS is exclusively a model-based reduction, in this work we employ only the first alternative (22a), to allow for a direct comparison between AUS and the others. Regardless of how they are computed, the equations (22) are used to formulate projected versions of particle filters.

### 3.2. Projected Particle Filter (Proj-PF)

Using the formulated projected models in the previous section, we can now formulate projected versions of the standard particle filter and the optimal proposal particle filter. The basic idea is to use either the original physical model (1) or the physical model-based projection only, (13) together with (23), for the particle update and the full projected models (13) and (16) for the weight update. We also will employ a projected resampling technique based on both the physical model and data model projections. Let  $\mathbf{v}_t^\ell := \mathbf{V}_t^\top \mathbf{u}_t^\ell$  for  $\ell = 1, \dots, L$  denote the  $\ell$ th projected particle at time  $t$ .

The following formulations detail the alterations to the particle update and weight update routines of the particle filter utilized in our projected particle filter:

- **Particle update:** Use (13) to form

$$\mathbf{v}_t^\ell = \mathbf{F}_{t-1}^q(\mathbf{v}_{t-1}^\ell) + \boldsymbol{\omega}_t^q, \quad \ell = 1, \dots, L. \quad (24)$$

- **Weight update:** Using the projected data model (16) or (21), (22a)

$$w_t^\ell \propto \exp\left(-\frac{1}{2}(\mathbf{I}_t^\ell)^\top (\mathbf{R}_t^q)^{-1} (\mathbf{I}_t^\ell) w_{t-1}^\ell\right), \quad \ell = 1, \dots, L, \quad (25)$$

$$\text{where } \mathbf{I}_t^\ell := \mathbf{z}_t - \mathbf{H}_t^q \mathbf{v}_t^\ell.$$

### 3.3. Projected Optimal Proposal Particle Filter (Proj-OP-PF)

In addition to a projected particle filter, we also employ a projected OP-PF. Accordingly, the following formulations detail the alterations to the particle update and weight update routines of the OP-PF utilized in our projected OP-PF:

- **Particle update:** Use the optimal proposal particle update (5), (6), (7) applied to the projected physical model (13) together with the corresponding data model (23) to form

$$\mathbf{v}_t^\ell = \mathbf{m}_t^\ell + \boldsymbol{\varphi}, \quad \boldsymbol{\varphi} \sim \mathcal{N}(0, \mathbf{Q}_p) \quad (26)$$

where

$$\mathbf{Q}_p^{-1} = (\mathbf{Q}_t^q)^{-1} + (\mathbf{H} \mathbf{V}_t)^\top \mathbf{R}^{-1} (\mathbf{H} \mathbf{V}_t), \quad (27)$$

$$\mathbf{m}_t^\ell = \mathbf{F}_{t-1}^q(\mathbf{v}_{t-1}^\ell) + \mathbf{Q}_p (\mathbf{H} \mathbf{V}_t)^\top \mathbf{R}^{-1} (\mathbf{y}_t - \mathbf{H} \mathbf{V}_t \mathbf{F}_{t-1}^q(\mathbf{v}_{t-1}^\ell)). \quad (28)$$

Alternatively, the particle update corresponding to the unprojected physical model ( $\mathbf{V}_t \equiv \mathbf{I}$ ) could be employed.

- **Weight update:** We employ the projected physical model (13) and either the state space based projected data model (16) or the observation space based projected data model (21), (22a), their corresponding covariance matrices  $\mathbf{Q}_t^q = \mathbf{V}_t^\top \mathbf{Q}_t \mathbf{V}_t$  and  $\mathbf{R}_t^q =$

$\mathbf{U}_i^\top \mathbf{H}^\top \mathbf{R}_i (\mathbf{H}^\dagger)^\top \mathbf{U}_i$  or  $\mathbf{R}_i^q = \mathbf{U}_i^\top \mathbf{R}_i \mathbf{U}_i$ , respectively, and the projected observation operator,  $\mathbf{H}_i^q = \mathbf{V}_i^\top \mathbf{H}^\dagger \mathbf{H} \mathbf{V}_i$  or  $\mathbf{H}_i^q = \mathbf{U}_i^\top \mathbf{H} \mathbf{V}_i$ , respectively. Form the matrix

$$\mathbf{Z}_i^q := (\mathbf{H}_i^q) \mathbf{Q}_i^q (\mathbf{H}_i^q)^\top + \mathbf{R}_i^q \quad (29)$$

and then update the weights as

$$w_i^\ell \propto \exp[-\frac{1}{2} (\mathbf{I}_i^\ell)^\top (\mathbf{Z}_i^q)^{-1} (\mathbf{I}_i^\ell)] w_{i-1}^\ell, \quad \ell = 1, \dots, L, \quad (30)$$

where  $\mathbf{I}_i^\ell := \mathbf{z}_i - \mathbf{H}_i^q \mathbf{v}_i^\ell$ .

We employ an extension of the projected resampling scheme proposed in [12]. When the Effective Sample Size (ESS), given by

$$\text{ESS} = \frac{(\sum_{\ell=1}^L w^\ell)^2}{\sum_{\ell=1}^L (w^\ell)^2}, \quad (31)$$

falls below a threshold (e.g.,  $\text{ESS} < \frac{1}{2}L$ ), then we resample. For a given  $\alpha \in [0, 1]$ , noise of the following form is added to resampled particles

$$\mathbf{V}_i^\top [\alpha \mathbf{U}_i \mathbf{U}_i^\top + (1 - \alpha) \mathbf{I}] \boldsymbol{\eta} \quad (32)$$

with  $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \omega \mathbf{I})$ , where  $\omega \geq 0$  is a tuneable parameter. The pseudocode summary of the algorithm is given in Algorithm 1.

---

**Algorithm 1:** Projected Optimal Proposal Particle Filter (Proj-OP-PF).

---

```

 $\alpha \leftarrow$  user input;
 $\omega \leftarrow$  user input;
for  $t = 1, \dots, T$  do
  for  $\ell = 1, \dots, L$  do
     $\mathbf{m}_i^\ell = \mathbf{F}_{i-1}^\top (\mathbf{v}_{i-1}^\ell) + \mathbf{Q}_p \{ \mathbf{H} \mathbf{V}_{i-1}^\top \mathbf{R}^{-1} [\mathbf{y}_i - \mathbf{H} \mathbf{V}_{i-1} \mathbf{F}_{i-1} (\mathbf{v}_{i-1}^\ell)] \}$ ;
     $\mathbf{v}_i^\ell \leftarrow \mathbf{m}_i^\ell + \boldsymbol{\varphi}, \quad \boldsymbol{\varphi} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_p)$ ;
     $w_i^\ell \propto \exp[-\frac{1}{2} (\mathbf{I}_i^\ell)^\top (\mathbf{Z}_i^q)^{-1} (\mathbf{I}_i^\ell)] w_{i-1}^\ell$ ;
  end
  if  $\text{ESS} < \frac{1}{2}L$  // Resample if ESS below threshold
  then
     $\mathbf{V}_i^\top [\alpha \mathbf{U}_i \mathbf{U}_i^\top + (1 - \alpha) \mathbf{I}] \boldsymbol{\eta}, \quad \boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \omega \mathbf{I})$ ;
  end
end

```

---

Significant reductions in computational complexity may be achieved via the reduced state space and the observation space dimensions employed in the Proj-OP-PF algorithm developed here. If the number of particles  $L$  is fixed, then in unprojected OP-PF the innovation vectors  $\mathbf{I}_i^\ell$  are multiplied by  $\mathbf{Q}_p \mathbf{H}^\top \mathbf{R}^{-1}$  for the particle updates and by  $(\mathbf{H} \mathbf{Q}_p^\top + \mathbf{R})^{-1}$  in the weight update. If these matrices are independent of time, then the cost becomes the cost of multiplying the  $L$  innovation vectors by these matrices, operations of order  $\mathcal{O}(MDL)$  and  $\mathcal{O}(M^2L)$ , respectively. Similarly, for Proj-OP-PF if the projections do not depend on time, then the computational cost of the operations is of order  $\mathcal{O}(D^q M^q L)$  and  $\mathcal{O}((D^q)^2 L)$ , respectively. If the projections depend on time, as with the AUS projection, then there is an additional cost of forming or factoring matrices, e.g., using LU or Cholesky, in addition to multiplication by the innovation vectors.

#### 4. Techniques for model reduction

Development of particle filters on a subspace of the state or observation, detailed in section 3, does not depend on any particular technique for computing the dimension reduction matrices  $\mathbf{V}_i$  and  $\mathbf{U}_i$ . In this section we outline three techniques for computing these matrices. POD and DMD are data-driven (model-free) techniques that only require a set of simulation snapshots to calculate the reduction subspace, while the Lyapunov Vectors (LV) computation requires access to derivatives of the deterministic part of the model update equation  $\mathbf{F}$  (see (1)).

##### 4.1. Proper Orthogonal Decomposition (POD)

Proper Orthogonal Decomposition (POD) is the model-reduction technique based on computation of a parsimonious orthogonal basis for the state space subspace occupied by a given evolution a dynamical system. It is ubiquitous in applied mathematics, and in other contexts is known as principal component analysis, Karhunen–Lo  ve decomposition, and empirical orthogonal function decomposition. Here, we review only the necessary about POD; an excellent short review of the main features with a wealth of references can be found in [50, §22.4]. In the context of fluid dynamics [51], the orthogonal basis is calculated using eigenvectors of the cross-correlation matrix of the simulated data.

Given a recording of evolution of state vectors<sup>2</sup> (called *snapshots*)  $\mathbf{u}_t \in \mathbb{R}^M$ , over time  $t = 1, \dots, T$ , stored as a *snapshot matrix*

$$\mathbf{X} := [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \dots \quad \mathbf{u}_T], \quad (33)$$

POD amounts to a separation-of-variables ansatz

$$\mathbf{u}_t \approx \sum_{m=1}^M \boldsymbol{\phi}_m \sigma_m \boldsymbol{\psi}_{t,m}. \quad (34)$$

Here vectors  $\boldsymbol{\phi}_m$ ,  $m = 1, \dots, M$ , are the normalized ‘‘spatial’’ profiles of the state  $\mathbf{u}_t \in \mathbb{R}^M$ , vectors  $\boldsymbol{\psi}_{t,m} := [\psi_{1,m} \quad \psi_{2,m} \quad \dots \quad \psi_{T,m}]^\top$  are the normalized time evolutions, while  $\sigma_m$  are the linear combination coefficients, i.e., magnitudes. While there are many possible separation-of-variable decompositions, POD is specified by the requirement that  $\{\boldsymbol{\phi}_m\}_{m=1}^M$  and  $\{\boldsymbol{\psi}_m\}_{m=1}^M$  should be orthogonal sets.

In matrix notation, and over a fixed period of time, this ansatz corresponds to Singular Value Decomposition (SVD) of the snapshot matrix  $\mathbf{X}$

$$\mathbf{X} = [\boldsymbol{\phi}_1 \quad \boldsymbol{\phi}_2 \quad \dots] \begin{bmatrix} \sigma_1 & \sigma_2 & \dots \\ & & \ddots \end{bmatrix} [\boldsymbol{\psi}_1 \quad \boldsymbol{\psi}_2 \quad \dots]^\top. \quad (35)$$

The rank of the snapshot matrix  $\mathbf{X}$  is equal to the number of nonzero singular values,  $\sigma_m$ . It is common to order the singular values in decreasing values, and refer to those  $\sigma_m$  and vectors  $\boldsymbol{\phi}_m$  as *dominant* if they have a low index. Furthermore, singular values that are equal to zero, and associated singular vectors, are sometimes omitted to form the ‘‘economy’’ version of SVD.

To reduce the dimension of  $\mathbf{X}$ , while preserving the character of dynamics, the reduction matrix  $\mathbf{V}$  (see (9)) is formed,

$$\mathbf{V}^{(r)} = [\boldsymbol{\phi}_1 \quad \dots \quad \boldsymbol{\phi}_r], \quad (36)$$

containing the first  $r < M$  left singular vectors  $\boldsymbol{\phi}_m$  (spatial profiles). By the Eckart–Young theorem [52], the projected snapshot matrix  $\mathbf{V}^{(r)} \mathbf{V}^{(r)\top} \mathbf{X}$  is the best approximation of  $\mathbf{X}$  among all matrices of rank  $r$  as measured by the Frobenius norm, i.e., element-wise  $\ell^2$  norm [53, §2.4].

In general, the choice  $r$  that obtains a parsimonious, yet usable, reduced-order approximation can be problem dependent, although there are prescriptions of optimal rank in absence of problem-dependent guidances [54]. Ideally, a gap or a jump in a singular value plot is an indication that there is a sharp change in the approximation error as the number of dimensions retained is changed across the gap. In other cases, no such gap may be seen, which can be in certain cases traced to model-agnostic application of the technique [55, §19.4].

##### 4.2. Dynamic Mode Decomposition (DMD)

Dynamic Mode Decomposition (DMD) [56,57] provides a route to order reduction by approximating the evolution of snapshots  $\mathbf{u}_t$  by a separation-of-variables ansatz in the form

<sup>2</sup> If the data-based model reduction is needed, as in (22b), then evolution of observations should be used instead.

**Algorithm 2:** Data projection using Proper Orthogonal Decomposition (POD).

---

```

 $\mathbf{X} \leftarrow [\mathbf{u}_1 \ \mathbf{u}_2 \ \dots \ \mathbf{u}_T]$  // Form snapshot matrix
 $\mathbf{X} = \Phi \Sigma \Psi^T$  // Singular Value Decomposition
 $r \leftarrow$  user input
 $\mathbf{V}_{\text{POD}} \leftarrow [\phi_1 \ \phi_2 \ \dots \ \phi_r]$  // Dimension reduction matrix

```

---

$$\mathbf{u}_i \approx \sum_{m=1}^M \mathbf{v}_m e^{i\omega_m \tau_M} b_m, \quad (37)$$

where  $\tau_M$  is the timestep separating adjacent snapshots,  $\mathbf{v}_m$  are DMD modes, corresponding to a spatial profile of the component of dynamics,  $\omega_m \in \mathbb{C}$  are complex-valued DMD frequencies governing growth, decay, and oscillation of time evolution, while  $b_m \in \mathbb{C}$  are linear combination coefficients. For real-valued input vectors  $\mathbf{u}_i$ , complex-valued DMD frequencies and modes come in conjugate pairs.

The primary assumption is that there exists a high-dimensional time-invariant matrix  $\mathbf{A}$  that relates pairs of snapshots

$$\mathbf{u}_{i+1} = \mathbf{A} \mathbf{u}_i. \quad (38)$$

While this assumption does not hold exactly, the Koopman operator theory [58–62] asserts that for time-invariant systems, including (1), the state  $\mathbf{u}_i$  can be embedded in an infinite-dimensional space on which the matrix  $\mathbf{A}$  is the linear infinite-dimensional Koopman operator and the equation analogous to (37) does hold for expected values of these state variables.

If (38) holds and we have access to  $\mathbf{A}$  then modes  $\mathbf{v}$  can be computed as eigenvectors of  $\mathbf{A}$

$$\mathbf{A} \mathbf{v}_m = \lambda_m \mathbf{v}_m, \quad (39)$$

while frequencies  $\omega_m$  are derived from eigenvalues  $\lambda_m$  by the formula

$$\lambda_m = \exp(i\omega_m \tau_M), \quad (40)$$

where  $\tau_M$  again is the timestep that separates adjacent snapshots  $\mathbf{u}_m$ . As is well-known from linear systems analysis,  $\text{Re}(\omega_m)$  determines the rate of exponential decay ( $\text{Re}(\omega_m) < 0$ ) or growth ( $\text{Re}(\omega_m) > 0$ ) of DMD modes, while  $\text{Im}(\omega_m)$  corresponds to the angular frequency of oscillations. Modes with frequencies at the origin  $\omega_m = 0$  correspond to constant components in the evolution.

Typically, though, we do not have access to  $\mathbf{A}$  directly and we need to approximate its eigenvalues and eigenvectors. DMD is a family of numerical procedures approximates  $\mathbf{v}_m$  and  $\omega_m$  while avoiding both the explicit embedding of (38) in the high-dimensional space, and the computation and storage of the full matrix  $\mathbf{A}$ , which in practice is prohibitively large, and theoretically infinite. Here we present the so-called exact DMD algorithm [59] which is commonly a starting point for the DMD analysis.

Regression of the discrete dynamics equation (38) onto the snapshots can be solved approximately for all adjacent time steps as an  $\ell^2$  minimization problem

$$\min_{\mathbf{A}} \sum_{i=0}^{T-1} \|\mathbf{u}_{i+1} - \mathbf{A} \mathbf{u}_i\|_2^2 = \min_{\mathbf{A}} \|\mathbf{X}_2 - \mathbf{A} \mathbf{X}_1\|. \quad (41)$$

The second formula is the equivalent matrix notation using involving two submatrices of the snapshot matrix,  $\mathbf{X}_1$  and  $\mathbf{X}_2$ , formed by respectively erasing the first and the last column of  $\mathbf{X}$ . In principle, this problem could be solved by a Moore–Penrose inverse as  $\mathbf{A}_{\text{DMD}} := \mathbf{X}_2 \mathbf{X}_1^\dagger$ , but this is generally avoided as the size of  $\mathbf{A}_{\text{DMD}}$  is quadratic in the dimension of snapshot vectors, and in practice may be prohibitively large store. Furthermore, any numerical errors arising in the process could fatally affect the well-posedness of the computation.

As the ultimate goal is not the calculation of  $\mathbf{A}$  but rather a (small) subset of its eigenvectors and eigenvalues (39), most variants of DMD employ an order reduction step to improve numerical robustness and

reduce the size of numerical linear algebra calculations. Here, we use a POD based order reduction, essentially using section 4.1 as the substep of the DMD analysis. Compute SVD of the “left” snapshot matrix

$$\mathbf{X}_1 = \Phi \Sigma \Psi^T \quad (42)$$

and truncate the involved matrices to  $R$  dominant vectors, forming  $\Phi_R$ ,  $\Sigma_R$ ,  $\Psi_R$ . Since the goal of this step is not to perform the full order reduction, but merely to numerically stabilize the problem and reduce the size of numerical problems solved below,  $R$  could be fairly large, even  $R = 0.9M$ .

Using the matrices just computed, form

$$\mathbf{A}_R = \Phi_R \mathbf{X}_2 \Psi_R \Sigma_R^{-1}, \quad (43)$$

and compute its eigenvalues and eigenvectors  $\mathbf{A}_R \hat{\mathbf{v}}_m = \lambda_m \hat{\mathbf{v}}_m$ . Matrix  $\mathbf{A}_R$  is a  $R \times R$  matrix whose eigenvalues are the same as eigenvalues of the  $M \times M$  matrix  $\mathbf{A}$ , and whose eigenvectors  $\hat{\mathbf{v}}_m$  can be used to reconstruct DMD modes by

$$\mathbf{v}_m = \lambda_m^{-1} \mathbf{X}_2 \Psi_R \Sigma_R^{-1} \hat{\mathbf{v}}_m. \quad (44)$$

Depending on the algorithm to compute eigenvectors  $\hat{\mathbf{v}}_m$  may need to be normalized to unit  $\ell^2$  norm.

Combination coefficients  $b_m$  can be used to rank the DMD modes by importance. They can be solved by regression, that is solving (37) at the initial snapshot, or even many steps

$$\mathbf{b} \leftarrow \mathbf{Y}^\dagger \mathbf{u}_0, \quad \text{or} \quad \mathbf{b} \leftarrow \arg \min_{\mathbf{x} \in \mathbb{C}^M} \sum_{i=0}^T \left\| \mathbf{u}_i - \sum_{m=1}^M \mathbf{v}_m e^{i\omega_m \tau_M} x_m \right\|. \quad (45)$$

Solving (45) at only the initial state can result in a relative approximation error that is unevenly spread between growing and decaying modes; conversely, using all steps in the regression balances the error, but is numerically expensive. As a compromise, implementation used below uses five steps equally spaced across all available snapshots.

One advantage of using DMD is in additional information that can be used to choose what subset of modes to use for dimension reduction. POD modes are ranked solely by their  $L^2$  norms (singular values), therefore most approaches simply choose some number of dominant modes. A similar effect can be achieved by ranking DMD modes by absolute values  $|b_m|$ , which represent contributions of modes to the initial condition. Alternatively, DMD modes can be ordered by  $L^2$  norms of time evolution for each DMD mode

$$\bar{b}_m^2 = \int_0^T |e^{i\omega_m t} b_m|^2 dt = |b|^2 \frac{\exp(2 \text{Re}(\omega_m T)) - 1}{2 \text{Re}(\omega_m T)}, \quad \text{and} \quad \bar{b}_m^2 := |b_m|^2 \text{ if } \text{Re}(\omega_m) = 0, \quad (46)$$

which give the same weight to modes that grow and those that decay at rates, everything else being the same. In the applications below, we rank DMD modes in the descending order of  $\bar{b}_m$ , and refer to those with large  $\bar{b}_m$  as *dominant*.

Alternatively, DMD modes can be chosen based on the real or imaginary parts of  $\omega_m$ , e.g., if there is a reason to choose only modes corresponding to a certain frequency band. We did not pursue this direction further in this paper as we did not suspect that evolutions of either L96 or SWE were concentrated in a particular frequency band.

Choosing the  $M^q$  dominant DMD modes  $\mathbf{v}_m$ , the final step is to form the orthogonal projection matrix  $\Pi_{\text{DMD}}$ . The truncated  $\mathbf{Y}$  is not an orthogonal projection because DMD modes, unlike POD modes, do not form an orthonormal system. Additionally,  $\mathbf{v}_m$  are complex-valued, so conjugate pairs of columns  $\mathbf{v}_m, \mathbf{v}_{m+1} = \mathbf{v}_m^*$  should be replaced by their real-valued cartesian components  $\text{Re } \mathbf{v}, \text{Im } \mathbf{v}$  in  $\mathbf{Y}$  matrix. Care should be taken to always include either both element of a pair, or neither. After these steps that prepare a truncated real-valued basis for a subspace of DMD modes  $\hat{\mathbf{Y}}$ , we compute orthogonal dimension reduction matrix  $\mathbf{V}_{\text{DMD}}$  as left singular vectors of  $\hat{\mathbf{Y}}$ , or state projection matrix  $\Pi_{\text{DMD}}$  using QR decomposition of  $\hat{\mathbf{Y}}$ .

**Algorithm 3:** Dynamic Mode Decomposition.

---

```

 $X_1 \leftarrow [u_0 \ u_1 \ \dots \ u_{T-1}], X_2 \leftarrow [u_1 \ u_2 \ \dots \ u_T], // \text{ Left/Right}$ 
Snapshot Matrix
 $X_1 = \Phi \Sigma \Psi^T // \text{ SVD}$ 
 $R \leftarrow \text{user input} // \text{ Choice of the DMD problem size } R < \text{rank } X$ 
 $\Phi_R, \Sigma_R, \Psi_R // \text{ Truncation of SVD matrices}$ 
 $A_R \leftarrow \Phi_R X_2 \Psi_R \Sigma_R // \text{ Compressed DMD matrix}$ 
 $A_R \hat{D}_m = \lambda_m \hat{D}_m // \text{ Eigendecomposition of DMD matrix}$ 
 $v_m \leftarrow \lambda_m^{-1} X_2 \Psi_R \Sigma_R^{-1} \hat{D}_m // \text{ Computing and normalizing DMD modes}$ 
 $b = \arg \min_{x \in \mathbb{C}^M} \sum_{i=0}^T \|u_i - \sum_{m=1}^M v_m e^{i\omega_m T} x_m\| // \text{ DMD coefficients via } \ell^2$ 
regression
 $\hat{b}_m = |b|^2 [\exp(2\text{Re}\omega_m T) - 1] / (2\text{Re}\omega_m T) // \text{ Rank DMD modes in descending}$ 
order of  $b$ 
 $M^q \leftarrow \text{user input} // \text{ Choice of size of dimension reduction}$ 
subspace  $M^q < R$ 
 $\hat{Y} \leftarrow [Y_1 \ Y_2 \ \dots \ Y_{M^q}] // \text{ Truncate DMD modes and convert to}$ 
real vectors
 $V_{\text{DMD}} \leftarrow \text{SVD}(\hat{Y}) // \text{ Left singular vectors form the dimension}$ 
reduction matrix

```

---

**4.3. Assimilation in the Unstable Subspace (AUS)**

The last of the model reduction methods employed here, which is restricted to projecting in physical model space, is based on computational techniques for Lyapunov exponents and Finite Time Lyapunov exponents. Again, calculated modes are used to select which dimensions are most dynamically significant and should be retained in the reduced basis. In AUS, these modes are determined by employing the discrete QR algorithm [63,64]. For the discrete time model  $u_{t+1} = F_t(u_t) + \omega_t$  with  $u_t \in \mathbb{R}^M$ , let  $U_0 \in \mathbb{R}^{M \times M^q}$  ( $M^q \leq M$ ) denote a random matrix such that  $U_0^T U_0 = I$  and

$$U_{t+1} T_t = F'_t(u_t) U_t \approx \frac{1}{\epsilon} [F_t(u_t + \epsilon U_t) - F_t(u_t)], \quad t = 0, 1, \dots \quad (47)$$

where  $U_{t+1}^T U_{t+1} = I$  and  $T_t$  is upper triangular with positive diagonal elements. With a finite difference approximation the cost is that of an ensemble of size  $p$  plus a reduced QR via modified Gram–Schmidt to re-orthogonalize. Lyapunov exponents or finite time approximations can be formed and monitored by taking time averages of the natural logarithm of the diagonal elements of the upper triangular  $p \times p$  matrices  $T_t$  (see, e.g., [63,64]). Time dependent orthogonal projections to decompose state space are given by  $\Pi_t = U_t U_t^T$  and can be employed to form both physical model and data model projections. While AUS traditionally refers to projection/restriction of the physical model onto the neutral and unstable subspace, here we use AUS more generally to refer to model-space-based projections using an approximate Lyapunov basis: accordingly, our implementation of AUS may include only some unstable modes, or may include the entire neutral-unstable subspace and some stable modes.

**5. Numerical results**

To evaluate the performance of Proj-OP-PF, we apply the presented techniques to two commonly used models: Lorenz'96 model (L96) and a configuration of the Shallow Water Equations (SWE) corresponding to a barotropic instability.

The experiments were chosen to evaluate how a particular choice of the order reduction technique, and the dimensions of reduced model and data dimensions, resp.  $M^q$  and  $D^q$ , influence the accuracy of assimilation, as well as protect the particle filter from weight collapse. In all cases, we use the model-based reduction of the observation space (see section 3.1) with a full row-rank linear observation operator, which allows for a fair comparison between order reduction techniques.

**5.1. Common experimental setup**

For both L96 and SWE we use a common setup to evaluate the effectiveness of the data assimilation scheme. We run one simulation of the

model that serves as the “truth”,  $u_t^{\text{truth}}$ . Noisy observations of the “truth” are used as inputs into assimilation, and the success of the assimilation is measured by how accurately the state of the estimator matches the state of the “truth” simulation. In all cases we use Proj-OP-PF as the data assimilation scheme, summarized in Algorithm 1. The scheme uses a set of  $L$  particles about the initial condition of the truth  $u^{\text{truth}}$  with added noise from the same Gaussian distribution employed to simulate noise in the physical model. All particles are initialized with equal weights ( $1/L$ ) and are propagated forward in time using the chosen model. The Effective Sample Size (ESS) is then calculated (31) and projected resampling is performed with the spread of particles governed by (32), where the proportion of resampling variance inside the projection subspace is always taken to be  $\alpha = 0.99$ , and total resampling variances  $\omega = 10^{-2}$ , for L96, and  $\omega = 10^{-4}$ , for SWE when  $\text{ESS} < \frac{1}{2}L$ .

To evaluate Proj-OP-PF we report on two quantities:

- Root Mean Squared Error (RMSE) between estimate of the state and the true state,

$$\text{RMSE}(u^{\text{truth}}, u^{\text{ens}}) := \|u^{\text{truth}} - u^{\text{ens}}\|_2 / \sqrt{M}, \quad (48)$$

where  $u^{\text{truth}}$  denotes the truth and  $u^{\text{ens}}$  denotes the particle ensemble mean, and

- Resampling Percentage (RES%), which measures the proportion of observation times in which the particle population needed to be resampled.

The lower each of the quantities are, the better the assimilation scheme is performing. We also report in some experiments on the projected RMSE,

$$\begin{aligned} \text{RMSE}^q(u^{\text{truth}}, u^{\text{ens}}) &:= \text{RMSE}(V V^T u^{\text{truth}}, V v^{\text{ens}}) \\ &= \|V V^T u^{\text{truth}} - V v^{\text{ens}}\|_2 / \sqrt{M^q}, \end{aligned} \quad (49)$$

which measures the error in the subspace of the reduced model. A low projected RMSE together with a significantly higher “full” RMSE (48) is an indication that the assimilation is being effective when restricted to the subspace of the reduced model, but the projected model does not sufficiently resolve the full model. In several of the figures, we will compare with the results of the optimal proposal particle filter with no model or data reduction and we will use (NON) to denote these results. All numerical results are obtained by averaging over 10 randomized trials.

**5.2. Lorenz '96 equations****5.2.1. Model and parameters**

We first consider Proj-OP-PF applied to the extensively used medium-dimensional dynamical system Lorenz'96 model (L96). This model, developed by Edward Lorenz [65], represents a nonlinear chaotic system that captures some multiscale features of the global horizontal circulation of the atmosphere. L96 has become one of the most commonly used test problems in data assimilation since its introduction.

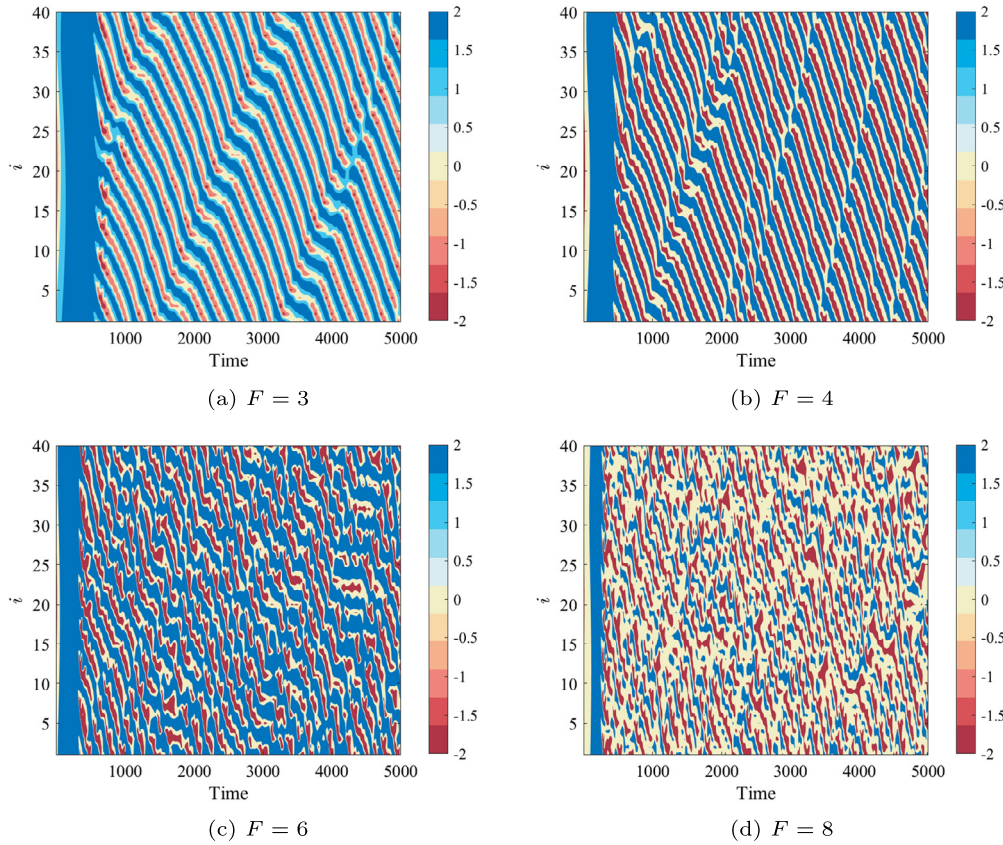
The model is presented as a system of ordinary differential equations (ODEs) in  $u = (u_i)_{i=1}^M$  of an arbitrary dimension  $M$ ,

$$\frac{du_i}{dt} = (u_{i+1} - u_{i-2}) u_{i-1} - u_i + F, \quad i = 1, \dots, M, \quad (50)$$

where the value of a constant (typically positive) forcing term  $F$  determines qualitatively whether the evolution will be regular or chaotic. In its original form, the model was introduced as a variable-order system of ODEs, but it can be interpreted as a 2nd order finite-difference discretization of a viscous Burgers-type equation with periodic boundary conditions (see [66] for derivation):

$$\partial_t u = -u \partial_x u - \frac{1}{3} (\partial_x u)^2 - \frac{1}{6} u \partial_{xx} u - u + F. \quad (51)$$





**Fig. 1.** Solutions  $u_i(t)$ ,  $i = 1, \dots, M \equiv 40$  of (50) demonstrating regular and chaotic spatiotemporal patterns in L96 model for various values of forcing  $F$ . Initial condition is a cosine bump in all cases.

In this form, the model has a nonlinear convective term corresponding to  $u\partial_x u$ , a diffusive term, and a dissipative term ( $-u$ ). Note, however, that the discretization does introduce nonlinear effects not present in the full model [67], so that the ODE–PDE correspondence is not straightforward.

To produce the true evolution  $u_i^{\text{truth}}$ , the L96 model (50) is evolved in time using the Dormand–Prince pair (MATLAB’s `ode45`), with solution resampled at multiples of the fixed time step  $\tau_M = 0.01$ . Fig. 1 illustrates the space-time behavior of typical solutions of L96 ranging from  $F = 3$  (regular structure) to  $F = 8$  (chaotic structure). For  $M = 40$  used in most calculations here, the onset of chaos is between  $F = 3$  and  $F = 4$  [67]; the similar behavior appears to hold for  $M = 400$  as well. In the numerical experiments, the initial condition is a random vector to reduce the burn-in time, and provide some variability between trials.

In the setup used here, all model variables are observed so that  $\mathbf{H} = \mathbf{I}$ . The noise is uncorrelated among any two variables of the system (within and between state and observation vectors). Moreover, it is uniform for all states and for all observation variables, which means that the correlation matrices  $\mathbf{Q}, \mathbf{R}$  are scaled identity matrices. We compare several levels of such noise in the physical model, yielding  $\mathbf{Q} = \alpha \cdot \mathbf{I}$  with scalar variances  $\alpha = 0.1, 1.0$ . The observation error covariance is always fixed to  $\mathbf{R} = 0.01 \mathbf{I}$  resulting in the standard deviation of observation error of 0.1 which is included for comparison in the figures when reporting on RMSE.

The number of particles is fixed with  $L = 20$ . The observations are computed every 5 steps, yielding the effective time step of assimilation  $\tau_D = 0.05$ . For the purposes of assimilation, the 4th order Dormand–Prince integrator is used but with a fixed step size  $\tau_M = 0.01$ . The assimilation is performed over 10,000 observation times, after 1000 time steps have elapsed. The average RMSE over time is calculated based on the RMSE over the last 5,000 observation times (see Fig. 5 where the last 5,000 observation times correspond to the second half of

**Table 1**

Lyapunov dimensions for the L96 model for various forcing values and model dimensions  $M$ , estimated by the Kaplan–Yorke formula (52).

$F$	3	4	6	8
$D_L$ ( $M = 40$ )	1	3	22	28
$D_L$ ( $M = 400$ )	5	12	224	270

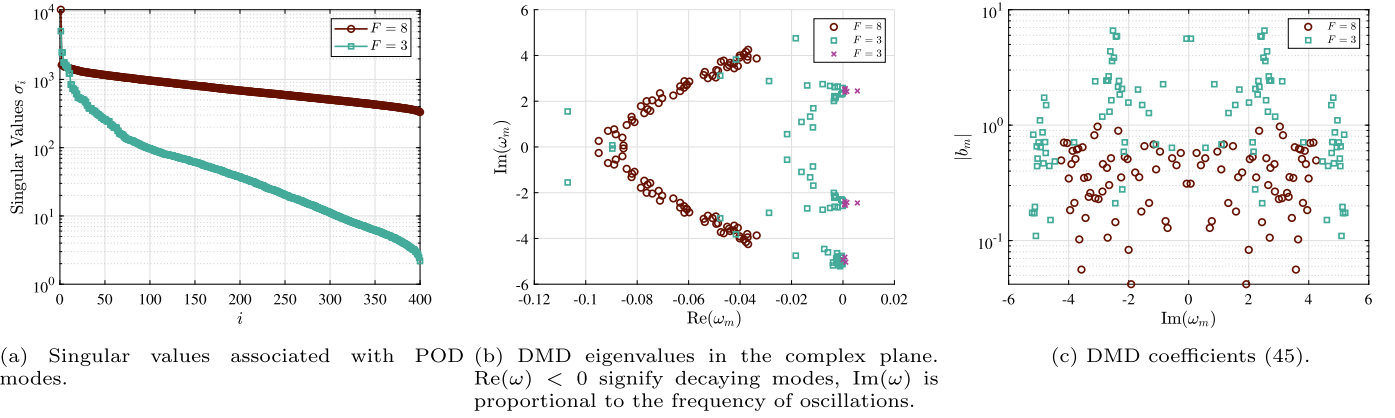
the assimilation window), to more accurately represent the asymptotic value of the RMSE. The resampling percentage RES% is computed as the proportion of all 10,000 observation times in which resampling was performed.

For the classical values of the L96 system where  $M = 40$  and  $F = 8$ , the system is chaotic with 13 positive and 1 neutral Lyapunov exponent. The percentage of positive Lyapunov exponents approximately scales with the dimension  $M$  and there are generally a smaller percentage of positive Lyapunov exponents for smaller values of  $F$ . For  $M = 40$ ,  $M = 400$  and  $F \in [3, 8]$  we approximate the Lyapunov dimension by Kaplan–Yorke formula  $D_L$  defined for ordered Lyapunov exponents  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_M$  given by

$$D_L = k + \frac{\lambda_1 + \lambda_2 + \dots + \lambda_k}{|\lambda_{k+1}|} \quad (52)$$

where  $k$  is the maximum value of  $i$  such that  $\lambda_1 + \lambda_2 + \dots + \lambda_i > 0$ . Approximations are computed over a time interval of length 10,000 with the integer part tabulated in Table 1 and indicates that for smaller values  $F$  the dynamics can be thought of as inherently low dimensional as  $D_L/M \ll 1$ , but for larger values of  $F$  this is not the case.

To determine suitable POD and DMD modes for L96, we simulate the model with a random initial condition (different from those used to compute the “truth”), over the entire time interval over which the



**Fig. 2.** Singular values and DMD eigenvalues spectra for L96 model with  $M = 400$ . Lower values of  $F$ , associated with more-regular behavior, demonstrate a faster decay of the singular values, and clustering of DMD frequencies. Higher values of  $F$ , associated with disordered behavior, have a slower decay of the singular values and more evenly distributed spectrum.

assimilation is performed at the model time step  $\tau_M = 0.01$ . Since the model evolution is largely in a steady state, both POD and DMD modes from such simulation can be used for the assimilation process. This setting can be justified in cases where the model evolution is largely in a steady state, that is, not undergoing a regime change. A more realistic setup could employ decompositions of assimilated data over some moving interval, which is something we will explore in future work. For L96 the POD and DMD projections are computed separately for each of the 10 trial using snapshots obtained from a random initial condition.

Order reduction techniques are effective when the dynamics can be reproduced by a relatively small number of modes. For POD, this is indicated by jumps (“gaps”) in the singular value spectrum. Additionally, sharp changes in slope between segments of the singular values plot indicate the presence of mechanisms evolving at different scales. Fig. 2 demonstrates that the regular  $F = 3$  evolution results in an “elbow” around subspace of dimension 70, while for  $F = 8$  there is no discernible boundary between system scales (elbows in Fig. 2a).

The duration of the time window over which POD and DMD are computed depends on the type of dynamics. For regular steady-state behavior, the time window should be large enough for the trajectory to trace out the orbit at least once. For irregular steady-state behavior, e.g., chaotic attractors, the time window should be long enough so that the points in the trajectory sample the attractor well. All parameters  $F$  considered for L96 lead to steady state behavior, although we leave the investigation of the effects of window duration for future work.

For transient behavior, e.g., trajectories shadowing heteroclinic orbits between two invariant sets, the duration of the window has to be chosen carefully. In those cases, a common strategy is to employ a sliding-window computation of POD or DMD [68,69], leading to time-varying reduction/projection matrices. The choice of the duration of the window leads to delicate effects, further explored in [70]. Theoretical connection between sliding window DMD and the stability theory of time-varying dynamical systems has been developed in [71–73]. A follow-up publication will investigate how time-varying projection matrices perform in particle filters, and we plan to explore these matters in more detail.

For DMD, regular steady-state dynamics are indicated by DMD frequencies  $\omega_m$ , see (40), concentrated along the vertical axis, and by a “peaked” graph of combination coefficients  $b$  against the frequency. Fig. 2 shows the singular values and DMD spectrum for the range of forcing parameters  $F$  distinguishing between DMD spectrum with positive and negative real parts for  $F = 3$  in Fig. 2b. It is evident that we can expect a better performance of order reduction techniques for lower values of  $F$ , which are associated with regular behavior of the spatiotemporal evolution.

**Table 2**

Parametrization of experiments used to test Proj-OP-PF for L96. In all cases, we set observation noise covariance to  $\mathbf{R} = 0.01 \mathbf{I}$ , number of particles  $L = 20$ , and observe all model variables  $\mathbf{H} = \mathbf{I}$ .

Exp.	$F$	$M = D$	$\mathbf{Q}$	Reduction	$M^q$	$D^q$
1	3,8	40	0.1I, 1.0I	AUS, POD, DMD	5 – 40	5
2	3,8	400	0.1I	POD	100	1 – 100
3	3,4,6,8	400	0.1I, 1.0I	POD, DMD	10–400	5

To evaluate the performance of Proj-OP-PF we conduct three experiments with the intent of comparing how three reduction schemes performed over a range of parameters. Details of experiments are given in Table 2.

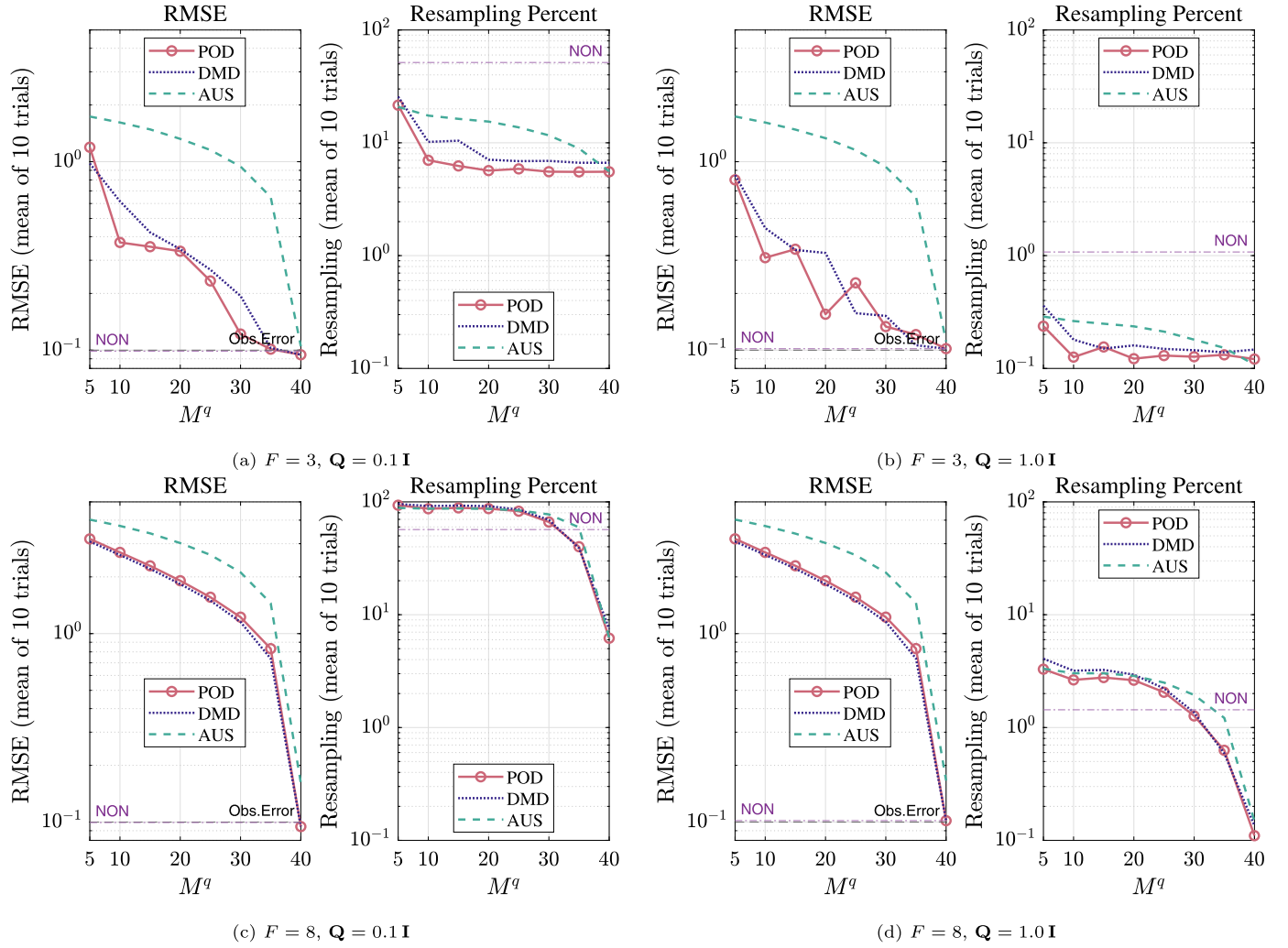
### 5.2.2. Experiment 1 ( $F = 3, 8$ , $M = 40$ )

Using AUS for reducing the order of the data model was investigated in [12,74]; here we investigate the efficacy when AUS is used for the reduction of both the physical and data models. Additionally, we compare AUS with order reductions derived using POD and DMD. Fig. 3 shows the mean RMSE and RES% trends with increasing model projection rank for the three projection methods. For all projection types the RMSE does not approach the observation noise until the rank of the projections are at least 35 although the RMSEs for POD and DMD are slightly lower than for AUS. With  $\mathbf{Q} = 1.0 \mathbf{I}$  the RES% are much smaller than for  $\mathbf{Q} = 0.1 \mathbf{I}$  since the optimal proposal particle filter with larger model error covariance leads to greater particle diversity and hence smaller RES%. We focus on dimension reduction using POD and DMD for the remaining experiments.

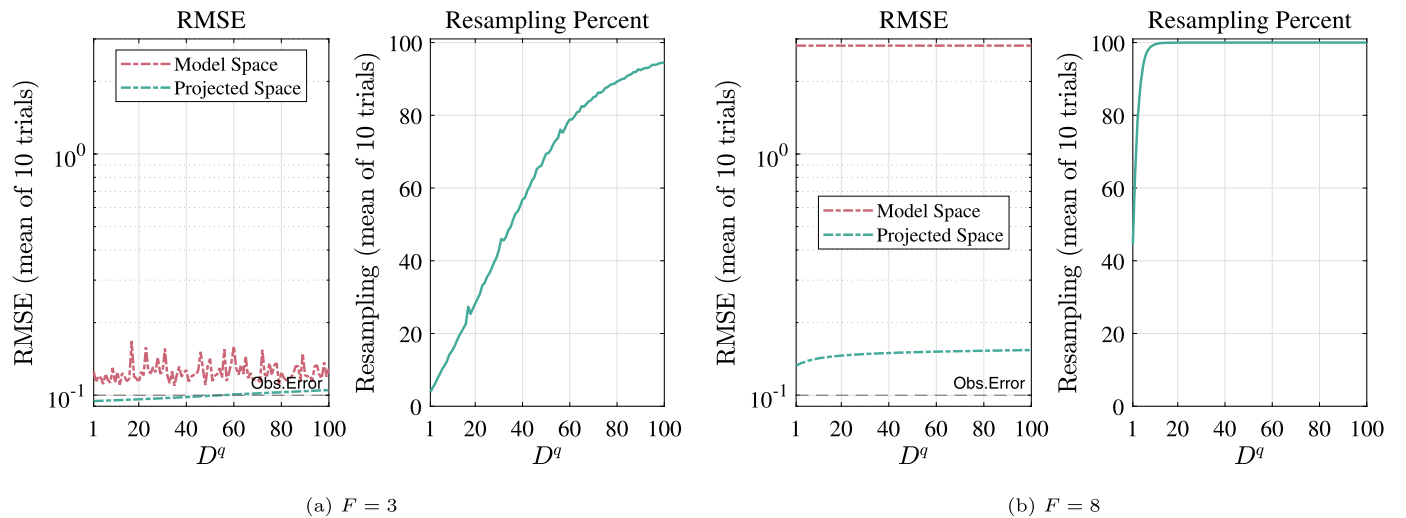
### 5.2.3. Experiment 2 ( $F = 3, 8$ , $M = 400$ )

We set the model dimension to  $M = 400$  and fix  $M^q = 100$ , then vary  $D^q$  between 1 and 100. The results are shown in Fig. 4. The RMSE and projected RMSE are both relatively constant over the range of data dimensions. The projected RMSE is at the level of standard deviation of observation noise, which indicates that the assimilation is effective in the states  $\mathbf{u}$  belonging to the subspace of the reduced model. The higher value of the “full” RMSE indicates that this is not sufficient to constrain the full state, meaning that the projected model does not sufficiently resolve the full model.

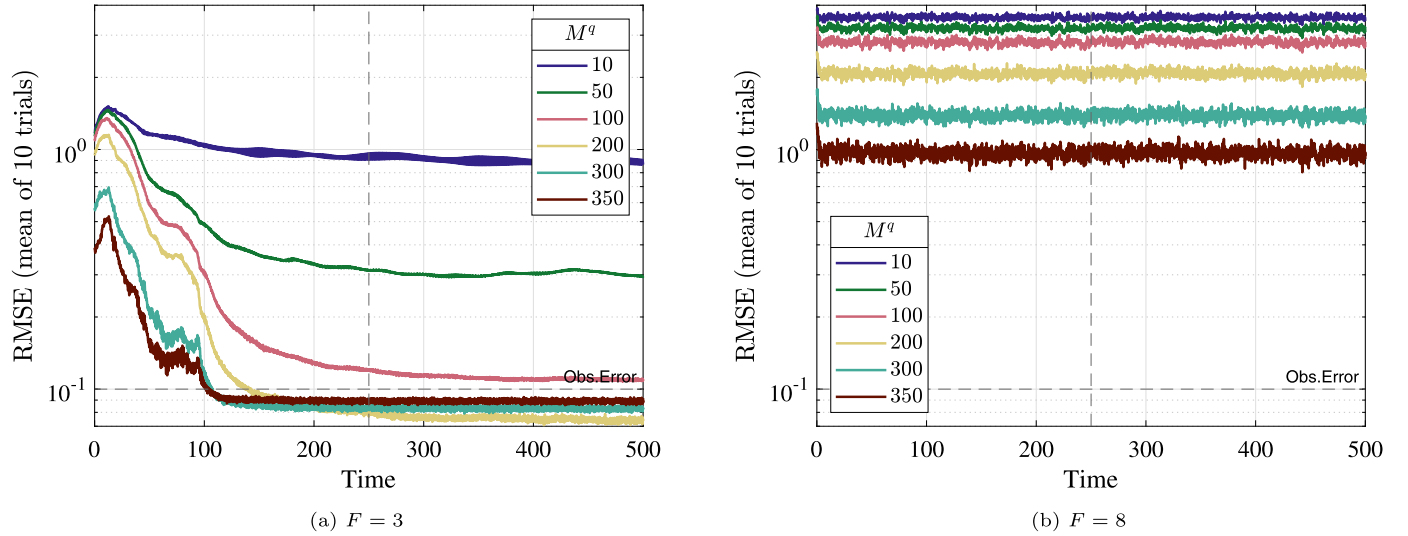
The proportion of time steps in which particle resampling was performed, RES%, increases steadily with the dimension of the projection of the data space. This indicates that the primary effect of  $D^q$  is to mitigate the weight collapse of the particle filter, without significantly affecting the accuracy of the assimilation. This is a confirmation of the effectiveness of using data-driven order reduction techniques, POD and DMD, instead of the model-based AUS, as detailed in [12].



**Fig. 3.** Experiment 1 for the L96 model with  $M = 40$ , Proj-OP-PF using DMD, POD, and AUS for physical and data models with fixed data projection dimension  $D^q = 5$  and varying the rank of the physical model projection dimension ( $M^q = 5, 10, \dots, M = 40$ ). Standard deviation of the observation noise ( $\mathbf{R} = 0.01 \mathbf{I}$ ) is given by the dash-dot line in RMSE panels. NON corresponds to Proj-OP-PF with the identity projection ( $M^q = M$  and  $D^q = D$ ) or equivalently no reduction of both the model and the data. See section 5.2.2 for further details.



**Fig. 4.** Influence of the reduction of order of the observation space for assimilation of L96 model with  $F = 3$  and  $F = 8$  for the experiment 2 (see Table 2). Assimilation was performed using Proj-OP-PF with POD projection of physical and data models  $M = D = 400$  with fixed physical projection dimension  $M^q = 100$  and varying the data model projection with rank  $D^q = 1, 2, \dots, M^q = 100$ . See section 5.2.3 for further details.



**Fig. 5.** Time evolution of RMSE for Proj-OP-PF using POD for the L96 model in regular  $F = 3$  and chaotic  $F = 8$  regime. Covariances  $\mathbf{Q} = 0.1\mathbf{I}$ ,  $\mathbf{R} = 0.01\mathbf{I}$ , and the remaining parameters as in Experiment 3 (see Table 2). Horizontal line at  $0.1 = \sqrt{0.01}$  denotes the standard deviation of the observation noise, while vertical line at  $t = 250$  indicates the start of the period used to compute mean RMSE in Fig. 4 and Fig. 6.

#### 5.2.4. Experiment 3 ( $F = 3, 4, 6, 8$ , $M = 400$ )

Next we consider L96 system with  $M = 400$  and varying values of the forcing term  $F$  and focus on time-independent POD and DMD based projections. We revert to  $D^q = 5$ . As before, we average the results over 10 trials with randomized initial conditions and noise realization.

Time evolution of RMSE shown in Fig. 5 for  $D^q = 5$  shows two representative cases of reduced-order assimilation in regular  $F = 3$  and chaotic  $F = 8$  regimes. For  $F = 3$ , the dynamics are regular, and evolve in lower-dimensional subspace of the state space, as indicated by the calculation of singular values and Lyapunov dimensions mentioned in Section 5.2.1. Consequently, fewer models, likely between  $M^q = 50$  and  $M^q = 100$  are needed for accurate data assimilation, as indicated by the RMSE converging to, or below, standard deviation of observation error once sufficiently many modes are selected. Increasing the reduced model dimension  $M^q > 100$  gives only a slight improvement in the asymptotic RMSE and in reducing the time to reach the asymptotic value.

In contrast,  $F = 8$  evolution has no similar low-dimensional structure, in addition to being chaotic, in which case only small improvements are seen in RMSE as more modes are added, but even for fairly large  $M^q = 350$ , the RMSE remains an order of magnitude larger than the observation error. Fig. 6 shows trends in dependence of RMSE and RES% across more values of forcing  $F$ , and additionally compares POD and DMD model reduction. In all cases, RMSE shown is a time-average of values in the second half of the assimilation period, after the transient (to the right of the vertical dashed line in Fig. 6).

We see that for larger values of  $F$  the results are of the order of the observation error only for projection ranks near the underlying model dimension  $M = 400$ . However, for  $F = 3$  and  $F = 4$  we obtain RMSE less than one for relatively small physical model projection ranks and RMSE of less than 0.25 for POD for physical model projections with ranks of approximately 50 and higher. Overall, as with  $M = 40$  and  $F = 8$ , we observe little gain in reduction of the physical model for  $M = 400$  and  $F = 6, 8$ . On the other hand, for  $F = 3, 4$  we observe plateauing ( $F = 4$ ) and a minimum ( $F = 3$ ) as a function of the reduced model dimension  $M^q$ .

### 5.3. Shallow Water Equations (SWE)

#### 5.3.1. Model, simulation, and parameters

The SWE are frequently used in science and engineering applications to model free-surface flows where the depth is small compared

to the horizontal scale(s) of the domain. Successful applications of the SWEs include modeling dam breaks, hurricane storm surges, tsunamis and atmospheric flows [75]. Motivated by this wide utility, our second example features SWE on a rectangular domain, configured to approximate a barotropic instability. Detailed derivation of SWE model can be found in standard textbooks on geophysical fluid dynamics, e.g. [76, §3].

The governing equations for this system are

$$\begin{aligned}\frac{\partial u}{\partial t} &= \left(-\frac{\partial u}{\partial y} + f\right)v - \frac{\partial}{\partial x}\left(\frac{1}{2}u^2 + gh\right) + \nu\nabla^2 u - c_b u, \\ \frac{\partial v}{\partial t} &= -\left(\frac{\partial v}{\partial x} + f\right)u - \frac{\partial}{\partial y}\left(\frac{1}{2}v^2 + gh\right) + \nu\nabla^2 v - c_b v, \\ \frac{\partial h}{\partial t} &= -\frac{\partial}{\partial x}((h + \underline{h})u) - \frac{\partial}{\partial y}((h + \underline{h})v).\end{aligned}\quad (53)$$

Here  $u$  and  $v$  are the  $x$ - and  $y$ -components of the velocity field. The total height of the water column is  $h + \underline{h}$ , where  $h$  is the height of the wave, and  $\underline{h}$  is the depth of the ocean, although we employ the flat orography  $\underline{h} \equiv 0$ . The parameter  $g$  is the gravitational constant,  $f$  is the Coriolis parameter,  $c_b$  the bottom friction coefficient, and  $\nu$  is the viscosity coefficient.

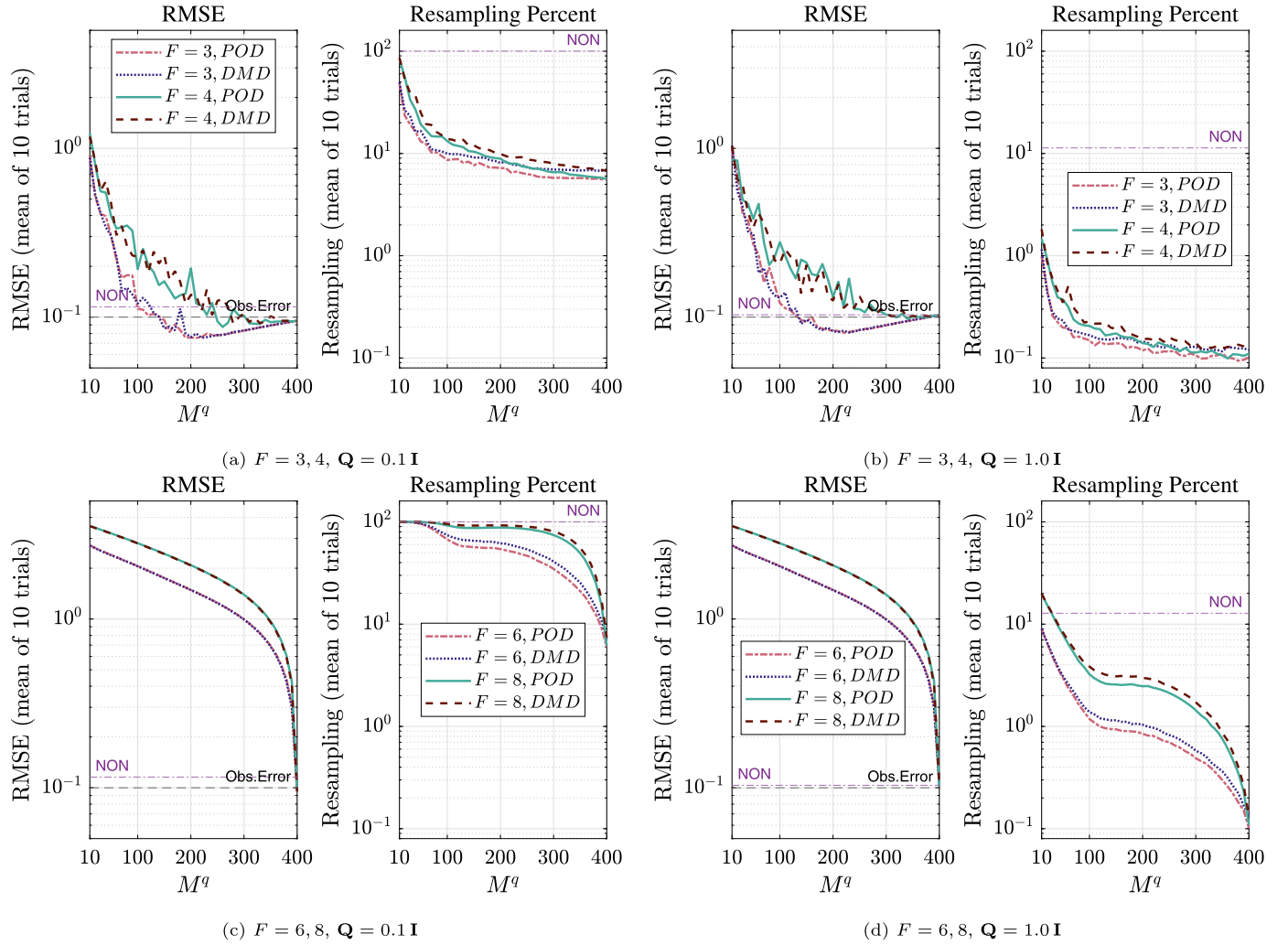
The initial value problem for (53) is solved using R. Hogan's finite difference code [77]. The three fields  $u, v, h$  are evaluated at a grid of  $254 \times 50$  points in the  $(x, y)$  plane, resulting in  $M = 38100$  state variables. Solutions are approximated using a standard finite difference scheme in space and a Lax–Wendroff finite-difference scheme in time with a fixed time step of  $\tau_M = 1 \text{ min} = 1/60 \text{ h}$  so that the Courant–Friedrichs–Lewy condition is satisfied. We let the system evolve in time over a total of 96 h. Fig. 8a shows an example of the non-assimilated simulation output showing the barotropic instability with the flat orography.

For SWE we consider both complete observations with all model variables observed ( $p = 100\%$ ) or sparse observations with every 100th variable observed ( $p = 1\%$ ). We additionally consider three basic observation scenarios for experiments (see [78]):

- (i) only  $u$  and  $v$  variables are observed, so  $D = \frac{2}{3}pM$ ,
- (ii) all variables are observed, so  $D = pM$ , and
- (iii) only  $h$  variable is observed, so  $D = \frac{1}{3}pM$ .

Since in all cases the observed variables are not transformed in any other way, the  $\mathbf{H}$  amounts to an identity matrix with a portion of the rows removed.





**Fig. 6.** Experiment 3 using L96 model with  $F = 3, 4, 6, 8$  and  $M = 400$ , Proj-OP-PF using DMD, and POD for the physical model and the data model with fixed data projection dimension  $D^q = 5$  and varying the rank of the physical model projection  $M^q = 10, 20, \dots, 400$ . Columns correspond to two different model noise correlations  $\mathbf{Q}$ . NON corresponds to the optimal proposal particle filter with no reduction of the model and the data. In the case of no reduction (NON), there is little difference between  $F = 3$  and  $F = 4$  ((a) and (b)) and similarly between  $F = 6$  and  $F = 8$  ((c) and (d)). The displayed RMSE is a time average after the initial transient (second half of the assimilation period shown in Fig. 5). Standard deviation of the observation noise is fixed and given in the dash-dot line in RMSE panels. See section 5.2.4 for further details.

As with L96, model and observation error noise are uncorrelated and affect all variables equally, so correlation matrices  $\mathbf{Q}, \mathbf{R}$  are chosen to be scaled identity matrices. Specifically, we employ error covariance matrices  $\mathbf{Q} = 1.0 \mathbf{I}, 0.1 \mathbf{I}$  and  $\mathbf{R} = 0.01 \mathbf{I}$  and subsequently (see 5.3.5) observation error covariance matrices  $\mathbf{R} = 1.0 \mathbf{I}, 0.1 \mathbf{I}$ .

Although we will not employ AUS projections with SWE, we calculated the approximate Lyapunov spectrum for the version of SWE we are employing. In particular, the largest computed approximate Lyapunov exponents over the 4 day time interval are relatively smaller than those for L96 and of the order  $10^{-3} \text{ min}^{-1}$ . We found  $\approx 30$  positive exponents and a computed Kaplan–Yorke dimension of  $\approx 110$ . For POD and DMD type projections, Fig. 7 shows the spectrum of singular values and DMD frequencies. Changes in the slope of the graph of singular values are sometimes used as an indicator of an inherent dimensionality of the problem, with the assumption that different component phenomena, such as multiscale oscillations or noise sources, may have different slopes of variance associated with them. For SWE the POD and DMD projections are obtained using a fixed reference solution (the truth) for each of the 10 trials.

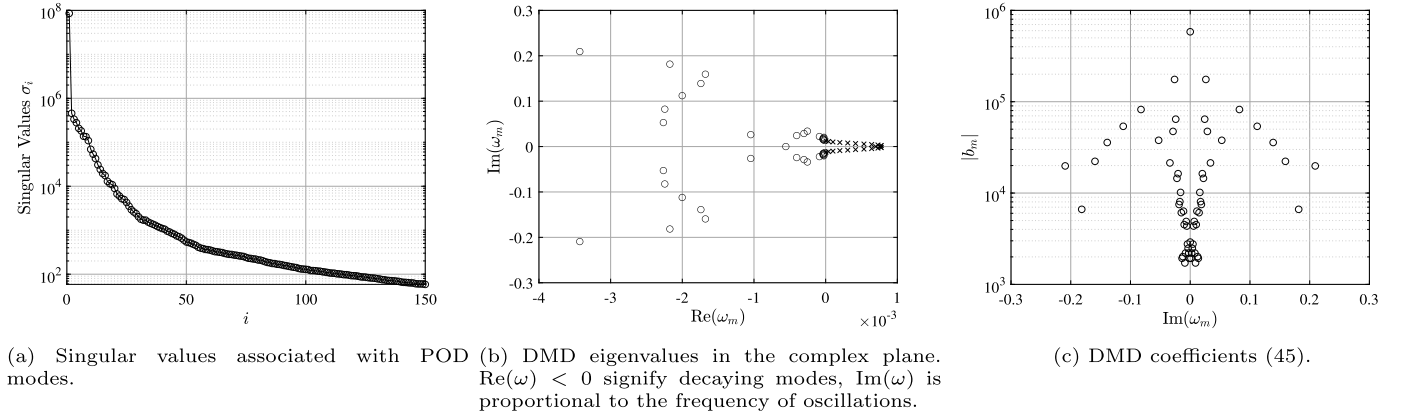
The DMD coefficients  $b_i$  do not show many isolated peaks, suggesting that the dynamics is not prominently low-dimensional. The same

conclusion can be drawn from a lack of jumps or gaps in the spectrum of singular values. Nevertheless, we can evaluate how well the reduced-order assimilation performs for various choices of  $M^q$ .

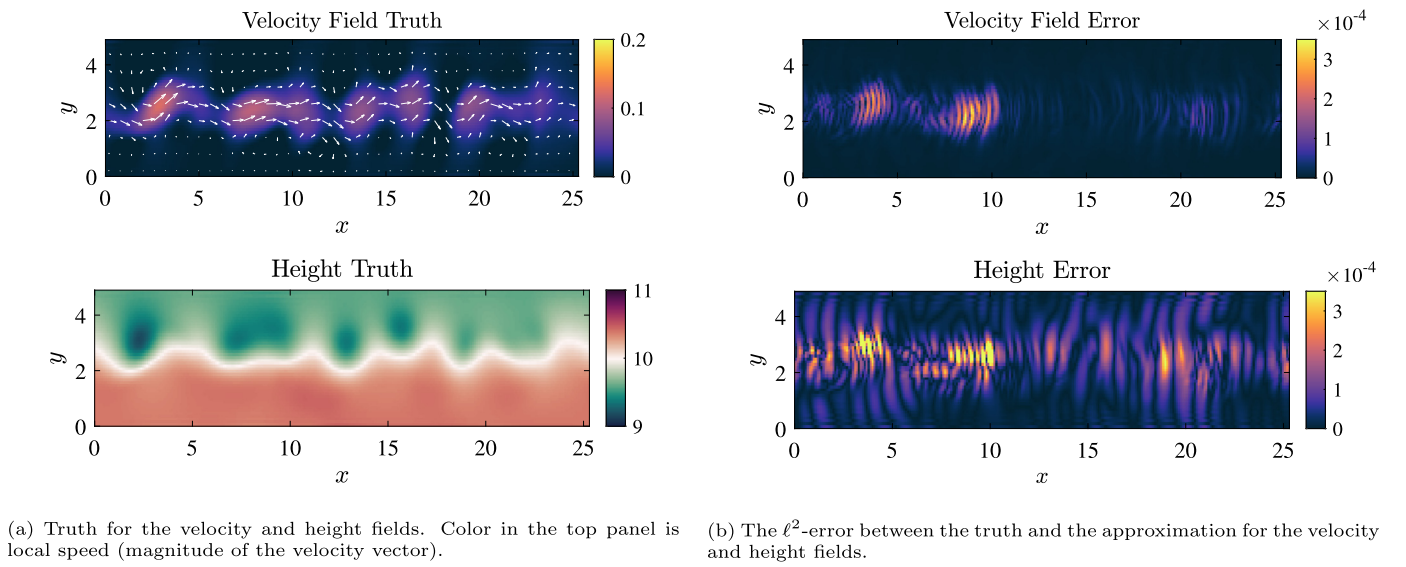
With the projection matrices computed, we assimilate using Proj-OP-PF starting at  $t = 48 \text{ h}$  and continuing for the next 24 h, with observations performed and assimilated with the timestep of  $\tau_D = 60 \text{ min}$  (as compared to a one-minute observational time scale in [78]). This effectively discards the first day as transient between the initial condition and the development of coherent structures.

The time scales, from the length of the simulation down to the observation timescale, were chosen to allow for an efficient proof-of-principle demonstration of the use of data-driven model reduction with data assimilation. We make no claim here that the same choices should be made in general when the SWE is used as the model. In general, the parametrization of the model, the broader context in which DA is used, and the variation of the computed projection matrices with respect to the duration and start of the window would all be driving the choice of the chosen timescales. We expect to further explore these issues in future work.

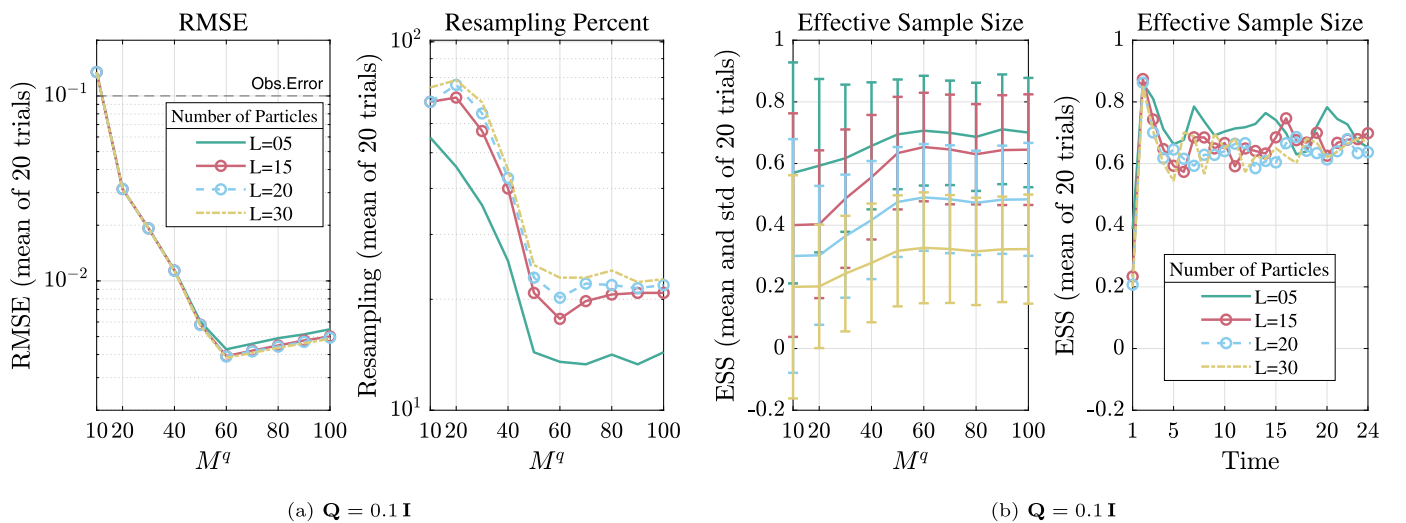
Fig. 8 shows the true state of the model at the end of the assimilation period ( $t = 72$ ) and the assimilation error at the same time. Here



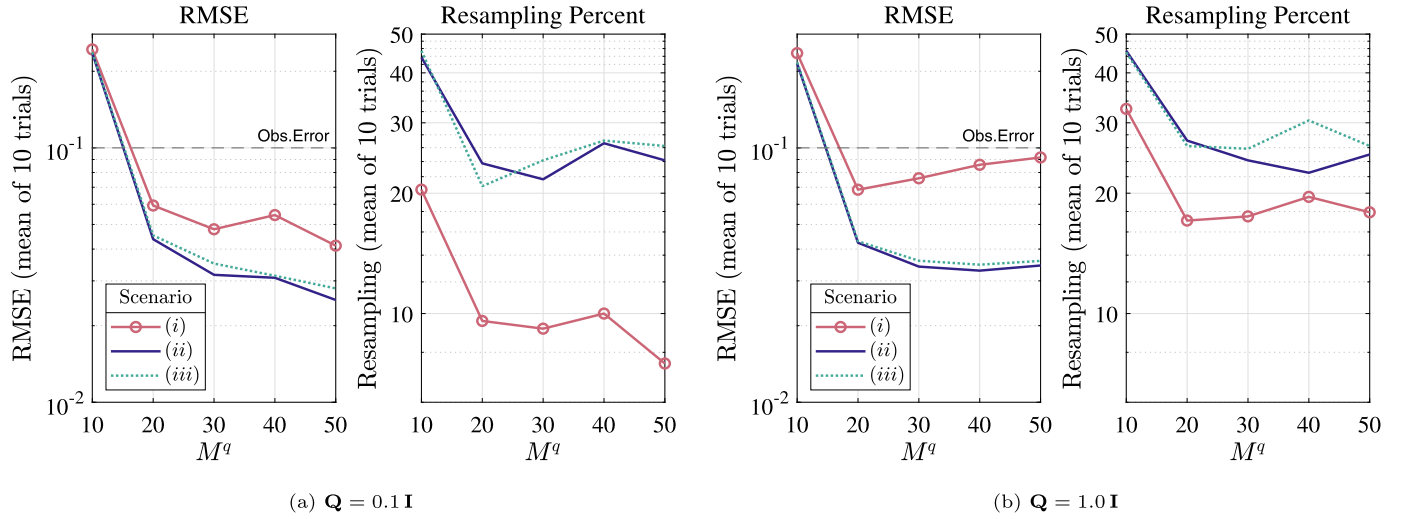
**Fig. 7.** Singular values and DMD eigenvalues spectra for SWE model between  $t = 24$  h and 48 h.



**Fig. 8.** “Truth” for the state variables  $h, u, v$  of the SWE models, and the spatial structure of the assimilation error at the end of assimilation period ( $t = 72$ ). Observation operator included all state variables at 1% of spatial nodes. Particle filter used 5 particles with error covariances  $\mathbf{Q} = 0.1 \mathbf{I}$ , and  $\mathbf{R} = 0.01 \mathbf{I}$ . Reduction performed using POD with  $M^q = 20$ , and  $D^q = 10$ .



**Fig. 9.** SWE model, Proj-OP-PF using DMD for the physical and data models with fixed data projection dimension  $D^q = 10$  and varying the physical model projection from  $M^q = 10$  to  $M^q = 100$ . Observation scenario (ii) with number of particles  $L = 5, 15, 20, 30$ . In (b) the mean scaled ESS ( $\text{ESS}/L$ ) is plotted against the reduced model dimension  $M^q$  (error bars correspond to a single standard deviation) and against time.



**Fig. 10.** SWE model, Proj-OP-PF using POD for both the physical and data models with fixed data projection rank  $D^q = 10$  and varying the rank of the physical model projection  $M^q$ . In all cases, assimilation uses  $L = 5$  particles.

**Table 3**

Parametrization of experiments used to test Proj-OP-PF for SWE.

Exp.	Obs. scenario	$p$	$\mathbf{Q}$	$\mathbf{R}$	Reduction	$M^q$	$D^q$	$L$
1	(ii)	100%	$0.1 \mathbf{I}$	$0.01 \mathbf{I}$	DMD	10 – 100	10	5,15,20,30
2	(i), (ii), (iii)	1%	$0.1 \mathbf{I}, 1.0 \mathbf{I}$	$0.01 \mathbf{I}$	POD	10 – 50	10	5
3	(ii)	1%, 100%	$0.1 \mathbf{I}, 1.0 \mathbf{I}$	$0.01 \mathbf{I}$	POD, DMD	10 – 100	10	5
4	(ii)	1%	$0.1 \mathbf{I}, 1.0 \mathbf{I}$	$0.1 \mathbf{I}, 1.0 \mathbf{I}$	POD, DMD	10 – 100	10	5

we used POD to reduce the dimension of the model space to  $M^q = 20$  and the data space to  $D^q = 10$ . We observed  $p = 1\%$  of spatial nodes, used  $L = 5$  particles, and employed error covariances of  $\mathbf{Q} = 0.1 \mathbf{I}$  and  $\mathbf{R} = 0.01 \mathbf{I}$ . The magnitude of error in both components several orders of magnitude smaller than the value of the state, indicating that the assimilation was successful. The ideal error fields would show no structure, since the observation noise was homogeneous across all spatial nodes. However, the error field in Fig. 8b shows a striated pattern which is likely related to the structure of the first POD mode removed from the model. The peaks in the pattern are concentrated along the midline, where the state variables change rapidly, which is expected as sensitivity to the spatial location above/below the midline would likely result in different assimilation particles taking different values of state variables at those spatial nodes.

For a systematic evaluation of OP-PF across ranges of parameters, we performed experiments in which the quality of assimilated state was measured by RMSE and RES%, as explained in section 5.1. Details of the setup of each experiment are given in Table 3. The RMSE and RES% are calculated based on hourly observations in the 24 h observation window. We also tested the dependence on the projected data dimension ( $D^q$ ) similar to what we illustrated in Fig. 4 for L96. For example, for SWE with  $M^q = 40$  with  $p = 1\%$  spatial nodes observed in scenario (ii) with  $\mathbf{Q} = 0.1 \mathbf{I}$  and  $\mathbf{R} = 0.01 \mathbf{I}$  we found little variation in the RMSE and RES% as the projected observation dimension  $D^q$  varied from 1 to 40. In particular, with  $L = 5$  particles we found a mean Effective Sample Size (ESS) of nearly 4, no resampling, and mean RMSE of 0.0465 in model space and mean RMSE of 0.0250 in the projected model space.

### 5.3.2. Experiment 1 (number of particles)

In Fig. 9, we vary the number of particles  $L$  and find that we obtain similar results for  $L = 5, 15, 20, 30$  particles. Fig. (9a) represents the RMSE and RES% where we obtain minimum RMSE for reduced model dimension of  $M^q = 60$ . Fig. (9b) shows the scaled ESS (ESS/ $L$ ) where the left graph is showing the mean and standard deviation of 20 trials for  $M^q = 10, 20, \dots, M$ . Note that although the scaled ESS is in general

larger for  $L = 5$  particles, in an absolute sense for  $L = 15, 20, 30$  the ESS is much larger than with  $L = 5$ . The right graph is showing the mean of ESS of 20 trials over time for  $L = 5, 15, 20, 30$ . For the remaining experiments with SWE we employ  $L = 5$  particles since it is computationally more efficient.

### 5.3.3. Experiment 2 (observation scenarios)

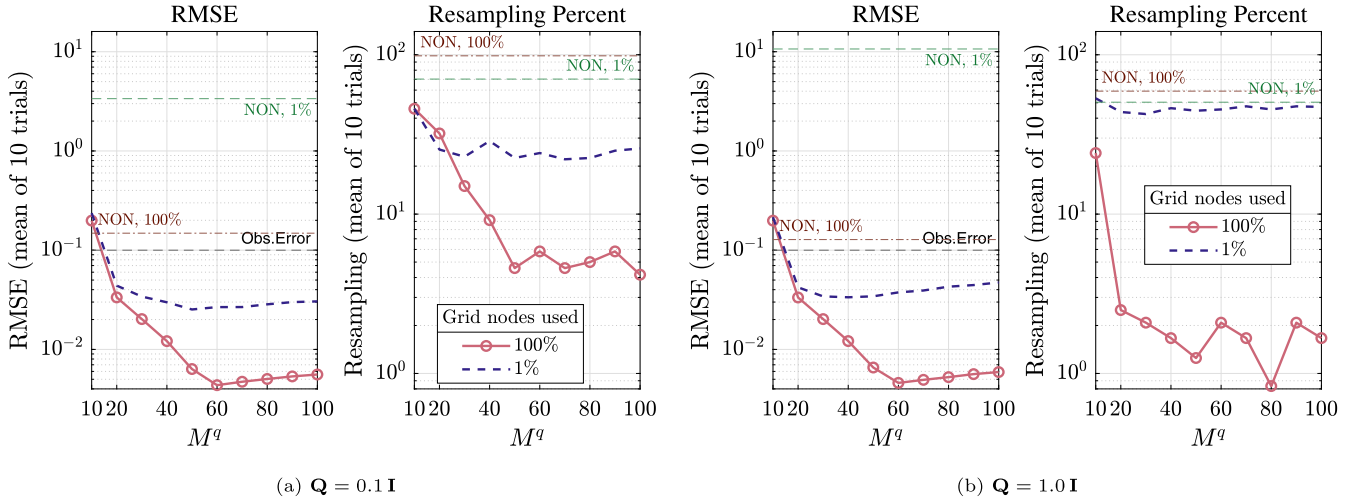
Fig. 10, illustrates the performance of OP-PF on SWE where POD is used to derive the reduced order physical and data models. It shows the mean RMSE the resampling parentage for each of the three scenarios considered here. Scenarios (ii) and (iii), in which variables  $u, v$  and/or  $h$  are observed, can be seen to have a lower RMSE than the case of scenario (i), in which just  $u$  and  $v$  are observed.

### 5.3.4. Experiment 3 (sparse and complete observations)

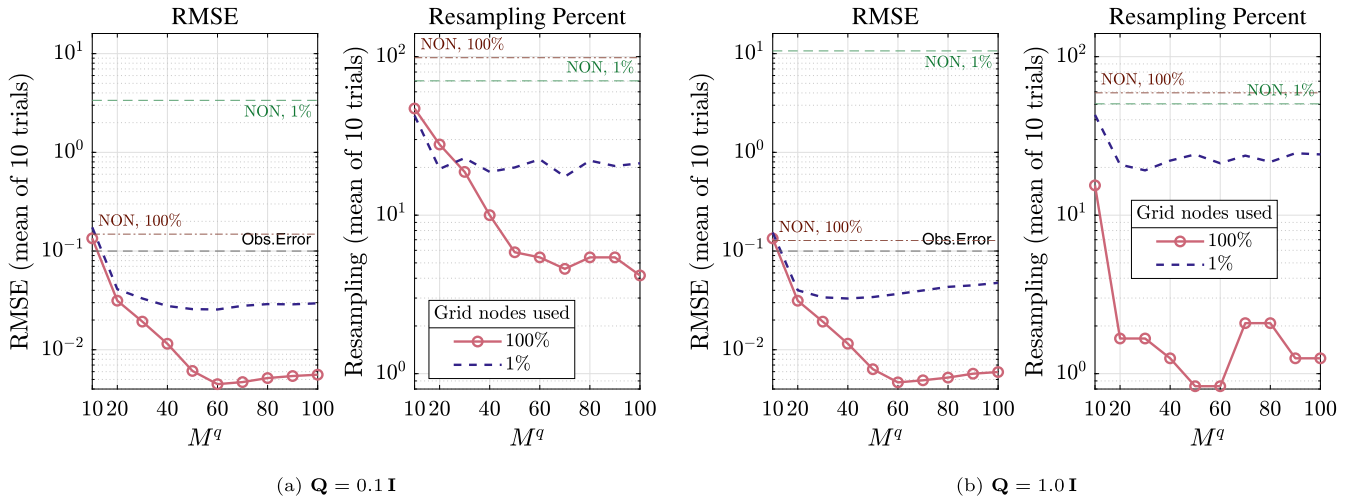
Figs. 11 and 12 compare the RMSE and RES% for the case when all variables ( $u, v, h$ ) are observed at only a fraction of spatial nodes,  $p = 1\%$  and  $p = 100\%$  of nodes observed, and implying that the dimension of the observation space scales as  $D = pM$ . In all cases the minimum RMSE obtained over 10 trials occurs for reduced model dimension  $M^q = 60$ . Both RMSE and RES% are much smaller compared to the optimal proposal particle filter with no model and data reduction (NON) in all cases. The resampling percentage RES% is lower for  $p = 100\%$  as are the calculated RMSE.

### 5.3.5. Experiment 4 (larger observation error covariances $\mathbf{R}$ )

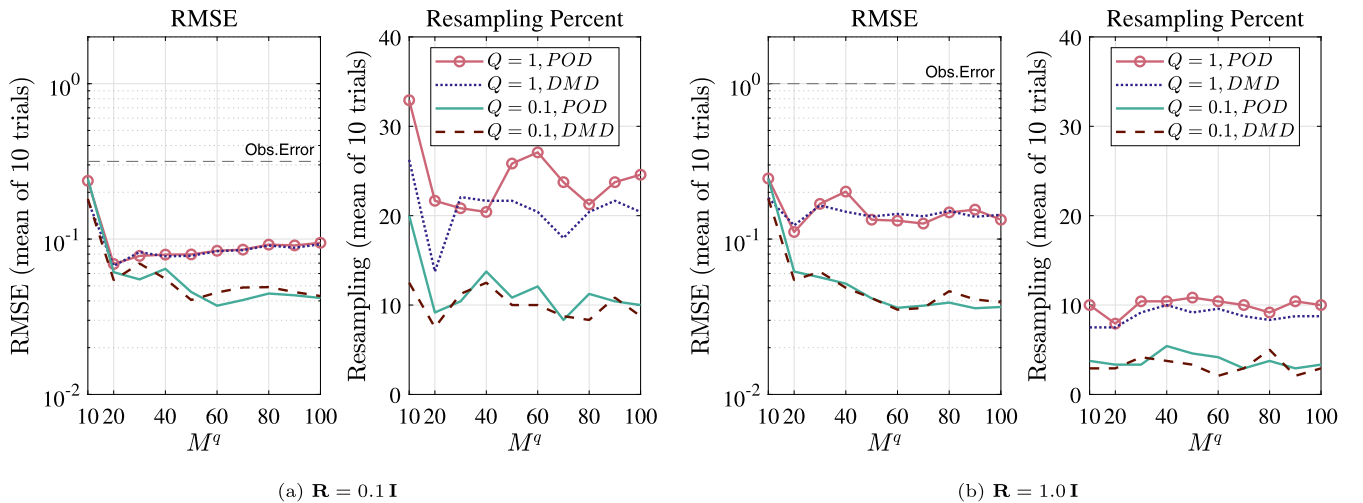
In Fig. 13, POD and DMD are employed to reduce the dimension of the model space to  $M^q = 10, 20, \dots, 100$  and the data space to  $D^q = 10$ . We observe  $p = 1\%$  of spatial nodes and use  $L = 5$  particles. We compare POD and DMD with error covariance matrices  $\mathbf{Q} = 1.0 \mathbf{I}$  and  $0.1 \mathbf{I}$  and  $\mathbf{R} = 0.1 \mathbf{I}$  and  $1.0 \mathbf{I}$ . We observe a plateauing of the RMSE starting from approximately  $M^q = 60$  with model error covariance  $\mathbf{Q} = 0.1 \mathbf{I}$  and a minimum at  $M^q = 20$  for  $\mathbf{Q} = 1.0 \mathbf{I}$ . The resampling percentages are relatively constant as a function of  $M^q$  with larger values for  $\mathbf{Q} = 1.0 \mathbf{I}$  and  $\mathbf{R} = 0.1 \mathbf{I}$ .



**Fig. 11.** SWE model, Proj-OP-PF using POD for the physical and data models with fixed Data projection  $D^q = 10$  and varying the physical model projection from  $M^q = 10$  to  $M^q = 100$ . NON corresponds to the optimal proposal particle filter with no reduction of the model and the data. All variables ( $u, v, h$ ) are observed (observation scenario (ii)) at each observation time but at varying percentages  $p = 1\%$  and  $p = 100\%$  of spatial nodes.



**Fig. 12.** SWE model, Proj-OP-PF using DMD for the physical and data models with fixed Data projection  $D^q = 10$  and varying the physical model projection from dimension  $M^q = 10$  to  $M^q = 100$ . NON corresponds to optimal proposal particle filter with no model and no data reduction. All variables ( $u, v, h$ ) are observed (observation scenario (ii)) at each observation time but at varying percentages  $p = 1\%$  and  $p = 100\%$  of spatial nodes.



**Fig. 13.** Influence of larger observation error covariances on assimilation for the SWE model.



## 6. Discussion and conclusions

In this paper, we have derived a projected bootstrap and optimal proposal Particle Filters for physical and data models used in a standard data assimilation framework. Our focus is on state space based projections formed using Assimilation in the Unstable Subspace (AUS), Proper Orthogonal Decomposition (POD), or Dynamic Mode Decomposition (DMD). This framework provides a basis for employing Particle Filters for high dimensional nonlinear problems, and extensively tests a projected optimal proposal particle filter algorithm, Projected Optimal Proposal Particle Filter (Proj-OP-PF), that combines projected and unprojected models. It is shown that stable assimilation results are obtained for the Lorenz'96 model (L96) model and Shallow Water Equations (SWE) in terms of Root Mean Squared Error (RMSE) and resampling percentage. The results are particularly promising for the SWE, where Proj-OP-PF with minimal tuning provides good results for severely truncated physical models and low dimensional observation operators from full to sparse observations. That is, we have successfully applied a Particle Filter to a 38, 100-dimensional multi-component physical system. Essentially, the methods developed here perform effectively when either the physical model or the observational data have a lower effective dimension. If these effective dimensions are sufficiently small, then the resampling percentage is low due to working with lower dimensional projected models. When there is sufficient resolution in the reduced dimensional physical model solutions and in the reduced dimensional data, then RMSEs are obtained on the order of the observation error.

There are several interesting avenues for further exploration. These include the application of more sophisticated ocean, atmosphere, and coupled models. Extension to nonlinear observation operators would require a suitable linearization of the nonlinear observation operator be obtained, to employ in the projected data models. Incorporating localization techniques for particle filters is theoretically trivial (although we have not yet tried it), as the projected approach is phrased as a Bayesian problem suitable for any data assimilation algorithm: a projected, localized Particle Filter would illustrate the value of these techniques in a more realistic context. Another avenue we are planning to explore is to develop time dependent POD and DMD modes using appropriately sized moving windows of snapshots and an updating/downdating procedure. We also plan to explore observation space projections using windowed snapshots of the observations.

## References

- [1] A. Chorin, M. Morzfeld, X. Tu, Implicit particle filters for data assimilation, *Comput. Methods Appl. Mech. Eng.* 318 (2) (2010) 221–240.
- [2] P.J. van Leeuwen, Nonlinear data assimilation in geosciences: an extremely efficient particle filter, *Q. J. R. Meteorol. Soc.* 136 (653) (2010) 1991–1999.
- [3] C. Snyder, Particle filters, the “optimal” proposal and high-dimensional systems, in: *Proceedings of the ECMWF Seminar on Data Assimilation for Atmosphere and Ocean*, 2011, pp. 1–10.
- [4] M. Morzfeld, X. Tu, E. Atkins, A.J. Chorin, A random map implementation of implicit filters, *J. Comput. Phys.* 231 (4) (2012) 2049–2066.
- [5] C. Snyder, T. Bengtsson, P. Bickel, J. Anderson, Obstacles to high-dimensional particle filtering, *Mon. Weather Rev.* 136 (12) (2008) 4629–4640.
- [6] P.J. van Leeuwen, Aspects of particle filtering in high-dimensional spaces, in: S. Ravela, A. Sandu (Eds.), *Dynamic Data-Driven Environmental Systems Science*, in: *Lecture Notes in Computer Science*, Springer International Publishing, Cham, 2015, pp. 251–262.
- [7] A. Trevisan, F. Uboldi, Assimilation of standard and targeted observations within the unstable subspace of the observation–analysis–forecast cycle system, *J. Atmos. Sci.* 61 (1) (2004) 103–113.
- [8] F. Uboldi, A. Trevisan, A. Carrassi, Developing a dynamically based assimilation method for targeted and standard observations, *Nonlinear Process. Geophys.* 12 (1) (2005) 149–156.
- [9] A. Trevisan, L. Palatella, On the Kalman filter error covariance collapse into the unstable subspace, *Nonlinear Process. Geophys.* 18 (2) (2011) 243–250.
- [10] L. Palatella, A. Carrassi, A. Trevisan, Lyapunov vectors and assimilation in the unstable subspace: theory and applications, *J. Phys. A, Math. Theor.* 46 (25) (2013).
- [11] L. Palatella, A. Trevisan, Interaction of Lyapunov vectors in the formulation of the nonlinear extension of the Kalman filter, *Phys. Rev. E* (3) 91 (4) (2015) 042905, 10.
- [12] J. Maclean, E. Van Vleck, Particle filters for data assimilation based on reduced order data models, *Q. J. R. Meteorol. Soc.* 147 (2021) 1892–1907.
- [13] A. Carrassi, A. Trevisan, L. Descamps, O. Talagrand, F. Uboldi, Controlling instabilities along a 3dvar analysis cycle by assimilating in the unstable subspace: a comparison with the enf, *Nonlinear Process. Geophys.* 15 (4) (2008) 503–521.
- [14] M. Bocquet, A. Carrassi, Four-dimensional ensemble variational data assimilation and the unstable subspace, *Tellus, Ser. A Dyn. Meteorol. Oceanogr.* 69 (1) (2017) 1304504, publisher: Taylor & Francis.
- [15] B. de Leeuw, S. Dubinkina, J. Frank, A. Steyer, X. Tu, E. Van Vleck, Projected shadowing-based data assimilation, *SIAM J. Appl. Dyn. Syst.* 17 (2018) 2446–2477.
- [16] A. Farchi, M. Bocquet, Comparison of local particle filters and new implementations, *Nonlinear Process. Geophys.* 25 (4) (2018) 765–807.
- [17] J. Poterjoy, J.L. Anderson, Efficient assimilation of simulated observations in a high-dimensional geophysical system using a localized particle filter, *Mon. Weather Rev.* 144 (5) (2016) 2007–2020.
- [18] R. Potthast, A. Walter, A. Rhodin, A localized adaptive particle filter within an operational nwp framework, *Mon. Weather Rev.* 147 (1) (2019) 345–362.
- [19] J. Poterjoy, A localized particle filter for high-dimensional nonlinear systems, *Mon. Weather Rev.* 144 (1) (2016) 59–76.
- [20] T.P. Sapsis, P.F.J. Lermusiaux, Dynamically orthogonal field equations for continuous stochastic dynamical systems, *Physica D* 238 (23–24) (2009) 2347–2360.
- [21] T. Sapsis, Dynamically orthogonal field equations, Ph.D. thesis, Massachusetts Institute of Technology, Department of Mechanical Engineering, 2010.
- [22] T. Sondergaard, P.F. Lermusiaux, Data assimilation with Gaussian mixture models using the dynamically orthogonal field equations. Part I: Theory and scheme, *Mon. Weather Rev.* 141 (6) (2013) 1737–1760.
- [23] A.J. Majda, D. Qi, T.P. Sapsis, Blended particle filters for large-dimensional chaotic dynamical systems, *Proc. Natl. Acad. Sci. USA* 111 (21) (2014) 7511–7516.
- [24] D. Qi, A.J. Majda, Blended particle methods with adaptive subspaces for filtering turbulent dynamical systems, *Physica D* 298/299 (2015) 21–41.
- [25] G.V. Iungo, C. Santoni-Ortiz, M. Abkar, F. Porté-Agel, M.A. Rotea, S. Leonardi, Data-driven Reduced Order Model for prediction of wind turbine wakes, *J. Phys. Conf. Ser.* 625 (2015) 012009, <https://doi.org/10.1088/1742-6596/625/1/012009>.
- [26] P.M. Mehta, R. Linares, A new transformative framework for data assimilation and calibration of physical ionosphere-thermosphere models, *Space Weather* 16 (8) (2018) 1086–1100, <https://doi.org/10.1029/2018SW001875>.
- [27] Z. Wang, X. Luo, S.S.-T. Yau, Z. Zhang, Proper orthogonal decomposition method to nonlinear filtering problems in medium-high dimension, *IEEE Trans. Autom. Control* 65 (4) (2020) 1613–1624.
- [28] A.A. Popov, C. Mou, A. Sandu, T. Iliescu, A multifidelity ensemble Kalman filter with reduced order control variates, *SIAM J. Sci. Comput.* 43 (2) (2021) A1134–A1162, <https://doi.org/10.1137/20M1349965>.
- [29] Y. Cao, J. Zhu, I.M. Navon, Z. Luo, A reduced-order approach to four-dimensional variational data assimilation using proper orthogonal decomposition, *Int. J. Numer. Methods Fluids* 53 (10) (2007) 1571–1583.
- [30] G. Dimitriu, N. Apreutesei, Comparative study with data assimilation experiments using proper orthogonal decomposition method, in: *Lecture Notes in Computer Science*, 2008, pp. 393–400.
- [31] J. Du, I.M. Navon, J. Zhu, F. Fang, A.K. Alekseev, Reduced order modeling based on POD of a parabolized Navier–Stokes equations model II: trust region POD 4D VAR data assimilation, *Comput. Math. Appl.* 65 (3) (2013) 380–394, <https://doi.org/10.1016/j.camwa.2012.06.001>.
- [32] F. Fang, C.C. Pain, I.M. Navon, M.D. Piggott, G.J. Gorman, P.E. Farrell, P.A. Allison, A.J.H. Goddard, A POD reduced-order 4D-Var adaptive mesh ocean modelling approach, *Int. J. Numer. Methods Fluids* 60 (7) (2009) 709–732, <https://doi.org/10.1002/fld.1911>.
- [33] R. Ștefănescu, A. Sandu, I.M. Navon, POD/DEIM reduced-order strategies for efficient four dimensional variational data assimilation, *J. Comput. Phys.* 295 (2015) 569–595.
- [34] N. Panda, M.G. Fernández-Godino, H.C. Godínez, C. Dawson, A data-driven nonlinear assimilation framework with neural networks, *Comput. Geosci.* 25 (1) (2021) 233–242, <https://doi.org/10.1007/s10596-020-10001-6>.
- [35] T. Nonomura, H. Shibata, R. Takaki, Dynamic mode decomposition using a Kalman filter for parameter estimation, *AIP Adv.* 8 (10) (2018) 105106.
- [36] T. Nonomura, H. Shibata, R. Takaki, Extended-Kalman-filter-based dynamic mode decomposition for simultaneous system identification and denoising, *PLoS ONE* 14 (2) (2019) e0209836.
- [37] J. Maclean, N. Santitissadeekorn, C.K.R.T. Jones, A coherent structure approach for parameter estimation in Lagrangian data assimilation, *Physica D* 360 (2017) 36–45.
- [38] M. Morzfeld, J. Adams, S. Lunderman, R. Orozco, Feature-based data assimilation in geophysics, *Nonlinear Process. Geophys.* 25 (2) (2018) 355–374.
- [39] E. Kalnay, *Atmospheric Modeling, Data Assimilation and Predictability*, Cambridge University Press, 2003.
- [40] C.K. Wikle, L.M. Berliner, A bayesian tutorial for data assimilation, *Physica D* 230 (1) (2007) 1–16.
- [41] M. Bocquet, P. Savov, An iterative ensemble Kalman smoother, *Q. J. R. Meteorol. Soc.* 140 (682) (2014) 1521–1535.
- [42] K. Law, A. Stuart, K. Zygalakis, *Data Assimilation: Mathematical Introduction*, Springer-Verlag, 2015.

- [43] S. Vetra-Carvalho, J. van Leeuwen, N. Peter, B. Lars, A. Alexander, M. Umer, P. Brasseur, P. Kirchgeßner, J.-M. Becker, State-of-the-art stochastic data assimilation methods for high-dimensional non-gaussian problems, *Tellus, Ser. A Dyn. Meteorol. Oceanogr.* 70 (2018) 1–43.
- [44] P.-J. van Leeuwen, Nonlinear data assimilation for high-dimensional systems, in: *Nonlinear Data Assimilation*, in: *Frontiers in Applied Dynamical Systems: Reviews and Tutorials*, vol. 2, Springer-Verlag, Cham, 2015, pp. 1–73.
- [45] A. Budhiraja, E. Friedlander, C. Guider, C.K. Jones, J. Maclean, Data assimilation, inference for linking physical and probabilistic models for complex nonlinear dynamic systems, 1 edn., in: A.E. Gelfand, M. Fuentes, J.A. Hoeting, R.L. Smith (Eds.), *Handbook of Environmental and Ecological Statistics*, CRC Press, 2017, pp. 687–708.
- [46] S.C. Surace, A. Kutschireiter, J.-P. Pfister, How to avoid the curse of dimensionality: scalability of particle filters with and without importance weights, *SIAM Rev.* 61 (1) (2019) 79–91.
- [47] C. Snyder, T. Bengtsson, P. Bickel, J. Anderson, Obstacles to high-dimensional particle filtering, *Mon. Weather Rev.* 136 (12) (2008) 4629–4640.
- [48] C. Snyder, T. Bengtsson, M. Morzfeld, Performance bounds for particle filters using the optimal proposal, *Mon. Weather Rev.* 143 (11) (2015) 4750–4761.
- [49] P.-J. van Leeuwen, H.R. Künsch, L. Nerger, R. Potthast, S. Reich, Particle filters for high-dimensional geoscience applications: a review, *Q. J. R. Meteorol. Soc.* 145 (723) (2019) 2335–2365.
- [50] C. Tropea, A.L. Yarin, *Springer Handbook of Experimental Fluid Mechanics*, Springer Science & Business Media, 2007.
- [51] G. Berkooz, P. Holmes, J.L. Lumley, The proper orthogonal decomposition in the analysis of turbulent flows, *Annu. Rev. Fluid Mech.* 25 (1) (1993) 539–575.
- [52] C. Eckart, G. Young, The approximation of one matrix by another of lower rank, *Psychometrika* 1 (3) (1936) 211–218.
- [53] G.H. Golub, C.F. Van Loan, *Matrix Computations*, 4th edition, Johns Hopkins Studies in the Mathematical Sciences, Johns Hopkins University Press, Baltimore, MD, 2013.
- [54] M. Gavish, D.L. Donoho, The optimal hard threshold for singular values is  $4/\sqrt{3}$ , *IEEE Trans. Inf. Theory* 60 (8) (2014) 5040–5053.
- [55] J.N. Kutz, *Data-Driven Modeling & Scientific Computation: Methods for Complex Systems & Big Data*, OUP, Oxford, 2013.
- [56] P.J. Schmid, Dynamic mode decomposition of numerical and experimental data, *J. Fluid Mech.* 656 (2010) 5–28.
- [57] C. Rowley, I. Mezić, S. Bagheri, P. Schlatter, D. Henningson, Spectral analysis of nonlinear flows, *J. Fluid Mech.* 641 (2009) 115–127.
- [58] M. Budišić, R. Mohr, I. Mezić, Applied koopmanism, chaos, *Chaos, Interdiscip. J. Nonlinear Sci.* 22 (4) (2012) 047510, 33.
- [59] J.H. Tu, C.W. Rowley, D.M. Luchtenburg, S.L. Brunton, J.N. Kutz, On dynamic mode decomposition: theory and applications, *J. Comput. Dyn.* 1 (2) (2014) 391–421.
- [60] M. Korda, I. Mezić, On convergence of extended dynamic mode decomposition to the Koopman operator, *J. Nonlinear Sci.* 28 (2) (2018) 687–710, arXiv:1703.04680.
- [61] J. Kutz, S. Brunton, B. Brunton, J. Proctor, Dynamic mode decomposition, in: *Other Titles in Applied Mathematics*, Society for Industrial and Applied Mathematics, 2016.
- [62] I. Mezić, Analysis of fluid flows via spectral properties of the Koopman operator, *Annu. Rev. Fluid Mech.* 45 (1) (2013) 357–378.
- [63] L. Dieci, E. Van Vleck, Lyapunov and Sacker-Sell spectral intervals, *J. Dyn. Differ. Equ.* 19 (2007) 265–293.
- [64] L. Dieci, E. Van Vleck, Lyapunov exponents: computation, in: B. Engquist (Ed.), *Encyclopedia of Applied and Computational Mathematics*, Springer-Verlag, 2015, pp. 834–838.
- [65] E.N. Lorenz, Predictability - a problem partly solved, in: T. Palmer, R. Hagedorn (Eds.), *Proceedings of Seminar on Predictability*, vol. 1, UK ECMWF, Cambridge University Press, Reading, 1996, pp. 1–18.
- [66] R. Blender, J. Wouters, V. Lucarini, Avalanches, breathers, and flow reversal in a continuous Lorenz-96 model, *Phys. Rev. E* 88 (1) (2013) 013201.
- [67] D.L. van Kekem, A.E. Sterk, Travelling waves and their bifurcations in the Lorenz-96 model, *Physica D* 367 (2018) 38–60, <https://doi.org/10.1016/j.physd.2017.11.008>.
- [68] J.N. Kutz, X. Fu, S.L. Brunton, Multiresolution dynamic mode decomposition, *SIAM J. Appl. Dyn. Syst.* 15 (2) (2016) 713–735, <https://doi.org/10.1137/15M1023543>.
- [69] B.W. Brunton, L.A. Johnson, J.G. Ojemann, J.N. Kutz, Extracting spatial-temporal coherent patterns in large-scale neural recordings using dynamic mode decomposition, *J. Neurosci. Methods* 258 (2016) 1–15, <https://doi.org/10.1016/j.jneumeth.2015.10.010>.
- [70] J. Page, R.R. Kerswell, Koopman mode expansions between simple invariant solutions, *J. Fluid Mech.* 879 (2019) 1–27, <https://doi.org/10.1017/jfm.2019.686>.
- [71] S. Mačević, N. Črnjarić-Žic, I. Mezić, Koopman operator family spectrum for nonautonomous systems, *SIAM J. Appl. Dyn. Syst.* 17 (4) (2018) 2478–2515, <https://doi.org/10.1137/17M1133610>.
- [72] S. Mačević, N. Črnjarić-Žic, Koopman operator theory for nonautonomous and stochastic systems, in: A. Mauroy, I. Mezić, Y. Susuki (Eds.), *The Koopman Operator in Systems and Control: Concepts, Methodologies, and Applications*, in: *Lecture Notes in Control and Information Sciences*, Springer International Publishing, Cham, 2020, pp. 131–160.
- [73] N. Črnjarić-Žic, S. Mačević, I. Mezić, Koopman operator spectrum for random dynamical systems, *J. Nonlinear Sci.* 30 (5) (2020) 2007–2056, <https://doi.org/10.1007/s00332-019-09582-z>.
- [74] A. Carrassi, M. Bocquet, J. Demaeyer, C. Grudzien, P. Raanes, S. Vannitsem, Data assimilation for chaotic dynamics, arXiv preprint, arXiv:2010.07063, 2020.
- [75] A. Lozovskiy, M. Farthing, C. Kees, Evaluation of Galerkin and Petrov-Galerkin model reduction for finite element approximations of the shallow water equations, *Commun. Appl. Math. Comput. Sci.* 5 (2017) 537–571.
- [76] J. Pedlosky, *Geophysical Fluid Dynamics*, Springer Science & Business Media, 2013.
- [77] R. Hogan, Shallow water model in Matlab, [http://www.met.reading.ac.uk/~swrhgnrj/shallow\\_water\\_model/](http://www.met.reading.ac.uk/~swrhgnrj/shallow_water_model/), 2014.
- [78] D. Paulin, A. Jasra, A. Beskos, D. Crisan, A 4d-var method with flow-dependent background covariances for the shallow-water equations, arXiv:1710.11529, 2017.

## Acronyms

**4D-Var:** Four-dimensional variational data assimilation  
**AUS:** Assimilation in the Unstable Subspace  
**DA:** Data Assimilation  
**DMD:** Dynamic Mode Decomposition  
**EnKF:** Ensemble Kalman Filter  
**ESS:** Effective Sample Size  
**KF:** Kalman Filter  
**L96:** Lorenz'96 model  
**LV:** Lyapunov Vectors  
**ODE:** ordinary differential equation  
**OP-PF:** Optimal Proposal Particle Filter  
**PDE:** partial differential equation  
**PDF:** Probability Density Function  
**PF:** Particle Filter  
**POD:** Proper Orthogonal Decomposition  
**Proj-OP-PF:** Projected Optimal Proposal Particle Filter  
**Proj-PF:** Projected Particle Filter  
**RES%:** Resampling Percentage  
**RMSE:** Root Mean Squared Error  
**ROM:** Reduced Order Models  
**SVD:** Singular Value Decomposition  
**SWE:** Shallow Water Equations