

A MULTIFIDELITY ENSEMBLE KALMAN FILTER WITH REDUCED ORDER CONTROL VARIATES*

ANDREY A. POPOV[†], CHANGHONG MOU[‡], ADRIAN SANDU[†], AND TRAIAN ILIESCU[‡]

Abstract. This work develops a new multifidelity ensemble Kalman filter (MFEnKF) algorithm based on a linear control variate framework. The approach allows for rigorous multifidelity extensions of the EnKF, where the uncertainty in coarser fidelities in the hierarchy of models represents control variates for the uncertainty in finer fidelities. Small ensembles of high-fidelity model runs are complemented by larger ensembles of cheaper, lower-fidelity runs to obtain much improved analyses at only small additional computational costs. We investigate the use of reduced order models as coarse fidelity control variates in the MFEnKF and provide analyses to quantify the improvements over the traditional ensemble Kalman filters. We apply these ideas to perform data assimilation with a quasi-geostrophic test problem, using direct numerical simulation and a corresponding proper orthogonal decomposition–Galerkin reduced order model. Numerical results show that the two-fidelity MFEnKF provides better analyses than existing EnKF algorithms at comparable or reduced computational costs.

Key words. Bayesian inference, control variates, data assimilation, multifidelity ensemble Kalman filter, reduced order modeling

AMS subject classifications. 62F15, 62M20, 65C05, 65M60, 76F70, 86A22, 93E11

DOI. 10.1137/20M1349965

1. Introduction. Data assimilation [5, 42, 58] aims to improve forecasting power of dynamical systems [66] by fusing information from mathematical models and observations from nature. Ensemble Kalman filters (EnKF) [14, 15, 18, 38] have gained widespread popularity for large-scale data assimilation. They use a Monte Carlo approach to propagate covariance information and take advantage of ensemble forecasting to remove the linear model assumption in conventional Kalman filtering.

The idea of leveraging a hierarchy of models for increasing the efficiency of Monte Carlo estimation algorithms was proposed in [20, 21], and the multilevel Monte Carlo approach was successfully applied for inference with low-dimensional models. Recent work extended the multilevel idea to operational EnKF algorithms for stochastic models [9, 29, 39], proposed multilevel sampling ensemble smoothers [6], and developed multilevel particle filters [23, 24].

Reduced order modeling is the approach of constructing inexpensive surrogates able to capture the dominant dynamics of large-scale systems. Previous work has employed reduced order models (ROMs) [7, 27, 56] to speed up variational data assimilation [8, 12, 36, 44, 64, 67, 69, 71]. The underlying idea is to perform the optimization in a reduced order space and then to reconstruct this subspace around the new point

*Submitted to the journal's Methods and Algorithms for Scientific Computing section July 2, 2020; accepted for publication (in revised form) November 25, 2020; published electronically March 25, 2021.

<https://doi.org/10.1137/20M1349965>

Funding: This work was supported by NSF through grant CDS&E-MSS-1953113. The work of the first and third authors was supported by NSF through grant CCF-1613905, by DOE through grant ASCR DE-SC0021313, and by the Computational Science Laboratory at Virginia Tech. The work of the second and fourth authors was supported by NSF through grant DMS-1821145.

[†]Computational Science Laboratory, Department of Computer Science, Virginia Tech, Blacksburg VA 24061 USA (apopov@vt.edu, sandu@cs.vt.edu).

[‡]Department of Mathematics, Virginia Tech, Blacksburg VA 24061 USA (cmou@vt.edu, iliescu@vt.edu).

in the full state space. It has been shown in [64] that the reduced order basis needs to include snapshot information from both the forward and the adjoint models in order for the reduced space optimization to progress to the full order optimal point.

ROMs have also been used to develop new Markov chain Monte Carlo [11, 19, 28], Kalman filter (KF) [13], and EnKF [26, 43, 50, 70] algorithms. As opposed to variational methods, in the KF and EnKF settings ROMs have been used as replacements to traditional physics-based models. To quantify the effect of replacing the physics-based models with ROMs, rigorous error bounds were derived for both the KF [13] and the EnKF [50] algorithms.

This work develops a new multifidelity ensemble KF algorithm building on the theory of multivariate control variates [60] and on ROM data assimilation approaches [64]. Small ensembles of high-fidelity model runs are complemented by larger ensembles of cheaper, lower-fidelity runs to obtain much improved analyses at only small additional computational costs. New contributions of this work include rederiving the EnKF data assimilation approach from a multivariate linear control variate theory perspective. This perspective allows for rigorous multifidelity extensions of the EnKF, where the uncertainty in coarser levels in the hierarchy of models represents control variates for the uncertainty in finer levels. Moreover, the state of different control variates can reside in different spaces (e.g., those with different dimension and/or different inner product), which justifies the “multifidelity” [51] EnKF name given to our approach. The mapping between different spaces (i.e., the mapping of each control variate to the space of the corresponding principal variates) is done by coupling operators that can be computed in an optimal way. The paper derives an optimal statistical estimation framework in order to show significant reduction both in the cost of the method and in error.

Key innovations of the multifidelity EnKF approach as compared to the standard multilevel EnKF (MLEnKF) proposed in [9, 29] include the use of multivariate linear control variate theory [60] to rigorously incorporate all model levels in the statistical estimation approach. MLEnKF [9, 29] incorporates different model levels using signed empirical measures, which makes the multilevel covariances possibly nonpositive; in our approach the multilevel empirical covariances are always nonnegative. The use of signed empirical measures over the fine space requires MLEnKF to represent the states from all model levels into the same (fine level) space. In the proposed multifidelity EnKF approach, different control variates represent states from different model levels that live in their own spaces; the mapping between spaces is done explicitly by coupling operators that are computed in a statistically optimal manner, obtained from the theory to the optimal gains required to compute the covariance estimates. (We note that these operators are equal to identity in MLEnKF.)

The remainder of the paper is organized as follows. The data assimilation problem, control variate theory, and the ensemble KF are reviewed in section 2. Properties of the ROM as a control variate are analyzed in section 3. The multifidelity ROM EnKF algorithm is derived in section 4. The quasi-geostrophic test problem and the corresponding models of different fidelity are introduced in section 5. Numerical experiments are reported in section 6, and closing remarks are made in section 7.

2. Control variates and the ensemble KF. Consider the data assimilation problem of predicting the state of a natural phenomenon through modeling and sparse noisy nonlinear observations.

Let X_{i-1}^a be a random variable whose distribution represents (our uncertain knowledge of) the true state of the physical system, projected onto model space, at

time t_{i-1} . This knowledge is propagated to time t_i through the model dynamics $\mathcal{M}$,

$$(2.1) \quad X_i^b = \mathcal{M}_{i-1,i}(X_{i-1}^a, \mathcal{E}_i^{\mathcal{M}}) \stackrel{\text{assumed}}{=} \mathcal{M}_{i-1,i}(X_{i-1}^a),$$

where the distribution of the random variable X_i^b represents the prior knowledge of the state as time i and $\mathcal{E}_i^{\mathcal{M}}$ is a random variable quantifying stochastic effects and model errors. In this paper we assume that the model (2.1) is deterministic and exact, meaning that $\mathcal{E}_i^{\mathcal{M}} = 0$. Noisy observations of the true state $\mathbf{X}_i^t$ are collected at time t_i ,

$$(2.2) \quad Y_i = \mathcal{H}_i(\mathbf{X}_i^t, \mathcal{E}_i^{\mathcal{H}}) \stackrel{\text{assumed}}{=} \mathcal{H}_i(\mathbf{X}_i^t) + \eta_i, \quad \eta_i \sim \mathcal{N}(\mathbf{0}, \Sigma_{\eta_i, \eta_i}),$$

where $\mathcal{H}_i$ is the (non)linear observation operator and $\mathcal{E}_i^{\mathcal{H}}$ is a random variable quantifying uncertainty in the observations. We assume that the observation errors η_i are additive, unbiased, and Gaussian, with observation error covariance matrices Σ_{η_i, η_i} .

Using the prior knowledge of the state (2.1) at time t_i described by the probability density $\pi(X_i^b)$ and the likelihood of observations (2.2) described by the probability density $\pi(Y_i|X_i^b)$, the Bayesian approach gives the posterior knowledge of the state:

$$(2.3) \quad \pi(X_i^b|Y_i) \propto \pi(Y_i|X_i^b) \pi(X_i^b).$$

We seek to approximate this posterior probability density in an ensemble KF framework.

As some concepts in this paper are not present in traditional data assimilation literature, we use the following notation (slightly different from [32]) for presentation clarity. Let χ and v be random variables. The exact mean of χ is denoted by μ_χ and the empirical (sample) mean by $\tilde{\mu}_\chi$. Similarly, the exact covariance is denoted by $\Sigma_{\chi, v}$ and the sample covariance by $\tilde{\Sigma}_{\chi, v}$. An ensemble of samples of χ is denoted by $\mathbf{E}_\chi$, the i th sample by $\chi^{[i]}$, and the scaled ensemble anomalies (defined later) by $\mathbf{A}_\chi$.

2.1. Linear control variates. Consider a random variable χ with support $\Omega_\chi = \mathbb{R}^n$ with a distribution that represents the uncertainty in the state. Its mean μ_χ represents the minimum variance estimator of the true state, and its higher moments quantify the confidence in this estimator. We call χ the *principal variate*.

Consider a second random variable $\hat{v}$ with support $\Omega_{\hat{v}} = \mathbb{R}^r$, which is highly correlated with χ and has a known mean $\mu_{\hat{v}}$. This second random variable $\hat{v}$ is a *control variate* used to improve the estimate μ_χ of the true state of the system. Here we consider $r \ll n$, though this is not required in general.

Our goal is to estimate μ_χ , and the direct way is to sample the principal variate χ . The linear control variate approach seeks to obtain better estimates by taking samples of *both* the principal variate χ and the control variate $\hat{v}$. Specifically, in a multivariate linear control variate approach [60], one constructs the new random variable

$$(2.4) \quad \zeta = \chi - \mathbf{S}(\hat{v} - \mu_{\hat{v}}),$$

which we call the *total variate*, which has the same mean as the principal variate $\mu_\zeta = \mu_\chi$ but whose other moments have been modified by the knowledge of the control variate $\hat{v}$. The deterministic *gain matrix* $\mathbf{S} \in \mathbb{R}^{n \times r}$ is chosen such as to minimize the generalized variance of the total variate. We recall the following result [60, Lemma 1 in the appendix].

LEMMA 2.1 (optimal gain). *The optimal gain that minimizes the generalized variance of the total variate ζ (the determinant of $\Sigma_{\zeta,\zeta}$) is*

$$(2.5) \quad \mathbf{S} = \Sigma_{\chi,\hat{v}} \Sigma_{\hat{v},\hat{v}}^{-1}.$$

Consider now the case where the mean $\mu_{\hat{v}}$ of the control variate is unknown. However, one can sample a random variable $v \in \mathbb{R}^r$ that has the same mean and support as $\hat{v}$ but is independent of both χ and $\hat{v}$. Using (a realization of) what we call the *ancillary variate* v as a proxy for the exact mean $\mu_{\hat{v}} = \mu_v$, the total variate (2.4) becomes

$$(2.6) \quad \zeta = \chi - \mathbf{S}(\hat{v} - v).$$

Letting $\omega = \hat{v} - v + \mu_{\hat{v}}$, $\mu_{\omega} = \mu_{\hat{v}}$, and $\hat{v} - v = \omega - \mu_{\omega}$, (2.6) reduces to (2.4) with $\mu_{\hat{v}}$ replaced by ω . By Lemma 2.1 the optimal gain is

$$(2.7) \quad \mathbf{S} = \Sigma_{\chi,\hat{v}} (\Sigma_{\hat{v},\hat{v}} + \Sigma_{v,v})^{-1},$$

where $\Sigma_{\hat{v},\hat{v}} + \Sigma_{v,v} = \Sigma_{\omega,\omega}$ and $\Sigma_{\chi,\hat{v}} = \Sigma_{\chi,\omega}$.

If the control variate $\hat{v}$ and its ancillary variate v share not only the same mean but also the same covariance, $\Sigma_{v,v} = \Sigma_{\hat{v},\hat{v}}$, then (2.7) becomes

$$(2.8) \quad \mathbf{S} = \frac{1}{2} \Sigma_{\chi,\hat{v}} \Sigma_{\hat{v},\hat{v}}^{-1}.$$

Remark 1 (total variate covariance). The covariance of the total variate (2.6) using the optimal gain (2.7) is

$$(2.9) \quad \begin{aligned} \Sigma_{\zeta,\zeta} &= \Sigma_{\chi,\chi} - \Sigma_{\chi,\hat{v}} \mathbf{S}^{\top} - \mathbf{S} \Sigma_{\hat{v},\chi} + \mathbf{S} \Sigma_{\hat{v},\hat{v}} \mathbf{S}^{\top} + \mathbf{S} \Sigma_{v,v} \mathbf{S}^{\top} \\ &= \Sigma_{\chi,\chi} - \Sigma_{\chi,\hat{v}} (\Sigma_{\hat{v},\hat{v}} + \Sigma_{v,v})^{-1} \Sigma_{\hat{v},\chi}. \end{aligned}$$

Note that this is always a symmetric semipositive definite (s.p.d.) matrix that is smaller (in the s.p.d. matrix sense) than the principal variate covariance, $0 \leq \Sigma_{\zeta,\zeta} \leq \Sigma_{\chi,\chi}$. In contrast, the multilevel covariance formula for variables that live in the same space, $\Sigma_{\zeta,\zeta} = \Sigma_{\chi,\chi} - \Sigma_{\hat{v},\hat{v}} + \Sigma_{v,v}$ proposed in [29], does not necessarily enjoy these properties, as the signed empirical measure ignores cross covariances. The covariance (2.9) is s.p.d. for any matrix $\mathbf{S}$ by the construction in the proof of Lemma 2.1.

2.2. Multiple fidelities of control variates. One can recursively apply the control variate approach (2.6) to improve estimation of the mean $\mu_{\hat{v}}$. To this end, in (2.6) we identify $v_0 \equiv \chi$ and the first fidelity control and ancillary variate with $\hat{v}_1 \equiv \hat{v}$ and $v_1 \equiv v$, respectively. Next, we consider v_1 as a principal variate and use a control variate $\hat{v}_2$ to build a total variate $v_1 - \mathbf{S}_2(\hat{v}_2 - \mu_{\hat{v}_2})$. Next, we replace the exact mean $\mu_{\hat{v}_2}$ by a realization of the ancillary variate v_2 and repeat until we reach the $\mathcal{L}$ th fidelity control variate $\hat{v}_{\mathcal{L}}$ with the ancillary variate $v_{\mathcal{L}}$. This telescopic structure replaces the ancillary variate at fidelity $\ell - 1$ by a total variate constructed using the next fidelity ℓ control and ancillary variates:

$$(2.10) \quad v_{\ell-1} \leftarrow v_{\ell-1} - \mathbf{S}_{\ell}(\hat{v}_{\ell} - v_{\ell}), \quad \mathbf{S}_{\ell} = \Sigma_{v_{\ell-1},\hat{v}_{\ell}} (\Sigma_{\hat{v}_{\ell},\hat{v}_{\ell}} + \Sigma_{v_{\ell},v_{\ell}})^{-1}, \quad \ell = 1, \dots, \mathcal{L}.$$

The total variate ζ , representing a multifidelity control variate approach for the top fidelity principal variate χ , is

$$(2.11) \quad \zeta = \chi - \sum_{\ell=1}^{\mathcal{L}} \overline{\mathbf{S}}_{\ell} (\hat{v}_{\ell} - v_{\ell}), \quad \overline{\mathbf{S}}_{\ell} = \prod_{\lambda=1}^{\ell} \mathbf{S}_{\lambda}.$$

2.3. Implementation of linear control variates using ensembles. In practice, the exact distributions of χ , $\hat{v}$, and v are not available; therefore, computing the exact moments of the total variate ζ is not possible. However, we assume that one can sample from these distributions and seek to estimate the statistics of ζ .

For this, take N_χ pairwise samples $(\chi^{[k]}, \hat{v}^{[k]})$ of the principal and control variates (to be able to derive correlated statistics), and construct the ensembles $\mathbf{E}_\chi = [\chi^{[1]}, \dots, \chi^{[N_\chi]}] \in \mathbb{R}^{n \times N_\chi}$ and $\mathbf{E}_{\hat{v}} = [\hat{v}^{[1]}, \dots, \hat{v}^{[N_\chi]}] \in \mathbb{R}^{r \times N_\chi}$. Take N_v samples $\mathbf{v}^{[k]}$ of the ancillary variate, and construct the ensemble $\mathbf{E}_v = [\mathbf{v}^{[1]}, \dots, \mathbf{v}^{[N_v]}] \in \mathbb{R}^{r \times N_v}$.

The empirical means and the ensembles of anomalies are defined as

$$(2.12) \quad \begin{aligned} \tilde{\boldsymbol{\mu}}_\chi &:= N_\chi^{-1} \mathbf{E}_\chi \mathbf{1}_{N_\chi}, \quad \tilde{\boldsymbol{\mu}}_{\hat{v}} := N_\chi^{-1} \mathbf{E}_{\hat{v}} \mathbf{1}_{N_\chi}, \quad \tilde{\boldsymbol{\mu}}_v := N_v^{-1} \mathbf{E}_v \mathbf{1}_{N_v}, \\ \mathbf{A}_\chi &:= (N_\chi - 1)^{-\frac{1}{2}} \left(\mathbf{E}_\chi - \tilde{\boldsymbol{\mu}}_\chi \mathbf{1}_{N_\chi}^\top \right), \quad \mathbf{A}_{\hat{v}} := (N_\chi - 1)^{-\frac{1}{2}} \left(\mathbf{E}_{\hat{v}} - \tilde{\boldsymbol{\mu}}_{\hat{v}} \mathbf{1}_{N_\chi}^\top \right), \\ \mathbf{A}_v &:= (N_v - 1)^{-\frac{1}{2}} \left(\mathbf{E}_v - \tilde{\boldsymbol{\mu}}_v \mathbf{1}_{N_v}^\top \right), \end{aligned}$$

where $\mathbf{1}_N \in \mathbb{R}^N$ is a column vector of ones. This leads to the empirical covariances

$$(2.13) \quad \tilde{\boldsymbol{\Sigma}}_{\chi, \chi} = \mathbf{A}_\chi \mathbf{A}_\chi^\top, \quad \tilde{\boldsymbol{\Sigma}}_{\hat{v}, \hat{v}} = \mathbf{A}_{\hat{v}} \mathbf{A}_{\hat{v}}^\top, \quad \tilde{\boldsymbol{\Sigma}}_{\chi, \hat{v}} = \mathbf{A}_\chi \mathbf{A}_{\hat{v}}^\top = \tilde{\boldsymbol{\Sigma}}_{\hat{v}, \chi}^\top, \quad \tilde{\boldsymbol{\Sigma}}_{v, v} = \mathbf{A}_v \mathbf{A}_v^\top.$$

The empirical mean and covariance estimates of the total variate (2.6) are

$$(2.14) \quad \begin{aligned} \tilde{\boldsymbol{\mu}}_\zeta &= N_\chi^{-1} \sum_{k=1}^{N_\chi} \left(\chi^{[k]} - \mathbf{S} \hat{v}^{[k]} \right) + N_v^{-1} \sum_{k=1}^{N_v} \mathbf{S} \mathbf{v}^{[k]}, \\ \tilde{\boldsymbol{\Sigma}}_{\zeta, \zeta} &= \tilde{\boldsymbol{\Sigma}}_{\chi, \chi} + \mathbf{S} \tilde{\boldsymbol{\Sigma}}_{\hat{v}, \hat{v}} \mathbf{S}^\top - \tilde{\boldsymbol{\Sigma}}_{\chi, \hat{v}} \mathbf{S}^\top - \mathbf{S} \tilde{\boldsymbol{\Sigma}}_{\hat{v}, \chi} + \mathbf{S} \tilde{\boldsymbol{\Sigma}}_{v, v} \mathbf{S}^\top. \end{aligned}$$

When the exact covariances $\boldsymbol{\Sigma}_{\chi, \hat{v}}$ and $\boldsymbol{\Sigma}_{\hat{v}, \hat{v}}$ are not known but the exact covariance of the ancillary variate $\boldsymbol{\Sigma}_{v, v}$ is known, the optimal gain matrix (2.7) is approximated by

$$(2.15) \quad \mathbf{S} \approx \tilde{\mathbf{S}} = \tilde{\boldsymbol{\Sigma}}_{\chi, \hat{v}} \left(\tilde{\boldsymbol{\Sigma}}_{\hat{v}, \hat{v}} + \boldsymbol{\Sigma}_{v, v} \right)^{-1},$$

which is well defined when $\boldsymbol{\Sigma}_{v, v}$ is full rank. In the case where the underlying random variables are Gaussian, the expected value of the sampled gain matrix, $\tilde{\mathbf{S}}$, is not the exact gain matrix, even in the scalar case [54].

When $\boldsymbol{\Sigma}_{v, v}$ is also unknown and all empirical covariance estimates are undersampled, meaning that the rank of the sampled covariance is lower than the rank of the covariance of the underlying random variable, then the approximation

$$(2.16) \quad \mathbf{S} \approx \tilde{\mathbf{S}} = \tilde{\boldsymbol{\Sigma}}_{\chi, \hat{v}} \left(\tilde{\boldsymbol{\Sigma}}_{\hat{v}, \hat{v}} + \tilde{\boldsymbol{\Sigma}}_{v, v} \right)^{-1}$$

can be ill-defined, and a better approach is required to estimate the optimal gain matrix. In this case our goal will be to determine a control variate whose relation with the principal variate leads to a good approximation of the gain matrix with minimal reliance on sampling.

If the cost of obtaining one sample of the principal variate is C_χ and the cost of a sample from either the ancillary or the control variate is C_v , then the cost of a two-fidelity estimator is

$$(2.17) \quad N_\chi C_\chi + (N_\chi + N_v) C_v,$$

which, if the cost of sampling the coarser random variables is negligible $C_v \ll C_\chi$, is roughly equal to the cost of sampling the principal variate.

2.4. Ensemble KF. The traditional KF [37] aims to optimally solve the Bayesian inference problem under the assumption that the probability distributions of the prior knowledge about the state, observations, and the resulting posterior knowledge are all Gaussian. The KF also makes the assumptions that $\boldsymbol{\mu}_{X_i^b} = \mathbf{X}_i^t$ and $\boldsymbol{\mu}_{\mathcal{H}(X_i^b)} = \boldsymbol{\mu}_{Y_i}$. We now rederive the EnKF framework from a multivariate linear control variate theory perspective.

The principal variate represents our prior knowledge $\chi \equiv X_i^b$, the control variate is the model-predicted observations $\hat{v} \equiv \mathcal{H}(X_i^b)$, and the ancillary variate is the observations $v \equiv Y_i$. The goal is to estimate the true state, which is the mean of the principal variate $\boldsymbol{\mu}_\chi \equiv \boldsymbol{\mu}_{X_i^b} = \mathbf{X}_i^t$. The posterior knowledge is represented by the new, reduced variance total variate $\zeta \equiv X_i^a$ (2.6),

$$(2.18) \quad X_i^a = X_i^b - \mathbf{K}_i (\mathcal{H}(X_i^b) - Y_i),$$

where the control variate gain matrix $\mathbf{S} \equiv \mathbf{K}_i$ is the Kalman gain. The mean of the total variate is also the true state $\boldsymbol{\mu}_\zeta \equiv \boldsymbol{\mu}_{X_i^a} = \mathbf{X}_i^t = \boldsymbol{\mu}_\chi$, but its covariance is smaller.

EnKF represents the random variables by ensembles of N samples, with $\mathbf{E}_{X^b}$ and $\mathbf{E}_{\mathcal{H}(X^b)}$ defined in the usual way. The perturbed observation version of the EnKF [18] also constructs an ensemble of independent samples from the observation distribution,

$$(2.19) \quad \mathbf{E}_{Y_i} = \mathbf{Y}_i \mathbf{1}_N^T + \mathbf{A}_{\eta_i},$$

where the anomalies $\mathbf{A}_{\eta_i}$ are derived from an ensemble of N independent samples from the observation error distribution (2.2).

It is typically assumed that the only variable whose covariance is known is Y_i , meaning that the Kalman gain is approximated using (2.15):

$$(2.20) \quad \tilde{\mathbf{K}}_i = \tilde{\boldsymbol{\Sigma}}_{X_i^b, \mathcal{H}(X_i^b)} \left(\tilde{\boldsymbol{\Sigma}}_{\mathcal{H}(X_i^b), \mathcal{H}(X_i^b)} + \boldsymbol{\Sigma}_{\eta_i, \eta_i} \right)^{-1}.$$

Thus, the EnKF analysis formulas are

$$(2.21) \quad \begin{aligned} \tilde{\boldsymbol{\mu}}_{X_i^a} &= \tilde{\boldsymbol{\mu}}_{X_i^b} - \tilde{\mathbf{K}}_i \tilde{\mathbf{d}}_i, & \mathbf{A}_{X_i^a} &= \mathbf{A}_{X_i^b} - \tilde{\mathbf{K}}_i (\mathbf{A}_{\mathcal{H}(X_i^b)} - \mathbf{A}_{\eta_i}), \\ \tilde{\mathbf{d}}_i &= \tilde{\boldsymbol{\mu}}_{\mathcal{H}(X_i^b)} - \tilde{\boldsymbol{\mu}}_{Y_i}, & \mathbf{E}_{X_i^a} &= \tilde{\boldsymbol{\mu}}_{X_i^a} \mathbf{1}_N^T + (N-1)^{\frac{1}{2}} \mathbf{A}_{X_i^a}, \end{aligned}$$

with the ensemble $\mathbf{E}_{X_i^a}$ representing the posterior uncertainty at time t_i .

The number of ensemble members is usually significantly smaller than the dimension of the state space, $N \ll n$, and the covariance matrix estimate is affected by sampling errors. In order to alleviate these errors and probabilistically inaccurate assumptions about the statistical Kalman gain (2.20), methods such as inflation [1, 3, 54, 68], localization [2, 48, 52, 53], and covariance shrinkage [47, 49, 55] have been developed.

3. Spaces, projections, information, and control variates. Bayes' rule requires using all information in the inference process [34]; in particular, if additional information about the dynamics of the system is known, it must be used in the inference in order to increase confidence in the inference results. The assumption of linearity (in KF and linear control variates), however, precludes the inclusion of important information about the manifold on which nonlinear model dynamics live. ROMs construct linear subspaces that capture the most important (in some well-defined sense) features and modes of the full order dynamics. For this reason we seek to build enhanced ensemble KF with ROMs as control variates.

To this end we consider finite-dimensional random variables. Without loss of generality, the principal variate χ lives in the space $\mathcal{S}_\chi = \mathbb{R}^n$, endowed with the canonical Euclidean basis and the canonical Euclidean inner product $\langle \cdot, \cdot \rangle_{\mathcal{S}_\chi} = \langle \cdot, \cdot \rangle_{\mathbb{R}^n}$. The control and ancillary variates ($\hat{v}$ and v , respectively) are vectors in $\mathcal{S}_v = \mathbb{R}^r$, endowed with the canonical Euclidean basis and the canonical inner product $\langle \cdot, \cdot \rangle_{\mathcal{S}_v} = \langle \cdot, \cdot \rangle_{\mathbb{R}^r}$.

We consider the natural idea of utilizing a control variate that is the projection of the principal variate $\chi \in \mathcal{S}_\chi$ onto an r -dimensional subspace $\hat{\mathcal{S}}_v \subset \mathcal{S}_\chi$ that captures the dominant features of the nonlinear dynamics of the system.

We identify the space of control and ancillary variates $\mathcal{S}_v = \mathbb{R}^r$ with an r -dimensional subspace $\hat{\mathcal{S}}_v \subset \mathcal{S}_\chi$ equipped with the $\langle \cdot, \cdot \rangle_{\hat{\mathcal{S}}_v} = \langle \cdot, \cdot \rangle_{\mathbf{M}_v}$ inner product, where $\mathbf{M}_v \in \mathbb{R}^{n \times n}$ is a s.p.d. matrix. Specifically, let $\Phi = [\Phi_1, \dots, \Phi_n] \in \mathbb{R}^{n \times n}$ be an $\mathbf{M}_v$ -orthogonal basis of $\mathbb{R}^n$; we identify the control space with the span of the first r vectors in the basis $\hat{\mathcal{S}}_v = \text{Span}\{\Phi_1, \dots, \Phi_r\}$. Consider two vectors in the control space $\mathbf{u}, \mathbf{v} \in \mathbb{R}^r$; their representations as n -dimensional vectors in $\mathcal{S}_v$ are $\Phi_r \mathbf{u}$ and $\Phi_r \mathbf{v}$, respectively, where $\Phi_r = [\Phi_1, \dots, \Phi_r] \in \mathbb{R}^{n \times r}$. The dot product is preserved in both representations:

$$(3.1) \quad \langle \mathbf{u}, \mathbf{v} \rangle_{\mathbb{R}^r} = \mathbf{u}^\top \mathbf{v} = \mathbf{u}^\top \Phi_r^\top \mathbf{M}_v \Phi_r \mathbf{v} = \langle \Phi_r \mathbf{u}, \Phi_r \mathbf{v} \rangle_{\mathbf{M}_v}.$$

Remark 2. There is no loss of generality with the above formulation. Consider the control space $\mathcal{S}_v = \mathbb{R}^r$ endowed with the general inner product $\langle \cdot, \cdot \rangle_{\mathbf{N}_v}$ and the $\mathbf{N}_v$ -orthonormal basis $\{\varphi_1, \dots, \varphi_r\}$. Identify the control space with the r -dimensional subspace $\hat{\mathcal{S}}_v = \text{span}\{\hat{\Phi}_1, \dots, \hat{\Phi}_r\} \subset \mathbb{R}^n$, where $\hat{\Phi}_r = [\hat{\Phi}_1, \dots, \hat{\Phi}_r] \in \mathbb{R}^{n \times r}$ are the control basis vectors represented as vectors in $\mathbb{R}^n$. The following change of basis casts this general case in our formulation:

$$\Phi_r = \mathbf{M}_v^{-\frac{1}{2}} \hat{\Phi}_r \mathbf{N}_v^{-\frac{1}{2}}, \quad \Phi_r^\top \mathbf{M}_v \Phi_r = \mathbf{I}_{r \times r}, \quad \text{Range}(\Phi_r) = \text{Range}(\hat{\Phi}_r) = \hat{\mathcal{S}}_v,$$

where $\mathbf{M} = \mathbf{M}^{\frac{1}{2}} \mathbf{M}^{\frac{1}{2}}$ is a square root factorization of the s.p.d. matrix $\mathbf{M}$.

Remark 3. The transformed vectors $\Psi = \mathbf{M}_v^{\frac{1}{2}} \Phi$ form an orthonormal basis of $\mathcal{S}_\chi$, and the first r vectors of Ψ form an orthonormal basis of $\hat{\mathcal{S}}_v$ w.r.t. the Euclidean dot product:

$$\Psi_r = \mathbf{M}_v^{\frac{1}{2}} \Phi_r, \quad \Psi_r^\top \Psi_r = \mathbf{I}_{r \times r}, \quad \text{Range}(\Psi_r) = \text{Range}(\Phi_r) = \hat{\mathcal{S}}_v.$$

In summary, a control vector $\mathbf{u} \in \mathbb{R}^r$ is represented in the principal space $\mathbb{R}^n$ as

$$(3.2a) \quad \mathbf{x} = \Phi_r \mathbf{u}.$$

Vice versa, a vector in the principal space $\mathbf{x} \in \mathbb{R}^n$ is projected $\mathbf{M}_v$ -orthogonally onto the control space $\mathbb{R}^r$ as follows:

$$(3.2b) \quad \Phi_r^* := \Phi_r^\top \mathbf{M}_v, \quad \mathbf{u} = \Phi_r^* \mathbf{x} = \Phi_r^\top \mathbf{M}_v \mathbf{x}.$$

We note that $\Phi_r^* \in \mathbb{R}^{r \times n}$ is the adjoint operator of $\Phi_r \in \mathbb{R}^{n \times r}$ w.r.t. the control dot products: $\langle \Phi_r^* \mathbf{x}, \mathbf{u} \rangle_{\mathbb{R}^r} = \langle \mathbf{x}, \Phi_r \mathbf{u} \rangle_{\mathbf{M}_v}$.

Remark 4. The method of snapshots [63] that underpins the ROM finds an $\mathbf{M}_v$ -orthonormal basis $\Phi_1, \dots, \Phi_n$ of $\mathbb{R}^n$ with vectors sorted in decreasing order of importance (e.g., with respect to the energy of the dynamical system solution projected

onto that vector). In the method of snapshots, Φ_i are the eigenvectors of the temporal covariance of the dynamics (discretely approximated by the snapshot covariance). A full-state vector $\mathbf{x} = \sum_{i=1}^n u_i \Phi_i \in \mathbb{R}^n$ is given a reduced order approximation by keeping only the main r components: $\mathbf{x}_r = \sum_{i=1}^r u_i \Phi_i$. This is equivalent to projecting the vector $\mathbf{M}_v$ -orthonormally onto the first r basis vectors, $\mathbf{u} = [u_1 \dots u_r]^\top = \Phi_r^* \mathbf{x}$.

Consider the case where the ensemble size of the principal variate–control variate pair is insufficient to accurately determine their statistical covariances. In this case one cannot accurately determine the statistical analogue of the optimal gain (2.16) at any given point in time. To overcome this difficulty we leverage the projection operators defined in this section in order to describe both the control variate and the corresponding optimal gain.

THEOREM 3.1. *Let the control variate (2.4) be the projection of the principal variate over $\hat{\mathcal{S}}_v$:*

$$(3.3) \quad \hat{v} = \Phi_r^* \chi.$$

The principal and control variates in the Φ basis read

$$(3.4) \quad \chi = \Phi_r \Phi_r^* \chi + (\mathbf{I} - \Phi_r \Phi_r^*) \chi = \Phi_r \hat{v} + \Delta \chi_r.$$

The optimal gain for the total variate (2.5) is

$$(3.5) \quad \mathbf{S}^{\text{opt}} = \Phi_r + \Sigma_{\Delta \chi_r, \hat{v}} (\Sigma_{\hat{v}, \hat{v}})^{-1}.$$

Using the approximate gain matrix

$$(3.6) \quad \mathbf{S} = \Phi_r$$

in (2.4) removes the variability of χ within $\hat{\mathcal{S}}_v$.

Proof. From (3.4) we have

$$(\chi - \mu_\chi)(\hat{v} - \mu_{\hat{v}})^T = \Phi_r (\hat{v} - \mu_{\hat{v}})(\hat{v} - \mu_{\hat{v}})^T + (\Delta \chi_r - \mu_{\Delta \chi_r})(\hat{v} - \mu_{\hat{v}})^T.$$

Taking expected values and replacing in the optimal gain formula (2.5) gives (3.5). From (2.4),

$$\zeta = \chi - \mathbf{S}(\hat{v} - \mu_{\hat{v}}) = (\mathbf{I} - \mathbf{S} \Phi_r^*) \chi + \mathbf{S} \mu_{\hat{v}} = (\Phi_r - \mathbf{S}) \hat{v} + \Delta \chi_r + \mathbf{S} \mu_{\hat{v}},$$

and the approximate gain (3.6) leads to $\Sigma_{\zeta, \zeta} = \Sigma_{\Delta \chi_r, \Delta \chi_r}$. Since $\Delta \chi_r = (\mathbf{I} - \Phi_r \Phi_r^*) \chi$ is $\mathbf{M}_v$ -orthogonal to $\hat{\mathcal{S}}_v$, the variability of χ within $\hat{\mathcal{S}}_v$ has been removed. $\square$

A consequence of Theorem 3.1 is that the approximate gain (3.6) is constant in time.

Remark 5 (approximation of optimal gain). If the mean of the control variate is unknown, then an ancillary variate is used (2.6). If the ancillary variate has a second moment that is equal to that of the control variate, then by (2.8) the optimal gain is approximately $\mathbf{S} \approx \Phi_r/2$.

Remark 6 (gain error). In (3.4) the reduced order approximation error is

$$\Delta \chi_r = \chi - \Phi_r \hat{v}.$$

The fixed gain (3.6) is a good approximation of the optimal gain (3.5) when the term $\Sigma_{\Delta \chi_r, \hat{v}} (\Sigma_{\hat{v}, \hat{v}})^{-1}$ is small, i.e., when the covariance between the approximation error $\Delta \chi_r$ and the reduced order projection $\hat{v}$ is small relative to the covariance of the reduced order projection.

Remark 7. Upper bounds for the error between the forecasted full model state and the forecasted ROM state that functions as a control variate are available in the literature [33, 41, 62]. Assume that the deviations from the mean of $\Delta\chi_r$ and $\hat{v}$ are bounded by a moderate constant times the respective means. A simple scale analysis in (3.5) shows that

$$\|\mathbf{S}^{\text{opt}} - \mathbf{S}\| = \|\Sigma_{\Delta\chi_r, \hat{v}} (\Sigma_{\hat{v}, \hat{v}})^{-1}\| \sim \frac{\|\Delta\chi_r\|}{\|\hat{v}\|} = \frac{\|\Delta\chi_r\|}{\|\chi - \Delta\chi_r\|} \leq \frac{\|\Delta\chi_r\|/\|\chi\|}{1 - (\|\Delta\chi_r\|/\|\chi\|)},$$

so the smaller the ROM error is, the closer the fixed approximate gain (3.6) is to the optimal one (3.5).

Remark 8. The discussion in this section applies with minor changes to the infinite-dimensional case. Consider an infinite-dimensional principal space $\mathcal{S}_\chi$ with an inner product $\langle \cdot, \cdot \rangle_{\mathcal{S}_\chi}$ and a possibly infinite-dimensional control space $\mathcal{S}_v$ with the inner product $\langle \cdot, \cdot \rangle_{\mathcal{S}_v}$. Consider a second dot product $\langle \cdot, \cdot \rangle_{\hat{\mathcal{S}}_v}$ on $\mathcal{S}_\chi$ (motivated by the physics of the problem). A linear bounded operator $\Phi_r : \mathcal{S}_v \rightarrow \mathcal{S}_\chi$ links the control and primal spaces; let $\hat{\mathcal{S}}_v = \text{Range}\{\Phi_r\}$. The adjoint operator $\Phi_r^* : \mathcal{S}_\chi \rightarrow \mathcal{S}_v$ defined by $\langle \Phi_r^* \mathbf{x}, \mathbf{u} \rangle_{\mathcal{S}_v} = \langle \mathbf{x}, \Phi_r \mathbf{u} \rangle_{\hat{\mathcal{S}}_v}$ gives the control variate relation (3.3).

Algorithm 4.1 Pseudocode for two-level EnKF with a ROM control variate.

Require: The ensembles $\mathbf{E}_{X_i^b}$, $\mathbf{E}_{\hat{U}_i^b}$, and $\mathbf{E}_{U_i^b}$, the projection operator Φ_r . Computes the analysis ensembles $\mathbf{E}_{X_i^a}$, $\mathbf{E}_{\hat{U}_i^a}$, and $\mathbf{E}_{U_i^a}$.

- 1: $\tilde{\mu}_{Z_i^b} := \tilde{\mu}_{X_i^b} - \frac{1}{2} \Phi_r (\tilde{\mu}_{\hat{U}_i^b} - \tilde{\mu}_{U_i^b})$
 - 2: $\tilde{\Sigma}_{Z_i^b, \mathcal{H}_i(Z_i^b)} := \tilde{\Sigma}_{X_i^b, \mathcal{H}_i(X_i^b)} + \frac{1}{4} \Phi_r \tilde{\Sigma}_{\hat{U}_i^b, \mathcal{H}(\Phi_r \hat{U}_i^b)} - \frac{1}{2} \tilde{\Sigma}_{X_i^b, \mathcal{H}(\Phi_r \hat{U}_i^b)} - \frac{1}{2} \Phi_r \tilde{\Sigma}_{\hat{U}_i^b, \mathcal{H}(X_i^b)} + \frac{1}{4} \Phi_r \tilde{\Sigma}_{\hat{U}_i^b, \mathcal{H}(\Phi_r U_i^b)}$
 - 3: $\tilde{\Sigma}_{\mathcal{H}_i(Z_i^b), \mathcal{H}_i(Z_i^b)} := \tilde{\Sigma}_{\mathcal{H}_i(X_i^b), \mathcal{H}_i(X_i^b)} + \frac{1}{4} \tilde{\Sigma}_{\mathcal{H}(\Phi_r \hat{U}_i^b), \mathcal{H}(\Phi_r \hat{U}_i^b)} - \frac{1}{2} \tilde{\Sigma}_{\mathcal{H}(X_i^b), \mathcal{H}(\Phi_r \hat{U}_i^b)} - \frac{1}{2} \tilde{\Sigma}_{\mathcal{H}(\Phi_r \hat{U}_i^b), \mathcal{H}(X_i^b)} + \frac{1}{4} \tilde{\Sigma}_{\mathcal{H}(\Phi_r U_i^b), \mathcal{H}(\Phi_r U_i^b)}$
 - 4: $\tilde{\mathbf{K}}_i := \tilde{\Sigma}_{Z_i^b, \mathcal{H}_i(Z_i^b)} \left(\tilde{\Sigma}_{\mathcal{H}_i(Z_i^b), \mathcal{H}_i(Z_i^b)} + \Sigma_{\eta_i, \eta_i} \right)^{-1}$
 - 5: Generate $\mathbf{E}_{\eta_i^X}$, $\mathbf{E}_{\eta_i^U}$, $\eta_i^X \sim \mathcal{N}(\mathbf{0}, \Sigma_{\eta_i^X, \eta_i^X})$, $\eta_i^U \sim \mathcal{N}(\mathbf{0}, \Sigma_{\eta_i^U, \eta_i^U})$
 - 6: $\mathbf{A}_{X_i^a} := \mathbf{A}_{X_i^b} - \tilde{\mathbf{K}}_i \left(\mathbf{A}_{\mathcal{H}_i(X_i^b)} - \mathbf{A}_{\eta_i^X} \right)$
 - 7: $\mathbf{A}_{\hat{U}_i^a} := \mathbf{A}_{\hat{U}_i^b} - \Phi_r^* \tilde{\mathbf{K}}_i \left(\mathbf{A}_{\mathcal{H}_i(\Phi_r \hat{U}_i^b)} - \mathbf{A}_{\eta_i^X} \right)$
 - 8: $\mathbf{A}_{U_i^a} := \mathbf{A}_{U_i^b} - \Phi_r^* \tilde{\mathbf{K}}_i \left(\mathbf{A}_{\mathcal{H}_i(\Phi_r U_i^b)} - \mathbf{A}_{\eta_i^U} \right)$
 - 9: $\tilde{\mu}_{Z_i^a} := \tilde{\mu}_{Z_i^b} - \tilde{\mathbf{K}}_i (\tilde{\mu}_{\mathcal{H}_i(Z_i^b)} - \mathbf{Y}_i)$
 - 10: $\mathbf{E}_{X_i^a} := \tilde{\mu}_{Z_i^a} + \sqrt{N_X - 1} \mathbf{A}_{X_i^a}$, $\mathbf{E}_{\hat{U}_i^a} := \Phi_r^* \tilde{\mu}_{Z_i^a} + \sqrt{N_X - 1} \mathbf{A}_{\hat{U}_i^a}$
 - 11: $\mathbf{E}_{U_i^a} := \Phi_r^* \tilde{\mu}_{Z_i^a} + \sqrt{N_U - 1} \mathbf{A}_{U_i^a}$
-

4. Multifidelity EnKF with ROM control variates. We now build a multifidelity EnKF using the multivariate control variate framework, with the transitions between fidelities defined in terms of optimal gains, leading to the multifidelity approach, which is different than the MLEnKF idea discussed in [29].

For ease of exposition, a two-fidelity variant of the MFEnKF with ROM control variates is discussed first, and a telescopic generalization to $\mathcal{L}$ fidelities is presented later. The schematic working of a two levels of fidelity MFEnKF is illustrated in Figure 1. The pseudocode is provided in Algorithm 4.1.

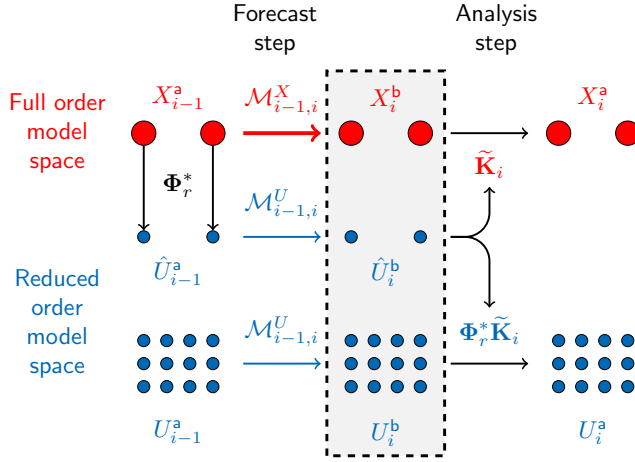


FIG. 1. Conceptual working of a two-fidelity MFEnKF with a ROM control variate.

Assumption 1 (setting for constructing MFEnKF).

1. Two numerical models (2.1) of the same natural phenomenon are available. The first one is the full order model (FOM), which propagates a state $X \in \mathbb{R}^n$ in the full order space via the FOM dynamics $\mathcal{M}^X$. The second one is a ROM that propagates a reduced order state $U \in \mathbb{R}^r$ via the ROM dynamics $\mathcal{M}^U$. The distribution of X embodies our knowledge about the state represented in the FOM space, and the distribution of U embodies our knowledge about the state (represented in the ROM space).
2. Projection operators (3.2) are available that map the full space onto the reduced one, $U = \Phi_r^* X$, and the reduced space into the full one, $X = \Phi_r U$, such that $\Phi_r^* \Phi_r = \mathbf{I}_r$.
3. A full space observation operator $\mathcal{H}_i$ maps the FOM state space onto the observation space. A reduced space observation operator $\mathcal{H}_{r,i}$ maps the ROM state space onto the observation space. The observation errors have covariances Σ_{η_i, η_i} and $\Sigma_{\eta_i^U, \eta_i^U}$, respectively. The reduced space observation operator is assumed to be consistent with the full space observation operator in the sense that

$$\mathcal{H}_{r,i}(U_i^b) \approx \mathcal{H}_i(\Phi_r U_i^b) \Rightarrow \mathcal{H}'_{r,i}|_{U_i^b} \approx \mathcal{H}'_i|_{\Phi_r U_i^b} \Phi_r.$$

Our goal is to build an ensemble KF that takes advantage of two models and that can leverage the higher accuracy of the FOM and the lower cost of the ROM by using small ensembles of FOM runs in conjunction with large ensembles of ROM runs. A first possible approach is to use EnKF in the FOM space and employ multimodel ensembles to build empirical covariances. However, building empirical moments from ensemble members of different sizes is challenging. For example, one needs to project the ROM ensemble members into the full space and carry out the inference there. A second possible approach is to “stack” the two models and obtain a supermodel that advances the combined ROM and FOM states, stack the observation operators, and apply EnKF in the aggregated state space. This second approach, however, cannot employ different numbers of FOM and ROM ensemble members and is likely to suffer when the ROM and FOM solutions are poorly correlated. A third approach is multilevel EnKF (MLEnKF) [9, 29], where EnKF is applied in the FOM space and ROM runs are (only) used to improve the empirical covariance estimates.

MFEnKF takes advantage of the availability of two models by employing a control variate framework. The FOM state $X \equiv \chi$ is the principal variate. The ROM state $\hat{U} \equiv \hat{v}$ is the control variate, and an independent ROM state $U \equiv v$ is its ancillary variate. We focus on projection control variates $\hat{U} = \Phi_r^* X$. The total variates Z_i^b and Z_i^a represent our combined prior and posterior knowledge, respectively, at time t_i through the linear control variate technique (2.4):

$$(4.1) \quad Z_i^b = X_i^b - \mathbf{S}_i (\hat{U}_i^b - U_i^b), \quad Z_i^a = X_i^a - \mathbf{S}_i (\hat{U}_i^a - U_i^a).$$

The main idea of MFEnKF is to apply EnKF (2.21) to the total variate (4.1) under the following restrictions.

Assumption 2 (restrictions in constructing MFEnKF).

1. One can run the FOM and the ROM, but there is no dynamical model associated with the total variate (4.1). Consequently, one does not sample Z directly. Rather, the uncertainty in the total variate is represented by the three ensembles of principal, control, and ancillary variates:

$$(4.2) \quad Z^b \approx (\mathbf{E}_{X^b}, \mathbf{E}_{\hat{U}^b}, \mathbf{E}_{U^b}), \quad Z^a \approx (\mathbf{E}_{X^a}, \mathbf{E}_{\hat{U}^a}, \mathbf{E}_{U^a}).$$

The MFEnKF forecast step propagates the three ensembles forward in time to obtain a representation of the prior total variate, and the analysis step produces three ensembles representing the posterior total variate.

2. One can observe the principal variate using the full space observation operator $\mathcal{H}_i$ and the control and ancillary variates using the reduced space observation operator $\mathcal{H}_{r,i}(U_i^b)$. However, one does not observe the total variate Z_i^b directly. Instead, we consider the following indirect observation operator:

$$(4.3) \quad \bar{\mathcal{H}}_i(Z_i^b) = \mathcal{H}_i(X_i^b) - \mathbf{T}_i (\mathcal{H}_{r,i}(\hat{U}_i^b) - \mathcal{H}_{r,i}(U_i^b)).$$

We are interested in indirect observations (4.3) that approximate, to first order, the nonlinear observation operator $\mathcal{H}_i$ applied to the total variate (4.1):

$$(4.4a) \quad \begin{aligned} \mathcal{H}_i(Z_i^b) &= \mathcal{H}_i(X_i^b) - \mathcal{H}'_i|_{X_i^b} \mathbf{S}_i (\hat{U}_i^b - U_i^b) + \text{h.o.t.}, \\ \bar{\mathcal{H}}_i(Z_i^b) &= \mathcal{H}_i(X_i^b) - \mathbf{T}_i \mathcal{H}'_{r,i}|_{\Phi_r^* X_i^b} (\hat{U}_i^b - U_i^b) + \text{h.o.t.} \end{aligned}$$

This is achieved by choosing a matrix $\mathbf{T}_i$ such that

$$(4.4b) \quad \mathcal{H}'_i|_{X_i^b} \mathbf{S}_i \approx \mathbf{T}_i \mathcal{H}'_{r,i}|_{\Phi_r^* X_i^b} \approx \mathbf{T}_i \mathcal{H}'_i|_{\Phi_r \Phi_r^* X_i^b} \Phi_r.$$

4.1. Forecast step. In order to ensure that the analysis control variate $\hat{U}_{i-1}^a$ is highly correlated with the corresponding principal variate X_{i-1}^a , $\hat{U}_{i-1}^a$ is not obtained through EnKF analysis (2.21); rather, it is obtained by projecting the principal variate (the FOM analysis state) onto the reduced space:

$$(4.5a) \quad \hat{U}_{i-1}^a := \Phi_r^* X_{i-1}^a.$$

The MFEnKF forecast step propagates each of the three analysis ensembles (4.2) at time t_{i-1} forward to time t_i :

$$(4.5b) \quad \begin{aligned} \mathbf{X}_i^{b,[k]} &= \mathcal{M}_{i-1,i}^X(\mathbf{X}_{i-1}^{a,[k]}), & \hat{\mathbf{U}}_i^{b,[k]} &= \mathcal{M}_{i-1,i}^U(\hat{\mathbf{U}}_{i-1}^{a,[k]}), & k &= 1, \dots, N_X; \\ \mathbf{U}_i^{b,[k]} &= \mathcal{M}_{i-1,i}^U(\mathbf{U}_{i-1}^{a,[k]}), & & & k &= 1, \dots, N_U. \end{aligned}$$

The MFEnKF forecast step (4.5) is illustrated in Figure 1.

Remark 9 (assumption of independence). In the control variate framework the control $\hat{U}_i^b$ and ancillary U_i^b variates are independent random variables. In MFEnKF the analysis step will correlate the principal X_i^b and the ancillary variates. Nevertheless, using typical statistical Kalman gain independence assumptions in the EnKF, we will treat $\hat{U}_i^b$ and U_i^b as independent in MFEnKF calculations.

Remark 10 (forecast step and ROM bias). For linear models, another control variate highly correlated with the principal variate can be obtained by direct projection

$$(4.6) \quad \hat{U}_i^b = \Phi_r^* X_i^b,$$

saving the additional ROM runs for $\hat{U}_i^b$ required by (4.5b). In general, however, ROMs are affected by systematic bias

$$\mathcal{M}_{i-1,i}^U(\Phi_r^* X_{i-1}) - \Phi_r^* \mathcal{M}_{i-1,i}^X(X_{i-1}) = \beta_i.$$

While the ancillary variate U_i^b computed using the ROM (4.5b) is affected by this bias, the control variate (4.6) obtained by direct projection is not, and in general (4.6) violates the underlying probabilistic assumptions (4.1).

In contrast, the forecasting strategy (4.5) computes both the control variate $\hat{U}_i^b$ as well as the ancillary variate U_i^b as solutions of the same ROM model (4.5b). Consequently, they are both affected by the ROM bias. If β_i is independent of the ROM state, then the biases in control and ancillary variates cancel each other out in (4.1); if the bias is not constant, this strategy is still likely to significantly reduce it.

4.2. Analysis step. We focus on the case where the control variate is $\hat{U}_i^b = \Phi_r^* X_i^b$, the reduced observation operator is $\mathcal{H}_{r,i}(U_i^b) = \mathcal{H}_i(\Phi_r U_i^b)$, and the gain matrix is $\mathbf{S}_i = \Phi_r/2$, per Remark 5. In this case (4.4b) is satisfied exactly by $\mathbf{T}_i = 1/2$, and the indirect observation operator (4.3) reads

$$(4.7) \quad \bar{\mathcal{H}}_i(Z_i^b) = \mathcal{H}_i(X_i^b) - \frac{1}{2}\mathcal{H}_{r,i}(\hat{U}_i^b) + \frac{1}{2}\mathcal{H}_{r,i}(U_i^b).$$

Using Remark 9 we have that

$$\begin{aligned} \tilde{\mu}_{\bar{\mathcal{H}}_i(Z_i^b)} &= \tilde{\mu}_{\mathcal{H}_i(X_i^b)} - \frac{1}{2}\tilde{\mu}_{\mathcal{H}_i(\Phi_r \hat{U}_i^b)} + \frac{1}{2}\tilde{\mu}_{\mathcal{H}_i(\Phi_r U_i^b)}, \\ \tilde{\Sigma}_{\bar{\mathcal{H}}_i(Z_i^b), \bar{\mathcal{H}}_i(Z_i^b)} &= \tilde{\Sigma}_{\mathcal{H}_i(X_i^b), \mathcal{H}_i(X_i^b)} + \frac{1}{4}\tilde{\Sigma}_{\mathcal{H}_i(\Phi_r \hat{U}_i^b), \mathcal{H}_i(\Phi_r \hat{U}_i^b)} \\ &\quad - \frac{1}{2}\tilde{\Sigma}_{\mathcal{H}_i(X_i^b), \mathcal{H}_i(\Phi_r \hat{U}_i^b)} - \frac{1}{2}\tilde{\Sigma}_{\mathcal{H}_i(\Phi_r \hat{U}_i^b), \mathcal{H}_i(X_i^b)} + \frac{1}{4}\tilde{\Sigma}_{\mathcal{H}_i(\Phi_r U_i^b), \mathcal{H}_i(\Phi_r U_i^b)}. \end{aligned}$$

The covariance $\tilde{\Sigma}_{Z_i^b, \bar{\mathcal{H}}_i(Z_i^b)}$ is defined in a similar manner, and the empirical Kalman gain for the total variate is computed as follows:

$$(4.8) \quad \tilde{\mathbf{K}}_i := \tilde{\Sigma}_{Z_i^b, \bar{\mathcal{H}}_i(Z_i^b)} \left(\tilde{\Sigma}_{\bar{\mathcal{H}}_i(Z_i^b), \bar{\mathcal{H}}_i(Z_i^b)} + \Sigma_{\eta_i^Z, \eta_i^Z} \right)^{-1}.$$

The perturbed observation EnKF (2.21) is applied using the indirect observations (4.7) to estimate the total variate (4.1),

$$(4.9) \quad Z_i^a = Z_i^b - \tilde{\mathbf{K}}_i (\bar{\mathcal{H}}_i(Z_i^b) - Y_i - \eta_i^Z),$$

where η_i^Z is an independent variable that represents the perturbations to be added to the indirect observations. We make the ansatz

$$(4.10) \quad \eta_i^Z = \eta_i^X - \frac{1}{2}\eta_i^{\hat{U}} + \frac{1}{2}\eta_i^U$$

such that we have

$$\bar{\mathcal{H}}_i(Z_i^b) - \eta_i^Z = (\mathcal{H}_i(X_i^b) - \eta_i^X) - \frac{1}{2}(\mathcal{H}_i(\Phi_r \hat{U}_i^b) - \eta_i^{\hat{U}}) + \frac{1}{2}(\mathcal{H}_i(\Phi_r U_i^b) - \eta_i^U).$$

The MFEnKF analysis step transforms the three background ensembles (4.2) into three posterior ensembles. Using the EnKF update (4.9), the representation of the total variates (4.1), and the representation of the observation error (4.10), we have

$$\begin{aligned} & \underbrace{X_i^a - \frac{1}{2}\Phi_r(\hat{U}_i^a - U_i^a)}_{Z_i^a} \\ &= \underbrace{X_i^b - \frac{1}{2}\Phi_r(\hat{U}_i^b - U_i^b)}_{Z_i^b} \\ & \quad - \underbrace{\tilde{\mathbf{K}}_i \left((\mathcal{H}_i(X_i^b) - Y_i + \eta_i^X) - \frac{1}{2}(\mathcal{H}_{r,i}(\hat{U}_i^b) - Y_i - \eta_i^{\hat{U}}) + \frac{1}{2}(\mathcal{H}_{r,i}(U_i^b) - Y_i - \eta_i^U) \right)}_{\bar{\mathcal{H}}_i(Z_i^b) - Y_i - \eta_i}. \end{aligned}$$

Under the assumption that the all the information of Z^b in the orthogonal complement control space $\hat{\mathcal{S}}_U^\perp$ does not affect the analysis control and ancillary variates, the MFEnKF transforms each of the variables (4.2) as follows:

$$\begin{aligned} X_i^a &= X_i^b - \tilde{\mathbf{K}}_i(\mathcal{H}_i(X_i^b) - Y_i + \eta_i^X), \\ \hat{U}_i^a &= \hat{U}_i^b - \Phi_r^* \tilde{\mathbf{K}}_i(\mathcal{H}_{r,i}(\hat{U}_i^b) - Y_i + \eta_i^{\hat{U}}), \\ U_i^a &= U_i^b - \Phi_r^* \tilde{\mathbf{K}}_i(\mathcal{H}_{r,i}(U_i^b) - Y_i + \eta_i^U). \end{aligned}$$

The background and analysis means of the total variate (4.1) are, respectively,

$$(4.11) \quad \tilde{\mu}_{Z_i^b} = \tilde{\mu}_{X_i^b} - \frac{1}{2}\Phi_r(\tilde{\mu}_{\hat{U}_i^b} - \tilde{\mu}_{U_i^b}), \quad \tilde{\mu}_{Z_i^a} = \tilde{\mu}_{Z_i^b} - \tilde{\mathbf{K}}_i(\tilde{\mu}_{\bar{\mathcal{H}}_i(Z_i^b)} - Y_i).$$

The MFEnKF analysis step (4.5) is illustrated in Figure 1.

We consider two interpretations of the error in the indirect observations, which lead to different distributions of observation perturbations. Approach (i), called “total variate uncertainty consistency,” interprets inference as occurring only on Z_i^b , $\mathcal{H}_i(Z_i^b)$, and Y , with all other variates being a means to an end. Approach (ii), called “control space uncertainty consistency,” interprets the total variate as a means to an end and focuses on the inference on primary and ancillary variates.

We first discuss approach (i), the total variate uncertainty consistency. We require that $\bar{\mathcal{H}}_i(\cdot) \approx \mathcal{H}_i(\cdot)$ (4.4a) and that both operators have the same distribution of the observation errors:

$$(4.12) \quad \Sigma_{\eta_i^Z, \eta_i^Z} = \Sigma_{\eta_i^X, \eta_i^X} = \Sigma_{\eta_i, \eta_i}.$$

To maintain the independence of the ancillary variate of both the principal and the control variates, we make the natural assumption that η_i^U is independent of η_i^X and $\eta_i^{\hat{U}}$. Consequently,

$$(4.13) \quad \Sigma_{\eta_i^Z, \eta_i^Z} = \Sigma_{\eta_i^X, \eta_i^X} + \frac{1}{4} \Sigma_{\eta_i^{\hat{U}}, \eta_i^{\hat{U}}} + \frac{1}{4} \Sigma_{\eta_i^U, \eta_i^U} - \frac{1}{2} \Sigma_{\eta_i^{\hat{U}}, \eta_i^X} - \frac{1}{2} \Sigma_{\eta_i^X, \eta_i^{\hat{U}}}.$$

To support the projection assumption (4.5a), we select $\eta^{\hat{U}} = \eta^X$. From (4.13) and (4.12) we infer that $\Sigma_{\eta_i^U, \eta_i^U} = 3\Sigma_{\eta_i, \eta_i}$, and therefore

$$(4.14) \quad \eta_i^X = \eta_i^{\hat{U}} \sim \mathcal{N}(\mathbf{0}, \Sigma_{\eta_i, \eta_i}), \quad \eta_i^U \sim \mathcal{N}(\mathbf{0}, 3\Sigma_{\eta_i, \eta_i}).$$

Note that replacing the analysis control variate with the projection of the analysis principal variate (4.5a) leads to a second possible definition of the analysis total variate:

$$Z_i^a = X_i^a - \frac{1}{2} \Phi_r(\hat{U}_i^a - U_i^a), \quad \tilde{Z}_i^a = X_i^a - \frac{1}{2} \Phi_r(\Phi_r^* X_i^a - U_i^a).$$

The choice of observation perturbations η^X , $\eta_i^{\hat{U}}$, and η^U in method (i) ensures the total variate uncertainty consistency:

$$(4.15) \quad \Sigma_{Z_i^a, Z_i^a} = \Sigma_{\tilde{Z}_i^a, \tilde{Z}_i^a}.$$

However, in this view the inference on the ancillary variate has no direct physical meaning, and the assumed ancillary observation error differs from the one used to construct the Kalman gain.

We now discuss approach (ii). In this view the total variate is a means to an end. One runs multiple EnKFs for X_i^a , $\hat{U}_i^a$, and U_i^a . Observations are taken in the full order space and in the reduced order space. The observations of control and ancillary variates use the same operator $\mathcal{H}_{r,i}(\cdot)$, and therefore the errors have the same covariance Σ_{η^U, η^U} . The observation errors are assumed to be

$$\eta_i^X \sim \mathcal{N}(\mathbf{0}, \Sigma_{\eta_i, \eta_i}), \quad \eta_i^U \sim \mathcal{N}(\mathbf{0}, \Sigma_{\eta_i^U, \eta_i^U}), \quad \eta_i^{\hat{U}} = \Sigma_{\eta_i^U, \eta_i^U}^{1/2} \Sigma_{\eta_i, \eta_i}^{-1/2} \eta_i^X \sim \mathcal{N}(\mathbf{0}, \Sigma_{\eta_i^U, \eta_i^U})$$

such that $\eta_i^{\hat{U}}$ and η_i^X are highly correlated but η_i^U and η_i^X are independent. From (4.13) the covariance of the total variate observation error is

$$\Sigma_{\eta_i^Z, \eta_i^Z} = \Sigma_{\eta_i, \eta_i} + \frac{1}{2} \Sigma_{\eta_i^U, \eta_i^U} - \frac{1}{2} \Sigma_{\eta_i^U, \eta_i^U}^{1/2} \Sigma_{\eta_i, \eta_i}^{1/2} - \frac{1}{2} \Sigma_{\eta_i, \eta_i}^{1/2} \Sigma_{\eta_i^U, \eta_i^U}^{1/2}.$$

If the errors of the reduced space observations are specified, then the above formula can be used to construct the empirical Kalman gain. For simplicity, we consider in this paper that $\Sigma_{\eta_i^U, \eta_i^U} = s_i^2 \Sigma_{\eta_i, \eta_i}$ such that $\eta_i^{\hat{U}} = s_i \eta_i^X$ and $\Sigma_{\eta_i^Z, \eta_i^Z} = (1 - s_i + s_i^2/2) \Sigma_{\eta_i, \eta_i}$. If $s_i = 1$, then then we have the nice property that

$$(4.16) \quad \Sigma_{\eta_i, \eta_i} = \Sigma_{\eta_i^X, \eta_i^X} = \Sigma_{\eta_i^{\hat{U}}, \eta_i^{\hat{U}}} = \Sigma_{\eta_i^U, \eta_i^U},$$

and the projection assumption (4.5a) is supported. This choice leads to $\Sigma_{\eta_i^Z, \eta_i^Z} = (1/2) \Sigma_{\eta_i, \eta_i}$ and requires updating the observation error covariance in the Kalman gain calculation (4.8) accordingly. Since the same Kalman gain is used for all variables, this change is not optimal for the analysis of principal, control, and ancillary variates. If $s_i = 2$, then $\Sigma_{\eta_i^Z, \eta_i^Z} = \Sigma_{\eta_i, \eta_i}$, and the inference performed on the total and principal

variates has the correct observation error. Moreover, the control and ancillary variates share the same observation error covariance. However, the projection assumption (4.5a) is unsupported.

A third approach would be to slightly relax the projection assumption (4.5a) by allowing perturbations to it and by additionally weakening the correlation structure of η_i^X and $\eta_i^{\hat{U}}$. Such methods are outside the scope of this paper.

In the remainder of the paper we primarily focus on method (i), where the Kalman gain is consistent for the total variate and the projection assumption is satisfied. Analysis of the ancillary variate uses a suboptimal noise level in the Kalman gain. The authors have not noticed a difference in the error between the two different methods for a two-fidelity EnKF; therefore, the choice of perturbed observation method is ultimately left to the practitioner.

Remark 11. The control and ancillary variates have to have the same mean, and the control variate needs to remain strongly correlated with the principal variate. In order to satisfy the first condition, we perform a recentering procedure around the mean of the combined analysis (4.11):

$$(4.17) \quad \tilde{\mu}_{X_i^a} \leftarrow \tilde{\mu}_{Z_i^a}, \quad \tilde{\mu}_{\hat{U}_i^a} \leftarrow \Phi_r^* \tilde{\mu}_{Z_i^a}, \quad \tilde{\mu}_{U_i^a} \leftarrow \Phi_r^* \tilde{\mu}_{Z_i^a}.$$

The approach (4.17) is not the only way to ensure that U and $\hat{U}$ have a common mean; however, it is a natural choice. An alternative approach is to not correct the mean of the ancillary ensemble at all but to recenter the control ensemble: $\tilde{\mu}_{\hat{U}_i^a} \leftarrow \tilde{\mu}_{U_i^a}$. In this second approach one runs two KF side by side, one for X and one for U .

The MFEnKF anomaly updates are as follows:

$$(4.18) \quad \begin{aligned} A_{X_i^a} &= A_{X_i^b} - \tilde{\mathbf{K}}_i \left(A_{\mathcal{H}_i(X_i^b)} - A_{\eta_i^X} \right), \\ A_{\hat{U}_i^a} &= \Phi_r^* A_{X_i^a}, \\ A_{U_i^a} &= A_{U_i^b} - \Phi_r^* \tilde{\mathbf{K}}_i \left(A_{\mathcal{H}_i(\Phi_r U_i^b)} - A_{\eta_i^U} \right). \end{aligned}$$

Note that the anomaly updates for $A_{U_i^a}$ are done solely in reduced (and observation) space. The only significant additional cost in the analysis step is the calculation of the statistical Kalman gain (4.8).

Remark 12 (computational cost). Consider the application of the Kalman gain to the observation anomalies. The dominant terms in the cost of the EnKF analysis (2.21) are

$$(4.19) \quad \mathcal{O}(m^3 + N(m^2 + mn)),$$

where a full Cholesky decomposition is assumed. Similarly, for the MFEnKF, the cost of applying the Kalman gain to the analysis anomalies in (4.18) is dominated by

$$(4.20) \quad \mathcal{O}(m^3 + (N_X + N_U)(m^2 + mn + rn + rm)).$$

In an operational setting, the observations are almost always assumed to be independent. Consequently, the dominant terms are most likely to be Nmn and $(N_X + N_U)mn$ for the methods, respectively, meaning that, ignoring the constants in front of the cost, the analysis step of the MFEnKF is as costly as that of an EnKF with $N = N_X + N_U$.

4.3. Statistical analysis of the MFEnKF. For analysis we consider the case where all uncertainties are Gaussian and the observation operators are linear, $\mathcal{H} = \mathbf{H}$, $\mathcal{H}_r = \mathbf{H}\Phi_r$. We assume that $\mu_{X^b} = X^t$ and $\mu_{U^b} = \mu_{\hat{U}^b}$. We consider the transformed total variate and principal variate parametrized by the Kalman gain,

$$(4.21) \quad Z^a(\mathbf{K}) = Z^b - \mathbf{K}(\overline{\mathcal{H}}(Z^b) - Y), \quad X^a(\mathbf{K}) = X^b - \mathbf{K}(\mathcal{H}(X^b) - Y),$$

and denote by $\mathbf{K}_Z$, $\mathbf{K}_X$ the optimal gains given by Kalman's formula that minimize the covariances of $Z^a(\mathbf{K}_Z)$ and $X^a(\mathbf{K}_X)$, respectively.

THEOREM 4.1. *Under the assumption that the statistical Kalman gain is independent of all other variates, the MFEnKF analysis is an unbiased estimator:*

$$\mu_{Z^a(\mathbf{K}_Z)} = \mu_{X^a(\mathbf{K}_Z)} = \mu_{X^a(\mathbf{K}_X)} = X^t.$$

Proof. The result follows from applying a Kalman formula and taking means. $\square$

The following theorem shows that performing the exact analysis in the total variate leads to better estimates than performing the analysis in the principal variate.

THEOREM 4.2. *The analysis total variate is $Z^a(\mathbf{K}_Z)$, and the principal component of the analysis total variate is $X^a(\mathbf{K}_Z)$. Application of the KF to the principal variate leads to the analysis $X^a(\mathbf{K}_X)$. It holds that*

$$(4.22) \quad \Sigma_{Z^a(\mathbf{K}_Z), Z^a(\mathbf{K}_Z)} \leq \Sigma_{X^a(\mathbf{K}_X), X^a(\mathbf{K}_X)} \leq \Sigma_{X^a(\mathbf{K}_Z), X^a(\mathbf{K}_Z)},$$

where inequalities are interpreted in the symmetric positive definite matrix sense.

Proof. From the optimality of the KF we have that

$$\Sigma_{Z^a(\mathbf{K}_Z), Z^a(\mathbf{K}_Z)} \leq \Sigma_{Z^a(\mathbf{K}_X), Z^a(\mathbf{K}_X)}, \quad \Sigma_{X^a(\mathbf{K}_X), X^a(\mathbf{K}_X)} \leq \Sigma_{X^a(\mathbf{K}_Z), X^a(\mathbf{K}_Z)},$$

which proves the second inequality in (4.22). From (2.9) we have that

$$(4.23) \quad \Sigma_{Z^b, Z^b} = \Sigma_{X^b, X^b} - \Sigma_{X^b, \hat{U}^b} (\Sigma_{\hat{U}^b, \hat{U}^b} + \Sigma_{U^b, U^b})^{-1} \Sigma_{\hat{U}^b, X^b} \leq \Sigma_{X^b, X^b}.$$

Next, we use the above equations and the Kalman analysis covariance formula to prove the first inequality in (4.22):

$$(4.24) \quad \begin{aligned} \Sigma_{Z^a(\mathbf{K}_Z), Z^a(\mathbf{K}_Z)} &\leq \Sigma_{Z^a(\mathbf{K}_X), Z^a(\mathbf{K}_X)} \\ &= (\mathbf{I} - \mathbf{K}_X \mathbf{H}) \Sigma_{Z^b, Z^b} (\mathbf{I} - \mathbf{K}_X \mathbf{H})^\top + \mathbf{K}_X \Sigma_{\eta, \eta} \mathbf{K}_X^\top \\ &= \Sigma_{X^a(\mathbf{K}_X), X^a(\mathbf{K}_X)} \\ &\quad - (\mathbf{I} - \mathbf{K}_X \mathbf{H}) \Sigma_{X^b, \hat{U}^b} \left(\Sigma_{\hat{U}^b, \hat{U}^b} + \Sigma_{U^b, U^b} \right)^{-1} \Sigma_{\hat{U}^b, X^b} (\mathbf{I} - \mathbf{K}_X \mathbf{H})^\top. \quad \square \end{aligned}$$

We next turn our attention to sampling errors.

THEOREM 4.3. *Assume that EnKF produces N_X i.i.d. samples of $X^a(\mathbf{K}_X)$. The covariance of the sample mean estimate about the true state is*

$$(4.25a) \quad \Sigma_{\tilde{\mu}_{X^a(\mathbf{K}_X)}, \tilde{\mu}_{X^a(\mathbf{K}_X)}} = N_X^{-1} \Sigma_{X^a(\mathbf{K}_X), X^a(\mathbf{K}_X)}.$$

Assume that MFEnKF produces N_X i.i.d. samples of $X^a(\mathbf{K}_Z)$ and $\hat{U}^a$ and N_U i.i.d. samples of U^a . Since

$$(4.26) \quad \begin{aligned} \tilde{\mu}_{Z^a(\mathbf{K}_Z)} &= (\mathbf{I} - \mathbf{K}_Z \mathbf{H}) [\tilde{\mu}_{X^b} - \mathbf{S} (\tilde{\mu}_{\hat{U}^b} - \tilde{\mu}_{U^b})] + \mathbf{K}_Z (Y + \eta) \\ &\approx \tilde{\mu}_{X^a(\mathbf{K}_Z)} - \mathbf{S} \tilde{\mu}_{\hat{U}^a} + \mathbf{S} \tilde{\mu}_{U^a} \end{aligned}$$

and we estimate the moments of X^a , $\hat{U}^a$ using N_X samples and the moments of U^a using N_U samples, the sample mean of the analysis total variate has less variance than the KF applied to the principal variate:

$$(4.27) \quad \Sigma_{\tilde{\mu}_{Z^a(\mathbf{K}_Z)}, \tilde{\mu}_{Z^a(\mathbf{K}_Z)}} \leq \Sigma_{\tilde{\mu}_{X^a(\mathbf{K}_X)}, \tilde{\mu}_{X^a(\mathbf{K}_X)}}.$$

Proof. A direct calculation shows that the variance of the empirical mean estimate about the truth is

$$(4.28) \quad \begin{aligned} & \Sigma_{\tilde{\mu}_{Z^a(\mathbf{K}_Z)}, \tilde{\mu}_{Z^a(\mathbf{K}_Z)}} \\ &= N_X^{-1} \Sigma_{Z^a(\mathbf{K}_Z), Z^a(\mathbf{K}_Z)} \\ & \quad + (N_U^{-1} - N_X^{-1}) (\mathbf{I} - \mathbf{K}_Z \mathbf{H}) \mathbf{S} \Sigma_{U^b, U^b} \mathbf{S}^\top (\mathbf{I} - \mathbf{K}_Z \mathbf{H})^\top \\ & \leq \Sigma_{\tilde{\mu}_{X^a(\mathbf{K}_X)}, \tilde{\mu}_{X^a(\mathbf{K}_X)}} + (N_U^{-1} - N_X^{-1}) (\mathbf{I} - \mathbf{K}_X \mathbf{H}) \mathbf{S} \Sigma_{U^b, U^b} \mathbf{S}^\top (\mathbf{I} - \mathbf{K}_X \mathbf{H})^\top \\ & \quad - N_X^{-1} (\mathbf{I} - \mathbf{K}_X \mathbf{H}) \mathbf{S} \left(\Sigma_{\hat{U}^b, \hat{U}^b} + \Sigma_{U^b, U^b} \right) \mathbf{S}^\top (\mathbf{I} - \mathbf{K}_X \mathbf{H})^\top, \\ &= \Sigma_{\tilde{\mu}_{X^a(\mathbf{K}_X)}, \tilde{\mu}_{X^a(\mathbf{K}_X)}} \\ & \quad - (\mathbf{I} - \mathbf{K}_X \mathbf{H}) \mathbf{S} \left(N_X^{-1} \Sigma_{\hat{U}^b, \hat{U}^b} + (2N_X^{-1} - N_U^{-1}) \Sigma_{U^b, U^b} \right) \mathbf{S}^\top (\mathbf{I} - \mathbf{K}_X \mathbf{H})^\top, \end{aligned}$$

where for the inequality we used (4.24) and (4.25a). $\square$

Theorem 4.3 shows that MFEnKF provides an estimate that is always at least as good as the corresponding EnKF estimate for the same number N_X of high-fidelity model runs. The difference comes from the smaller variance of $Z^a(\mathbf{K}_Z)$ compared to $X^a(\mathbf{K}_X)$ (first term in (4.28)), from the use of control variates in covariance estimates, and from using the data to assimilate the reduced space variables (second term in (4.28)).

Remark 13. EnKF produces an ensemble that quantifies the posterior uncertainty in the FOM state. From (4.22), the posterior ensemble of principal variables $\{\mathbf{X}^{a,[e]}(K_Z)\}_{e=1,\dots,N_X}$ constructed by MFEnKF provides (only) an upper bound for the analysis state error covariance. For posterior uncertainty quantification, one can use N_X members of the $\mathbf{U}^{a,[e]}$ ensemble to construct an ensemble of total variates.

4.4. Cost analysis of the MFEnKF. We seek to find an equivalent EnKF running an ensemble size of M_X FOM that gives the same analysis sampling error as MFEnKF with N_X full order and N_U reduced order ensemble sizes. We measure sampling errors by the trace generalized variance $\sigma_W = \text{tr}(\Sigma_{W,W})$.

By (4.25a) the sampling error for EnKF is $M_X^{-1} \sigma_X$, and by (4.28) sampling error for MFEnKF is $N_X^{-1} \sigma_Z + (N_U^{-1} - N_X^{-1}) \sigma_{\mathbf{S}U}$. By matching these generalized variances, the effective ensemble size of the EnKF is

$$(4.29) \quad M_X = \frac{N_X N_U \sigma_X}{N_U \sigma_Z - \sigma_{\mathbf{S}U} (N_U - N_X)}.$$

We see by direct calculation that $M_X \geq N_X$ whenever $N_U \geq N_X$ and $\sigma_Z \geq \sigma_{\mathbf{S}U} (N_U - N_X) / N_U$.

Let C_X be the cost of running a FOM and C_U the cost of running a lower-fidelity model. To obtain similar analyses, the cost of running the EnKF is $C_X M_X$, and the cost of running the MFEnKF is $C_X N_X + C_U (N_X + N_U)$. Consequently, the MFEnKF

algorithm is more efficient than EnKF whenever the cost of running the lower-fidelity model satisfies

$$(4.30) \quad C_U \leq \frac{C_X(M_X - N_X)}{N_X + N_U}.$$

Knowledge of the empirical variance gains with the MFEnKF procedure and the respective costs of the two forward models can help decide whether the MFEnKF is a cost-effective replacement for the EnKF.

4.5. Telescopic extension. We now discuss the telescopic extension from the two-fidelity to the multifidelity ensemble KF, by utilizing the multivariate control variate extensions discussed in subsection 2.2. Consider a sequence of projection operators $\Phi_{r_\ell}^\ell$ and $\Phi_{r_\ell}^{\ell,*}$ for $\ell = 1, \dots, \mathcal{L}$, and denote $\bar{\Phi}_{r_\ell}^\ell = \prod_{\lambda=1}^\ell \Phi_{r_\lambda}^\lambda$. The control variate relation between fidelity $\ell - 1$ and ℓ is $\hat{U}_{\ell,i} = \Phi_{r_\ell}^{\ell,*} U_{\ell-1,i}$, with $U_{0,i} \equiv X_i$. The corresponding gain from fidelity ℓ to the fidelity of the principal variate is $\bar{S}_\ell = 2^{-\ell} \bar{\Phi}_{r_\ell}^\ell$. The random variables representing the control variate and the ancillary variate in the two-fidelity scheme now represent the corresponding first fidelity variates. Extending the total variate to $\mathcal{L}$ fidelities gives

$$(4.31) \quad Z_i^b = X_i^b - \sum_{\ell=1}^{\mathcal{L}} 2^{-\ell} \bar{\Phi}_{r_\ell}^\ell (\hat{U}_{\ell,i}^b - U_{\ell,i}^b).$$

The empirical Kalman gain is computed through a natural extension to the two-fidelity approach. The MFEnKF anomaly updates are defined as

$$(4.32) \quad A_{U_{\ell,i}^a} = A_{U_{\ell,i}^b} - \bar{\Phi}_{r_\ell}^{\ell,*} \tilde{K}_i \left(A_{\mathcal{H}_i(\bar{\Phi}_{r_\ell}^\ell U_{\ell,i}^b)} - A_{\eta_i} \right).$$

The additive perturbed observation errors are chosen in a fashion similar to the methods described above. Note that for a large number of fidelities, from a practical perspective, it might be beneficial to choose method (ii) with $s_i = 1$, thereby making the synthetic observation error equal for all variates, at the cost of the total variate observation error being reduced to $\Sigma_{\eta_i^Z, \eta_i^Z} = ((1 + 2^{1-2\mathcal{L}})/3) \Sigma_{\eta_i, \eta_i}$.

5. The test model hierarchy. One salient feature of our MFEnKF framework is that it can employ different spaces to represent the models at different resolutions. In our numerical tests we employ the following models of the quasi-geostrophic equations (QGE). The highest-resolution model, called the *truth*, represents the reference solution and provides $\mathbf{X}_i^t$ and the observation data via (2.2). In subsection 5.2 the truth corresponds to a direct numerical simulation (DNS) on a fine mesh. The FOM is an accurate approximation of the truth and is obtained in subsection 5.3 by performing DNS on a coarser spatial mesh. The ROM is a low-cost approximation of the FOM and is obtained in subsection 5.4 by performing a proper orthogonal decomposition in the FOM space, the number of modes that represent the dynamics.

Figure 2 presents a comparison between the truth, the FOM, and the ROM solutions for a 10-day forecast with QGE. All discrete models are implemented in our test suite [10, 59]. Over a 1-day time interval, the DNS computational cost (20.7 seconds for one model run) is the highest, the high-res FOM computational cost (0.39 seconds) is 53 times lower than the DNS cost, and the low-res ROM ($r = 50$) computational cost (0.00094 seconds) is 418 times lower than the FOM cost. As our implementations are naïve MATLAB-based models, these values only represent rough order-of-magnitude estimates.

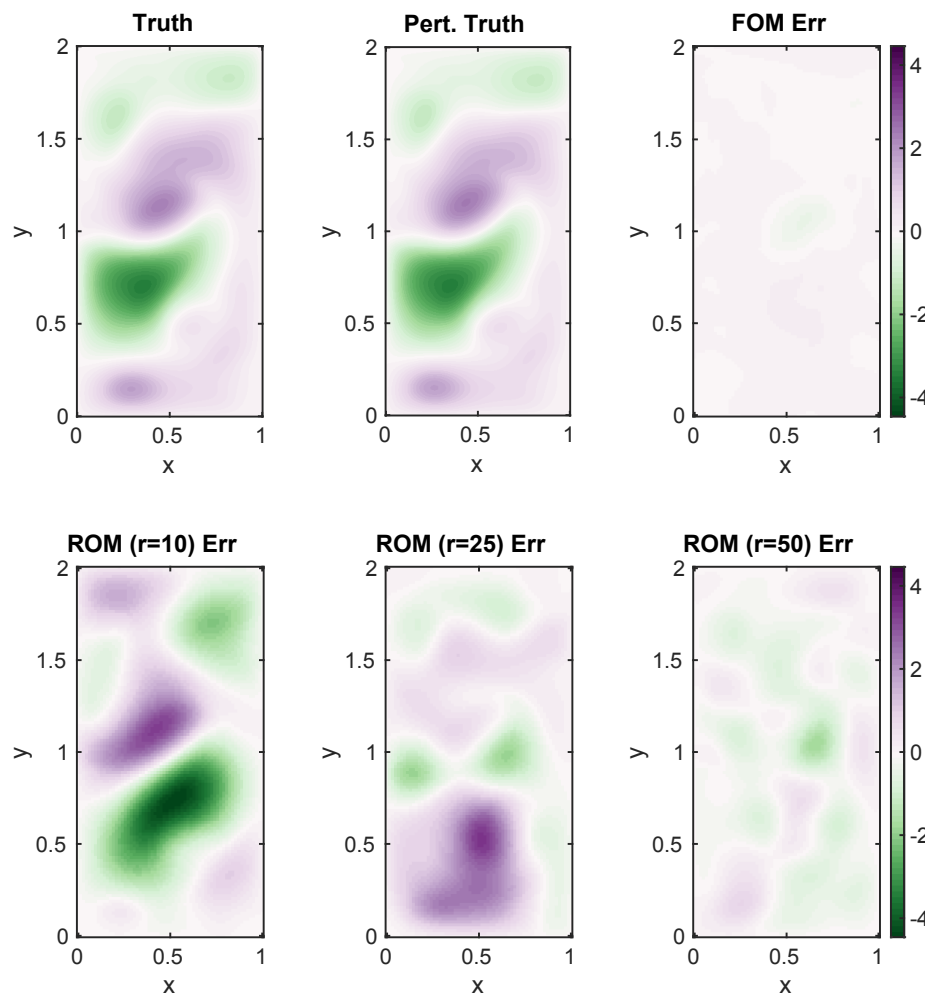


FIG. 2. A 10-day forecast of the QGE. We show the true DNS, a perturbation of the true DNS, and err in propagating the perturbed truth through both the high-res FOM and the low-res ROM ($r = 10, 25, 50$). The plotted values represent the streamfunction (and errors therein), with purple representing positive values and green representing negative values.

5.1. The QGE. The QGE [16, 17, 22, 45] are based on the barotropic vorticity equations and are widely used in both data assimilation and reduced order modeling, thereby providing an excellent test problem for MFEnKF. Here we follow the formulation given in [46, 61],

$$(5.1) \quad \begin{aligned} \omega_t + J(\psi, \omega) - \text{Ro}^{-1} \psi_x &= \text{Re}^{-1} \Delta \omega + \text{Ro}^{-1} F, \\ J(\psi, \omega) &\equiv \psi_y \omega_x - \psi_x \omega_y, \quad \omega = -\Delta \psi, \end{aligned}$$

where ω is the vorticity, ψ is the streamfunction, Re is the Reynolds number, Ro is the Rossby number, and F is a forcing term. We use a symmetric double gyre for the forcing term [22, 46, 61] $F = \sin(\pi(y-1))$ and homogeneous Dirichlet boundary conditions $\omega(x, y) = 0$, $\psi(x, y) = 0$, $(x, y) \in \partial\Omega$, where the computational domain is $\Omega = [0, 1] \times [0, 2]$. The constants are $\text{Re} = 450$ and $\text{Ro} = 0.0036$. The timescale of the problem uses 80 time units to represent 20.12 years [61].

TABLE 1
Relative kinetic energy for the first r ROM modes.

r	10	25	50	100
Relative KE	0.9071	0.9679	0.9871	0.9963

5.2. The DNS. The truth involves a DNS simulation of the QGE (5.1) on a “fine” spatial mesh with 255 interior points in the x direction and 511 points in the y direction. Second-order finite difference discretization are used for both first- and second-order spatial derivatives, together with the Arakawa approximation [4, 35] for the Jacobian term J in (5.1). The embedded Poisson equation is solved using a precomputed sparse Cholesky decomposition.

Time integration for this and subsequent discretizations is performed using a fourth-order “almost Runge–Kutta” method with adaptive time stepping [57]. We take observations every 24 hours (approximately 0.0109 model time units) of 150 equally spaced variables.

5.3. The FOM. The FOM performs a numerical simulation of the QGE (5.1) on a “coarse” spatial mesh with 63 interior points in the x direction and 127 points in the y direction. The same spatial and temporal discretizations as for the truth simulation are used. As illustrated in Figure 2, although the FOM approximation is qualitatively similar to the DNS approximation, the former does not capture all the physical details displayed by the latter. The changes of grid (from the DNS to FOM state-space) are performed through canonical multigrid techniques [73].

5.4. The ROM. The construction of ROM for the QGE (5.1) follows [46, 61, 65]. We start by building the ROM vorticity basis using the proper orthogonal decomposition [30] and the method of snapshots [63]. We collect $\omega_1, \dots, \omega_M$ snapshots of FOM vorticity at $M = 700$ different times along a model trajectory, with each snapshot roughly 6 months apart in model time. The snapshot trajectory is unrelated to the trajectory of the truth in order to simulate more realistic operational conditions. We build the snapshot covariance matrix $[C]_{ij} = \langle \omega_i, \omega_j \rangle$, $i, j = 1, \dots, M$, using a quadrature approximate integration. The eigendecomposition of C yields the ROM vorticity basis $\{\varphi_1, \dots, \varphi_r\}$, where r is the ROM dimension. The relative kinetic energy of the first r modes [46] is listed in Table 1, where the relative kinetic energy is calculated based on FOM data over the time $[10, 80]$ (units). In numerical simulations we consider $r = 10, 25$, and 50 .

The ROM streamfunction basis is obtained from the ROM vorticity basis by the relationship $-\Delta\phi_i = \varphi_i$, $i = 1, \dots, r$. The ROM vorticity and streamfunction approximations are

$$(5.2) \quad \tilde{\omega}(t) = \sum_{i=1}^r a_i(t) \varphi_i \quad \text{and} \quad \tilde{\psi}(t) = \sum_{i=1}^r a_i(t) \phi_i,$$

respectively. The dynamics of the unknown ROM coefficients $\mathbf{a}(t) = [a_1(t), \dots, a_r(t)]^\top$ is determined by using a Galerkin projection of the equations (5.2),

$$(5.3) \quad \mathbf{a}_t = \mathbf{b} + \mathbf{A}\mathbf{a} + \mathbf{a}^\top \mathbf{B}\mathbf{a},$$

where

$$(5.4) \quad \begin{aligned} [\mathbf{b}]_i &= \text{Ro}^{-1} \langle F, \varphi_i \rangle, & [\mathbf{A}]_{i,j} &= \text{Ro}^{-1} \left\langle \frac{\partial \phi_j}{\partial x}, \varphi_i \right\rangle - \text{Re}^{-1} \langle \nabla \varphi_j, \nabla \varphi_i \rangle, \\ [\mathcal{B}]_{i,m,n} &= -\langle J(\varphi_m, \phi_n), \varphi_i \rangle, \end{aligned}$$

with the inner products implemented using the two-dimensional Simpson's rule discretization. As illustrated in Figure 2, although the ROM approximation is qualitatively similar to the DNS and FOM, it does not capture all the physical details.

5.5. Projection operators. We now explicitly define the space projection operators from section 3 for the QGE and its corresponding ROM. Let the matrix $\mathbf{D}$ represent the two-dimensional Simpson's rule discretization of the spatial inner product and Δ the discrete version of the Laplacian. Then $\mathbf{M}_v = \Delta \mathbf{D} \Delta$ in (3.2). Let Ψ_r be the r dominant eigenvectors of the temporal covariance matrix $\mathcal{C}$. The following operators preserve the relationship between the vorticity and streamfunction bases:

$$(5.5) \quad \Phi_r = -\mathbf{D}^{-1/2} \Delta^{-1} \Psi_r, \quad \Phi_r^* = -\Psi_r^T \mathbf{D}^{1/2} \Delta.$$

6. Numerical experiments. The numerical experiments aim to investigate the performance of MFEnKF compared against other EnKF methods to assess how the analysis accuracy depends on the accuracy of the underlying ROM and the usefulness of the ensembles underlying MFEnKF to represent probability distributions of interest. We consider two fidelities; see section 5. The principal variate represents the uncertainty in QGE FOM, and our control and ancillary variates represent the uncertainty in QGE ROM. The truth is provided by the QGE DNS model.

In order to create synthetic observations, the truth solution is relaxed onto the FOM space, and the states corresponding to 150 equally spaced indices are observed. The observation error covariance is $\Sigma_{\eta,\eta} = \mathbf{I}_{150}$. In terms of EnKF correction techniques, the experiments use inflation, as there is strong evidence [54] that it is an explicit probabilistic requirement in EnKF-based methods. The same inflation factor is used for the principal and control variate ensembles, and an independently chosen value is used for the ancillary ensemble. All experiments run for 350 observation steps. The first 50 are discarded to account for model spin-up, and the rest are used to compute the analysis quality results.

6.1. Comparison with other techniques. We assess the performance of MFEnKF compared to three other perturbed observation filters: a heavily corrected and localized version of the original MLEnKF [9, 29], a localized EnKF, and a shrinkage covariance-corrected EnKF.

MFEnKF uses a ROM of dimension $r = 50$, ancillary ensemble size $N_U = 40$, and ancillary inflation factor $\alpha_U = 1.1$. Since the standard MLEnKF [9, 29] did not converge for this test problem, we consider a modified version and correct it by augmenting the MLEnKF formulas with the forecast correction (4.5a), the mean correction (4.17), and Gaussian kernel localization with a radius of 20 grid units. The ROM and ROM space ensembles have an identical configuration to that used with MFEnKF. All other implementation details follow [29]. We note that the standard MLEnKF does not converge for any ensemble size that we used without applying all of these heuristic corrections, especially localization. The localized EnKF uses Gaussian localization with a radius of 20 grid units. We again note that for the chosen ensemble sizes of fewer than 30 members, the EnKF does not converge at all without localization. The covariance shrinkage EnKF [47] uses the target matrix to

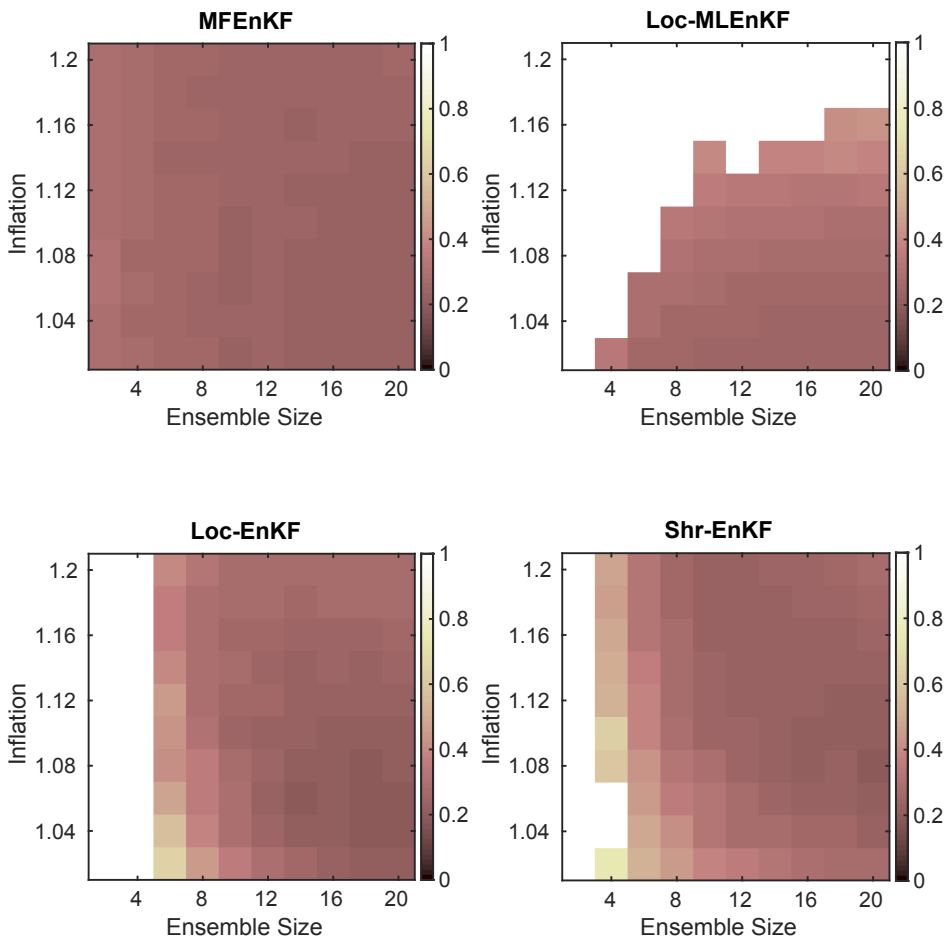


FIG. 3. Analysis RMSE comparison of two-fidelity MFEnKF versus a corrected and localized MLEnKF [9, 29] (Loc-MLEnKF), a localized EnKF (Loc-EnKF), and a covariance shrinkage-based EnKF (Shr-EnKF). Darker shades represent lower error, with lighter shades representing higher error.

be a snapshot-derived localized background covariance matrix and the (normalized) Rao–Blackwellized Ledoit and Wolf estimator.

For all methods we employ different FOM ensemble sizes $N_X = 2, 4, \dots, 20$ and inflation factors $\alpha_X = 1.02, 1.04, \dots, 1.2$ and calculate the spatiotemporal root mean square error (RMSE) (averaged over three model runs) of the analysis with respect to the truth. Results reported in Figure 3 show that MFEnKF outperforms the heavily corrected localized MLEnKF, meaning that our derivation of the MFEnKF from a robust control variate framework indeed has merit. We additionally outperform standard correction techniques such as localization and covariance shrinkage in standard EnKF. The combination of a FOM and a ROM in ensemble-based methods could be used as a replacement to (or in conjunction with) such methods.

It is interesting to note that our method is stable for the QG model even with $N_X = 2$, while an EnKF run that uses only $N = N_U = 40$ ROM runs would not be. Even small corrections to the ROM by using a few FOM can potentially stabilize the EnKF.

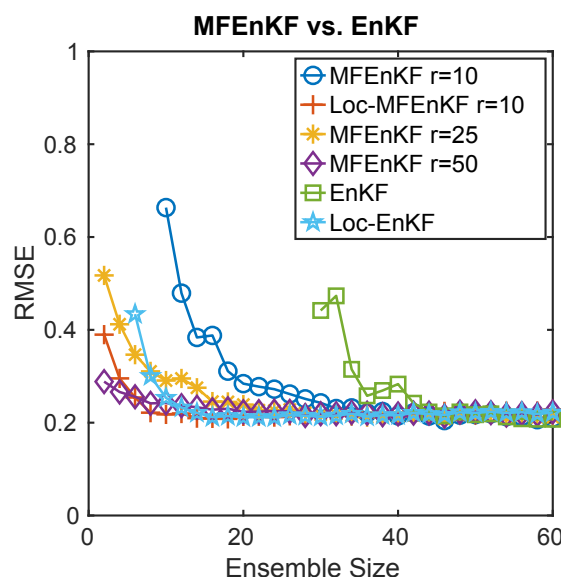


FIG. 4. Comparison RMSE of the two-fidelity MFEnKF for various values of the ROM dimension r and both a localized and a standard EnKF.

We perform a simple computational cost analysis. For $r = 50$ the normalized cost of one ROM run is 1 unit, and the cost of one FOM run is approximately 418 units (empirically measured wall-clock time). The cost of one MFEnKF forecast is $418N_X + (N_X + N_U)$ normalized wall-clock units by (2.17). If $N_X + N_U \approx 418$, then the cost roughly equals that of one extra FOM ensemble member. For any reasonable N_U , we obtain a stable algorithm for the cost of less than one extra FOM run. For $N_X = 4$ we maintain the accuracy of a (nonlocalized, not pictured) EnKF with an ensemble size $N_X = 40$ and that of a localized EnKF with an ensemble size of $N_X = 12$. This results in twofold to 10-fold cost savings in terms of forward model runs. Additionally, a lot of additional computation is saved by not having to apply localization.

6.2. Impact of ROM dimension. The second numerical experiment assesses the impact of ROM basis size. We consider $r = 10$, $r = 25$, and $r = 50$, representing a severely underrepresented system, an underrepresented system, and a system with a medium level of representation, respectively. For the severely underrepresented system we use a localized (Gaussian with radius of 20 grid units) implementation. The ROM ensemble sizes are $N_U = 9$, $N_U = 20$, and $N_U = 40$, respectively, in order to always have undersampled ensembles. For comparison we consider both a localized and a standard EnKF. The inflation factors $\alpha_X = 1.1$ and $\alpha_U = 1.1$ are used in all experiments. Spatiotemporal analysis RMSE (averaged over three runs) for different FOM ensemble sizes N_X are shown in Figure 4. Larger ROM bases lead to more accurate analyses. Even with the particularly small basis size $r = 10$ MFEnKF is significantly superior to a standard EnKF; this substantiates Remark 1 that the magnitude of the analysis covariance can only be improved when an optimal gain is used, even if the quality of the ROM is poor. A basis of size $r = 25$ leads to results very similar to that of the localized EnKF, and even using only $r = 10$ basis vectors with a localized variant of the MFEnKF algorithm is almost as good as using $r = 50$ basis vectors.

We note that if we take the cost of running models forward in time as our independent variable, compared to RMSE, then we would get a figure identical to Figure 4. This is because even in the case of $r = 50$, the ROM ensemble size of $N_U = 40$ only incurs negligible runtime which does not add up to even one extra FOM run.

6.3. Rank histograms. A rank histogram measures the reliability with which an ensemble forecast captures the probability distributions of certain quantities of interest [25]. Consider an ensemble of scalar quantities representing independent draws from the exact distribution (here, normal); tallying the number of ensemble members that underestimate each of them should result in a uniformly distributed histogram.

We consider the rank histograms of the ensembles representing the principal variate, control variate, and the ancillary variate and measure the KL divergence [40] between these histograms (Q) and an ideal uniform distribution (P),

$$(6.1) \quad D_{KL}(P||Q) = - \sum_i P_i \log \left(\frac{P_i}{Q_i} \right),$$

where the result represents the information (in nats) required to transform one distribution to the other. A value close to zero nats implies that the distributions are essentially indistinguishable.

We construct the rank histograms using the truth values of all 150 observed variables and assuming their independence from each other. Multiple data assimilation experiments are carried out using inflation factors from 1.02 to 1.2 for each of the ensembles considered herein. We compare two-fidelity MFEnKF with $r = 50$, $N_X = 20$, and $N_U = 40$ to a vanilla perturbed observation EnKF with $N_X = 60$. For each experiment, ensemble, and algorithm, we collect the KL divergence value (6.1) and the analysis RMSE.

Figure 5 shows the KL divergence values versus RMSE, where each point corresponds to a different experiment. It can be seen that the EnKF preserves predictability (low KL divergence value) for almost all values of inflation and that inflation mainly affects the RMSE. In contrast, for the MFEnKF, inflation does not have such a dramatic impact, especially for the principal and control variates. For the ancillary variate, inflation plays a key role in lowering the KL divergence of the rank histogram from the normal and has much less impact on RMSE. This means that in terms of predictability, the ensemble of the principal variates is more reliable than that produced by EnKF.

7. Conclusions and future work. This work develops the new multifidelity ensemble KF algorithm based on a linear control variate framework. The multivariate linear control variate theory perspective allows for rigorous multifidelity extensions of the EnKF, where the uncertainty in coarser levels in the hierarchy of models represents control variates for the uncertainty in finer levels. Thus, complementing a small ensemble of high-fidelity model runs with larger ensembles of cheaper, lower-fidelity runs results in improved analyses with only small additional computational costs. Different models in the hierarchy can have different state spaces, with different dimensions and/or different inner products. The mapping between different spaces (i.e., the mapping of each control variate to the space of the corresponding principal variates) is done by gain matrices that can be computed in an optimal way. The analysis of the new algorithm shows that it always produces better analyses than EnKF with the same number of high-fidelity ensemble members.

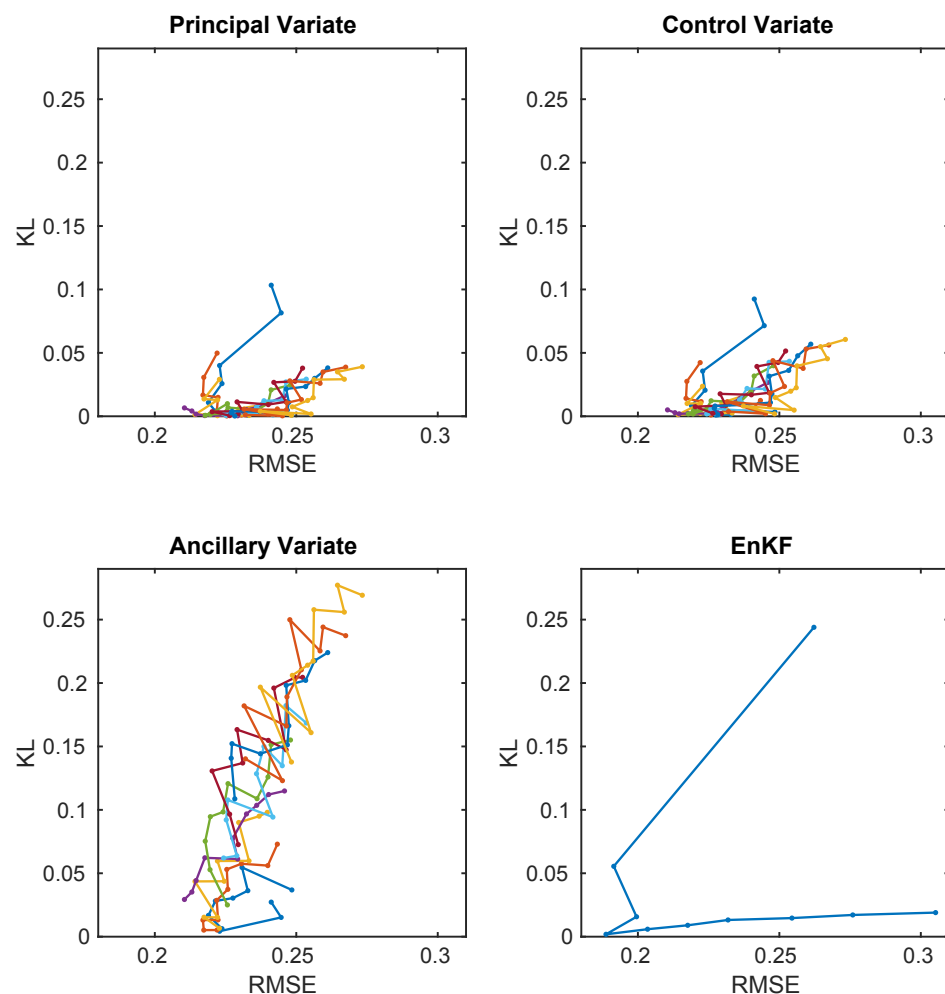


FIG. 5. KL divergence (in nats) of rank histogram from uniform distribution compared with spatiotemporal RMSE. For MFEnKF results each line represents a constant value of inflation of the principal variate ensemble and each point a different value of ancillary variate ensemble inflation. For EnKF result each point represents a different value of inflation.

MFEnKF has several advantages over other approaches to couple information from different models in data assimilation. Using multimodel ensembles to build empirical covariances in EnKF faces the challenge that different ensemble members have different dimensions. The strategy of stacking models to formally construct a supermodel and applying EnKF in the aggregated space cannot employ different numbers of ensemble members of different models. MLEnKF [9, 29] applies EnKF in the high-fidelity space and uses different model levels to improve the empirical covariance estimates. Incorporating different model levels using signed empirical measures leads to possibly non-positive-definite multilevel covariance estimates and requires all models to share the same state space.

Numerical experiments with a quasi-geostrophic model reveal that MFEnKF provides significant improved analysis over the standard MLEnKF and is competitive with other EnKF correction methods such as localization and covariance shrinkage.

Moreover, the ensembles underlying the MFEnKF technique are useful in representing the probability distributions of given quantities of interest.

The algorithm discussed herein is a multifidelity variant of the perturbed observation EnKF. An interesting future research direction is to develop multifidelity square root filters, e.g., multifidelity LETKF [31]. Another interesting direction is extending the control variate approach to the case where different models are not hierarchically organized. Experiments with an operational model could reveal that more complex models could be considered with the same amount of ensemble members without compromising accuracy. Proving more rigorous error bounds [13, 50, 72] for the new MFEnKF framework could provide further insight into parameter model choices. Exploring the error induced in the algorithm by the discarded error term in (3.5) could also prove a fruitful endeavor.

Acknowledgments. The first author would like to thank Steven Roberts for his help and knowledge with regard to time integration methods. The authors would like to thank the two anonymous referees for their insights, making the manuscript significantly better.

REFERENCES

- [1] J. ANDERSON, *An ensemble adjustment Kalman filter for data assimilation*, Monthly Weather Rev., 129 (2001), pp. 2884–2903.
- [2] J. ANDERSON, *Localization and sampling error correction in EnKF data assimilation*, Monthly Weather Rev., 140 (2012), pp. 2359–2371.
- [3] J. L. ANDERSON, *An adaptive covariance inflation error correction algorithm for ensemble filters*, Tellus A Dyn. Meteorol. Oceanogr., 59 (2007), pp. 210–224.
- [4] A. ARAKAWA, *Computational design for long-term numerical integration of the equations of fluid motion: Two-dimensional incompressible flow. Part I*, J. Comput. Phys., 1 (1966), pp. 119–143.
- [5] M. ASCH, M. BOCQUET, AND M. NODET, *Data Assimilation: Methods, Algorithms, and Applications*, SIAM, Philadelphia, 2016.
- [6] A. ATTIA, R. STEFANESCU, AND A. SANDU, *The reduced-order hybrid Monte-Carlo sampling smoother*, Int. J. Numer. Methods Fluids, 83 (2016), pp. 28–51, <https://doi.org/10.1002/fld.4255>.
- [7] S. L. BRUNTON AND J. N. KUTZ, *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*, Cambridge University Press, Cambridge, 2019.
- [8] Y. CAO, J. ZHU, I. M. NAVON, AND Z. LUO, *A reduced-order approach to four-dimensional variational data assimilation using proper orthogonal decomposition*, Int. J. Numer. Methods Fluids, 53 (2007), pp. 1571–1583.
- [9] A. CHERNOV, H. HOEL, K. LAW, F. NOBILE, AND R. TEMPONE, *Multilevel ensemble Kalman filtering for spatially extended models*, MATHICSE technical report 22.2017, EPFL, 2017, https://www.epfl.ch/labs/mathicse/wp-content/uploads/2018/10/Report-22.2017-AC_HAH_KL_FN_RT.pdf.
- [10] COMPUTATIONAL SCIENCE LABORATORY, *ODE Test Problems*, 2020, <https://github.com/ComputationalScienceLaboratory/ODE-Test-Problems>.
- [11] T. CUI, Y. M. MARZOUK, AND K. E. WILLCOX, *Data-driven model reduction for the Bayesian solution of inverse problems*, Internat. J. Numer. Methods Engrg., 102 (2015), pp. 966–990.
- [12] D. N. DAESCU AND I. M. NAVON, *Efficiency of a POD-based reduced second-order adjoint model in 4D-Var data assimilation*, Int. J. Numer. Methods Fluids, 53 (2007), pp. 985–1004.
- [13] M. DIHLMANN AND B. HAASDONK, *A reduced basis Kalman filter for parametrized partial differential equations*, ESAIM Control Optim. Calc. Var., 22 (2016), pp. 625–669.
- [14] G. EVENSEN, *Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics*, J. Geophys. Res. Oceans, 99 (1994), pp. 10143–10162.
- [15] G. EVENSEN, *Data Assimilation: The Ensemble Kalman Filter*, Springer, New York, 2009.
- [16] J. FERGUSON, *A Numerical Solution for the Barotropic Vorticity Equation Forced by an Equatorially Trapped Wave*, Master's thesis, University of Victoria, 2008.

- [17] E. L. FOSTER, T. ILIESCU, AND Z. WANG, *A finite element discretization of the streamfunction formulation of the stationary quasi-geostrophic equations of the ocean*, Comput. Methods Appl. Mech. Engrg., 261 (2013), pp. 105–117.
- [18] G. BURGERS, P. J. VAN LEEUWEN AND G. EVENSEN, *Analysis scheme in the ensemble Kalman filter*, Monthly Weather Rev., 126 (1998), pp. 1719–1724.
- [19] D. GALBALLY, K. FIDKOWSKI, K. WILLCOX, AND O. GHATTAS, *Non-linear model reduction for uncertainty quantification in large-scale inverse problems*, Internat. J. Numer. Methods Engrg., 81 (2010), pp. 1581–1608.
- [20] M. B. GILES, *Multilevel Monte Carlo path simulation*, Oper. Res., 56 (2008), pp. 607–617.
- [21] M. B. GILES, *Multilevel Monte Carlo methods*, Acta Numer., 24 (2015), pp. 259–328.
- [22] R. J. GREATBATCH AND B. T. NADIGA, *Four-gyre circulation in a barotropic model with double-gyre wind forcing*, J. Phys. Oceanogr., 30 (2000), pp. 1461–1471.
- [23] A. GREGORY AND C. COTTER, *A seamless multilevel ensemble transform particle filter*, SIAM J. Sci. Comput., 39 (2017), pp. A2684–A2701, <https://doi.org/10.1137/16M1102021>.
- [24] A. GREGORY, C. COTTER, AND S. REICH, *Multilevel ensemble transform particle filtering*, SIAM J. Sci. Comput., 38 (2016), pp. A1317–A1338, <https://doi.org/10.1137/15M1038232>.
- [25] T. M. HAMILL, *Interpretation of rank histograms for verifying ensemble forecasts*, Monthly Weather Rev., 129 (2001), pp. 550–560.
- [26] J. HE, P. SARMA, AND L. J. DURLOFSKY, *Use of reduced-order models for improved data assimilation within an EnKF context*, in SPE Reservoir Simulation Symposium, Society of Petroleum Engineers, 2011.
- [27] J. S. HESTHAVEN, G. ROZZA, AND B. STAMM, *Certified Reduced Basis Methods for Parametrized Partial Differential Equations*, Springer, New York, 2015.
- [28] C. HIMPE AND M. OHLBERGER, *Data-driven combined state and parameter reduction for inverse problems*, Adv. Comput. Math., 41 (2015), pp. 1343–1364.
- [29] H. HOEL, K. J. H. LAW, AND R. TEMPONE, *Multilevel ensemble Kalman filtering*, SIAM J. Numer. Anal., 54 (2016), <https://doi.org/10.1137/15M100955X>.
- [30] P. HOLMES, J. L. LUMLEY, AND G. BERKOOZ, *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*, Cambridge University Press, Cambridge, 1996.
- [31] B. R. HUNT, E. J. KOSTELICH, AND I. SZUNYOGH, *Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter*, Phys. D, 230 (2007), pp. 112–126.
- [32] K. IDE, P. COURTIER, M. GHIL, AND A. C. LORENC, *Unified notation for data assimilation: Operational, sequential and variational*, J. Meteorol. Soc. Jpn., Ser. II, 75 (1997), pp. 181–189.
- [33] T. ILIESCU AND Z. WANG, *Variational multiscale proper orthogonal decomposition: Navier-Stokes equations*, Numer. Methods Partial Differential Equations, 30 (2014), pp. 641–663.
- [34] E. T. JAYNES, *Probability Theory: The Logic of Science*, Cambridge University Press, Cambridge, 2003.
- [35] D. C. JESPERSEN, *Arakawa’s method is a finite-element method*, J. Comput. Phys., 16 (1974), pp. 383–390.
- [36] M. KAERCHER, S. BOYAVAL, M. A. GREPL, AND K. VEROY, *Reduced basis approximation and a posteriori error bounds for 4D-Var data assimilation*, Optim. Eng., (2018), pp. 1–33.
- [37] R. KALMAN, *A new approach to linear filtering and prediction problems*, Trans. ASME J. Basic Eng., 82 (1960), pp. 35–45.
- [38] E. KALNAY, *Atmospheric Modeling, Data Assimilation, and Predictability*, Cambridge University Press, Cambridge, 2003.
- [39] R. KIKUCHI, T. MISAKA, AND S. OBAYASHI, *Assessment of probability density function based on POD reduced-order model for ensemble-based data assimilation*, Fluid Dyn. Res., 47 (2015), 051403, <https://doi.org/10.1088/0169-5983/47/5/051403>.
- [40] S. KULLBACK AND R. A. LEIBLER, *On information and sufficiency*, Ann. Math. Stat., 22 (1951), pp. 79–86.
- [41] K. KUNISCH AND S. VOLKWEIN, *Galerkin proper orthogonal decomposition methods for parabolic problems*, Numer. Math., 90 (2001), pp. 117–148.
- [42] K. LAW, A. STUART, AND K. ZYGALAKIS, *Data Assimilation: A Mathematical Introduction*, Vol. 62, Springer, New York, 2015.
- [43] B. LIN AND D. McLAUGHLIN, *Efficient characterization of uncertain model parameters with a reduced-order ensemble Kalman filter*, SIAM J. Sci. Comput., 36 (2014), pp. B198–B224.
- [44] Y. MADAY, A. T. PATERA, J. D. PENN, AND M. YANO, *A parameterized-background data-weak approach to variational data assimilation: Formulation, analysis, and application to acoustics*, Internat. J. Numer. Methods Engrg., 102 (2015), pp. 933–965.
- [45] A. J. MAJDA AND X. WANG, *Nonlinear Dynamics and Statistical Theories for Basic Geophysical Flows*, Cambridge University Press, Cambridge, 2006.

- [46] C. MOU, H. LIU, D. R. WELLS, AND T. ILIESCU, *Data-driven correction reduced order models for the quasi-geostrophic equations: A numerical investigation*, Internat. J. Comput. Fluid Dyn., (2020), pp. 1–13.
- [47] E. NINO-RUIZ AND A. SANDU, *Ensemble Kalman filter implementations based on shrinkage covariance matrix estimation*, Ocean Dynam., 65 (2015), pp. 1423–1439, <https://doi.org/10.1007/s10236-015-0888-9>.
- [48] E. NINO-RUIZ AND A. SANDU, *An ensemble Kalman filter implementation based on modified Cholesky decomposition for inverse covariance matrix estimation*, SIAM J. Sci. Comput., 40 (2018), pp. A867–A886, <https://doi.org/10.1137/16M1097031>.
- [49] E. NINO-RUIZ AND A. SANDU, *Efficient parallel implementation of DDDAS inference using an ensemble Kalman filter with shrinkage covariance matrix estimation*, Cluster Comput., 22 (2019), pp. 2211–2221, <https://doi.org/10.1007/s10586-017-1407-1>.
- [50] S. PAGANI, A. MANZONI, AND A. QUARTERONI, *Efficient state/parameter estimation in nonlinear unsteady PDEs by a reduced basis ensemble Kalman filter*, SIAM/ASA J. Uncertain. Quantif., 5 (2017), pp. 890–921.
- [51] B. PEHERSTORFER, K. WILLCOX, AND M. GUNZBURGER, *Survey of multifidelity methods in uncertainty propagation, inference, and optimization*, SIAM Rev., 60 (2018), pp. 550–591.
- [52] R. PETRIE, *Localization in the ensemble Kalman filter*, MSc Atmos. Ocean Climate Univ. Reading, 2008.
- [53] A. A. POPOV AND A. SANDU, *A Bayesian approach to multivariate adaptive localization in ensemble-based data assimilation with time-dependent extensions*, Nonlinear Proce. Geophys., 26 (2019), pp. 109–122.
- [54] A. A. POPOV AND A. SANDU, *An Explicit Probabilistic Derivation of Inflation in a Scalar Ensemble Kalman filter for Finite Step, Finite Ensemble Convergence*, preprint, arXiv:2003.13162, 2020.
- [55] A. A. POPOV, A. SANDU, E. D. NINO-RUIZ, AND G. EVENSEN, *A Stochastic Covariance Shrinkage Anseml Transform Kalman Filtering*, preprint, arXiv:2003.00354, 2020.
- [56] A. QUARTERONI, A. MANZONI, AND F. NEGRI, *Reduced Basis Methods for Partial Differential Equations: An Introduction*, Vol. 92, Springer, New York, 2015.
- [57] N. RATTENBURY, *Almost Runge-Kutta Methods for Stiff and Non-Stiff Problems*, Ph.D. thesis, University of Auckland, 2005.
- [58] S. REICH AND C. COTTER, *Probabilistic Forecasting and Bayesian Data Assimilation*, Cambridge University Press, Cambridge, 2015.
- [59] S. ROBERTS, A. A. POPOV, AND A. SANDU, *ODE Test Problems: A MATLAB Suite of Initial Value Problems*, preprint, arXiv:1901.04098, 2019.
- [60] R. Y. RUBINSTEIN AND R. MARCUS, *Efficiency of multivariate control variates in Monte Carlo simulation*, Oper. Res., 33 (1985), pp. 661–677.
- [61] O. SAN AND T. ILIESCU, *A stabilized proper orthogonal decomposition reduced-order model for large scale quasigeostrophic ocean circulation*, Adv. Comput. Math., (2015), pp. 1289–1319.
- [62] J. R. SINGLER, *New POD error expressions, error bounds, and asymptotic results for reduced order models of parabolic PDEs*, SIAM J. Numer. Anal., 52 (2014), pp. 852–876.
- [63] L. SIROVICH, *Turbulence and the dynamics of coherent structures. I. Coherent structures*, Quart. Appl. Math., 45 (1987), pp. 561–571.
- [64] R. STEFANESCU, A. SANDU, AND I. NAVON, *POD/DEIM strategies for reduced data assimilation systems*, J. Comput. Phys., 295 (2015), pp. 569–595, <https://doi.org/10.1016/j.jcp.2015.04.030>.
- [65] M. STRAZZULLO, F. BALLARIN, R. MOSETTI, AND G. ROZZA, *Model reduction for parametrized optimal control problems in environmental marine sciences and engineering*, SIAM J. Sci. Comput., 40 (2018), pp. B1055–B1079.
- [66] S. H. STROGATZ, *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering*, CRC Press, Boca Raton, FL, 2018.
- [67] X. TIAN, Z. XIE, AND Q. SUN, *A POD-based ensemble four-dimensional variational assimilation method*, Tellus A Dyn. Meteorol. Oceanogr., 63 (2011), pp. 805–816, <https://doi.org/10.1111/j.1600-0870.2011.00529.x>.
- [68] X. T. TONG, A. J. MAJDA, AND D. KELLY, *Nonlinear Stability of the Ensemble Kalman Filter with Adaptive Covariance Inflation*, preprint, arXiv:1507.08319, 2015.
- [69] P. T. M. VERMEULEN AND A. W. HEEMINK, *Model-reduced variational data assimilation*, Monthly Weather Rev., 134 (2006), pp. 2888–2899, <https://doi.org/10.1175/MWR3209.1>.
- [70] D. XIAO, J. DU, F. FANG, C. C. PAIN, AND J. LI, *Parameterised non-intrusive reduced order methods for ensemble Kalman filter data assimilation*, Comput. Fluids, 177 (2018), pp. 69–77.

- [71] M. YAREMCHUK, D. NECHAEV, AND G. PANTELEEV, *A method of successive corrections of the control subspace in the reduced-order variational data assimilation*, Monthly Weather Rev., 137 (2009), pp. 2966–2978, <https://doi.org/10.1175/2009MWR2592.1>.
- [72] C. ZERFAS, L. G. REBHOLZ, M. SCHNEIER, AND T. ILIESCU, *Continuous data assimilation reduced order models of fluid flow*, Computer Methods Appl. Mech. Engrg., 357 (2019), 112596.
- [73] H. B. ZUBAIR, *Efficient Multigrid Methods Based on Improved Coarse Grid Correction Techniques*, Ph.D. thesis, Delft University of Technology, 2009.