

Multifidelity Data Assimilation for Physical Systems



Andrey A. Popov and Adrian Sandu

Abstract Multifidelity methods aim to leverage the availability of models at different levels of fidelity describing the same physical phenomena and are receiving growing attention in computational science. One field that can considerably benefit from statistical multifidelity approaches is data assimilation. This chapter presents a broad overview of multifidelity methods in data assimilation for hierarchies of models and hierarchies of observations. We introduce the theoretical multifidelity Kalman filter, and discuss its practical implementation using an ensemble-based framework as the multifidelity ensemble Kalman filter (MFEEnKF). The discussion builds upon the theory of linear and nonlinear control variates. Numerical examples compare the multifidelity and the traditional EnKF.

1 Introduction

An often ignored principle in Bayesian inference is that the inference requires the utilization of all available knowledge and all the relevant information available (Jaynes 2003). In the context of data assimilation, especially for physical systems, one often has access to hierarchies of multiple models, each one more accurate than its predecessor in the hierarchy; higher resolution models can be obtained by simply refining the simulation grid or step-size, or through the ability to more accurately capture the underlying physical phenomena. In addition, one frequently has access to observations of the same physical variable through different types of sensors. The principle of Bayesian inference asks to not discard this information, but to incorporate it whenever possible.

Multifidelity data assimilation refers to methods that merge information about the same underlying natural truth obtained through the use of multiple models or obser-

A. A. Popov (✉) · A. Sandu
Virginia Tech, Blacksburg, VA, USA
e-mail: apopov@vt.edu

A. Sandu
e-mail: sandu@cs.vt.edu

vation operators at different levels of fidelity. For a survey of general multifidelity methods and types of multifidelity models see Peherstorfer et al. (2018).

First introduced by Giles (2008) and then later more formally defined in Giles (2015) the ‘multilevel’ Monte Carlo filter describes the optimal linear coupling between different ‘levels’ of simulations (which we generalize to calling fidelities). Here we aim to generalize the notion of levels to fidelities. We introduce a formal notion of structure in the relation between fidelities, and also within the structure that optimally combines the information contained therein.

This chapter is organized as follows. The rest of the introduction is concerned with describing the data assimilation problem, and the idea of multifidelity models. Control variate theory is introduced in Sect. 2. The problem of multifidelity inference is introduced in Sect. 3, with the multifidelity Kalman filter discussed in Sect. 3.1 and the multifidelity ensemble Kalman filter in Sect. 3.2. We then introduce the concept of multifidelity observations in Sect. 4. A trivial numerical example with the Lorenz ‘96 system is shown in Sect. 5. We conclude with some remarks in Sect. 6.

1.1 Notation

Consider a random variable χ . The distribution of χ is denoted by π_χ , and an ensemble representing N samples from the distribution by $\mathbf{E}_\chi = [\chi^{(1)}, \dots, \chi^{(N)}]$. We (exactly) denote the mean by $\boldsymbol{\mu}_\chi$, and the empirical sample mean by $\tilde{\boldsymbol{\mu}}_\chi$. The covariance between random variables χ and ν is denoted by $\Sigma_{\chi,\nu}$, and the empirical sample covariance by $\tilde{\Sigma}_{\chi,\nu}$.

1.2 The Data Assimilation Problem

We seek to model the state X of a dynamical system with an imperfect model,

$$X_i = \mathcal{M}_i(X_{i-1}) + \Xi_i, \quad (1)$$

where the model errors at different times Ξ_i are independent of each other. We assume the errors have mean zero, $\boldsymbol{\mu}_{\Xi_i} = 0$, and covariances Σ_{Ξ_i, Ξ_i} .

Observations of the true state X_i^t are available at discrete time moments i :

$$Y_i = \mathcal{H}_i(X_i^t) + \eta_i, \quad (2)$$

where we again assume that the observation errors η_i at different times are independent of each other, have mean zero, $\boldsymbol{\mu}_{\eta_i} = 0$, and covariances Σ_{Y_i, Y_i} .

Given prior information X_i^b about the state at time i , and noisy observations of the truth, the filtering problem consists of sequentially computing the posterior, X_i^a in some (usually Bayesian) inference sense.

Explicitly, the Bayesian formulation (Reich and Cotter 2015) aims to find

$$\pi(X_i^a) = \pi(X_i^b|Y_i) = \frac{\pi(Y_i|X_i^b)\pi(X_i^b)}{\pi(Y_i)}, \quad (3)$$

in some approximate way, as the problem, more often than not, is computationally intractable.

1.3 Multifidelity Models

The focus of multifidelity data assimilation is to compute the inference (3) using not a single model (1), but leveraging a hierarchy of models at different fidelities. Consider two levels of fidelity, and assume that our high fidelity state variable is X and coarse fidelity variable is U . The two models that propagate these quantities in time are:

$$X_i = \mathcal{M}_i^X(X_{i-1}), \quad (4)$$

$$U_i = \mathcal{M}_i^U(U_{i-1}). \quad (5)$$

The goal of multifidelity data assimilation is to make use of these different models to incorporate as much information as possible.

An important aspect of multifidelity models, which is a generalization of multi-level hierarchies, is that the the state spaces of the different models do not necessarily have to be the same. In fact, we will assume that the fine fidelity model state can be embedded into \mathbb{R}^n and that the coarse fidelity model state can be embedded into \mathbb{R}^r , where typically $r < n$, though this is not necessarily the case.

Alternatively, we can think of the word ‘model’ as describing the operator about which we are optimizing. In a data assimilation context this will often be our observations. Assume that there exist two ways of obtaining observations (2) of the same fundamental phenomenon, one defined by a fine fidelity operator \mathcal{H}^X , and the other defined by a coarse fidelity observation operator, \mathcal{H}^U , such that,

$$Y^X = \mathcal{H}^X(\mathbf{X}^t) + \eta^X, \quad (6)$$

$$Y^U = \mathcal{H}^U(\mathbf{X}^t) + \eta^U, \quad (7)$$

wherein the goal would shift to either combining and utilizing the observations in some optimal way without loss of information, but also without duplication of information.

2 Control Variates

The linear control variate technique (Rubinstein and Marcus 1985) is a method for reducing the variance of an estimator by making use of highly correlated data about which additional information is known. Assume our quantity of interest is described by the distribution of the *principal variate* χ . The distribution of the highly correlated *control variate* \hat{v} describes information in an alternate way (such as in a different space), and the distribution of an independent (or more weaker, uncorrelated) *ancillary variate* v describes information related to that of the control variate and shares the same mean, $\mu_{\hat{v}} = \mu_v$. The linear control variate approach builds a *total variate* ζ

$$\zeta = \chi - \mathbf{S}(\hat{v} - v), \quad (8)$$

where the free parameter \mathbf{S} , known as the *gain operator*, is chosen to minimize the generalized variance of ζ . The three variates that make up the total variate will be collectively called the *constituent variates*.

Theorem 1 (Unbiased nature of linear control variates) *Without proof, the mean of the total variate equals the mean of the principal variate,*

$$\mu_{\zeta} = \mu_{\chi}. \quad (9)$$

Theorem 2 (Optimal gain for linear control variates) *The optimal gain matrix \mathbf{S} that minimizes the trace of the covariance of (8) is*

$$\mathbf{S} = \Sigma_{\chi, \hat{v}} (\Sigma_{\hat{v}, \hat{v}} + \Sigma_{v, v})^{-1}. \quad (10)$$

Proof Observe that the covariance of (8) is

$$\Sigma_{\zeta, \zeta} = \Sigma_{\chi, \chi} - \mathbf{S}\Sigma_{\hat{v}, \chi} - \Sigma_{\chi, \hat{v}}\mathbf{S}^T + \mathbf{S}\Sigma_{\hat{v}, \hat{v}}\mathbf{S}^T + \mathbf{S}\Sigma_{v, v}\mathbf{S}^T. \quad (11)$$

Taking the derivative with respect to \mathbf{S} of the trace of (11),

$$\frac{\partial}{\partial \mathbf{S}} \text{tr}(\Sigma_{\zeta, \zeta}) = -2\Sigma_{\chi, \hat{v}} + 2\mathbf{S}(\Sigma_{\hat{v}, \hat{v}} + \Sigma_{v, v}), \quad (12)$$

the local minimum is found at (10), as required.

Corollary 1 *By simple manipulation, the covariance of the total variate under the optimal gain from Theorem 2 is:*

$$\Sigma_{\zeta, \zeta} = \Sigma_{\chi, \chi} - \Sigma_{\chi, \hat{v}} (\Sigma_{\hat{v}, \hat{v}} + \Sigma_{v, v})^{-1} \Sigma_{\hat{v}, \chi}, \quad (13)$$

from which it is clear that $\Sigma_{\zeta, \zeta} \leq \Sigma_{\chi, \chi}$ in the symmetric semi-positive definite sense.

The linear control variate technique can be derived in a parallel but completely alternate way. Taking the principal, control, and ancillary variates as Gaussian random variables, the total variate is the solution to the Bayesian inference problem,

$$\pi(\zeta) = \pi(\chi | \hat{v}, \nu). \quad (14)$$

which is a well-known result due to Kalman (1960).

Following Nelson (1987) we now attempt to introduce the idea of non-linear control variates. Instead of searching for function approximations that follow a set of rules, we will instead view the problem of finding the total variate ζ in terms of the principal, control, and ancillary variates as an inference problem, generalizing (14) to arbitrary distributed variables. Specifically, we seek to cast the general inference problem (and specific approximations thereof) into an application of some problem-specific transform,

$$\zeta = \mathcal{T}(\chi, \hat{v}, \nu), \quad (15)$$

with the function \mathcal{T} represents a distribution transformation on the principal variate, built making use of the information given by the control and ancillary variates.

For the remainder of this chapter we will assume that the control variate is related to the principal variate through a deterministic function (coupling),

$$\hat{v} = \theta(\chi), \quad (16)$$

which implies that there is necessarily some loss of information from the space of the total and principal variates to the space of the control and ancillary variates.

An important generalization of the control variate concept is its ability to be applied in a nested form. This means that the total variate ζ can itself be an ancillary variate for a finer fidelity principal variate. Assume that we have to have \mathcal{F} fidelities, with $\nu_{\mathcal{F}}$ being the coarsest fidelity ancillary variate. Its corresponding control variate is $\hat{v}_{\mathcal{F}}$, and its principal and total variates are on level $\mathcal{F} - 1$: $\chi_{\mathcal{F}-1}$ and $\zeta_{\mathcal{F}-1}$. The total variate is then also the ancillary variate for the next set,

$$\nu_{\mathcal{F}-1} \leftarrow \zeta_{\mathcal{F}-1}, \quad (17)$$

which can be generalized all the way up the chain, until we reach the constituent variates χ_1 , \hat{v}_2 , and ν_2 , that represent the full information content through the total variate ζ_1 . Explicitly, from Popov et al. (2020), the total variate for \mathcal{F} fidelities and the corresponding optimal gain matrices can be written as,

$$\zeta = \chi - \sum_{f=1}^{\mathcal{F}} \bar{\mathbf{S}}_f (\hat{v}_f - \nu_f), \quad \bar{\mathbf{S}}_f = \prod_{\psi=1}^f \mathbf{S}_{\psi}. \quad (18)$$

2.1 Ensemble Control Variates

Instead of employing the exact distribution of a random variable, which usually is considered to be an intractable task, an ensemble of samples is typically used.

We will discuss the ways in which ensemble multifidelity inference is performed in a later section. Here we concern ourselves with the problem of finding an ensemble representation of the total variate (8) given an ensemble of N_χ samples of the principal variate $\mathbf{E}_\chi = [\boldsymbol{\chi}^{(1)}, \dots, \boldsymbol{\chi}^{(N_\chi)}]$ and corresponding pairwise samples of the control variate $\mathbf{E}_{\hat{v}} = [\hat{v}^{(1)}, \dots, \hat{v}^{(N_\chi)}]$. We seek to find an ensemble of the total variate $\mathbf{E}_\zeta = [\boldsymbol{\zeta}^{(1)}, \dots, \boldsymbol{\zeta}^{(N_\chi)}]$.

We will define the ensemble means as

$$\tilde{\boldsymbol{\mu}}_\chi = \frac{1}{N_\chi} \mathbf{E}_\chi \mathbf{1}_{N_\chi} \in \mathfrak{R}^n, \quad \tilde{\boldsymbol{\mu}}_{\hat{v}} = \frac{1}{N_\chi} \mathbf{E}_{\hat{v}} \mathbf{1}_{N_\chi} \in \mathfrak{R}^r, \quad (19)$$

and the anomalies as,

$$\mathbf{A}_\chi = \mathbf{E}_\chi - \tilde{\boldsymbol{\mu}}_\chi \mathbf{1}_{N_\chi}^T \in \mathfrak{R}^{n \times N_\chi}, \quad \mathbf{A}_{\hat{v}} = \mathbf{E}_{\hat{v}} - \tilde{\boldsymbol{\mu}}_{\hat{v}} \mathbf{1}_{N_\chi}^T \in \mathfrak{R}^{r \times N_\chi}. \quad (20)$$

Assume that we are given either the mean and covariance of the ancillary variate ($\boldsymbol{\mu}_v$ and $\Sigma_{v,v}$), or that we are able to derive empirical approximations $\tilde{\boldsymbol{\mu}}_v$ and $\tilde{\Sigma}_{v,v}$ an ensemble of N_v samples of v , \mathbf{E}_v . In the first approach we utilize the linear control variate framework (8).

There are numerous ways in which to derive the ensemble of the total variate. One way is to create a synthetic ensemble of N_χ samples of the ancillary variate sampled from its known distribution. Denote this ensemble $\tilde{\mathbf{E}}_v$. Under the linear control variate approach,

$$\mathbf{E}_\zeta = \mathbf{E}_\chi - \mathbf{S}(\mathbf{E}_{\hat{v}} - \tilde{\mathbf{E}}_v), \quad (21)$$

where the optimal gain is approximated by

$$\tilde{\mathbf{S}} = \tilde{\Sigma}_{\chi,\hat{v}} (\tilde{\Sigma}_{\hat{v},\hat{v}} + \Sigma_{v,v})^{-1}. \quad (22)$$

The astute reader will recognize this as the ‘perturbed observations’ ensemble Kalman filter (Houtekamer and Mitchell 1998).

An alternate formulation that does away with the synthetic ensemble assumes a Gaussian prior on the ancillary variate, and uses the optimal empirical gain,

$$\mathbf{E}_\zeta = \tilde{\boldsymbol{\mu}}_\chi \mathbf{1}_{N_\chi}^T + \mathbf{A}_\chi \mathbf{T} \left(\mathbf{I}_{N_\chi} - \frac{1}{N_\chi - 1} \mathbf{T}^T \mathbf{A}_{\hat{v}}^T \Sigma_{v,v}^{-1} (\tilde{\boldsymbol{\mu}}_{\hat{v}} - \boldsymbol{\mu}_v) \mathbf{1}_{N_\chi}^T \right), \quad (23)$$

$$\mathbf{T} = \left(\mathbf{I}_{N_\chi} - \frac{1}{N_\chi - 1} \mathbf{A}_{\hat{v}}^T (\tilde{\Sigma}_{\hat{v},\hat{v}} + \Sigma_{v,v})^{-1} \mathbf{A}_{\hat{v}} \right)^{\frac{1}{2}}, \quad (24)$$

which the astute reader will recognize as the ensemble transform Kalman filter (ETKF) (Bishop et al. 2001).

Note that if it is not possible to represent the covariance of the ancillary variate exactly, then one needs to compute the optimal gain in alternate ways.

Another interesting approach to ensemble inference is the importance sampling optimal transport procedure (Reich 2013). In essence, one constructs the posterior mean from the importance sampling procedure,

$$\boldsymbol{\mu}_\zeta = \mathbf{E}_\chi \mathbf{w}, \quad (25)$$

$$[\mathbf{w}]_i \propto \pi_\nu(\hat{\boldsymbol{\theta}}^{(i)}). \quad (26)$$

The ansatz is made that the optimal transportation into an equally weighted ensemble with the same mean defines an ensemble with the same empirical moments as those defined by the importance sampling weights,

$$\mathbf{E}_\zeta = \mathbf{E}_\chi \mathbf{T}^*, \quad (27)$$

where the optimal transport is defined in the Monge-Kantorovich sense,

$$\mathbf{T}^* = \arg \min_{\mathbf{T}} \sum_{1 \leq i, j \leq N_\chi} [\mathbf{T}]_{ij} \|\boldsymbol{\chi}^{(i)} - \boldsymbol{\chi}^{(j)}\|_2^2 \quad (28)$$

$$\text{subject to: } \mathbf{T} \mathbf{1}_{N_\chi} = N_\chi \mathbf{w}, \mathbf{T}^T \mathbf{1}_{N_\chi} = \mathbf{1}_{N_\chi}, [\mathbf{T}]_{ij} \geq 0,$$

which ensures that the weights of the new posterior ensemble are equal.

Second order accurate (preserving the weighted ensemble covariance) extensions to this formulation exist (Acevedo et al. 2017) and should be used if this methodology is to be attempted operationally.

3 Multifidelity Filtering

For ease of exposition we primarily focus on the case of two fidelities; multifidelity extensions will be described separately.

Assume now that the state of a dynamical system is our quantity of interest, and that there are two different fidelities in which we can represent it: fine and coarse. Let the distribution of the principal variate X^b represent the prior information about the state at fine fidelity. Let \hat{U}^b be its corresponding control variate, and U^b be the ancillary variate, the distributions of which describe information about the state at coarse fidelity.

Assume that the prior total variate Z^b represents the general posterior of the multifidelity inference procedure (14). Note that it is possible, with some abuse of notation, to represent the inference as an application of some nonlinear function,

$$Z^b = \mathcal{C}(X^b, \hat{U}^b, U^b), \quad (29)$$

with the function \mathcal{C} defining an implicit assumption about the relationship between the four variates, such as the linear control variate assumption (8) or an optimal transport based assumption (27) if our variates are represented by ensembles. In the most general sense, \mathcal{C} can represent some non-linear variance reduction technique that is informed by the distributions of the constituent variates (Nelson 1987).

The prior total variate is Z^b and the posterior total variate is Z^a , defined by the same function applied to its component variates:

$$Z^a = \mathcal{C}(X^a, \hat{U}^a, U^a). \quad (30)$$

The inference step from the prior total to the posterior total variates is a filtering step which explicitly combines information,

$$Z^a = \mathcal{F}(Z^b, \mathcal{H}(Z^b), Y), \quad (31)$$

with the function \mathcal{F} standing in for some filter, such as the Kalman filter.

The principal variate can be propagated by some constituent filter,

$$X^a = \tilde{\mathcal{F}}_X^Z(X^b, \mathcal{H}(X^b), Y), \quad (32)$$

which is dependent on Z^b , Z^a and the filter \mathcal{F} that is implicitly applied between them. Similar formulations can be made for the other constituent variates.

Note that the goal of one step of a multifidelity filter is not to find the posterior total variate Z^a , but rather to find posteriors of its constituent variates, X^a , \hat{U}^a , and U^a . In fact, as the total variate is merely a synthetic construction, the multifidelity inference reduces to performing virtual inference on the total variate by manipulating the principal, control, and ancillary variates. In this way the explicit filtering of the total variate (31) is not performed, but only the constituent filtering problems (32) are explicitly solved.

While the general problem of finding the analysis principal variate X^a given only the analysis total variate Z^a is not well posed, the combined problem of finding the distributions of X^a , \hat{U}^a , and U^a may be posed in terms of a minimum cross entropy problem:

$$X^a, \hat{U}^a, U^a = \arg \min_{X^a, \hat{U}^a, U^a} H(X^b, X^a) + H(\hat{U}^b, \hat{U}^a) + H(U^b, U^a), \quad (33)$$

subject to the constraints,

$$Z^a = \mathcal{C}(X^a, \hat{U}^a, U^a), \quad Z^a = \mathcal{F}(Z^b, \mathcal{H}(Z^b), Y), \quad (34)$$

$$\mathbb{E}[\hat{U}^a] = \mathbb{E}[U^a], \quad \hat{U}^a = \theta(X^a), \quad (35)$$

from which the constituent filters (32) are implicitly defined.

A powerful assumption that can be made is that the same control structure imposed on the prior is also imposed on the posterior. We call this the ‘control structure consistency assumption’. One way in which this holds in the linear control variate approach is:

$$Z^b = X^b - \mathbf{S}^b(\hat{U}^b - U^b), \quad (36)$$

$$Z^a = X^a - \mathbf{S}^a(\hat{U}^a - U^a), \quad (37)$$

where we impose the assumption that the (approximately) optimal prior and posterior gains are equivalent,

$$\mathbf{S}^b = \mathbf{S}^a, \quad (38)$$

meaning that we restrict all possible posterior constituent variates to ones that obey the same structure as their prior counterparts. One way in which this is achieved is by assuming a particular structure on the relationship between the principal and control variate (16) from Popov et al. (2020).

3.1 Multifidelity Kalman Filter

We now introduce the multifidelity Kalman filter (MFKF), fleshed out from Popov et al. (2020). As Gaussian random variables can be trivially combined through known formulas involving their means and covariances, the MFKF is not an algorithm that needs to exist for the purposes of practical implementation, but merely needs to exist to explain derivations of practical extensions thereof.

We restrict ourselves to a linear principal-control variate coupling (16),

$$\hat{U}_i = \theta(X_i) = \Theta X_i, \quad (39)$$

with Θ a projection operator from the n -dimensional space of the principal variate onto the r -dimensional space of the control variate. The corresponding interpolating

operator is denoted Φ (such that $\Theta \Phi = \mathbf{I}_r$). We decompose the principal variate into its control variate and residual variate components:

$$X_i = \Theta \hat{U}_i + \Delta X_i. \quad (40)$$

Additionally, as is canonical, we restrict ourselves to the case of a linear observation operator \mathbf{H}_i .

For the rest of this chapter we assume that we seek to propagate the total variate,

$$Z_i = X_i - \mathbf{S}_i (\hat{U}_i - U_i), \quad (41)$$

through both a dynamical model (forecast step), and through the analysis step conditioned by observations.

We express the moments of the total variate in terms of the moments of the corresponding constituents:

$$\boldsymbol{\mu}_{Z_i} = \boldsymbol{\mu}_{X_i} - \mathbf{S}_i (\boldsymbol{\mu}_{\hat{U}_i} - \boldsymbol{\mu}_{U_i}), \quad (42)$$

$$\boldsymbol{\Sigma}_{Z_i, Z_i} = \boldsymbol{\Sigma}_{X_i, X_i} + \mathbf{S}_i \boldsymbol{\Sigma}_{\hat{U}_i, \hat{U}_i} \mathbf{S}_i^T + \mathbf{S}_i \boldsymbol{\Sigma}_{U_i, U_i} \mathbf{S}_i^T - \mathbf{S}_i \boldsymbol{\Sigma}_{\hat{U}_i, X_i} - \boldsymbol{\Sigma}_{X_i, \hat{U}_i} \mathbf{S}_i^T. \quad (43)$$

We are now ready to look at the MFKF. For the forecast step, assume that we have a linear fine fidelity model \mathbf{M}_i^X , and a linear coarse fidelity model \mathbf{M}_i^U . Assume that the error Ξ_i of the fine fidelity model is unbiased and is known to have covariance $\boldsymbol{\Sigma}_{\Xi_i, \Xi_i}$. Assume additionally that the coarse fidelity model has no error in the coarse subspace in relation to the truth. This could be because the coarse fidelity model was built to capture this error through data driven closures.

Assume that we have the posterior information at the previous step $i - 1$ about the principal, control, and ancillary variates, and that the relation between the principal and control variate (39) holds. We propagate the means as follows:

$$\boldsymbol{\mu}_{X_i^b} = \mathbf{M}_i^X \boldsymbol{\mu}_{X_{i-1}^a}, \quad (44)$$

$$\boldsymbol{\mu}_{\hat{U}_i^b} = \mathbf{M}_i^U \boldsymbol{\mu}_{\hat{U}_{i-1}^a}, \quad (45)$$

$$\boldsymbol{\mu}_{U_i^b} = \mathbf{M}_i^U \boldsymbol{\mu}_{U_{i-1}^a}, \quad (46)$$

with the covariances propagated as,

$$\boldsymbol{\Sigma}_{X_i^b, X_i^b} = \mathbf{M}_i^X \boldsymbol{\Sigma}_{X_{i-1}^a, X_{i-1}^a} \mathbf{M}_i^{X,T} + \boldsymbol{\Sigma}_{\Xi_i, \Xi_i}, \quad (47)$$

$$\boldsymbol{\Sigma}_{X_i^b, \hat{U}_i^b} = \mathbf{M}_i^X \boldsymbol{\Sigma}_{X_{i-1}^a, \hat{U}_{i-1}^a} \mathbf{M}_i^{U,T}, \quad (48)$$

$$\boldsymbol{\Sigma}_{\hat{U}_i^b, \hat{U}_i^b} = \mathbf{M}_i^U \boldsymbol{\Sigma}_{\hat{U}_{i-1}^a, \hat{U}_{i-1}^a} \mathbf{M}_i^{U,T}, \quad (49)$$

$$\boldsymbol{\Sigma}_{U_i^b, U_i^b} = \mathbf{M}_i^U \boldsymbol{\Sigma}_{U_{i-1}^a, U_{i-1}^a} \mathbf{M}_i^{U,T}, \quad (50)$$

We note that unless the principal variate residual is not propagated in control space by the fine fidelity model,

$$\Theta \mathbf{M}_i^X \Delta X_{i-1}^a = 0, \quad (51)$$

then the above propagation will violate (39). Therefore, as a useful heuristic, the propagation of the control variate moments can be replaced by the propagation of the projected principal variate moments in order for the relation (39) to hold more strongly at each step,

$$\boldsymbol{\mu}_{\hat{U}_i^b} = \mathbf{M}_i^U \Theta \boldsymbol{\mu}_{X_{i-1}^a}, \quad (52)$$

$$\Sigma_{X_i^b, \hat{U}_i^b} = \mathbf{M}_i^X \Sigma_{X_{i-1}^a, X_{i-1}^a} (\mathbf{M}_i^U \Theta)^T, \quad (53)$$

$$\Sigma_{\hat{U}_i^b, \hat{U}_i^b} = \mathbf{M}_i^U \Theta \Sigma_{X_{i-1}^a, X_{i-1}^a} (\mathbf{M}_i^U \Theta)^T. \quad (54)$$

This is especially useful if the models are non-linear, generalizing to the multifidelity extended Kalman filter, or in the case of the multifidelity ensemble Kalman filter later in the chapter in Sect. 3.2.

Lemma 1 *The fine fidelity model, coarse fidelity model, posterior optimal gain at step $i - 1$, and prior optimal gain at time i are related as follows:*

$$\mathbf{M}_i^X \mathbf{S}_{i-1}^a = \mathbf{S}_i^b \mathbf{M}_i^U. \quad (55)$$

Proof By simple manipulation,

$$\mathbf{M}_i^X \mathbf{S}_{i-1}^a = \mathbf{M}_i^X \Sigma_{X_{i-1}^a, \hat{U}_{i-1}^a} \left(\Sigma_{\hat{U}_{i-1}^a, \hat{U}_{i-1}^a} + \Sigma_{U_{i-1}^a, U_{i-1}^a} \right)^{-1} \quad (56)$$

$$= \Sigma_{X_i^b, \hat{U}_{i-1}^a} \mathbf{M}_i^{U,T} \left[\mathbf{M}_i^U \left(\Sigma_{\hat{U}_{i-1}^a, \hat{U}_{i-1}^a} + \Sigma_{U_{i-1}^a, U_{i-1}^a} \right) \mathbf{M}_i^{U,T} \right]^{-1} \mathbf{M}_i^U \quad (57)$$

$$= \Sigma_{X_i^b, \hat{U}_i^b} \left(\Sigma_{\hat{U}_i^b, \hat{U}_i^b} + \Sigma_{U_i^b, U_i^b} \right)^{-1} \mathbf{M}_i^U = \mathbf{S}_i^b \mathbf{M}_i^U, \quad (58)$$

as required.

Theorem 3 *The MFKF forecast is the total variate forecast:*

$$\boldsymbol{\mu}_{Z_i^b} = \mathbf{M}_i^X \boldsymbol{\mu}_{Z_{i-1}^a}, \quad (59)$$

$$\Sigma_{Z_i^b, Z_i^b} = \mathbf{M}_i^X \Sigma_{Z_{i-1}^a, Z_{i-1}^a} \mathbf{M}_i^{X,T} + \Sigma_{\Xi_i, \Xi_i}. \quad (60)$$

Proof Using lemma 1, manipulate the formulation for the mean of Z in (42),

$$\mathbf{M}_i^X \boldsymbol{\mu}_{Z_{i-1}^a} = \mathbf{M}_i^X \boldsymbol{\mu}_{Z_{i-1}^a}, \quad (61)$$

$$= \mathbf{M}_i^X \boldsymbol{\mu}_{X_{i-1}^a} - \mathbf{M}_i^X \mathbf{S}_{i-1}^a (\boldsymbol{\mu}_{\hat{U}_{i-1}^a} - \boldsymbol{\mu}_{U_{i-1}^a}), \quad (62)$$

$$= \boldsymbol{\mu}_{X_i^b} - \mathbf{S}_i^b \mathbf{M}_i^U (\boldsymbol{\mu}_{\hat{U}_{i-1}^a} - \boldsymbol{\mu}_{U_{i-1}^a}), \quad (63)$$

$$= \boldsymbol{\mu}_{X_i^b} - \mathbf{S}_i^b (\boldsymbol{\mu}_{\hat{U}_i^b} - \boldsymbol{\mu}_{U_i^b}) = \boldsymbol{\mu}_{Z_i^b}, \quad (64)$$

as required. A similar manipulation can be performed for the covariance.

In order to obtain an efficient implementation of the analysis step in the MFKF, we need to restrict the projection operator (39) to a class that has ‘nice’ properties. We assume that the joint variability of the principal variate in the orthogonal complement space and control variate is negligible,

$$\Sigma_{\Delta X, \hat{U}} \approx 0, \quad (65)$$

or alternatively that the projection operator Θ captures the dominant linear modes of the variability in the dynamics of X . Common methods by which such operators can be obtained are POD and DMD, and variants thereof (Brunton and Kutz 2019).

Theorem 4 *If the first two moments of the control and ancillary variate are identical, and assumption (65) holds, then the optimal gain is,*

$$\mathbf{S} = \frac{1}{2} \Theta. \quad (66)$$

Proof Observe by Theorem 2 and (65),

$$\mathbf{S} = \Sigma_{X, \hat{U}} (\Sigma_{\hat{U}, \hat{U}} + \Sigma_{U, U})^{-1} = \frac{1}{2} \Theta \Sigma_{\hat{U}, \hat{U}} \Sigma_{\hat{U}, \hat{U}}^{-1} = \frac{1}{2} \Theta, \quad (67)$$

as required.

If we choose a projection operator for which (65) holds, then the optimal gain is constant and does not have to be estimated. Moreover this provides for a clear relationship between the projection operator Θ and the optimal gain, such that $\Theta \mathbf{S} = \frac{1}{2} \mathbf{I}_r$. For the rest of this section we assume that \mathbf{S} is constant.

We next discuss the analysis step of the MFKF. Note first that the Kalman gain is the optimal gain when the principal variate is the prior information about the state of the dynamics, the control variate is that information cast into observation space, and the ancillary variate are the independent observations. Assume that the arbitrary variate W_i^b represents some prior information, we write the Kalman gain as a function of W_i^b ,

$$\mathbf{K}_{W_i^b} = \Sigma_{W_i^b, \mathbf{H}_i} W_i^b \left(\Sigma_{\mathbf{H}_i, W_i^b, \mathbf{H}_i} W_i^b + \Sigma_{Y_i, Y_i} \right)^{-1}. \quad (68)$$

The standard Kalman filter analysis step applied to the total variate, as described by (31):

$$Z_i^a(\mathbf{K}_{Z_i^b}) = Z_i^b - \mathbf{K}_{Z_i^b}(\mathbf{H}_i Z_i^b - Y_i), \quad (69)$$

can be decomposed into its constituent variates:

$$X^a - \mathbf{S}(\hat{U}^a - U^a) = X^b - \mathbf{S}(\hat{U}^b - U^b) - \mathbf{K}_{Z_i^b} \left(\mathbf{H} \left(X^b - \mathbf{S}(\hat{U}^b - U^b) \right) - Y \right). \quad (70)$$

Taking the ‘natural’ decomposition of this relation into components leads to, the constituent filters (32):

$$X^a = X^b - \mathbf{K}_{Z_i^b} (\mathbf{H} X^b - Y), \quad (71)$$

$$\hat{U}^a = \hat{U}^b - \Theta \mathbf{K}_{Z_i^b} (\mathbf{H} \Phi \hat{U}^b - Y), \quad (72)$$

$$U^a = U^b - \Theta \mathbf{K}_{Z_i^b} (\mathbf{H} \Phi U^b - Y), \quad (73)$$

which assumes that the control and ancillary variates do not carry any additional information from the orthogonal complement space of the principal variate.

The authors conjecture that the decomposition (71) approximately minimizes the cross entropy functional (33) out of all such decompositions, though there is no strong evidence for this claim as of yet.

The propagation of the total mean through its constituent variate means is:

$$\boldsymbol{\mu}_{X^a} = \boldsymbol{\mu}_{X^b} - \mathbf{K}_{Z_i^b} (\mathbf{H} \boldsymbol{\mu}_{X^b} - \mathbf{y}) \quad (74)$$

$$\boldsymbol{\mu}_{\hat{U}^a} = \boldsymbol{\mu}_{\hat{U}^b} - \Theta \mathbf{K}_{Z_i^b} (\mathbf{H} \Phi \boldsymbol{\mu}_{\hat{U}^b} - \mathbf{y}), \quad (75)$$

$$\boldsymbol{\mu}_{U^a} = \boldsymbol{\mu}_{U^b} - \Theta \mathbf{K}_{Z_i^b} (\mathbf{H} \Phi \boldsymbol{\mu}_{U^b} - \mathbf{y}). \quad (76)$$

The corresponding covariance update formulas are:

$$\Sigma_{X_i^a, X_i^a} = (\mathbf{I} - \mathbf{K}_{Z_i^b} \mathbf{H}) \Sigma_{X_i^b, X_i^b} (\mathbf{I} - \mathbf{K}_{Z_i^b} \mathbf{H})^T + \mathbf{K}_{Z_i^b} \Sigma_{Y_i, Y_i} \mathbf{K}_{Z_i^b}^T, \quad (77)$$

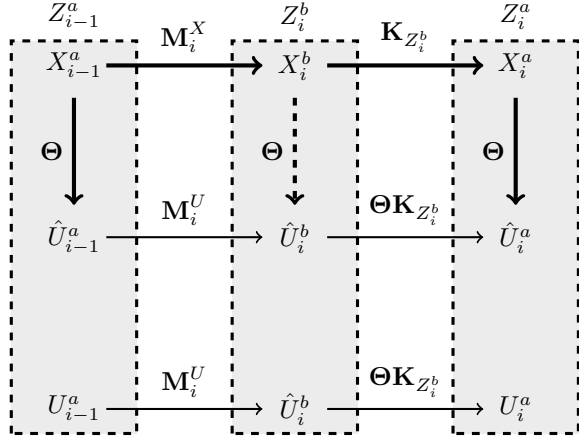
$$\Sigma_{X_i^a, \hat{U}_i^a} = (\mathbf{I} - \mathbf{K}_{Z_i^b} \mathbf{H}) \Sigma_{X_i^b, \hat{U}_i^b} (\mathbf{I} - \Theta \mathbf{K}_{Z_i^b} \mathbf{H}_i \Phi)^T, \quad (78)$$

$$+ \mathbf{K}_{Z_i^b} \Sigma_{Y_i, Y_i} (\Theta \mathbf{K}_{Z_i^b})^T, \quad (79)$$

$$\Sigma_{\hat{U}_i^a, \hat{U}_i^a} = (\mathbf{I} - \Theta \mathbf{K}_{Z_i^b} \mathbf{H}_i \Phi) \Sigma_{\hat{U}_i^b, \hat{U}_i^b} (\mathbf{I} - \Theta \mathbf{K}_{Z_i^b} \mathbf{H}_i \Phi)^T \quad (80)$$

$$+ \Theta \mathbf{K}_{Z_i^b} \Sigma_{Y_i, Y_i} (\Theta \mathbf{K}_{Z_i^b})^T, \quad (81)$$

Fig. 1 A diagram of the MFKF computations, including the heuristic re-correlation of the principal and control variates (52), and the assumed relation between the forecast variates (16)



$$\Sigma_{U_i^a, U_i^a} = \left(\mathbf{I} - \Theta \mathbf{K}_{Z_i^b} \mathbf{H}_i \Phi \right) \Sigma_{U_i^b, U_i^b} \left(\mathbf{I} - \Theta \mathbf{K}_{Z_i^b} \mathbf{H}_i \Phi \right)^T \quad (82)$$

$$+ \Theta \mathbf{K}_{Z_i^b} \Sigma_{Y_i, Y_i} \left(\Theta \mathbf{K}_{Z_i^b} \right)^T. \quad (83)$$

The inner working of the MFKF is illustrated in Fig. 1.

Theorem 5 *Without proof, if the optimal gain interpolation projection step does not remove additional information from the control and ancillary variate (that is (65) is exact), then the ‘natural’ decomposition (71) is exact, thus the linear control variate combination of the mean is the total variate analysis mean,*

$$\boldsymbol{\mu}_{Z_i^a} = \boldsymbol{\mu}_{X_i^a} - \mathbf{S}(\boldsymbol{\mu}_{\hat{U}_i^a} - \boldsymbol{\mu}_{U_i^a}). \quad (84)$$

Similarly for the covariances.

Theorem 6 *Without proof, if \mathbf{S} is the optimal gain (Theorems 2 and 4), then the simple relation between the covariances of the principal and total variates is,*

$$\Sigma_{Z_i, Z_i} = (\mathbf{I}_n - \mathbf{S}\Theta) \Sigma_{X_i, X_i}. \quad (85)$$

Thus we are able to obtain a covariance for the total variate by only knowing the covariance of the principal variate.

An alternate decomposition for which Theorem 5 is exact without qualification, that we will not be analyzing is:

$$X^a = X^b - \mathbf{K}_{Z_i^b} (\mathbf{H}Z^b - Y), \quad (86)$$

$$\hat{U}^a = \hat{U}^b - \Theta \mathbf{K}_{Z_i^b} (\mathbf{H}Z^b - Y), \quad (87)$$

$$U^a = U^b - \Theta \mathbf{K}_{Z_i^b} (\mathbf{H}Z^b - Y), \quad (88)$$

These formulas, however, are difficult to implement using ensembles.

We show next how the variability of the total variate $Z_i^a(\mathbf{K}_{Z_i^b})$, the principal variate $X_i^a(\mathbf{K}_{Z_i^b})$, and the principal variate analyzed by itself $X_i^a(\mathbf{K}_{X_i^b})$ are related.

Theorem 7 *The covariances of $Z_i^a(\mathbf{K}_{Z_i^b})$, $X_i^a(\mathbf{K}_{Z_i^b})$, $X_i^a(\mathbf{K}_{X_i^b})$ are such that:*

$$\Sigma_{Z^a(\mathbf{K}_{Z_i^b}), Z^a(\mathbf{K}_{Z_i^b})} \leq \Sigma_{X^a(\mathbf{K}_{X_i^b}), X^a(\mathbf{K}_{X_i^b})} \leq \Sigma_{X^a(\mathbf{K}_{Z_i^b}), X^a(\mathbf{K}_{Z_i^b})}. \quad (89)$$

Proof By the optimality of the Kalman gain $\mathbf{K}_{Z_i^b}$ in Theorem 2,

$$\Sigma_{Z^a(\mathbf{K}_{Z_i^b}), Z^a(\mathbf{K}_{Z_i^b})} \leq \Sigma_{Z^a(\mathbf{K}_{X_i^b}), Z^a(\mathbf{K}_{X_i^b})}, \quad (90)$$

and by the optimality of the control variate relation \mathbf{S} from Corollary 1,

$$\Sigma_{Z^a(\mathbf{K}_{X_i^b}), Z^a(\mathbf{K}_{X_i^b})} \leq \Sigma_{X^a(\mathbf{K}_{X_i^b}), X^a(\mathbf{K}_{X_i^b})}. \quad (91)$$

The second inequality similarly relies on the optimality of the Kalman gain $\mathbf{K}_{X_i^b}$.

Theorem 7 shows that the principal variate covariance is an upper bound on the covariance of the total variate.

Relations (89) in Theorem 7 are valid only when the means of the constituent variates are roughly equivalent. This is especially important in the ‘extended’ and ‘ensemble’ extensions to the MFKF. To achieve this, at each step we apply the following heuristic correction:

$$\mu_{X_i^a} \leftarrow \mu_{Z_i^a}, \quad \mu_{\hat{U}_i^a} \leftarrow \Theta \mu_{Z_i^a}, \quad \mu_{U_i^a} \leftarrow \Theta \mu_{Z_i^a}, \quad (92)$$

which additionally enforces the control variate relation (39), ensures that the principal and total variate means are equivalent, and that the control and ancillary variate means are equivalent.

3.2 Multifidelity Ensemble Kalman Filter

Following Popov et al. (2020), we present the multifidelity ensemble Kalman filter (MFEnKF).

Assume now that instead of manipulating the first two moments of our variates, we manipulate ensembles. Assume that we have N_X pairwise correlated samples of the principal and control variates $(\mathbf{X}^{(i)}, \mathbf{U}^{(i)})$ represented by the ensembles \mathbf{E}_X and $\mathbf{E}_{\hat{U}}$, respectively, and $N_U \geq N_X$ samples of the ancillary variate $\mathbf{U}^{(i)}$, represented by the ensemble \mathbf{E}_U . We wish to construct practical ensemble-based generalizations to the MFKF.

The forecast step, similar to the standard EnKF, and MFKF ((44) and (47)), propagates the ensemble members individually through their respective models,

$$\mathbf{X}_i^{b,(j)} = \mathcal{M}_i^X(\mathbf{X}_{i-1}^{a,(j)}) + \boldsymbol{\xi}_i^{(j)}, \quad 1 \leq j \leq N_X \quad (93)$$

$$\hat{\mathbf{U}}_i^{b,(j)} = \mathcal{M}_i^U(\hat{\mathbf{U}}_{i-1}^{a,(j)}), \quad 1 \leq j \leq N_X \quad (94)$$

$$\mathbf{U}_i^{b,(j)} = \mathcal{M}_i^U(\mathbf{U}_{i-1}^{a,(j)}), \quad 1 \leq j \leq N_U, \quad (95)$$

where each $\boldsymbol{\xi}_i^{(j)}$ is a random sample accounting for the fine fidelity model error. The coarse fidelity model bias is accounted for by the propagation of both the control variate ensemble and ancillary ensemble through the coarse fidelity model.

Assume that the sample means, anomalies, and covariances are readily available for the three constituent ensembles, from which it is possible to derive the empirical estimates of the first two moments of the total variate,

$$\tilde{\boldsymbol{\mu}}_{Z_i} = \tilde{\boldsymbol{\mu}}_{X_i} - \mathbf{S}_i(\tilde{\boldsymbol{\mu}}_{\hat{U}_i} - \tilde{\boldsymbol{\mu}}_{U_i}), \quad (96)$$

$$\tilde{\boldsymbol{\Sigma}}_{Z_i, Z_i} = \tilde{\boldsymbol{\Sigma}}_{X_i, X_i} + \mathbf{S}_i \tilde{\boldsymbol{\Sigma}}_{\hat{U}_i, \hat{U}_i} \mathbf{S}_i^T + \mathbf{S}_i \tilde{\boldsymbol{\Sigma}}_{U_i, U_i} \mathbf{S}_i^T - \mathbf{S}_i \tilde{\boldsymbol{\Sigma}}_{\hat{U}_i, X_i} - \tilde{\boldsymbol{\Sigma}}_{X_i, \hat{U}_i} \mathbf{S}_i^T, \quad (97)$$

where once again we assume that the optimal gain is constant (66) from Theorem 4.

Similar to the standard EnKF, it is not explicitly required to compute the full total background covariance, but merely the related cross-covariances with respect to the observation operator:

$$\tilde{\boldsymbol{\Sigma}}_{Z_i^b, \mathbf{H}_i Z_i^b} = \tilde{\boldsymbol{\Sigma}}_{Z_i^b, Z_i^b} \mathbf{H}_i^T, \quad (98)$$

$$\tilde{\boldsymbol{\Sigma}}_{\mathbf{H}_i Z_i^b, \mathbf{H}_i Z_i^b} = \mathbf{H}_i \tilde{\boldsymbol{\Sigma}}_{Z_i^b, Z_i^b} \mathbf{H}_i^T, \quad (99)$$

which can efficiently be computed by utilizing the observation ensemble anomalies. From this the sample Kalman gain can be computed.

Applying the MFKF formulas (71) to the MFEnKF statistics it is possible to gain access to the corresponding empirical Kalman gain,

$$\tilde{\mathbf{K}}_{Z_i^b} = \tilde{\boldsymbol{\Sigma}}_{Z_i^b, \mathbf{H}_i Z_i^b} \left(\tilde{\boldsymbol{\Sigma}}_{\mathbf{H}_i Z_i^b, \mathbf{H}_i Z_i^b} + \Sigma_{Y_i, Y_i} \right)^{-1}, \quad (100)$$

and the corresponding analysis of the anomalies,

$$\mathbf{A}_{X_i^a} = \mathbf{A}_{X_i^b} - \tilde{\mathbf{K}}_{Z_i^b} (\mathbf{A}_{\mathbf{H}_i} X_i^b - \mathbf{E}_{Y_i^X}), \quad (101)$$

$$\mathbf{A}_{\hat{U}_i^a} = \mathbf{A}_{\hat{U}_i^b} - \Theta \tilde{\mathbf{K}}_{Z_i^b} (\mathbf{A}_{\mathbf{H}_i} \Theta \hat{U}_i^b - \mathbf{E}_{Y_i^{\hat{U}}}), \quad (102)$$

$$\mathbf{A}_{U_i^a} = \mathbf{A}_{U_i^b} - \Theta \tilde{\mathbf{K}}_{Z_i^b} (\mathbf{A}_{\mathbf{H}_i} \Theta U_i^b - \mathbf{E}_{Y_i^U}), \quad (103)$$

where each \mathbf{E}_{Y_i} is an ensemble of perturbed observations. In Popov et al. (2020) it was shown that there is no unique ‘nice’ solution to the problem of perturbed observations in the MFEnKF, thus we will leave this discussion aside in this chapter.

In order to get an ensemble of Z^a , we can look towards the ensemble transform Kalman filter, specifically at the ‘left transform’ variant (Sakov and Bertino 2011). Using known properties of the matrix shift lemma (Asch et al. 2016) and the linearity of the control variate relation (39) one can write the transformation of the ensemble of X^a into anomalies of Z^a given by the ETKF (23) as:

$$\mathbf{E}_{Z^a} = \tilde{\boldsymbol{\mu}}_{Z^a} \mathbf{1}_{N_X}^T + (\mathbf{I}_n - \mathbf{S}\Theta)^{\frac{1}{2}} \mathbf{A}_{X^a}, \quad (104)$$

which can be implemented in any number of computationally efficient ways (Allen et al. 2000) beyond the scope of this chapter.

Note however that this methodology relies on the equality of the covariances of the control and ancillary variates, which especially in the ensemble case will be violated.

We now discuss the main advantage of the MFEnKF: utilizing the same amount of samples of the fine fidelity model as the standard EnKF, the MFEnKF provides a more accurate mean analysis.

Theorem 8 *Assume that we have access to the exact Kalman gains $K_{Z_i^b}$, $K_{X_i^b}$ from (68), of the theoretical Kalman filters. The variance of the empirical mean of the analysis total variate computed with the total variate Kalman gain is less than the variance in the empirical mean in the analysis principal variate computed with the principal variate Kalman gain,*

$$\Sigma_{\tilde{\boldsymbol{\mu}}_{Z_i^a}(\mathbf{K}_{Z_i^b}), \tilde{\boldsymbol{\mu}}_{Z_i^a}(\mathbf{K}_{Z_i^b})} \leq \Sigma_{\tilde{\boldsymbol{\mu}}_{X_i^a}(\mathbf{K}_{X_i^b}), \tilde{\boldsymbol{\mu}}_{X_i^a}(\mathbf{K}_{X_i^b})}. \quad (105)$$

Proof Assuming again that $N_U \geq N_X$, and by Theorem 7,

$$\Sigma_{\tilde{\mu}_{Z_i^a(\mathbf{K}_{Z_i^b})}, \tilde{\mu}_{Z_i^a(\mathbf{K}_{Z_i^b})}} \leq N_X^{-1} \Sigma_{Z_i^a(\mathbf{K}_{Z_i^b}), Z_i^a(\mathbf{K}_{Z_i^b})} \quad (106)$$

$$+ (N_U^{-1} - N_X^{-1}) \mathbf{S} \Sigma_{U_i^a(\mathbf{K}_{Z_i^b}), U_i^a(\mathbf{K}_{Z_i^b})} \mathbf{S}^T, \quad (107)$$

$$\leq N_X^{-1} \Sigma_{X_i^a(\mathbf{K}_{X_i^b}), X_i^a(\mathbf{K}_{X_i^b})}, \quad (108)$$

$$= \Sigma_{\tilde{\mu}_{X_i^a(\mathbf{K}_{X_i^b})}, \tilde{\mu}_{X_i^a(\mathbf{K}_{X_i^b})}}, \quad (109)$$

as required.

The perturbed observations MFEnKF is similar to a typical EnKF algorithm in the way in which inflation and B-localization can be applied. An important salient difference is that there is now an additional hyperparameter, namely the inflation factor for the ancillary ensemble α_U . Inflation for the principal and control ensembles α_X should be the same in order to keep them highly correlated. As optimal inflation is known to depend on the ensemble size (Popov and Sandu 2020), it should generally be the case that $\alpha_U < \alpha_X$.

3.3 Other ‘Multi- x ’ Data Assimilation Algorithms

In this section we discuss other data assimilation algorithms in the ‘multi-’ family that do not, as-of-now, have rigorous multifidelity counterparts.

3.3.1 Particle Filters

In Giles (2008, 2015), Giles discusses ‘multilevel’ Monte-Carlo simulations. The case of projection and interpolation is ignored, and the optimal gain is explicitly set to be identity. The author examines the component variates as being coupled through their differences, which in a two-level control variate framework is equivalent to examining the variates:

$$\begin{cases} U \\ X - \hat{U} \end{cases}, \quad (110)$$

treating each as an independent source of information, with means:

$$\begin{cases} \mu_U \\ \mu_X - \mu_{\hat{U}} \end{cases}. \quad (111)$$

The implicit assumption is that the only important source of information is the mean-estimate, and that μ_U carries negligible uncertainty. The intended use of such algorithms is in small-dimensional cases where large ensembles on coarse fidelity models can be created, thus this is not an unreasonable assumption.

In Gregory et al. (2016), Gregory and Cotter (2017), Gregory and co-authors propose ‘multilevel’ ensemble transform particle filters (ETPF). The authors again employ a linear control variate structure where the optimal gain is assumed to be the identity, and in which all variable operations are performed on the same empirical measures. The authors pay attention to the need for their principal and control variate ensemble to be related, but do not pay attention to the optimality of the couplings. Furthermore the authors utilize a coupling that is optimal for Gaussian random variables; an optimal coupling based on optimal transport could be utilized, while at the same time performing transformations between ensembles through optimal transport techniques.

It is of independent interest to develop more rigorous ‘multifidelity’ generalizations of such algorithms using the couplings outlined in this chapter.

3.3.2 Ensemble Kalman Filters

In Chernov et al. (2017), Hoel et al. (2016), the authors propose a ‘multilevel’ EnKF. The authors extend the empirical measures (110) to spatial relations. In a two-level framework the authors analyze the variables

$$\begin{cases} U \\ X - \Phi \hat{U} \end{cases}, \quad (112)$$

again treating each as an independent source of information, with the means:

$$\begin{cases} \mu_U \\ \mu_X - \Phi \mu_{\hat{U}} \end{cases}, \quad (113)$$

and the signed empirical measure covariance estimates:

$$\begin{cases} \Sigma_{U,U} \\ \Sigma_{X,X} - \Sigma_{\Phi \hat{U}, \Phi \hat{U}} \end{cases}. \quad (114)$$

This covariance estimate is not guaranteed to be semi-positive definite. Additionally, no attention is paid to utilize an optimal gain linear control variate structure, and the enforcement the principal-control variate relation.

4 Multifidelity Observations

We now discuss an optimal way in which to combine observations from different sources at roughly similar physical locations of the same phenomenon. In operational literature this is commonly dubbed ‘super-observations’ (Cummings 2005; Oke et al. 2008), though such formulations are largely heuristic in that they take naive averages of interpolations of similar observations. The chief reason why observations are combined instead of used separately is to reduce the observation space dimension, making similar information represented in a denser format.

Assume that the true state is \mathbf{X}^t , and recall the multifidelity observation definition (6) where the observations Y^x and Y^v have the observation errors η^x , and η^v that are assumed to be unbiased and independent. We make the additional assumption that the fine fidelity and coarse fidelity observation operators are deterministically related by the coupling

$$\mathcal{H}^v(\mathbf{x}) = \theta(\mathcal{H}^x(\mathbf{x})), \quad (115)$$

similar to the state relation assumed in (16).

The truth in observation space is assumed to be the expected value of the observation for each fine and coarse observation. This can be alternatively reformulated as the truth in observation space is distributed according to a distribution with mean \mathbf{Y}^x and \mathbf{Y}^v for the fine and coarse observations respectively.

A canonical way of dealing with such a scenario is by ‘stacking’ the observations, and creating the observation operator

$$\mathcal{H}(\mathbf{X}) = \begin{bmatrix} \mathcal{H}^x(\mathbf{X}) \\ \mathcal{H}^v(\mathbf{X}) \end{bmatrix}. \quad (116)$$

We will not pursue this approach, as it increases the dimensionality of the observations without increasing the information content.

Under the linear control variate approach the total variate observation mean is defined to be:

$$\boldsymbol{\mu}_{Y^\zeta} = \boldsymbol{\mu}_{Y^x} - \mathbf{S}(\boldsymbol{\mu}_{Y^x} - \boldsymbol{\mu}_{Y^v}), \quad (117)$$

where one implicitly assumes that $\mathbb{E}[\theta(Y^x)] = \mathbb{E}[Y^v]$. The optimal gain is,

$$\mathbf{S} = \Sigma_{Y^x, \theta(Y^x)} (\Sigma_{\theta(Y^x), \theta(Y^x)} + \Sigma_{Y^v, Y^v})^{-1}, \quad (118)$$

with the new covariance of the total observation given by

$$\Sigma_{Y^\zeta, Y^\zeta} = \Sigma_{Y^x, Y^x} - \Sigma_{Y^x, \theta(Y^x)} (\Sigma_{\theta(Y^x), \theta(Y^x)} + \Sigma_{Y^v, Y^v})^{-1} \Sigma_{\theta(Y^x), Y^x}. \quad (119)$$

Evaluation of this formula, however, requires knowledge of both $\Sigma_{Y^x, \theta(Y^x)}$ and $\Sigma_{\theta(Y^x), \theta(Y^x)}$, which might not be readily available.

An alternate approach is to utilize the importance sampling framework. Assume we have an ensemble of perturbed observations, $\mathbf{E}_{Y^\chi} = [\mathbf{Y}^{\chi,(1)}, \dots, \mathbf{Y}^{\chi,(M)}]$, representing M independent samples from the assumed distribution of the fine fidelity observation π_{Y^χ} . Apply the importance sampling procedure to generate the weights:

$$[\mathbf{w}]_i \propto \pi_{Y^v}(\theta(\mathbf{Y}^{\chi,(i)})). \quad (120)$$

The unbiased mean and covariance estimates of the total observation are given by

$$\tilde{\boldsymbol{\mu}}_{Y^\zeta} = \mathbf{E}_{Y^\chi} \mathbf{w}, \quad \tilde{\boldsymbol{\Sigma}}_{Y^\zeta, Y^\zeta} = \frac{M}{M-1} \mathbf{E}_{Y^\chi} (\text{diag}(\mathbf{w}) - \mathbf{w}\mathbf{w}^T) \mathbf{E}_{Y^\chi}^T. \quad (121)$$

Alternatively, an ensemble of equally weighted perturbed observations to be used with a perturbed observations EnKF can be derived by the optimal transport framework,

$$\mathbf{E}_{Y^\zeta} = \mathbf{E}_{Y^\chi} \mathbf{T}^*, \quad (122)$$

given by (27).

As many of these methods rely on empirical estimates of the total observation covariance matrix, methods such as localization can trivially be applied, especially since in most operational algorithms for physical systems the observation covariance is typically assumed to be diagonal.

5 Numerical Experiments

For the sake of completeness we provide a simple twin experiment on a simple dynamical system to test a two-fidelity MFEnKF.

For the fine fidelity model we use the 40-variable Lorenz '96 system (Lorenz 1996), posed as an ODE:

$$\mathbf{x}'_i = f_\chi(\mathbf{x}) = -\mathbf{x}_{i-1}(\mathbf{x}_{i-2} - \mathbf{x}_{i+1}) - \mathbf{x}_i + 8, \quad 1 \leq i \leq 40, \quad (123)$$

where $\mathbf{x}_0 := \mathbf{x}_{40}$, $\mathbf{x}_{-1} := \mathbf{x}_{39}$ and $\mathbf{x}_{41} := \mathbf{x}_1$.

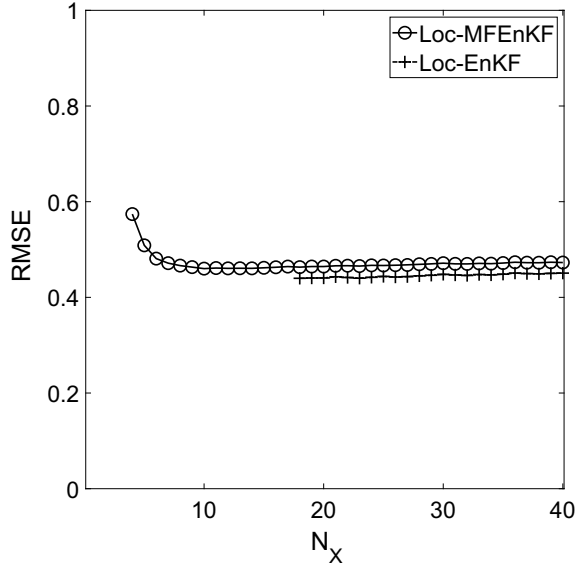
We use the method of snapshots (Sirovich 1987) to construct linear projection and interpolation operators, Φ and Θ , utilizing 20000 snapshots over an expressive time interval of 1000 units.

For the coarse fidelity we consider a reduced order model built using a naive approach, where we evaluate the derivative in the full space and then project onto the reduced space:

$$\mathbf{u}' = f_U(\mathbf{u}) = \Theta f_\chi(\Phi \mathbf{u}). \quad (124)$$

For the Lorenz '96 system this can be written equivalently as a multivariate quadratic equation.

Fig. 2 Comparison of the analysis empirical mean RMSE of a localized perturbed observations MFEnKF with a localized perturbed observations EnKF, for various fine fidelity (full order) model ensemble sizes



The Lorenz '96 system is known to have a Kaplan-Yorke dimension of 27.1 (Popov and Sandu 2019). For this reason we take $r = 28$ reduced modes to describe the whole system (though this is only possible non-linearly). In the reduced model, this represents about 90% of the total energy of the system, as represented by the ratio of the captured eigenvalues to the total eigenvalues. In this context it is actually relatively difficult to build a reduced order model for the Lorenz '96 system.

We compare the algorithm to the standard perturbed observations ensemble Kalman filter. Both algorithms will use forecast anomaly inflation and Gaspari-Cohn covariance localization (Gaspari and Cohn 1999).

We observe every other variable every $\Delta t = 0.05$ time units, with a Gaussian error (2) of $\Sigma_{Y,Y} = \mathbf{I}_{20}$.

We perform localization and inflation as follows. For forecast anomaly inflation for the full system we will take $\alpha_X = 1.1$ and for the coarse system $\alpha_U = 1.00$ as the reduced order model is less stable than the full order model, thus not requiring inflation. To retain an undersampled ensemble for the ancillary variate, we choose an ensemble size of $N_U = 25$. The inner parameter of the localization function is selected to match that of a Gaussian kernel, and set the radius to be equal to 4 (Petrie and Dance 2010).

Figure 2 shows the relationship between the principal variate ensemble size and the spatio-temporal RMSE of the empirical analysis mean of the MFEnKF and the EnKF. As can be seen, the problem is comparatively difficult for the EnKF, as it requires at least 18 fine fidelity ensemble members for a stable behavior. The same RMSE can be achieved with less than 10 fine fidelity ensemble members in the MFEnKF

framework. Assuming the coarse fidelity model runs are significantly cheaper (not true in this trivial contrived example) then the MFEnKF is clearly superior.

We note that there is some loss of accuracy in the results, due in part to several assumptions that are violated. One is that the orthogonal complement space is uncorrelated with that of the full space (65). As we are capturing 90% of the energy of the system, the rest of the energy is not that negligible, and is no doubt highly correlated with the what is captured. Methods to diminish the influence of this error, would be needed for operational systems.

6 Discussion

Multifidelity data assimilation, and multifidelity inference in general, seek to leverage the availability of information about reality at multiple resolution levels. The field is still in its infancy, but the multifidelity methods are highly promising. This chapter provides a general philosophical and theoretical framework for the development of such methods. New multifidelity data assimilation approaches should utilize efficient coarse fidelity models to speed up high fidelity inference. The new methods should be grounded in sound statistical and probabilistic theory.

In this chapter we focus on the multifidelity stochastic EnKF. Variational multifidelity approaches have been developed in Stefanescu et al. (2015). Square root multifidelity Kalman filters, analogues to the perturbed observations MFEnKF, must be developed in the future. Particle filters that are appropriate for non-Gaussian probability densities, or even hybrid EnKF-PF systems where different variates are assimilated with different algorithms, might provide an avenue for development of multifidelity particle filtering. Multifidelity hybrid data assimilation, that combines multifidelity EnKF and multifidelity variational methods, are also a promising future venue. Finally, the construction of a hierarchy of coarser models to support data assimilation should be carefully investigated. For example, methods based on machine learning (e.g., as discussed in Moosavi et al. 2018a, b, or non-linear projections using autoencoders) are of considerable interest.

Acknowledgements The authors would like to acknowledge Traian Iliescu and Changhong Mou who have helped make some of the work underlying this possible. This work was supported by awards NSF CCF-1613905, NSF ACI-1709727, NSF CDS&E-MSS-1953113, and by the Computational Science Laboratory at Virginia Tech.

References

- Acevedo W, de Wiljes J, Reich S (2017) Second-order accurate ensemble transform particle filters. *SIAM J Sci Comput* 39(5):A1834–A1850
- Allen E, Baglama J, Boyd S (2000) Numerical approximation of the product of the square root of a matrix with a vector. *Linear Algebra Appl* 310(1–3):167–181

- Asch M, Bocquet M, Nodet M (2016) Data assimilation: methods, algorithms, and applications. SIAM
- Bishop C, Etherton B, Majumdar S (2001) Adaptive sampling with the ensemble transform Kalman filter. Part I: theoretical aspects. *Mon Weather Rev* 129:420–436
- Brunton SL, Kutz JN (2019) Data-driven science and engineering: machine learning, dynamical systems, and control. Cambridge University Press
- Chernov A, Hoel H, Law K, Nobile F, Tempone R (2017) Multilevel ensemble Kalman filtering for spatio-temporal processes. MATHICSE technical report 22.2017, EPFL
- Cummings JA (2005) Operational multivariate ocean data assimilation. *Q J R Meteorol Soc* 131(613):3583–3604
- Gaspari G, Cohn S (1999) Construction of correlation functions in two and three dimensions. *Q J R Meteorol Soc* 125:723–757
- Giles MB (2008) Multilevel Monte Carlo path simulation. *Oper Res* 56(3):607–617
- Giles MB (2015) Multilevel Monte Carlo methods. *Acta Numer* 24:259–328
- Gregory A, Cotter CJ (2017) A seamless multilevel ensemble transform particle filter. *SIAM J Sci Comput* 39(6):A2684–A2701
- Gregory A, Cotter CJ, Reich S (2016) Multilevel ensemble transform particle filtering. *SIAM J Sci Comput* 38(3):A1317–A1338
- Hoel H, Law KJH, Tempone R (2016) Multilevel ensemble Kalman filtering. *SIAM J Numer Anal* 54(3). <https://doi.org/10.1137/15M100955X>
- Houtekamer P, Mitchell H (1998) Data assimilation using an ensemble Kalman filter technique. *Mon Weather Rev* 126:796–811
- Jaynes ET (2003) Probability theory: the logic of science. Cambridge University Press
- Kalman R (1960) A new approach to linear filtering and prediction problems. *Trans ASME J Basic Eng* 82:35–45
- Lorenz EN (1996) Predictability: a problem partly solved. In: Proceedings of seminar on predictability, vol 1
- Moosavi A, Stefanescu R, Sandu A (2018a) Efficient construction of local parametric reduced order models using machine learning techniques. *Int J Numer Methods Eng* 113(3):512–533
- Moosavi A, Stefanescu R, Sandu A (2018b) Parametric domain decomposition for accurate reduced order models: applications of MP-LROM methodology. *J Comput Appl Math* 340:629–644
- Nelson BL (1987) On control variate estimators. *Comput Oper Res* 14(3):219–225
- Oke PR, Brassington GB, Griffin DA, Schiller A (2008) The Bluelink ocean data assimilation system (BODAS). *Ocean Model* 21(1–2):46–70
- Peherstorfer B, Willcox K, Gunzburger M (2018) Survey of multifidelity methods in uncertainty propagation, inference, and optimization. *SIAM Rev* 60(3):550–591
- Petrie RE, Dance S (2010) Ensemble-based data assimilation and the localisation problem. *Weather* 65(3):65–69
- Popov AA, Mou C, Iiescu T, Sandu A (2020) A multifidelity ensemble Kalman filter with reduced order control variates. [arXiv:2007.00793](https://arxiv.org/abs/2007.00793)
- Popov AA, Sandu A (2019) A Bayesian approach to multivariate adaptive localization in ensemble-based data assimilation with time-dependent extensions. *Nonlinear Process Geophys* 26(2):109–122
- Popov AA, Sandu A (2020) An explicit probabilistic derivation of inflation in a scalar ensemble Kalman filter for finite step, finite ensemble convergence. [arXiv:2003.13162](https://arxiv.org/abs/2003.13162)
- Reich S (2013) A nonparametric ensemble transform method for Bayesian inference. *SIAM J Sci Comput* 35(4):A2013–A2024
- Reich S, Cotter C (2015) Probabilistic forecasting and Bayesian data assimilation. Cambridge University Press
- Rubinstein RY, Marcus R (1985) Efficiency of multivariate control variates in Monte Carlo simulation. *Oper Res* 33(3):661–677

- Sakov P, Bertino L (2011) Relation between two common localisation methods for the EnKF. *Comput Geosci* 15(2):225–237
- Sirovich L(1987) Turbulence and the dynamics of coherent structures. I. Coherent structures. *Q Appl Math* 45(3), 561–571
- Stefanescu R, Sandu A, Navon I (2015) POD/DEIM strategies for reduced data assimilation systems. *J Comput Phys* 295:569–595