Notes on: Data Assimilation with Model Error from Unresolved Scales

John Harlim*
Department of Mathematics and Department of Meteorology
The Pennsylvania State University

January 31, 2014

Abstract

In this note, we provide an overview of two important competing data assimilation schemes that were developed in the past two-decade, discuss the current methods that are operationally used in weather forecasting applications, and point out one major challenge in data assimilation community: utilize these existing schemes in the presence of model error from unresolved scales. The main goal of this note is to discuss a simple mathematical analysis to understand why this problem is difficult and to provide a mathematically justifiable stochastic parameterization to mitigate this type of model error in practical multiscale filtering problems. This is a prototypical situation in many applications due to limited ability in resolving the smaller scale processes as well as the difficulty to model the interaction across scales. We will present simple examples to point out the importance of accounting for model error when the separation of scales are not apparent. These examples also elucidate the necessity of treating model error as a stochastic process in a nontrivial fashion for optimal filtering. In general, it is difficult to guess the appropriate stochastic models to represent model error and/or even to estimate higher order statistics. For practical consideration, we only discuss “optimality” in the sense that the first two moments are as accurate as those of the true filter. Although this is just a necessary (and not sufficient) condition for optimal nonlinear filtering, we shall see that many of the existing approaches do not even meet this minimum criteria. Several classical approaches to estimate model error statistics are briefly reviewed and reassessed, and several recent developments are overviewed, including important open problems/questions. We hope that this review note can make this important issue accessible to inspire more applied mathematicians to work on this exciting open problem. More importantly, we hope that the theoretical results reviewed

*The bibliography is far from complete. If one is interested to discuss this research topic, please don’t hesitate to contact me at: jharlim@psu.edu.
in this note can help the future design of more robust methods for coping with model error.

1 Introduction

Given noisy observations from nature, filtering (or data assimilation) is a numerical scheme to find the best statistical estimate of the true signal and unknown parameters. Bayesian filtering consists of a two-step predictor-corrector scheme that adjusts prior forecast (background) statistical estimates from a predictor (or dynamical) model to be more consistent with the current observations; this correction step is referred to as analysis in the atmospheric and ocean science (AOS) community. Subsequently, the posterior (revised or analysis) statistical estimates are fed into the model as initial conditions for future time prior statistical estimates.

In the past two-decade, many practical data assimilation approaches were developed to reduce the computational cost in the analysis step and improve the statistical prediction skill. In the AOS data assimilation community, two important schemes are: (i) ensemble-based methods [26, 48, 1, 92, 2, 12, 27, 49, 87] which rely on empirical statistical estimates from ensemble forecasts and a Kalman-based formulation; (ii) variational-based methods [86, 63, 21, 64] that rely on linear tangent and adjoint models. Operationally, most of the weather prediction centers, including the European Center for Medium-range Weather Forecasts (ECMWF), the UK Met Office, and the National Centers for Environmental Prediction (NCEP), are adopting hybrid approaches, taking advantages from both the ensemble and variational based methods [51, 16, 15, 20, 90].

Mathematically, convergence of these methods are not well understood in filtering complex dynamical systems despite the fact that they are being used with great successes in real applications, assimilating high-dimensional, Global Circulation Models for the atmospheric and ocean dynamics with nearly $10^9$ state variables (depending on the model resolution) with abundant data collected from radiosonde, scatterometer, satellite, and radar measurements. Few papers that rigorously analyze the convergence of these schemes in idealistic settings include [39] for ensemble Kalman filter and [59, 19, 13] for variational based methods.

The estimation from these schemes is mostly accurate in the midlatitude atmospheric region, where the dynamics are nearly geostrophic. This is a physical balance between the pressure gradient and Coriolis forces that occur in the midlatitude atmosphere due to earth rotation corresponding to a single length scale, the Rossby deformation radius. In the Tropics, this physical balance does not exist since the Coriolis force vanishes at the equator and the dynamics is dominated by vertical heating/cooling in response to diabatic heating caused primarily by latent heat release. Despite some improvement in tropical weather forecasting [8], the forecast error for the zonal (east-west direction) wind component remains the largest in the Tropics (e.g., see Fig 1 in [96]). On the other hand, recent observations (see [97] and the references therein) suggest that an improvement of state estimation in the Tropics extend the global weather prediction beyond two weeks. The difficulty in predicting the Tropics is primarily caused by limited representation of the tropical convection and its multiscale organization in the contemporary convection parameterization [79, 8]. This is an example of “intrinsic
information barrier” that prevents one from capturing the large-scale phenomena with a coarse model, as pointed out in [67].

Given the practical issues above as examples, an important challenge in data assimilation is to intelligently utilize the existing methods (ensemble, variational, and any hybrid based approaches) in the presence of model errors, which are unavoidable when the physics of the underlying dynamics are not completely understood, when the model parameters are not specified correctly, or when the model is under-resolved. The goal of this note is to make this important issue accessible and we hope to inspire more applied mathematicians to work on this exciting open problem.

This note is organized as follows: In Section 2, we review practical filtering methods. Subsequently, we discuss the filtering problem in the presence of model error in Section 3. Here, we show two simple examples that elucidate the importance of accounting for model error in state estimation of multiscale dynamical systems with moderate scale gap. We will review the mathematical theory developed in [10] for “optimal filtering” with this form of model error. In Section 4, we review and classify some standard approaches for filtering with model errors. We also discuss a simple approach that takes into account the theoretical insight discussed in Section 2 to filter simulated turbulent signals with intermittent instabilities. In Section 5, we conclude the note by proposing a practical approach based on the theory discussed in Section 3 and discuss some remaining open problems.

2 Data assimilation: Linear and Gaussian methods

Here we review practical Gaussian and linear based data assimilation methods that are being used operationally. None of these methods yield optimal nonlinear filtering in the sense of minimum variances.

Consider the following discrete-time filtering problem,

\[
\begin{align*}
\frac{du}{dt} &= f(u, t), \\
v_m &= g(u_m) + \sigma_m^o,
\end{align*}
\]

where the underlying truth signal, \(u(t)\), solved a general dynamical system (1) given initial condition, \(u(t_0)\). In (2), \(v_m\) denotes the observations that are measured at discrete-time, corrupted by noise \(\sigma_m^o\). Here, \(g\) denotes the observation operator that maps the true solution in \(\mathbb{R}^N\) to the observation space in \(\mathbb{R}^M\). In (2), \(\sigma_m^o\) denotes the measurement error which can follow any kind of distribution in general. In practical numerical weather prediction application, almost every method is developed under the assumption that \(\sigma_m^o\) are i.i.d., Gaussian noises with mean zero and variance \(R\). The data assimilation schemes (i) and (ii), mentioned in Section 1, are developed to solve the Bayesian formula,

\[
p(u_m|v_m) \propto p(u_m)p(v_m|u_m),
\]

to obtain the first two moments of the posterior distribution \(p(u_m|v_m)\) at time \(t_m\). Here, \(p(u_m)\) denotes distribution of a prior (or background) estimate \(\bar{u}_m^b\) at time \(t_m\).
If we assume that the prior estimate error is Gaussian (which is not necessary in general), unbiased, and uncorrelated with the observation error, then we can write

\[ p(u_m) \propto \exp \left( -\frac{1}{2} (u_m - \bar{u}_b)^\top (P_b)^{-1} (u_m - \bar{u}_b) \right) \equiv \exp \left( -\frac{1}{2} J_b(u_m) \right), \] (4)

where \( P_b = \mathbb{E}[(u_m - \bar{u}_b)(u_m - \bar{u}_b)^\top] \) denotes the prior error covariance matrix at time \( t_m \), which characterizes the error of the mean estimates, \( \bar{u}_b \). In (3), \( p(v_m|u_m) \) denotes the observation likelihood function associated with the observation model in (2), that is,

\[ p(v_m|u_m) \propto \exp \left( -\frac{1}{2} (v_m - g(u_m))^\top R^{-1} (v_m - g(u_m)) \right) \equiv \exp \left( -\frac{1}{2} J_o(u_m) \right). \]

There are various methods that assume non-Gaussian, exponential family for either the prior and the likelihood functions such as in [30, 31, 42, 29]. The posterior (or analysis) mean and covariance estimates, \( \bar{u}_a \) and \( P_a \), are obtained by maximizing the posterior density in (3), which is equivalent to solving the following optimization problem,

\[ \min_{u_m} J_b(u_m) + J_o(u_m), \] (5)

for \( u_m \) that solves (1). These posterior statistics are fed into the model in (1) to estimate the prior statistical estimates at the next time step \( t_{m+1} \), \( \bar{u}_b \) and \( P_b \), when observations become available.

If the dynamical model in (1) and the observation operator \( g \) are linear, then the unbiased posterior mean and covariance estimates are given by the Kalman filter solutions [52]. In general, the nonlinear minimization problem in (5) is nontrivial when the state vector \( u_m \in \mathbb{R}^N \) is high-dimensional; the major difficulty is in obtaining accurate prior statistical estimates \( \bar{u}_b \) and \( P_b \). The ensemble Kalman filter empirically approximates these prior statistical solutions with an ensemble of solutions and uses the Kalman filter formula to obtain the posterior statistics, implicitly assuming that these ensemble based prior statistics are Gaussian. Alternatively, the variational approach solves this minimization problem [21, 64], often assuming that matrix \( P_b = B \) is time independent. In practice, the variational approach that is used minimizes,

\[ \min_{u_{m0}} J_b(u_{m0}) + \sum_{j=0}^T J_o(u_{m_j}), \] (6)

for initial condition \( u_{m0} \), accounting for observations at times \( \{t_{m_j}, j = 0, \ldots, T\} \) and constraining \( u_{m_j} \) to satisfy the model in (1). This method (also known as the strong constrained 4D-VAR) is typically solved with an incremental approach that relies on linear tangent and adjoint models [21] and it is sensitive to the choice of \( B \) [88]. To alleviate this sensitivity issue, many operational centers such as the ECMWF, UKMet Office, and NCEP are adopting hybrid methods [51, 16, 15, 20, 90] that use ensemble of solutions to estimate \( P_b \) in each minimization step.
3 Model error

Let \( \tilde{u}(t) \) be solutions of an imperfect model and \( u(t) \) be solutions of the perfect model in (1), given initial condition, \( \tilde{u}(t_0) = u(t_0) \). For simplicity, assume that both solutions are random variables in the same probability space, such that all of the expectation operators, \( \mathbb{E}(\cdot) \), below are defined with respect to the same probability measure. Define model error as follows,

\[
e(t) \equiv u(t) - \tilde{u}(t),
\]

where \( e(t) \) is a random variable with mean \( \bar{b}(t) = \mathbb{E}(e(t)) \) and covariance \( Q^b(t) = \text{Cov}(e(t)) \). In this section, we discuss only the estimation of the first two moments which is readily difficult to estimate in practice. Ideally, we want to estimate all higher-order moments of the stochastic process, \( e(t) \), but it is nontrivial (hopeless) since finding these moments require solving an infinite dimensional PDE for general nonlinear problems.

Define \( \tilde{u}^b_m \equiv \tilde{u}(t_m) \equiv \mathbb{E}(u(t_m)) \) as the prior mean estimate from the perfect model. Similarly, define also \( \tilde{u}^b_m \equiv \tilde{u}(t_m) \equiv \mathbb{E}(\tilde{u}(t_m)) \), as the prior mean estimate from the imperfect model. Assume that these estimates are initiated with the same initial conditions, \( u_{m-1} \) at previous time step \( t_{m-1} \). One can show that the mean model error,

\[
\tilde{b}_m \equiv \tilde{b}(t_m) = \tilde{u}(t_m) - \tilde{u}(t_m) = \tilde{u}^b_m - \tilde{u}^b_m,
\]

is equivalent to the “bias forecast error”, defined in [24]. From (7) and (8),

\[
u_m - \bar{b}_m^b = (\tilde{u}_m + \bar{b}_m) - (\bar{b}_m + \bar{b}_m^b) = (\tilde{u}_m - \bar{u}_m^b) + (\bar{b}_m - \bar{b}_m),
\]

and therefore, we can deduce,

\[
\mathbb{E}[(\mathbf{u}_m - \bar{u}_m^b)(\mathbf{u}_m - \bar{u}_m^b)^\top] = \mathbb{E}[(\tilde{u}_m - \bar{u}_m^b)(\tilde{u}_m - \bar{u}_m^b)^\top]
+ \mathbb{E}[(\mathbf{u}_m - \bar{u}_m^b)(\bar{b}_m - \bar{b}_m)^\top]
+ \mathbb{E}[(\bar{b}_m - \bar{b}_m^b)(\bar{u}_m^b - \tilde{u}_m^b)^\top]
+ \mathbb{E}[(\bar{b}_m - \bar{b}_m^b)(\bar{b}_m - \bar{b}_m)^\top].
\]

Notice that by the definition of the prior distribution in (4),

\[
\mathbb{E}[(\mathbf{u}_m - \bar{u}_m^b)(\mathbf{u}_m - \bar{u}_m^b)^\top] = \mathbb{E}[(\mathbf{u}_m - \mathbb{E}(\mathbf{u}_m))(\mathbf{u}_m - \mathbb{E}(\mathbf{u}_m))^\top] = P^b_m.
\]

Let us define \( P^b_m \equiv \mathbb{E}[(\mathbf{u}_m - \bar{u}_m^b)(\bar{u}_m^b - \tilde{u}_m^b)^\top] \) to be the error covariance of the estimate from the imperfect model. Then, the expression in (9) becomes

\[
P^b_m = \tilde{P}^b_m + Q^b_{nb} + (Q^b_{nb})^\top + Q^b_m,
\]

where \( Q^b_{nb} \) denotes the cross covariances between the forecasts from imperfect model, \( \tilde{u}_m^b \), and the model error estimator, \( \bar{b}_m \). Equations (8) and (10) suggest that “optimal” filtered solutions can only be attained when the mean model errors, \( \bar{b}_m \), are accounted for in the prior mean estimates and the prior error covariances, \( \tilde{P}^b_m \), are appropriately adjusted by inflation factors \( Q^b_{nb} + (Q^b_{nb})^\top + Q^b_m \). By optimal here, we imply the mean
and covariance filter estimates are comparable to those of the true filtered solutions, obtained from the perfect prior filter model.

For Kalman filter based assimilation methods, one can simply apply the standard Kalman filter formula to the adjusted prior mean and covariance, accounting for the model error statistics if they are available [24]. For the 4D-VAR implementation, one can add new constraints associated with the model error terms to the minimization problem in (6). In [88], he proposed to solve the following weak constrained 4D-VAR minimization problem for \( \{u_{mj}, j = 0, \ldots, T\} \), accounting for the model error with an additional term shown below,

\[
\min_{u_{mj}} J^b(u_{m0}) + \sum_{j=0}^{T} J^o(u_{mj}) + \ldots
\]

\[
\sum_{j=1}^{T} (u_{mj} - \varphi_{mj}(u_{mj-1}) - \bar{b}_{mj})^\top (Q_{mj}^b)^{-1} (u_{mj} - \varphi_{mj}(u_{mj-1}) - \bar{b}_{mj}),
\]

assuming that \( Q^{bu} = 0 \) and the model error \( e(t) \) is Gaussian. In (11), we define \( \varphi_{mj}(u_{mj-1}) \equiv \tilde{u}_{mj} \) as the solutions of the imperfect model, given initial condition \( u_{mj-1} \). This is obviously a more expensive optimization problem compared to the strong constrained 4D-VAR in (6), even when \( \bar{b}_{mj} \) and \( Q_{mj} \) are known. Various pragmatic approximations were suggested for estimating the solutions of this minimization problem [88]. Recent approaches to reduce the computational costs for solving the weak constrained 4D-VAR minimization problem in (11) employ a time parallelization through a saddle-point formulation [28].

Recently, various reduced filter models were proposed for filtering complex turbulence systems [69]. Some of these reduced filtering methods were designed to handle observations that involve the small-scale variables [45]. A non-trivial numerical study on filtering nonlinear slow-fast test systems was proposed in [37, 38]. An important approach to extract information from sparse observations of “compressed” (or non-separable) multiscale processes was called “stochastic superresolution” [56, 18]. While these approaches are very accurate and numerically cheap for estimating the mean, they underestimate the covariance statistics.

In the examples below, we discuss two simple dynamical models with only observations of the large-scale variables. The first one is the linear example studied in [40, 10] and the second one is a nonlinear problem introduced in [36, 35]. With these simple examples, we hope to: (1) Elucidate the importance of accounting for model error mean and covariance statistics for accurate filtering of multiscale dynamical systems with moderate scale gap in practice. Note that this is simply a weak condition for any filtering method to meet, indeed satisfying this condition does not guarantee optimal nonlinear filtering; (2) Provide theoretical insight for optimal filtering in the sense that the mean and covariance estimates are as accurate as the true filter; (3) Demonstrate that model errors can be mitigated with appropriate stochastic parameterization in the prior filter model.

Example 1: Consider filtering a partially observed two-scale linear system of stochas-
tic differential equations \[40\],

\[
\begin{align*}
\frac{dx}{dt} &= (a_{11} x + a_{12} y) dt + \sigma_x dW_x, \\
\frac{dy}{dt} &= \frac{1}{\epsilon}(a_{21} x + a_{22} y) dt + \frac{\sigma_y}{\sqrt{\epsilon}} dW_y.
\end{align*}
\]

(12)

(13)

Here, \( W_x, W_y \) are independent Wiener processes, the parameter \( \epsilon \) characterizes the
time scale gap between the variables \( x \in \mathbb{R} \) and \( y \in \mathbb{R} \). We assume throughout that
\( \sigma_x, \sigma_y \neq 0 \) and that the eigenvalues of the matrix,

\[
A = \begin{pmatrix} a_{11} & a_{12} \\ \frac{1}{\epsilon} a_{21} & \frac{1}{\epsilon} a_{22} \end{pmatrix},
\]

are strictly negative, to assure the existence of a unique joint invariant density \( \rho_\infty(x, y) \).
Furthermore we require \( \tilde{a} = a_{11} - a_{12} a_{22}^{-1} a_{21} < 0 \) to assure that the leading order slow
dynamics,

\[
\frac{d\tilde{x}}{dt} = \tilde{a} \tilde{x} dt + \sigma_x dW_x,
\]

(14)
supports an invariant density. It is well known that solutions of the one-dimensional
SDE in (14) converge to solutions, \( x^\epsilon(t) \), of (12) pathwise up to finite time, assuming
\( \epsilon \to 0 \). The convergence rate is on the order of \( \epsilon \) (see e.g.,[82] for detail).

**Reduced Stochastic Filter (RSF):** Consider (14) as the prior model to assimilate
noisy observations,

\[
z_m = x(t_m) + \epsilon_m^o, \quad \epsilon_m^o \sim \mathcal{N}(0, R), \tag{15}
\]
of the slow variable \( x \) at discrete time step \( t_m \) with constant observation time interval
\( \Delta t = t_{m+1} - t_m \). Since this example is linear, the optimal solutions can be obtained by
the Kalman filter formula, in the sense that the solutions minimize the posterior error
covariance [52]. In discrete form, the prior mean and error covariance estimates [34, 69]
are given by

\[
\begin{align*}
\tilde{x}_m^b &= e^{\tilde{a} \Delta t} \tilde{x}_{m-1}^a, \\
\tilde{P}_m^b &= e^{2\tilde{a} \Delta t} \tilde{P}_{m-1}^a + \frac{\sigma_x^2}{2\tilde{a}}(1 - e^{2\tilde{a} \Delta t}).
\end{align*}
\]

The posterior mean and covariance update are given by,

\[
\begin{align*}
\tilde{x}_m^a &= \tilde{x}_m^b + K_m(z_m - \tilde{x}_m^b), \\
\tilde{P}_m^a &= (1 - K_m)\tilde{P}_m^b, \\
K_m &= \tilde{P}_m^b (\tilde{P}_m^b + R)^{-1}.
\end{align*}
\]

We will refer to this filtering scheme as the reduced stochastic filter (RSF) as in [40].
It has been shown that the posterior filtered estimates of such a reduced stochastic
filter converge to the true filtered solutions, with a convergence rate of \( \sqrt{\epsilon} \) for general
nonlinear filtering problems, see [50].
Now we discuss results from a numerical simulation with $a_{11} = a_{21} = a_{22} = -1, a_{12} = 1$, $\sigma_x^2 = \sigma_y^2 = 2$, $\Delta t = 1$, and $R = 50\% Var(x)$ and compare them with the true filtered solutions, obtained with the perfect prior model in (12)-(13). In Figure 1, we show the filter accuracy (left panel), quantified by the Mean-Square-Error (MSE) between the posterior mean state estimate, $\bar{x}_m$, and the truth, $x_m$, and the asymptotic error covariance estimate (right panel) of the posterior mean estimate, $\bar{x}_m$, as functions of scale gap $\epsilon$. Note that the asymptotic posterior error covariance estimate is constant for this linear problem after $m = 10,000$ iterations. Notice also that when $\epsilon \ll 1$ is small ($x$ is much slower than $y$), the MSE are almost identical to those of the true filter. For moderate scale gap with larger $\epsilon$, notice that the filter accuracy degrades (with higher MSE) and the true prior error covariance $P_m^b$ is significantly underestimated (see the solid line with circles in Figure 1).
RSF with an additive noise correction (RSFA): Let’s rewrite the fast equation in (13) as follows,

$$y dt = \frac{a_{21}}{a_{22}} x dt - \sqrt{\epsilon \sigma_y} \frac{a_{12}}{a_{22}} dW_y + \mathcal{O}(\epsilon).$$  \hspace{1cm} (16)$$

Substitute this expression into the slow equation in (12), we obtain:

$$d\hat{x} = \tilde{\epsilon} x dt + \sigma_x dW_x - \sqrt{\epsilon \sigma_y} \frac{a_{12}}{a_{22}} dW_y.$$

One can check [40] for a more concise formal asymptotic expansion; therein, they also showed that solutions of (17) converge pathwise to solutions, $x^\epsilon(t)$, of (12) up to finite time, with convergence rate of order $\epsilon^2$. We will refer to the filtering strategy with the prior model in (17) as the reduced stochastic filter with an additive noise correction (RSFA), following the notation in [40]. The additional additive noise correction in (17) essentially inflates the prior covariance estimates in each filtering step, so RSFA is an analog of an additive covariance inflation method [81, 93].

Mathematically, this reduced filter model is equivalent to defining the following stochastic process as an estimator for model error,

$$d\hat{e} = \tilde{\epsilon} \hat{e} dt - \sqrt{\epsilon \sigma_y} \frac{a_{12}}{a_{22}} dW_y.$$

where $\hat{e} \equiv \hat{x} - \tilde{x}$; here, $\tilde{x}$ solves (14) and $\hat{x}$ solves (17).

Our numerical simulations suggest that while the filter accuracy is improved (notice in Figure 1 that the MSE are almost identical to those of the true filter), the true posterior error covariances, $P^b_m$, are still underestimated.

Optimal Reduced Stochastic Filter: In [10], they analyzed the linear filtering problem in (12)-(13) for a continuous-time observation model,

$$dz = x(t) dt + \sqrt{R} dV,$$

where $V$ denotes standard a Wiener process.* In this setting, they rigorously proved that there exists a unique choice of estimator of model error, $e \equiv x - \tilde{x}$, such that the filtered solutions are optimal in the sense that both the mean and covariance estimates are as accurate as the those of the true filter. The model error estimator satisfies the following dynamics,

$$d\hat{e} = \hat{a} \hat{e} dt - \sqrt{\epsilon \sigma_y} \frac{a_{12}}{a_{22}} dW_y - \epsilon \hat{a} \tilde{a} (\hat{e} + \tilde{x}) dt - \epsilon \sigma_x \hat{a} dW_x.$$

where $\hat{a} \equiv a_{12} a_{21} / a_{22}^2$. With this model error estimator, the reduced filter prior model is given by

$$d\hat{x} = \hat{a} (1 - \epsilon \hat{a}) \hat{x} dt + \sigma_x (1 - \epsilon \hat{a}) dW_x - \sqrt{\epsilon \sigma_y} \frac{a_{12}}{a_{22}} dW_y,$$

\hspace{1cm} (20)

*The only reason they consider continuous-time filter is because the Kalman-Bucy solutions have much simpler expressions compared to the discrete-time counterpart. We should note that for the discrete-time observation model in (15), one will obtain a slightly different model error estimator for optimal filtering.
where $\hat{x} \equiv \tilde{x} + \hat{e}$.

We numerically confirm the accuracy of both the mean and covariance estimates with this optimal reduced model in Figure 1. We should also point out that this result was found by enforcing linear optimality condition, $E(e \cdot \hat{x}) = 0$ (which is satisfied when a filtered mean estimate is optimal [80]). With this choice of parameters, the reduced filtered solutions become consistent in the sense that the actual error covariance of the filtered mean estimate matches the filtered error covariance estimate, $E[e^2] = E[(x - \hat{x})^2] + O(\epsilon^2)$. Numerically, notice that the MSE (a numerical estimate for the actual error covariance estimate) and the posterior error covariance estimate in Figure 1 are very similar for only the true filter and the optimal one-dimensional filter. In this example, these are the only consistent filters for linear problem.

We should point out that the same reduced model in (20) can be determined by fitting the reduced filter model to the equilibrium covariance statistics and the correlation time of the underlying true signal that solves (12)-(13) for the slow variable $x$. As a consequence, the optimal reduced model in (20) produces, both, an optimal filtering and an optimal equilibrium statistical prediction; this is the linear theory established in [10].

**Example 2:** Consider the nonlinear filtering problem [40] of noisy observations,

$$z_m = u(t_m) + \varepsilon_m^o, \quad \varepsilon_m^o \sim \mathcal{N}(0, R), \quad (21)$$

where

$$\frac{du}{dt} = - (\gamma + \hat{\lambda}) u + \tilde{b} + f(t) + \sigma_u \dot{W}_u,$$

$$\frac{d\tilde{b}}{dt} = - \lambda b \tilde{b} + \frac{\sigma_b}{\sqrt{\epsilon}} \dot{W}_b,$$

$$\frac{d\tilde{\gamma}}{dt} = - \frac{d_\gamma}{\epsilon} \tilde{\gamma} + \frac{\sigma_\gamma}{\sqrt{\epsilon}} \dot{W}_\gamma,$$

with $\hat{\lambda} = \gamma - i \omega$ and $\lambda_b = \gamma_b - i \omega_b$. The model in (22) was introduced as a stochastic parameterization for filtering a turbulent mode in the presence of model error in [36, 35] (we will review this aspect in Example 3 below). The solutions for the nonlinear filtering problem in (22), (21), was called SPEKF, which stands for Stochastic Parameterized Extended Kalman Filter [36, 35, 71, 69]. In particular, SPEKF posterior statistical solutions are obtained by applying Kalman update to the exactly solvable prior statistical solutions of (22). We should point out that the SPEKF solutions are not the true filtered solutions. For general continuous-time nonlinear filtering problems, the true filtered solutions are characterized by the conditional distribution $p(u_t, b_t, \gamma_t | z_{\tau}, 0 \leq \tau \leq t)$, which solves a stochastically forced partial differential equation known as the Kushner equation [58]. Solving the Kushner equations is nontrivial for general high-dimensional nonlinear problems. It turns out that the posterior solutions of SPEKF for discrete observation time are the analog of the Gaussian closure on the first two-moments of this conditional distribution for the corresponding continuous-time filter [10]. In this sense, one can refer to SPEKF solutions as the best approximate solutions that are numerically attainable since the true filtered solutions are not accessible.
The nonlinear system in (22) has many attractive features as a test model. First, it has exactly solvable statistical solutions which are non-Gaussian. Thus, it allows one to study non-Gaussian prior statistics conditional to the Gaussian posterior statistical solutions of the Kalman update and to verify uncertainty quantification methods [18]. Second, a recent study by [17] suggests that the system in (22) can reproduce signals in various turbulent regimes such as intermittent instabilities in a turbulent energy transfer range and in a dissipative range as well as laminar dynamics.

As in the linear example 1 above, the $O(1)$ dynamics are given by the averaged dynamics, where the average is taken over the unique invariant density generated by the fast dynamics of $\tilde{b}$ and $\tilde{\gamma}$ [40], which results in a linear SDE,

$$\frac{d\tilde{u}}{dt} = -\hat{\lambda}\tilde{u} + f(t) + \sigma_u \dot{W}_u. \tag{23}$$

In the numerical simulation below, we will refer to the filtering scheme with the prior model in (23) as the Reduced Stochastic Filter (RSF). In [40], they defined a reduced stochastic filter with an additive noise correction (RSFA) given by the following model error estimator,

$$\frac{d\hat{e}}{dt} = -\hat{\lambda}\hat{e} + \sqrt{\epsilon}\frac{\sigma_b}{\lambda_b} \dot{W}_b. \tag{24}$$

In [10], they found that the best reduced one-dimensional filtering (best in the sense that the errors in mean and covariance are of order $\epsilon$ from the solutions of SPEKF) can be achieved with a damping and combined, additive and multiplicative, noise corrections,

$$\frac{d\hat{e}}{dt} = -\hat{\lambda}\hat{e} + \sqrt{\epsilon}\frac{\sigma_b}{\lambda_b} \dot{W}_b - \frac{\sigma_\gamma}{\sqrt{d_\gamma(d_\gamma + \epsilon \hat{\gamma})}} (\tilde{u} + \hat{e}) \circ \dot{W}_\gamma, \tag{25}$$

where the multiplicative noise term in (25) is in Stratonovich sense. Notice that we refrain from calling the model estimator in (25) the optimal estimator since the optimal filtered solutions are not accessible unless one can solve the Kushner equation for the full conditional distribution as we explained above. We will refer to the filtered solutions corresponding to model error estimator in (25) as the reduced SPEKF solutions.

Notice that when $\epsilon \hat{\gamma} \ll d_\gamma$, the noise correction model in (25) can be approximated by,

$$\frac{d\hat{e}}{dt} = -\hat{\lambda}\hat{e} + \sqrt{\epsilon}\frac{\sigma_b}{\lambda_b} \dot{W}_b - \sqrt{\epsilon}\frac{\sigma_\gamma}{d_\gamma} (\tilde{u} + \hat{e}) \circ \dot{W}_\gamma, \tag{26}$$

which yields the reduced stochastic prior model RSFC, introduced in [40]. We should point out that the multiplicative noise in [40] is also in the Stratonovich sense. In fact, the statistical solutions for the resulting prior model, accounting for the model error estimate in (26),

$$\frac{d\tilde{u}}{dt} = -\hat{\lambda}\tilde{u} + \hat{b} + f(t) + \sigma_u \dot{W}_u + \sqrt{\epsilon} \left( \frac{\sigma_b}{\lambda_b} \dot{W}_b - \frac{\sigma_\gamma}{d_\gamma} \tilde{u} \circ \dot{W}_\gamma \right), \tag{27}$$
Figure 2: Trajectory of the posterior mean estimates (in grey) compared to the truth (dashes).

were computed in the Stratonovich sense as shown in their Appendix, see the electronic supplementary material of [40]. We should also mention that the reduced model in (27) converges pathwise to solutions, \( u^\epsilon(t) \), of (22).

Here, we only show numerical results for the parameter set corresponding to the turbulent transfer energy range regime \([17, 40]\), \( \epsilon = 1, \hat{\gamma} = 1.2, \gamma_b = 0.5, d_\gamma = 20, \sigma_u = 0.5, \sigma_b = 0.5, \sigma_\gamma = 20 \). In this regime, \( u(t) \) exhibits frequent rapid transient instabilities, and \( \hat{\gamma} \) decays faster than \( u \), that is, \( \epsilon \hat{\gamma} < d_\gamma \), such that the RFSC prior model in (27) is a good approximation of the reduced SPEKF with model error estimator (25). The noisy observations in (21) are sampled at every time interval \( \Delta t = 0.5 \) (shorter than the decay time 0.833) and the noise variance is \( R = 0.5Var(u) \). We will show the numerical results of three reduced filters, where the analyses are updated by the Kalman filter formula with prior models: (i) RSF in (23), (ii) RSFA, accounting for model error with the stochastic model in (24), and (iii) reduced SPEKF, accounting for model error with the stochastic model in (25). We compare the estimates from these three filters with those from solutions of SPEKF in Figure 2.

Notice that the reduced SPEKF, which accounts for model error with the combined additive and multiplicative noise in (25), is the only one method which produces filtered solutions with accuracy that is comparable to that of SPEKF solutions (see Figure 2); the average RMS errors (over 2000 iterations) are 0.7730 for the true filter, 0.7861 for the optimal filter, 1.1356 for RSFA, and 1.5141 for RSF. In Figure 3, we show the corresponding posterior error covariance estimates from various reduced filters, \( \hat{B}_m \), compared to that from SPEKF, \( P^b_m \) (in grey). Notice that RSF and RSFA significantly underestimate the posterior error covariances. The reduced SPEKF, on the other hand,
Figure 3: Trajectory of the posterior covariance estimates corresponding to the filtered mean estimates in Figure 2.
tracks the covariance estimates from SPEKF, quite accurately. More comprehensive results based on various parameter regimes are shown in [10]; in there, they also showed that the same model in (27) produces accurate long term covariance solutions, with accuracy of order-ε.

4 Numerical methods for estimating model error statistics

One way to reduce model error is to simply improve the model and hope that, at one point, the underlying physics can be simulated with an appropriate model. In this situation, one would hope that the model error can be as simple as misspecification of parameters. If this is the case, then one can apply various statistical methods to estimate the unknown parameters; for e.g., one can apply Kalman filter based methods on an augmented vector of state-parameter as described below (cf. (28)-(29)). Unfortunately, it is impossible to avoid model error since we only have access to empirical measurements of nature, as described in [67]. In general nonlinear problems, it is unclear how to choose the appropriate estimator for model error, as opposed to examples 1 and 2 above. Here, we will mention several classical numerical methods to mitigate model errors.

4.1 Estimating mean model error

The classical approach for estimating mean model error is based on the strategy for estimating parameters introduced by [32, 33]. The basic idea of this approach is to implement the filtering scheme on an augmented state-parameter space which consists of the dynamical system for ˜u with the unknown parameters ˜b, assuming a certain dynamical model for ˜b,

\[ \frac{d\tilde{u}}{dt} = F(\tilde{u}, \tilde{b}), \]
\[ \frac{d\tilde{b}}{dt} = G(\tilde{u}, \tilde{b}). \]

(28)
(29)

In practice, this approach was implemented with an additional assumption for the model error covariances. The typical choice is to assume the model error covariances to be proportional to the forecast error covariances [24],

\[ Q_m^b \approx \alpha \bar{D}_m^b, \]

(30)

with an empirically chosen scalar α. Moreover, the parameter model G is often chosen on an ad hoc basis, such as G = 0 or white noise forcing with a small variance [32, 33]. Various models for G were proposed in [6] with empirical choices of $Q_m^b$.

4.2 Estimating model error covariance

In the weak constrained 4D-VAR implementation [89], they assumed unbiased model error, that is, $\bar{b}_m = 0$ in (11), in addition to an empirical model error covariance
estimator as in (30) with constant $\hat{P}_m^b = B$. In the ensemble Kalman filter community, such practice (setting $b_m = 0$ and modeling error covariance with (30)) is known as “multiplicative covariance inflation”; this practical approach was introduced to mitigate covariance underestimation due to unresolved scales model error [41, 91] or when small a ensemble size is used [4]. An alternative approach known as “additive covariance inflation” was also used to account for inhomogeneity of the underestimated covariance matrix [81, 95, 53, 93]. In practice, one prefers the multiplicative covariance inflation rather than the additive covariance inflation since it is difficult to specify an ansatz for the additive inflation matrix with appropriate scaling when the system variables have different quantifying units (personal communication with J.L. Anderson). There is also a relaxation-to-prior method [98] that was used in various applications; this method adjusts the analysis error covariance to be closer to its prior covariance estimate based on,

$$
(u_m^a - \bar{u}_m^a) \leftarrow (1 - \alpha)(u_m^a - \bar{u}_m^a) + \alpha(u_m^b - \bar{u}_m^b),
$$

(31)

with an empirically choice of $\alpha$. Note that this approach implicitly approximates $Q_{bu}^m$ with empirical cross-covariances between the two terms in (31), in addition to $Q_{bm}^b$. A more systematic Bayesian approach that alleviates the covariance undersampling in the ensemble Kalman filter context was studied by [14].

There are also classical approaches [75, 76, 9, 23, 70] that adaptively estimate model error covariance $Q_m^b$, assuming unbiased model error, $e_m$. These adaptive methods basically utilize the information from the innovation vector, $d_m \equiv v_m - g(\bar{u}_m^b)$, to estimate the model error covariance $Q_m^b$ in the linear Kalman filter or extended Kalman filter setting. Recent extension of these noise estimation methods to the ensemble Kalman filter setting was proposed in [11, 43]. Other model error covariance estimation methods, known as adaptive covariance inflation method, were also proposed to estimate the multiplicative factor $\alpha$ in (30) on-the-fly [3, 61, 78].

### 4.3 Estimating both model error statistics (mean and covariance)

Recall that the mathematical analysis in the examples in Section 2 suggests that we should represent model error as a stochastic process for optimal filtering. This implies that practical filtering methods should account for, at least, both model error statistics (mean and covariance) simultaneously. Practically, estimating higher order model error statistics beyond the second moments will be the next important scientific problem once we know how to estimate the first two moments accurately. The numerical approaches discussed above, in contrast, estimate only one of the model error statistics, either mean or covariance, and impose various assumptions on the other statistics which are not estimated. Furthermore, many of these methods ignore estimating the cross-covariances between the prior model and the model error.

As of the author’s knowledge, there are few adhoc methods that simultaneously account for, both, the forecast bias (or mean model error) as well as the model error covariance. For example, the multiphysics (or multimodel) ensemble approach was proposed for simulating surrogate statistics for the model error (see the review article
Another method that was proposed by [62] simultaneously estimates a certain parametric form of mean model error estimator and applies an empirical choice of additive covariance inflation. Independent from filtering, there are linear regression based methods that fit these parameters to long time series, see e.g., [94, 5]. In particular, these methods fit the mean model error to a cubic polynomial and then use then to model the residual with an autoregressive model. While these methods are claimed to be useful for climate statistical prediction, we found that they don’t produce accurate climatological variance and correlation time, and furthermore, they are not very useful for filtering [10]. In fact, the filter solutions with such models blow up when observation time intervals are not so short.

In the next example, we will show a simple example of filtering simulated turbulent signals with intermittent instabilities, in which, a stochastic parameterization is used to mitigate model error. We should point out that this stochastic parameterization approach is motivated by the augmented state-parameter approach in (28)-(29). We will provide a simple mathematical justification for choosing the appropriate model for \( G \) in (29) such that not only are both model error statistics simultaneously accounted, but also the cross-covariances between the prior estimate of the imperfect model and the model error will be accounted as well.

Example 3: This example reviews the test filtering problem considered in [36, 69]. Consider the signal from nature be the solution of the complex scalar Langevin equation with a time dependent damping,

\[
\frac{du}{dt} = -\gamma(t)u(t) + i\omega u(t) + \sigma \dot{W}(t),
\]

where \( \dot{W}(t) \) is a complex white noise forcing, \( \gamma \) is the damping coefficient, characterizing the decaying (correlation) time, and \( \sigma \) is the noise strength, characterizing the variance of the model. To generate significant model errors as well as to mimic intermittent chaotic instability as often occurs in nature, we allow \( \gamma(t) \) to switch between stable \((\gamma > 0)\) and unstable \((\gamma < 0)\) regimes according to a two-state Markov jump process (see e.g., [60]). Here we regard \( u(t) \) as one of the modes from nature in a turbulent signal as is often done in turbulence models [66, 25, 84, 68], and the switching process can mimic physical features such as intermittent baroclinic instability [83]. As often occurs in practice, we assume that the switching process details are not known and only averaged properties are modeled. For the filtering simulation below, we consider noisy observations of \( u \) through the observation model,

\[
v_m = u_m + \sigma^o_m,
\]

at discrete time intervals \( \Delta t = 0.25 \) (chosen to be shorter than the average decaying time, \( 1/\bar{\gamma} \) with \( \bar{\gamma} = 1.5 \)) and noise error variance \( R = Var(u) = 0.008 \).

Mean Stochastic Model (MSM): A simple approach for filtering signals with intermittent instability is to use the Mean Stochastic Model (MSM), [44, 46, 69], which is a linear stochastic model with parameters specified by two equilibrium statistical quantities, energy spectrum and correlation time. In particular, the MSM solves

\[
\frac{d\tilde{u}}{dt} = -\tilde{\gamma}\tilde{u}(t) + i\omega\tilde{u}(t) + \sigma \tilde{W}(t)
\]
for prediction and its first and second order statistics are used for filtering. Here \( \gamma \) is the average damping, characterizing the decaying (correlation) time. With the prior model in (34), model errors are introduced naturally through the unknown two-state continuous time Markov damping coefficient. We should point out that for linear problem filtering with MSM as the reduced prior model will produce optimal filtered solutions, in the sense that the mean and covariance estimates are comparable to those of the true filter, when accurate equilibrium statistics, correlation time and variance, are known (see [10] for rigorous proof of this result). In the numerical simulation here, one can see that the posterior statistical estimates are not optimal since the truth is not linear, as the damping coefficient switches between \( \gamma = 2.27 > 0 \) and \( \gamma = -0.04 < 0 \) at random times with appropriate switching rates that produces average damping of \( \bar{\gamma} = 1.5 \) [36, 69]. Notice that the mean posterior estimates miss the peak of the intermittent instabilities (see Figure 4) and the posterior covariance estimates are constant as functions of time (see the solid line in the bottom panel of Figure 5).

In this numerical simulation, the corresponding average RMS error between the truth and the posterior mean estimate is 0.069 for the MSM filter, much worse than the RMS error for the true filter, 0.041. For a reference, the observation error in this numerical simulation is 0.089. By true filter, we refer to the perfectly specified filter, in which we prescribe the underlying true damping coefficients used in generating the true signals.

**Stochastic Parameterized Extended Kalman Filter (SPEKF):** Here, we will consider a simple version of the nonlinear model in (22), that involves only a multiplicative noise correction to estimate the damping coefficient that randomly switches between positive and negative values. The design of the nonlinear model in (22) for stochastic parameterization was motivated by the augmented state-parameter approach in (28)-(29) with an appropriate choice of stochastic differential equation for the parameter model, \( G \).

If one chooses \( G = 0 \) as proposed in [32], then the deterministic part of the augmented model,

\[
\begin{align*}
\frac{d\tilde{u}}{dt} &= -\gamma \tilde{u} + i\omega \tilde{u} + \sigma \tilde{W}(t) \\
\frac{d\gamma}{dt} &= 0
\end{align*}
\]

is linearly unstable (the linearized dynamics have one zero eigenvalue). The nonlinear controllability condition [47] for the system in (35)-(36) is not satisfied: let \( f = (-(\gamma + i\omega)\tilde{u}, 0)^\top \) and \( g = (\sigma, 0)^\top \),

\[
\det(g, [f, g]) = \det \begin{pmatrix} \sigma & \sigma(\gamma - i\omega) \\ 0 & 0 \end{pmatrix} = 0,
\]

where the Lie bracket for vector fields \( f, g \) is defined as usual, \( [f, g] \equiv (Dg)f - (Df)g \), and notice that the higher order Lie brackets \( [[f, g], g], [[[f, g], g], g], \ldots \) are zero. In this filtering problem, the observation model in (33) is defined by a deterministic operator \( h(\tilde{u}, \gamma) = \tilde{u} \). Therefore, the local observability condition [47] for the nonlinear filtering problem in (35), (33) is satisfied except at \( \tilde{u} = 0 \) since,

\[
\det \begin{pmatrix} \frac{\partial h}{\partial \tilde{u}} & \frac{\partial h}{\partial \gamma} \\ \frac{\partial h}{\partial \tilde{u}} & \frac{\partial h}{\partial \gamma} \end{pmatrix} = \det \begin{pmatrix} 1 & 0 \\ (-\gamma + i\omega) & \tilde{u} \end{pmatrix} = \tilde{u}.
\]
Based on the linear Kalman filter theory [52], the filtering problem in (35), (36), (33) is stable (except on a set of measure zero, $\tilde{u} = 0$) but the filtered solution is not necessarily accurate since the controllability condition is not satisfied.

Another popular choice for the parameter model is white noise [33],

$$\frac{d\gamma}{dt} = \sigma_\gamma \dot{W}_\gamma,$$  

(37)

which makes the system in (35), (37) controllable. However, the filter model in (37) is still linearly unstable and sensitive to the choice of the amplitude of the noise, $\sigma_\gamma$, since $\text{Var}(\gamma) = \sigma_\gamma^2 t$ can be very large at finite time $t$ so $\sigma_\gamma$ can’t be too large.

To avoid the instability issues encountered with the parameter models in (36) and (37), a simple choice for stable filtering is with an Ornstein-Uhlenbeck process. This choice leads us to the nonlinear model, introduced in [36, 35],

$$\frac{d\tilde{u}}{dt} = -\gamma \tilde{u} + i\omega \tilde{u} + \sigma \dot{W}(t),$$  

(38)

$$\frac{d\gamma}{dt} = -d_\gamma \gamma + \sigma_\gamma \dot{W}_\gamma.$$  

(39)

The empirical choice of the dynamical model in (39) does not only guarantee a stable filter model with an invariant measure when $d_\gamma > 0$; the nonlinear system in (38)-(39) has explicit statistical solutions [36, 35, 69] such that no linear tangent approximation is needed to obtain prior statistical solutions. In principle, any filtering method can be used to update these non-Gaussian prior statistics. If the classical Kalman filter formula is used to update these non-Gaussian prior statistics to obtain Gaussian posterior statistics; this nonlinear filter was called SPEKF. For continuous time observations as in (19), the SPEKF solutions are simply the Gaussian closure of the first two moments of the Kushner equation [10].

One drawback with this model is that one has to choose parameters $d_\gamma, \sigma_\gamma$ in (39). A physically sensible choice for $d_\gamma$ depends on the decaying time of parameter $\gamma$, while $\sigma_\gamma$ depends on the allowable error tolerance of the estimate, $\gamma$. In our numerical simulation, we choose $d_\gamma = 0.01$ (much slower decaying time compared to $\tilde{u}$) and $\sigma_\gamma = 0.5\sigma$; we found that the filter accuracy is quite robust for a wide-range of parameters [69]. With these parameters, the filter posterior mean estimates for $\tilde{u}$ are almost identical to those of the true filter (see Figure 4); the average RMS error, 0.045, is also comparable to that of the true filter, 0.041. The posterior mean estimate for the hidden variable, $\gamma$, is not so accurate (see Figure 5) but it fluctuates around the true damping parameter.

Similarly, the posterior error covariance estimate for $\tilde{u}$ is also not perfectly accurate but it tracks the covariance of the true filter closely.

In principle, SPEKF is not very different from the existing approaches mentioned in the beginning of this section; apply the augmented state-parameter space with an empirical choice of model error covariance. Mathematically, however, this approach differs substantially from the others in the following sense: SPEKF uses a stochastically forced differential equation to account for model error in $\gamma$, and it simultaneously corrects the mean model error, the variance of the parameters, and the cross-covariances between the state and parameters. Note that these statistics are coupled to each other since the system in (22)-(39) is nonlinear. The other approaches, in contrast, treat the mean
Figure 4: Trajectory of the posterior mean estimates (in grey) compared to the truth (dashes) of $u$.

Figure 5: Trajectories of the posterior estimate for $\gamma$ from SPEKF (top) and the covariance estimates corresponding to the filtered estimates in Figure 4.
and covariance of the model error separately and many of these methods do not even account for the cross-covariances between the prior estimates and model error.

5 Summary and discussions

In this article, we discussed one major challenge in data assimilation community: filtering multiscale dynamical systems in the presence of model error from unresolved scales. This is a prototypical situation in many applications due to limitations in resolving the smaller scale processes as well as the difficulty to model the interaction across scales. We used simple examples to point out the importance of accounting for model error when the separation of scales are moderate. These examples also elucidate the necessity of treating model error as a stochastic process in a nontrivial fashion for optimal filtering. In general, however, it is difficult to guess the appropriate ansatz for the model error estimator. Several classical approaches to estimate the model error statistics were briefly reviewed. We pointed out that almost all of these methods were designed to estimate one of the model error statistics, either mean or covariance, and impose various assumptions on the other statistics that is not estimated. Furthermore, many of these methods ignore estimating the cross-covariances between the prior model and the model error. We showed a simple example where with slight modification from the classical approach, we can simultaneously account for both the model error mean and covariance as well as the cross covariances between the prior estimate from the imperfect model and the model error.

The main results in the two idealized examples in Section 2 suggest the following simple stochastic model to represent model error,

$$\text{d}e = \alpha \text{e} \text{d}t + \sigma \text{d}W_1 + \beta (\tilde{u} + \text{e}) \circ \text{d}W_2,$$

(40)

where \(\alpha, \beta, \gamma\) are to be determined parameters, and \(\tilde{u}\) are solutions from imperfect model as defined in Section 2.2. Such combined additive and multiplicative noise corrections were already proposed in the context of reduced climate modeling (see \([72, 73]\)). For filtering problems, however, such a model error estimator was never thoroughly tested in real applications. Many questions remained to be answered before one can readily use such stochastic parameterization, they include:

1. Will the ansatz in (40) be useful for filtering general nonlinear problems?
2. If not, what will be the alternative ansatz?
3. What will be the best way to estimate the parameters \(\alpha, \beta, \gamma\) in (40)?
4. Should these parameters be spatially dependent?

In a numerical example of filtering partially observed solutions of the two-layer L96 model \([65]\) with a reduced, one-layer L96 filter model, \([10]\) found that the ansatz in (40) significantly improves the filtered estimates, even with \(\beta = 0\). In their simulations, they use the augmented state-parameter space method in (28)-(29) to estimate \(\alpha\) and Mehra’s method to estimate \(\sigma\) in an ensemble Kalman filter context (see \([11]\) for the detail of the scheme). Furthermore, they also found that the resulting filtered estimates are more consistent, in the sense that the filtered covariance estimate is closer to the
actual error covariance of the mean estimate. With such encouraging results, we plan to test this stochastic parameterization on more realistic applications in near future.

As of the our knowledge, there is no single stochastic parameterization that can give optimal estimates for filtering high-dimensional problems in the presence of model error, since the true dynamical systems are unknown in real applications. We should also mentioned various stochastic parameterization strategies, which can potentially be useful for data assimilation application, but have not been thoroughly tested. They include: (i) stochastic superparameterization [74]; (ii) the reduced order modified quasilinear Gaussian algorithm [85]; (iii) the physics-constrained multilevel nonlinear regression model [70, 43]; (iv) Markov chain type modeling [22, 57]; (v) Autoregressive models [54, 7]; (vi) Heterogeneous Multiscale Methods-based reduced models [55].

We hope that this note provides a better understanding of model error from under-resolving smaller scale processes. Furthermore, we hope to convince the reader’s that we can use simple mathematical analysis to provide guidelines to mitigate model error. There are, of course, other sources of model error that are yet to be understood.

Acknowledgment

The author thanks Jeffrey L. Anderson, Fuqing Zhang, Andrew J. Majda, Tyrus Berry, and anonymous readers for their valuable inputs and criticisms that help improving this note.

References

[1] J.L. Anderson. An ensemble adjustment Kalman filter for data assimilation. *Monthly Weather Review*, 129:2884–2903, 2001.

[2] J.L. Anderson. A local least squares framework for ensemble filtering. *Monthly Weather Review*, 131(4):634–642, 2003.

[3] J.L. Anderson. An adaptive covariance inflation error correction algorithm for ensemble filters. *Tellus A*, 59:210–224, 2007.

[4] J.L. Anderson and S.L. Anderson. A Monte Carlo implementation of the non-linear filtering problem to produce ensemble assimilations and forecasts. *Monthly Weather Review*, 127:2741–2758, 1999.

[5] H. M. Arnold, I. M. Moroz, and T. N. Palmer. Stochastic parametrizations and model uncertainty in the lorenz 96 system. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1991), 2013.

[6] S.-J. Baek, B.R. Hunt, E. Kalnay, E. Ott, and I. Szunyogh. Local ensemble Kalman filtering in the presence of model bias. *Tellus A*, 58(3):293–306, 2006.

[7] E.S. Bakunova and J. Harlim. Optimal filtering of complex turbulent systems with memory depth through consistency constraints. *J. Comput. Physics*, 237:320–343, 2013.
[8] P. Bechtold, M. Köhler, T. Jung, F. Doblas-Reyes, M. Leutbecher, M.J. Rodwell, F. Vitart, and G. Balsamo. Advances in simulating atmospheric variability with the ECMWF model: From synoptic to decadal time-scales. *Quarterly Journal of the Royal Meteorological Society*, 134(634):1337–1351, 2008.

[9] P.R. Belanger. Estimation of noise covariance matrices for a linear time-varying stochastic process. *Automatica*, 10(3):267 – 275, 1974.

[10] T. Berry and J. Harlim. Linear Theory for Filtering Nonlinear Multiscale Systems with Model Error. *Proc. Roy. Soc. A (in press)*, 2014.

[11] T. Berry and T. Sauer. Adaptive ensemble Kalman filtering of nonlinear systems. *Tellus A*, 65:20331, 2013.

[12] C.H. Bishop, B. Etherton, and S.J. Majumdar. Adaptive sampling with the ensemble transform Kalman filter part I: the theoretical aspects. *Monthly Weather Review*, 129:420–436, 2001.

[13] D. Bloemker, K.J.H. Law, A.M. Stuart, and K. Zygalakis. Accuracy and stability of the continuous-time 3dvar filter for the navier-stokes equation. Submitted, 2013.

[14] M. Bocquet. Ensemble kalman filtering without the intrinsic need for inflation. *Nonlinear Processes in Geophysics*, 18(5):735–750, 2011.

[15] M. Bonavita, L. Isaksen, and E. Hólm. On the use of EDA background error variances in the ECMWF 4D-Var. *Quarterly Journal of the Royal Meteorological Society*, 138(667):1540–1559, 2012.

[16] Massimo Bonavita, Laure Raynaud, and Lars Isaksen. Estimating background-error variances with the ECMWF Ensemble of Data Assimilations system: some effects of ensemble size and day-to-day variability. *Quarterly Journal of the Royal Meteorological Society*, 137(655):423–434, 2011.

[17] M. Branicki, B. Gershgorin, and A.J. Majda. Filtering skill for turbulent signals for a suite of nonlinear and linear extended kalman filters. *Journal of Computational Physics*, 231(4):1462 – 1498, 2012.

[18] M. Branicki and A.J. Majda. Fundamental limitations of polynomial chaos for uncertainty quantification in systems with intermittent instabilities. *Comm. Math. Sci.*, 11(1):55–103, 2013.

[19] C.E.A. Brett, K.F. Lam, K.J.H. Law, D.S. McCormick, M.R. Scott, and A.M. Stuart. Accuracy and stability of filters for the navier-stokes equation. *Physica D*, 245:34–45, 2013.

[20] A. M. Clayton, A. C. Lorenc, and D. M. Barker. Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office. *Quarterly Journal of the Royal Meteorological Society*, 139(675):1445–1461, 2013.

[21] P. Courtier, J.-N. Thépaut, and A. Hollingsworth. A strategy for operational implementation of 4D-VAR, using an incremental approach. *Quarterly Journal of Royal Meteorological Society*, 120:1367–1387, 1994.

[22] D. Crommelin and E. Vanden-Eijnden. Subgrid-scale parameterization with conditional Markov chains. *J. Atmos. Sci.*, 65:2661–2675, 2008.
[23] D. Dee, S. Cohn, A. Dalcher, and M. Ghil. An efficient algorithm for estimating noise covariances in distributed systems. *Automatic Control, IEEE Transactions on*, 30(11):1057 – 1065, nov 1985.

[24] D.P. Dee and A.M. da Silva. Data assimilation in the presence of forecast bias. *Quarterly Journal of the Royal Meteorological Society*, 124:269–295, 1998.

[25] T. DeSole. Stochastic model of quasigeostrophic turbulence. *Surveys in Geophysics*, 25(2):107–149, 2004.

[26] G. Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, 99:10143–10162, 1994.

[27] G. Evensen. The ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean Dynamics*, 53:343–367, 2003.

[28] M. Fisher, Y. Tremolet, H. Auvinen, D. Tan, and P. Poli. Weak-constraint and long window 4dvar. Technical Report 655, ECMWF, 2011.

[29] S. Fletcher. Mixed gaussian-lognormal four-dimensional data assimilation. *Tellus A*, 62(3), 2010.

[30] S. J. Fletcher and M. Zupanski. A data assimilation method for log-normally distributed observational errors. *Quarterly Journal of the Royal Meteorological Society*, 132(621):2505–2519, 2006.

[31] Steven J. Fletcher and Milija Zupanski. A hybrid multivariate normal and lognormal distribution for data assimilation. *Atmospheric Science Letters*, 7(2):43–46, 2006.

[32] B. Friedland. Treatment of bias in recursive filtering. *IEEE Trans. Automat. Contr.*, AC-14:359–367, 1969.

[33] B. Friedland. Estimating sudden changes of biases in linear dynamical systems. *IEEE Trans. Automat. Contr.*, AC-27:237–240, 1982.

[34] C.W. Gardiner. *Handbook of Stochastic Methods for physics, chemistry, and the natural sciences*. Springer-Verlag New York, 1997.

[35] B. Gershgorin, J. Harlim, and A.J. Majda. Improving filtering and prediction of spatially extended turbulent systems with model errors through stochastic parameter estimation. *J. Comput. Phys.*, 229(1):32–57, 2010.

[36] B. Gershgorin, J. Harlim, and A.J. Majda. Test models for improving filtering with model errors through stochastic parameter estimation. *J. Comput. Phys.*, 229(1):1–31, 2010.

[37] B. Gershgorin and A.J. Majda. A nonlinear test model for filtering slow-fast systems. *Comm. Math. Sci.*, 6(3):611–649, 2008.

[38] B. Gershgorin and A.J. Majda. Filtering a nonlinear slow-fast system with strong fast forcing. *Comm. Math. Sci.*, 8(1):67–92, 2010.

[39] Cecilia Gonzalez-Tokman and Brian R. Hunt. Ensemble data assimilation for hyperbolic systems. *Physica D: Nonlinear Phenomena*, 243(1):128 – 142, 2013.
[40] G.A. Gottwald and J. Harlim. The role of additive and multiplicative noise in filtering complex dynamical systems. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, 469(2155), 2013.

[41] T.M. Hamill and J.S. Whitaker. Accounting for the error due to unresolved scales in ensemble data assimilation: A comparison of different approaches. *Monthly Weather Review*, 133(11):3132–3147, 2013/11/12 2005.

[42] J. Harlim and B.R. Hunt. A non-Gaussian ensemble filter for assimilating infrequent noisy observations. *Tellus A*, 59(2):225–237, 2007.

[43] J. Harlim, A. Mahdi, and A.J. Majda. An ensemble kalman filter for statistical estimation of physics constrained nonlinear regression models. *Journal of Computational Physics*, 257, Part A:782 – 812, 2014.

[44] J. Harlim and A.J. Majda. Filtering nonlinear dynamical systems with linear stochastic models. *Nonlinearity*, 21(6):1281–1306, 2008.

[45] J. Harlim and A.J. Majda. Mathematical strategies for filtering complex systems: Regularly spaced sparse observations. *Journal of Computational Physics*, 227(10):5304–5341, 2008.

[46] J. Harlim and A.J. Majda. Catastrophic filter divergence in filtering nonlinear dissipative systems. *Comm. Math. Sci.*, 8(1):27–43, 2010.

[47] R. Hermann and Arthur J. Krener. Nonlinear controllability and observability. *Automatic Control, IEEE Transactions on*, 22(5):728–740, 1977.

[48] P.L. Houtekamer and H.L. Mitchell. Data assimilation using an ensemble Kalman filter technique. *Monthly Weather Review*, 126:796–811, 1998.

[49] B.R. Hunt, E.J. Kostelich, and I. Szunyogh. Efficient data assimilation for spatiotemporal chaos: a local ensemble transform Kalman filter. *Physica D*, 230:112–126, 2007.

[50] P. Imkeller, N.S. Namachchivaya, N. Perkowski, and H.C. Yeong. Dimensional reduction in nonlinear filtering: a homogenization approach. *to appear in Annals of Applied Probability*, 2013.

[51] L. Isaksen, M. Bonavita, M. Buizza, M. Fisher, J. Haseler, M. Leutbecher, and L. Raynaud. Ensemble of data assimilations at ECMWF. Technical Report 636, ECMWF, 2010.

[52] R.E. Kalman and R. Bucy. New results in linear filtering and prediction theory. *Trans. ASME J. Basic Eng.*, 83D:95–108, 1961.

[53] E. Kalnay, H. Li, T. Miyoshi, S.-C. Yang, and J. Ballabrera-Poy. 4D-Var or ensemble Kalman filter? *Tellus A*, 59A:758–773, 2007.

[54] E.L. Kang and J. Harlim. Filtering nonlinear spatio-temporal chaos with autoregressive linear stochastic models. *Physica D: Nonlinear Phenomena*, 241(12):1099 – 1113, 2012.

[55] E.L. Kang and J. Harlim. Filtering partially observed multiscale systems with heterogeneous multiscale methods-based reduced climate models. *Monthly Weather Review*, 140(3):860–873, 2012.
[56] Shane R. Keating, Andrew J. Majda, and K. Shafer Smith. New methods for estimating ocean eddy heat transport using satellite altimetry. *Monthly Weather Review*, 140(5):1703–1722, 2012.

[57] B. Khoudier, J. A. Biello, and A. J. Majda. A stochastic multicloud model for tropical convection. *Comm. Math. Sci.*, 8:187–216, 2010.

[58] H. Kushner. On the differential equations satisfied by conditional probability densities of markov processes, with applications. *Journal of the Society for Industrial and Applied Mathematics Series A Control*, 2(1):106–119, 1964.

[59] K.J.H. Law, A. Shukla, and A.M. Stuart. Analysis of the 3dvar filter for the partially observed lorentz ’63 model. *Discrete and Continuous Dynamical Systems A*, 34:1061–1078, 2014.

[60] G. Lawler. *Introduction to Stochastic Processes*. Chapman & Hall/CRC, 1995.

[61] H. Li, E. Kalnay, and T. Miyoshi. Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter. *Quarterly Journal of the Royal Meteorological Society*, 135(639):523–533, 2009.

[62] Hong Li, Eugenia Kalnay, Takemasa Miyoshi, and Christopher M. Danforth. Accounting for model errors in ensemble data assimilation. *Monthly Weather Review*, 137(10):3407–3419, 2013/11/15 2009.

[63] A.C. Lorenc. Analysis methods for numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, 112:1177–1194, 1986.

[64] A.C. Lorenc. The potential of the ensemble Kalman filter for NWP-a comparison with 4D-Var. *Quarterly Journal of Royal Meteorological Society*, 129:3183–3203, 2003.

[65] E.N. Lorenz. Predictability - a problem partly solved. In *Proceedings on predictability*, held at ECMWF on 4-8 September 1995, pages 1–18, 1996.

[66] A.J. Majda, R.V. Abramov, and M.J. Grote. *Information theory and stochastics for multiscale nonlinear systems*. CRM Monograph Series v.25, American Mathematical Society, Providence, Rhode Island, USA, 2005.

[67] A.J. Majda and B. Gershgorin. Link between statistical equilibrium fidelity and forecasting skill for complex systems with model error. *Proc. Nat. Acad. Sci. USA*, 108(31):12599–12604, 2011.

[68] A.J. Majda and M.J. Grote. Explicit off-line criteria for stable accurate time filtering of strongly unstable spatially extended systems. *Proceedings of the National Academy of Sciences*, 104:1124–1129, 2007.

[69] A.J. Majda and J. Harlim. *Filtering Complex Turbulent Systems*. Cambridge University Press, UK, 2012.

[70] A.J. Majda and J. Harlim. Physics constrained nonlinear regression models for time series. *Nonlinearity*, 26:201–217, 2013.

[71] A.J. Majda, J. Harlim, and B. Gershgorin. Mathematical strategies for filtering turbulent dynamical systems. *Discrete and Continuous Dynamical Systems A*, 27(2):441–486, 2010.
[72] A.J. Majda, I. Timofeyev, and E. Vanden-Eijnden. Models for stochastic climate prediction. Proc. Nat. Acad. Sci., 96:15687–15691, 1999.

[73] A.J. Majda, I. Timofeyev, and E. Vanden-Eijnden. A mathematical framework for stochastic climate models. Comm. Pure Appl. Math., 54:891–974, 2001.

[74] Andrew J. Majda and Ian Grooms. New perspectives on superparameterization for geophysical turbulence. Journal of Computational Physics (in press), 2013.

[75] R.K. Mehra. On the identification of variances and adaptive kalman filtering. Automatic Control, IEEE Transactions on, 15(2):175–184, 1970.

[76] R.K. Mehra. Approaches to adaptive filtering. Automatic Control, IEEE Transactions on, 17(5):693–698, 1972.

[77] Zhiyong Meng and Fuqing Zhang. Limited-area ensemble-based data assimilation. Monthly Weather Review, 139(7):2025–2045, 2013/11/15 2011.

[78] T. Miyoshi. The gaussian approach to adaptive covariance inflation and its implementation with the local ensemble transform kalman filter. Monthly Weather Review, 139(5):1519–1535, 2011.

[79] M.W. Moncrieff, M.V. Shapiro, J.M. Slingo, and F. Molteni. Collaborative research at the intersection of weather and climate. World Meteorological Organization Bulletin, 56(3):1–9, 2007.

[80] B. Øksendal. Stochastic Differential Equations: An Introduction with Applications. Universitext. Springer, 2003.

[81] E. Ott, B.R. Hunt, I. Szunyogh, A.V. Zimin, E.J. Kostelich, M. Corrazza, E. Kalnay, and J.A. Yorke. A local ensemble Kalman filter for atmospheric data assimilation. Tellus A, 56:415–428, 2004.

[82] G.A. Pavliotis and A.M. Stuart. Multiscale Methods: Averaging and Homogenization, volume 53 of Texts in Applied Mathematics. Springer, 2000.

[83] J. Pedlosky. Geophysical Fluid Dynamics. Springer, New York, 1979.

[84] R. Salmon. Lectures on geophysical fluid dynamics, volume 378. Oxford University Press, 1998.

[85] T.P. Sapsis and A.J. Majda. Statistically accurate low-order models for uncertainty quantification in turbulent dynamical systems. Proceedings of the National Academy of Sciences, 110(34):13705–13710, 2013.

[86] Y. Téromelet. Accounting for an imperfect model in 4D-Var. Quarterly Journal of the Royal Meteorological Society, 132(621):2483–2504, 2006.
[89] Yannick Trémolet. Model-error estimation in 4d-var. *Quarterly Journal of the Royal Meteorological Society*, 133(626):1267–1280, 2007.

[90] Xuguang Wang, David Parrish, Daryl Kleist, and Jeffrey Whitaker. GSI 3DVar-based Ensemble-Variational Hybrid Data Assimilation for NCEP Global Forecast System: Single Resolution Experiments. *Monthly Weather Review*, 2013.

[91] Jeffrey S. Whitaker and Thomas M. Hamill. Evaluating methods to account for system errors in ensemble data assimilation. *Monthly Weather Review*, 140(9):3078–3089, 2013/11/12 2012.

[92] J.S. Whitaker and T.M. Hamill. Ensemble data assimilation without perturbed observations. *Monthly Weather Review*, 130:1913–1924, 2002.

[93] J.S. Whitaker, T.M. Hamill, X. Wei, Y. Song, and Z. Toth. Ensemble data assimilation with the ncep global forecast system. *Monthly Weather Review*, 136(2):463–482, 2013/10/11 2008.

[94] D.S. Wilks. Effects of stochastic parameterizations in the lorenz 96 model. *Quart. J. Roy. Meteor. Soc.*, 131:389–407, 2005.

[95] S.-C. Yang, D. Baker, K. Cordes, M. Huff, F. Nagpal, E. Okereke, J. Villafane, E. Kalnay, and G.S. Duane. Data assimilation as synchronization of truth and model: experiments with the three-variable lorenz system. *Journal of the Atmospheric Sciences*, 63(9):2340–2354, 2006.

[96] N. Žagar, L. Isaksen, D. Tan, and J. Tribbia. Balance properties of the short-range forecast errors in the ecmwf 4d-var ensemble. *Quarterly Journal of the Royal Meteorological Society*, 139(674):1229–1238, 2013.

[97] C. Zhang. Madden–Julian Oscillation. *Reviews of Geophysics*, 43:G2003+, June 2005.

[98] F. Zhang, C. Snyder, and J. Sun. Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble kalman filter. *Monthly Weather Review*, 132(5):1238–1253, 2004.