The role of additive and multiplicative noise in filtering complex dynamical systems

By Georg A. Gottwald[1] and John Harlim[2]

[1] *School of Mathematics and Statistics, University of Sydney, NSW 2006, Australia,* georg.gottwald@sydney.edu.au

[2] *Department of Mathematics, North Carolina State University, BOX 8205, Raleigh, NC 27695, U.S.A.,* jharlim@ncsu.edu

Covariance inflation is an ad-hoc treatment that is widely used in practical real-time data assimilation algorithms to mitigate covariance underestimation due to model errors, nonlinearity, or/and, in the context of ensemble filters, insufficient ensemble size. In this paper, we systematically derive an effective "statistical" inflation for filtering multi-scale dynamical systems with moderate scale gap, $\epsilon = \mathcal{O}(10^{-1})$, to the case of no scale gap with $\epsilon = \mathcal{O}(1)$, in the presence of model errors through reduced dynamics from rigorous stochastic subgrid-scale parametrizations.

We will demonstrate that for linear problems, an effective covariance inflation is achieved by a systematically derived additive noise in the forecast model, producing superior filtering skill. For nonlinear problems, we will study an analytically solvable stochastic test model, mimicking turbulent signals in regimes ranging from a turbulent energy transfer range to a dissipative range to a laminar regime. In this context, we will show that multiplicative noise naturally arises in addition to additive noise in a reduced stochastic forecast model. Subsequently, we will show that a "statistical" inflation factor that involves mean correction in addition to covariance inflation is necessary to achieve accurate filtering in the presence of intermittent instability in both the turbulent energy transfer range and the dissipative range.

Keywords: data assimilation; filtering; multi-scale systems; covariance inflation; stochastic parametrisation; additive noise; multiplicative noise; model error; averaging

1. Introduction

An integral part of each numerical weather forecasting scheme is the estimation of the most accurate possible initial conditions given a forecast model with possible model error and noisy observations at discrete observation intervals; this process is called *data assimilation* (see e.g. Kalnay 2002, Majda & Harlim 2012). The presence of the often chaotic multi-scale nature of the underlying nonlinear dynamics significantly complicates this process. Each forecast estimate, or *background state*, is an approximation of the current atmospheric state, with a whole range of sources for error and uncertainty such as model error, errors due to the nonlinear chaotic nature of the model, the unavoidable presence of unresolved subgrid-scales, as well as errors caused by the numerical discretisation (see e.g. Palmer 2001).

An attractive method for data assimilation is the ensemble Kalman filter introduced by Evensen (1994, 2006) which computes a Monte-Carlo approximation of the forecast error covariance on-the-fly as an estimate for the degree of uncertainty of the forecast caused by the model dynamics. The idea behind ensemble based methods is that the nonlinear chaotic dynamics of the underlying forecast model and the associated sensitivity to initial conditions cause an ensemble of trajectories to explore sufficiently large parts of the phase space in order to deduce meaningful statistical properties of the dynamics. A requirement for a reliable estimate is an adequate size of the ensemble (Houtekamer & Mitchell 1998, Ehrendorfer 2007, Petrie & Dance 2010). However, all currently operational ensemble systems suffer from insufficient ensemble sizes causing sampling errors. The associated underdispersiveness implies that the true atmospheric state is on average outside the statistically expected range of the forecast or analysis (e.g. Buizza et al. (2005), Hamill & Whitaker (2011)). An underdispersive ensemble usually underestimates the forecast error covariance which potentially leads to filter divergence whereby the filter trusts its own forecast and ignores the information provided by the observations. This filter divergence is caused by ensemble members aligning with the most unstable direction of the forecast dynamics as shown by Ng et al. (2011). To mitigate this underdispersiveness of the forecast ensemble and avoid filter divergence the method of covariance inflation was introduced (Anderson & Anderson 1999) whereby the prior forecast error covariance is artificially increased in each assimilation cycle. Covariance inflation can either be done in a multiplicative (Anderson 2001) or in an additive fashion (Sandu et al. 2007, Houtekamer et al. 2009). This is usually done globally and involves careful and computationally expensive tuning of the inflation factor (for recent methods on adaptive estimation of the inflation factor from the innovation statistics, see Anderson 2007, 2009, Li et al. 2009, Miyoshi 2011). Although covariance inflation is widely used in the context of ensemble based filters, it can also be used to mitigate covariance underestimation in the presence of model errors in non-ensemble based setting. For example, Majda & Harlim (2012) discuss an effective covariance inflation using an information theory criterion to compensate model errors due to numerical discretization in linear filtering problems. From the perspective of linear Kalman filter theory, the covariance inflation improves the linear controllability condition for accurate filtered solutions (Castronovo et al. 2008). Law & Stuart (2012) inflate the static background covariance in a 3DVAR setting to stabilise the filter.

The goal of this paper is to compute an effective "statistical inflation" to achieve accurate filtering in the presence of model errors. In particular, we study filtering multi-scale dynamics with model errors arising from a misrepresentation of unresolved sub-grid scale processes in the regime of moderate time-scale separation. The filtering problem, where the fast degrees of freedom are not observable, has been considered by Pavliotis & Stuart (2007), Zhang (2011), Mitchell & Gottwald (2012), Imkeller et al. (2012), with a large scale gap assumption, $\epsilon \ll 1$. Several numerical methods to address the same problem for moderate scale gap were developed by Harlim (2011), Kang & Harlim (2012), Harlim & Majda (2013). While these methods produce encouraging results on nontrivial applications, unfortunately, they are difficult to be justified in rigorous fashion. Averaging and homogenisation techniques developed by Khasminskii (1968), Kurtz (1973), Papanicolaou (1976) provide a rigorous backbone for developing reduced slow dynamics for certain classes of multi-scale systems. Following the idea of Mitchell & Gottwald (2012) we will study how these stochastic model reductions can be used in data assimilation and how they act as a dynamically consistent way of statistical inflation. Our particular emphasis here is to understand the role of additive and multiplicative noise in stochastic model reduction in improving filtering skill. We employ theoretically well-understood stochastic model reductions such as averaging, formulated in the framework of singular perturbation theory, and stochastic invariant manifold theory. We consider multi-scale systems whose averaging limit does not generate additional stochastic diffusion terms; as an innovative feature, we reintroduce stochastic effects induced by the fast dynamics by going one order higher in the time scale separation parameter ϵ in the classical averaging theory. The question we address here is, in what way do these stochastic effects improve the process of data assimilation for multi-scale systems.

The outline of the paper is the following. In Section 2 we consider a linear model and show how additive inflation naturally arises as judicious model error for the resolved slow scales. We will show that while singular perturbation expansion allows us to improve the estimates of the statistical moments when compared to the averaged system, the reduced Fokker-Planck equation for the slow variable does not support a non-negative probability density function. Consequently, this method does not provide any means to estimate the temporal evolution of the mean and the variance needed for data assimilation. As a remedy, we will derive an approximate reduced stochastic model via invariant manifold theory (Berglund & Gentz 2005, Fenichel 1979, Boxler 1989, Roberts 2008). This yields a Langevin equation for the slow variable the variance of which is one order less accurate than that obtained with the singular perturbation theory, but provides an explicit expression for the temporal evolution which can be used to propagate mean and variance in the data assimilation step. The reduced stochastic differential equation obtained through stochastic invariant manifold theory produces significantly better filtering skills when compared to the classical averaged equation. The additional additive noise correction of the reduced stochastic system provides an effective covariance inflation factor. Nonlinear problems will be discussed in Section 3. There, we will consider an analytically solvable nonlinear stochastic test model, mimicking turbulent signals in regimes ranging from the turbulent energy transfer range to the dissipative range to laminar regime (Majda &

Harlim 2012, Branicki et al. 2012). We will show that the approximate reduced model involves both, additive and multiplicative noise, and improves the accuracy in the actual error convergence rate of solutions when compared to the classic averaged equations. For data assimilation applications, we will derive an analytically solvable non-Gaussian filter based on the approximate reduced model with additive and multiplicative noise corrections which effectively inflates the first and second order statistics. We will demonstrate that this non-Gaussian filter produces accurate filtered solutions comparable to those of the true filter in the presence of intermittent instability in both the turbulent energy transfer range and the dissipative range. We conclude with a short summary and discussion in Section 4. We accompany this article with an electronic supplementary material that discusses the detailed calculations.

2. Linear model

We first consider the linear multi-scale model

$$dx = (a_{11}x + a_{12}y)\,dt + \sigma_x dW_x, \tag{2.1}$$

$$dy = \frac{1}{\epsilon}(a_{21}x + a_{22}y)\,dt + \frac{\sigma_y}{\sqrt{\epsilon}}dW_y, \tag{2.2}$$

for a slow variable $x \in \mathbb{R}$ and fast variable $y \in \mathbb{R}$. Here, W_x, W_y are independent Wiener processes and the parameter ϵ characterizes the time scale gap. 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 invariant joint density. Furthermore we require $\tilde{a} = a_{11} - a_{12}a_{22}^{-1}a_{21} < 0$ to assure that the leading order slow dynamics supports an invariant density (cf.(2.3)).

(a) *Singular perturbation theory: averaging and beyond*

The multi-scale linear system (2.1)-(2.2) is amenable to averaging. A recent exposition of the theory of averaging and its applications is provided for example in Givon et al. (2004) and in Pavliotis & Stuart (2008). The goal of this section is to apply singular perturbation theory to obtain one order higher than the usual averaging limit. We will derive an $\mathcal{O}(\epsilon)$ approximation for the joint probability density function $\rho(x, y, t)$ of the full multi-scale system (2.1)-(2.2), when marginalised over the fast variable y. The resulting higher-order approximation of the probability density will turn out to be not strictly non-negative which implies that it is not a proper probability density function, and therefore cannot be associated with a stochastic differential equation for the slow variable x. However, it will enable us to find corrections to the mean and variance and all higher order moments.

The standard effective averaged slow dynamics,

$$dX = \tilde{a}X\,dt + \sigma_x dW_x, \tag{2.3}$$

is obtained by averaging the slow equation (2.1) over the conditional invariant measure $\rho_\infty(y; x)$ induced by the fast dynamics where the slow variable x is assumed fixed in the limit of infinite time scale separation $\epsilon \to 0$. Formally, the effective averaged slow dynamics can be derived by finding a solution of the Fokker-Planck equation associated with (2.1)-(2.2) with the following multiscale expansion,

$$\rho(x, y, t) = \rho_0 + \epsilon\rho_1 + \epsilon^2\rho_2 + \cdots . \tag{2.4}$$

and the reduced system in (2.3) is the Langevin equation corresponding to marginal distribution, $\int \rho_0(x, y, t)\,dy$ (see Appendix A in the electronic supplementary material). Here, we are looking for a higher order correction through the contribution from $\rho_1(x, y, t)$. In order for $\rho_0 + \epsilon\rho_1$ to be a density we require

$$\int dx\,dy\,\rho_1 = 0. \tag{2.5}$$

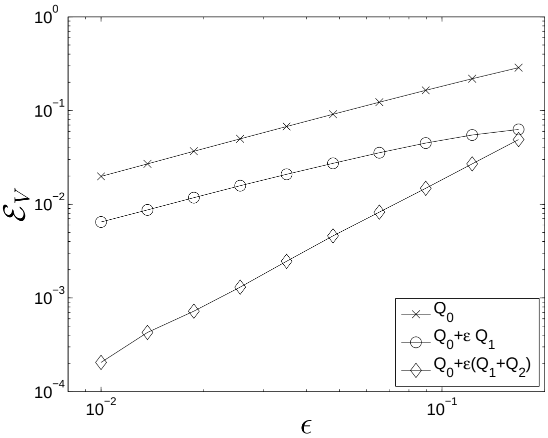

Figure 1. Absolute error of the variance of the slow variable x when calculated from the full multi-scale system (2.1)-(2.2) and from the singular perturbation theory result (2.6) evaluated at $t = 1.3$. The variances were evaluated at time $t = 2.8$. Parameters were with $a_{11} = 1, a_{12} = -1, a_{21} = -1, a_{22} = -1$ and $\sigma_x = \sigma_y = \sqrt{2}$. Linear regression of the upper two data sets reveals a slope of 0.95 and 0.90, respectively, the slope of the lower line is estimated as 1.80.

Again, see Appendix A (and eqn (A14)) for the detailed discussion and an explicit solution for ρ_1 that satisfies (2.5).

We find that for each value of ϵ the expression $\rho = \rho_0 + \epsilon\rho_1$ is not a strictly positive function of x and y. This implies that, contrary to ρ_0, the next order probability density function $\rho = \rho_0 + \epsilon\rho_1$ does not satisfy a Fokker-Planck equation which describes the time evolution of a probability density function. As such it is not possible to find a reduced Langevin equation corresponding to $\rho = \rho_0 + \epsilon\rho_1$. This indicates that in general memory effects are too significant to allow for a decorrelation at that order. In the next subsection we will explore a different approximative approach which allows to derive approximate slow Langevin equations. This as we will see will come at the cost of less accurate estimate of the variance of the slow variable.

Despite the lack of a probabilistic interpretation of the Fokker-Planck equation associated with $\rho_0 + \epsilon\rho_1$, we can use the higher-order approximation of the probability density function to calculate $\mathcal{O}(\epsilon)$-corrections to the variance of the slow variable. We find that $\mathbb{E}_{\rho_1}(x) = 0$, which implies that ρ_1 does not contribute to the mean solution, and

$$\begin{aligned}\mathbb{E}(x^2) &= \mathbb{E}_{\rho_0}(x^2) + \epsilon\mathbb{E}_{\rho_1}(x^2) + \mathcal{O}(\epsilon^2)\\ &= -\frac{\sigma_x^2}{2\tilde{a}}(1 - e^{2\tilde{a}t}) - \epsilon\frac{a_{12}^2\sigma_y^2}{2\tilde{a}a_{22}^2}(1 - e^{2\tilde{a}t}) + \epsilon\frac{a_{12}a_{21}\sigma_x^2}{2\tilde{a}a_{22}^2}(1 - e^{4\tilde{a}t}) + \mathcal{O}(\epsilon^2)\\ &= Q_0 + \epsilon\,(Q_1 + Q_2) + \mathcal{O}(\epsilon^2). \qquad (2.6)\end{aligned}$$

Note that for sufficiently small values of ϵ the approximation to the variance is always positive. In Figure 1 we show how the approximation to the variance (2.6) converges with ϵ to the variance of the slow variable x of the full multi-scale system (2.1)-(2.2). We plot the absolute error of the variances $\mathcal{E}_V = |\mathbb{E}(x^2) - \big(\mathbb{E}_{\rho_0}(x^2) + \epsilon\mathbb{E}_{\rho_1}(x^2)\big)\,|$ using only the averaged system (i.e. Q_0), with partial correction (i.e., $Q_0 + \epsilon Q_1$), and with full higher order correction (i.e. $Q_0 + \epsilon(Q_1 + Q_2)$). As expected from the singular perturbation theory we observe linear scaling behaviour for the averaged system without $\epsilon(Q_1 + Q_2)$ correction and quadratic scaling behaviour with the higher-order correction in (2.6). We should also point out that the covariances $Q_1 + Q_2$ are not Gaussian statistics (one can verify that the kurtosis is not equal to three times of the square of the variance for a generic choice of parameters) and therefore it is mathematically unclear how to use this statistical correction in the data assimilation setting since we have no information about the temporal evolution of these non-Gaussian statistics without a Fokker-Planck equation which supports a probabilistic solution.

(*b*) *Stochastic invariant manifold theory: An approximate reduced slow diffusive model*

In this Section we will derive an approximate reduced stochastic differential equation for the general linear system (2.1)-(2.2). The resulting equation will produce inferior second order statistics compared to the singular perturbation theory described in the previous subsection (cf. (2.6)), but will be advantageous in the context of data assimilation as we will demonstrate in next section.

We employ here invariant manifold theory (see for example Berglund & Gentz 2005, Fenichel 1979, Boxler 1989, Roberts 2008). We perform a formal perturbation theory directly with equations (2.1)-(2.2). Ignoring terms of $\mathcal{O}(\epsilon)$ in (2.2) we obtain

$$y = h(x,t) = -\frac{a_{21}}{a_{22}}x - \sqrt{\epsilon}\frac{\sigma_x}{a_{22}}\dot{W}_y + \mathcal{O}(\epsilon).$$

The slow stochastic invariant manifold $h(x,t)$ is hyperbolic, satisfying the requirements outlined in Fenichel (1979). Substituting into (2.1) and ignoring the $\mathcal{O}(\epsilon)$ contribution, we arrive at a reduced equation for the slow variable

$$d\tilde{X} = \tilde{a}\tilde{X}\,dt + \sigma_x dW_x - \sqrt{\epsilon}\sigma_y\frac{a_{12}}{a_{22}}dW_y, \tag{2.7}$$

which contains an additional additive noise term when compared to the averaged reduced equation (2.3). Note that whereas in the framework of the Fokker-Planck equation employed in singular perturbation theory the diffusion term and the drift term of the fast equation are of the same order in ϵ, invariant manifold theory works directly with the stochastic differential equation and hence allocates different orders of ϵ to the drift and diffusion terms.

We now show that solutions of (2.7) converge to solutions $x^\epsilon(t)$ of the full system (2.1)-(2.2) in a pathwise sense. There exists a vast literature on stochastic invariant manifold theory (see for example the monograph by Berglund & Gentz (2005)). There are, however, not many convergence results which establish the scaling with the time scale separation ϵ. In Appendix B in the electronic supplementary material, we show that the error $e(t) = x^\epsilon(t) - \tilde{X}(t)$ is bounded for finite time T by

$$\mathbb{E}\Big(\sup_{0\le t\le T}|e(t)|^2\Big) \le c\epsilon^2. \tag{2.8}$$

This states that the rate of convergence of solutions of the stochastically reduced system (2.7) is one order better than for the averaged system (2.3). Figure 2 illustrates the convergence of solutions of the averaged model (2.3) and the reduced model (2.7) obtained from stochastic invariant manifold theory confirming the estimate in (2.8). Figure 1, however, shows that this superior scaling of the solutions does not imply better scaling in the convergence of the variances. Recall that the variance of the reduced system (2.7) $\mathbb{E}(\tilde{X}^2) = Q_0 + \epsilon Q_1$ constitutes a truncation of the $\mathcal{O}(\epsilon)$-correction (see (2.6)), but note the improvement in the actual error when compared to the averaged system with $\mathbb{E}(X^2) = Q_0$ (again, see Figure 2).

(*c*) *Data assimilation for the linear model*

In this section, we compare numerical results of filtering the linear system in (2.1)-(2.2), assimilating noisy observations,

$$z_m = x(t_m) + \varepsilon_m^o = Gu(t_m) + \varepsilon_m^o, \quad \varepsilon_m^o \sim \mathcal{N}(0, r^o), \tag{2.9}$$

of the slow variable x at discrete time step t_m with constant observation time interval $\Delta t = t_{m+1} - t_m$. For convenience, we define $u = (x,y)^T$ and the observation operator $G = [1,0]$. The observations in (2.9) are corrupted by unbiased Gaussian noise with variance r^o. In our numerical simulation below, we define signal-to-noise ratio, $SNR = Var(x)/r^o$, such that SNR^{-1} indicates the relative observation error variance compared to the temporal (climatological) variance of x.

Suppose that u_m^- and R_m^- are the prior mean and error covariance estimates at time t_m. In this Gaussian and linear setting, the optimal posterior mean, u_m^+, and error covariance, R_m^+, estimates are obtained by the standard Kalman filter formula (Anderson & Moore 1979):

$$u_m^+ = u_m^- + K_m(z_m - Gu_m^-),$$

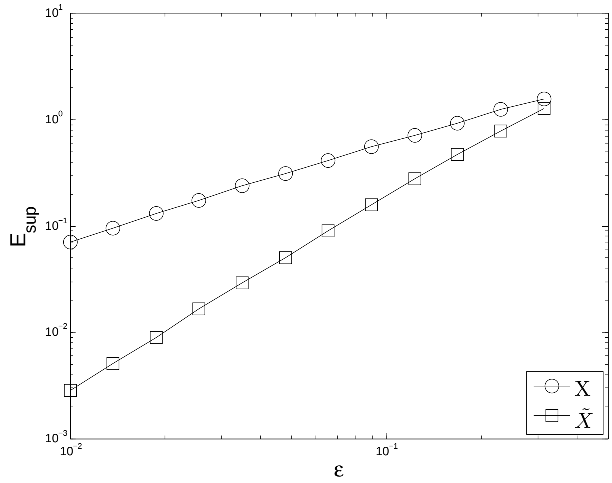

Figure 2. Convergence of solutions x^ϵ of the full model (2.1)-(2.2) to those of the reduced models, (2.3) and (2.7), on a time interval $0 \le t \le T = 250$. Parameters are $a_{11} = a_{21} = a_{22} = -1, a_{12} = 1$ and $\sigma_x = \sigma_y = \sqrt{2}$. The supremum errors, $E_{sup} \equiv \mathbb{E}(\sup_{0\le t\le T} |e(t)|^2)$, where $e(t) = x^\epsilon(t) - X(t)$ (circles) and $e(t) = x^\epsilon(t) - \tilde{X}(t)$ (squares), are plotted as functions of ϵ. Trajectories were averaged over 100 realisations of the Wiener processes. Linear regression of the two data sets reveals a slope of 0.9 and 1.8 for the error of X and $\tilde{X}$, respectively.

$$R_m^+ = (I - K_m G) R_m^-, \qquad (2.10)$$
$$K_m = R_m^- G^T (G R_m^- G^T + r^o)^{-1}.$$

To complete one cycle of filtering (or data assimilation), the next prior statistical estimates are obtained by propagating the posterior according to

$$u_{m+1}^- = F u_m^+, \qquad (2.11)$$
$$R_{m+1}^- = F R_m^+ F^T + Q, \qquad (2.12)$$

where for the *true filter*, F and Q are 2×2-matrices associated with the dynamical propagation operator, corresponding to the full linear multi-scale system (2.1)-(2.2), and the model error covariance, respectively. We will compare the true filter with two reduced filtering strategies: (i) *Reduced Stochastic Filter (RSF)* which implements a one-dimensional Kalman filter on Gaussian prior statistics produced by the standard averaging model (2.3); (ii) *Reduced Stochastic Filter with Additive correction (RSFA)* which implements a one-dimensional Kalman filter on Gaussian prior statistics produced by model (2.7) obtained from stochastic invariant manifold theory. For both reduced filters (RSF and RSFA), we implement the same formulae (2.10) and (2.11), replacing u with x, $G = [1, 0]$ with $G = 1$, and using $F = \exp(\tilde{a}\Delta t)$. The difference between RSF and RSFA is on the covariance update where $Q = Q_0$ for RSF and $Q = Q_0 + \epsilon Q_1$ for RSFA. Therefore, RSFA is nothing but applying RSF with an effective additive covariance inflation factor, ϵQ_1.

To measure the filter accuracy, we compute a temporally average root-mean-square (RMS) error between the posterior mean estimate, x_m^+, and the underlying true signal, $x(t_m)$, for numerical simulations up to time $T = 10,000$. In Figure 3, we show the average RMS errors as functions of ϵ for $\Delta t = 1$ and $SNR^{-1} = 0.5$ (left panel); as functions of SNR^{-1} for $\epsilon = 1$ and $\Delta t = 1$ (middle panel); and as functions of Δt for $\epsilon = 1$ and $SNR^{-1} = 0.5$ (right panel). Notice that the average RMS error for RSFA is almost identical to that of the true filter. More importantly, the additive inflation in RSFA improves the filtered accuracy compared to RSF.

3. Nonlinear model

In this section, we consider solutions of the nonlinear SPEKF model as true signals. This nonlinear test model was introduced in Gershgorin et al. (2010*a*,*b*) for filtering multi-scale turbulent signals with hidden

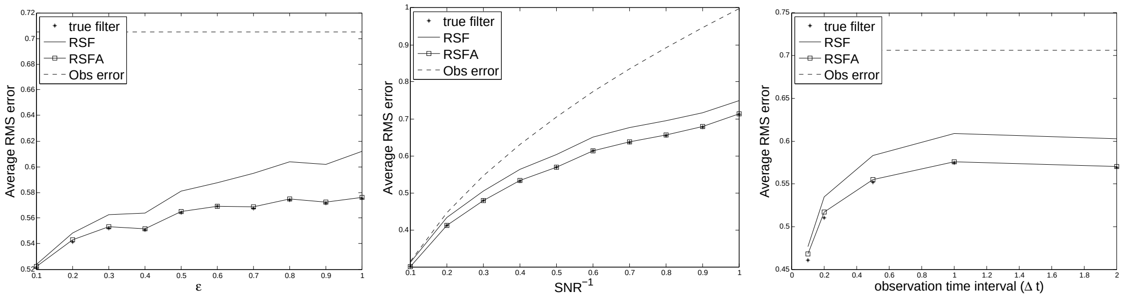

Figure 3. Filter accuracy: Average RMS errors as functions of ϵ for $\Delta t = 1$ and $SNR^{-1} = 0.5$ (left panel); as functions of SNR^{-1} for $\epsilon = 1$ and $\Delta t = 1$ (middle panel); as functions of Δt for $\epsilon = 1$ and $SNR^{-1} = 0.5$ (right panel). In these simulations, the model parameters were $a_{11} = a_{21} = a_{22} = -1, a_{12} = 1$ and $\sigma_x = \sigma_y = \sqrt{2}$.

intermittent instabilities. In particular, the SPEKF model is given by the following system of SDEs,

$$\begin{aligned}
\frac{du}{dt} &= -(\tilde{\gamma} + \hat{\lambda})u + \hat{b} + \tilde{b} + f(t) + \sigma_u \dot{W}_u, \\
\frac{d\tilde{b}}{dt} &= -\frac{\lambda_b}{\epsilon}\tilde{b} + \frac{\sigma_b}{\sqrt{\epsilon}}\dot{W}_b, \\
\frac{d\tilde{\gamma}}{dt} &= -\frac{d_\gamma}{\epsilon}\tilde{\gamma} + \frac{\sigma_\gamma}{\sqrt{\epsilon}}\dot{W}_\gamma,
\end{aligned} \tag{3.1}$$

with $\hat{\lambda} = \hat{\gamma} - \mathrm{i}\omega$ and $\lambda_b = \gamma_b - \mathrm{i}\omega_b$. Here, u represents a resolved mode in a turbulent signal with the nonlinear mode-interaction terms replaced by an additive complex-valued noise term $\tilde{b}(t)$ and multiplicative stochastic damping $\tilde{\gamma}(t)$ as it is often done in turbulence modeling (Delsole 2004, Majda et al. 1999, 2001, 2003, 2008). In (3.1), W_u, W_b are independent complex valued Wiener processes and W_γ is a real valued standard Wiener process. The nonlinear stochastic system in (3.1) involves the following parameters: stationary mean damping $\hat{\gamma}$ and mean bias $\hat{b}$, and two oscillation frequencies ω and ω_b for the slow mode u; two damping parameters γ_b and d_γ for the fast variables $\tilde{b}$ and $\tilde{\gamma}$; three noise amplitudes $\sigma_u, \sigma_b, \sigma_\gamma > 0$; and deterministic forcing $f(t)$ of the slow variable u. Here, two stochastic parameters $\tilde{b}(t)$ and $\tilde{\gamma}(t)$ are modeled with Ornstein-Uhlenbeck processes, rather than treated as constant or as a Wiener processes (Friedland 1969, 1982). For our purpose, we consider the temporal scales of $\tilde{b}, \tilde{\gamma}$ to be of order t/ϵ, which leaves the system amenable to averaging for $\epsilon \ll 1$.

The nonlinear system in (3.1) has several attractive features as a test model. First, it has exactly solvable statistical solutions which are non-Gaussian, allowing to study non-Gaussian prior statistics conditional on the Gaussian posterior statistics in a Kalman filter. This filtering method is called “Stochastic Parameterized Extended Kalman Filter” (SPEKF). Note that it is different from the classical extended Kalman filter that produces Gaussian prior covariance matrix through the corresponding linearized tangent model (see e.g., Anderson & Moore 1979, Kalnay 2002). Second, a recent study by Branicki et al. (2012) suggests that the system (3.1) 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. In Figure 4, we show pathwise solutions $\mathrm{Re}[u(t)]$ of the system in (3.1) without deterministic forcing ($f(t) = 0$), unbiased $\hat{b} = 0$, frequencies $\omega = 1.78$ and $\omega_b = 1$, for a non-time scale separated situation with $\epsilon = 1$, in three turbulent regimes considered in Branicki et al. (2012):

I. **Turbulent transfer energy range.** The dynamics of $u(t)$ is dominated by frequent rapid transient instabilities. The parameters are: $\hat{\gamma} = 1.2$, $\gamma_b = 0.5$, $d_\gamma = 20$, $\sigma_u = 0.5, \sigma_b = 0.5, \sigma_\gamma = 20$. Here, $\tilde{\gamma}$ decays faster than u.

II. **Dissipative range.** The dynamics of $u(t)$ exhibits intermittent burst of transient instabilities, followed by quiescent phases. The parameters are: $\hat{\gamma} = 0.55$, $\gamma_b = 0.4$, $d_\gamma = 0.5$, $\sigma_u = 0.1, \sigma_b = 0.4, \sigma_\gamma = 0.5$. Here, u and $\tilde{\gamma}$ have comparable decaying time scales.

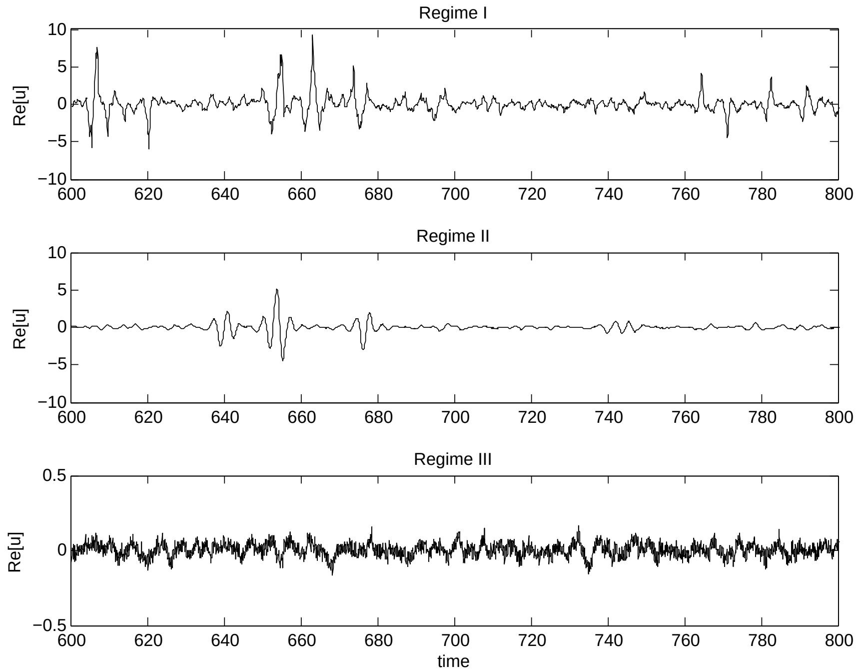

Figure 4. Pathwise solutions of SPEKF model (real part of u) in three turbulent regimes considered in Branicki et al. (2012). Notice the vertical scale of regime III is much smaller than the other two regimes.

III. **Laminar mode.** Here, $u(t)$ has no transient instabilities (almost surely). The parameters are: $\hat{\gamma} = 20$, $\gamma_b = 0.5$, $d_\gamma = 0.25$, $\sigma_u = 0.25, \sigma_b = 0.5, \sigma_\gamma = 1$. Here, u decays much faster than $\tilde{\gamma}$.

We remark that regimes I and II exist only for sufficiently large values of ϵ. For smaller ϵ, the solutions in these two regimes qualitatively look like a laminar mode. Second, for $\epsilon = 1$, the following inequality is satisfied,

$$\Xi_n \equiv -n\hat{\gamma} + \epsilon \frac{n^2\sigma_\gamma^2}{2d_\gamma^2} < 0, \tag{3.2}$$

for $n = 1$ in all three regimes; this is a sufficient condition for stable mean solutions (Branicki et al. 2012). For $n = 2$, the condition in (3.2) is only satisfied in Regimes I and III. In Regime I, where $\tilde{\gamma}$ decays much faster than u, this condition implies bounded first and second order statistics (see Appendix D of Branicki & Majda 2013).

Next, we will find a reduced stochastic prior model corresponding to the nonlinear system in (3.1). Subsequently, we will discuss the strong error convergence rate. Finally, we will compare the filtered solutions from the proposed reduced stochastic model with the true filter with exact statistical prior solutions and various classical approximate filters.

(*a*) *Reduced stochastic models*

The goal of this section is to derive an approximate $\mathcal{O}(\epsilon)$-correction for the effective stochastic model in (3.1). As in the linear case, the $\mathcal{O}(1)$ dynamics is given by the averaged dynamics, where the average is taken over the unique invariant density generated by the fast dynamics of $\tilde{b}$ and $\tilde{\gamma}$, which is

$$\frac{dU}{dt} = -\hat{\lambda}U + \hat{b} + f(t) + \sigma_u \dot{W}_u. \tag{3.3}$$

To recover stochastic effects of the fast dynamics neglected in (3.3), we employ invariant manifold theory. We do not pursue here singular perturbation theory as it does not yield information on the temporal

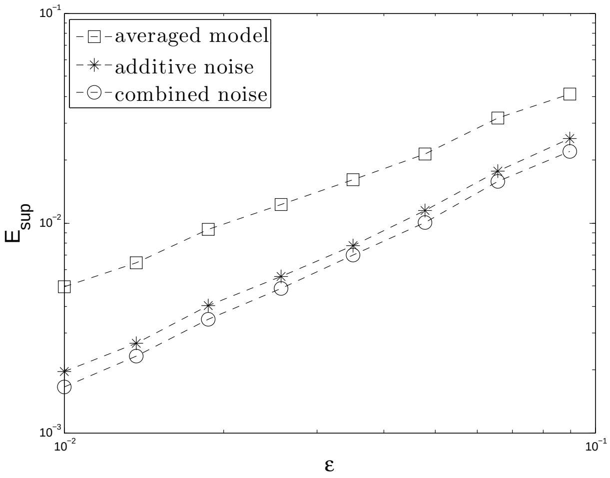

Figure 5. Convergence of solutions u^ϵ of the full model (3.1) to those of several reduced models on a time interval $0 \leq t \leq T = 0.5$. The supremum error, $E_{sup} \equiv \mathbb{E}(\sup_{0\leq t\leq T} |e(t)|^2)$, where $e(t) = u^\epsilon(t) - U(t)$, is plotted as a function of ϵ. Trajectories were averaged over 500 realizations of the Wiener processes. Linear regression of the data sets reveal slopes close to 1. Parameters are for Regime I, where $\tilde{\gamma}$ decays faster than u, and $\Xi_4 < 0$ for these values of ϵ.

evolution of the statistics as already encountered in the linear case of the previous section. We will show that invariant manifold theory naturally produces order-ϵ correction factors that consist of both additive and multiplicative noise.

Proceeding as in Section 2.2, we consider an $\mathcal{O}(\sqrt{\epsilon})$-approximation of the invariant manifold of the fast subsystem

$$\tilde{b} = \sqrt{\epsilon}\frac{\sigma_b}{\lambda_b}\dot{W}_b + \mathcal{O}(\epsilon),$$
$$\tilde{\gamma} = \sqrt{\epsilon}\frac{\sigma_\gamma}{d_\gamma}\dot{W}_\gamma + \mathcal{O}(\epsilon).$$

Inserting into the equation for the slow complex-valued variable u in (3.1) we obtain an approximate reduced model,

$$\frac{dU}{dt} = -\hat{\lambda}U + \hat{b} + f(t) + \sigma_u\dot{W}_u + \sqrt{\epsilon}\Big(\frac{\sigma_b}{\lambda_b}\dot{W}_b - \frac{\sigma_\gamma}{d_\gamma}U\dot{W}_\gamma\Big). \tag{3.4}$$

Note that the reduced system (3.4) can be written as the averaged system (3.3) with added $\mathcal{O}(\sqrt{\epsilon})$ additive and multiplicative noise. This model allows for a complete analytical description of the statistics. In Appendix C (see the electronic supplementary material) we present explicit expressions for the first and second moments of (3.4). In Figure 5, we show the numerically simulated errors of solutions of the approximate reduced models when compared to solutions of the full multi-scale system (3.1). All reduced models, the classical averaged model (3.3), the reduced stochastic model (3.4) and the reduced model (3.4) with only additive noise exhibit linear scaling with ϵ. Note that the absolute error is smaller though for the higher-order models (3.4). Notice that the multiplicative noise does not contribute here much to the closeness of solutions; however, we will see below that it will be significant for filtering turbulent signals when $\epsilon = 1$. In the electronic supplementary material (Appendix D), we present a convergence proof for solutions of the reduced model (3.4).

(*b*) *Data assimilation for the nonlinear model*

In this section, we report numerical results from filtering the SPEKF model (3.1). We assume partial observations v_m of u only,

$$v_m = u(t_m) + \varepsilon_m^o, \quad \varepsilon_m^o \sim \mathcal{N}(0, r^o), \tag{3.5}$$

at discrete time step t_m with observation time interval $\Delta t = t_{m+1} - t_m$. We choose $\Delta t = 0.5$ in regimes I and II so that it is shorter than the decorrelation times, 0.833 and 1.81, respectively. In regime III, since the decorrelation time is much shorter, 0.12, we choose $\Delta t = 0.05$. The observations in (3.5) are corrupted by unbiased Gaussian noise with variance r^o. We report on observation noise variances corresponding to $SNR^{-1} = 0.5$, that is, r^o is 50% of the climatological variance of u. We have checked that our results are robust when changing to smaller observational noise with $SNR^{-1} = 0.1$.

We will investigate the performance of our reduced models as forecast models in the Kalman filter for strong, moderate, and no time scale separation. We recall that the intermittent regime II does only exist for sufficiently large values of $\epsilon \sim \mathcal{O}(1)$. We compare four filtering schemes, one perfect and three imperfect model experiments:

1. The true filter. This filtering scheme applies the classical Kalman filter formula (2.10) to update the first and second order non-Gaussian prior statistical solutions at each data assimilation step. These non-Gaussian prior statistics are semi-analytical statistical solutions of the nonlinear SPEKF model in (3.1) (see Gershgorin et al. 2010*b*, Majda & Harlim 2012, for the detailed statistics).

2. Reduced stochastic filter (RSF). This approach implements a one-dimensional Kalman filter on Gaussian prior statistical solutions of the averaged system (3.3).

3. Reduced stochastic filter with additive inflation (RFSA). This method implements a one-dimensional Kalman filter on Gaussian prior statistical solutions of the reduced stochastic model in (3.4) ignoring the $\mathcal{O}(\sqrt{\epsilon})$ multiplicative noise term. Technically, this method inflates the prior covariance of the reduced filtering approach in 2 with an additive correction factor, $\epsilon \frac{\sigma_b^2}{2|\lambda_b|^2 \hat{\gamma}}(1 - e^{-2\hat{\gamma}\Delta t})$, associated to the order-$\sqrt{\epsilon}$ additive noise term in (3.4).

4. Reduced stochastic filter with combined, additive and multiplicative, inflations (RSFC). This filtering scheme applies a one-dimensional Kalman filter formula to update the non-Gaussian statistical solutions (see Appendix C in the electronic supplementary material) of the reduced stochastic model in (3.4). The presence of the multiplicative noise correction term yields a statistical correction on both the mean and covariance. This is what we refer as to "statistical inflation" in the abstract and introduction.

In Figure 6 we show the average RMS error as a function of the time scale separation parameter ϵ in regimes I and III. For sufficiently small values of ϵ all filters perform equally well, consistent with the asymptotic theory; here, the higher order correction is negligible since ϵ is small. For moderate time-scale separation with $\epsilon = 0.5$ the two higher-order filters produce better skills than RSF which was based on classical averaging theory. RSFC, which includes multiplicative noise, performed slightly better than RSFA; here, the transient instabilities have smaller magnitude compared to those with $\epsilon = 1$ shown in Figure 4 and the RMS error suggests that RSFA has relatively high filtering skill. Here, we ignore showing regime II for smaller ϵ since the solutions look like a laminar mode.

In the remainder of this section, we consider the three turbulent regimes in a tough test bed without time-scale separation with $\epsilon = 1$. In the turbulent energy transfer regime I, $\tilde{\gamma}$ decays much faster than u. The numerical result suggests that RSFC, including multiplicative noise, produces significantly better filtered estimates than all other filters with RMS errors comparable to the true filter. Note that for this regime, the optimal inflation factor for which the smallest RMS errors are obtained is unphysical large with a value of 7.1, and still produces RMS errors 10% larger than those of the true filter (results are not shown). This is analogous to the situation in ensemble filters found by Mitchell & Gottwald (2012). In Figure 7, we show the posterior estimates for Regime I for observation error corresponding to $SNR^{-1} = 0.5$, compared to the true signal. Notice that the estimates from RFSC and the true filter are nearly identical; they capture almost every intermittent instability in this turbulent energy transfer range. The other two methods, RSF and RSFA, do not capture these intermittent instabilities. In this regime, we also see an improved covariance estimate with RSFC compared to RSF and RSFA (results are not shown).

In the dissipative regime II, we consider observation error $\sqrt{r^o} = 1.2786$; this choice corresponds to $SNR^{-1} = 0.5$ of the temporal average over time $t = [0, 400]$, ignoring the immense intermittent burst

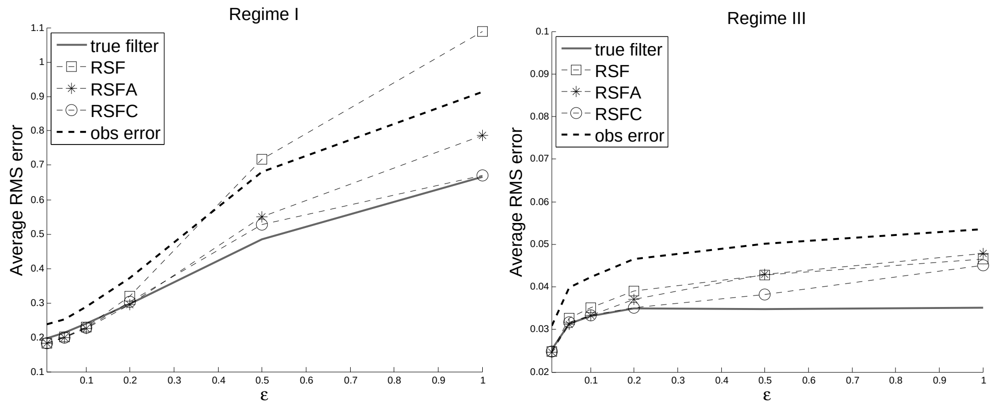

Figure 6. Average RMS errors as functions of ϵ for $SNR^{-1} = 0.5$ in regimes I and III, for observation interval $\Delta t = 0.5$ for regime I and $\Delta t = 0.05$ for regime III. We assimilated until $T = 1000$.

that occurs at time interval $[540, 580]$ (see Fig. 8). This regime is an extremely difficult test bed since the variance for the chosen parameters is unstable (that is, $\Xi_2 < 0$), and furthermore the decaying time scales for u and $\tilde{\gamma}$ are comparable. Nevertheless, RSFC performs extremely well for observation times not too large when compared to the growth rate of the instability, Ξ_2. RSFC exhibits an RMS error of 0.70 which is close to the true filter with an RMS error of 0.58 and much lower than the observation error, $\sqrt{r^o} = 1.2786$. The filters without multiplicative noise, RSF and RSFA, have RMS errors of one magnitude higher with 14.9 and 13.3, respectively. In Figure 8 we show the posterior estimates for Regime II. The multiplicative, i.e. amplitude dependent, nature of the noise clearly enables the filter to follow abrupt large amplitude excursions of the signal. This illustrates well the necessity to incorporate multiplicative noise in modelling highly non-Gaussian intermittent signals.

In the laminar regime III, all reduced filters are doing equally well with RMS errors lying between the true filter and the observational noise (see Fig 6). The filtered estimate from RSFC is slightly less accurate than those of the true filter; we suspect that this deterioration is because u decays much faster than $\tilde{\gamma}$ in this parameter regime, and hence the reduced stochastic systems are not dynamically consistent with the full dynamics. We should also mention that for all these three regimes with $\epsilon = 1$, we also simulate the classical EKF with linear tangent model and its solutions diverge to infinity in finite time; this is due to the practical violation of the observability condition in filtering intermittent unstable dynamics in these turbulent regimes (see Chapter 3 of Majda & Harlim 2012, in a simpler context).

4. Summary and discussion

In this paper, we presented a study of filtering partially observed multi-scale systems with reduced stochastic models obtained from a systematic closure on the unresolved fast processes. In particular, we considered stochastic reduced models derived by singular perturbation theory as well as invariant manifold theory. Here, we were not only showing convergence of solutions in the limit of large time scale separation, but we also tackled the question of how the stochasticity induced by the unresolved scales can enhance the filtering skill, and how their diffusive behaviour can be translated into effective inflation. Our work suggests that by incorporating results from stochastic perturbation theory one may guide inflation strategies which so far have been mostly ad-hoc to more efficiently filter multi-scale systems in regimes where the asymptotic theory usually fails.

We focussed here on the realistic case of moderate and non-existing time scale separation in our data assimilation experiments, and showed how judicious model error, in particular additive and multiplicative noise, can enhance filter performance by providing sufficient dynamically adaptive statistical inflation. We systematically derived higher order additive noise in a reduced stochastic model that does not only improve the path-wise estimate of solutions, but also improves the filtered estimates through a covariance inflation for the Kalman update in the linear case and in a nonlinear example that mimics a laminar

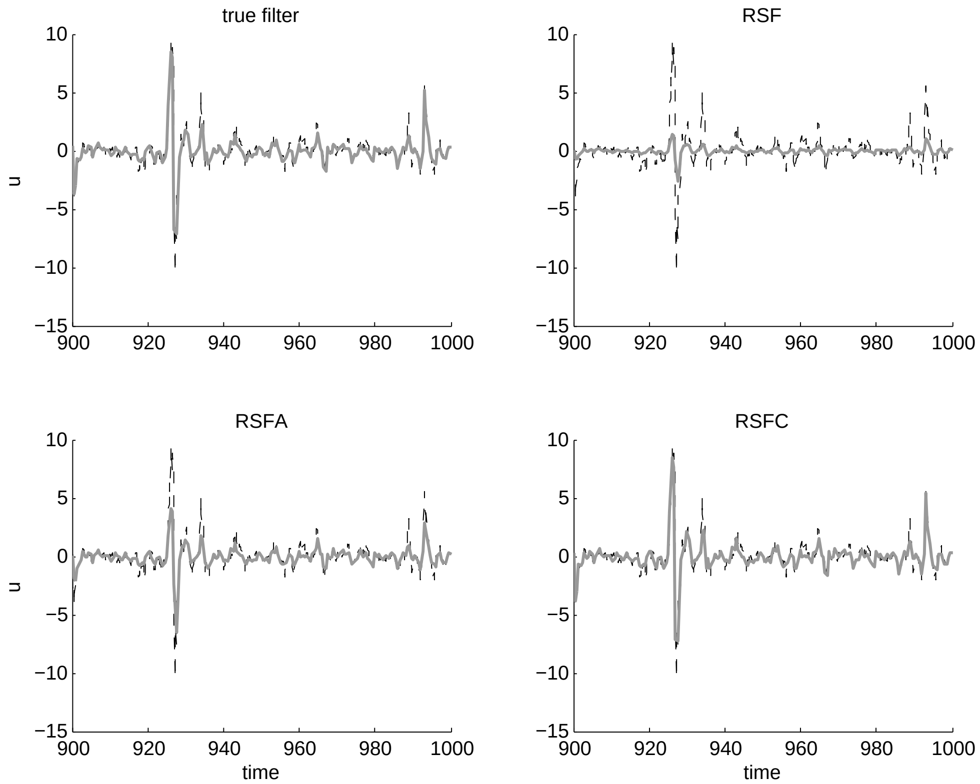

Figure 7. Filtered posterior estimates (grey) compared to the truth (black dashes) for regime I with $SNR^{-1} = 0.5$.

mode of a turbulent signal. We also systematically derived multiplicative noise, which implies both mean and covariance corrections for the Kalman update, with a significant improvement of the filtering skill for intermittent nonlinear dynamics with sudden large amplitude excursions.

The main message of this work here is that reduced stochastic models can be viewed as dynamically consistent way to introduce covariance inflation as well as mean correction, guiding the filtering process. As already found in Mitchell & Gottwald (2012) in the context of ensemble filters, the improved skill of reduced stochastic models is not due to their ability to accurately reproduce the slow dynamics of the full system, but rather by providing additional covariance inflation. It is pertinent to mention though that skill improvement is not observed in regimes where the stochastic reduced model fails to sufficiently approximate the statistics of the full dynamics. For example, the stochastic reduced model (3.4) produces inferior skill to the classical averaged system (3.3) for large vales of ϵ in the laminar regime in which resolved variables decay much faster than the unresolved variables. Hence it is an important feature of the effective stochastic inflation that it is dynamically consistent with the full dynamics and state-adaptable, contrary to adhoc inflation techniques.

Finally, we should mention that although the study in this paper demonstrates the importance of a statistical inflation in the form of additive and multiplicative noise stochastic forcings for optimal filtering, it does not tell us how to choose the parameters associated with these noises $(\epsilon, \gamma_b, \sigma_b, d_\gamma, \sigma_\gamma)$ for real applications. This issue will be addressed in the future.

Acknowledgments

The authors thank Andrew Stuart for insightful discussions. GAG acknowledges support from the Australian Research Council. The research of JH was partially supported by the ONR grant N00014-11-1-0310 and the ONR MURI grant N00014-12-1-0912.

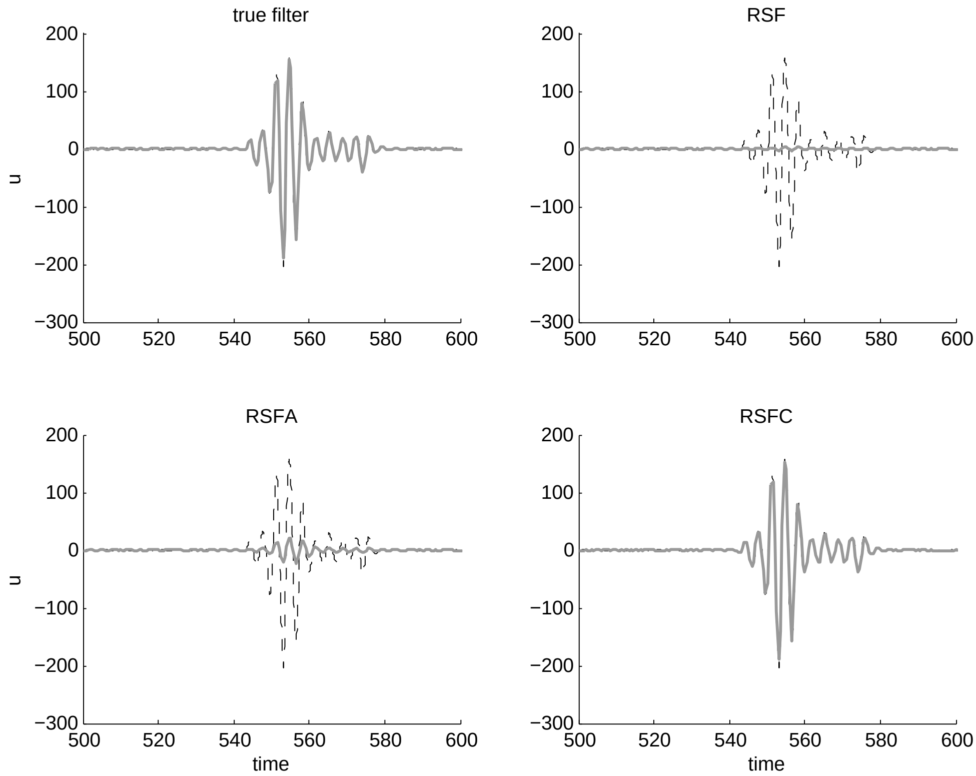

Figure 8. Filtered posterior estimates (grey) compared to the truth (black dashes) for regime II with $SNR^{-1} = 0.5$.

References

Anderson, B. & Moore, J. (1979), *Optimal filtering*, Prentice-Hall Englewood Cliffs, NJ.

Anderson, J. L. (2001), 'An ensemble adjustment Kalman filter for data assimilation', *Monthly Weather Review* **129**(12), 2884–2903.

Anderson, J. L. (2007), 'An adaptive covariance inflation error correction algorithm for ensemble filters', *Tellus A* **59**(2), 210–224.

Anderson, J. L. (2009), 'Spatially and temporally varying adaptive covariance inflation for ensemble filters', *Tellus A* **61**(2), 72–83.

Anderson, J. L. & Anderson, S. L. (1999), 'A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts', *Monthly Weather Review* **127**(12), 2741–2758.

Berglund, N. & Gentz, B. (2005), *Noise-Induced Phenomena in Slow-Fast Dynamical Systems*, Springer.

Boxler, P. (1989), 'A stochastic version of the centre manifold theorem', *Probab. Th. Rel. Fields* **83**, 509–545.

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

Branicki, M. & Majda, A. (2013), 'Fundamental limitations of polynomial chaos for uncertainty quantification in systems with intermittent instabilities', *Comm. Math. Sci.* **11**(1), 55–103.

Buizza, R., Houtekamer, P. L., Pellerin, G., Toth, Z., Zhu, Y. & Wei, M. (2005), 'A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems', *Monthly Weather Review* **133**(5), 1076–1097.

Castronovo, E., Harlim, J. & Majda, A. (2008), 'Mathematical test criteria for filtering complex systems: plentiful observations', *Journal of Computational Physics* **227**(7), 3678–3714.

Delsole, T. (2004), 'Stochastic models of quasigeostrophic turbulence', *Surveys in Geophysics* **25**, 107–149.

Ehrendorfer, M. (2007), 'A review of issues in ensemble-based Kalman filtering', *Meteorologische Zeitschrift* **16**(6), 795–818.

Evensen, G. (1994), 'Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics', *Journal of Geophysical Research* **99**(C5), 10143–10162.

Evensen, G. (2006), *Data Assimilation: The Ensemble Kalman Filter*, Springer, New York.

Fenichel, N. (1979), 'Geometric singular perturbation theory for ordinary differential equations', *Journal of Differential Equations* **31**(1), 53–98.

Friedland, B. (1969), 'Treatment of bias in recursive filtering', *IEEE Trans. Automat. Contr.* **AC-14**, 359–367.

Friedland, B. (1982), 'Estimating sudden changes of biases in linear dynamical systems', *IEEE Trans. Automat. Contr.* **AC-27**, 237–240.

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

Gershgorin, B., Harlim, J. & Majda, A. (2010*b*), 'Test models for improving filtering with model errors through stochastic parameter estimation', *J. Comput. Phys.* **229**(1), 1–31.

Givon, D., Kupferman, R. & Stuart, A. (2004), 'Extracting macroscopic dynamics: Model problems and algorithms', *Nonlinearity* **17**(6), R55–127.

Hamill, T. M. & Whitaker, J. S. (2011), 'What constrains spread growth in forecasts initialized from ensemble Kalman filters', *Monthly Weather Review* **139**, 117–131.

Harlim, J. (2011), 'Numerical strategies for filtering partially observed stiff stochastic differential equations', *Journal of Computational Physics* **230**(3), 744–762.

Harlim, J. & Majda, A. (2013), 'Test models for filtering with superparameterization', *Multiscale Model. Simul.* **11**(1), 282 – 307.

Houtekamer, P. L. & Mitchell, H. L. (1998), 'Data assimilation using an ensemble Kalman filter technique', *Monthly Weather Review* **126**(3), 796–811.

Houtekamer, P. L., Mitchell, H. L. & Deng, X. (2009), 'Model error representation in an operational ensemble Kalman filter', *Monthly Weather Review* **137**(7), 2126–2143.

Imkeller, P., Namachchivaya, N. S., Perkowski, N. & Yeong, H. C. (2012), 'A homogenization approach to multiscale filtering', *Procedia IUTAM* **5**(0), 34 – 45. IUTAM Symposium on 50 Years of Chaos: Applied and Theoretical.

Kalnay, E. (2002), *Atmospheric Modeling, Data Assimilation and Predictability*, Cambridge University Press, Cambridge.

Kang, E. & Harlim, J. (2012), 'Filtering partially observed multiscale systems with heterogeneous multiscale methods-based reduced climate models', *Monthly Weather Review* **140**(3), 860–873.

Khasminskii, R. (1968), 'On averaging principle for it stochastic differential equations', *Kybernetika, Chekhoslovakia (in Russian)* **4**(3), 260–279.

Kurtz, T. G. (1973), 'A limit theorem for perturbed operator semigroups with applications to random evolutions', *Journal of Functional Analysis* **12**(1), 55–67.

Law, K. & Stuart, A. (2012), 'Evaluating data assimilation algorithms', *Monthly Weather Review* **140**, 3757–3782.

Li, H., Kalnay, E. & Miyoshi, T. (2009), 'Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter', *Quarterly Journal of the Royal Meteorological Society* **135**(639), 523–533.

Majda, A. J., Franzke, C. & Khouider, B. (2008), 'An applied mathematics perspective on stochastic modelling for climate', *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **366**(1875), 2427–2453.

Majda, A. J. & Harlim, J. (2012), *Filtering Complex Turbulent Systems*, Cambridge University Press, Cambridge.

Majda, A. J., Timofeyev, I. & Vanden Eijnden, E. (1999), 'Models for stochastic climate prediction', *Proceedings of the National Academy of Sciences* **96**(26), 14687–14691.

Majda, A. J., Timofeyev, I. & Vanden Eijnden, E. (2001), 'A mathematical framework for stochastic climate models', *Communications on Pure and Applied Mathematics* **54**(8), 891–974.

Majda, A. J., Timofeyev, I. & Vanden-Eijnden, E. (2003), 'Systematic strategies for stochastic mode reduction in climate', *Journal of the Atmospheric Sciences* **60**(14), 1705–1722.

Mitchell, L. & Gottwald, G. A. (2012), 'Data assimilation in slow-fast systems using homogenized climate models', *Journal of the Atmospheric Sciences* **69**(4), 1359–1377.

Miyoshi, T. (2011), 'The Gaussign approach to adaptive covariance inflation and its implementation with the Local Eensemble Transform Kalman Filter', *Monthly Weather Review* **139**, 1519–1535.

Ng, G.-H. C., McLaughlin, D., Entekhabi, D. & Ahanin, A. (2011), 'The role of model dynamics in ensemble Kalman filter performance for chaotic systems', *Tellus A* **63**, 958–977.

Palmer, T. N. (2001), 'A nonlinear dynamical perspective on model error: A proposal for non-local stochastic-dynamic parametrization in weather and climate prediction models', *Quarterly Journal of the Royal Meteorological Society* **127**(572), 279–304.

Papanicolaou, G. (1976), 'Some probabilistic problems and methods in singular perturbations', *Rocky Mountain J. Math* **6**, 653–673.

Pavliotis, G. A. & Stuart, A. M. (2008), *Multiscale Methods: Averaging and Homogenization*, Springer, New York.

Pavliotis, G. & Stuart, A. (2007), 'Parameter estimation for multiscale diffusions', *Journal of Statistical Physics* **127**(4), 741–781.

Petrie, R. E. & Dance, S. L. (2010), 'Ensemble-based data assimilation and the localisation problem', *Weather* **65**(3), 65–69.

Roberts, A. J. (2008), 'Normal form transforms separate slow and fast modes in stochastic dynamical systems', *Physica A* **387**, 12–38.

Sandu, A., Constantinescu, E. M., Carmichael, G. R., Chai, T., Seinfeld, J. H. & Daescu, D. (2007), 'Localized ensemble kalman dynamics data assimilation for atmospheric chemistry', *Lecture Notes Comput. Sci.* **4487**, 1018–1490.

Zhang, F. (2011), Parameter estimation and model fitting of stochastic processes, Phd thesis, University of Warwick.