Preprint of: W. Lee and C. L. Farmer. 2014. "Data Assimilation by Conditioning of Driving Noise on Future Observations". IEEE Transactions on Signal Processing, 62 (15), pp 3887-3896.

Data Assimilation by Conditioning of Driving Noise on Future Observations

Wonjung Lee and Chris Farmer

***Abstract*—Conventional recursive filtering approaches, designed for quantifying the state of an evolving stochastic dynamical system with intermittent observations, use a sequence of (i) an uncertainty propagation step followed by (ii) a step where the associated data is assimilated using Bayes' rule. Alternatively, the order of the steps can be switched to (i) one step ahead data assimilation followed by (ii) uncertainty propagation. In this paper we apply this smoothing-based sequential filter to systems driven by random noise, however with the conditioning on future observation not only to the system variable but to the driving noise. Our research reveals that, for the nonlinear filtering problem, the conditioned driving noise is biased by a nonzero mean and in turn pushes forward the filtering solution in time closer to the true state when it drives the system. As a result our proposed method can yield a more accurate approximate solution for the state estimation problem.**

***Index Terms*—Bayesian statistics, Gaussian approximation filter, cubature measure**

I. Introduction

There are many examples in science and engineering where the state of a system has to be identified from a set of noisy observations [1], [2], [3], [4]. In general, this state estimation problem involves (i) a forward model governing the evolution of the underlying dynamical system and (ii) observational data associated with the system state. Let the unknown system state $\mathbf{x} \in \mathbb{R}^d$ satisfy

$$\textbf{Forward model} \quad \mathbf{x}_{n+1} = \Phi^n(\mathbf{x}_n, \xi_n), \quad \xi_n \sim \mathcal{N}(\mathbf{0}, \Gamma_n) \tag{1}$$

where $n \in \mathbb{N} \cup \{0\}$ labels the time step, $\xi_n \in \mathbb{R}^D$ is an independent and identically distributed (i.i.d.) Gaussian driving noise and $\mathbf{0}$ denotes a zero vector (or later, a zero matrix). Let the data $\mathbf{y}_n \in \mathbb{R}^{d'}$, associated with $\mathbf{x}_n$, satisfy

$$\textbf{Observation} \quad \mathbf{y}_n = \phi^n(\mathbf{x}_n) + \eta_n, \quad \eta_n \sim \mathcal{N}(\mathbf{0}, R_n) \tag{2}$$

for a measurement function ϕ^n and i.i.d. Gaussian η_n. Here $\mathbf{x}_0$, $\{\xi_n\}_{n\geq 0}$ and $\{\eta_n\}_{n\geq 0}$ are independent variables. Note $\mathbf{x}_n$ and ξ_n are also independent because ξ_n is i.i.d.

One normally combines the process equation (1) and the measurement equation (2) to form an effective solution of the state estimation problem. Let $\mathbf{x}_{n|n'} \equiv \mathbf{x}_n | Y_{n'}$ be the random vector conditioned on the collection of observations $Y_{n'} \equiv \{\mathbf{y}_1, \cdots, \mathbf{y}_{n'}\}$. It is called *smoothing* when seeking the probability distribution of $\mathbf{x}_{n|n'}$ when $n < n'$, *filtering* when $n = n'$, and *prediction* when $n > n'$. Given Eqs. (1), (2) and the probability distribution of $\mathbf{x}_0$, sequential filtering calculates the probability distribution of $\mathbf{x}_{n|n}$ to characterise the system state $\mathbf{x}_n$ for $n \geq 1$.

Wonjung Lee and Chris Farmer are with the Oxford Centre for Collaborative Applied Mathematics (OCCAM) in the Mathematical Institute, University of Oxford, Oxford, U.K. (email:leew@maths.ox.ac.uk; farmer@maths.ox.ac.uk).

$$\begin{array}{ccc} \mathcal{X}_{n|n} = \left[\mathbf{x}_{n|n}, \xi_{n|n}\right] & \longrightarrow & \mathbf{x}_{n+1|n} \\ \Downarrow & & \Downarrow \\ \mathcal{X}_{n|n+1} = \left[\mathbf{x}_{n|n+1}, \xi_{n|n+1}\right] & \longrightarrow & \mathbf{x}_{n+1|n+1} \end{array}$$

Fig. 1. Conventional filter ($\rightarrow$, $\Rightarrow$) and Noise-Smoothing filter ($\Rightarrow$, $\rightarrow$).

For practical reasons, online estimation via a recursive method is desired. The conventional way to achieve this real-time filtering algorithm is to alternate application of (i) $\mathbf{x}_{n|n} \rightarrow \mathbf{x}_{n+1|n}$; the uncertainty quantification (UQ) or time update and (ii) $\mathbf{x}_{n+1|n} \Rightarrow \mathbf{x}_{n+1|n+1}$; the data assimilation (DA) or measurement update, in a sequential fashion. We here use the notation $\rightarrow$ to increase the first index by one or one step UQ and the notation $\Rightarrow$ to increase the second index by one or one step DA. We call any filtering algorithm following the conventional approach, a *conventional filter*.

Examples of conventional filters include the Kalman filter [5], extended Kalman filter [6], unscented Kalman filter [7], cubature Kalman filter [8] and Gaussian particle filter [9]. All of these filters characterise the probability distribution of the filtering solution through its mean and covariance. There are other kinds of filters, called Gaussian sum filters, where the filtering solution is approximated using multiple positively weighted Gaussian kernels rather than a single one (see for example [10], [11]). The ensemble Kalman filter [4] and the bootstrap filter and its variants [3], [12], [13], [14] are sequential Monte Carlo methods where a number of Dirac masses are employed to approximate the probability distribution. Some of these Monte Carlo methods can be viewed as specific cases of one version of Gaussian sum filters [11].

While most of the prevailing filters fall into the conventional filtering category, it is also possible to solve the sequential filtering problem in the switched manner, i.e., successive application of a measurement update $\mathbf{x}_{n|n} \Rightarrow \mathbf{x}_{n|n+1}$ to achieve smoothing followed by a time update $\mathbf{x}_{n|n+1} \rightarrow \mathbf{x}_{n+1|n+1}$ [15]. In this paper we go further by conditioning the driving noise ξ_n on future observations $\mathbf{y}_{n+1}$ to form a new class of filtering algorithms named *noise-smoothing filters*. In order to achieve this, the $(d + D)$-dimensional augmented system $\mathcal{X}_n \equiv \begin{bmatrix} \mathbf{x}_n \\ \xi_n \end{bmatrix}$ is introduced. Let $\xi_{n|n'} \equiv \xi_n | Y_{n'}$ and

$\mathcal{X}_{n|n'} \equiv \begin{bmatrix} \mathbf{x}_{n|n'} \\ \xi_{n|n'} \end{bmatrix}$ be conditioned random vectors. Our filtering strategy is to alternate application of $\mathcal{X}_{n|n} \Rightarrow \mathcal{X}_{n|n+1}$ and $\mathcal{X}_{n|n+1} \rightarrow \mathbf{x}_{n+1|n+1}$. The following two different sequential methods:

1) Conventional filter : $\mathcal{X}_{n|n} \rightarrow \mathbf{x}_{n+1|n} \Rightarrow \mathbf{x}_{n+1|n+1}$
2) Noise-Smoothing filter : $\mathcal{X}_{n|n} \Rightarrow \mathcal{X}_{n|n+1} \rightarrow \mathbf{x}_{n+1|n+1}$

are illustrated in Fig. 1.

It is worth remarking that our noise-smoothing filter can yield an accurate approximate solution for the state estimation problem when a possible bias of the conditioned noise $\xi_{n|n+1}$, in the sense $\mathbb{E}(\xi_{n|n+1}) \neq \mathbf{0}$ with $\mathbb{E}(\cdot)$ denoting the statistical average, makes the approximation of $\mathbf{x}_{n+1|n+1}$ closer to $\mathbf{y}_{n+1}$ through the UQ step $\mathbf{x}_{n+1|n+1} = \Phi^n(\mathbf{x}_{n|n+1}, \xi_{n|n+1})$. This is in contrast to the cases of conventional filters and traditional smoothing-based filters for which $\xi_{n|n} = \xi_n$, satisfying $\mathbb{E}(\xi_{n|n}) = \mathbf{0}$, is the driving noise. Note that, by utilising future observations to push forward the filtering solution in time, our noise-smoothing filter has a similarity with sequential Monte Carlo importance sampling [16], [17], [18], [19], [20]. However $\mathbf{y}_{n+1}$ is not directly, but implicitly involved via $\xi_{n|n+1}$ in noise-smoothing filters, and the algorithm does not require an auxiliary step corresponding to the importance reweighting in sequential importance sampling. Thus noise-smoothing filters can be viewed as an automatic form of importance sampling.

One way of building a noise-smoothing filter is to mimic an existing conventional filter, i.e., to let its UQ and DA methods be basically of the same kind as those of a conventional filter, but with the ordering of UQ and DA reversed. We note that there are many conventional filters that adopt a sum of Gaussian kernels (or Dirac masses) to approximate the conditioned probability distribution, and although these may ultimately be the approximations of choice, in this paper concern is confined to the problem of approximation using a unimodal distribution so that the mean and covariance would be used to characterise the filtering solution. This is because the aim is to develop simple but efficient filtering methods, similar to traditional methods, but with enhanced accuracy. However, we hope that any success in this aim will help direct future efforts toward the wider aim of developing a rigorous and convergent method, perhaps using a Gaussian sum approximation.

The rest of the paper is organised as follows. We develop a number of conventional filters based on traditional filters in Section II and formulate corresponding noise-smoothing filters in Section III. With the help of the test problems gathered in Section IV, numerical simulations are performed to examine the accuracy of noise-smoothing filters in Section V. We conclude our results and discuss future work in Section VI.

II. Conventional Filtering

By combining the UQ methods presented in subsection II-A and the DA methods presented in subsection II-B, a number of conventional filters are developed in subsection II-C. The algorithm provides the mapping of the first two moments.

A. Time Update ($\mathcal{X}_{n|n} \rightarrow \mathbf{x}_{n+1|n}$)

From now on we use the notation $\mathbf{x}_{n+1} = \Phi^n(\mathcal{X}_n)$ in place of Eq. (1). In order to quantify the uncertainty propagation, the following two methods would be adopted to approximate the mean and covariance of $\mathbf{x}_{n+1|n'} = \Phi^n(\mathcal{X}_{n|n'})$.

1) Linear approximation: Let the mean and covariance of $\mathbf{x}_{n|n'}$ be denoted by $\bar{\mathbf{x}}_{n|n'}$ and $\mathbf{C}_{n|n'}$. Let the mean and covariance of $\mathcal{X}_{n|n'}$ be denoted by $\bar{\mathcal{X}}_{n|n'}$ and $\mathcal{C}_{n|n'}$. Let $\nabla_i = \partial_i$ be the gradient operator. The first order Taylor approximation

$$\Phi^n(\mathcal{X}_{n|n'}) \simeq \Phi^n(\bar{\mathcal{X}}_{n|n'}) + \nabla\Phi^n|_{\bar{\mathcal{X}}_{n|n'}}(\mathcal{X}_{n|n'} - \bar{\mathcal{X}}_{n|n'})$$

leads to

$$\begin{aligned} \bar{\mathbf{x}}_{n+1|n'} &\simeq \Phi^n(\bar{\mathcal{X}}_{n|n'}), \\ \mathbf{C}_{n+1|n'} &\simeq \nabla\Phi^n|_{\bar{\mathcal{X}}_{n|n'}}\mathcal{C}_{n|n'}\,(\nabla\Phi^n|_{\bar{\mathcal{X}}_{n|n'}})^T \end{aligned} \tag{3}$$

where the superscript T denotes the matrix transpose. An application of Eq. (3) with $n' = n$, $\bar{\mathcal{X}}_{n|n} = \begin{bmatrix} \bar{\mathbf{x}}_{n|n} \\ \mathbf{0} \end{bmatrix}$ and $\mathcal{C}_{n|n} = \begin{bmatrix} \mathbf{C}_{n|n} & \mathbf{0} \\ \mathbf{0} & \Gamma_n \end{bmatrix}$ produces one prediction algorithm for a conventional filter.

We here mention that the extended Kalman filter utilises the linearisation of Eq. (1) with respect to $\mathbf{x}_{n|n}$ and the known statistics of ξ_n to obtain a similar equation to Eq. (3) where $n' = n$.

2) Point-based approximation: Let δ_x denote a Dirac mass centered at x. Let $\sum_j \Lambda^j_{n|n'}\delta_{\mathcal{X}^j_{n|n'}}$ with $\Lambda^j_{n|n'} > 0$ be a normalised measure approximating the probability distribution of $\mathcal{X}_{n|n'}$. To construct this point-based approximation of $\mathcal{X}_{n|n'}$, one can use either (i) cubature measure supported on deterministically placed points or (ii) empirical measure which is a set of equally weighted Dirac masses supported on random draws. Recall that a set of positively weighted Dirac masses is called a cubature measure of degree r with respect to the given probability distribution provided the moments of these two measures agree with one another up to total degree r [21].

Taking $\sum_j \Lambda^j_{n|n'}\delta_{\Phi^n(\mathcal{X}^j_{n|n'})}$ as one approximation of the probability distribution of $\Phi^n(\mathcal{X}_{n|n'})$, we obtain

$$\begin{aligned} &\bar{\mathbf{x}}_{n+1|n'} \simeq \sum_j \Lambda^j_{n|n'}\Phi^n(\mathcal{X}^j_{n|n'}), \\ &\mathbf{C}_{n+1|n'} \simeq \\ &\quad \sum_j \Lambda^j_{n|n'}\left(\Phi^n(\mathcal{X}^j_{n|n'}) - \bar{\mathbf{x}}_{n+1|n'}\right)\left(\Phi^n(\mathcal{X}^j_{n|n'}) - \bar{\mathbf{x}}_{n+1|n'}\right)^T \end{aligned} \tag{4}$$

which provides one prediction algorithm for a conventional filter when $n' = n$.

We here mention that the prediction methods in the cubature Kalman filter [8] and Gaussian particle filter [9] make use of cubature measure and empirical measure approximating the distribution of $\mathbf{x}_{n|n}$ respectively, in addition to the known statistics of ξ_n, to obtain a similar equation to Eq (4) where $n' = n$.

We also mention that there are more sophisticated approaches than Eq. (4) that use the point-based approximation for the uncertainty propagation [22], but we will not apply them in this paper as our primary concern is not to develop accurate conventional filters, but to compare the performances of the noise-smoothing filters and the elementary conventional filters upon which the noise-smoothing filters are modelled.

B. Measurement Update ($\mathbf{x}_{n+1|n} \to \mathbf{x}_{n+1|n+1}$)

Bayes' rule

$$\mathbb{P}(X|Y) = \frac{\mathbb{P}(X,Y)}{\mathbb{P}(Y)} \tag{5}$$

for random vectors X and Y can be employed to perform the data assimilation. Eq. (5) implies that if X and Y are jointly Gaussian, i.e., $Z = \begin{bmatrix} X \\ Y \end{bmatrix}$ is Gaussian with mean $\begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix}$ and covariance $\begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix}$, then the conditioned variable $X|Y$ with $Y = y$ is Gaussian with mean and covariance given by

$$\begin{aligned} \bar{x}' &= \bar{x} + \Sigma_{xy}\Sigma_{yy}^{-1}(y - \bar{y}), \\ \Sigma_{xx}' &= \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}, \end{aligned} \tag{6}$$

respectively [2].

If both $\mathbf{x}_{n+1|n}$ and $\phi^n(\mathbf{x}_{n+1|n})$ are Gaussian, one can apply Eq. (6) with $X = \mathbf{x}_{n+1|n}$, $Y = \phi^n(\mathbf{x}_{n+1|n}) + \eta_{n+1}$ and $y = \mathbf{y}_{n+1}$ to obtain the first two moments of $\mathbf{x}_{n+1|n}|\mathbf{y}_{n+1} = \mathbf{x}_{n+1|n+1}$. However $\phi^n(\mathbf{x}_{n+1})$ is not a Gaussian, unless ϕ^n is a linear function and $\mathbf{x}_{n+1}$ is Gaussian. As in the case of the time update, we consider two approximations of $\phi^n(\mathbf{x}_{n+1})$.

1) Linear approximation: The Taylor approximation of

$$\phi^n(\mathbf{x}_{n+1|n}) \simeq \phi^n(\bar{\mathbf{x}}_{n+1|n}) + \nabla\phi^n|_{\bar{\mathbf{x}}_{n+1|n}}(\mathbf{x}_{n+1|n} - \bar{\mathbf{x}}_{n+1|n}),$$

would be used in place of $\phi^n(\mathbf{x}_{n+1|n})$. In this case, we use Eq. (6) to obtain

$$\begin{aligned} \bar{\mathbf{x}}_{n+1|n+1} &\simeq \bar{\mathbf{x}}_{n+1|n} + G_{\mathbf{x}}\left(\mathbf{y}_{n+1} - \phi^n(\bar{\mathbf{x}}_{n+1|n})\right), \\ \mathbf{C}_{n+1|n+1} &\simeq \mathbf{C}_{n+1|n} - G_{\mathbf{x}}\nabla\phi^n|_{\bar{\mathbf{x}}_{n+1|n}}\mathbf{C}_{n+1|n} \end{aligned} \tag{7}$$

where

$$\begin{aligned} G_{\mathbf{x}} \equiv\ & \mathbf{C}_{n+1|n}(\nabla\phi^n|_{\bar{\mathbf{x}}_{n+1|n}})^T \\ & \left(\nabla\phi^n|_{\bar{\mathbf{x}}_{n+1|n}}\mathbf{C}_{n+1|n}(\nabla\phi^n|_{\bar{\mathbf{x}}_{n+1|n}})^T + R_{n+1}\right)^{-1}. \end{aligned}$$

2) Point-based approximation: Let a normalised measure $\sum_j \lambda^j_{n+1|n}\delta_{\mathbf{x}^j_{n+1|n}}$ with $\lambda^j_{n+1|n} > 0$ be distributed according to the probability distribution of $\mathbf{x}_{n+1|n}$, then $\sum_j \lambda^j_{n+1|n}\delta_{\phi^n(\mathbf{x}^j_{n+1|n})}$ is one approximation for the probability distribution of $\phi^n(\mathbf{x}_{n+1|n})$. We approximate the mean and covariance of $\phi^n(\mathbf{x}_{n+1|n})$ by those from the point-based approximation. In this case, we use Eq. (6) to obtain

$$\begin{aligned} \bar{\mathbf{x}}_{n+1|n+1} &\simeq \bar{\mathbf{x}}_{n+1|n} + L_{\mathbf{x}}\left(\mathbf{y}_{n+1} - \mathbf{z}\right), \\ \mathbf{C}_{n+1|n+1} &\simeq \mathbf{C}_{n+1|n} - L_{\mathbf{x}}P_{\mathbf{xz}}^T \end{aligned} \tag{8}$$

where

$$\begin{aligned} L_{\mathbf{x}} &\equiv P_{\mathbf{xz}}\left(P_{\mathbf{zz}} + R_{n+1}\right)^{-1}, \\ \mathbf{z} &\equiv \sum_j \lambda^j_{n+1|n}\phi^n(\mathbf{x}^j_{n+1|n}), \\ P_{\mathbf{xz}} &\equiv \sum_j \lambda^j_{n+1|n}\left(\mathbf{x}^j_{n+1|n} - \sum_j \lambda^j_{n+1|n}\mathbf{x}^j_{n+1|n}\right) \\ &\qquad \left(\phi^n(\mathbf{x}^j_{n+1|n}) - \mathbf{z}\right)^T, \\ P_{\mathbf{zz}} &\equiv \sum_j \lambda^j_{n+1|n}\left(\phi^n(\mathbf{x}^j_{n+1|n}) - \mathbf{z}\right)\left(\phi^n(\mathbf{x}^j_{n+1|n}) - \mathbf{z}\right)^T. \end{aligned}$$

In addition to the above two methods, we consider an algorithm that does not require an approximation for the mean and covariance of $\phi^n(\mathbf{x}_{n+1})$. It is motivated by the variational data assimilation widely used in weather forecasting [23].

3) Variational approximation: Let the probability density function of a Gaussian, centered at $\mathbf{y}_{n+1}$ with covariance R_{n+1}, be denoted by $\Theta(\mathbf{y}_{n+1}, R_{n+1})$. Eq. (5) then implies that $\mathbb{P}(X|Y) = \mathbb{P}(X)\mathbb{P}(Y|X)/\mathbb{P}(Y)$ and we approximate

$$\begin{aligned} \mathbb{P}(\mathbf{x}_{n+1|n+1}) &\propto \mathbb{P}(\mathbf{x}_{n+1|n})\,\Theta\left(\mathbf{y}_{n+1} - \phi^n(\mathbf{x}_{n+1}), R_{n+1}\right) \\ &\simeq \exp(-\mathbf{J}_{n+1|n+1}(\mathbf{x_{n+1}})) \end{aligned}$$

where the misfit function $\mathbf{J}$ is given by

$$\begin{aligned} &\mathbf{J}_{n+1|n+1}(\mathbf{x}_{n+1}) \\ &\equiv \frac{1}{2}\left\{\| \mathbf{x}_{n+1} - \bar{\mathbf{x}}_{n+1|n} \|^2_{\mathbf{C}_{n+1|n}} + \| \mathbf{y}_{n+1} - \phi^n(\mathbf{x}_{n+1}) \|^2_{R_{n+1}}\right\}. \end{aligned} \tag{9}$$

Here the notation $\| X \|^2_\Sigma \equiv X^T\Sigma^{-1}X$ is used for a positive definite quadratic form with matrix Σ. Assuming a unimodal distribution of $\mathbf{x}_{n+1|n}$, we approximate $\bar{\mathbf{x}}_{n+1|n+1}$ by the minimiser of $\mathbf{J}_{n+1|n+1}$ and we approximate $\mathbf{C}_{n+1|n+1}$ by the inverse of the Hessian of the misfit function at $\bar{\mathbf{x}}_{n+1|n+1}$,

$$\begin{aligned} \bar{\mathbf{x}}_{n+1|n+1} &\simeq \text{minimiser of Eq. (9)}, \\ \mathbf{C}_{n+1|n+1} &\simeq \left(\nabla\nabla\mathbf{J}_{n+1|n+1}|_{\bar{\mathbf{x}}_{n+1|n+1}}\right)^{-1}, \end{aligned} \tag{10}$$

where $\nabla\nabla_{ij} = \partial_i\partial_j$.

C. Construction of Conventional Filters

We can choose one from the two UQ methods (Eqs. (3), (4) with $n' = n$) and independently one from the three DA methods (Eqs. (7), (8), (10)) to construct a conventional filter. In this paper we intend to make the UQ and DA methods consistent, if possible, and not to simultaneously use the non-point-based algorithm and point-based algorithm. As a result, we define the linear conventional filter (**LCF**) as the combination of UQ with a linear approximation and DA with a linear approximation; the variational conventional filter (**VCF**) as the combination of UQ with linear approximation and DA with a variational approximation; the cubature conventional filter (**CCF**) and the particle conventional filter (**PCF**) as the combination of UQ with a point-based approximation and DA with a point-based approximation, for which cubature measure and empirical measure are employed respectively.

III. Noise-Smoothing Filters

By combining the DA methods presented in subsection III-A and the UQ methods presented in subsection III-B, we develop a number of noise-smoothing filters in subsection III-C.

A. Measurement Update ($\mathcal{X}_{n|n} \Rightarrow \mathcal{X}_{n|n+1}$)

The methodology for the measurement update in a noise-smoothing filter is the same as in the case of conventional filtering, except for the use of $\mathcal{X}_{n|n'}$ in place of $\mathbf{x}_{n+1|n'}$ (hence $\bar{\mathcal{X}}_{n|n'}$ and $\mathcal{C}_{n|n'}$ in place of $\bar{\mathbf{x}}_{n+1|n'}$ and $\mathbf{C}_{n+1|n'}$ respectively) and $\Psi^n \equiv \phi^n \circ \Phi^n$ (due to $\mathbf{y}_{n+1} = \Psi^n(\mathcal{X}_n) + \eta_{n+1}$) in place of ϕ^n. We mention that Ψ^n in a noise-smoothing filter might be a nonlinear function even when ϕ^n is linear.

1) Linear approximation: As with Eq. (7), we obtain

$$\begin{aligned}\bar{\mathcal{X}}_{n|n+1} &\simeq \bar{\mathcal{X}}_{n|n} + G_{\mathcal{X}}\left(\mathbf{y}_{n+1} - \Psi^n(\bar{\mathcal{X}}_{n|n})\right), \\ \mathcal{C}_{n|n+1} &\simeq \mathcal{C}_{n|n} - G_{\mathcal{X}} \nabla\Psi^n|_{\bar{\mathcal{X}}_{n|n}} \mathcal{C}_{n|n}\end{aligned} \tag{11}$$

where

$$G_{\mathcal{X}} \equiv \mathcal{C}_{n|n}(\nabla\Psi^n|_{\bar{\mathcal{X}}_{n|n}})^T \left(\nabla\Psi^n|_{\bar{\mathcal{X}}_{n|n}}\mathcal{C}_{n|n}(\nabla\Psi^n|_{\bar{\mathcal{X}}_{n|n}})^T + R_{n+1}\right)^{-1}.$$

2) Point-based approximation: Recall that $\sum_j \Lambda^j_{n|n'} \delta_{\mathcal{X}^j_{n|n'}}$ is distributed according to the probability distribution of $\mathcal{X}_{n|n'}$. As with Eq. (8), we obtain

$$\begin{aligned}\bar{\mathcal{X}}_{n|n+1} &\simeq \bar{\mathcal{X}}_{n|n} + L_{\mathcal{X}}\left(\mathbf{y}_{n+1} - \mathcal{Z}\right), \\ \mathcal{C}_{n|n+1} &\simeq \mathcal{C}_{n|n} - L_{\mathcal{X}} P^T_{\mathcal{X}\mathcal{Z}}\end{aligned} \tag{12}$$

where

$$\begin{aligned}L_{\mathcal{X}} &\equiv P_{\mathcal{X}\mathcal{Z}}\left(P_{\mathcal{Z}\mathcal{Z}} + R_{n+1}\right)^{-1}, \\ \mathcal{Z} &\equiv \sum_j \Lambda^j_{n|n'} \Psi^n(\mathcal{X}^j_{n|n}), \\ P_{\mathcal{X}\mathcal{Z}} &\equiv \sum_j \Lambda^j_{n|n'}\left(\mathcal{X}^j_{n|n} - \sum_j \Lambda^j_{n|n'}\mathcal{X}^j_{n|n}\right)\left(\Psi^n(\mathcal{X}^j_{n|n}) - \mathcal{Z}\right)^T, \\ P_{\mathcal{Z}\mathcal{Z}} &\equiv \sum_j \Lambda^j_{n|n'}\left(\Psi^n(\mathcal{X}^j_{n|n}) - \mathcal{Z}\right)\left(\Psi^n(\mathcal{X}^j_{n|n}) - \mathcal{Z}\right)^T.\end{aligned}$$

3) Variational approximation: Let

$$\begin{aligned}&\mathcal{J}_{n|n+1}(\mathcal{X}_n) \\ &= \frac{1}{2}\left\{\| \mathcal{X}_n - \bar{\mathcal{X}}_{n|n} \|^2_{\mathcal{C}_{n|n}} + \| \mathbf{y}_{n+1} - \Psi^n(\mathcal{X}_n) \|^2_{R_{n+1}}\right\}\end{aligned} \tag{13}$$

be the misfit function. As with Eq. (10), we obtain

$$\begin{aligned}\bar{\mathcal{X}}_{n|n+1} &\simeq \text{minimiser of Eq. (13)}, \\ \mathcal{C}_{n|n+1} &\simeq \left(\nabla\nabla\mathcal{J}_{n|n+1}|_{\bar{\mathcal{X}}_{n|n+1}}\right)^{-1}.\end{aligned} \tag{14}$$

B. Time Update ($\mathcal{X}_{n|n+1} \rightarrow \mathbf{x}_{n+1|n+1}$)

Having obtained $\bar{\mathcal{X}}_{n|n+1}$ and $\mathcal{C}_{n|n+1}$, the moment approximations of $\mathcal{X}_{n|n+1}$, we now apply Eq. (3) or Eq. (4) with $n' = n+1$ to achieve the UQ of the noise-smoothing filter. Notice that, unlike the case of $\mathcal{X}_{n|n}$ with $\bar{\mathcal{X}}_{n|n} = \begin{bmatrix} \bar{\mathbf{x}}_{n|n} \\ \mathbf{0} \end{bmatrix}$ and $\mathcal{C}_{n|n} = \begin{bmatrix} \mathbf{C}_{n|n} & \mathbf{0} \\ \mathbf{0} & \Gamma_n \end{bmatrix}$, the driving noise of the conditioned variable $\mathcal{X}_{n|n+1}$ is biased in the sense $\mathbb{E}(\xi_{n|n+1}) \neq \mathbf{0}$ and $\mathcal{C}_{n|n+1}$ is not a block diagonal matrix due to a non-vanishing correlation between $\mathbf{x}_{n|n+1}$ and $\xi_{n|n+1}$.

C. Construction of Noise-Smoothing Filters

We can choose one from the three DA methods (Eqs. (11), (12), (14)) and independently one from the two UQ methods (Eqs. (3), (4) with $n' = n+1$) to construct a noise-smoothing filter. In this paper we intend to develop a noise-smoothing filter modelled upon a given conventional filter or to make a one-to-one correspondence between conventional filters and noise-smoothing filters. As a result, we define the linear noise-smoothing filter (**LNSF**) as the combination of DA with a linear approximation and UQ with a linear approximation; the variational noise-smoothing filter (**VNSF**) as the combination of DA with a variational approximation and UQ with a linear approximation; the cubature noise-smoothing filter (**CNSF**) and the particle noise-smoothing filter (**PNSF**) as the combination of DA with a point-based approximation and UQ with a point-based approximation, for which cubature measure and empirical measure are employed respectively.

We mention that LNSF, VNSF, CNSF and PNSF correspond to LCF, VCF, CCF and PCF, respectively, and vice versa. We also mention that the computational difference between two corresponding conventional and noise-smoothing filters lies at the DA step. Because the noise-smoothing filters perform conditioning of $(d+D)$-dimensional augmented variable $\mathcal{X}$, $\mathcal{X}_{n|n} \Rightarrow \mathcal{X}_{n|n+1}$, each individual's computational cost is in general more expensive than that of the corresponding conventional filter where the state vector $\mathbf{x}$ in $\mathbf{x}_{n+1|n} \Rightarrow \mathbf{x}_{n+1|n+1}$ is d-dimensional. In case of the linear system with linear measurement function, LCF and LNSF become the Kalman filter and the Kalman smoother. Their computational costs are determined by Eq. (7) and Eq. (11).

IV. Practical Implementation

In this section some practical issues, encountered while implementing noise-smoothing filters, are resolved.

1) In the case that the forward model is derived from a discrete-time approximation of the stochastic differential equation,

$$d\mathbf{x}(t) = b(t, \mathbf{x}(t))dt + s(t, \mathbf{x}(t))\, dB(t), \tag{15}$$

where $b \in \mathbb{R}^d$ is the drift, $s \in \mathbb{R}^{d \times N}$ is the volatility and $B = (B_1, \cdots, B_N)$ is the set of independent Brownian motions, describing the evolution of the underlying system to be estimated, one can construct the forward model of Eq. (1) as follows. Let $\delta t > 0$ be the numerical simulation time step and let the observations arrive at the times $\Delta t = M \times \delta t$. A finite difference approximation of Eq. (15) using the Euler-Maruyama or Milstein method [24] yields

$$\mathbf{x}_{n,m+1} = \mathcal{F}^{n,m}(\mathbf{x}_{n,m}, w_{n,m}), \quad w_{n,m} \sim \mathcal{N}(\mathbf{0}, Q_{n,m}) \tag{16}$$

where $\mathbf{x}_{n,m}$ denotes an approximation of $\mathbf{x}(n\Delta t+m\delta t)$. The repeated application of Eq. (16) from $m = 0$ to $m = M-1$ defines $\Phi^n(\cdot)$ of the forward model, the mapping from $\mathbf{x}_n = \mathbf{x}_{n,0}$ to $\mathbf{x}_{n+1} = \mathbf{x}_{n,M}$, along with the augmented vector $\xi_n = \begin{bmatrix} w_{n,0} \\ \cdots \\ w_{n,M-1} \end{bmatrix}$ and the block diagonal matrix $\Gamma_n = \begin{bmatrix} Q_{n,0} & & \mathbf{0} \\ & \cdots & \\ \mathbf{0} & & Q_{n,M-1} \end{bmatrix}$.

2) Let $\sum_j \omega_j \delta_{x^j}$ be a cubature measure or an empirical measure approximating the standard Gaussian. Then $\sum_j \omega_j \delta_{m+Sx^j}$, for which S satisfies $C = SS^T$, becomes an approximation for $\mathcal{N}(m, C)$.
3) Some cubature formulae with respect to the standard Gaussian can be found in [21], [25], [26]. In one dimension, a cubature measure is more commonly referred to as a quadrature measure. A general multi-dimensional Gaussian cubature can be constructed via the tensor product of a quadrature formula [8]. Using Gauss-Hermite quadrature with support size $s = (r+1)/2$ for degree r, one can develop a cubature formula of degree r with respect to a k-dimensional standard Gaussian whose support size is s^k. Because the computational cost increases as the support size of the Dirac masses increases, it is important to use cubature measure supported on a smaller set. For the numerical simulations performed in the next section we use the standard Gaussian cubature formula of degree 3 and 5 introduced in [27], whose support size is $2k$ and $2k^2+1$, respectively.
4) The Broyden-Fletcher-Goldfarb-Shanno (BFGS) iterative method [28], [29], is used to solve the nonlinear optimisation problem for the variational Gaussian approximation. Numerical derivatives are employed as the dimension of our test cases is low. More generally an adjoint derivative method could be used for greater efficiency.

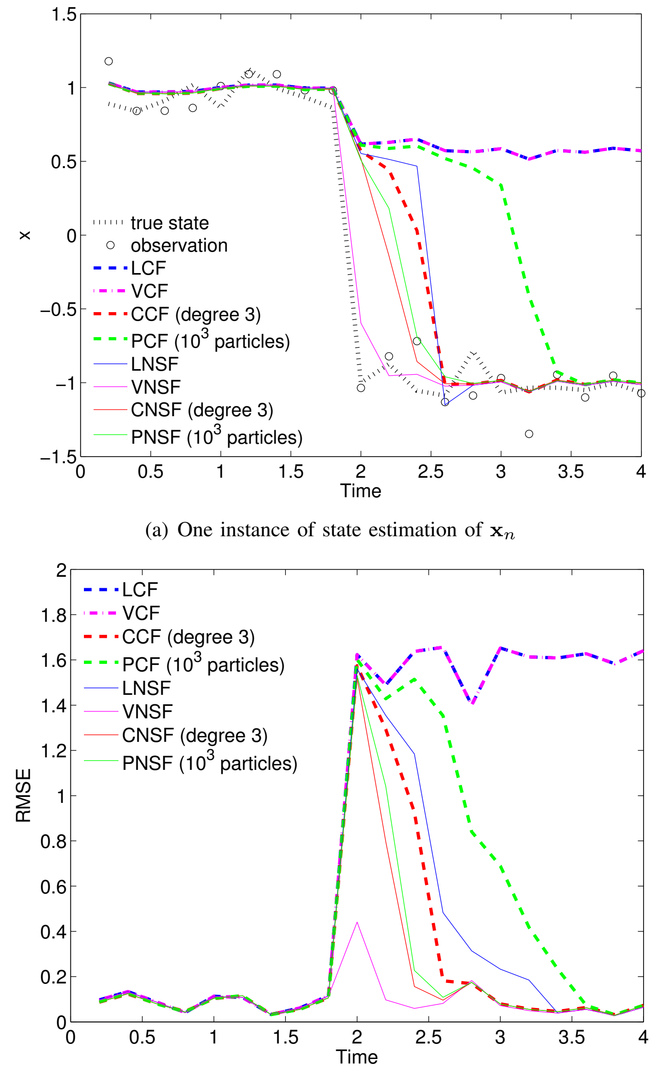
V. Numerical Simulations

In this section the feasibility of our noise-smoothing filters is investigated. The performances of LNSF, VNSF, CNSF and PNSF are compared with those of LCF, VCF, CCF and PCF. The metric used to compare the performance of various filters is the root mean square error (RMSE). The RMSE between $A = \{A_i\}_{i=1}^N$ and $B = \{B_i\}_{i=1}^N$ is defined by

$$\text{RMSE}(A, B) = \sqrt{\tfrac{1}{N}\sum_{i=1}^N |A_i - B_i|^2}$$

where A_i and B_i are vectors. The index i will specify either simulation number or time step.

The following comparison study uses various examples with different starting states and problem data. In the examples we studied, we see the noise-smoothing filters generally yield more accurate estimates than the corresponding conventional filters. The test examples consist of (i) a bistable system (subsection V-A), (ii) a prototypical chaotic dynamical system (subsection V-B) and (iii) a target tracking problem (subsection V-C).



(a) One instance of state estimation of $\mathbf{x}_n$

(b) The average of 50 independent state estimates

Fig. 2. The performances of various filters applied to the bistable system with the identity measurement function.

A. Bistable System

We consider the one-dimensional differential equation

$$d\mathbf{x}(t) = \beta\mathbf{x}(1-\mathbf{x}^2)dt + \sigma dB(t), \quad \beta > 0. \tag{17}$$

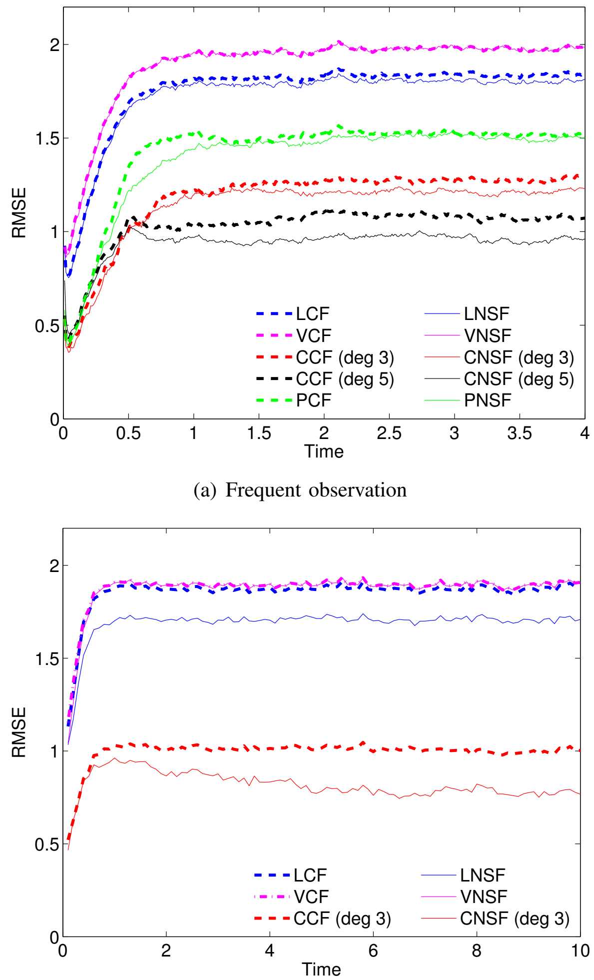
The deterministic equation with $\sigma = 0$ has two stable equilibria, -1 and $+1$, and one unstable equilibrium, 0. In this case the process $\mathbf{x}(t)$ is distributed around one of the stable equilibria. The stochastic system ($\sigma \neq 0$) however shows sudden transitions between the two stable equilibria due to the presence of random perturbations [30]. The Euler approximation of Eq. (17) is used to produce

$$\mathbf{x}_{n,m+1} = \mathbf{x}_{n,m} + \delta t \times \beta\mathbf{x}_{n,m}(1-\mathbf{x}_{n,m}^2) + \mathcal{N}(0, \sigma^2\delta t) \tag{18}$$

which corresponds to Eq. (16).

We first study the case for which the measurement function is the identity function, i.e.,

$$\phi^n(\mathbf{x}) = \mathbf{x}. \tag{19}$$



(a) Frequent observation

(b) Sparse observation

Fig. 3. The performances of various filters applied to the bistable system with the shifted quadratic measurement function. For PCF and PNSF, empirical measures consisting of 10^3 random samples are used (Top). The CCF and CNSF with cubature degree 5 are very similar to ones with degree 3 respectively and not depicted (Bottom).

For the system parameters $\beta = 10$ and $\sigma = 0.5$, a single realisation of Eq. (17) starting at $\mathbf{x}(0) = 0.8$ is simulated using Eq. (18) with $\delta t = 0.01$ and regarded as the true state. We see this trajectory has a jump from 1 to -1 at $t = 2.0$ (see Fig. 2(a)). We perform 50 independent numerical approximations to estimate the true state. In each case the observational data is generated with $R_n = 0.03$ at every $\Delta t = 20 \times \delta t = 0.2$. We then apply the conventional filters as well as the noise-smoothing filters with the initial probability distribution $\mathbf{x}_0 \sim \mathcal{N}(0.8, 0.02)$. Fig. 2(a) shows a representative case from our numerical simulations. There, we depict the conditioned mean of the filtering solutions together with the true state. Fig. 2(b) shows the average RMSEs of 50 state estimates for each time step. In this example we are interested in state estimates obtained from the filters since the transition takes place, i.e., for $t \geq 2.0$. We see the non-point-based conventional filters (LCF and VCF) completely lose the true trajectory. The point-based conventional filters (CCF and PCF) eventually become able to catch the trajectory but only after a number of assimilation time steps. Differently from conventional filters, the noise-smoothing filters successfully build accurate reconstructions of the entire state evolution despite the jump. We here notice that the conditioned noise $\xi_{n|n+1}$ in the VNSF is particularly far from zero when the jump happens, and as a result the VNSF clearly outperforms the other filters. However our subsequent numerical simulations show that this is not general but is a very specific case.

We next study the case for which the measurement function is the square of the shifted distance from the origin

$$\phi^n(\mathbf{x}) = (\mathbf{x} - 0.05)^2.$$

The observation distinguishes the two stable equilibria marginally. Fig. 3 uses the system parameters $\beta = 5$, $\sigma = 0.5$, the initial state $\mathbf{x}(0) = -0.2$, the numerical simulation time step $\delta t = 0.01$, and the observation noise covariance $R_n = 1.0$. Along with the initial condition $\mathbf{x}_0 \sim \mathcal{N}(0.8, 2.0)$, the filters are applied at various inter-observation times $\Delta t = M \times \delta t$. The average RMSEs committed by each filter across 100 independent simulations are depicted with frequent observations ($M = 1$) in Fig. 3(a), and sparse observations ($M = 10$) in Fig. 3(b). In this example one can see that the point-based conventional filters outperform the non-point-based conventional filters and that the accuracy of noise-smoothing filters improves compared with corresponding conventional filters. Furthermore, as the time between two successive measurements increases, the noise-smoothing filters become more accurate compared with corresponding conventional filters. This improvement of noise-smoothing filters for temporally sparse observations can be understood from the bias of the conditioned noise. We here mention that we do not depict the PCF and PNSF in the case of sparse observations, or for the other examples, because their performances are sensitively dependent on the number of random draws, and should be very similar to the ones from the cubature methods with degree 3 or 5 when they employ a sufficiently large number of samples.

B. *Lorenz-63 System*

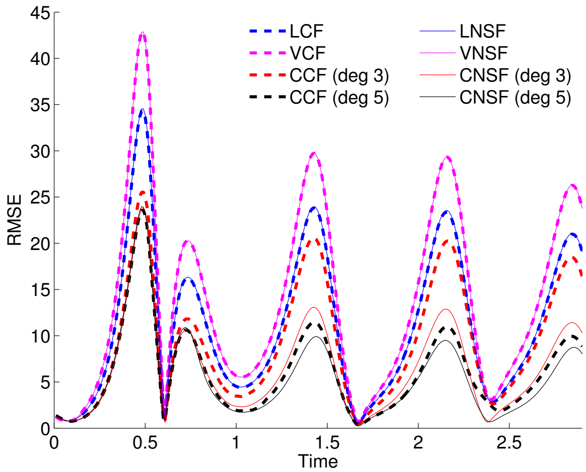
Let $\mathbf{x}(t) = [x(t), y(t), z(t)]^T$ be the state vector. We use the Euler approximation of the chaotic dynamical system

$$dx = \sigma(y - x)dt + g_1 dB_1,$$
$$dy = (\rho x - y - xz)dt + g_2 dB_2,$$
$$dz = (xy - \beta z)dt + g_3 dB_3,$$

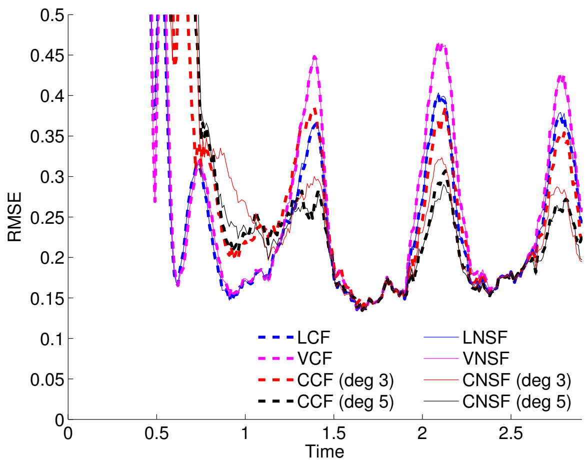
with $\delta t = \Delta t = 0.01$ as the forward model [31], [32]. We choose the system parameters $\sigma = 10$, $\rho = 28$, $\beta = 8/3$ and $g_1 = g_2 = 0$, $g_3 = 0.5$. The starting state is $\mathbf{x}(0) = [-0.2,\ -0.3,\ -0.5]^T$ and the initial condition is $\mathbf{x}_0 \sim \mathcal{N}([1.35,\ -3,\ 6]^T, 0.35 I_3)$ where I_3 denotes the 3×3 identity matrix. The observation process is determined by the measurement function

$$\phi^n(\mathbf{x}) = \sqrt{(x - 0.5)^2 + y^2 + z^2}$$

and the noise covariance $R_n = 0.5$. Fig. 4 depicts the average RMSEs from 120 simulations for each component of system



(a) x_1



(b) x_3

Fig. 4. The performance of various filters applied to the Lorenz-63 model with the shifted quadratic measurement function. The plot for x_2 is very similar to that of x_1 and is not shown.

variables, obtained from the conventional filters and noise-smoothing filters. The ordering of filtering accuracy among the different methods is very similar to that of the bistable system with frequent squared observation.

C. Target Tracking

Here we consider a model air-traffic monitoring scenario, where an aircraft executes a maneuvering turn in a horizontal plane at an unknown turn rate Ω_n at time n. The dynamical system is governed by the equation

$$\mathbf{x}_{n+1} = \begin{bmatrix} 1 & \frac{\sin(\Omega_n \Delta t)}{\Omega_n} & 0 & \frac{\cos(\Omega_n \Delta t)-1}{\Omega_n} & 0 \\ 0 & \cos(\Omega_n \Delta t) & 0 & -\sin(\Omega_n \Delta t) & 0 \\ 0 & \frac{1-\cos(\Omega_n \Delta t)}{\Omega_n} & 1 & \frac{\sin(\Omega_n \Delta t)}{\Omega_n} & 0 \\ 0 & \sin(\Omega_n \Delta t) & 0 & \cos(\Omega_n \Delta t) & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \mathbf{x}_n + \xi_n$$

where $\mathbf{x}_n = [x_n,\ \dot{x}_n,\ y_n,\ \dot{y}_n,\ \Omega_n]^T$; $[x_n,\ y_n]$ and $[\dot{x}_n,\ \dot{y}_n]$ are the position and velocity of the aircraft at time n; Δt is the time interval between two consecutive measurements; the driving noise $\xi_n \in \mathbb{R}^5$ is the zero mean Gaussian with covariance matrix

$$\Gamma_n = \begin{bmatrix} \frac{\Delta t^3}{3} & \frac{\Delta t^2}{2} & 0 & 0 & 0 \\ \frac{\Delta t^2}{2} & \Delta t & 0 & 0 & 0 \\ 0 & 0 & \frac{\Delta t^3}{3} & \frac{\Delta t^2}{2} & 0 \\ 0 & 0 & \frac{\Delta t^2}{2} & \Delta t & 0 \\ 0 & 0 & 0 & 0 & q\Delta t \end{bmatrix}.$$

Here the scalar parameter q controls the random walk behaviour of the turn rate from $\Omega_{n+1} = \Omega_n + \mathcal{N}(0, q\Delta t)$.

We assume a radar is fixed at the origin of the plane and equipped to measure the range, ρ_n, and bearing, θ_n, at time n. Hence the observation process is

$$\mathbf{y}_n = \begin{bmatrix} \rho_n \\ \theta_n \end{bmatrix} + \eta_n = \begin{bmatrix} \sqrt{x_n^2 + y_n^2} \\ \tan^{-1}\left(\frac{y_n}{x_n}\right) \end{bmatrix} + \eta_n$$

where the measurement noise is $\eta_n \sim \mathcal{N}(\mathbf{0}, R_n)$ with

$$R_n = \begin{bmatrix} \sigma_{\rho_n}^2 & 0 \\ 0 & \sigma_{\theta_n}^2 \end{bmatrix}.$$

Due to the inherent nonlinearity of the observation function, target tracking is a good problem for testing the performance of noise-smoothing filters.

With the parameters $\Delta t = 1$, $q = 1.75 \times 10^{-3}$, $\sigma_{\rho_n}^2 = 10^2$, $\sigma_{\theta_n}^2 = 10^{-5}$ and

$$\mathbf{x}_0 \sim \mathcal{N}\left(\begin{bmatrix} 10^3 \\ 3\times 10^2 \\ 10^3 \\ 0 \\ -\frac{3\pi}{180} \end{bmatrix}, \begin{bmatrix} 10^2 & 0 & 0 & 0 & 0 \\ 0 & 10 & 0 & 0 & 0 \\ 0 & 0 & 10^2 & 0 & 0 \\ 0 & 0 & 0 & 10 & 0 \\ 0 & 0 & 0 & 0 & 10^{-4} \end{bmatrix} \right),$$

we perform 200 independent simulations. In each case the target trajectory, whose initial state is an independent draw from $\mathbf{x}_0$, and the associated observations over $1 \le n \le 200$ time steps are randomly generated. We then apply the filters to reconstruct the evolution of the dynamical variables.

Fig. 5 displays one instance of the aircraft trajectory together with the various filtering estimates. In this example the

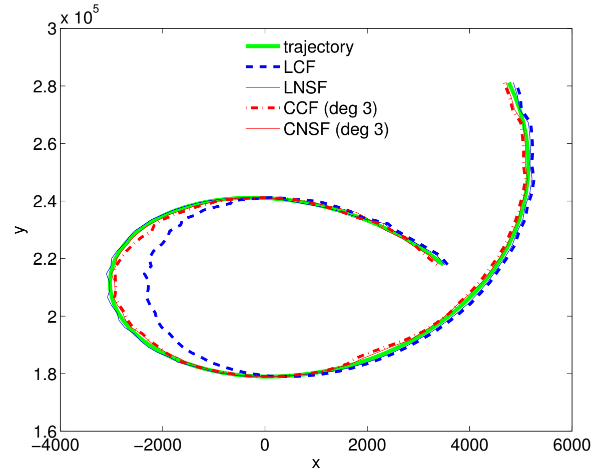
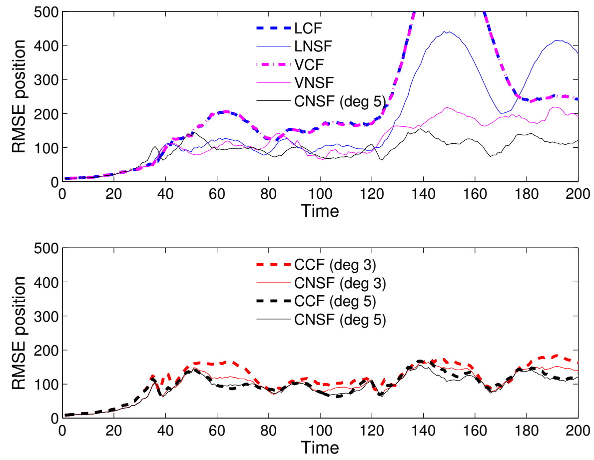
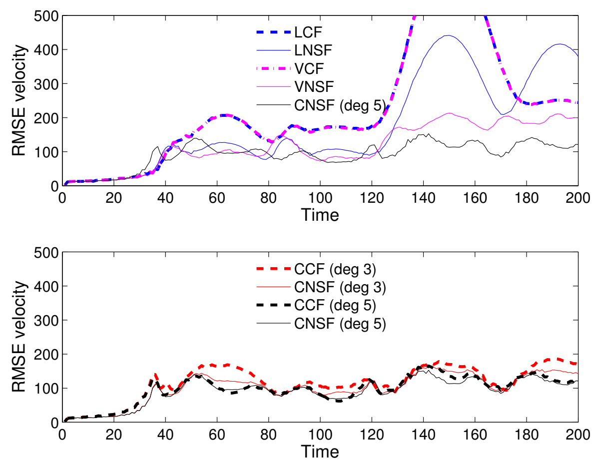


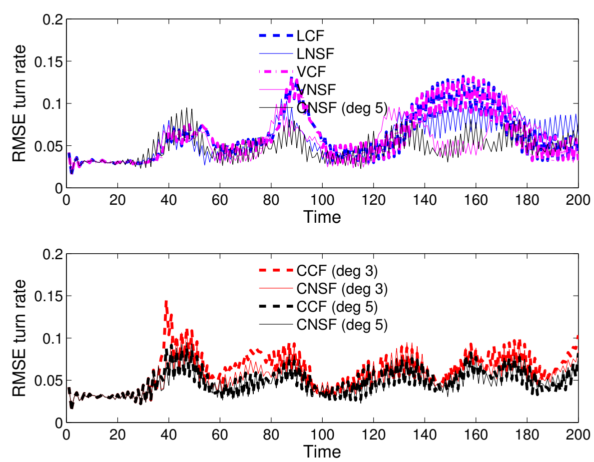
Fig. 5. A trajectory of the aircraft together with filtering estimates for $120 \le n \le 200$. The forward time direction is counterclockwise.



(a) position



(b) velocity



(c) turn rate

Fig. 6. The RMSEs between target and filtering estimates obtained from averaging over 200 independent simulations.

LCF cannot accurately estimate the target position when the system undergoes strong nonlinearities (the curvature of the position trajectory is large near $n = 150$) but the corresponding noise-smoothing filter accurately follows the target. Fig. 6 shows the average RMSEs, committed by each filter across 200 independent simulations, with respect to position, velocity and turn rate. We see that the non-point-based conventional filters (LCF and VCF) become quite in error around $n = 150$, whereas the CCF estimates keep reasonable accuracy for the entire time period. The application of noise-smoothing filters leads to accuracy improvements in all cases. To quantify the improvement we turn our attention to the time average. We depict, in Fig. 7, the RMSEs averaged over time intervals $50 \leq n \leq 200$. Although the overall accuracy of CNSF is superior to that of LNSF and VNSF, these two non-point-based noise-smoothing algorithms sometimes reach very high accuracy in the sense of a reduced time average, i.e., the enhanced but less uniform accuracy when compared with conventional filters.

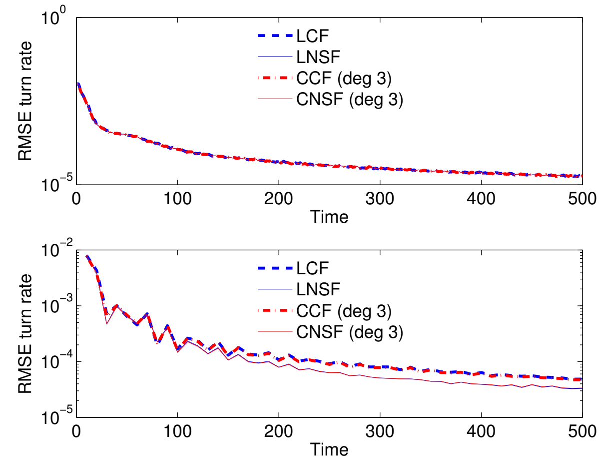
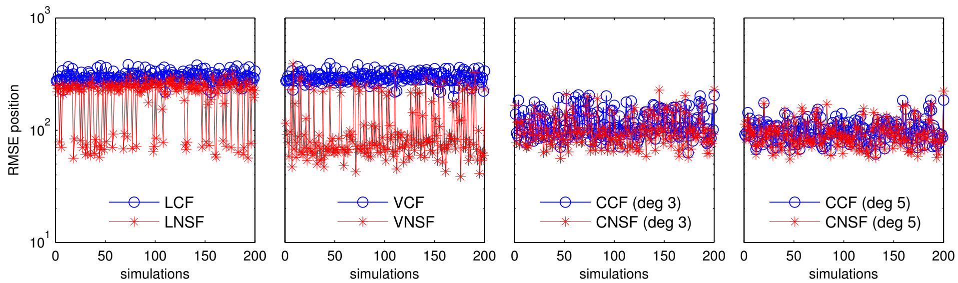


Fig. 8. The RMSEs between the turn rate parameter Ω_n and its estimates, obtained from averaging of 200 independent simulations. The top case uses frequent observations ($\Delta t = \delta t$) and the lower case, sparse observations ($\Delta t = 10 \times \delta t$).

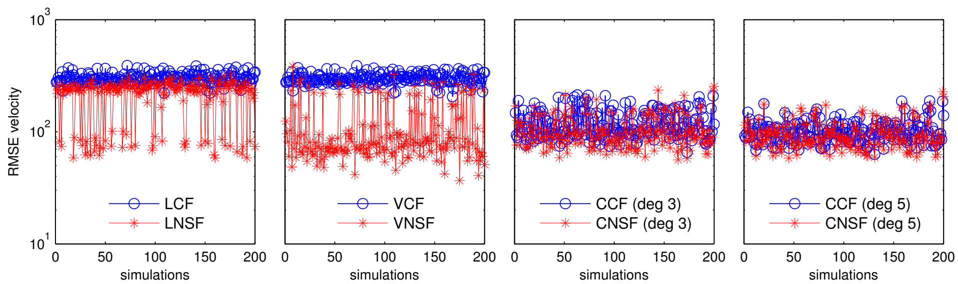
Finally we study the system with $q = 0$. In this case Ω_n is constant and the filtering solution can be used for parameter estimation. In Fig. 8, we see the noise-smoothing filters outperform conventional filters particularly with temporally sparse observations.

VI. Concluding Discussion

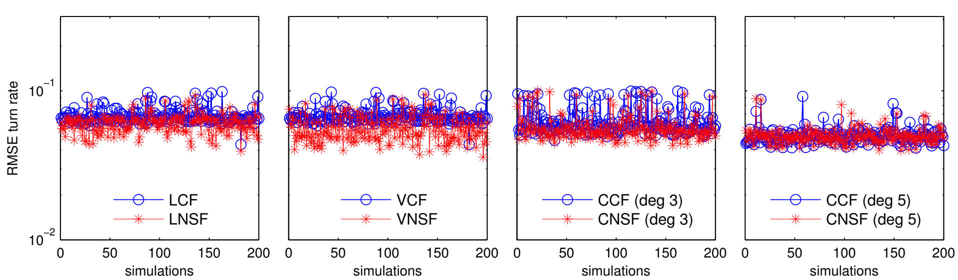
This paper formulates a family of sequential filters that, in contrast to the conventional approach, achieve the data assimilation via one step backward smoothing for both the variable and the driving noise to estimate a stochastic system with intermittent observations. The approximate solutions obtained from the proposed noise-smoothing filters tend to be closer to the observation forward in time due to the bias of the driving noise conditioned on future observations, and as a result can be more accurate than the conventional filters upon which the noise-smoothing filters are modelled. Our numerical simulations, performed on some examples widely used in the data assimilation community, reveal that this is indeed the case as far as the nonlinearity is present in either the time process equation or the measurement function. The result is encouraging and leads us to conjecture similar improvements in accuracy when the noise-smoothing filters are generalised to use Gaussian sum approximations or particles in solving



(a) position



(b) velocity



(c) turn rate

Fig. 7. The RMSEs between target and filtering estimates, obtained from averaging over the time period $50 \leq n \leq 200$ for each 200 simulations.

the nonlinear filtering problem. The authors plan to perform a comparison test between noise-smoothing filters and sequential Monte Carlo importance sampling or other smoothing-based filters such as a predictive filter [33]. Analytical results as to why in some particular circumstances noise-smoothing filters outperform conventional filters, would also be of interest and such understanding is currently being sought.

Acknowledgment

The authors would like to thank King Abdullah University of Science and Technology (KAUST) Award No. KUK-C1-013-04 for its financial support of this research. We also thank the anonymous referees and the editor for their helpful criticism of our work.

References

[1] A. Jazwinski, "Stochastic processes and filtering theory, vol. 64. San Diego, California: Mathematics in science and engineering," 1970.
[2] B. Anderson and J. Moore, *Optimal filtering*. Prentice-hall Englewood Cliffs, NJ, 1979, vol. 11.
[3] A. Doucet, N. De Freitas, and N. Gordon, *Sequential Monte Carlo methods in practice*. Springer Verlag, 2001.
[4] G. Evensen, *Data assimilation: the ensemble Kalman filter*. Springer Verlag, 2009.
[5] R. Kalman *et al.*, "A new approach to linear filtering and prediction problems," *Journal of Basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960.
[6] A. Gelb, *Applied optimal estimation*. MIT press, 1974.
[7] S. J. Julier and J. K. Uhlmann, "Unscented filtering and nonlinear estimation," *Proceedings of the IEEE*, vol. 92, no. 3, pp. 401–422, 2004.
[8] I. Arasaratnam and S. Haykin, "Cubature Kalman filters," *Automatic Control, IEEE Transactions on*, vol. 54, no. 6, pp. 1254–1269, 2009.
[9] J. H. Kotecha and P. M. Djuric, "Gaussian particle filtering," *Signal Processing, IEEE Transactions on*, vol. 51, no. 10, pp. 2592–2601, 2003.
[10] R. Chen and J. Liu, "Mixture Kalman filters," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 62, no. 3, pp. 493–508, 2000.
[11] A. Stordal, H. Karlsen, G. Nævdal, H. Skaug, and B. Vallès, "Bridging the ensemble Kalman filter and particle filters: the adaptive Gaussian mixture filter," *Computational Geosciences*, vol. 15, no. 2, pp. 293–305, 2011.
[12] N. Gordon, D. Salmond, and A. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," in *Radar and Signal Processing, IEE Proceedings F*, vol. 140, no. 2. IET, 1993, pp. 107–113.
[13] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking," *Signal Processing, IEEE Transactions on*, vol. 50, no. 2, pp. 174–188, 2002.
[14] O. Cappé, E. Moulines, and T. Rydén, *Inference in hidden Markov models*. Springer, 2005.
[15] F. Desbouvries, Y. Petetin, and B. Ait-El-Fquih, "Direct, prediction- and smoothing-based Kalman and particle filter algorithms," *Signal Processing*, vol. 91, no. 8, pp. 2064–2077, 2011.

[16] A. Doucet, S. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statistics and computing*, vol. 10, no. 3, pp. 197–208, 2000.
[17] R. Van Der Merwe, A. Doucet, N. De Freitas, and E. Wan, "The unscented particle filter," in *NIPS*, 2000, pp. 584–590.
[18] P. van Leeuwen, "Nonlinear data assimilation in geosciences: an extremely efficient particle filter," *Quarterly Journal of the Royal Meteorological Society*, vol. 136, no. 653, pp. 1991–1999, 2010.
[19] N. Whiteley and A. M. Johansen, "Recent developments in auxiliary particle filtering," *Barber, Cemgil, Chiappa Q*, vol. 2, 2011.
[20] S. Saha, P. K. Mandal, Y. Boers, H. Driessen, and A. Bagchi, "Gaussian proposal density using moment matching in smc methods," *Statistics and Computing*, vol. 19, no. 2, pp. 203–208, 2009.
[21] R. Cools and P. Rabinowitz, "Monomial cubature rules since Stroud: a compilation," *Journal of Computational and Applied Mathematics*, vol. 48, no. 3, pp. 309–326, 1993.
[22] T. Lyons and N. Victoir, "Cubature on wiener space," *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 460, no. 2041, pp. 169–198, 2004.
[23] M. Fisher, M. Leutbecher, and G. Kelly, "On the equivalence between Kalman smoothing and weak-constraint four-dimensional variational data assimilation," *Q. J. R. Meteorol. Soc.*, vol. 131, pp. 3235–3246, 2005.
[24] P. Kloeden and E. Platen, *Numerical solution of stochastic differential equations*. Springer, 2011, vol. 23.
[25] N. Victoir, "Asymmetric cubature formulae with few points in high dimension for symmetric measures," *SIAM journal on numerical analysis*, vol. 42, no. 1, pp. 209–227, 2004.
[26] D. Xiu, "Numerical integration formulas of degree two," *Applied Numerical Mathematics*, vol. 58, no. 10, pp. 1515–1520, 2008.
[27] B. Jia, M. Xin, and Y. Cheng, "High-degree cubature Kalman filter," *Automatica*, 2012.
[28] A. Mordecai, "Nonlinear programming: Analysis and methods," *Mineola, NY: Dover Publishing*, 2003.
[29] *MATLAB and Optimization Toolbox Release 2012b*. Natick, Massachusetts, United States.
[30] M. Freidlin and A. D. Wentzell, *Random perturbations of dynamical systems*. Springer, 2012, vol. 260.
[31] E. Lorenz, "Deterministic nonperiodic flow," *Journal of the atmospheric sciences*, vol. 20, no. 2, pp. 130–141, 1963.
[32] R. Miller, E. Carter Jr, and S. Blue, "Data assimilation into nonlinear stochastic models," *Tellus A*, vol. 51, no. 2, pp. 167–194, 1999.
[33] J. L. Crassidis and F. L. Markley, "Predictive filtering for nonlinear systems," *Journal of Guidance, Control, and Dynamics*, vol. 20, no. 3, pp. 566–572, 1997.

Wonjung Lee received the Ph.D. degree in mathematics from Courant Institute of Mathematical Sciences, New York, U.S., in 2009. Spending one year as postdoctoral research fellow at University of Warwick, U.K., he has worked at University of Oxford, U.K., since 2010.

Chris Farmer was awarded a Ph.D. degree in theoretical physics by the University of Exeter, U.K., in 1978. After a career in the nuclear and oil industries, mainly concerning uncertainty quantification, he moved to the University of Oxford, U.K., in 2009.