# ENLLVM: ENSEMBLE BASED NONLINEAR BAYESIAN FILTERING USING LINEAR LATENT VARIABLE MODELS

*Xiao Lin*

University of South Carolina
Department of Computer Science and Engineering
Columbia, SC, USA

*Gabriel Terejanu**

University of North Carolina at Charlotte
Department of Computer Science
Charlotte, NC, USA

## ABSTRACT

Real-time nonlinear Bayesian filtering algorithms are overwhelmed by data volume, velocity and increasing complexity of computational models. In this paper, we propose a novel ensemble based nonlinear Bayesian filtering approach which only requires a small number of simulations and can be applied to high-dimensional systems in the presence of intractable likelihood functions. The proposed approach uses linear latent projections to estimate the joint probability distribution between states, parameters, and observables using a mixture of Gaussian components generated by the reconstruction error for each ensemble member. Since it leverages the computational machinery behind linear latent variable models, it can achieve fast implementations without the need to compute high-dimensional sample covariance matrices. The performance of the proposed approach is compared with the performance of ensemble Kalman filter on a high-dimensional Lorenz nonlinear dynamical system.

***Index Terms***— dimensionality reduction, multi-view model, model error, intractable likelihood, Gaussian mixture

## 1. INTRODUCTION

Bayesian filtering is concerned with obtaining the posterior distribution of quantities of interest (QoI) such as state variables of a dynamical systems as well as unknown parameters in computational models using observational data. Its applicability ranges from data assimilation with applications in weather forecasting [1] to training recurrent neural networks for computer vision applications [2]. For a linear system, under the assumption of Gaussian probability distributions, the problem of estimating the states of the system has an exact closed-form solution given by the Kalman filter [3]. If the probability distributions are non-Gaussian or the system is nonlinear, in general no closed-form solutions are available. Among the most frequently cited approximate nonlinear filters are the extended Kalman filter (EKF) [4], the unscented

Kalman filter (UKF) [5, 6, 7], the ensemble Kalman filter (EnKF) [8], and the particle filter (PF) [9].

Both EKF and UKF provide Gaussian approximations to the posterior distributions. In addition, the EKF suffers from poor approximations for high nonlinear systems due to its reliance on model linearization, and the number of samples/sigma points in UKF scales linearly with the number of QoIs. On the other hand EnKF is a reduced rank sampling filter which propagates the uncertainty through the system nonlinearities and updates a relatively small ensemble of samples. Due to its simplicity in both theory and implementation, EnKF has been widely used, especially in meteorology [10, 11] where high dimensional data assimilation is performed. Nonetheless, EnKFs have been shown to be sensitive to the violation of the Gaussian assumption [12].

The particle filter uses a Monte Carlo sampling approach by propagating and updating a number of particles without the assumption of Gaussian statistics. Compared with Kalman filter variations, particle filter works with any arbitrary prior distributions and is capable to capture multimodal posterior distributions. However, particle filter has mostly been applied to low dimensional problems. To cope with high dimensional problems, variations of particle filter have been developed. One class is multiple particle filter [13, 14, 15]. Multiple particle filter uses standard particle filters to update state variables in subspaces. In this way high dimensional filtering is converted to a number of easier lower dimensional problems that require a smaller number of particles. However, problems arise by ignoring when states are coupled in the observation model or when only partial states are observed.

Often, general formulations that exploit the structure of physics-based models by using internal discrepancies to capture structural errors yield intractable likelihood functions [16]. In fact, in many cases, as compared with the most common additive observational noise, the distribution of observational noise is unknown and embedded into the nonlinear observational model which yields an intractable likelihood function. These make particle filter unsuitable as it relies on the evaluation of the likelihood function.

We propose a novel Bayesian filtering approach which

only requires a small number of samples even in high dimensional systems. The proposed approach uses linear latent projections to approximate the non-Gaussian joint probability distribution between states, parameters, and observables using a mixture of Gaussian components generated by the reconstruction error for each ensemble member. The method is performed without evaluating the likelihood of observations, thus it can be applied to filtering problems in which the likelihood function is intractable.

The rest of paper is organized as follows. The proposed nonlinear filter is detailed in Section 2. The method is not restricted to a particular type of linear latent models and several options are provided. In Section 3, two numerical experiments are performed to assess its performance as compared with EnKF on several different scenarios. Conclusions are provided in Section 4.

## 2. METHODOLOGY

Consider the following general parameterized nonlinear dynamical system perturbed by process noise $\mathbf{w}_k$, measurement noise $\mathbf{v}_k$ and uncertain initial conditions. Given a set of measurements, the goal is to infer the state of the system and parameters to improve model predictions and as a result decision making under uncertainty.

$$\mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \boldsymbol{\theta}, \mathbf{w}_k) \tag{1}$$
$$\mathbf{d}_k = \mathbf{h}(\mathbf{x}_k, \boldsymbol{\theta}, \mathbf{v}_k) \tag{2}$$
$$\mathbf{x}_0 \sim p(\mathbf{x}_0) \tag{3}$$

Given a set of observation $\mathbf{D}_k = \{\tilde{\mathbf{d}}_i | 1 \leq i \leq k\}$, Bayesian filtering is the problem of finding the joint probability density function (pdf) of the states $\mathbf{x}_k$ and parameters $\boldsymbol{\theta}$ conditioned on all the observations up to and including current time $t_k$, $p(\mathbf{x}_k, \boldsymbol{\theta} \mid \mathbf{D}_k)$.

Here, we tackle Bayesian filtering from a different angle. Instead of relying on Eq. (2) to calculate the likelihood, we operate on the joint distribution of $\mathbf{x}_k$, $\boldsymbol{\theta}$ and $\mathbf{d}_k$ directly. For simplicity, in the followings, all the quantities of interest (QoIs) at any time $k$ will be denoted by $\mathbf{q}_k$.

$$\mathbf{q}_k = [\mathbf{x}_k, \boldsymbol{\theta}]^T \tag{4}$$

Furthermore, $\mathbf{q}_{k+1|k}$ is obtained through Eq. (1) after observing $\mathbf{D}_k$ but before $\tilde{\mathbf{d}}_{k+1}$. Thus $p(\mathbf{x}_{k+1}, \boldsymbol{\theta} \mid \mathbf{D}_k)$ can be written as $p(\mathbf{q}_{k+1|k})$. Similarly, $\mathbf{d}_{k+1|k}$ is obtained by propagating $\mathbf{q}_{k+1|k}$ through Eq. (2).

The proposed EnLLVM is a sampling based approach, where $N$ samples from the QoI prior distribution $p(\mathbf{q}_0)$ are propagated through the nonlinear system and updated at every time step. While the propagation of samples is straightforward, the update requires first to obtain a Gaussian mixture approximation of the predictive joint distribution $p(\mathbf{q}_{k+1|k}, \mathbf{d}_{k+1|k})$ based on the reconstruction error of a linear latent variable model (LLVM), see Figure 1. Second, the

posterior distribution of the QoIs is obtained analytically as a Gaussian mixture by conditioning the newly approximated joint pdf with the observational data.
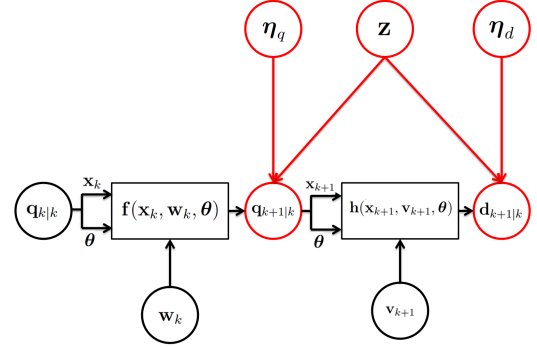


**Fig. 1**. Linear latent variable model construction

**EnLLVM: Gaussian mixture approximation**. The linear latent variable model is shown in Eq. (5)-(6). $\mathbf{W}_q$ and $\mathbf{W}_d$ are the QoI and observable coefficient matrices, $\mathbf{z}$ is the latent variable, $\boldsymbol{\mu}_q$ and $\boldsymbol{\mu}_d$ are the QoI and observable bias, and $\boldsymbol{\eta}_q$ and $\boldsymbol{\eta}_d$ are the noise variables for the QoI and observable. Assume $\mathbf{q}_{k+1|k}$ and $\mathbf{d}_{k+1|k}$ have a dimensionality of $H_q$ and $H_d$ respectively, and denote the dimensionality of the joint space as $H = H_q + H_d$. Let $M$ be the dimensionality of latent variable $\mathbf{z}$, where $M$ is much smaller than $H$.

$$\mathbf{q}_{k+1|k} = \mathbf{W}_q \mathbf{z} + \boldsymbol{\mu}_q + \boldsymbol{\eta}_q \tag{5}$$
$$\mathbf{d}_{k+1|k} = \mathbf{W}_d \mathbf{z} + \boldsymbol{\mu}_d + \boldsymbol{\eta}_d \tag{6}$$
$$\mathbf{z} \sim \mathcal{N}(0, \mathbf{I}_{M \times M})$$
$$\boldsymbol{\eta}_q \sim \mathcal{N}(0, \boldsymbol{\Psi}_q)$$
$$\boldsymbol{\eta}_d \sim \mathcal{N}(0, \boldsymbol{\Psi}_d)$$

The parameters of the LLVM can be obtained via maximum likelihood estimation [17] from the samples propagated through the nonlinear system, $\{\mathbf{q}_{k+1|k}^i, \mathbf{d}_{k+1|k}^i\}_{i=1...N}$. We may restrict the structure of the parameters, such that the latent space will preserve certain properties of the original data set. This will lead to different linear latent variable models: (PPCA) probabilistic principal component analysis where $\text{blkdiag}(\boldsymbol{\Psi}_q, \boldsymbol{\Psi}_d) = \sigma^2 \mathbf{I}_{H \times H}$, (FA) factor analysis where $\text{blkdiag}(\boldsymbol{\Psi}_q, \boldsymbol{\Psi}_d) = \text{diag}(\sigma_1^2 \dots \sigma_H^2)$, and (PCCA) probabilistic canonical correlation analysis where $\boldsymbol{\Psi}_q, \boldsymbol{\Psi}_d$ have off-diagonal elements [18].

Using the above LLVM, the Gaussian mixture approximation of the predicted joint distribution is obtained in the following way. For each sample $\{\mathbf{q}_{k+1|k}^i, \mathbf{d}_{k+1|k}^i\}$ of $p(\mathbf{q}_{k+1|k}, \mathbf{d}_{k+1|k})$, we project it into the latent space where there is a corresponding latent variable $\mathbf{z}_i$ which follows a Gaussian distribution:

$$\mathbf{z}_i \sim \mathcal{N}(\text{E}[\mathbf{z}|\mathbf{q}_{k+1|k}^i, \mathbf{d}_{k+1|k}^i], \text{Cov}[\mathbf{z}|\mathbf{q}_{k+1|k}^i, \mathbf{d}_{k+1|k}^i])$$

Projecting the Gaussian latent distribution back we obtain a Gaussian distribution in the original space corresponding to the reconstruction error of the above $i$th sample. Then, the joint pdf $p(\mathbf{q}_{k+1|k}, \mathbf{d}_{k+1|k})$ can be approximated by combining all the Gaussian reconstructions corresponding to the original samples, which results in a Gaussian mixture provided by the EnLLVM, Eq. (7), see also Figure 2.

$$\hat{p}_{en}(\mathbf{q}_{k+1|k}, \mathbf{d}_{k+1|k}) = \frac{1}{N}\sum_{i=1}^{N}\mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma_i}) \tag{7}$$

$$\boldsymbol{\mu}_i = \mathbf{W}\mathrm{E}[\mathbf{z}|\mathbf{q}_{k+1|k}^i, \mathbf{d}_{k+1|k}^i] + \boldsymbol{\mu} \tag{8}$$

$$\boldsymbol{\Sigma_i} = \mathbf{W}\mathrm{Cov}[\mathbf{z}|\mathbf{q}_{k+1|k}^i, \mathbf{d}_{k+1|k}^i]\mathbf{W}^T + \boldsymbol{\Psi} \tag{9}$$

Here, the LLVM parameters are $\mathbf{W} = [\mathbf{W}_q^T, \mathbf{W}_d^T]^T$, $\boldsymbol{\Psi} = \mathrm{blkdiag}(\boldsymbol{\Psi}_q, \boldsymbol{\Psi}_d)$, and the overall Gaussian approximation of LLVM is given by $\boldsymbol{\mu} = [\boldsymbol{\mu}_q^T, \boldsymbol{\mu}_d^T]^T$ and $\boldsymbol{\Sigma} = \mathbf{W}\mathbf{W}^T + \boldsymbol{\Psi}$. The mean and covariance of the Gaussian distribution corresponding to the $i$th latent variable are given by:

$$\mathrm{E}[\mathbf{z}|\mathbf{q}_{k+1|k}^i, \mathbf{d}_{k+1|k}^i] = \mathbf{W}^T\boldsymbol{\Sigma}^{-1}([\mathbf{q}_{k+1|k}^i, \mathbf{d}_{k+1|k}^i]^T - \boldsymbol{\mu}) \tag{10}$$

$$\mathrm{Cov}[\mathbf{z}|\mathbf{q}_{k+1|k}^i, \mathbf{d}_{k+1|k}^i] = \mathbf{I}_{M\times M} - \mathbf{W}^T\boldsymbol{\Sigma}^{-1}\mathbf{W} . \tag{11}$$

Computationally, we assume that the $H \gg N \gg M$, where $H$ is the dimensionality of the joint space, $N$ the number of samples, and $M$ the dimensionality of the latent space. Since we can leverage the entire computational machinery of the LLVM, different implementations will yield different computational costs. Nonetheless, expectation maximization (EM) and its variants are preferred. Every iteration of EM will result in $\mathcal{O}(HNM)$, and online EM can process one sample at a time, which is preferred when both $H$ and $N$ are large and samples may require different run times (e.g. different parameter values may increase the stiffness of a system).
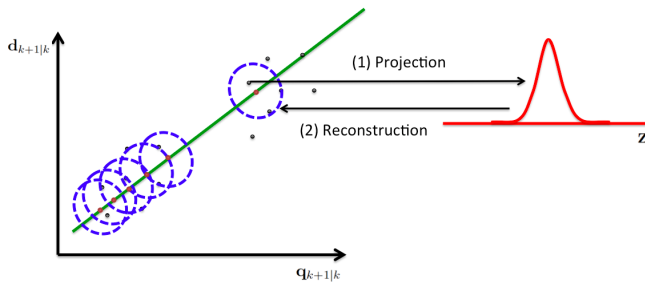


**Fig. 2**. EnLLVM: Gaussian mixture approximation

*Lemma*. EnLLVM provides the same estimate for the mean and covariance of the samples as LLVM, while it captures higher order statistics as compared with LLVM.

$$\mathrm{E}_{en}[\mathbf{q}_{k+1|k}, \mathbf{d}_{k+1|k}] = \boldsymbol{\mu} \tag{12}$$

$$\mathrm{Cov}_{en}[\mathbf{q}_{k+1|k}, \mathbf{d}_{k+1|k}] = \boldsymbol{\Sigma} \tag{13}$$

**EnLLVM: Bayesian update**. Given the observation data $\tilde{\mathbf{d}}_{k+1}$ and the Gaussian mixture approximation in Eq.7, the QoI posterior approximation is easily obtained via conditioning of the joint and results in the following Gaussian mixture.

$$\hat{p}_{en}(\mathbf{q}_{k+1|k+1}) = \sum_{i=1}^{N}w_i\mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma_i}) \tag{14}$$

$$\boldsymbol{\mu}_i = \mathbf{W}_q\mathrm{E}[\mathbf{z}_i|\tilde{\mathbf{d}}_{k+1}] + \boldsymbol{\mu}_q \tag{15}$$

$$\boldsymbol{\Sigma_i} = \mathbf{W}_q\mathrm{Cov}[\mathbf{z}_i|\tilde{\mathbf{d}}_{k+1}]\mathbf{W}_q^T + \boldsymbol{\Psi}_q . \tag{16}$$

Each Gaussian component is assigned with a weight, $w_i$, which is calculated according to the corresponding likelihood under the Gaussian mixture approximation. The mean and covariance matrix of $\mathbf{z}_i$ given $\tilde{\mathbf{d}}_{k+1}$ are as follows.

$$\mathrm{E}[\mathbf{z}_i|\tilde{\mathbf{d}}_{k+1}] = \mathrm{Cov}[\mathbf{z}_i|\tilde{\mathbf{d}}_{k+1}](\mathbf{W}^T\boldsymbol{\Psi}_d^{-1}(\tilde{\mathbf{d}}_{k+1} - \boldsymbol{\mu}_d) \tag{17}$$
$$+ \mathrm{Cov}[\mathbf{z}|\mathbf{q}_{k+1|k}^i, \mathbf{d}_{k+1|k}^i]^{-1}\mathrm{E}[\mathbf{z}|\mathbf{q}_{k+1|k}^i, \mathbf{d}_{k+1|k}^i])$$

$$\mathrm{Cov}[\mathbf{z}_i|\tilde{\mathbf{d}}_{k+1}] = (\mathrm{Cov}[\mathbf{z}|\mathbf{q}_{k+1|k}^i, \mathbf{d}_{k+1|k}^i]^{-1} + \mathbf{W}_d^T\boldsymbol{\Psi}_d^{-1}\mathbf{W}_d)^{-1}$$

**EnLLVM: Noise inflation**. To avoid component degeneracy and accommodate for potential model error, noise inflation is performed. We introduce a new hyperparameter $\alpha$ in the distribution of the data noise $\boldsymbol{\eta}_d$:

$$\boldsymbol{\eta}_d \sim \mathcal{N}(0, \alpha\boldsymbol{\Psi}_d) \tag{18}$$

Here, $\alpha$ is obtained via maximum likelihood:

$$\alpha^* = \arg_\alpha\max\mathcal{N}(\tilde{\mathbf{d}}_{k+1}; \boldsymbol{\mu}_d, \mathbf{W}_d\mathbf{W}_d^T + \alpha\boldsymbol{\Psi}_d) \tag{19}$$

$$= \frac{1}{H}\mathrm{tr}(\boldsymbol{\Psi}_d^E\boldsymbol{\Psi}_d^{-1}) \tag{20}$$

If $\alpha^* > 1$, the noise in Eq. (6) will be inflated, which will lead to larger uncertainty in the latent variable. This uncertainty will eventually be reflected in the uncertainty of QoIs through Eq. (5). This means the model error will be absorbed in the uncertainty of QoIs, which provides a way to investigate model error and develop models.

## 3. NUMERICAL RESULTS

### 3.1. Example 1 - Lorenz 63

The Lorenz 63 is a three dimensional system and it is used here to show that EnLLVM can capture non-Gaussian distributions [19, 20].

$$dx_1/dt = -cx_1 + cx_2 \tag{21}$$

$$dx_2/dt = -x_1x_3 + rx_1 - y_2 \tag{22}$$

$$dx_3/dt = x_1x_2 - bx_3 \tag{23}$$

Here $c = 10$, $b = \frac{8}{3}$ and $r = 28$. The discrete measurement model is given in Eq. (24)

$$d_k = \sqrt{x_1(t_k)^2 + x_2(t_k)^2 + x_3(t_k)^2} + v_k, \ v_k \sim \mathcal{N}(0, 1) \tag{24}$$

The state prior distribution is set to:

$$p(x(t_0)) \sim 0.5\mathcal{N}([-0.2, -0.2, 8]^T, \sqrt{0.35}\mathbf{I}_3) + 0.5\mathcal{N}([0.2, 0.2, 8]^T, \sqrt{0.35}\mathbf{I}_3) \tag{25}$$

The total simulation time is 4sec and the system is discretized with a time step $\Delta t = 0.1$sec. Synthetic measurements are generated by randomly selected initial states from Eq. (25) and propagated through Eq. (24). States are updated every 4 time steps.

The joint space of states and observations have a dimensionality of four, and in this simulation, we use two components for PPCA and 30 samples for Bayesian update. Fig. 3 shows the posterior distribution of states after each update. The horizontal lines indicate the true states. As we can see, EnPPCA is capable to capture the true states and the bimodal distribution of the states.
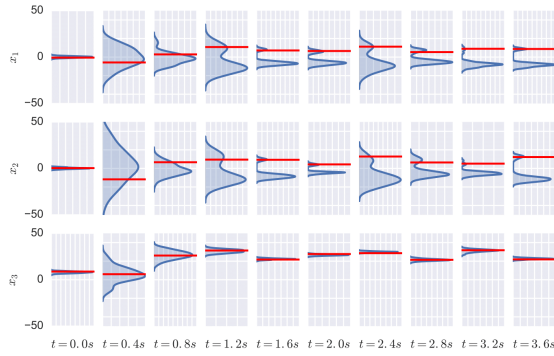


**Fig. 3**. Posterior distribution of states in Lorenz 63 system after each update.

## 3.2. Example 2 - Lorenz 96

In this example, EnPPCA and EnKF will be applied to track the states of Lorenz 96, a 40 dimensional nonlinear system and commonly used in benchmark studies.

$$dx_j(t)/dt = (x_{j+1} - x_{j-2})x_{j-1} - x_j + 8, \ j = 1 \ldots 40, \quad (26)$$

where $x_0 = x_{40}, x_{-1} = x_{39}$, and $x_{41} = x_1$. The system is discretized with a time step of $\Delta t = 0.001$sec. The prior is given by $\mathcal{N}(0,1)$ and the observation model provides an incomplete observation of the state of the system at every $\Delta t = 0.1$sec. Besides the linear measurement model, two different nonlinear measurement models are also used. Another important scenario is the presence of model error. As previously discussed, EnPPCA has the property to absorb potential model error and to reflect it in the posterior distribution of the QoI. Here, four different models are used to generate the observational data and they only differ in the constant forcing.

The joint space between the QoIs (states of the system) and the observable has dimensionality of 60. For this study, 5 components are used for EnPPCA, which means the dimensionality of the latent space is 5. Here, we use 30 samples to track the states of the system. Root mean square error (RMSE) is used as the metric to measure the predictability of EnPPCA and EnKF. RMSE is calculated between the mean of posterior samples and the true state value. Table. 1 shows

**Table 1**. EnKF vs EnPPCA: RMSE statistics of 100 trials

| Model Error | EnKF | | EnPPCA | |
| --- | --- | --- | --- | --- |
| | Mean | STD | Mean | STD |
| **Linear Measurement Model** | | | | |
| No err. | **2.71** | 0.61 | 2.92 | **0.23** |
| M1: 9 | 3.71 | 0.63 | **2.99** | **0.23** |
| M2: 10 | 4.63 | 0.71 | **3.07** | **0.22** |
| M3: 11 | 5.31 | 0.81 | **3.18** | **0.21** |
| M4: 12 | 5.9 | 0.84 | **3.27** | **0.20** |
| **Nonlinear Measurement Model (I)** | | | | |
| No err. | **1.87** | **0.53** | 2.85 | 0.26 |
| M1: 9 | 3.3 | 0.63 | **2.91** | **0.22** |
| M2: 10 | 4.37 | 0.47 | **3.02** | **0.21** |
| M3: 11 | 5.01 | 0.44 | **3.17** | **0.17** |
| M4: 12 | 5.51 | 0.49 | **3.32** | **0.17** |
| **Nonlinear Measurement Model (II)** | | | | |
| No err. | 2.87 | 0.76 | **2.84** | **0.24** |
| M1: 9 | 3.78 | 0.75 | **2.94** | **0.22** |
| M2: 10 | 4.78 | 0.56 | **3.03** | **0.20** |
| M3: 11 | 5.43 | 0.53 | **3.16** | **0.19** |
| M4: 12 | 6.1 | 0.55 | **3.37** | **0.17** |

the RMSE statistics of 100 trials. If no model error exists, the performance of EnKF is comparable with or better than EnPPCA in average RMSE for all three measurement models, otherwise the performance of EnLLVM is statistically significant better than EnKF as the model error increases.

$$
\begin{cases}
\text{Data generation process:} \\
M1: \ dx_j(t)/dt = (x_{j+1} - x_{j-2})x_{j-1} - x_j + 9 \\
M2: \ dx_j(t)/dt = (x_{j+1} - x_{j-2})x_{j-1} - x_j + 10 \\
M3: \ dx_j(t)/dt = (x_{j+1} - x_{j-2})x_{j-1} - x_j + 11 \\
M4: \ dx_j(t)/dt = (x_{j+1} - x_{j-2})x_{j-1} - x_j + 12 \\
\text{Linear measurement model:} \\
d_j(t) = x_{2j-1}(t) + v_j(t), \quad v_j(t) \sim \mathcal{N}(0, I_{20}) \\
\text{Nonlinear measurement models:} \\
(I): \ d_j(t) = x_{2j-1}(t)x_{2j}(t) + v_j(t), \quad v_j(t) \sim \mathcal{N}(0, I_{20}) \\
(II): \ d_j(t) = x_{2j-1}(t)^2 + v_j(t), \quad v_j(t) \sim \mathcal{N}(0, I_{20})
\end{cases}
$$

## 4. CONCLUSIONS

A general and fast computational framework is developed for nonlinear filtering of high dimensional nonlinear dynamical systems. This is achieved by developing a robust probabilistic model to approximate the Bayesian inference problem and obtain samples from high dimensional posterior distributions. In addition, the proposed approach addresses one of the main challenges in uncertainty quantification, namely dealing with modeling errors and their effect on inference and prediction.

# 5. REFERENCES

[1] L. Isaksen, M. Bonavita, R. Buizza, M. Fisher, J. Haseler, M. Leutbecher, and L. Raynaud, "Ensemble of data assimilations at ecmwf," *ECMWF Technical Memorandum*, , no. 636, pp. 45, 12/2010 2010.

[2] J. Gu, X. Yang, S. D. Mello, and J. Kautz, "Dynamic facial analysis: From Bayesian filtering to recurrent neural network," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017, pp. 1531–1540.

[3] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME–Journal of Basic Engineering*, vol. 82, pp. 35–45, 1960, Series D.

[4] J. Crassidis and J. Junkins, *Optimal Estimation of Dynamic Systems*, CRC Press, 2004.

[5] S.J. Julier and J.K. Uhlmann, "Unscented Filtering and Nonlinear Estimation," *Proceedings of the IEEE*, vol. 92, no. 3, pp. 401–422, 2004.

[6] E. A. Wan and Rudolph v.d.M., *The Unscented Kalman Filter*, chapter 7, pp. 221–280, Wiley-Blackwell, 2002.

[7] P. Stano, Z. Lendek, J. Braaksma, R. Babuka, C. de Keizer, and A. J. den Dekker, "Parametric Bayesian filters for nonlinear stochastic dynamical systems: A survey," *IEEE Transactions on Cybernetics*, vol. 43, no. 6, pp. 1607–1624, 2013.

[8] G. Evensen, "The Ensemble Kalman Filter: Theoretical Formulation and Practical Implementation," *Ocean Dynamics*, vol. 53, pp. 343–367, 2003.

[9] P. Del Moral, "Non linear filtering: Interacting particle solution," *Markov Processes and Related Fields*, vol. 2, pp. 555–580, 03 1996.

[10] P. L. Houtekamer and Herschel L. Mitchell, "A sequential ensemble Kalman filter for atmospheric data assimilation," *Monthly Weather Review*, vol. 129, no. 1, pp. 123–137, 2001.

[11] P. L. Houtekamer and Fuqing Zhang, "Review of the ensemble Kalman filter for atmospheric data assimilation," *Monthly Weather Review*, vol. 144, no. 12, pp. 4489–4532, 2016.

[12] J. Lei, P. Bickel, and C. Snyder, "Comparison of ensemble Kalman filters under non-Gaussianity," *Monthly Weather Review*, vol. 138, no. 4, pp. 1293–1306, 2010.

[13] P. M. Djuric, T. Lu, and M. F. Bugallo, "Multiple particle filtering," in *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*, April 2007, vol. 3, pp. III–1181–III–1184.

[14] M. F. Bugallo and P. M. Djuric, "Particle filtering in high-dimensional systems with Gaussian approximations," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 8013–8017.

[15] B. Ait-El-Fquih and I. Hoteit, "A variational Bayesian multiple particle filtering scheme for large-dimensional systems," *IEEE Transactions on Signal Processing*, vol. 64, no. 20, pp. 5409–5422, Oct 2016.

[16] T. A. Oliver, G. Terejanu, C. S. Simmons, and R. D. Moser, "Validating Predictions of Unobserved Quantities," *Computer Methods in Applied Mechanics and Engineering*, vol. 283, no. 0, pp. 1310 – 1335, 2015.

[17] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer Verlag, Aug. 2006.

[18] F. R. Bach and M. I. Jordan, "A probabilistic interpretation of canonical correlation analysis," Tech. Rep. Technical Report 688, Department of Statistics, University of California, Berkeley, 2005.

[19] G. Terejanu, P. Singla, T. Singh, and P. D. Scott, "Adaptive Gaussian sum filter for nonlinear Bayesian estimation," *IEEE Transactions on Automatic Control*, vol. 56, no. 9, pp. 2151–2156, 2011.

[20] L. Dovera and E. Della Rossa, "Multimodal ensemble Kalman filtering using Gaussian mixture models," *Computational Geosciences*, vol. 15, no. 2, pp. 307–323, 2011.