A Gaussian mixture ensemble transform filter[*]

Sebastian Reich

August 14, 2011

Abstract

We generalize the popular ensemble Kalman filter to an ensemble transform filter where the prior distribution can take the form of a Gaussian mixture or a Gaussian kernel density estimator. The design of the filter is based on a continuous formulation of the Bayesian filter analysis step. We call the new filter algorithm the ensemble Gaussian mixture filter (EGMF). The EGMF is implemented for three simple test problems (Brownian dynamics in one dimension, Langevin dynamics in two dimensions, and the three dimensional Lorenz-63 model). It is demonstrated that the EGMF is capable to track systems with non-Gaussian uni- and multimodal ensemble distributions.

1 Introduction

We consider dynamical models given in the form of ordinary differential equations (ODEs)

$$\dot{\mathbf{x}} = f(\mathbf{x}, t) \tag{1}$$

with state variable $\mathbf{x} \in \mathbb{R}^N$. Initial conditions at time t_0 are not precisely known and are treated as a random variable instead, i.e., we assume that

$$\mathbf{x}(t_0) \sim \pi_0,$$

where $\pi_0(\mathbf{x})$ denotes a given probability density function (PDF). The solution of (1) at time t with initial condition $\mathbf{x}_0$ at t_0 is denoted by $\mathbf{x}(t; t_0, \mathbf{x}_0)$.

The evolution of the initial PDF π_0 under the ODE (1) up to a time $t > t_0$ is provided by the continuity equation

$$\frac{\partial \pi}{\partial t} = -\nabla_{\mathbf{x}} \cdot (\pi f), \tag{2}$$

which is also called Liouville's equation in the statistical mechanics literature (Gardiner, 2004). Let us denote the solution of Liouville's equation at observation time t by $\pi(\mathbf{x}, t)$. In other words, solutions $\mathbf{x}(t; t_0, \mathbf{x}_0)$ with $\mathbf{x}_0 \sim \pi_0$ constitute a random variable with PDF $\pi(\cdot, t)$.

For a chaotic ODE (1), i.e. for an ODE with positive Lyapunov exponents, the PDF $\pi(\cdot, t)$ will be spread out over the whole chaotic attractor for $t \to \infty$. This in turn implies a limited solution predictability in the sense that the time-evolved PDF will become increasingly independent of the initial PDF π_0. Furthermore, even if the initial PDF is nearly Gaussian

---

[*]Universität Potsdam, Institut für Mathematik, Am Neuen Palais 10, D-14469 Potsdam, Germany

with mean $\overline{\mathbf{x}}_0$ and small covariance matrix $\mathbf{P}$, the solution $\mathbf{x}(t; t_0, \overline{\mathbf{x}}_0)$ will become increasingly unrepresentative of the expectation value of the underlying random variable it is supposed to represent.

To counteract the divergence of nearby trajectories under chaotic dynamics, we assume that we have uncorrelated measurements $\mathbf{y}_{\mathrm{obs}}(t_j) \in \mathbb{R}^K$ at times t_j, $j \geq 1$ with measurement error covariance matrix $\mathbf{R} \in \mathbb{R}^{K \times K}$, i.e.

$$\mathbf{y}_{\mathrm{obs}}(t_j) - \mathbf{H}\mathbf{x}(t_j) \sim \mathrm{N}(\mathbf{0}, \mathbf{R}), \tag{3}$$

where the notation $\mathrm{N}(\overline{\mathbf{y}}, \mathbf{B})$ is used to denote a normal distribution in $\mathbf{y} \in \mathbb{R}^K$ with mean $\overline{\mathbf{y}}$ and covariance matrix $\mathbf{B} \in \mathbb{R}^{K \times K}$. The matrix $\mathbf{H} \in \mathbb{R}^{K \times N}$ is called the forward operator. The task of combining solutions to (1) with intermittent measurements (3) is called data assimilation in the geophysical literature (Evensen, 2006) and filtering in the statistical literature (Bain and Crisan, 2009).

A first step to perform data assimilation for nonlinear ODEs (1) is to approximate solutions to the associated Liouville equation (2). In this paper, we rely exclusively on particle methods (Bain and Crisan, 2009) for which Liouville's equation is naturally approximated by the evolving empirical measure. More precisely, particle or ensemble filters rely on the simultaneous propagation of M independent solutions $\mathbf{x}_i(t) \in \mathbb{R}^N$, $i = 1, \ldots, M$, of (1) (Evensen, 2006). We associate the empirical measure

$$\pi_{\mathrm{em}}(\mathbf{x}, t) = \sum_{i=1}^{M} \gamma_i \delta(\mathbf{x} - \mathbf{x}_i(t)) \tag{4}$$

with weights $\gamma_i > 0$ satisfying

$$\sum_{i=1}^{M} \gamma_i = 1.$$

Here $\delta(\cdot)$ denotes the Dirac delta function. Hence our statistical model is given by the empirical measure (4) and is parametrized by the particle weights $\{\gamma_i\}$ and the particle locations $\{\mathbf{x}_i\}$. In the absence of measurements, the empirical measure π_{em} with constant weights γ_i is an exact (weak) solution to Liouville's equation (2) provided the $\mathbf{x}_i(t)$'s are solutions to the ODE (1). Optimal statistical efficiency is achieved with equal particle weights $\gamma_i = 1/M$.

The assimilation of a measurement at t_j leads via Bayes' theorem to a discontinuous change in the statistical model (4). Sequential Monte Carlo methods (Bain and Crisan, 2009) are primarily based on a discontinuous change in the weight factors γ_i. To avoid a subsequent degeneracy in the particle weights one re-samples or uses other techniques which essentially lead to a redistribution of particle positions $\mathbf{x}_i$. See, for example, Bain and Crisan (2009) for more details. The ensemble Kalman filter (EnKF) relies on the alternative idea to replace the empirical measure (4) by a Gaussian PDF prior to an assimilation step (Evensen, 2006). This approach allows for the application of the Kalman analysis formulas to the ensemble mean and covariance matrix. The final step of an EnKF is the re-interpretation of the Kalman analysis step in terms of modified particle positions while the weights are held constant at $\gamma_i = 1/M$. We call filter algorithms that rely on modified particle/ensemble positions and fixed particle weights *ensemble transform filters*. A new ensemble transform filter has recently been proposed by Anderson (2010). The filter is based on an appropriate transformation step in observation space and subsequent linear regression of the transformation onto the full state space. The

approach developed in this paper relies instead on a general methodology for deriving ensemble transform filters as proposed by Reich (2011). See Section 2 below for a summary. The same methodology has been developed for continuous-in-time observations by Crisan and Xiong (2010). In this paper, we demonstrate how our ensemble transform filter framework can be used to generalize EnKFs to Gaussian mixture models and Gaussian kernel density estimators. The essential steps are summarized in Section 3 while an algorithmic summary of the proposed ensemble Gaussian mixture filter (EGMF) is provided in Section 4. The EGMF can also be viewed as a generalization of the continuous formulation of ensemble square root filters (Tippett et al., 2003) as provided by Bergemann and Reich (2010a,b) and the EnKF with perturbed observations, as demonstrated by Reich (2011). The paper concludes with three numerical examples in Section 5. We first demonstrate the properties of the newly proposed EGMF for one-dimensional Brownian dynamics under a double-well potential. This simulation is extended to the associated two-dimensional Langevin dynamics model with only velocities being observed. Finally we consider the three variable model of Lorenz (1963).

We mention that alternative extensions of the EnKF to Gaussian mixtures have recently been proposed, for example, by Smith (2007), Stordal et al. (2011), and Frei and Künsch (2011). However, while the cluster EnKF of Smith (2007) is an example of an ensemble transform filter, it fits the posterior (analysed) ensemble distribution back to a single Gaussian PDF and, hence, only partially works with a Gaussian mixture. Both the mixture ensemble Kalman filter of Frei and Künsch (2011) and the adaptive Gaussian mixture filter of Stordal et al. (2011) approximate the model uncertainty by a sum of Gaussian kernels and utilize the ensemble Kalman filter as a particle update step under a single Gaussian kernel. Resampling or a re-weighting of particles is required to avoid a degeneracy of particle weights due to changing kernel weights. A related filter algorithm based on Gaussian kernel density estimators has previously been considered by Anderson and Anderson (1999).

2   A general framework for ensemble transform filters

Bayes' formula can be interpreted as a discontinuous change of a forecast PDF π_f into an analyzed PDF π_a at each observation time t_j. On the other hand, one can find a continuous embedding $\pi(\mathbf{x}, s)$ with respect to a fictitious time $s \in [0, 1]$ such that $\pi(\cdot, 0) = \pi_f$ and $\pi_a = \pi(\cdot, 1)$. As proposed by Reich (2011), this embedding can be viewed as being induced by a continuity (Liouville) equation

$$\frac{\partial \pi}{\partial s} = -\nabla_{\mathbf{x}} \cdot (\pi g) \tag{5}$$

for an appropriate vector field $g(\mathbf{x}, s) \in \mathbb{R}^N$. The vector field g is not uniquely determined for a given continuous embedding $\pi(\cdot, s)$ unless we also require that it is the minimizer of the kinetic energy

$$\mathcal{T}(v) = \frac{1}{2} \int d\pi \, v^T \mathbf{M} v$$

over all admissible vector fields $v \in L^2(d\pi, \mathbb{R}^N)$, where $\mathbf{M} \in \mathbb{R}^{N \times N}$ is a positive definite mass matrix (Villani, 2003). Admissibility means that $g = v$ satisfies (5) for given π and $\partial \pi / \partial s$.

Under these assumptions, a constrained variational principle (Villani, 2003) implies that the desired vector field is given by $g = \mathbf{M}^{-1} \nabla_{\mathbf{x}} \psi$, where the potential $\psi(\mathbf{x}, s)$ is the solution of

the elliptic partial differential equation (PDE)

$$\nabla_{\mathbf{x}} \cdot \left(\pi \mathbf{M}^{-1} \nabla_{\mathbf{x}} \psi \right) = \pi \left(S - \mathbb{E}_{\pi}[S] \right) \tag{6}$$

for given PDF π, mass matrix $\mathbf{M}$, and negative log-likelihood function

$$S(\mathbf{x}; \mathbf{y}_{\text{obs}}(t_j)) = \frac{1}{2}(\mathbf{H}\mathbf{x} - \mathbf{y}_{\text{obs}}(t_j))^T \mathbf{R}^{-1}(\mathbf{H}\mathbf{x} - \mathbf{y}_{\text{obs}}(t_j)). \tag{7}$$

Here $\mathbb{E}_{\pi}[f]$ denotes the expectation value of a function $f(\mathbf{x})$ with respect to a PDF $\pi(\mathbf{x})$. We finally replace (5) by

$$\frac{\partial \pi}{\partial s} = -\nabla_{\mathbf{x}} \cdot \left(\pi \mathbf{M}^{-1} \nabla_{\mathbf{x}} \psi \right) \tag{8}$$

with an underlying ODE formulation

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}s} = \mathbf{M}^{-1} \nabla_{\mathbf{x}} \psi(\mathbf{x}, s) \tag{9}$$

in fictitious time $s \in [0, 1]$. As for the ODE (1) and its associated Liouville equation (2), we may approximate (9) and its associated Liouville equation (8) by an empirical measure of type (4). Furthermore, one and the same empirical measure approximation can now be used for both the ensemble propagation step under the model dynamics (1) and the data assimilation step (8) using constant and equal weights $\gamma_i = 1/M$. The particle filter approximation is closed by finding an appropriate numerical solution to the elliptic PDE (6). This is the crucial step which will lead to different ensemble transform filter algorithms.

The basic numerical approach to the data assimilation step within an ensemble transform filter formulation consists then of the following sequence of steps. (i) Given a current ensemble of solutions $\mathbf{x}_i(s)$, $i = 1, \ldots, M$, one fits a statistical model $\widehat{\pi}(\mathbf{x}, s)$. (ii) Solve the elliptic PDE

$$\nabla_{\mathbf{x}} \cdot (\widehat{\pi}\widehat{g}) = \widehat{\pi} \left(S - \mathbb{E}_{\widehat{\pi}}[S] \right) \tag{10}$$

for a vector field $\widehat{g}(\mathbf{x}, s)$. The solution is not uniquely determined and an appropriate choice needs to be made. See the discussion above. (iii) Propagate the ensemble members under the ODE

$$\frac{\mathrm{d}\mathbf{x}_i}{\mathrm{d}s} = \widehat{g}(\mathbf{x}_i, s). \tag{11}$$

We assume that a forecast ensemble of M members $\mathbf{x}_i \in \mathbb{R}^N$, $i = 1, \ldots, M$, is available at an observation time t_j which provides the initial conditions for the ODE (11). Solutions at $s = 1$ yield the analyzed ensemble members, which are then used as the new initial conditions for (1) at time $t = t_j$ and (1) is solved over $[t_j, t_{j+1}]$ up to the next observation point.

If the statistical model is a Gaussian with mean $\overline{\mathbf{x}} \in \mathbb{R}^N$ and covariance matrix $\mathbf{P} \in \mathbb{R}^{N \times N}$, then the outlined approach leads to a continuous formulation of the ensemble square-root ensemble filter analysis step at time t_j (Bergemann and Reich, 2010a,b), i.e.

$$\frac{\mathrm{d}\mathbf{x}_i}{\mathrm{d}s} = -\frac{1}{2}\mathbf{P}\mathbf{H}^T \mathbf{R}^{-1} \left(\mathbf{H}\mathbf{x}_i + \mathbf{H}\overline{\mathbf{x}} - 2\mathbf{y}_{\text{obs}}(t_j) \right) \tag{12}$$

for $s \in [0, 1]$. It follows that $\mathbf{M} = \mathbf{P}^{-1}$ and

$$\psi(\mathbf{x}) = -\frac{1}{4} \left(\mathbf{H}\mathbf{x} + \mathbf{H}\overline{\mathbf{x}} - 2\mathbf{y}_{\text{obs}}(t_j) \right)^T \mathbf{R}^{-1} \left(\mathbf{H}\mathbf{x} + \mathbf{H}\overline{\mathbf{x}} - 2\mathbf{y}_{\text{obs}}(t_j) \right). \tag{13}$$

3 An ensemble transform filter based on Gaussian mixture statistical models

We now develop an ensemble transform filter algorithm based on a $L \geq 1$ component Gaussian mixture model, i.e.

$$\widehat{\pi}(\mathbf{x}) = \sum_{l=1}^{L} \frac{\alpha_l}{(2\pi)^{N/2} \det \mathbf{P}_l^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \overline{\mathbf{x}}_l)^T \mathbf{P}_l^{-1}(\mathbf{x} - \overline{\mathbf{x}}_l)\right) = \sum_{l=1}^{L} \alpha_l \pi_{\mathrm{Gauss},l}(\mathbf{x}), \qquad (14)$$

where $\pi_{\mathrm{Gauss},l}(\mathbf{x})$ denotes the normal distribution in $\mathbf{x} \in \mathbb{R}^N$ with mean $\overline{\mathbf{x}}_l$ and covariance matrix $\mathbf{P}_l$. The Gaussian mixture parameters, i.e. $\alpha_l, \overline{\mathbf{x}}_l, \mathbf{P}_l$, $l = 1, \ldots, L$, need to be determined from the ensemble $\mathbf{x}_i$, $i = 1, \ldots, M$, in an appropriate manner using, for example, the expectation-maximization (EM) algorithm (Dempster et al., 1977; Smith, 2007). See Section 3.3 for more details. Note that $\sum_{l=1}^{L} \alpha_l = 1$ and $\alpha_l \geq 0$. To simplify notation, we write π_l instead of $\pi_{\mathrm{Gauss},l}$ from now on.

An implementation of the associated continuous formulation of the Bayesian analysis step proceeds as follows. To simplify the discussion we derive our filter formulation for a scalar observation variable, i.e. $K = 1$, $y_{\mathrm{obs}}(t_j) - \mathbf{H}\mathbf{x}(t_j) \sim \mathrm{N}(0, R)$, and

$$S(\mathbf{x}; y_{\mathrm{obs}}(t_j)) = \frac{1}{2R}\left(y_{\mathrm{obs}}(t_j) - \mathbf{H}\mathbf{x}\right)^2. \qquad (15)$$

The vector-valued case can be treated accordingly provided the error covariance matrix $\mathbf{R}$ is diagonal. We first decompose the vector field $\widehat{g}(\mathbf{x}, s) \in \mathbb{R}^N$ in (11) into two contributions, i.e.

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}s} = \widehat{g}(\mathbf{x}, s) = u_{\mathrm{A}}(\mathbf{x}, s) + u_{\mathrm{B}}(\mathbf{x}, s). \qquad (16)$$

To simplify notation we drop the explicit s dependence in the following calculations. We next decompose the right hand side of the elliptic PDE (10) also into two contributions

$$\widehat{\pi}\left(S(\mathbf{x}) - \mathbb{E}_{\widehat{\pi}}[S]\right) = \left\{\sum_{l=1}^{L} \alpha_l \pi_l \left(S(\mathbf{x}) - \mathbb{E}_{\pi_l}[S]\right)\right\} + \left\{\sum_{l=1}^{L} \alpha_l \pi_l \left(\mathbb{E}_{\pi_l} - \mathbb{E}_{\widehat{\pi}}[S]\right)\right\}. \qquad (17)$$

We now derive explicit expressions for $u_{\mathrm{A}}(\mathbf{x})$ and $u_{\mathrm{B}}(\mathbf{u})$.

3.1 Gaussian mixture Kalman filter contributions

We define the vector field $u_{\mathrm{A}}(\mathbf{x})$ through the equation

$$u_{\mathrm{A}}(\mathbf{x}) = \sum_{l=1}^{L} \frac{\alpha_l \pi_l(\mathbf{x})}{\widehat{\pi}(\mathbf{x})} \mathbf{P}_l \nabla_{\mathbf{x}} \psi_{\mathrm{A},l}(\mathbf{x}), \qquad (18)$$

together with

$$\nabla_{\mathbf{x}} \cdot \{\widehat{\pi}(\mathbf{x}) u_{\mathrm{A}}(\mathbf{x})\} = \nabla_{\mathbf{x}} \cdot \left\{\sum_{l=1}^{L} \alpha_l \pi_l(\mathbf{x}) \mathbf{P}_l \nabla_{\mathbf{x}} \psi_{\mathrm{A},l}(\mathbf{x})\right\} = \sum_{l=1}^{L} \alpha_l \pi_l(\mathbf{x})(S(\mathbf{x}) - \mathbb{E}_{\pi_l}[S]) \qquad (19)$$

5

which implies that the potentials $\psi_{A,l}(\mathbf{x})$, $l = 1, \ldots, L$, are uniquely determined by

$$\nabla_{\mathbf{x}} \cdot \{\pi_l(\mathbf{x})\mathbf{P}_l\nabla_{\mathbf{x}}\psi_{A,l}(\mathbf{x})\} = \pi_l(\mathbf{x})(S(\mathbf{x}) - \mathbb{E}_{\pi_l}[S]) \tag{20}$$

for all $l = 1, \ldots, L$. It follows that the potentials $\psi_{A,l}(\mathbf{x})$ are equivalent to the ensemble Kalman filter potentials for the l-th Gaussian component. Hence, using (12) and (18), we obtain

$$u_A(\mathbf{x}, s) = -\frac{1}{2}\sum_{l=1}^{L}\frac{\alpha_l(s)\pi_l(\mathbf{x}, s)}{\widehat{\pi}(\mathbf{x}, s)}\mathbf{P}_l(s)\mathbf{H}^T R^{-1}\left[\mathbf{H}\mathbf{x}(s) + \mathbf{H}\overline{\mathbf{x}}_l(s) - 2y_{\text{obs}}(t_j)\right]. \tag{21}$$

3.2 Gaussian mixture exchange contributions

The remaining contributions for solving (5) are collected in the vector field

$$u_B(\mathbf{x}) = \sum_{l=1}^{L}\frac{\alpha_l\pi_l(\mathbf{x})}{\widehat{\pi}(\mathbf{x})}\mathbf{P}_l\nabla_{\mathbf{x}}\psi_{B,l}(\mathbf{x}), \tag{22}$$

which therefore needs to satisfy

$$\nabla_{\mathbf{x}} \cdot \{\widehat{\pi}(\mathbf{x})u_B(\mathbf{x})\} = \nabla_{\mathbf{x}} \cdot \left\{\sum_{l=1}^{L}\alpha_l\pi_l(\mathbf{x})\mathbf{P}_l\nabla_{\mathbf{x}}\psi_{B,l}(\mathbf{x})\right\} = \sum_{l=1}^{L}\alpha_l\pi_l(\mathbf{x})(\mathbb{E}_{\pi_l}[S] - \mathbb{E}_{\widehat{\pi}}[S]) \tag{23}$$

and, hence, we may set

$$\nabla_{\mathbf{x}} \cdot \{\pi_l(\mathbf{x})\mathbf{P}_l\nabla_{\mathbf{x}}\psi_{B,l}(\mathbf{x})\} = \pi_l(\mathbf{x})(\mathbb{E}_{\pi_l}[S] - \mathbb{E}_{\widehat{\pi}}[S]) \tag{24}$$

for all $l = 1, \ldots, L$. To find a solution of (24) we introduce functions $\widehat{\psi}_{B,l}$ such that

$$\psi_{B,l}(\mathbf{x}) = \widehat{\psi}_{B,l}(\mathbf{H}\mathbf{x} - \mathbf{H}\overline{\mathbf{x}}_l) = \widehat{\psi}_{B,l}(y - \overline{y}_l) \tag{25}$$

with $y := \mathbf{H}\mathbf{x}$ and $\overline{y}_l = \mathbf{H}\overline{\mathbf{x}}_l$. Now the right hand side of (24) gives rise to

$$\nabla_{\mathbf{x}} \cdot \{\pi_l(\mathbf{x})\mathbf{P}_l\nabla_{\mathbf{x}}\psi_{B,l}(\mathbf{x})\} = \pi_l(\mathbf{x})\left(-(y - \overline{y}_l)\frac{d\widehat{\psi}_{B,l}}{dy}(y - \overline{y}_l) + \mathbf{H}\mathbf{P}_l\mathbf{H}^T\frac{d^2\widehat{\psi}_{B,l}}{dy^2}(y - \overline{y}_l)\right) \tag{26}$$

and (24) simplifies further to the scalar PDE

$$-(y - \overline{y}_l)\frac{d\widehat{\psi}_{B,l}}{dy}(y - \overline{y}_l) + \mathbf{H}\mathbf{P}_l\mathbf{H}^T\frac{d^2\widehat{\psi}_{B,l}}{dy^2}(y - \overline{y}_l) = \mathbb{E}_{\pi_l}[S] - \mathbb{E}_{\widehat{\pi}}[S]. \tag{27}$$

The PDE (27) can be solved for

$$f(z) = \frac{d\widehat{\psi}_{B,l}}{dy}(y - \overline{y}_l), \qquad z = y - \overline{y}_l, \tag{28}$$

under the condition $f(0) = 0$ by explicit quadrature and one obtains

$$f(y - \overline{y}_l) = \frac{1}{2}\frac{\mathbb{E}_{\pi_l}[S] - \mathbb{E}_{\widehat{\pi}}[S]}{\mathbf{H}\mathbf{P}_l\mathbf{H}}\frac{\text{erf}\left((y - \overline{y}_l)/\sqrt{2\sigma_l^2}\right)}{\pi_l(y)} \tag{29}$$

with marginalized PDF

$$\pi_l(y) := \frac{1}{\sqrt{2\pi\sigma_l^2}} \exp\left(-\frac{(y - \overline{y}_l)^2}{2\sigma_l^2}\right), \tag{30}$$

$\sigma_l = \sqrt{\mathbf{H}\mathbf{P}_l\mathbf{H}^T}$, and the standard error function

$$\mathrm{erf}(y) = \frac{2}{\sqrt{\pi}} \int_0^y e^{-s^2} \mathrm{d}s. \tag{31}$$

Note that

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}y}\mathrm{erf}\left((y - \overline{y}_l)/\sqrt{2\sigma_l^2}\right) = \pi_l(y). \tag{32}$$

We finally obtain the expression

$$u_{\mathrm{B}}(\mathbf{x}, s) = \frac{1}{2}\sum_{l=1}^{L}\frac{\alpha_l(s)\pi_l(\mathbf{x}, s)}{\widehat{\pi}(\mathbf{x}, s)}\mathbf{P}_l(s)\mathbf{H}^T\frac{\mathbb{E}_{\pi_l}[S](s) - \mathbb{E}_{\widehat{\pi}}[S](s)}{\sigma_l^2}\frac{\mathrm{erf}\left((y - \overline{y}_l)/\sqrt{2\sigma_l^2}\right)}{\pi_l(y)} \tag{33}$$

for the vector field $u_{\mathrm{B}}(\mathbf{x}, s)$.

The expectation values $\mathbb{E}_{\pi_l}[S]$, $l = 1, \ldots, L$, can be computed analytically, i.e.

$$\mathbb{E}_{\pi_l}[S] = \frac{1}{2R}\left((y_{\mathrm{obs}}(t_j) - \overline{y}_l)^2 + \sigma_l^2\right) \tag{34}$$

or estimated numerically. Recall that $\sum \alpha_l = 1$ and, therefore,

$$\mathbb{E}_{\widehat{\pi}}[S] = \sum_{l=1}^{L}\alpha_l\mathbb{E}_{\pi_l}[S]. \tag{35}$$

It should be kept in mind that the Gaussian mixture parameters $\alpha_l, \overline{\mathbf{x}}_l, \mathbf{P}_l$ can be updated directly using the differential equations

$$\frac{\mathrm{d}\overline{\mathbf{x}}_l}{\mathrm{d}s} = -\mathbf{P}_l\mathbf{H}^T R^{-1}(\mathbf{H}\overline{\mathbf{x}}_l - \mathbf{y}_{\mathrm{obs}}(t_j)), \tag{36}$$

$$\frac{\mathrm{d}\mathbf{P}_l}{\mathrm{d}s} = -\mathbf{P}_l\mathbf{H}^T R^{-1}\mathbf{H}\mathbf{P}_l, \tag{37}$$

$$\frac{\mathrm{d}\alpha_l}{\mathrm{d}s} = -\frac{1}{2}\alpha_l\left\{(\mathbf{H}\overline{\mathbf{x}}_l - \mathbf{y}_{\mathrm{obs}}(t_j))^T R^{-1}(\mathbf{H}\overline{\mathbf{x}}_l - \mathbf{y}_{\mathrm{obs}}(t_j)) + \lambda\right\}, \tag{38}$$

for $l = 1, \ldots, L$. Here $\lambda \in \mathbb{R}$ is chosen such that

$$\sum_{l=1}^{L}\frac{\mathrm{d}\alpha_l}{\mathrm{d}s} = 0. \tag{39}$$

Furthermore, $u_{\mathrm{A}}(\mathbf{x}, s)$ exactly mirrors the update of the Gaussian mixture means $\overline{\mathbf{x}}_l$ and co-variance matrices $\mathbf{P}_l$, while $u_{\mathrm{B}}(\mathbf{x}, s)$ mimics the update of the weight factors α_l by rearranging the particle positions accordingly.

As already eluded to, we can treat each uncorrelated observation separately and sum the individual contributions in $u_{\mathrm{A}}(\mathbf{x}, s)$ and $u_{\mathrm{B}}(\mathbf{x}, s)$, respectively, to obtain the desired total vector field (16).

3.3 Implementation aspects

Given a set of ensemble members $\mathbf{x}_i$, $i = 1, \ldots, M$, there are several options for implementing a Gaussian mixture filter. The trivial case $L = 1$ leads back to the continuous formulations of Bergemann and Reich (2010a,b). More interestingly, one can chose $L \ll M$ and estimate the mean and the covariance matrices for the Gaussian mixture model using the EM algorithm (Dempster et al., 1977; Smith, 2007). The EM algorithm self-consistently computes the mixture mean $\overline{\mathbf{x}}_l$ and covariance matrix $\mathbf{P}_l$ via

$$\overline{\mathbf{x}}_l = \frac{1}{\sum_{i=1}^{M} \beta_{i,l}} \sum_{i=1}^{M} \beta_{i,l} \mathbf{x}_i, \qquad \mathbf{P}_l = \frac{1}{\sum_{i=1}^{M} \beta_{i,l}} \sum_{i=1}^{M} \beta_{i,l} \left(\mathbf{x}_i - \overline{\mathbf{x}}_l \right) \left(\mathbf{x}_i - \overline{\mathbf{x}}_l \right)^T \tag{40}$$

for $l = 1, \ldots, L$ using weights $\beta_{i,l}$ defined by

$$\beta_{i,l} = \frac{\alpha_l \pi_l(\mathbf{x}_i)}{\sum_{k=1}^{M} \alpha_k \pi_k(\mathbf{x}_i)}, \qquad \alpha_l = \frac{1}{M} \sum_{i=1}^{M} \beta_{i,l}. \tag{41}$$

The EM algorithm can fail to converge and a possible remedy is to add a constant contribution $\delta \mathbf{I}$ to the empirical covariance matrix in (40) with the parameter $\delta > 0$ appropriately chosen. We mention that more refined implementations of the EM algorithm, such as those discussed by Fraley and Raftery (2007), could also be considered. It is also possible to select the number of mixture components adaptively. See, for example, Smith (2007). The resulting vector fields for the ith ensemble member are given by

$$u_{\mathrm{A}}(\mathbf{x}_i, s) = -\frac{1}{2} \sum_{l=1}^{L} \beta_{i,l}(s) \mathbf{P}_l(s) \mathbf{H}^T R^{-1} \left[\mathbf{H} \mathbf{x}_i(s) + \mathbf{H} \overline{\mathbf{x}}_l(s) - 2 y_{\mathrm{obs}}(t_j) \right] \tag{42}$$

and, using (33),

$$u_{\mathrm{B}}(\mathbf{x}_i, s) = \frac{1}{2} \sum_{l=1}^{L} \beta_{i,l}(s) \mathbf{P}_l(s) \mathbf{H}^T \frac{\mathbb{E}_{\pi_l}[S](s) - \mathbb{E}_{\widehat{\pi}}[S](s)}{\sigma_l^2} \frac{\mathrm{erf}\left((y_i - \overline{y}_l)/\sqrt{2\sigma_l^2} \right)}{\pi_l(y_i)} \tag{43}$$

with weights $\beta_{i,l}$ given by (41) and $y_i = \mathbf{H} \mathbf{x}_i$.

Another option to implement an EGMF is to set the number of mixture components equal to the number of ensemble members, i.e. $L = M$, and to use a prescribed covariance matrix $\mathbf{B}$ for all mixture components, i.e. $\mathbf{P}_l = \mathbf{B}$ and $\overline{\mathbf{x}}_l = \mathbf{x}_l$, $l = 1, \ldots, L$. We also give all mixture components equal weights $\alpha_l = 1/M$. In this setting, it is more appropriate to call (14) a kernel density estimator (Wand and Jones, 1995). Then

$$u_{\mathrm{A}}(\mathbf{x}_i, s) = -\frac{1}{2} \sum_{l=1}^{M} \beta_{i,l}(s) \mathbf{B} \mathbf{H}^T R^{-1} \left[\mathbf{H} \mathbf{x}_i(s) + \mathbf{H} \mathbf{x}_l(s) - 2 y_{\mathrm{obs}}(t_j) \right] \tag{44}$$

and

$$u_{\mathrm{B}}(\mathbf{x}_i, s) = \frac{1}{2} \sum_{l=1}^{M} \beta_{i,l}(s) \mathbf{B} \mathbf{H}^T \frac{\mathbb{E}_{\pi_l}[S](s) - \mathbb{E}_{\widehat{\pi}}[S](s)}{\mathbf{H} \mathbf{B} \mathbf{H}^T} \frac{\mathrm{err}\left((y_i - y_l)/\sqrt{2\sigma_l^2} \right)}{\pi_l(y_i)} \tag{45}$$

with weights $\beta_{i,l}$ given by (41), $\sigma_l = \sqrt{\mathbf{HBH}^T}$, and $\alpha_l = 1/M$. The Kalman filter like contributions (44) can be replaced by a formulation with perturbed observations (Evensen, 2006; Reich, 2011) which yields

$$u_{\mathrm{A}}(\mathbf{x}_i, s) = -\mathbf{BH}^T R^{-1} \left[\mathbf{Hx}_i(s) - y_{\mathrm{obs}}(t_j) + d_i \right], \tag{46}$$

where $d_i \in \mathbb{R}$, $i = 1, \ldots, m$, are independent, identically distributed Gaussian random numbers with mean zero and variance R. A particular choice is $\mathbf{B} = c\mathbf{P}$, where $\mathbf{P}$ is the empirical covariance matrix of the ensemble and $c > 0$ is an appropriate constant. Assuming that the underlying probability density is Gaussian with covariance $\mathbf{P}$, the choice

$$c = (2/(N+2))^{4/(N+4)} M^{-2/(N+4)} \tag{47}$$

is optimal for large ensemble sizes M in the sense of kernel density estimation (see, e.g. Wand and Jones (1995)). Recall that N denotes the dimension of phase space. The resulting filter is then similar in spirit to the rank histogram filter (RHF) suggested by Anderson (2010) with the RHF increments in observation space being replaced by those generated from a Gaussian kernel density estimator. Another choice is $\mathbf{B} \approx \mathbf{P}$ in which case (45) can be viewed as a correction term to the standard ensemble Kalman filter (46). We will explore the kernel estimator in the numerical experiment of Section 5.4.

Note that localization, as introduced by Houtekamer and Mitchell (2001) and Hamill et al. (2001), can be combined with (42)-(43) and (44)-(45), respectively, as outlined in Bergemann and Reich (2010a). For example, one could set the covariance matrix $\mathbf{B}$ in (44)-(45) equal to the localized ensemble covariance matrix. Localization will be essential for a successful application of the proposed filter formulations to high-dimensional systems (1). The same applies to ensemble inflation (Anderson and Anderson, 1999).

We also note that the computation of the particle-mixture weight factors (41) can be become expensive since it requires the computation of $\mathbf{P}_l^{-1}$. This can be avoided by either using only the diagonal part of $\mathbf{P}_l$ in $\pi_l(\mathbf{x}_i)$ (Smith, 2007) or by using a marginalized density such as (30), i.e. $\pi_l(y_i)$, $y_i := \mathbf{Hx}_i$, instead of the full Gaussian PDF values $\pi_l(\mathbf{x}_i)$. Some other suitable marginalization could also be performed.

The vector field $u_{\mathrm{B}}(\mathbf{x}, s)$ is responsible for a transfer of particles between different mixture components according to the observation adjusted relative weight α_l of each mixture component. These transitions can be rather rapid implying that $\|u_{\mathrm{B}}(\mathbf{x}, s)\|_\infty$ can become large in magnitude. This can pose numerical difficulties and we eliminated those by limiting the l_∞-norm of $u_{\mathrm{B}}(\mathbf{x}, s)$ through a cut-off value u_{cut}. Alternatively, we might want to replace (30) by a PDF which leads to less stiff contributions to the vector field $u_{\mathrm{B}}(\mathbf{x}, s)$ such as the Student's t-distributions (Schaefer, 1997). Hence a natural approximative PDF is provided by the the scaled t-distribution with three degrees of freedom, i.e.

$$\phi(y; \bar{y}, \sigma) = \frac{2\sigma^3}{\pi} \frac{1}{\left(\sigma^2 + (y - \bar{y})^2\right)^2}. \tag{48}$$

We also introduce the shorthand $\phi_l(y) = \phi(y; \bar{y}_l, \sigma_l)$ with $\bar{y}_l = \mathbf{H\bar{x}}_l$ and $\sigma_l = \sqrt{\mathbf{HP}_l\mathbf{H}^T}$. A first observation is that

$$\mathbb{E}_{\phi_l}[y] = \bar{y}_l, \qquad \mathbb{E}_{\phi_l}[(y - \bar{y}_l)^2] = \sigma_l^2, \tag{49}$$

i.e., the first two moments of ϕ_l match those of (30). The second observation is that ϕ_l can be integrated explicitly, i.e.

$$\Phi_l(y) = \int_{\bar{y}_l}^{y} \phi_l(u)\mathrm{d}u = \frac{1}{\pi}\arctan\left(\frac{(y-\bar{y}_l)}{\sigma_l}\right) + \frac{\sigma_l}{\pi}\frac{(y-\bar{y}_l)}{\sigma_l^2 + (y-\bar{y}_l)^2}. \tag{50}$$

Hence the relation (32) suggests the alternative formulation

$$u_{\mathrm{B}}(\mathbf{x},s) = \sum_{l=1}^{L}\frac{\alpha_l(s)\pi_l(\mathbf{x},s)}{\widehat{\pi}(\mathbf{x},s)}\mathbf{P}_l(s)\mathbf{H}^T\frac{\mathbb{E}_{\pi_l}[S](s) - \mathbb{E}_{\widehat{\pi}}[S](s)}{\sigma_l^2}\frac{\Phi_l(y,s))}{\phi_l(y,s)}, \tag{51}$$

where $\phi_l(y,s) = \phi(y; \bar{y}_l(s), \sigma_l(s))$.

The differential equation (16) needs to be approximated by a numerical time-stepping procedure. In this paper, we use the forward Euler method for simplicity. However, the limited region of stability of an explicit method such as forward Euler implies that the step-size Δs needs to be made sufficiently small. This issue has been investigated by Amezcua et al. (2011) for the formulation with $L = 1$ (standard continuous ensemble Kalman filter formulation) and a diagonally implicit scheme has been proposed which overcomes the stability restrictions of the forward Euler method while introducing negligible computational overhead. The computational cost of a single evaluation of (16) for given mixture components should be slightly lower than for a single EnKF step since no matrix inversion is required. Additional expenses arise from fitting the mixture components (e.g. using the EM algorithm) and from having to use a number of time-steps $1/\Delta s > 1$.

4 Algorithmic summary of the ensemble Gaussian mixture filter (EGMF)

Since the proposed methodology for treating nonlinear filter problems is based on an extension of the EnKF approach, we call the new filter the ensemble Gaussian mixture filter (EGMF). We now provide an algorithmic summary.

First a set of M ensemble members $\mathbf{x}_i(0)$ is generated at time t_0 according to the initial PDF π_0.

In between observations, the ensemble members are propagated under the ODE model (1), i.e.

$$\dot{\mathbf{x}}_i = f(\mathbf{x}_i, t), \tag{52}$$

for $i = 1, \ldots, M$ and $t \in [t_{j-1}, t_j]$.

At an observation time t_j, a Gaussian mixture model (14) is fitted to the ensemble members $\mathbf{x}_i$, $i = 1, \ldots, M$. One can, for example, use the classic EM algorithm (Dempster et al., 1977; Smith, 2007) for this purpose. In this paper we use a simple heuristic to determine the number of components $L \in \{1, 2\}$. An adaptive selection of L is, however, feasible (see, e.g., Smith (2007)). Alternatively, one can set $L = M$ and implement the EGMF with a Gaussian kernel density estimator with $\overline{\mathbf{x}}_l = \mathbf{x}_l$, $\alpha_l = 1/M$. The covariance matrix $\mathbf{B}$ can be based on the empirical covariance matrix $\mathbf{P}$ of the whole ensemble via $\mathbf{B} = c\mathbf{P}$ with the constant $c > 0$ appropriately chosen. At this stage covariance localization can also be applied.
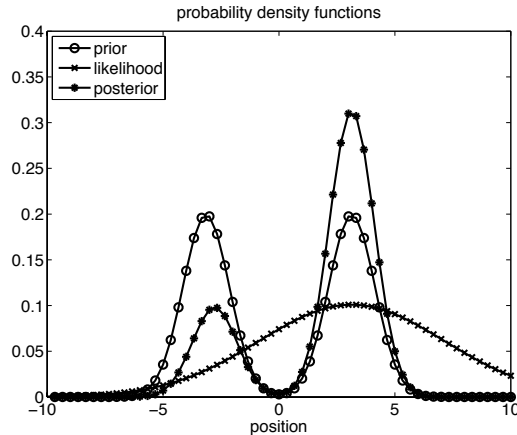
Figure 1: Displayed are the prior distribution, the likelihood from a measurement and the resulting posterior distribution. The prior as well as the posterior are bimodal Gaussian.

The vector fields $u_{\mathrm{A}}(\mathbf{x}, s)$ and $u_{\mathrm{B}}(\mathbf{x}, s)$ are computed according to (42) and (43), respectively, (or, alternatively, use (45)-(46)) for each independent observation and the resulting contributions are summed up to a total vector field $\widehat{g}(\mathbf{x}, s)$. Next the ensemble members are updated according to (16) for $\mathbf{x} = \mathbf{x}_i$, $i = 1, \ldots, M$. Here we use a simple forward Euler discretization with step-size Δs chosen sufficiently small. After each time-step the Gaussian mixture components are updated, if necessary, using the EM algorithm. The analyzed ensemble members $\mathbf{x}_i(t_j)$ are obtained after $1/\Delta s$ time-steps as the numerical solutions at $s = 1$ and provide the new initial conditions for (52) with time t now in the interval $[t_j, t_{j+1}]$.

Ensemble induced estimates for the expectation value of a function $f(\mathbf{x})$ can be computed via

$$\overline{f} = \frac{1}{M} \sum_{i=1}^{M} f(\mathbf{x}_i) \tag{53}$$

without reference to a Gaussian mixture model.

5   Numerical experiments

In this section we provide results from several numerical simulations and demonstrate the performance of the proposed EGMF in comparison with standard implementations of the EnKF and an implementation of the RHF (Anderson, 2010). We first investigate the Bayesian assimilation step without model equations.

5.1   Single Bayesian assimilation step

We test our formulation first for a single assimilation step where the prior is a bimodal Gaussian

$$\pi_{\mathrm{prior}}(x) = \frac{1}{2} \frac{1}{\sqrt{2\pi}} e^{-(x-\pi)^2/2} + \frac{1}{2} \frac{1}{\sqrt{2\pi}} e^{-(x+\pi)^2/2} \tag{54}$$

and the likelihood is

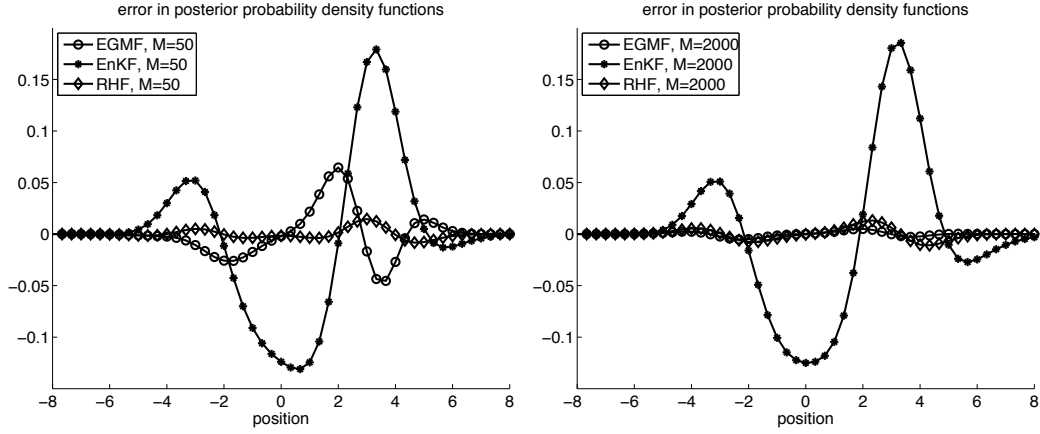$$\pi(y_{\mathrm{obs}}|x) = \frac{1}{\sqrt{2\pi}4} e^{-(x-\pi)^2/32}. \tag{55}$$

Figure 2: Numerically obtained posterior for ensemble sizes $M = 50$ (left panel) and $M = 2000$ (right panel). Shown are results from the EGMF, the RHF, and an EnKF analysis step. While the EGMF and the RHF converge to the correct posterior distribution, the EnKF leads to qualitatively incorrect results for both ensemble sizes.
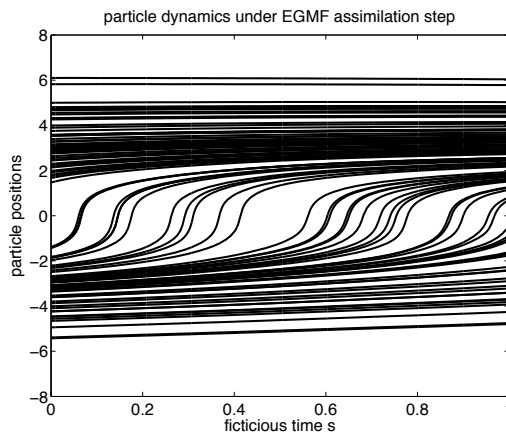

Figure 3: Displayed is the rearrangement of the particles under the dynamics of the EGMF analysis step. Rapid transitions between the Gaussian mixture components are induced by the vector field u_B.
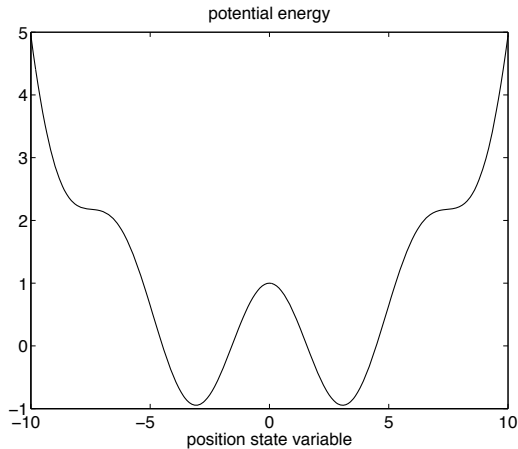
Figure 4: Shown is the potential energy V used in both the Brownian and Langevin dynamics model.
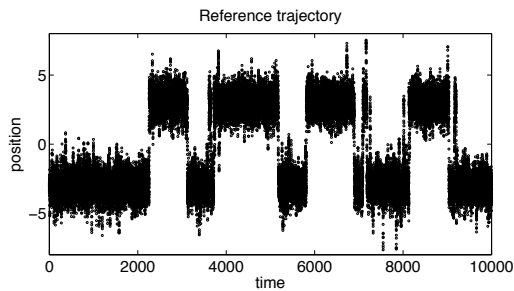
Figure 5: Shown is the reference solution from which observations are generated by adding Gaussian noise with mean zero and variance R.

The posterior distribution is again bimodal Gaussian and can be computed analytically. See Fig. 1. Here we demonstrate how an EnKF, the RHF, and the proposed EGMF approximate the posterior for ensemble sizes $M = 50, 2000$ and for $x_i(0) \sim \pi_{\mathrm{prior}}$, $i = 1, \ldots, M$. See Fig. 2. Both the RHF and the EGMF are capable of reproducing the Bayesian assimilation step correctly for M sufficiently large while the EnKF leads to a qualitatively incorrect result. The transformation of the ensemble members (particles) under the dynamics (16) is displayed in Fig. 3.

We now present results from three increasingly complex filtering problems.

5.2   A one-dimensional model problem

As a first numerical example we consider Brownian dynamics under a one-dimensional potential $V(x)$, i.e.

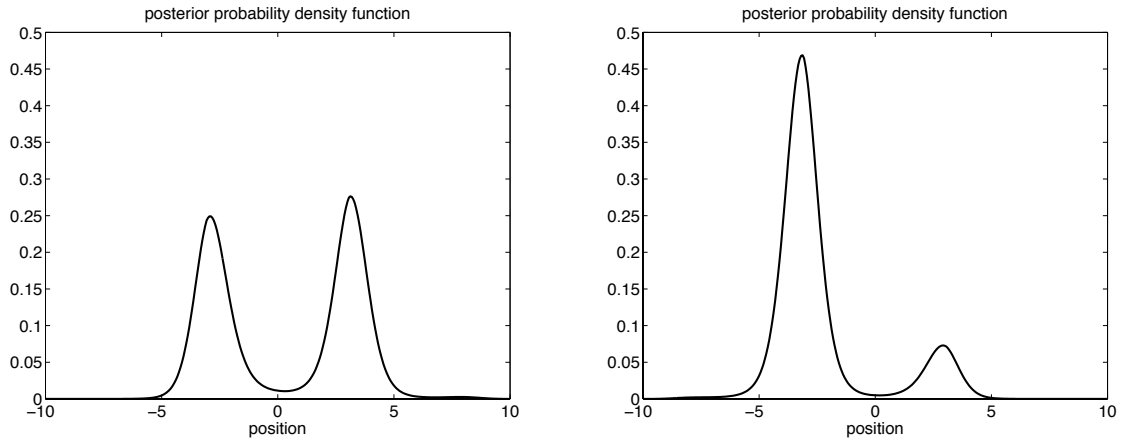$$\mathrm{d}x = -V'(x)\mathrm{d}t + \mathrm{d}w(t), \tag{56}$$

13

Figure 6: Displayed are two posterior PDFs for measurement error variance $R = 36$ as obtained from the Fokker-Planck approach. A distinct bimodal behavior can be observed which motivates the use of a binary Gaussian mixture model for the EGMF.
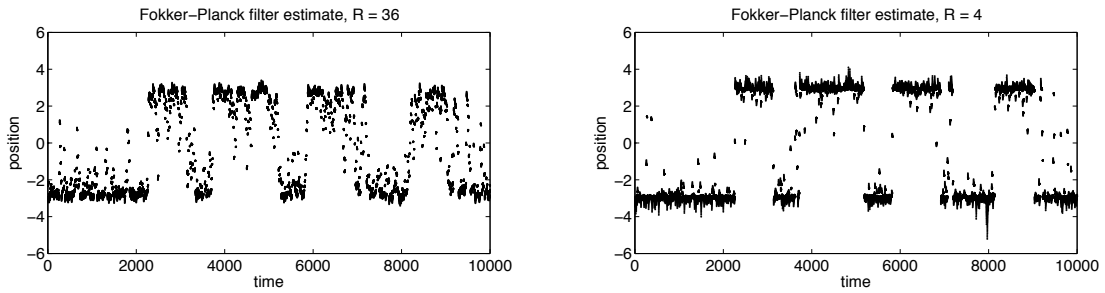

Figure 7: Estimated ensemble mean computed from a direct simulation of the assimilation problem using a discretized Fokker-Planck equation for measurement error variance $R = 36$ (left panel) and $R = 4$ (right panel).
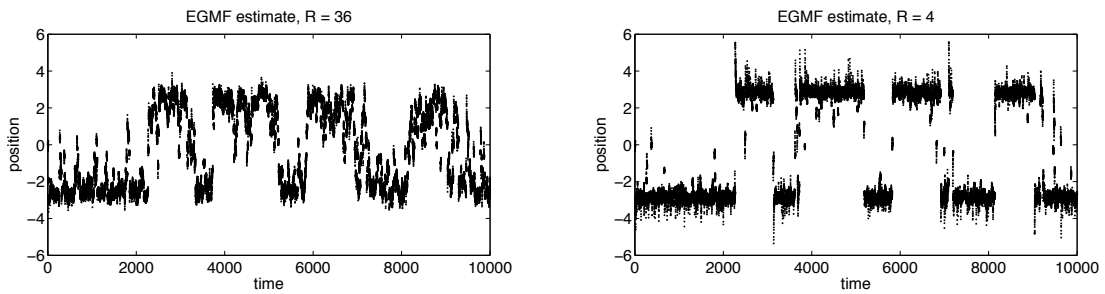

Figure 8: Ensemble mean from the EGMF for measurement error variance $R = 36$ (left panel) and $R = 4$ (right panel). It can be observed that the EGMF leads to results similar to those from the discrete Fokker-Planck approach (Fig. 7).
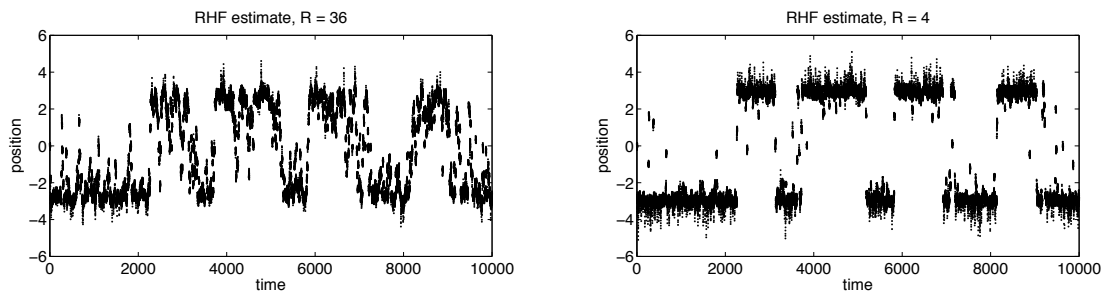
Figure 9: Ensemble mean from a RHF for measurement error variance $R = 36$ (left panel) and $R = 4$ (right panel). The results are for $M = 50$ particles. It can be observed that the RHF leads to results similar to those from the discrete Fokker-Planck approach (Fig. 7).
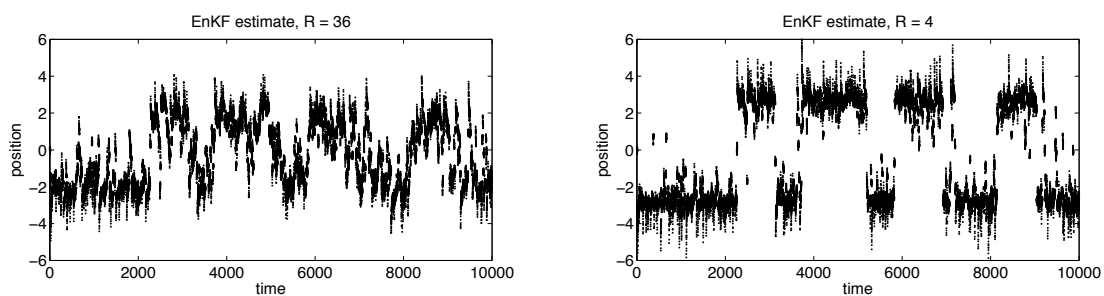

Figure 10: Ensemble mean from an EnKF with perturbed observations for measurement error variance $R = 36$ (left panel) and $R = 4$ (right panel). The results are for $M = 50$ particles. The EnKF behaves not as well as the RHF, the EGMF, and the discretized Fokker-Planck approach. Similar results are obtained for an ensemble square root filter.

where $w(t)$ denotes standard Brownian motion and the potential is given by

$$V(x) = \cos(x) + \frac{3}{4}(x/6)^4. \tag{57}$$

See Fig. 4. The true reference trajectory is started at $x(0) = -3.14$. See Fig. 5. Measurements of $x(t)$ are collected every 10 time units with two different values of the measurement error variance R ($R = 4, 36$).

The initial PDF is given by the bimodal distribution

$$\pi_0(x) \propto e^{-(x-3.14)^2/2} + e^{-(x+3.14)^2/2}. \tag{58}$$

Depending on the distribution of ensemble members we either use a single Gaussian ($L = 1$) or a bi-Gaussian mixture ($L = 2$) in the EGMF assimilation step. A single Gaussian is used whenever more than 90% of the particles are located near either the right (i.e. $x_i > 0$) or left potential well (i.e. $x_i < 0$). The computed variances are modified such that $\sigma_l^2 \geq 0.0005$ to avoid singularities in the EM algorithm. The model equation (56) is discretized with the forward Euler method and time-step $\Delta t = 0.1$. The total simulation interval is $t \in [0, 100000]$. The assimilation equations with (42) and (43) are discretized with the forward Euler method and step-size $\Delta s = 0.05$. The l_∞-norm of $u_B(\mathbf{x}, s)$ is limited to a value of $u_{\mathrm{cut}} = 5/\Delta s$.

The performance of the EGMF is compared to an EnKF with perturbed observations, an ensemble square root filter, and a RHF. The particle positions are adjusted during each data assimilation step of a RHF such that the particles maintain equal weights $\gamma_i = 1/M$. The adjustment is done similar to what has been proposed by Anderson (2010) except that the posterior is approximated by piecewise constant functions.

For this simple, one-dimensional problem the densities can be directly propagated through a discretization of the associated Fokker-Planck equation. Bayes theorem reduces to a multiplication of the prior PDF approximation from the Fokker-Planck approximation with the likelihood at each grid point. We have used a grid with mesh-size $\Delta x = 0.125$ over $x \in [-10, 10]$ to provide an independent and accurate filtering result. Periodic boundary conditions are used for the diffusion operator such that the spatial discretization leads to a stochastic matrix. It is found from the numerical simulations that $R = 36$ leads to a pronounced bimodal behavior of the solution PDF π. See Figure 6 for two posterior PDF approximations from the Fokker-Planck approach.

Typical filter behaviors over the time interval $t \in [0, 10000]$ with regard to the reference trajectory can be found in Figures 7, 8, 9, and 10, respectively, for $M = 50$ ensemble members. The EGMF and the RHF display a behavior similar to that from the discretized Fokker-Planck approach while significantly different results are obtained from the EnKF implementation with perturbed observations. Similar results are obtained for an ensemble square root filter (not displayed). The EGMF uses a bi-Gaussian approximation in 97% of the assimilation steps for $R = 36$ and in 47% of the assimilation steps for $R = 4$.

We also provide the root mean square (RMS) error between the computed mean from the three different filters and the mean computed from the Fokker-Planck approach in Table 1 for $R = 36$ and different ensemble sizes M. The RHF converges as $M \to \infty$ to the solution of the Fokker-Planck approach for this one-dimensional model problem. The EGMF provides better results than the EnKF but does not converge since the limiting distributions are not exactly bi-Gaussian. Note that the EGMF should converge for $M \to \infty$ and the number of mixture components sufficiently large. Note also that the RHF does not converge to the analytic filter

Table 1: RMS errors for ensemble means obtained from EnKF, RHF, and EGMF compared to the expected value computed by a Fokker-Planck discretization with error variance $R = 36$ and $M = 20, 50, 100$ particles/ensemble members.

	RHF	EGMF	EnKF
$M = 20$	0.6551	0.7683	1.0283
$M = 50$	0.3717	0.5127	0.8798
$M = 100$	0.2691	0.4033	0.8412

solution as $M \to \infty$ in case of more than one dynamic variable (i.e. $N > 1$). See also the following example.

5.3 A two-dimensional model problem

We discuss another example from classical mechanics. The evolution of a particle with position $q \in \mathbb{R}$ and velocity $v \in \mathbb{R}$ is described by Langevin dynamics (Gardiner, 2004) with equations of motion

$$
\begin{aligned}
dq &= v \, dt, &(59) \\
dv &= -V'(q) \, dt - \gamma v \, dt + \sigma \, dw(t), &(60)
\end{aligned}
$$

where the potential $V(q)$ is given by

$$
V(q) = \cos(q) + \frac{3}{4}(q/6)^4, \tag{61}
$$

the friction coefficient is $\gamma = 0.25$, $w(t)$ denotes standard Brownian motion, and $\sigma^2 = 0.35$. A reference solution, denoted by $(q_r(t), v_r(t))$, is obtained for initial condition $(q_0, v_0) = (1, 1)$ and a particular realization of $w(t)$.

Let us address the situation that the reference solution is not directly accessible to us and that instead we are only able to observe $Q(t)$ subject to

$$
dQ(t) = v_r(t) \, dt + c^{1/2} d\xi(t), \tag{62}
$$

where $\xi(t)$ denotes again standard Brownian motion and $c = 0.2$. In other words, we are effectively only able to observe particle velocities.

We now combine the model equations and the observations within the ensemble Kalman-Bucy framework. The ensemble filter relies on the simultaneous propagation of an ensemble of solutions $x_i(t) = (q_i(t), v_i(t)) \in \mathbb{R}^2$, $i = 1, \ldots, M$. In our experiment we set $M = 50$. The filter equations for an EGMF with a single Gaussian, ensemble Kalman-Bucy filter (Bergemann and Reich, 2011), respectively, become

$$
dq_i = v_i \, dt - \frac{P_{qv}}{2c}(v_i \, dt + \bar{v} \, dt - 2dQ(t)), \tag{63}
$$

$$
dv_i = -V'(q_i) \, dt - \gamma v_i \, dt + \sigma \, dw_i(t) - \frac{P_{vv}}{2c}(v_i \, dt + \bar{v} \, dt - 2dQ(t)) \tag{64}
$$

17

with mean

$$\bar{v} = \frac{1}{M} \sum_{i=1}^{M} v_i, \qquad \bar{q} = \frac{1}{M} \sum_{i=1}^{M} q_i \qquad (65)$$

and variance/covariance

$$P_{vv} = \frac{1}{M-1} \sum_{i=1}^{M} (v_i - \bar{v})^2, \qquad P_{qv} = \frac{1}{M-1} \sum_{i=}^{M} (q_i - \bar{q})(v_i - \bar{v}). \qquad (66)$$

The equations are solved for each ensemble member with different realizations $w_i(t)$ of standard Brownian motion and step-size $\Delta t = 0.01$. The observation interval in (62) is $\tau = \Delta t$. The extension of the EGMF to a Gaussian mixture with $L = 2$ is straightforward. One substitutes $y = v$, $R = c/\Delta t$, and $y_{\mathrm{obs}} = \Delta Q(t_n)/\Delta t$ with

$$\Delta Q(t_n) = v_r(t_n)\Delta t + \sqrt{c\Delta t}\xi_n, \qquad \xi_n \sim N(0,1), \qquad (67)$$

into (42) and (43) and sets $\Delta s = 1$ in the numerical time-stepping procedure for the assimilation step. The l_∞-norm of $u_{\mathrm{B}}(\mathbf{x}, s)$ is limited to a value of $u_{\mathrm{cut}} = 0.25/\Delta s$. Assimilation is performed at every model time-step. We perform a total of two million time-steps/data assimilation cycles. In the same manner one can implement a RHF for this problem.

The computed ensemble means $\bar{q}(t)$ and $\bar{v}(t)$ in comparison to the reference solution can be found in Fig. 11 for the continuous EGMF (using $L = 1$ and $L = 2$ mixture components as appropriate) and the ensemble Kalman-Bucy filter (continuous EGMF with $L = 1$). The root mean square (RMS) error with respect to the true reference solution is 2.3331 for the ensemble Kalman-Bucy filter and 1.9148 for the EGMF, which amounts to a relative improvement of about 20%. The EGMF uses a bi-Gaussian distribution in about 25% of the assimilation steps. For comparison we show the results from the RHF in Fig. 12 for $M = 50$ particles. To interpret the behavior of the RHF one needs to look at the potential energy function $V(q)$ (see Fig. 4). The RHF assimilation scheme apparently pushes the solutions occasionally into the flat side regions of the potential energy curve resulting in a relatively large RMS error of 3.9375. Qualitatively similar results are obtained for the RHF with $M = 800$ particles. Recall that we do observe velocities and not positions in this example and that the RHF uses the ensemble generated covariance matrix $\mathbf{P}$ to linearly regress filter increments onto state space.

5.4 Lorenz-63 model

The three variable model

$$\dot{x} = 10(y - x), \quad \dot{y} = x(28 - z) - y, \quad \dot{z} = xy - \frac{8}{3}z \qquad (68)$$

of Lorenz (1963) is used as a final test for the EGMF method. Only the x variable is observed every 0.20 time units with an observational error drawn from a normal distribution with mean zero and variance eight. The model time-step is $\Delta t = 0.01$. A total of 101000 assimilation steps is performed for each experiment with only the last 100000 steps being used for the computation of RMS errors. We have implemented an ensemble square root filter, a RHF, and an EGMF using formulation (45)-(46) with $\mathbf{B} = c\mathbf{P}$, $\mathbf{P}$ the empirical covariance matrix of the ensemble. The parameter c is chosen from the interval $c \in [0.4, 1.0]$. The number of ensemble members
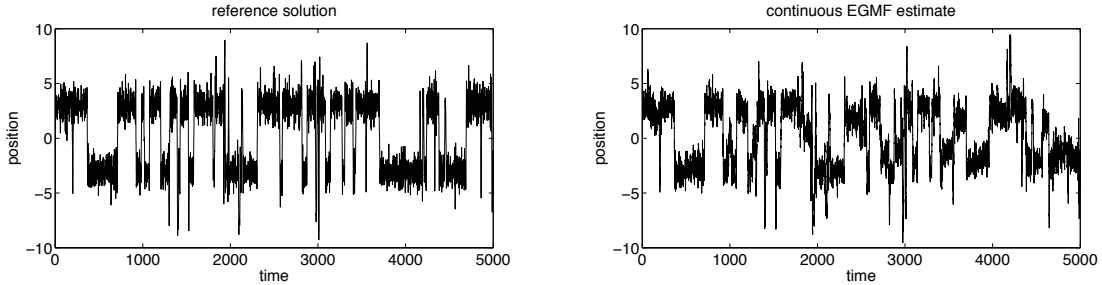
Figure 11: The reference solution $q_r(t)$ (left panel) and the estimated (ensemble mean) solution from the continuous EGMF (right panel) over the first quarter of the simulation interval $t \in [0, 20000]$. The estimated solution mostly follows the reference solution with the exception of a number of missed transitions.

is set to $M = 25$ and no covariance localization is applied. The internal assimilation step-size is $\Delta s = 1/4$ and the l_∞-norm of $u_B(\mathbf{x}, s)$ is limited to a value of $u_{\text{cut}} = 0.125/\Delta s$. We have computed the RMS errors for ensemble inflation factors between 1.0 and 1.3 and only report the optimal results in Fig. 13 as a function of c for the EGMF. The overall smallest RMS error is achieved for $c = 0.6$ with a value of 4.1114. The associated RMS errors for the ensemble square root filter are 4.4813 and 4.6596 for the RHF, respectively. An increase in the number of ensemble members to $M = 100$ leads to a reduction in the RMS error for the RHF to 4.3276 while the ensemble square root filter yields its optimal performance for $M = 50$ with a RMS error of 4.3785. Both values are significantly larger than the optimal RMS error for the EGMF with $M = 25$. A better performance is observed for the EnKF with perturbed observations and $M = 25$ for which we obtain 4.1775 as the smallest RMS error which, in fact, is close to the performance of the EGMF with $c = 1$.

6   Summary

We have extended the popular EnKF to statistical models provided by Gaussian mixtures. The EGMF is derived using a continuous reformulation of the Bayesian analysis step and consists of a combination of EnKF steps for each mixture component and an exchange term. The exchange term is determined for each measurement by a scalar elliptic PDE, which can be solved analytically. We have demonstrated by means of two numerical examples that the EGMF performs well when bimodal PDFs arise naturally from the model dynamics. The EGMF provides a valuable and easy to implement alternative to sequential Monte Carlo methods and other nonlinear filter algorithms. In this paper, we have used the standard EM algorithm to assign Gaussian mixture models to ensemble predictions. More refined methods such as those discussed by Fraley and Raftery (2007) will be considered in future work in order to provide a robust and accurate clustering of ensemble predictions. Alternatively, one can implement the EGMF with a Gaussian kernel density estimator. In this case, the empirical covariance matrix of the ensemble can be used as a base for kernel bandwidth selection (Wand and Jones, 1995). With this choice, the EGMF becomes closely related to the RHF of Anderson (2010). The feasibility of our approach has been demonstrated for the Lorenz-63 model. Further work is required to assess the merits of Gaussian kernel density estimators in comparison to EnKF
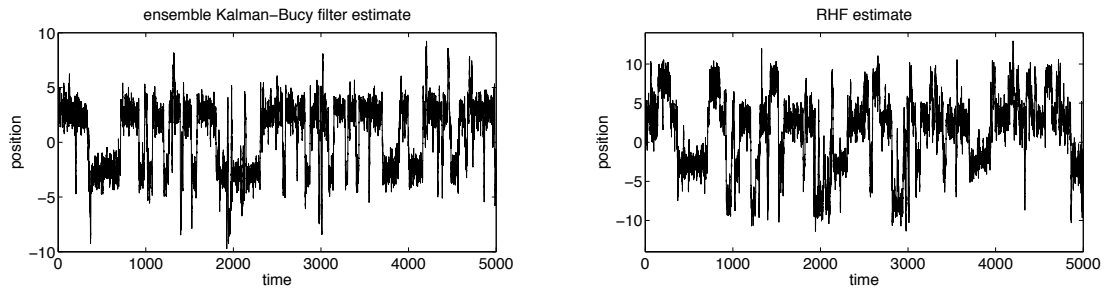
Figure 12: The estimated (ensemble mean) solution from the ensemble Kalman-Bucy filter over a quarter of the simulation interval is displayed in the left panel. The results look qualitatively similar to the results from the EGMF filter. However, in terms of RMS errors, the EGMF outperforms the ensemble Kalman-Bucy filter by about 20% over the complete simulation interval. The right panel displays the estimated (ensemble mean) solution from the RHF (left panel). The reader should note the enlarged range of the vertical axis. At several instances the filtered solution strongly deviates from the reference solution. To interpret this behavior we need to have a closer look at the potential energy $V(q)$ (compare Fig. 4). Apparently the RHF interprets the data as corresponding to solutions with positions in the flat side regions of the potential energy function.

and RHF implementations for high dimensional systems. Encouraging results have also been reported by Stordal et al. (2011) for their related adaptive Gaussian mixture filter applied to the Lorenz-96 model (Lorenz, 1996; Lorenz and Emanuel, 1998).

References

J. Amezcua, E. Kalnay, K. Ide, and S. Reich. Using the Kalman-Bucy filter in an ensemble framework. *Mon. Weath. Rev.*, submitted, 2011.

J.L. Anderson. A non-Gaussian ensemble filter update for data assimilation. *Monthly Weather Review*, 138:4186–4198, 2010.

J.L. Anderson and S.L. Anderson. A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.*, 127:2741–2758, 1999.

A. Bain and D. Crisan. *Fundamentals of stochastic filtering*, volume 60 of *Stochastic modelling and applied probability*. Springer-Verlag, New-York, 2009.

K. Bergemann and S. Reich. A localization technique for ensemble Kalman filters. *Q. J. R. Meteorological Soc.*, 136:701–707, 2010a.

K. Bergemann and S. Reich. A mollified ensemble Kalman filter. *Q. J. R. Meteorological Soc.*, 136:1636–1643, 2010b.

K. Bergemann and S. Reich. An ensemble Kalman-Bucy filter for continuous data assimilation. *Meteorolog. Zeitschrift*, submitted, 2011.
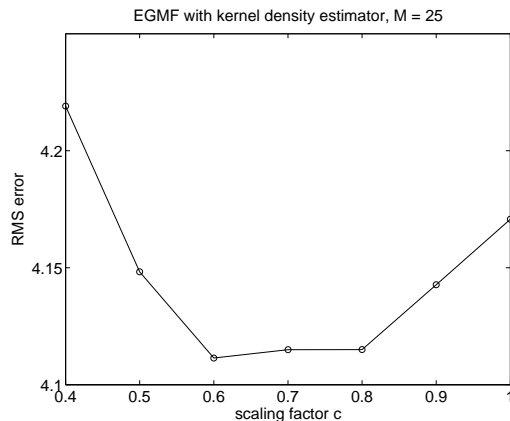
EGMF with kernel density estimator, M = 25

Figure 13: RMS errors from the EGMF for a range of values of the scaling parameter c and ensemble size $M = 25$. A series of experiments has been conducted with the Lorenz-63 model for each fixed scaling parameter $c \in [0.4, 1.0]$ and a range of ensemble inflation factors. We only display the optimal results. The overall smallest RMS error is achieved for $c = 0.6$ with a value of 4.1114. The corresponding optimal RMS error for an ensemble square root filter is 4.4813 and 4.6596 for the RHF, respectively. The EnKF with perturbed observations leads to a RMS error of 4.1775 which is only slightly worse than the performance of the EGMF with $c = 1$.

D. Crisan and J. Xiong. Approximate McKean-Vlasov representation for a class of SPDEs. *Stochastics*, 82:53–68, 2010.

A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J. Royal Statistical Soc.*, 39B:1–38, 1977.

G. Evensen. *Data assimilation. The ensemble Kalman filter.* Springer-Verlag, New York, 2006.

Ch. Fraley and A.E. Raftery. Bayesian regularization for normal mixture estimation and model-based clustering. *Journal of Classification*, 24:155–181, 2007.

M. Frei and H.R. Künsch. Mixture ensemble Kalman filters. *Computational Statistics and Data Analysis*, in press, 2011.

C.W. Gardiner. *Handbook on stochastic methods.* Springer-Verlag, 3rd edition, 2004.

Th.M. Hamill, J.S. Whitaker, and Ch. Snyder. Distance-dependent filtering of background covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.*, 129:2776–2790, 2001.

P.L. Houtekamer and H.L. Mitchell. A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.*, 129:123–136, 2001.

E.N. Lorenz. Deterministic non-periodic flows. *J. Atmos. Sci.*, 20:130–141, 1963.

E.N. Lorenz. Predictibility: A problem partly solved. In *Proc. Seminar on Predictibility*, volume 1, pages 1–18, ECMWF, Reading, Berkshire, UK, 1996.

E.N. Lorenz and K.E. Emanuel. Optimal sites for suplementary weather observations: Simulations with a small model. *J. Atmos. Sci.*, 55:399–414, 1998.

S. Reich. A dynamical systems framework for intermittent data assimilation. *BIT*, 51:235–249, 2011.

J.L. Schaefer. *Analysis of incomplete multivariate data by simulation.* Chapmann and Hall, London, 1997.

K.W. Smith. Cluster ensemble Kalman filter. *Tellus*, 59A:749–757, 2007.

A.S. Stordal, H.A. Karlsen, G. Nævdal, H.J. Skaug, and B. Vallés. Bridging the ensemble Kalman filter and particle filters: the adaptive Gaussian mixture filter. *Comput. Geosci.*, 15: 293–305, 2011.

M.K. Tippett, J.L. Anderson, G.H. Bishop, T.M. Hamill, and J.S. Whitaker. Ensemble square root filters. *Mon. Wea. Rev.*, 131:1485–1490, 2003.

C. Villani. *Topics in Optimal Transportation.* American Mathematical Society, Providence, Rhode Island, NY, 2003.

M.P. Wand and M.C. Jones. *Kernel smoothing.* Chapmann and Hall, London, 1995.