Marginalized Particle Filtering Framework for Tuning of Ensemble Filters

VÁCLAV ŠMÍDL * AND RADEK HOFMAN[†]

Institute of Information Theory and Automation, Prague, Czech Republic

December 1, 2010

Corresponding author address: Václav Šmídl, Institute of Information Theory and Automation, Pod Vodárenskou věží 4, Praha 8, 196 00 Czech Republic. His research was supported by grant GACR 102/08/P250. E-mail: smidl@utia.cas.cz

[†]Research supported by grant MSMT 1M0572. E-mail: hofman@utia.cas.cz

Abstract

Marginalized particle filtering (MPF), also known as Rao-Blackwellized particle filtering, has been recently developed as a hybrid method combining analytical filters with particle filters. In this paper, we investigate the prospects of this approach in enviromental modelling where the key concerns are nonlinearity, high-dimensionality, and computational cost. In our formulation, exact marginalization in the MPF is replaced by approximate marginalization yielding a framework for creation of new hybrid filters. In particular, we propose to use the MPF framework for on-line tuning of nuisance parameters of ensemble filters. Conditional independence based simplification of the MPF algorithm is proposed for computational reasons and its close relation to previously published methods is discussed. Strength of the framework is demonstrated on the joint estimation of the inflation factor, the measurement error variance and the length-scale parameter of covariance localization. It is shown that accurate estimation can be achieved with a moderate number of particles. Moreover, this result was achieved with naively chosen proposal densities leaving space for further improvements.

1. Introduction

We are concerned with the task of data assimilation using Bayesian approach. Bayesian interpretation of data assimilation is well established in the community, e.g. by Anderson et al. (1999). However, application of standard Bayesian methods in this domain is not straightforward due to its specific properties, most notably high-dimensionality of the state

1

variable. Successful application of Bayesian methods in this domain requires their adaptation. This was the case of two prominent Bayesian methods: the Kalman filter and the particle filter. Practical constraints of the domain motivated approximation of the formally correct Kalman filter into the form of ensemble Kalman filter, (Evensen 1994), which has evolved beyond the Kalman filter since then. Direct application of the original particle filter in the considered domain is problematic, Snyder et al. (2008), however, the ideas of particle filtering are incorporated in many new algorithms. Each of these filters has features attractive for the domain as well as it has limitations. Specifically, the KF (and EnKF) is limited by approximation of the posterior by a Gaussian density which is required for computational tractability in high dimensions. On the other hand, the particle filter is capable of approximating an arbitrary density via empirical density at the price of significantly higher computational cost, which is prohibitive in the considered dimensions. A combination of these approaches is an obvious goal. Indeed, various hybrid filters have been proposed to overcome the above mentioned difficulties, e.g. (Pham 2001; Nakano et al. 2007; Hoteit et al. 2008; Van Leeuwen 2009; Bocquet et al. 2010), etc. A common assumption of these methods is that the state variable is treated as a homogeneous variable, i.e., that all entries in the state vector are treated identically.

In this paper we investigate applicability of the marginalized particle filter (MPF), (Schön et al. 2005)—which is also known as Rao-Blackwellized PF, (Doucet et al. 2000)—in the field of environmental modeling. This filter arises as a combination of the conditional Kalman filter with the particle filter. The key difference from the above mentioned hybrid filters is that it partitions the state variable into two parts, and approximates each part by a different type of distribution. Partitioning of the space has been discussed by Van Leeuwen

(2009) in the context of hierarchical Bayesian models, where both partitions were estimated by the particle filter. The partitioning may seem as an arbitrary step in the considered domain, however, we will show that this approach may be attractive for addressing some specific scenarios such as on-line tuning of already existing filters. Particle filtering ideas have already been used for estimation of parameters by (Vossepoel and Van Leeuwen 2007), however, the resulting method was based on weighting of the ensemble members. The MPF approach is based on running several ensemble filters in parallel and it is thus closer to the hierarchical ensemble filter (Anderson 2007b).

The paper is organized as follows. In Section 2, we review the basic algorithms of Bayesian filtering, including the ensemble filter, the particle filter, and the marginalized particle filter. In Section 3, we apply the MPF approach to tuning of ensemble filters and discuss its potential extensions. Performance of the proposed methods is studied in Section 4 via simulation of the 40-dimensional Lorenz model.

2. Bayesian Filtering and its Approximation

The task of data assimilation can be interpreted as a problem of inference of a discrete-time stochastic process:

$$\mathbf{y}_t \sim p\left(\mathbf{y}_t | \mathbf{x}_t\right), \quad \mathbf{x}_t \sim p\left(\mathbf{x}_t | \mathbf{x}_{t-1}\right). \tag{1}$$

Here, $\mathbf{x}_t$ is a vector known as the state variable and $\mathbf{y}_t$ are the observations, $p(\cdot|\cdot)$ denotes the conditional probability density of the variable. By *Bayesian Filtering* we mean the recursive

evaluation of the filtering distribution, $p\left(\mathbf{x}_t|\mathbf{y}_{1:t}\right)$, using Bayes' rule (Peterka 1981):

$$p\left(\mathbf{x}_t|\mathbf{y}_{1:t}\right) \;=\; \frac{p\left(\mathbf{y}_t|\mathbf{x}_t\right)p\left(\mathbf{x}_t|\mathbf{y}_{1:t-1}\right)}{p(\mathbf{y}_t|\mathbf{y}_{1:t-1})}, \tag{2}$$

$$p\left(\mathbf{x}_t|\mathbf{y}_{1:t-1}\right) \;=\; \int p\left(\mathbf{x}_t|\mathbf{x}_{t-1}\right)p\left(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1}\right)\,d\mathbf{x}_{t-1}, \tag{3}$$

where $p\left(\mathbf{x}_1|\mathbf{y}_0\right)$ is the prior distribution, and $\mathbf{y}_{1:t} = [\mathbf{y}_1,\ldots,\mathbf{y}_t]$ denotes the set of all obser-vations. The integration in (3), and elsewhere in this paper, is over the whole support of the involved probability density functions.

Equations (2)–(3) are analytically tractable only for a limited set of models. The most notable example of an analytically tractable model is linear Gaussian (1) for which (2)–(3) are equivalent to the Kalman filter. For other models, (2)–(3) need to be evaluated approximately. Selected approximations are now briefly described.

a. Ensemble Kalman filtering (EnKF)

The ensemble Kalman filter arises as an approximation of Bayesian filtering for Gaussian stochastic system,

$$p\left(\mathbf{y}_t|\mathbf{x}_t\right) = \mathcal{N}\left(\mathbf{H}_t\mathbf{x}_t, \mathbf{R}_t\right), \quad p\left(\mathbf{x}_t|\mathbf{x}_{t-1}\right) = \mathcal{N}\left(\mathcal{M}(\mathbf{x}_{t-1}), \mathbf{Q}_t\right), \tag{4}$$

where $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is a Gaussian distribution with mean value $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. $\mathbf{H}_t$ is a linear observation operator that transforms vectors from the space of model into the space of observations; generally non-linear operator $\mathcal{M}(\cdot)$ represents a dynamics of the system. Matrices $\mathbf{R}_t$ and $\mathbf{Q}_t$ are error covariance matrices of observations and model, respectively. Optimal inference of model (4) for linear operator $\mathcal{M}(\cdot)$ yields the Kalman filter which implies potentially high computational cost for high-dimensional state vector. The Ensemble

Kalman filter (EnKF) is an approximation of the Bayesian filtering for model (4) with predictive and posterior densities (2)–(3) projected to Gaussian densities:

$$p\left(\mathbf{x}_t|\mathbf{y}_{1:t}\right) = \mathcal{N}(\overline{\mathbf{x}}_t^a, \mathbf{P}_t^a),$$

$$p\left(\mathbf{x}_t|\mathbf{y}_{1:t-1}\right) = \mathcal{N}(\overline{\mathbf{x}}_t^b, \mathbf{P}_t^b).$$

Here, vectors $\overline{\mathbf{x}}^b$ and $\overline{\mathbf{x}}^a$ are known as background (predictive) model estimate and analysis (posterior) estimate, in standard notation (Ide et al. 1997).

The predictive and posterior statistics are evaluated via a set of state vectors $\mathbf{x}_{i;t} \in \{\mathbf{x}_{i;t}\}_{i=1}^M$ known as the *ensemble members* as follows:

$$\overline{\mathbf{x}}_t^b \equiv \frac{1}{M} \sum_{i=1}^M \mathbf{x}_{i;t}^b, \tag{5}$$

$$\mathbf{P}_t^b \equiv \frac{1}{M-1} \sum_{i=1}^M \tilde{\mathbf{x}}_{i;t}^b \left(\tilde{\mathbf{x}}_{i;t}^b\right)^{\mathrm{T}}. \tag{6}$$

Here, $\{\tilde{\mathbf{x}}_{i;t}^b = \mathbf{x}_{i;t}^b - \overline{\mathbf{x}}_t^b\}_{i=1}^M$ are deviations of the ensemble members from the mean, $\overline{\mathbf{x}}^b$. The standard EnKF filter (Burgers et al. 1998) requires to draw random samples to perturb the observations to compute $\overline{\mathbf{x}}_t^a, \mathbf{P}_t^a$. This need for random sampling is eliminated by deterministic variants such as the ensemble adjustment Kalman filter (Anderson 2001) and the ensemble square root filter (EnSRF) (Whitaker and Hamill 2002; Tippett et al. 2003).

In the EnSRF, the posterior mean $\overline{\mathbf{x}}_t^a$ and the deviations $\tilde{\mathbf{x}}_{i;t}^a$ are updated separately, as follows:

$$\overline{\mathbf{x}}_t^a = \overline{\mathbf{x}}_t^b + \mathbf{K}_t \left(\mathbf{y}_t - \mathbf{H}_t \overline{\mathbf{x}}_t^b\right), \tag{7}$$

$$\tilde{\mathbf{x}}_{i;t}^a = \tilde{\mathbf{x}}_{i;t}^b - \tilde{\mathbf{K}}_t \mathbf{H}_t \tilde{\mathbf{x}}_{i;t}^b, \tag{8}$$

where $\mathbf{H}_t$ is an observation operator (which can be potentially non-linear), $\mathbf{K}_t$ is the Kalman gain matrix for updating of the ensemble mean (7)

$$\mathbf{K}_t = \mathbf{P}_t^b \mathbf{H}_t^{\mathrm{T}} (\mathbf{H}_t \mathbf{P}_t^b \mathbf{H}_t^{\mathrm{T}} + \mathbf{R}_t)^{-1}, \tag{9}$$

and $\tilde{\mathbf{K}}_t$ is the modified Kalman gain for updating departures from the mean (8)

$$\tilde{\mathbf{K}}_t = \mathbf{P}_t^b \mathbf{H}_t^{T} \left[\left(\sqrt{\mathbf{H}_t \mathbf{P}_t^b \mathbf{H}_t^{\mathrm{T}} + \mathbf{R}_t} \right)^{-1} \right]^{\mathrm{T}} \left[\sqrt{\mathbf{H}_t \mathbf{P}_t^b \mathbf{H}_t^{\mathrm{T}} + \mathbf{R}_t} + \sqrt{\mathbf{R}_t} \right]^{-1}. \tag{10}$$

For a diagonal covariance matrix of observations $\mathbf{R}_t$, (10) may be processed one element of $\mathbf{y}_t$ at a time without any significant increase of computational effort.

In all the equations above, the terms with $\mathbf{P}_t^b$ are replaced by empirical covariance in the sense of (6). The same replacement is used for predictive density of the observations,

$$p(\mathbf{y}_t | \mathbf{y}_{1:t-1}) = \mathcal{N}(\mathbf{H}_t \bar{x}_t^b, \mathbf{S}_t), \tag{11}$$

$$\mathbf{S}_t = \mathbf{H}_t \mathbf{P}_t^b \mathbf{H}_t^{\mathrm{T}} + \mathbf{R}_t, \tag{12}$$

which corresponds to the standard predictive density of the Kalman filter (Peterka 1981). This quantity is often called *marginal likelihood* (marginalization is with respect to x_t) and plays an important role in statistical model selection (Jeffreys 1961).

1) SMALL SAMPLE ISSUES

Due to approximation (6), the maximum rank of matrix $\mathbf{P}_t^b$ is M. The full rank covariance matrix can thus be obtained with the number of ensemble members equal to or greater than the dimensionality of the state variable. However, the number of ensemble members is typically much smaller yielding inaccurate approximation of the true background error covariance matrix. Basic techniques for improvement are covariance inflation and distance-

dependent covariance localization.

Ensemble inflation is a method that artificially increases the background error variance, $\mathbf{P}_t^b$, (Anderson and Anderson 1999). The forecast ensemble $\mathbf{x}_t = [\mathbf{x}_{i;t}^b, \ldots, \mathbf{x}_{M;t}^b]$ is replaced by $\mathbf{x}_t^* = [\mathbf{x}_{i;t}^{*b}, \ldots, \mathbf{x}_{M;t}^{*b}]$,

$$\mathbf{x}_{i;t}^{*b} = \sqrt{\Delta}\left(\mathbf{x}_{i;t}^b - \bar{\mathbf{x}}_t^b\right) + \bar{\mathbf{x}}_t^b, \; i = 1, \ldots, M, \tag{13}$$

where Δ is the inflation factor slightly greater than 1. (13) implies that the mean value of the ensemble remains the same but its variance is increased.

Covariance localization is a method suppressing spurious covariances representing false relations between distant states (Houtekamer and Mitchell 1998). The method artificially suppresses covariances between distant entries of the state vector and thus improves the conditioning of matrices $\mathbf{P}_t^b\mathbf{H}^{\mathrm{T}}$ and $\mathbf{H}\mathbf{P}_t^b\mathbf{H}^{\mathrm{T}}$. Localization of the forecast error covariance matrix is performed as a Schur product (element-by-element product) of a localization matrix with the model forecast error covariance matrix $\mathbf{P}_t^b$ (6). Localization matrices are constructed by using various correlation functions. The correlation function is equal to 1 at the observation location and gradually decreases to zero at a finite distance from the observation. The rate of decrease of the correlation function with distance is governed by the so called length-scale parameter, l, (Gaspari and Cohn 1999).

b. Particle Filtering (PF)

Particle filtering (PF), (Gordon et al. 1993; Doucet et al. 2001), refers to a range of techniques for generating an empirical approximation of $p\left(\mathbf{x}_{1:t}|\mathbf{y}_{1:t}\right)$, where $\mathbf{x}_{1:t} = [\mathbf{x}_1, \ldots, \mathbf{x}_t]$

is the state trajectory:

$$p\left(\mathbf{x}_{1:t}|\mathbf{y}_{1:t}\right) \approx \frac{1}{n} \sum_{i=1}^{n} \delta\left(\mathbf{x}_{1:t} - \mathbf{x}_{1:t}^{(i)}\right), \tag{14}$$

where $\left\{\mathbf{x}_{1:t}^{(i)}\right\}_{i=1}^{n}$ are i.i.d. samples from the posterior and $\delta(\cdot)$ denotes the Dirac δ-function.

Therefore, this approach is feasible only if we can sample from the exact posterior, $p\left(\mathbf{x}_{1:t}|\mathbf{y}_{1:t}\right)$.

If this is not the case, we can draw samples from a chosen proposal distribution (importance

function), $q\left(\mathbf{x}_{1:t}|\mathbf{y}_{1:t}\right)$, as follows:

$$p\left(\mathbf{x}_{1:t}|\mathbf{y}_{1:t}\right) = \frac{p\left(\mathbf{x}_{1:t}|\mathbf{y}_{1:t}\right)}{q\left(\mathbf{x}_{1:t}|\mathbf{y}_{1:t}\right)} q\left(\mathbf{x}_{1:t}|\mathbf{y}_{1:t}\right) \approx \frac{p\left(\mathbf{x}_{1:t}|\mathbf{y}_{1:t}\right)}{q\left(\mathbf{x}_{1:t}|\mathbf{y}_{1:t}\right)} \frac{1}{n} \sum_{i=1}^{n} \delta\left(\mathbf{x}_{1:t} - \mathbf{x}_{1:t}^{(i)}\right). \tag{15}$$

Approximation (15) can be written in the form of a *weighted* empirical distribution, as

follows:

$$p\left(\mathbf{x}_{1:t}|\mathbf{y}_{1:t}\right) \approx \sum_{i=1}^{n} w_t^{(i)} \delta\left(\mathbf{x}_{1:t} - \mathbf{x}_{1:t}^{(i)}\right), \tag{16}$$

$$w_t^{(i)} \propto \frac{p\left(\mathbf{x}_{1:t}^{(i)}|\mathbf{y}_{1:t}\right)}{q\left(\mathbf{x}_{1:t}^{(i)}|\mathbf{y}_{1:t}\right)}. \tag{17}$$

Under this *importance sampling* procedure, the true posterior distribution needs to be eval-

uated point-wise only, since (16) can be normalized trivially via a constant $c = \sum_{i=1}^{n} w_t^{(i)}$.

$\propto$ is used to denote equality up to the normalization constant.

The challenge for on-line algorithms is to achieve recursive generation of the samples and

evaluation of the importance weights. Using (1) and standard Bayesian calculus, (17) may

be written in the following recursive form:

$$w_t^{(i)} \propto \frac{p\left(\mathbf{y}_t|\mathbf{x}_t^{(i)}\right) p\left(\mathbf{x}_t^{(i)}|\mathbf{x}_{t-1}^{(i)}\right)}{q\left(\mathbf{x}_t^{(i)}|\mathbf{x}_{1:t-1}^{(i)}, \mathbf{y}_{1:t}\right)} w_{t-1}^{(i)}. \tag{18}$$

Now, $\mathbf{x}_t^{(i)}$ are drawn from the denominator of (18), which can be chosen as $p\left(\mathbf{x}_t|\mathbf{x}_{t-1}\right)$. A

successful application of the particle filter requires to resample the particles to preserve in-

formative weights $w_t^{(i)}$. While re-sampling can be done in each step, it is computationally advantageous to perform it only when the efficient number of particles $n_{eff} = 1/\sum_{i=1}^{n} \left(w_t^{(i)}\right)^2$ falls below a chosen threshold. See (Doucet et al. 2001) for more details and proofs of convergence of the filter to the true posterior density.

c. *Marginalized Particle Filtering (MPF)*

The main advantage of importance sampling is its generality. However, it may be computationally prohibitive to draw samples from the possibly high-dimensional state space of $\mathbf{x}_t$, (Snyder et al. 2008). These problems can be overcome in cases where the structure of the model (1) allows analytical marginalization over a subset, $\mathbf{x}_{c,t}$, of the full state vector $\mathbf{x}_t^{\mathrm{T}} = [\mathbf{x}_{c,t}^{\mathrm{T}}, \mathbf{x}_{p,t}^{\mathrm{T}}]$ (Doucet et al. 2001; Schön et al. 2005). Therefore, we consider the factorization

$$p\left(\mathbf{x}_{1:t}|\mathbf{y}_{1:t}\right) = p\left(\mathbf{x}_{c,1:t}|\mathbf{x}_{p,1:t}, \mathbf{y}_{1:t}\right) p\left(\mathbf{x}_{p,1:t}|\mathbf{y}_{1:t}\right), \tag{19}$$

where $p\left(\mathbf{x}_{c,1:t}|\mathbf{x}_{p,1:t}, \mathbf{y}_{1:t}\right)$ is analytically tractable, while $p\left(\mathbf{x}_{p,1:t}|\mathbf{y}_{1:t}\right)$ is not. We replace the latter by a weighted empirical distribution, in analogy to (15), yielding

$$p\left(\mathbf{x}_{1:t}|\mathbf{y}_{1:t}\right) \approx \sum_{i=1}^{n} w_t^{(i)} p\left(\mathbf{x}_{c,1:t}|\mathbf{x}_{p,1:t}^{(i)}, \mathbf{y}_{1:t}\right) \delta\left(\mathbf{x}_{p,1:t} - \mathbf{x}_{p,1:t}^{(i)}\right), \tag{20}$$

$$w_t^{(i)} \propto \frac{p\left(\mathbf{x}_{p,1:t}^{(i)}|\mathbf{y}_{1:t}\right)}{q\left(\mathbf{x}_{p,1:t}^{(i)}|\mathbf{y}_{1:t}\right)}. \tag{21}$$

Note that now we only have to sample from the space of $\mathbf{x}_{p,t}$. Recursive evaluation is achieved by application of the Bayes rule

$$p\left(\mathbf{x}_{1:t}|\mathbf{y}_{1:t}\right) \propto \int p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{x}_{t-1})p\left(\mathbf{x}_{1:t-1}|\mathbf{y}_{1:t-1}\right) dx_{t-1}, \tag{22}$$

9

and substitution of (20) in place of $p(\mathbf{x}_{1:t}|\mathbf{y}_{1:t})$ and $p(\mathbf{x}_{1:t-1}|\mathbf{y}_{1:t-1})$. Comparing elements in the summations on both sides of equation (22), we obtain:

$$w_t^{(i)} \quad \propto \quad \frac{p\left(\mathbf{y}_t, \mathbf{x}_{p,t}^{(i)}|\mathbf{x}_{p,t-1}^{(i)}, \mathbf{y}_{1:t-1}\right)}{q\left(\mathbf{x}_{p,t}^{(i)}|\mathbf{x}_{p,1:t-1}^{(i)}, \mathbf{y}_{1:t}\right)} w_{t-1}^{(i)}, \tag{23}$$

$$p(\mathbf{y}_t, \mathbf{x}_{p,t}^{(i)}|\mathbf{x}_{p,t-1}^{(i)}, \mathbf{y}_{1:t-1}) \quad = \quad \int p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{c,t-1}|\mathbf{x}_{p,1:t-1}, \mathbf{y}_{1:t-1})d\mathbf{x}_{c,t}d\mathbf{x}_{c,t-1}. \tag{24}$$

Analytical tractability of integrations in (24) allows to use the Rao-Blackwell theorem to show that the variance of weights (23) is lower than the variance of weights (18) for the same problem (Doucet et al. 2000).

d. Marginalized Particle Filtering Framework

The requirement of analytical tractability of integrations in (24) is always fulfilled when (1) contains a linear-Gaussian part, (Schön et al. 2005), giving rise to the MPF with the Kalman filter. Tractable solution also exists for discrete-variable models (Thrun et al. 2005) and models based on conjugate statistics (Saha et al. 2010). However, the range of models amendable to this approach is still rather small and does not contain any models common in atmospheric modeling.

We propose to relax the requirement of exact marginalization in (23) and replace it by an approximation. Specifically, we note that given numerical values of $\mathbf{x}_{p,t}^{(i)}, \mathbf{x}_{p,t-1}^{(i)}$, equation (24) is equivalent to the normalizing constant of a Bayesian filter (2). Hence, any Bayesian filter that is capable of evaluating its normalizing constant can be used to approximate (24). What results is an algorithm equivalent to the marginalized particle filtering where the analytical Kalman filters are replaced by approximate conditional filters. Specifically, the following

10

Algorithm 1 General MPF framework

---

   i. Generate initial particles $\left\{\mathbf{x}_{p,0}^{(i)}\right\}_{i=1}^{n}$ and set initial statistics of all associated conditional filters.

  ii. For each new data record $\mathbf{y}_t$ do:

     (a) Sample new value of particles $\left\{\mathbf{x}_{p,t}^{(i)}\right\}_{i=1}^{n}$, and update statistics of all associated conditional filters via (25).

     (b) Compute weights (23) of all particles and their associated conditional filters.

     (c) If the number of efficient particles, n_{eff}, is lower than the chosen threshold, re-sample the particles.

---

filters interact via Algorithm 1:

Conditional filter on variable $\mathbf{x}_{c,t}$, treating $\mathbf{x}_{p,t}$ as an observation, i.e.:

$$p(\mathbf{x}_{c,t}|\mathbf{x}_{p,1:t},\mathbf{y}_{1:t}) = \frac{p(\mathbf{y}_t,\mathbf{x}_{p,t}|\mathbf{x}_{c,t},\mathbf{x}_{p,1:t})p(\mathbf{x}_{c,t}|\mathbf{x}_{p,1:t-1},\mathbf{y}_{1,t-1})}{p(\mathbf{y}_t,\mathbf{x}_{p,t}|\mathbf{x}_{p,1:t-1},\mathbf{y}_{1,t-1})}. \tag{25}$$

Particle filter on variable $\mathbf{x}_p$, that handles sampling from the proposal function $q(\mathbf{x}_{p,t}|\mathbf{x}_{p,t-1},\mathbf{y}_{1:t})$ and re-sampling. Each particle is attached to one conditional filter.

In this general form, the algorithm is rather unspecific. This is due to the fact that arbitrary conditional filters can be combined with arbitrary particle filtering approaches. Therefore, we consider Algorithm 1 to represent a framework for designing specific filtering variants. The word framework is used to distinguish this approach from the analytical MPF.

11

3. MPF Framework for Tuning of Ensemble Filters

The key property of the MPF is partitioning of the state variable into two parts. In the original exact formulation, the choice of partitioning is fully determined by tractability of the Bayes rule (2). Finding a partitioning in the context of environmental modeling where the state variables typically obey the same equations is harder. However, the relaxed formalization of Section 2d allows to interpret $\mathbf{x}_{p,t}$ not as a partition of the full state but rather as an augmentation of the original state (now denoted by $x_{c,t}$) by nuisance parameters. What results is a framework for on-line tuning of existing filters.

The general algorithm of tuning is described in Algorithm 1, specific variants arise for the following choices:

Choice 1: choose a preferred variant of the conditional filter (e.g. a variant of the ensemble filter) estimating $\mathbf{x}_{c,t}$,

Choice 2: choose tuning parameters of interest, $\mathbf{x}_{p,t}$, use them to augment the original state $\mathbf{x}_{c,t}$ via the chosen model of their evolution, $p(\mathbf{x}_{p,t}|\cdot)$,

Choice 3: choose a proposal density $q(\mathbf{x}_{p,t}|\cdot)$, e.g. the evolution model, $q(\mathbf{x}_{p,t}|\cdot) \equiv p(\mathbf{x}_{p,t}|\cdot)$.

Different choices in each of the points above will lead to different properties of the resulting filter. The number of possible combinations of these choices is enormous, and finding guidelines for the best option in a given application context is a task for further research. In some applications, a physically motivated evolution model of $\mathbf{x}_{p,t}$ may be found, while heuristic or expert-chosen models may be more appropriate in others.

a. Estimation of inflation factor, observation error variance, and length scale parameter

The techniques for compensating small ensemble issues, Section 2a1), have tuning parameters that typically need to be set-up experimentally. Substantial effort has been put into on-line estimation of the inflation factor alone, (Anderson 2007a), or in tandem with the observation error (Li et al. 2009). In this Section, we approach the same problem using the MPF framework with the following specific choices.

Choice 1: We have chosen the EnSRF with multiplicative inflation (13) as our conditional filter.

Choice 2: The unknown tuning parameters are: (i) the time-variant inflation factor Δ_t, (ii) the time-variant variance of the observation error, r_t, for all observations, i.e. $\mathbf{R}_t = r_t \mathbf{I}$, where $\mathbf{I}$ is the identity matrix, and (iii) the time-varying length scale parameter l_t of the covariance localization function. Evolution of these parameters is modeled by truncated Gaussian random walks,

$$p(\Delta_t|\Delta_{t-1}) = t\mathcal{N}(\Delta_{t-1}, \sigma_\Delta^2, [1, \infty]),$$

$$p(r_t|r_{t-1}) = t\mathcal{N}(r_{t-1}, \sigma_r^2, [0, \infty]), \tag{26}$$

$$p(l_t|l_{t-1}) = t\mathcal{N}(l_{t-1}, \sigma_l^2, [0, \infty]).$$

where scalar parameters σ_Δ, σ_r, and σ_l denote the spread of the random walks, respectively. Non-negativity of all considered parameters motivates truncation of support of the random walks. The augmentation of the state vector is then $\mathbf{x}_{p,t} = [\Delta_t, r_t, l_t]^{\mathrm{T}}$.

Choice 3: The proposal density is chosen as $p(\mathbf{x}_{p,t}|\mathbf{x}_{p,t-1}) = p(\Delta_t|\Delta_{t-1})p(r_t|r_{t-1})p(l_t|l_{t-1})$,

under which (23) reduces to

$$w_t^{(i)} = w_{t-1}^{(i)} p(\mathbf{y}_t | \mathbf{y}_{1:t-1}, \mathbf{x}_{p,t}^{(i)}), \tag{27}$$

where $p(\mathbf{y}_t | \mathbf{y}_{1:t-1}, \mathbf{x}_{p,t}^{(i)})$ is defined by (11) now explicitly conditioned on the unknown parameters

$$p(\mathbf{y}_t | \mathbf{y}_{1:t-1}, \mathbf{x}_{p,t}^{(i)}) \propto \det\left(\mathbf{S}_t(\mathbf{x}_{p,t}^{(i)})\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{y}_t - \mathbf{H}_t \bar{\mathbf{x}}_t^b)^{\mathrm{T}} \mathbf{S}_t^{-1}(\mathbf{x}_{p,t}^{(i)})(\mathbf{y}_t - \mathbf{H}_t \bar{\mathbf{x}}_t^b)\right). \tag{28}$$

Under these choices, Algorithm 1 defines an adaptation scheme related to other approaches used in literature. Specifically, (28) is the same equation that was used for maximum likelihood estimation of covariance parameters (Dee 1995). Maximization of this function is achieved e.g. via simplex methods (Mitchell and Houtekamer 2000). In our approach, (28) serves as a likelihood function for Bayesian estimation of the tuning parameters, $\mathbf{x}_{p,t}$. The variance of the random walk then models our belief in time-variability of the tuning parameters. In the special case of stationary parameters, (i.e. $\sigma_\Delta = \sigma_r = \sigma_l = 0$), Algorithm 1 is reduced to a parallel run of n ensemble filters, each of which is accumulating the product of (28) in each step. After several hundreds of steps, majority of the weights will converge to zeros and one of them will converge to one. Such behavior is known as sample impoverishment in the particle filtering literature. The convergence of probability mass to a single point may be useful for finding the best tuned values in off-line phase. However, this degeneracy is undesirable for on-line application, and non-zero variances of random walks (26) have to be used.

For non-stationary parameters, each of the n filters follows a random walk of the tuning parameters. The re-sampling operation removes filters that diverged into unlikely regions, and replaces them by copies of the filters with parameters that are more likely. The area of

14

higher likelihood is then explored by more filters in detail. This of course requires to run n ensemble filters in parallel which is computationally expensive. However, the key advantage of this approach is that it is able to optimize non-convex and multi-modal likelihood functions.

b. Computational simplifications within the MPF framework

Parallel evaluation of n ensemble filters is computationally inefficient especially when the tuning parameters are close to each other, for example in cases when the posterior is very sharp. In that case, we may represent the posterior by a point estimate, $\hat{\mathbf{x}}_t$, and explore the space of (28) only locally. This is a common approach in adaptive ensemble filtering and we will discuss its implications from the MPF point of view.

Adaptive evaluation of a single ensemble filter corresponds to approximation of the joint posterior (19) by a conditionally independent posterior

$$p\left(\mathbf{x}_t|\mathbf{y}_{1:t}\right) \approx p\left(\mathbf{x}_{c,t}|\hat{\mathbf{x}}_{p,t}, \mathbf{y}_{1:t}\right) p\left(\mathbf{x}_{p,t}|\mathbf{y}_{1:t}\right). \tag{29}$$

The problem is then split in an interaction of an ensemble filter and a filter on parameters, as used e.g. in (Anderson 2007a). The key difficulty of this approach is that the normalized constant of the ensemble filter (28) is informative only about the point estimate $\hat{\mathbf{x}}_{p,t}$ and not about its posterior density. Optimization of (28) around this value at time t is independent of the value at time $t-1$. Restoration of temporal correlation between values of the parameters is achieved e.g. by the means of temporal smoothing (Li et al. 2009).

To resemble the same approach in the MPF framework, we impose additional approxi-

---
Algorithm 2 Computational simplification of the MPF framework (CI-MPF)
---
  i. Generate initial particles $\left\{\mathbf{x}_{p,0}^{(i)}\right\}_{i=1}^{n}$ and set the initial statistics of the ensemble filter

  ii. For each new data record $\mathbf{y}_t$ do:

    (a) Compute weights (23), each particle evaluates one temporary background step of the ensemble filter,

    (b) Evaluate expected value $\hat{\mathbf{x}}_{p,t}$ and use it to evaluate the background update of the ensemble filter,

    (c) Perform analysis step the ensemble filter via (25),

    (d) If the number of efficient particles, n_{eff}, is lower than the chosen threshold, resample the particles.

---

mation

$$p\left(\mathbf{x}_{c,t}|\mathbf{x}_{p,t}^{(i)}, \mathbf{y}_{1:t}\right) \approx p\left(\mathbf{x}_{c,t}|\hat{\mathbf{x}}_{p,t}, \mathbf{y}_{1:t}\right), \tag{30}$$

in (20) and (24). What results is a combination of a single ensemble filter with a particle filter, Algorithm 2. We note the following:

- The MPF framework (Algorithm 1) requires n background updates and n analysis steps of the ensemble filter, while the simplified algorithm CI-MPF requires one background update, $n+1$ evaluations of (28), and one analysis step.

- Using only one ensemble filter corresponds to local search of extrema in (28). However, the weights $w_t^{(i)}$ preserve temporal correlation between likelihood values at steps t and $t-1$.

- Experiments reveal that the values of (28) for different particles are very close to each other and it takes tens or hundreds of steps for the efficient sample size to fall below a threshold. As a heuristic, we propose to stop the temporal evolution of the particles during this period.

- When the efficient sample size reaches the threshold, the new particles are sampled from $p(\mathbf{x}_{p,t}|\hat{\mathbf{x}}_{p,t})$ rather than $p(\mathbf{x}_{p,t}|\mathbf{x}_{p,t}^{(i)})$. This choice is motivated by the strong influence of $\hat{\mathbf{x}}_{p,t}$ on the weights $w_t^{(i)}$.

Approximation (30) was first presented by Mustière et al. (2006). We will use the label conditionally-independent MPF (CI-MPF) in this paper. Approximation of the full posterior on $\mathbf{x}_{p,t}$ by the expected value may seem as a coarse approximation since it looses all information about variance and higher order moments. More advanced assumptions can be made, for example, we may combine the original MPF and the CI-MPF approximation in a way that Δ_t and r_t are approximated using conditional independence, while n parallel EnSRFs are run for different values of l_t. A way to use higher order moments within this approach has been presented in (Šmídl and Quinn 2008), however, it implies changes within the conditional filter, i.e. the ensemble filter in this case.

4. Simulation Studies

a. Lorenz 96 model

To demonstrate versatility of the method, we test the MPF approach in the Lorenz 96 model by Lorenz and Emanuel (1998) which has been widely used in simulation studies. The

model is given by

$$\frac{dx_j}{dt} = x_{j-1}(x_{j+1} - x_{j-2}) - x_j + F, \tag{31}$$

where F is the model forcing and x_i are variables forming a cyclic chain. We define $x_{-1} = x_{J-1}$, $x_0 = x_J$ and $x_{J+1} = x_1$ to make (31) meaningful for all values of $j = 1 \ldots J$. We use 40 variables, and $F = 8$ for the strength of forcing. The model (31) can be integrated forward with the fourth-order Runge–Kutta scheme. The system is computationally stable for step of 0.05 non-dimensional units, which is also the step of the analysis. All the experiments are performed as twin experiments.

b. Stationary parameters

To create a baseline for comparison of adaptive tuning strategies, we performed parallel run of EnSRFs for fixed values of Δ, l selected on a rectangular grid, as in (Whitaker and Hamill 2002). From Bayesian point of view, this setup corresponds to estimation of stationary parameters:

$$p(\Delta, l | y_{1:t}) \propto p(\Delta, l) \prod_{\tau=1}^{t} p(y_\tau | y_{1:\tau-1}, \Delta, l) \tag{32}$$

where $p(\Delta, l)$ is a prior probability density on discrete values of Δ, l at the grid points, which is uniform, and $p(y_\tau | y_{1:\tau-1}, \Delta, l)$ is given by (28). For numerical stability, (32) is often computed in logarithmic scale where the product is replaced by the sum of marginal log-likelihoods.

The observation data are generated from the perfect model scenario where the 'true' state was generated by integrating the Lorenz-96 model (31) for 100000 steps and observations are

generated from the true state by addition of zero-mean Gaussian noise with variance $r = 1$. The analysis was performed by the EnSRF with covariance localization constructed using a compactly supported fifth-order piecewise rational function given by (4.10) in (Gaspari and Cohn 1999) with length scale parameter l. The results of a simulation experiment with 132 EnSRFs with 15 ensemble members, $r = 1$, $\Delta = [1.0, 1.1, \ldots, 1.10]$, and $l = [0, 1, \ldots, 11]$ are displayed in Figure 1 in two modalities. First, the traditional RMSE is computed for each couple of parameters,

$$\text{RMSE} = \frac{1}{99000} \sum_{t=1000}^{100000} \sqrt{\frac{1}{40} (\mathbf{x}_t - \hat{\mathbf{x}}_{c,t})^{\text{T}} (\mathbf{x}_t - \hat{\mathbf{x}}_{c,t})}, \tag{33}$$

where

$$\hat{\mathbf{x}}_{c,t} = \sum_{i=1}^{n} w^{(i)} \overline{\mathbf{x}}_{c,t}^{(i)}, \tag{34}$$

for the MPF algorithm. Second, the sum of marginal log-likelihoods (32) within the same time intervals is displayed for illustration.

Note that the contours of the marginal log-likelihood (32) correspond closely to the contours of the RMSE. This suggests that the marginal likelihood (28) is a good measure to optimize for the best RMSE in cases where the true state values are not known. The three best choices within each modality are given in Table 1. Note that two choices— $\Delta = 1.05, l = 8$ and $\Delta = 1.04, l = 6$—are in the top three for both criteria. The relative differences are rather small, however, in terms of normalized posterior probability (32) the best parameters in Table 1 are e^{82} times more likely that the second best.

c. Adaptive Estimation in Perfect Model Scenario

The same observation data used for estimation of stationary parameters were used to estimate the time-varying parameters in two different scenarios: (i) fixed $r_t = 1.0$, estimated Δ_t, l_t, and (ii) estimated r_t, l_t, Δ_t. The particle filter is using multinomial re-sampling (Gordon et al. 1993) and $n_{eff} = 0.8n$. The variances of random walks (26) for the MPF algorithm were chosen as

$$
\begin{aligned}
\sigma_\Delta &= 0.01\Delta_{t-1}^{(i)} + 0.0001, \\
\sigma_l &= 0.01l_{t-1}^{(i)} + 0.0001, \\
\sigma_r &= 0.005r_{t-1}^{(i)} + 0.0001,
\end{aligned}
\tag{35}
$$

The CI-MPF algorithm was run with 10 times greater variances of the random walk than (35). The prior density of the tuning parameters is chosen as uniform on support $p(\sigma_{y,0}) = U(0.1, 4)$, $p(\Delta_0) = U(1.0, 1.10)$, $p(l_0) = U(0.11, 11.11)$.

Results of simulations for different number of particles, both algorithms (MPF and CI-MPF), and both scenarios are displayed in Tables 2 and 3, respectively. In accordance with (Whitaker and Hamill 2002; Li et al. 2009), we ignore the first 1000 steps and report the results only for the subsequent steps. Illustration of the behavior of the algorithm in the first 100 steps of scenario (ii) is displayed in Fig. 2 , via the mean value of the posterior density and the region between minimum and maximum value of the particles. Spatial distribution of the particles for Δ_t, l_t in scenario (i) at time steps $t = 1, 50000, 100000$ is displayed in Fig. 3 on the background of contours for the stationary marginal log-likelihood from Fig. 1. We note that alignment of the particles in the middle of the stationary contour at $t = 50000$ is a coincidence, in majority of all other time steps the cloud is a bit off the stationary optimum.

As expected, the RMSE of both tuning algorithms is steadily decreasing with increasing number of particles for all considered scenarios. The exact MPF algorithm clearly outperforms the conditionally-independent approximation CI-MPF. Note that for $n = 10$ and higher, the MPF filter achieves better performance than the best-tuned filter. On the other hand, the CI-MPF is coming close to the best-tuned values but never better. However, attractiveness of the CI-MPF lies in the fact that it imposes smaller computational overhead over the original EnSRF, while computational complexity of the MPF is n times that of the EnSRF.

In the more complex scenario of tuning all three parameters, we can observe much higher gap between performance of MPF and CI-MPF, Table 3. The MPF algorithm achieves only negligible increase of the RMSE over the scenario with known r, while CI-MPF is significantly worse. Especially with $n = 5$, the overall performance of the CI-MPF algorithm is severely degraded. This is caused by an instability in estimation that occurred around time $t = 90000$, see Fig. 4. Much smaller anomalies of a similar kind occur for all tested numbers of particles in the CI-MPF algorithm, while no such anomaly occurred in any run of the MPF algorithm. These anomalies also do not occur in any run of the CI-MPF algorithm of the simple scenario with known r. Therefore, we conjecture that with increasing uncertainty in $\mathbf{x}_{p,t}$ the enforced locality of approximation (29) is quickly loosing accuracy.

We note that good performance of the adaptive tuning was achieved for as low as 10 particles. This result is especially promising since it suggests that even more challenging assimilation scenarios can be handled at comparable computational complexity. This is particularly true for the exact MPF algorithm since addition of one extra tuning parameter had negligible impact on the performance. On the other hand, the CI-MPF algorithm does

not scale as well.

d. Model with Random Perturbations

For comparison with (Li et al. 2009), we tested the MPF algorithms on data simulated with model (31) with additive random perturbations

$$\frac{dx_j}{dt} = x_{j-1}(x_{j+1} - x_{j-2}) - x_j + F + \alpha e_t, \qquad (36)$$

where e_t is Gaussian distributed noise with zero-mean and unit variance. The observed data were generated using model (36) with $\alpha = 4$ for 100000 steps. The same setup of the EnSRF as in the previous experiments was used, including the same initial conditions. Results for estimation for the 100000 steps are displayed in Table 4. Since parameter α is stationary, the time evolution of the parameter estimates $\hat{\Delta}_t, \hat{l}_t$ using MPF is reaching the stationary values in Table 4 after the initial convergence period. However, convergence of the CI-MPF filter is much slower, causing large std in Table 4.

Note that the additive noise was compensated by higher values of $\hat{\Delta}_t$ and lower values of $\hat{l}_t$ than that of the perfect model, Table 3. This is in agreement with findings of Li et al. (2009) and also expected because α is increasing the background covariance (reflected by higher inflation) and decreasing correlation between elements of the state vector (reflected by lower length scale).

For testing the tracking properties of the MPF algorithms, we have designed a scenario with time varying α_t according to a triangular profile. Posterior densities of the parameters obtained using the MPF algorithm are displayed in Fig. 5. For this experiment, we

22

increased the variances of random walks (26) to

$$\sigma_\Delta = 0.01\Delta_{t-1}^{(i)} + 0.001,$$

$$\sigma_l = 0.01l_{t-1}^{(i)} + 0.001,$$

$$\sigma_r = 0.01r_{t-1}^{(i)} + 0.01.$$

In this case, the CI-MPF algorithm was unable to follow the change in the parameters, confirming the observation of its slower convergence from the experiment with stationary α. This experiment also confirms the trend of increasing Δ_t and decreasing l_t with increasing α_t. Note that when α returns to the stationary values, so do the estimates of the tuning parameters.

5. Discussion and Conclusion

The purpose of this paper is to present Marginalized Particle Filtering (also known as Rao-Blackwellized filtering) as an attractive tool for research of data assimilation methods in environmental modeling and especially for tuning of ensemble filters. The method is based on partitioning of the state (or unknown parameters) into two parts: (i) unknowns estimated by a conditional filter, and (ii) unknowns estimated by a particle filter. The original MPF assumes that the conditional filter is analytically tractable, which allows to prove advantages over a pure particle filter. In this paper, we propose to replace analytical filters by ensemble filters. The resulting algorithm is loosing its theoretical advantages, however it allows to address the problem of tuning of ensemble filters. We have shown that the number of particles needed to achieve acceptable performance is rather low, for example 10 particles

23

are sufficient to achieve on-line tuning of the inflation factor and the length scale parameter in the EnSRF for the 40-dimensional Lorenz model. Furthermore, we have shown that the approach easily extends to estimation of unknown variance of the observation error and potentially any other tuning parameters. Once again, 10 particles were sufficient to achieve performance comparable to that of the best-tuned filter.

Computational cost of the MPF framework is high since it requires running n EnSRFs in parallel. An attempt to reduce the cost was presented using the enforcement of conditional independence, resulting in an algorithm that adapts a point estimate of parameters for a single EnSRF. Similarities of this CI-MPF algorithm to previously published methods were discussed. The CI-MPF algorithm performed well only on simple scenarios with tuning of low number of stationary parameters where it was significantly faster than the exact MPF algorithm. However, in more demanding scenarios its performace significatly dropped. We expect that of parallel evaluation of ensemble filters over adaptation of a single ensemble filter will become apparent in even more demanding scenarios. Computational complexity may prevent its operational use, however, it may be an important tool for gaining insight into the ensemble filters in the same spirit as in (Anderson 2007b).

The potential of the framework has been demonstrated on on-line tuning of the ensemble filters. However, it is not the only scenario where it can be used. Since posterior density of the MPF is a mixture of Gaussians, the approach may be adapted for estimation of Gaussian mixture filters that have been studied e.g. by Bengtsson et al. (2003).

This paper is only an initial step in application of the MPF framework in this domain. Clearly, more work is required to discover full potential of the method. The open problems include justified design of suitable models of parameter evolution and approximations reduc-

24

ing the computational cost of the MPF. However, the existence of the exact solution allows

to design the necessary computational simplifications to resemble its behavior.

References

Anderson, J., 2001: An ensemble adjustment Kalman filter for data assimilation. *Monthly weather review*, **129 (12)**, 2884–2903.

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

Anderson, J., 2007b: Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter. *Physica D: Nonlinear Phenomena*, **230 (1-2)**, 99–111.

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

Anderson, J., S. Anderson, I. Metron, and V. Reston, 1999: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Monthly Weather Review*, **127**, 12.

Bengtsson, T., C. Snyder, and D. Nychka, 2003: Toward a nonlinear ensemble filter for high-dimensional systems. *Journal of Geophysical Research-Atmospheres*, **108 (D24)**, 8775.

Bocquet, M., C. Pires, and L. Wu, 2010: Beyond Gaussian statistical modeling in geophysical data assimilation. *Monthly Weather Review*, **138**, 2999–3023.

Burgers, G., P. Van Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Monthly Weather Review*, **126 (6)**, 1719–1724.

Dee, D., 1995: On-line estimation of error covariance parameters for atmospheric data assimilation. *Monthly weather review*, **123 (4)**, 1128–1145.

Doucet, A., N. de Freitas, and N. Gordon, (Eds.) , 2001: *Sequential Monte Carlo Methods in Practice.* Springer.

Doucet, A., N. de Freitas, K. Murphy, and S. Russell, 2000: Rao-Blackwellised particle filtering for dynamic Bayesian networks. *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence*, 176–183.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, **99**, 10 143–10 162.

Gaspari, G. and S. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, **125 (554)**, 723–758.

Gordon, N., D. Salmond, and A. Smith, 1993: Novel approach to nonlinear/non-Gaussian Bayesian state estimation. IEE, Vol. 140, 107–113.

Hoteit, I., D. Pham, G. Triantafyllou, and G. Korres, 2008: A new approximate solution of the optimal nonlinear filter for data assimilation in meteorology and oceanography. *Monthly Weather Review*, **136 (1)**, 317.

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

Ide, K., P. Courtier, M. Ghil, and A. Lorenc, 1997: Unified notation for data assimilation: Operational, sequential and variational. *J. Met. Soc. Japan*, **75 (1B)**, 181–189.

Jeffreys, H., 1961: *Theory of Probability*. 3d ed., Oxford University Press.

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

Lorenz, E. and K. Emanuel, 1998: Optimal sites for supplementary weather observations: Simulation with a small model. *Journal of the Atmospheric Sciences*, **55**, 399–414.

Mitchell, H. and P. Houtekamer, 2000: An adaptive ensemble Kalman filter. *Monthly Weather Review*, **128 (2)**, 416–433.

Mustière, F., M. Bolić, and M. Bouchard, 2006: A modified Rao-Blackwellised particle filter. *Proceedings of the IEEE conference on Acoustics, Speech, and Signal Processing*, Toulouse, France.

Nakano, S., G. Ueno, and T. Higuchi, 2007: Merging particle filter for sequential data assimilation. *Nonlinear Processes in Geophysics*, **14 (4)**, 395–408.

Peterka, V., 1981: Bayesian approach to system identification. *Trends and Progress in System identification*, P. Eykhoff, Ed., Pergamon Press, Oxford, 239–304.

Pham, D., 2001: Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Monthly Weather Review*, **129**, 1194–1207.

Saha, S., E. Özkan, V. Šmídl, and F. Gustafsson, 2010: Marginalized particle filters for Bayesian estimation of noise parameters. *Proceedings of the 13th International Conference on Information Fusion*, Edinburgh, UK.

Schön, T., F. Gustafsson, and P.-J. Nordlund, 2005: Marginalized particle filters for mixed linear/nonlinear state-space models. *IEEE Transactions on Signal Processing*, **53**, 2279–2289.

Šmídl, V. and A. Quinn, 2008: Variational bayesian filtering. *IEEE Transactions on Signal Processing*, **56 (10)**, 5020–5030.

Snyder, C., T. Bengtsson, P. Bickel, and J. Anderson, 2008: Obstacles to high-dimensional particle filtering. *Monthly Weather Review*, **136**, 4629–4640.

Thrun, S., W. Burgard, and D. Fox, 2005: *Probabilistic Robotics (Intelligent Robotics and Autonomous Agents)*. MIT press, Cambridge, Massachusetts, USA.

Tippett, M., J. Anderson, C. Bishop, T. Hamill, and J. Whitaker, 2003: Ensemble Square Root Filters*. *Monthly Weather Review*, **131**, 1485–1490.

Van Leeuwen, P., 2009: Particle Filtering in Geophysical Systems. *Monthly Weather Review*, **137 (12)**, 4089–4114.

Vossepoel, F. and P. Van Leeuwen, 2007: Parameter estimation using a particle method: Inferring mixing coefficients from sea level observations. *Monthly weather review*, **135 (3)**, 1006–1020.

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

List of Tables

TABLE 1. Best stationary choices of nuisance parameters for EnSRF according to two criteria.

	RMSE			marginalized log-likelihood		
rank	inflation	localization	mean value	inflation	localization	sum
#1	1.04	7.0	0.2074	1.05	7.0	-2079401
#2	1.04	6.0	0.2075	1.05	8.0	-2079483
#3	1.05	8.0	0.2076	1.04	6.0	-2079513

TABLE 2. Adaptive tuning of $\mathbf{x}_{p,t} = [\Delta_t, l_t]$ and the resulting analysis RMSE error, averaged over assimilation steps between $t = 1000$ and $t = 100000$, std denotes standard deviation of the estimates from the mean over time.

MPF with perfect observation error								
n	mean RMSE	std RMSE	mean$\hat{\Delta}_t$	std $\hat{\Delta}_t$	mean $\hat{l}_t$	std $\hat{l}_t$	mean $\hat{r}_t$	std $\hat{r}_t$
5	0.2089	0.0521	1.0334	0.0080	5.6942	0.9072	1.0	0.0
10	0.2071	0.0510	1.0337	0.0064	6.3431	0.9229	1.0	0.0
20	0.2065	0.0523	1.0331	0.0061	6.2933	0.5005	1.0	0.0
CI-MPF with perfect observation error								
n	mean RMSE	std RMSE	mean $\hat{\Delta}_t$	std $\hat{\Delta}_t$	mean $\hat{l}_t$	std $\hat{l}_t$	mean $\hat{r}_t$	std $\hat{r}_t$
5	0.2101	0.0503	1.0405	0.0097	5.2201	0.7541	1.0	0.0
10	0.2095	0.0503	1.0338	0.0059	4.9780	0.5387	1.0	0.0
20	0.2089	0.0498	1.0396	0.0061	5.1979	0.5335	1.0	0.0

TABLE 3. Adaptive tuning of $\mathbf{x}_{p,t} = [\Delta_t, r_t, l_t]^T$ and the resulting analysis RMSE error, averaged over assimilation steps between $t = 1000$ and $t = 100000$, std denotes standard deviation of the estimates from the mean over time.

n	mean RMSE	std RMSE	mean $\hat{\Delta}_t$	std $\hat{\Delta}_t$	mean $\hat{l}_t$	std $\hat{l}_t$	mean $\hat{r}_t$	std $\hat{r}_t$
MPF with unknown observation error								
5	0.2094	0.0521	1.0317	0.0075	5.6154	0.7289	1.0059	0.0215
10	0.2072	0.0509	1.0354	0.0058	6.7455	0.9541	1.0031	0.0230
20	0.2064	0.0511	1.0355	0.0055	6.7182	0.9202	1.0018	0.0193
CI-MPF with unknown observation error								
5	0.2295	0.1856	1.0458	0.0255	7.2192	3.0843	1.0394	0.2284
10	0.2149	0.0637	1.0399	0.0181	6.2032	1.9454	1.0048	0.0312
20	0.2122	0.0524	1.0440	0.0169	5.1734	1.2772	0.9996	0.0142

TABLE 4. Adaptive tuning of $\mathbf{x}_{p,t} = [\Delta_t, l_t, r_t]^T$ for system with random model errors (36) with variance $\alpha = 4$. The resulting analysis RMSE error is averaged over assimilation steps between $t = 1000$ and $t = 100000$, time averages of parameter estimates are displayed in tandem with standard deviation of the estimates from the mean over time. Both algorithms were run with $n = 10$ and $n_{eff} = 0.8n$.

	mean RMSE	std RMSE	mean $\hat{\Delta}_t$	std $\hat{\Delta}_t$	mean $\hat{l}_t$	std $\hat{l}_t$	mean $\hat{r}_t$	std $\hat{r}_t$
MPF	0.3563	0.0542	1.1944	0.0230	3.0810	0.8198	1.0055	0.0618
CI-MPF	0.4189	0.1539	1.1171	0.0343	3.6739	0.4772	1.1293	0.1811

List of Figures

FIG. 1. Estimation of stationary parameters on rectangular grid. **Top**: marginal log-likelihood (32) of the tuned parameters Δ, l. Labels of the countour lines denote difference from the maximum which is marked by a circle. **Bottom**: Time average of RMSE (33).
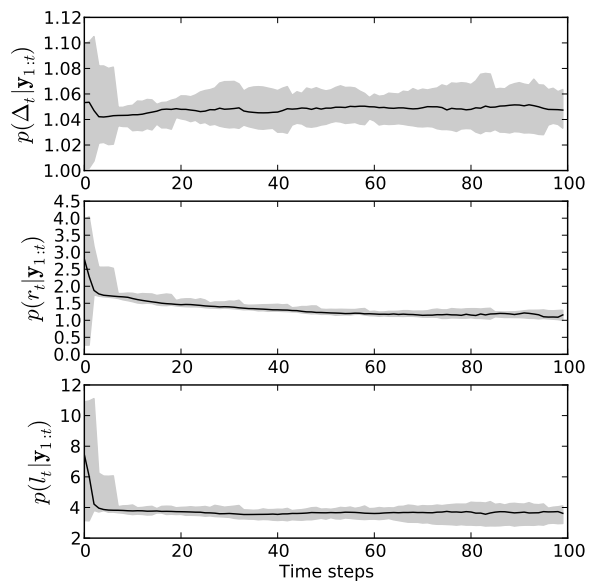
FIG. 2. Illustration of the posterior densities of the parameters via posterior mean (black line) and range between minimum and maximum value (gray area).
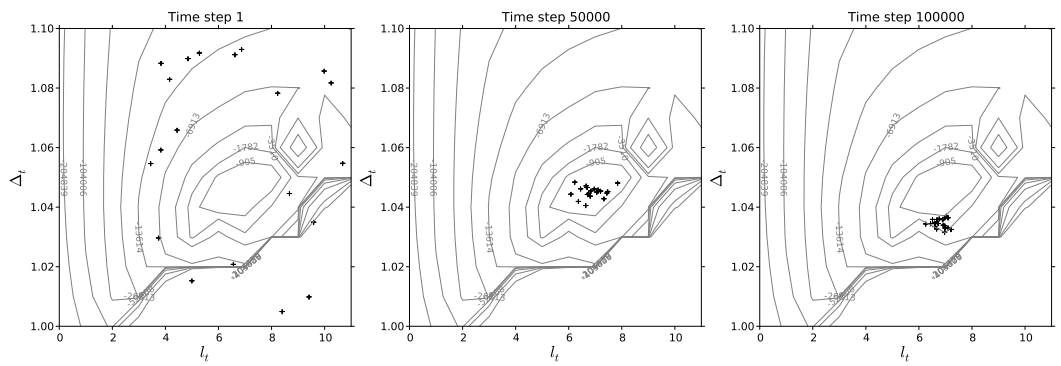
FIG. 3. Position of the particles $\Delta_t^{(i)}, l_t^{(i)}$ denoted by '+' on the backround of stationary contours of marginal log-likelihood.
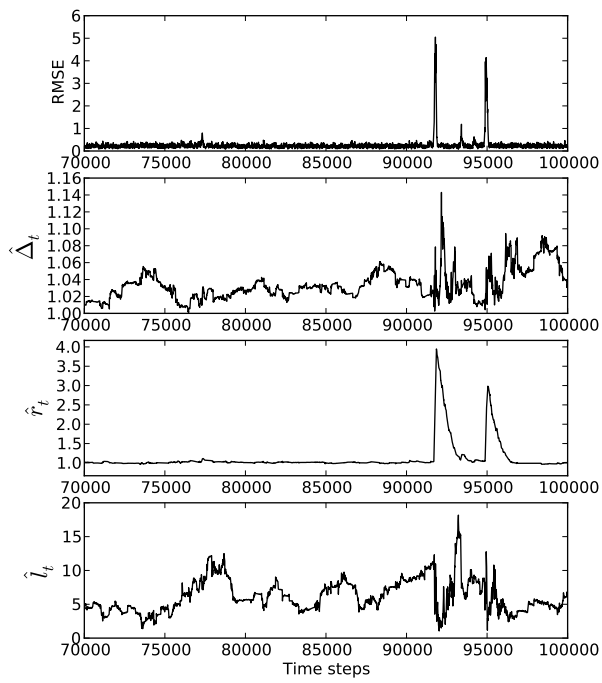
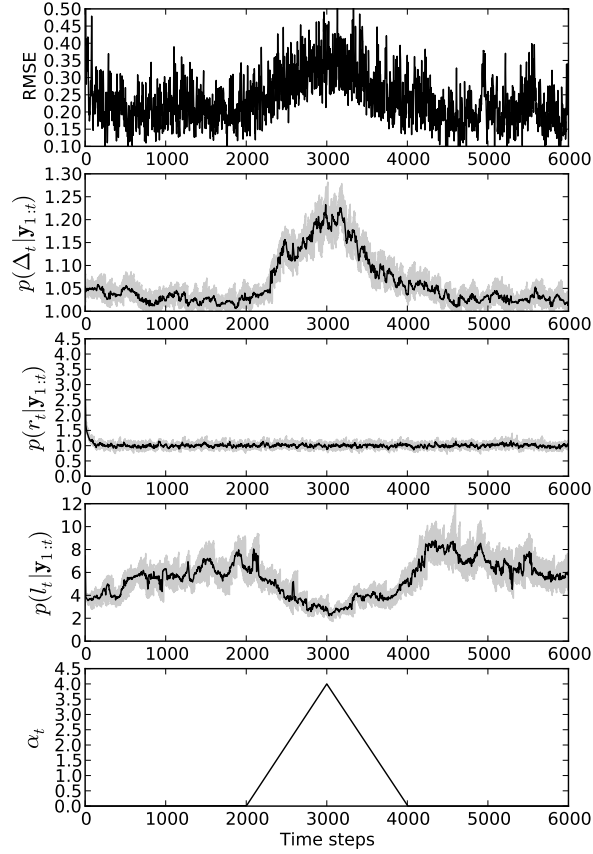FIG. 4. Instability in parameter estimation via the CI-MPF algorithm.

Fig. 5. Estimation of system (36) with time-varying α_t of triangular profile displeyd at the bottom. Posterior densities of the parameters are displayed via their mean value (black line) and region between minimum and maximum value of the particles (gray area).