Nonlin. Processes Geophys. Discuss., 2, 1631–1658, 2015 www.nonlin-processes-geophys-discuss.net/2/1631/2015/ doi:10.5194/npgd-2-1631-2015 © Author(s) 2015. CC Attribution 3.0 License.



This discussion paper is/has been under review for the journal Nonlinear Processes in Geophysics (NPG). Please refer to the corresponding final paper in NPG if available.

A local particle filter for high dimensional geophysical systems

S. G. Penny^{1,2,3} and T. Miyoshi^{3,1}

¹University of Maryland, College Park, Maryland, USA ²National Centers for Environmental Prediction, College Park, MD, USA ³RIKEN Advanced Institute for Computational Science, Kobe, Japan

Received: 31 October 2015 – Accepted: 5 November 2015 – Published: 7 December 2015

Correspondence to: S. G. Penny (steve.penny@noaa.gov)

Published by Copernicus Publications on behalf of the European Geosciences Union & the American Geophysical Union.

Discussion Pa	NPGD 2, 1631–1658, 2015 A local particle filter for high dimensional geophysical systems S. G. Penny and T. Miyoshi		
per Dis			
cussion Pap			
)er			
—	Abstract	Introduction	
Dis	Conclusions	References	
cussion	Tables	Figures	
n Pa	14	►I	
per	•	•	
—	Back	Close	
Discus	Full Screen / Esc		
sion	Printer-friendly Version		
Pape	Interactive Discussion		
- T			

Abstract

A local particle filter (LPF) is introduced that outperforms traditional ensemble Kalman filters in highly nonlinear/non-Gaussian scenarios, both in accuracy and computational cost. The standard Sampling Importance Resampling (SIR) particle filter is augmented with an observation-space localization approach, for which an independent analysis is computed locally at each gridpoint. The deterministic resampling approach of Kitagawa is adapted for application locally and combined with interpolation of the analysis weights to smooth the transition between neighboring points. Gaussian noise is applied with magnitude equal to the local analysis spread to prevent particle degeneracy while maintaining the estimate of the growing dynamical instabilities. The approach 10 is validated against the Local Ensemble Transform Kalman Filter (LETKF) using the 40-variable Lorenz-96 model. The results show that: (1) the accuracy of LPF surpasses LETKF as the forecast length increases (thus increasing the degree of nonlinearity), (2) the cost of LPF is significantly lower than LETKF as the ensemble size increases, and (3) LPF prevents filter divergence experienced by LETKF in cases with non-Gaussian observation error distributions.

1 Introduction

The Particle Filter (PF) has been explored in the data assimilation community since the introduction of its Gaussian linear variant, the Ensemble Kalman Filter (EnKF) in
 the mid-1990's (Evensen, 1994). While general PFs have been intractable for high dimensional systems, the EnKF has experienced great success in numerical weather prediction (NWP) (e.g., Kleist, 2012; Hamrud et al., 2014) and ocean data assimilation (e.g., Penny et al., 2015). However, at least two limitations are on the horizon for EnKFs. Perhaps counter-intuitively, these limitations arise due to *increased* computational resources, and have already become challenges at the RIKEN Advanced Institute for Computational Science (AICS, e.g., Miyamoto et al., 2013; Miyoshi et al., 2014, 2015).



First, global models will be pushed to higher resolutions in which they begin to resolve highly nonlinear processes. To maintain the Gaussian linear assumption required for the EnKF, much smaller timesteps are needed. For example, the standard 6 h analysis cycles used for the atmosphere may need to be decreased to 5 min or even 30 s. Second, large ensembles (e.g., with ensemble size $k > 10\,000$ members) will become

- ⁵ Second, large ensembles (e.g., with ensemble size $k > 10\,000$ members) will become feasible for lower-resolution models. While at first this may seem an advantage rather than a limitation, the computational cost of the local ensemble transform Kalman filter (LETKF) (Hunt et al., 2007), for example, increases at a rate $O(k^3)$ with the ensemble size k. Thus as the ensemble size increases, the cost of computing the analysis increases at a much greater rate. Alternative EnKFs feasible for large geophysical sys-
- tems scale in computational cost with the observation dimension, which is typically multiple orders of magnitude larger than the ensemble dimension.

The PF is generally applicable to nonlinear non-Gaussian systems, including cases with multi-modal distributions or nonlinear observation operators. With little difficulty,

- PFs can explicitly include representation of model error, nonlinear observation operators (Nakano, 2007; Lei and Bickel, 2011), non-diagonal observation error covariance matrices, and non-Gaussian likelihood functions. For example, observed variables such as precipitation are inherently non-Gaussian and cannot be effectively assimilated by standard EnKF techniques (e.g., Lien et al., 2013, 2015). In the expansion
- to sea-ice and land data assimilation applications, the non-Gaussian quantities such as ice concentration, ice thickness, snow cover, and soil moisture outnumber those that can be modeled with Gaussian error. Bocquet et al. (2010) further review the difficulties using observations with non-Gaussian error distributions. All of these problem-specific variations can create great difficulties for standard methods, such as the EnKF or variational approaches (3-D-Var/4-D-Var), as used in current operational systems.

Sampling Importance Resampling (SIR) (also known as the bootstrap filter, Gordon et al., 1993) is a commonly used enhancement to the basic Sequential Importance Sampling (SIS) particle filter. However, even with resampling the number of ensemble members required by the SIR particle filter to capture the high probability region of the



posterior in high-dimensional geophysical applications is too large to make SIR usable (Ades and van Leeuwen, 2013). Snyder et al. (2008) found that the number of required ensemble members scales exponentially with the size of the system, giving the example that a 200 dimensional system would require 10^{11} members. However, Snyder et al. note that clever choices of the proposal distribution could overcome the need for these exponentially large ensemble sizes in high-dimensional systems, which has been more recently explored by Synder et al. (2015). Applying such an approach, van Leeuwen (2003) considers a model for the Agulhas Current with dimension of roughly 2×10^5 . Further, Beskos et al. (2012) discuss recursive methods for estimating the proposal densities, similar to the Running-in-Place algorithm (Yang et al., 2012a, b; Penny

posal densities, similar to the Running-in-Place algorithm (Yang et al., 2012a, b; Penny et al., 2013) that has been used with LETKF in meteorological and oceanographic data assimilation.

Techniques such as localization and inflation are typically applied as modifications to make the EnKF operationally feasible. Inspired by this practice, we introduce a local particle filter (LPF) designed for geophysical systems that is scalable to high dimensions and has computational cost O(k). Spatial localization is typically justified by the fact that long distance correlations are either spurious or weak in comparison to nearby correlations, particularly when the ensemble is under-sampled. We use this same approach to reduce the required ensemble size for the LPF.

20 2 Methodology

Localization is used in most operational NWP data assimilation systems, either through a direct scaling of the background error covariance matrix (e.g., Whitaker and Hamill, 2002) or by a scaling of the observation error covariance matrix (Hunt et al., 2007). Because the computation of a background error covariance matrix is not needed for the

²⁵ PF, the latter approach is applied here to develop an efficient PF for high-dimensional geophysical systems. Localization reduces the dimensionality of the solution space,



thus requiring fewer ensemble members to sample the phase space. Gaussian noise is applied as an additive inflation to prevent particle degeneracy.

The standard SIR particle filter 2.1

There are many variations of the PF (Stewart and McCarty, 1992; Gordon et al., 1993; Kitagawa, 1996; Hurzeler and Kunsch, 1998; Liu and Chen, 1998). In essence it is simply a Monte Carlo estimation of Bayes Theorem, reformulated as a recursion (Doucet et al., 2001),

$$\rho\left(\boldsymbol{x}_{t}|\boldsymbol{y}_{1:t}\right) = \frac{\rho\left(\boldsymbol{y}_{t}|\boldsymbol{x}_{t}\right)\rho\left(\boldsymbol{x}_{t}|\boldsymbol{y}_{1:t-1}\right)}{\rho\left(\boldsymbol{y}_{t}|\boldsymbol{y}_{1:t-1}\right)}$$
(1)

where $p(\mathbf{x}_t | \mathbf{y}_{1:t})$ is the probability of the state x at time t, given all observations y up to time t. We consider the model domain to be described by vectors x with dimension m, the observation domain to be descried by vector \mathbf{y} with dimension *I*, and an ensemble of size k. The term in the numerator can be expressed using the Chapman-Kolmogorov equation as,

$$\rho\left(\boldsymbol{x}_{t}|\boldsymbol{y}_{1:t-1}\right) = \int \rho\left(\boldsymbol{x}_{t}|\boldsymbol{x}_{t-1}\right) \rho\left(\boldsymbol{x}_{t-1}|\boldsymbol{y}_{1:t-1}\right) \mathrm{d}\boldsymbol{x}_{t-1}$$
(2)

and similarly the term in the denominator can be expressed as, 15

$$\rho\left(\boldsymbol{y}_{t}|\boldsymbol{y}_{1:t-1}\right) = \int \rho\left(\boldsymbol{y}_{t}|\boldsymbol{x}_{t}\right) \rho\left(\boldsymbol{x}_{t}|\boldsymbol{y}_{1:t-1}\right) \mathrm{d}\boldsymbol{x}_{t}.$$
(3)

The two factors in the numerator of Eq. (1) are sampled using a numerical model f,

$$\rho\left(\boldsymbol{x}_{t}|\boldsymbol{y}_{1:t-1}\right) \approx \frac{1}{k} \sum_{i=1}^{k} \delta\left(\boldsymbol{x}_{t} - f\left(\boldsymbol{x}_{t-1}^{i}\right)\right)$$

$$\rho\left(\boldsymbol{y}_{t}|\boldsymbol{x}_{t}\right) = g\left(\boldsymbol{y}_{t}|\boldsymbol{x}_{t}\right).$$

$$(4)$$

$$(5)$$

Discussion Paper **NPGD** 2, 1631-1658, 2015 A local particle filter for high dimensional geophysical systems S. G. Penny and T. Miyoshi **Title Page** Introductio Abstract Conclusions References Tables **Figures** Close Back Full Screen / Esc **Printer-friendly Version** Interactive Discussion

Discussion

Paper

Discussion Paper

Discussion Paper

The term in Eq. (5) is typically called the likelihood, because the probability of y given x is equivalent to the likelihood of x given y, i.e., $p(y|x) \equiv \ell(x|y)$. The function g is general and can represent any distribution for the observations.

For the experiments here we generate two experiment cases, each with a different 5 likelihood function. First we use a Gaussian likelihood function corresponding to that used for EnKFs.

$$g\left(\boldsymbol{y}_{t}|\boldsymbol{x}_{t}\right) = \frac{1}{\sqrt{(2\pi)^{m}|\boldsymbol{\mathsf{R}}|}} \exp\left[-\frac{1}{2}(\boldsymbol{y}_{t} - h(\boldsymbol{x}_{t}))^{T}\boldsymbol{\mathsf{R}}^{-1}(\boldsymbol{y}_{t} - h(\boldsymbol{x}_{t}))\right]$$
(6)

where the function h is a general, possibly nonlinear, observation operator mapping from the model state space to the observation space. For the LPF, it is straightforward to generalize to arbitrary non-Gaussian likelihood functions. As an example, we also 10 apply a multivariate Gaussian mixture model (GM₂) following Fowler and van Leeuwen (2013) with pdf.

$$p(\mathbf{y}|\mathbf{x}) \propto v_{w} \exp\left[(\mathbf{y} + v_{1}\mathbf{1} - h(\mathbf{x}))^{T} \mathbf{R}^{-1} (\mathbf{y} + v_{1}\mathbf{1} - h(\mathbf{x}))\right] + (1 - v_{w}) \exp\left[(\mathbf{y} + v_{2}\mathbf{1} - h(\mathbf{x}))^{T} \mathbf{R}^{-1} (\mathbf{y} + v_{2}\mathbf{1} - h(\mathbf{x}))\right]$$
(7)

Let each ensemble member be identified with an index, *i*. Normalized weights are evaluated for each member. 15

$$w_t^i = \frac{\rho\left(\boldsymbol{y}_t | \boldsymbol{x}_t^i\right)}{\sum_{j=1}^k \rho\left(\boldsymbol{y}_t | \boldsymbol{x}t^j\right)}.$$

Then the posterior is,

$$\rho\left(\boldsymbol{x}_{t}|\boldsymbol{y}_{1:t}\right) \approx \sum_{i=1}^{k} w_{t}^{i} \delta\left(\boldsymbol{x}_{t} - f\left(\boldsymbol{x}_{t-1}^{i}\right)\right).$$
1636

NPGD 2, 1631-1658, 2015 A local particle filter for high dimensional geophysical systems **Discussion** Paper S. G. Penny and T. Miyoshi **Title Page** Introduction Abstract Conclusions References **Figures** Tables Close Back Full Screen / Esc **Printer-friendly Version** Interactive Discussion

Discussion Paper

Discussion Paper

Discussion Paper

(8)

(9)

Based on Liouville's theorem, the evolution of a probability measure in a dynamical system satisfies the property that "the probability of finding trajectories inside the time-variant volume W(t) is constant during the evolution of the dynamical system" (Property 2, http://www.ulb.ac.be/di/map/gbonte/ftp/bontempi_fpde.pdf). If the solution manifold ⁵ expands in some directions, so will the pdf represented by the particles. Thus, the fidelity of the distribution will quickly become insufficient to sample a solution manifold around the true trajectory. A resampling procedure is used to refocus the particles on the densest areas of the distribution at each analysis step. For the experiments here, we use a resampling procedure that resembles resampling with replacement. After resampling we have,

$$p\left(\boldsymbol{x}_{t}|\boldsymbol{y}_{1:t}\right) \approx \frac{1}{k} \sum_{i=1}^{k} \delta\left(\boldsymbol{x}_{t} - \boldsymbol{x}_{t}^{i}\right).$$

2.2 The transform interpretation

The PF can be interpreted similarly to the Ensemble Transform Kalman Filter (ETKF) of Bishop et al. (2001). Namely, we define the PF solution as a transformation of the background ensemble to the analysis ensemble,

$$\mathbf{X}^{\mathrm{a}} = \mathbf{X}^{\mathrm{b}}\mathbf{T}$$
(11)

where each column of \mathbf{X}^{b} is a background ensemble member, and each column of \mathbf{X}^{a} is an analysis ensemble member.

Let \boldsymbol{b} be the vector of background particle indices and \boldsymbol{a} be the vector of analysis particle indices,

$$\boldsymbol{b} = \left\{ \boldsymbol{z} \in \mathbb{Z}^k \, | \boldsymbol{z} = (1, 2, 3, \dots, k) \right\}$$
(12)

$$\boldsymbol{a} = \left\{ \boldsymbol{z} \in \mathbb{Z}^{k} | \boldsymbol{z} = (a_{1}, a_{2}, a_{3}, \dots, a_{k}), a_{i} \in [1, k] \right\}$$
(13)

$$\boldsymbol{e}_{j} = \left\{ \boldsymbol{z} \in \{0, 1\}^{k} \mid \boldsymbol{z} = (0, .., 1_{j}, ..., 0) \right\}.$$
(14)

Discussion Paper

Discussion Paper

Discussion Paper

Discussion Paper

(10)

If \boldsymbol{e}_i are the canonical basis vectors then we can define,

$$\mathbf{E}_{k\times k} = \begin{bmatrix} \boldsymbol{e}_{a_1} \boldsymbol{e}_{a_2} \cdots \boldsymbol{e}_{a_k} \end{bmatrix}.$$
(15)

For the standard PF, the indicator matrix **E** is made up of k (not necessarily unique) standard basis vectors e_i , with entries 0 and 1 that we will interpret as weights. Thus the analysis ensemble for the PF is defined simply by the transform,

$$\mathbf{X}_{m \times k}^{\mathsf{a}} = \mathbf{X}_{m \times k}^{\mathsf{b}} \mathbf{E}_{k \times k}$$

15

We note that by using this approach, each new analysis member, with index *i*, maintains the continuity properties of its associated background member, a_i .

For reference in the next section, the components of the analysis matrix \mathbf{X}^{a} will have the form,

$$\mathbf{X}^{a} = \begin{bmatrix} x_{1,i}e_{i,1} & x_{1,i}e_{i,2} & \cdots & x_{1,i}e_{i,k} \\ x_{2,i}e_{i,1} & x_{2,i}e_{i,2} & \cdots & x_{2,i}e_{i,k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m,i}e_{i,1} & x_{m,i}e_{i,2} & \cdots & x_{m,i}e_{i,k} \end{bmatrix}.$$
(17)

Here we have used the Einstein tensor notation for the elements, in which $x_{1,i}e_{i,1}$ represents a summation over the index *i* (i.e., the inner product of row 1 of **X**^b and column 1 of **E**). While the summation index could be represented generically by any symbol, we reuse the symbol "*i*" due to its correspondence with the background particle indices as defined above.

2.3 The localization approach

Snyder et al. (2008) note that when either the model dimension or observation count is large, the PF requires significantly more particles to give an adequate representation

ISCUSSION **NPGD** 2, 1631-1658, 2015 Paper A local particle filter for high dimensional geophysical systems **Discussion** Paper S. G. Penny and T. Miyoshi **Title Page** Introductio Abstract Discussion Paper Conclusions References Tables **Figures** Close Back Discussion Full Screen / Esc **Printer-friendly Version** Paper Interactive Discussion

(16)

of the system. Localization reduces both the model and observation dimensions by dividing the problem into a series of sub-domains, thus reducing the required number of particles for accurate filtering. The LPF uses the approach of Hunt et al. (2007) to select nearby observations for independent analysis at each grid point. Nearby grid ⁵ points thus assimilate nearly identical sets of observations to derive their analyses.

We use the deterministic resampling of Kitagawa (1996), with complexity O(k), adapted for local use as described next. A uniform partition of the interval [0,1] with width 1/k is first generated globally, with an offset applied from a uniform distribution over the interval [0,1/k]. The same partition is used locally for resampling at each grid point. Cumulative sums of the normalized weights (Eq. 8),

$$\breve{w}_t^j = \sum_{i=1}^j w_t^i$$

are compared with the elements of the partition. Traversing from j = 1, ..., k, all unassigned particles with index j having a corresponding cumulative sum with index j that surpasses the next element of the partition (ordered monotonically increasing) are as-

- ¹⁵ signed as particles of the resampled (analysis) ensemble. For a given grid point, when the cumulative sums of the particle weights are near one of the partition values, there may be sensitivity in neighboring grid points that lead to discontinuities between local analyses. The analysis ensemble at this grid point consists of a subset of background particle indices (1 through *k*) with repetitions. To eliminate the discontinuities
- with neighboring grid points, with the particle indices we associate weights of a local transform function T, nominally either 1.0 (full weight) or 0.0 (no weight). This is partially inspired by the "weight interpolation" of Bowler (2006), applied to LETKF by Yang et al. (2009), who found that interpolation of weights was more robust than interpolation of state values. At a single grid point, there are k pieces of background information
- ²⁵ about the possible system state at that point. In the standard PF, only 1 out of these k pieces of information is retained for each analysis ensemble member, based on the overall agreement with observations. In the LPF we use anywhere from 1 to k mem-

(18)

bers to construct each analysis member based on the agreement with observations within a local radius.

For the LPF, a new transform is defined for each point in the model domain to generate a set of *m* indicator matrices, $\mathbf{E}_{k\times k}^{(j)}$, so that for each point $(x_j, \text{ for } j = 1...m)$,

$${}_{5} \left(\mathbf{X}_{m \times k}^{\mathrm{a}} \right)^{j} = \mathbf{X}_{m \times k}^{\mathrm{b}} \mathbf{E}_{k \times k}^{(j)}.$$
 (19)

Using the summation tensor notation described in the previous section, the analysis ensemble can be written,

$$\mathbf{X}^{a} = \begin{bmatrix} x_{1,i}e_{i,1}^{(1)} & x_{1,i}e_{i,2}^{(1)} & \cdots & x_{1,i}e_{i,k}^{(1)} \\ x_{2,i}e_{i,1}^{(2)} & x_{2,i}e_{i,2}^{(2)} & \cdots & x_{2,i}e_{i,k}^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m,i}e_{i,1}^{(m)} & x_{m,i}e_{i,2}^{(m)} & \cdots & x_{m,i}e_{i,k}^{(m)} \end{bmatrix}.$$

10

The transform matrix may have any degree of sophistication. We apply a smoothing operator by modifying the weights \mathbf{e}_{a_i} associated with each analysis particle index $a_{(i)}$ from a binary value to a continuous value between 0 and 1, while maintaining all column sums equal to one, and call this new transform matrix **W**. This smoothing is performed only in the ensemble space; no explicit interpolation is applied in the model space.

We define the concept of a "neighbor point" abstractly as a point near the analyzed ¹⁵ grid point based on a specified distance metric. If there are *N* neighbor points, then there will be at most min(N + 1, k) collocated pieces of background information that can be utilized to construct each analysis ensemble member at this point. An example is given in Fig. 1. In our case, with a sufficiently large set of observations the indices for these neighbor points are calculated from nearly identical observation innovations.

²⁰ Therefore, when there is a sufficiently large ensemble size (k) the difference between the states associated with different particle indices will be small. The transform function T is applied across all background indices (i.e., for particles 1,...,k) at this grid point to compute the analysis.



(20)

2.4 Particle degeneracy

The particle selection process of the PF reduces the rank of the ensemble. For a linear deterministic system this leads to a rapid collapse of the ensemble and divergence of the filter. For a sufficiently stochastic nonlinear system the full rank is recovered af-

- ⁵ ter a single forecast step. If the nonlinear system is not sufficiently stochastic, then we must address the ensemble initialization problem at every analysis cycle. Pazo et al. (2010) discuss the desirable properties in an initial ensemble, namely the members: (1) should be well-embedded in the attractor, (2) should be statistically equivalent but have enough diversity to represent a significant portion of the phase space,
- (3) should adequately represent the error between the analysis and true state, and
 (4) should sample the fastest growing directions in phase space. We wish to avoid particle degeneracy while also engendering some of these qualities. Therefore we employ a simple approach: at each cycle we add Gaussian noise with variance scaled locally to a magnitude matching the analysis error variance and apply this to each analysis
 member prior to the subsequent ensemble forecast.

2.5 Computational complexity

20

A data assimilation system is comprised of many components. We simplify the cost analysis in order to gain an approximate relative measure of the algorithms presented here. Let *m* be the model dimension, *l* be the observation dimension, and \bar{l}_i be the average local observation dimension. The total cost (C_T) of an analysis cycle is equal to the overhead (C_H) of the assimilation system plus *m* times the average local cost (C_L) of the assimilation method plus *k* times the cost of one model forecast (C_M) of duration τ ,



$$C_{\mathsf{T}}(k,l,m) = C_{\mathsf{H}}(k,l,m) + m \cdot C_{\mathsf{L}}\left(k,\bar{l}_{i}\right) + k \cdot C_{\mathsf{M}}(\tau,m).$$
⁽²¹⁾

We will assume that between the two methods the overhead and model costs are approximately equal. The primary difference in cost between the two systems is then the average local cost,

$$C_{L}^{LPF} = O\left(k\overline{l}_{i}\right)$$

$$C_{L}^{LETKF} = O\left(k^{2}\overline{l}_{i} + k^{3}\right).$$

$$(22)$$

$$(23)$$

If as is typically the case, the system size m is large and the ensemble size k is small, then

$$C_{\mathsf{T}}(k,l,m) \approx C_{\mathsf{H}}(k,l,m) + O(m),$$

¹⁰ and the difference in cost between LETKF and LPF is small. However for large k, we see that the average local cost of LETKF,

$$C_{\mathrm{T}}^{\mathrm{LETKF}}(k,l,m) = C_{\mathrm{H}}(k,l,m) + O\left(mk^{3}\right)$$
(25)

exceeds that of the LPF,

$$C_{\rm T}^{\rm LPF}(k,l,m) = C_{\rm H}(k,l,m) + O(mk).$$
⁽²⁶⁾

¹⁵ Subtracting the overhead costs, in this case the LPF is a factor of k^2 cheaper than LETKF.

2.6 Data assimilation methods

We enumerate the benefits of the LPF vs. the benchmark LETKF, an ensemble square root filter that performs its analysis in the ensemble space at each grid point using



(24)

a geospatially local selection of observations. The LETKF approach is very efficient as long as the ensemble size is small relative to the number of observations and the model dimension.

We use LETKF as a proxy for a general EnKF. Nerger (2015) gives a comparison 5 between LETKF and the Ensemble Square Root Filter (ESRF) of Whitaker and Hamill (2002), while Tippett et al. (2003) indicate that the ESRF is identical to the Ensemble Adjustment Kalman Filter (EAKF) of Anderson (2001) when using serial single observation processing.

2.7 Experiment design

- ¹⁰ We demonstrate the algorithms on the Lorenz-96 system (Lorenz, 1996), composed of m = 40 grid points, using Lorenz's original forcing, F = 8.0. The Lorenz-96 system has frequently been used to demonstrate the PF and other data assimilation algorithms for the geosciences (e.g. Nakano et al., 2007; van Leeuwen, 2010; Lei and Bickel, 2011; Penny, 2014). Observations are sampled from a nature run of Lorenz-96 after running the model for 14 400 timesteps to allow the model to settle on the attractor. Gaussian noise is added to each observation with a standard deviation of 0.5. For various experiments, the Lorenz-96 system is observed either at every 0.05 or 0.5 timesteps, reflective of a 6 and 60 h forecast, respectively, based on Lorenz's original description of the system. Observations are sampled randomly on the interval [0,m], and a linear interpolation is used for the observation operator. The last experiment case
- uses a bimodal Gaussian mixture distribution to represent observational error.

3 Results

25

The standard SIR PF performs poorly with any ensemble size O(m). For example, using 1500 particles and 20 randomly chosen observations per analysis cycle leads to rapid filter divergence for the L96 system, even in a relatively linear regime (dt = 0.05)



of the system (Fig. 2). On the contrary, LETKF performs well even with few ensemble members and few observations per cycle (k = 20, l = 10). Localization is consistent between each method, using r = 2 gridpoints. For a given ensemble size, increasing the localization radius degraded the accuracy of both methods. To explore the relative advantages of each approach, we will describe a series of cases in which the LETKF outperforms the LPF, and in which the LPF outperforms LETKF.

Case 1: typical forecast lengths (dt = 0.05, or 6 h) 3.1

Lorenz (1996) introduced the dt = 0.05 timescale as being comparable to the error doubling taking place over 6 h in the operational forecasting systems of the early 1990's. In this relatively linear timescale of the L96 system, LETKF clearly outperforms the LPF at a given ensemble size. This is expected as EnKFs take advantage of the Gaussian/linear assumption. When the experiment parameters match such assumptions (even loosely), LETKF performs guite well. However, using localization, the LPF can perform adequately (i.e., avoid filter divergence) in a similar parameter regime (Fig. 3). Thus for this case, we find that LETKF attains higher accuracy than the LPF, but the 15 LPF improves upon the accuracy and stability of the standard SIR PF for a given en-

semble size.

3.2 Case 2: long forecast lengths (dt = 0.50, or 60 h)

To increase the degree of nonlinearity in a data assimilation system using L96, it is typical to increase the analysis cycle length (e.g., Lei and Bickel, 2011). The LPF has 20 superior performance for more nonlinear regimes of the L96 system (e.g., dt = 0.5) provided there are many ensemble members, e.g., O (100). Using 80 observations per cycle and 100 ensemble members, for example, LETKF produces occasional errors that propagate eastward (along the positive x direction). The LPF does not produce such effects, and the errors are generally lower than with LETKF (Fig. 4). We consider 25



this a relevant scenario because the majority of observations in operational weather forecasting are discarded (Ochatta et al., 2005).

Exploring a more complete parameter space, we examine the forecast error for LETKF over a range of observation coverage (l = 2, ..., 80 per analysis cycle) and en-⁵ semble sizes (k = 10, ..., 400), and compare the relative difference to LPF. Figure 5 shows the average absolute error over 600 analysis cycles of length dt = 0.5 for 1600 different parameter combinations of observation coverage (l) and ensemble size (k). The LPF is more accurate than LETKF when using many observations (e.g., l > 20) and large ensemble sizes (e.g., k > 100-200)

Further, when examining the computational cost of the LPF vs. LETKF, the relative costs reflect the analytical assessment given above in Sect. 2.5. Namely, the elapsed time of the LETKF experiments grows exponentially with ensemble size, while the elapsed time of the LPF is significantly lower at large ensemble sizes (Fig. 6).

3.3 Case 3: non-Gaussian observation error

- ¹⁵ The previous section examined the impacts of nonlinearity and non-Gaussianity on the forecast. We now examine the impacts of non-Gaussian observation error. Using a multivariate Gaussian mixture model (GM₂) following Fowler and van Leeuwen (2013), we apply a corresponding random error to each observation and compare the impacts on LETKF and LPF. We use the LETKF without modification, but modify the likelihood ²⁰ function of LPF to the definition of GM₂ as in Sect. 2.1, Eq. (7). Using $v_w = 0.1$, $v_1 = -1$, $v_n = 1$, we create a bimodal distribution biased toward the second Gaussian mode. The
- $v_2 = 1$, we create a bimodal distribution biased toward the second Gaussian mode. The analysis cycle is dt = 0.05 (6 h) as in experiment case 1, Sect. 3.1. Figure 7 compares LETKF and LPF using k = 100 ensemble members and l = 80 observations. An additional result is given for LPF with l = 20 observations. The introduction of a strong
- ²⁵ non-Gaussianity in the observation error distribution disrupts LETKF and eventually creates errors that propagate throughout the entire domain. Using the same ensemble size and observation count, the LPF gains significant advantage in its ability to explic-



itly account for the non-Gaussian error structure of the observations. Even reducing the observation count by 75%, the LPF maintains its advantage.

4 Conclusions

The Local Particle Filter (LPF) has been shown to outperform a state of the art ensem-⁵ ble Kalman filter (i.e., LETKF) in scenarios that violate the Gaussian/linear assumptions of the Kalman filter. We showed the advantage of the LPF when the forecast is more non-linear (via longer analysis cycles, or less frequent observations), and when the observation error is non-Gaussian (using a bimodal error distribution). Further, upon transitioning to large ensembles the LPF has a significant cost savings relative to LETKF.

The LPF maintains many of the attractive qualities that give Particle Filters (PFs) advantages over standard EnKFs. While the LPF is not optimal for all possible data assimilation scenarios, there is great potential for the LPF to be combined with more traditional approaches to create adaptive hybrid systems that can avoid catastrophic filter divergence and manage multi-modal forecast distributions, nonlinear observation operators, non-Gaussian observations.

We found that a large number of ensemble members (or particles) and observations are sufficient for the LPF to match or surpass the accuracy of LETKF. We find large ensemble sizes a relevant scenario for realistic systems running on large supercomputers such as the K computer at RIKEN. The use of large observation sets is relevant

²⁰ in operational weather forecasting because many of the dense satellite data are currently discarding in a thinning process. Further, in this parameter regime the LPF has significantly lower computational cost than LETKF.

In a realistic system, some mechanism is needed to drive the ensemble toward the observations in the event of the ensemble drifting away from the true state. The PF itself

has no inherent mechanism to do this other than the brute force generation of more particles. There are many techniques in the PF literature for managing filter divergence, but none of them are foolproof. One method to alleviate this problem is to introduce



proposal densities (van Leeuwen, 2009) that guide the particles toward the observed state prior to resampling. Another popular mechanism to achieve this is regularization, which uses a kernel to sample from a continuous distribution at the resampling stage.

Finally, while the inflation mechanism used here was effective for the Lorenz-96 sys-

- tem, it is not adequate for more realistic atmospheric or oceanic models. For such systems, either geospatially correlated noise or stochastic physics parameterizations may be capable of performing the same function. Stochastic physics parameterizations are an active area of research, and are under development for a number of operational center models, including NCEP (Hou et al., 2006, 2010; Kolczynski et al., 2015), ECMWF (Borner et al., 2009; Weisbeimer et al., 2014; Watern et al., 2015), and the Met Office.
- ¹⁰ (Berner et al., 2009; Weisheimer et al., 2014; Watson et al., 2015), and the Met Office (Tennant et al., 2011; Sanchez et al., 2014; Shutts and Paleres, 2014; Shutts, 2015).

Acknowledgements. We gratefully acknowledge the Japan Society for the Promotion of Science (JSPS) whose FY2013 fellowship supported this work. We would also like to thank the RIKEN Advanced Institute for Computational Science (AICS) for hosting Penny.

15 **References**

20

25

- Ades, M., van Leeuwen, P. J.: An exploration of the equivalent weights particle filter, Submitted to, Q. J. Roy. Meteorol. Soc., 139, 820–840, 2013.
- Berner, J., Shutts, G. J., Leutbecher, M., and Palmer, T. N.: A spectral stochastic kinetic energy backscatter scheme and its impact on flow-dependent predictability in the ECMWF ensemble
- prediction system, J. Atmos. Sci., 66, 603–626, doi:10.1175/2008JAS2677.1, 2009.
- Beskos, A., Crisan, D., and Jasra, A.: On the Stability of Sequential Monte Carlo Methods in High Dimensions, arXiv:1103.3965v2 [stat.CO], 2012.
- Bocquet, M., Pires, C. A., and Wu, L.: Beyond Gaussian Statistical Modeling in Geophysical Data Assimilation, Mon. Weather Rev., 138, 2997–3023, doi:10.1175/2010MWR3164.1, 2010.
- Bowler, N.: Comparison of error breeding, singular vectors, random perturbations and ensemble Kalman filter perturbation strategies on a simple model, Tellus A, 58, 538–548, 2006.



- 1648
- assimilation, Mon. Weather Rev., 139, 3964-3973, 2011. Lien, G.-Y., Kalnay, E., and Miyoshi, T.: Effective assimilation of global precipitation: simulation experiments, Tellus A, 65, 19915, doi:10.3402/tellusa.v65i0.19915, 2013.
- 27 Conference On Weather Analysis And Forecasting/23 Conference On Numerical Weather Prediction, recorded presentation available at: https://ams.confex.com/ams/27WAF23NWP/ webprogram/Paper273838.html, last access: 25 November 2015. Lei, J. and Bickel, P.: A moment matching ensemble filter for nonlinear non-Gaussian data
- Kitagawa, G.: Monte Carlo filter and smoother for non-Gaussian non-linear state space models, J. Comput. Graph. Stat., 5, 1–25, 1996. Kolczynski, W., Pegion, P., Hamill, T., Whitaker, J. S., Hou, D., Zhu, Y., and Zhou, X.: Investigat-

ing a New Stochastic Physics Suite for Use in the NCEP Global Ensemble. Am. Met. Soc.,

- Scheme Representing Model-Related Uncertainties in the NCEP Global Ensemble Forecast System, NOAA/NCEP/EMC, available at: http://www.emc.ncep.noaa.gov/gmb/yzhu/gif/pub/ Manuscript_STTP_Tellus_A_HOU-1.pdf (last access: 25 November 2015), 2010. Hurzeler, M. and Kunsch, H.: Monte Carlo approximations for general state-space models, J. 20 Comput. Grap. Stat., 7, 175–193, doi:10.2307/1390812, 1998.
- Hou, D., Toth, Z., and Zhu, Y.: A Stochastic Parameterization Scheme Within NCEP Global Ensemble Forecast System. Am. Met. Soc., 18 Conference on Probability and Statistics in the Atmospheric Sciences, available at: http://ams.confex.com/ams/pdfpapers/101401.pdf (last access: 25 November 2015), 2006. Hou, D., Toth, Z., Zhu, Y., Yang, W., and Wobus, R.: A Stochastic Total Tendency Perturbation

15

25

30

Bayesian state estimation, IEE Proc.-F., 140, 107, 1993. Hoffman, R. N. and Kalnay, E.: Lagged Average Forecasting: an alternative to Monte Carlo 10 Forecasting, Tellus, 35, 100–118, 1983.

Monte Carlo methods to forecast error statistics, J. Geophys. Res., 99, 10143–10162, 1994. 5 Fowler, A. and van Leeuwen, P. J.: Observation impact in data assimilation: the effect of non-Gaussian observation error, Tellus A, 65, 20035, doi:10.3402/tellusa.v65i0.20035, 2013.

Gordon, N. J., Salmond, D., and Smith, A. F. M.: Novel approach to nonlinear/non-Gaussian

Evenson, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using

ods, in: SMC in Practice, available at: http://www.stats.ox.ac.uk/~doucet/smc_resources.html (last access: 25 November 2015), 2001.

Doucet, A., De Freitas, N., and Gordon, N. J.: An introduction to Sequential Monte Carlo Meth-

Discussion

Paper

2, 1631-1658, 2015

A local particle filter for high dimensional geophysical systems

> S. G. Penny and T. Miyoshi



1649

a global ocean general circulation model, Nonlin. Processes Geophys., 20, 1031-1046,

 2139–2149, doi:10.1175/MWR-D-13-00131.1, 2014.
 Penny, S. G., Kalnay, E., Carton, J. A., Hunt, B. R., Ide, K., Miyoshi, T., and Chepurin, G. A.: The local ensemble transform Kalman filter and the running-in-place algorithm applied to

doi:10.5194/npg-20-1031-2013, 2013.

- tion, Tellus A, 56, 415–428, doi:10.1111/j.1600-0870.2004.00076.x, 2004.
 Pazo, D., Rodriguez, M. A., and Lopez, J. M.: Spatio-temporal evolution of perturbations in ensembles initialized by bred, Lyapunov, and singular vectors, Tellus A, 62, 10–23, 2010.
 Penny, S. G.: The Hybrid Local Ensemble Transform Kalman Filter, Mon. Weather Rev., 142,
- Ott, E., Hunt, B. R., Szunyogh, I., Zimin, A. V., Kostelich, E. J., Corazza, M., Kalnay, E., Patil, D. J., and Yorke, J. A.: A Local Ensemble Kalman Filter for atmospheric data assimilation, Tellus A, 56, 415–428, doi:10.1111/j.1600-0870.2004.00076.x, 2004.
- Ochatta, T., Gebhardt, C., Saupe, D., and Wergen, W.: Adaptive thinning of atmospheric observations in data assimilation with vector quantization and filtering methods, Q. J. Roy. Meteorol. Soc, 131, 3427–3437, 2005.
- Nakano, S., Ueno, G., and Higuchi, T.: Merging particle filter for sequential data assimilation, Nonlin. Processes Geophys., 14, 395–408, doi:10.5194/npg-14-395-2007, 2007.
 Nerger, L.: On serial observation processing in Localized Ensemble Kalman Filters, Mon. Weather Rev., 143, 1554–1567, doi:10.1175/MWR-D-14-00182.1, 2015.
- Miyoshi, T., Kondo, K., and Terasaki, K.: Big ensemble data assimilation in numerical weather prediction, Computer, 48, 15–21, doi:10.1109/MC.2015.332, 2015.
- Miyoshi, T., Kondo, K., and Imamura, T.: The 10240-member ensemble Kalman filtering with an intermediate AGC M, Geophys. Res. Lett., 41, 5264–5271, doi:10.1002/2014GL060863, 2014.
- atmospheric convection in a subkilometer global simulation, Geophys. Res. Lett., 40, 4922– 4926, doi:10.1002/grl.50944, 2013. Mivoshi T. Kondo K. and Imamura T. The 10240-member ensemble Kalman filtering with
- Miyamoto, Y., Kajikawa, Y., Yoshida, R., Yamaura, T., Yashiro, H., and Tomita, H.: Deep moist atmospheric convection in a subkilometer global simulation, Geophys. Res. Lett., 40, 4922–

Lien, G.-Y., Miyoshi, T., and Kalnay, E.: Assimilation of TRMM Multisatellite Precipitation

Analysis with a low-resolution NCEP Global Forecasting System, Mon. Weather Rev.,

doi:10.1175/MWR-D-15-0149.1, in press, 2015. Liu, J. S. and Chen, R.: Sequential Monte Carlo methods for dynamic systems, JAS A., 93, 1032–1044, doi:10.2307/2669847, 1998.

5

10

15

20

25

30

NPGD

2, 1631-1658, 2015

Discussion Paper

Discussion Paper

Discussion Paper

Discussion

Paper

A local particle filter for high dimensional geophysical systems

> S. G. Penny and T. Miyoshi

Title	Title Page		
Abstract	Introduction		
Conclusions	References		
Tables	Figures		
14	۰		
•	•		
Back	Close		
Full Scre	Full Screen / Esc		
Printer-frier	Printer-friendly Version		
Interactive	Interactive Discussion		

Sanchez, C., Williams, K. D., Shutts, G., and Collins, M.: Impact of a Stochastic Kinetic Energy Backscatter scheme across time-scales and resolutions, Q. J. Roy. Meteorol. Soc., 140, 2625–2637, 2014.

Shutts, G.: A stochastic convective backscatter scheme for use in ensemble prediction systems, Q. J. Roy. Meteorol. Soc., 141, 2602–2616, doi:10.1002/gi.2547, 2015.

tems, Q. J. Roy. Meteorol. Soc., 141, 2602–2616, doi:10.1002/qj.2547, 2015. Shutts, G. and Pallarès, A. C.: Assessing parametrization uncertainty associated with horizontal resolution in numerical weather prediction models, Philos. T. Roy. Soc. A, 372, 20130284, doi:10.1098/rsta.2013.0284, 2014.

Snyder, C., Bengtsson, T., Bickel, P., and Anderson, J.: Obstacles to High-Dimensional Particle Filtering, Mon. Weather Rev., 136, 4629–4640, 2008.

10

Snyder, C., Bengtsson, T., and Morzfeld, M.: Performance bounds for particle filters using the optimal proposal, Mon. Weather Rev., 143, 4750–4761, doi:10.1175/MWR-D-15-0144.1, 2015.

Stewart, L. and McCarty, P.: The use of Bayesian Belief Networks to fuse continuous and dis-

- ¹⁵ crete information for target recognition and discrete information for target recognition, tracking, and situation assessment, in Proc, P. Soc. Photo.-Opt. Ins., 1699, 177–185, 1992.
 - Tennant, W. J., Shutts, G. J., Arribas, A., and Thompson, S. A.: Using a Stochastic Kinetic Energy Backscatter Scheme to Improve MOGREPS Probabilistic Forecast Skill, Mon. Weather Rev., 139, 1190–1206, doi:10.1175/2010MWR3430.1, 2011.

van Leeuwen, P. J.: A variance-minimizing filter for large-scale applications, Mon. Weather Rev., 131, 2071–2084, 2003.

van Leeuwen, P. J.: Nonlinear data assimilation in geosciences: an extremely efficient particle filter, Q. J. Roy. Meteorol. Soc., 136, 1991–1999, 2010.

Watson, P. A. G., Christensen, H. M., and Palmer, T. N.: Does the ECMWF IFS Convection Pa-

rameterization with Stochastic Physics Correctly Reproduce Relationships between Convection and the Large-Scale State?, J. Atmos. Sci., 72, 236–242, doi:10.1175/JAS-D-14-0252.1, 2015.

Weisheimer, A., Corti, S., Palmer, T., and Vitart, F.: Addressing model error through atmospheric stochastic physical parametrizations: impact on the coupled ECMWF seasonal forecasting

system, Philos. T. Roy. Soc. A, 372, 20130290, doi:10.1098/rsta.2013.0290, 2014.
 Yang, S.-C., Kalnay, E., and Hunt, B. R.: Weight interpolation for efficient data assimilation with the Local Ensemble Transform Kalman Filter, Q. J. Roy. Meteorol. Soc., 135, 251–262, 2009.



- Yang, S.-C., Kalnay, E., and Hunt, B. R.: Handling nonlinearity in Ensemble Kalman Filter: experiments with the three-variable Lorenz model, Mon. Weather Rev., 140, 2628–2646, doi:10.1175/MWR-D-11-00313.1, 2012a.
- Yang, S.-C., Kalnay, E., and Miyoshi, T.: Improving EnKF spin-up for typhoon assimilation and prediction, Weather Forecast., 27, 878–897, 2012b.

5

NPGD 2, 1631-1658, 2015 A local particle filter for high dimensional geophysical systems S. G. Penny and T. Miyoshi **Title Page** Abstract Introduction References Conclusions Tables **Figures** ◄ Close Back Full Screen / Esc **Printer-friendly Version** Interactive Discussion

Discussion Paper

Discussion Paper

Discussion Paper

Discussion Paper



Figure 1. A hypothetical example depicting the construction of a single analysis member. Each level represents a different background ensemble member (particle), with a model space composed of a 3×3 grid. The nodes of the grid are circled if the member is chosen for the construction of analysis member 1 by the LPF. The numerals indicate the ids for the background members that will be averaged at the corresponding node, in this case based on the immediately adjacent neighbor points of that node.





Figure 2. Analysis error using an analysis cycle window length dt = 0.05 (6 h) for (a) the standard SIR PF using k = 1500 particles with l = 20 observations per analysis cycle, and (b) LETKF with localization radius r = 2, k = 20 ensemble members, and l = 10 observations per analysis cycle, sampled randomly on the domain.





Figure 3. Analysis error for (a) LETKF and (b) LPF, using an analysis cycle window length dt = 0.05 (6 h), localization radius r = 2 grid points, k = 40 ensemble members, and l = 20 observations sampled randomly on the domain (prescribed observation locations and errors are identical for both methods).





Figure 4. Forecast error for (a) LETKF and (b) LPF, using an analysis cycle window length dt = 0.5 (60 h), localization radius r = 2 grid points, k = 100 ensemble members, and l = 80 observations sampled randomly on the domain (observation locations and errors are identical for both methods).





Figure 5. Forecast error for **(a)** LETKF and **(b)** LPF minus LETKF. LPF reduces error in highly sampled cases with larger observation coverage. The LPF increases error in poorly sampled cases and with low observation coverage. Each cell represents one experiment case; absolute errors are averaged over the entire domain for 600 analysis cycles for each case.





Figure 6. Elapsed time in seconds for (a) LETKF and (b) LPF minus LETKF.





Figure 7. Analysis error for (a) LETKF and (b) LPF, using l = 80 observations and k = 100 ensemble members. The observations used between (a) and (b) are identical. In (c), the number of observations for LPF is reduced to l = 20, but improvement in accuracy vs. LETKF remains.

