Consortium

for

Small-Scale Modelling

Technical Report No. 10

Is the Particle Filtering Approach appropriate for Meso-Scale Data Assimilation ?

December 2007

DOI: 10.5676/DWD_pub/nwv/cosmo-tr_10

Deutscher Wetterdienst

MeteoSwiss

ΕΘΝΙΚΗ ΜΕΤΕΩΡΟΛΟΓΙΚΗ ΥΠΗΡΕΣΙΑ

Ufficio Generale Spazio Aereo e Meteorologia

Instytucie Meteorogii i Gospodarki Wodnej

Administratia Nationala de Meteorologie

Agenzia Regionale per la Protezione Ambientale del Piemonte

Agenzia Regionale per la Protezione Ambientale dell' Emilia-Romagna: Servizio Idro Meteo

Centro Italiano Ricerche Aerospaziali

Amt für GeoInformationswesen der Bundeswehr

www.cosmo-model.org

Editor: Massimo Milelli, ARPA Piemonte
Printed at Deutscher Wetterdienst, P.O. Box 100465, 63004 Offenbach am Main

Is the Particle Filtering Approach appropriate for Meso-Scale Data Assimilation ?

Mikhail D. Tsyrulnikov

Mikhail D. Tsyrulnikov
Hydrometeorological Centre of Russia
11-13 B. Predtechensky Lane
123242 Moscow
Russia

Contents

1 Introduction

This review aims to discuss the suitability of *particle filtering* as a method of choice for the future COSMO meso-scale data assimilation system.

2 Nonlinear filtering

In this introductory section some basic facts on nonlinear filtering are given in order to facilitate the subsequent discussion of the particle filtering technique.

2.1 Statement of the problem and notation

Let us be interested in the system (atmospheric, oceanic...) state, X, at the time instant t_k, X_k. The information we have on X_k is twofold:

1. We have the deterministic model,

$$X_k^f = \mathcal{M}_{k-1}(X_{k-1}^a), \tag{1}$$

   where the superscript "a" means "analysis" (forecast initial conditions) and "f" forecast.

2. We have observations Y_s, $s = k, k-1, k-2, ...$, which are related to the state in the known way:

$$Y_s = H_s(X_s) + \eta_s, \tag{2}$$

   where η_s is the observation error, which includes both the measurement error and the error in the observation operator H_s (representativeness error).

We assume that the deterministic model, Eq. (1), *approximates* the true system evolution, so that if we substitute, into Eq. (1), the *true system* states (not the *model* states), a discrepancy arises, which is called "model error", ξ:

$$X_k = \mathcal{M}_{k-1}(X_{k-1}) - \xi_{k-1}. \tag{3}$$

We stress that this equation is a stochastic dynamic model for the *true* system state.

In order to properly assimilate the information contained both in the observations and the model, we assume the error sources, η_s and ξ_s ($s = k, k-1, k-2, ...$), to be independent of X and governed by the respective stochastic models, so that their probability distributions are known, $p(\eta_s)$ and $p(\xi_s)$, respectively. For the assimilation to be *sequential*, we require that both η_s and ξ_s are *white* sequences. For simplicity, we also assume that observation errors and model errors are (probabilistically) independent.

2.2 Sequential non-linear filtering

In data assimilation, our goal is the conditional distribution of the state, given all current and past observations (called posterior distribution):

$$p(X_k|Y_{:k}), \tag{4}$$

where " $:k''$ means "up to k" (at the time moments $t_k, t_{k-1}, t_{k-2}, ...$).

To develop an assimilation technique, we break up the observations into the present and past batches: $Y_{:k} = (Y_{:k-1}, Y_k)$. Then, we transform Eq. (4) using the Bayes theorem:

$$p(X_k|Y_{:k}) = p(X_k|Y_{:k-1}, Y_k) \propto p(X_k|Y_{:k-1}) \cdot p(Y_k|X_k, Y_{:k-1}). \tag{5}$$

Here, $p(X_k|Y_{:k-1})$ is the *predictive* (prior) distribution (density) and

$$p(Y_k|X_k, Y_{:k-1}) = p(Y_k|X_k) \tag{6}$$

is the likelihood; the equality in Eq. (6) follows from the assumption that the observation error sequence is white.

Thus,

$$p(X_k|Y_{:k}) \propto p(X_k|Y_{:k-1}) \cdot p(Y_k|X_k). \tag{7}$$

The likelihood is assumed to be known (from the observation operator and the observation error probability distribution), so all it remains to be found is the predictive probability density. From Eq. (3), we have:

$$p(X_k|Y_{:k-1}) = p[\mathcal{M}_{k-1}(X_{k-1}) - \xi_{k-1} = X_k|Y_{:k-1}].$$

Here, both X_{k-1} and ξ_{k-1} are random and unknown, so we use the "complete probability rule": $p(X) = \mathbf{E}p(X|\vartheta)$, with ϑ being some random variable (vector):

$$p(X_k|Y_{:k-1}) = \mathbf{E}_{X_{k-1}} p[\xi_{k-1} = \mathcal{M}_{k-1}(X_{k-1}) - X_k|X_{k-1}, Y_{:k-1}] \tag{8}$$

(here the conditional expectation is taken over X_{k-1}). Using the *previous* posterior density, $p(X_{k-1}|Y_{:k-1})$, we obtain from Eq. (8):

$$p(X_k|Y_{:k-1}) = \int p(X_{k-1}|Y_{:k-1}) \cdot p[\xi_{k-1} = \mathcal{M}_{k-1}(X_{k-1}) - X_k] \mathrm{d}X_{k-1}. \tag{9}$$

The fact that the second p.d.f. in the r.h.s. of this equation is unconditional, follows from the assumptions indicated above in the last paragraph of the previous subsection.

Using the explicit transition probability density, $p(X_k|X_{k-1})$, we can rewrite Eq. (9) as

$$p(X_k|Y_{:k-1}) = \int p(X_{k-1}|Y_{:k-1}) \cdot p(X_k|X_{k-1}) \mathrm{d}X_{k-1}. \tag{10}$$

Thus, Eqs. (7) and (9) provide the desired sequential non-linear assimilation algorithm: having the posterior density at time t_{k-1}, $p(X_{k-1}|Y_{:k-1})$, we compute (in theory) the prior density at time t_k (using Eq. (9)) and then update it using current observations, Eq. (7). At the very beginning of the assimilation process, we specify some initial distribution $p_0(X)$.

3 Monte-Carlo filters

For high-dimensional systems like the atmosphere, the above non-linear sequential assimilation technique in its pure and non-approximated form is absolutely unsuitable, mainly, due to the complexity of the integral in Eq. (9) or (10). The analysis (measurement update) step, Eq. (7), may become feasible for high-dimensional systems if the observation operator is only weakly non-linear and the observation-error distribution is multivariate Gaussian (3D-Var). But the forecast step (time update) in its exact form is prohibitively expensive even for linearized (tangent-linear) dynamics and Gaussian model errors, when the Kalman filter is applicable. Simplified Kalman filters have not yet proven to be successful, so we resort to Monte-Carlo methods.

One of such methods, known as the Ensemble Kalman Filter (EnKF), relies on the Gaussian approximation of the predictive distribution (Evensen, 1994 and Houtekamer and Mitchell, 2005). But this approximation may become too restrictive if we attempt to treat highly non-linear atmospheric processes that involve moisture, clouds, precipitation etc. at fine spatial resolution. This leads us to turn our attention to fully non-linear assimilation methods known as "particle filters". The basic particle filter is called the bootstrap or SIR filter. The term SIR means "sequential importance resampling" or "sampling importance resampling".

4 The bootstrap (SIR) filter

The simplest particle filter proposed by Rubin (1987) relies on the *discrete* (point-support) approximation to the probability distributions in question (just like the statistical technique known as bootstrap (Efron and Tibshirani, 1993)). Namely, we proceed as follows

1. At the discrete time moment t_{k-1}, approximate the posterior distribution, $p(X_{k-1}|Y_{:k-1})$, by the *discrete* distribution, so that the whole probability mass is concentrated in N points (particles), $\{X^-_{k-1}(j)\}_{j=1}^N$ (the sign "$-$" means "before sampling from the posterior distribution"):

$$\hat{p}(X_{k-1}|Y_{:k-1}) = \sum_{j=1}^{N} w_{k-1}(j) \cdot \delta[X_k - X^-_{k-1}(j)], \tag{11}$$

   where the hat denotes an estimate and $w_{k-1}(j)$ are the weights.

2. Resampling: draw N samples from this discrete distribution: $\{X^+_{k-1}(j)\}_{j=1}^N$ (the sign "+" means "after sampling from the posterior distribution"). Now, all $\{X^+_{k-1}(j)\}$ are independent and identically distributed samples from the approximate posterior density Eq. (11).

3. Feed the time evolution model, Eq. (3), with $X_{k-1} = X^+_{k-1}(j)$, $j = 1, 2, ..., N$, and obtain (as it can be easily seen) a set of N samples from the approximate predictive (prior) distribution:

$$X^-_k(j) = \mathcal{M}_{k-1}[X^+_{k-1}(j)] - \xi_{k-1}(j), \tag{12}$$

   where $\xi_{k-1}(j)$ is the j-th pseudo-random realization of the model error on the (t_{k-1}, t_k) time interval.

4. Regard this sample from the approximate prior distribution as the discrete distribution with equal weights:

$$\hat{p}(X_k|Y_{:k-1}) = \frac{1}{N}\sum_{j=1}^{N}\delta[X_k - X_k^-(j)]. \tag{13}$$

   This is a virtual step, nothing is computed here. Actually, we can drop the assumption that the *prior* distribution is approximated by the discrete one, instead, we can derive the SIR equations by considering the analysis step as an importance sampler from the posterior (at time t_k) distribution (Pitt and Shephard, 1999), but we will not elaborate on this.

5. Accomplish the analysis step, Eq. (7), with the point-mass prior distribution, Eq. (13):

$$\hat{p}(X_k|Y_{:k}) \propto \frac{1}{N}\sum p(Y_k|X_k^-(j)) \cdot \delta[X_k - X_k^-(j)].$$

   Denoting $p(Y_k|X_k^-(j)) =: w_k'(j)$ and normalizing $w_k'(j)$ (in order to obtain a proper discrete distribution),

$$w_k(j) = \frac{w_k'(j)}{\sum_{i=1}^{N} w_k'(i)}, \tag{14}$$

   we get the point-mass approximation to the posterior distribution:

$$\hat{p}(X_k|Y_{:k}) = \sum_{j=1}^{N} w_k(j) \cdot \delta[X_k - X_k^-(j)]. \tag{15}$$

6. Now, we go to step 1 above and the process repeats.

Note that in the resampling step, if, for some j, the respective weight $w_k(j)$ is large, then, likely, this particle will be sampled several times, $N_j > 1$, but if $w_k(j)$ is small, then, likely, $N_j = 0$. So, the least likely (as identified by the likelihoods) particles are killed, whereas the most likely particles multiply ("survival of the fittest").

Finally, we discuss how the ordinary deterministic analysis can be obtained with the particle filtering assimilation scheme. As it can be easily seen from Eq. (15), the mean of the posterior distribution is

$$\hat{X}_k = \int X_k \cdot \hat{p}(X_k|Y_{:k})\mathrm{d}X_k = \sum w_k(j) \cdot X_k^-(j). \tag{16}$$

This mean provides the best estimate, in the mean-square sense, of the current state of the system, given all present and past observations. So, $\hat{X}_k$ can be used as the "analysis".

Eq. (16) implies that the analysis is just a linear combination of an inevitably <u>small number</u> of particles. This has very serious consequences for the applicability of a particle filtering data assimilation scheme for such a highly variable system as the atmosphere (see the discussion below).

5 Inherent advantages and disadvantages of the SIR filter

5.1 Advantages

The advantages of particle filtering include its great simplicity, substantial modularity, generality (no linearity, no Gaussianity needed). These are discussed in the referenced literature, so it is not necessary to dwell on this in the critical review here.

5.2 Disadvantages

In the literature, two main drawbacks of the SIR filter were indicated.

First, an "impoverishment" of the particle swarm is often observed. This effect arises if the predictive density appears to be substantially dissimilar to the likelihood. If this happens, most of the "forecast" particles are rejected, leaving only a small number of heavily weighted "survivors". As a consequence, the diversity of the swarm of particles becomes too weak, the point-mass approximation too poor, the filter ceases to follow the truth (degeneracy of the filter). The above dissimilarity of the predictive density and the likelihood may be a result of either very small model error or very small observational error.

Second, if the posterior distribution has heavy tails (outliers happen), then a very large ensemble size is needed in order to represent these distant areas in phase space. This is because their probability mass, p_{out}, is so small that, in order to obtain a single "representer" of these areas (provided we consider these outliers meteorologically important, say, indicating possible severe weather events), the ensemble size should satisfy the inequality $N \cdot p_{out} > 1$, so $N_{min} = p_{out}^{-1}$, which can be large. It becomes even larger if we wish to have several "representers" in these remote areas in phase space. If, moreover, a *difference* between various outlying areas is of interest to us (which is certainly the case for meso-scale modelling and filtering), p_{out} is divided into several even smaller probabilities, thus making N_{min} further larger.

6 Extensions to the SIR filter

A number of improvements and extensions to the basic bootstrap (SIR) filter have been proposed in order to mitigate the above problems of the basic scheme.

To overcome the "impoverishment" problem Gordon et al. (1993) proposed the technique called jittering. With this technique, one adds small random noise to the posterior sample, thus making coincident particles differ from each other. This introduces the desired diversity in the set ("swarm") of particles.

Another technique (Rubin, 1987), called "particle boosting" consists in multiplying, at the prediction step, the number of particle candidates $N_{+} = mN$, where m is an integer. Then, one randomly takes N particles from the set of mN candidates. This also increases the diversity, as desired. But this technique greatly increases the computational cost if the forecast step is expensive (as is the case in the atmospheric data assimilation problem).

Musso et al. (2001) proposed "regularized particle filters", in which the point-mass approximation is replaced by the kernel density approximation, see also the "smooth bootstrap" technique (Stavropoulos and Titterington, 2001). Again, the forecast particles become more different as a result of kernel smoothing, which can prevent the particles' swarm from "impoverishment" and even degeneracy.

Xiong et al. (2006), similarly, proposed to use a posterior Gaussian resampling technique in order to introduce variability in the ensemble sample (see below).

Liu et al. (2001) discuss reweighting as a means to control the balance between "diversity" and "focus" in the posterior swarm of particles. As a generic choice, they recommend to use

$$\pi_k(j) = \sqrt{w_k(j)} \quad (17)$$

as the probabilities of the particles $X_k(j)$. But then, in the prior discrete distribution, Eq. (13), $1/N$ should be replaced by $(1/N) \cdot w_k(j)/\pi_k(j)$. Using the square root in the definition of the probabilities $\pi_k(j)$, allows us to tolerate, to some extent, low-weight particles (outliers), which enhances diversity. Other than the square root choices are also possible.

So, even if the weights happen to differ considerably one from another, we may sample from a more uniform discrete distribution. This means that for a heavily weighted particle, we may not launch too many forecasts from the same (or similar) initial conditions (which looks as a wasting of computer time, even though model-error perturbations can make these forecasts slightly diverge). Instead, we may just remember that this particle has high weight and use this knowledge in the subsequent analysis. Similarly, we may let "light" particles "live", thus preventing the particles' swarm from collapsing.

Pitt and Shephard (1999) proposed to use a more complex sampling scheme, in which the so-called proposal distribution (importance distribution) for importance sampling from the posterior distribution becomes *dependent on current observations* (in the basic SIR formulation, the proposal distribution is the approximate prior distribution). Introducing current observations into the proposal density can, obviously, decrease the above dissimilarity between $p(X_k|Y_{:k-1})$ and $p(Y_k|X_k)$, thus alleviating the "impoverishment" problem. As they report, using the proposal distribution dependent on current observations significantly increases the efficiency of the sampling scheme (less very-low-probability particles). But their "auxiliary filter" requires multiple *evaluations* of the proposal density and the time transition density, which would be very time consuming in the atmospheric data assimilation context.

A simpler technique to use current observations with the aim to appropriately change the importance sampling distribution is "prior editing" proposed by Gordon et al. (1993). With this technique, one enhances the ensemble size at the prediction step and removes the particles with small likelihoods at the current time step. But, similarly to "particle boosting" (see above), this technique is computationally inefficient in our case in view of the high cost of the forecast.

Another technique to skip hopelessly bad (not supported by observations) particles was developed by Van Leeuwen (2003), see below.

Djuric (2001) proposed to combine the well-known from the seventies (Anderson and Moore, 1979) Gaussian sum filter with the SIR methodology. Such a combination can cope with the outliers problem and the "impoverishment" problem of the SIR technique.

More on different flavors of particle filtering can be found in Farnhead (1998).

The above brief review of the techniques developed to improve the basic SIR formulation demonstrates that there are a number of ways to make the SIR filter more efficient. However, Stavropoulos and Titterington (2001) note that *"in general the more sophisticated methods did not lead to noticeable improvement over the simpler versions of particle filtering"*.

7 Non-meteorological and non-oceanographic applications of particle filtering

In the last 15 years, a substantial literature arose devoted to applications of particle filtering. The main areas of applications include target tracking, navigation, positioning, computer vision, signal processing, financial time series modelling, bio-medical applications etc. The common point of all these papers is the low dimensionality of the dynamical systems under investigation. A few examples follow.

Marss (2001) reported on a successful implementation of a particle filter for a 3-D system. Koller and Lerner (2001) worked on a system with 33,000 discrete states. If we assume that in Euclidean space, about 30 points are enough to represent the function's behavior along each axis, we find that 33,000 points roughly correspond to the 3-D case. McGinnity and Irwin (2001) worked with systems whose dimensionality ranged from 1 to 4. It was not possible to find any high-dimensional (more than 10-D) particle filter applications in the statistical literature among tens of available papers.

For such low-dimensional systems, it is of interest for us to look at the ensemble sizes in use. About 40 papers have been checked and those with practical or near-practical applications have been selected. Then, the ensemble sizes reported by the authors has been indicated. The resulting distribution follows.

50

100

200, 200, 200

400

500, 500, 500

1000, 1000, 1000, 1000, 1000

3000

4000, 4000

5000, 5000, 5000

8000

9000

10000, 10000

50000, 50000

100000

50KF

The median of this distribution is about 3000, which is not encouraging for us. The $N = 50$ case corresponds to a simple 2-D mobile robot localization problem. 50KF means 50 Kalman filters in a mixture Kalman filtering scheme.

A personal interpretation of these results is that the particle filtering technique works if we can really *approximate* (in the weak sense) the true distribution by a point-mass distribution. In our effectively high-dimensional case this is certainly not possible.

8 Applications of particle filtering in meteorology and oceanography

8.1 A meteorological review

In the meteorological literature, the use of particle filters was proposed by Anderson and Anderson (1999), Pham (2001), Kim et al. (2003), Xiong at al. (2006), Eyink and Kim (2006), and Nakano et al. (2007).

Anderson and Anderson (1999) used a kernel smoothing (of the prior density) in a Gaussian sum filter without resampling. They used the three-variable Lorenz model and the ensemble size $N = 40$. They also showed that for a 3000-D global barotropic model in an idealized setting, their kernel filter with $N = 40$ was capable of doing the assimilation job. However, it is difficult to judge the relative merit of their filter as compared with other, more traditional approaches. Anderson and Anderson (1999) argue that a small ensemble size may be sufficient if the system "lives" on a low-dimensional attractor. But on the meso-scale, we hardly can assume that the attractor (if it exists) is low-dimensional — in view of the highly variable and diverse atmospheric behaviour on the meso scale. Besides, at every assimilation time step, observations will inevitably push the system off the attractor.

Pham (2001) proposed to use, in the meteorological and oceanographic context, a SIR filter with kernel smoothing of the posterior density. He applied the filter for the three-variable (just three degrees of freedom) Lorenz system and found that for good performance, 50 particles were needed. This is not a promising result for us because the dimensionality of our phase space is about one million.

Kim et al. (2003) used the SIS (sequential importance sampling) and SIR filters for a one-dimensional non-linear stochastic differential equation. They checked the ensemble sizes of 100 and 10^4 and found that 10^4 particles were needed for SIS and 100 for SIR. Again, this is not a good result for us, in view of the extremely low dimensionality of their phase space, $n = 1$. They also proposed a parametric particle filter, in which the prior density is approximated by a density from the exponential family. In their experiments, they found that this filter performed better than SIR (nothing is said about SIS). The parametrization idea seems fruitful, but in the implementation of their analysis step, only two free parameters of the parametric density were updated, which is not acceptable for us, since we wish to control, in the analysis, many meso-scale features across the domain of interest.

Xiong et al. (2006) proposed to use a posterior Gaussian resampling technique in order to fight "impoverishment" of the ensemble sample. This method is not useful for us because it is the non-Gaussianity that has drawn our attention to particle filtering. Besides, in other respects (especially in the analysis scheme), the technique of Xiong et al. (2006) does not differ from the basic SIR formulation. For the three-variable Lorenz system, the ensemble size in their numerical experiments ranged from 100 to 1000. Again, this is not encouraging for us, because we have to build a hyper-dimensional meso-scale data assimilation system.

Eyink and Kim (2006) proposed several methods based on the maximum entropy principle. The methods were checked on low-dimensional systems, but the authors claimed that they were preparing a paper in which they would report on a successful application of these methods for high-dimensional oceanic thermohaline circulation assimilation.

A device to fight the "impoverishment" problem was proposed by Nakano et al. (2007). Their suggestion is not to kill low-weight particles (thus, reducing the desirable diversity) but to use them in a particle merging scheme. The merging is a linear combination (with the weights independent of the physical-space coordinates) selected to preserve the first two moments of the multivariate posterior distribution. They tested their merging technique in experiments with the three-variable Lorenz model and the 40-variable Lorenz-96 model. While the merging technique itself does not seem to lead to vital improvements in the meso-scale data assimilation context, their results are interesting for us in the ensemble sizes they used. Namely, for the 40-D case, even as many as 260,000 members were not enough for the performance of their filter to converge (to stop improving). This result, again, indicates that in our 10^{-6}-D case, the ensemble size necessary for a particle filter to properly work will be enormous, thus prohibiting the use of the particle filtering idea for atmospheric data assimilation.

8.2 Work of P. J. Van Leeuwen

In the oceanographic context, Van Leeuwen (2003) tested three formulations of particle filters: first he used the standard SIR (with a slight modification of sampling from the discrete probability distribution), second he proposed the local SIR (LSIR) and third, the guided SIR (GSIR).

He noticed that the SIR analysis is essentially global, so that on one domain's sub-area, one subset of particles appears to be consistent with the observations, whereas in other sub-areas, quite different particles can be supported by observations. In order to let the observations influence not just the global weights but local neighborhoods, he proposed to make the weights spatially variable by performing the SIR analysis step many times for every grid point with local data selection, like in optimum interpolation (OI). This can introduce the necessary locality of the observational usage, but the great problem arises: what should be the posterior particles if at different grid points different particles were selected. Van Leeuwen (2003) used a simple gluing of the particles, which was not successful in his experiments and does not seem to be a suitable approach in meso-scale assimilation for the two following reasons.

1. From a theoretical point of view, this device violates the particle representation, so that we, seemingly, cannot justify how to *change* ("break") the particles at the analysis step. How can we find the posterior distribution if the prior particles are completely destroyed by "gluing" ? If "gluing" is an approximation, a question arises: is there a *proof* that the error of this approximation tends, in some sense, to zero as the ensemble size tends to infinity ? Without mathematically-based answers to these questions, the technique becomes largely *ad hoc*, which should be considered as a serious drawback.

2. From a practical point of view, "gluing", or equivalently, building the new posterior particles of patches from different ensemble members (presumably, as many patches as many different meso-scale features are present across the analysis domain) violates the dynamical and physical balances in the new particles. This will certainly lead to generation of spurious gravity waves (as noted by the author of the technique) and also to spurious fluxes of heat, water, and momentum, giving rise to the well-known "spin-up" phenomenon, which is highly undesirable in the data assimilation context because a *short-* or *very-short-range* forecast is used at the analysis step as a source of a priori information. In addition, like in OI, local data selection will lead to small-scale noise, which can mask important meteorological details on the meso scale.

So, in the present form, the LSIR technique seems to be not suitable for meso-scale data assimilation.

The GSIR technique of Van Leeuwen (2003) aims to reduce the ensemble size by removing "bad" particles at the very early stages of the forecast. This is possible because we can start the ensemble forecast step *after* we receive current observations (although, at the price of some time delay in issuing the analysis, which may be not desirable in the operation context). For meso-scale data assimilation, the GSIR idea can be useful, among others, but GSIR remains a *global* technique (the weights do not vary across the domain), which prohibits its direct use. The reason for this, as explained below, is that in the global ensemble, the number of global particles should be comparable with the number of effective degrees of freedom we wish to control in the analysis. In the meso-scale, the latter can be very large as compared with the affordable ensemble size.

As for the numerical experiments of Van Leeuwen (2003), it is worth noting that his assimilation/forecast results were not compared with observations. There was also no comparison with other assimilation methods. Besides, in his experiments the assimilated fields look smoother than atmospheric meso-scale patterns.

Concerning the applicability of the results obtained for *oceanic* data assimilation to atmospheric problems, it should be mentioned that the atmosphere is much better observed than the ocean. This implies that oceanic forecasts are relatively much worse than the atmospheric ones. As a result of this, oceanic data assimilation works in a rather different "regime" than the atmospheric one, which reduces the applicability of oceanic results to atmospheric data assimilation, in particular, on the meso-scale.

9 Summary of the disadvantages of a hypothetical particle-filtering based meso-scale data assimilation scheme

The main problem: critically insufficient ensemble size to resolve meso-scale details of meteorological fields and to allow efficient observational usage.

1. Mathematically, we cannot even imagine to approximate a million-dimensional density by a handful of phase-space points (particles). This is completely "unmathematical". Even if we take into account that not all of these numerous degrees of freedom are equally important and the effective number of degrees of freedom, n_{eff} is substantially less, n_{eff} remains, as it follows from physical arguments for meso-scale modelling, orders of magnitude larger than any practically imaginable ensemble size.

2. Meteorologically, an inevitably small ensemble size does not allow us to *resolve* small-scale meso-scale details, as discussed below in this section.

3. Practically, a particle filtering based analysis scheme would imply **critically inefficient use of observations** (as compared to OI and Var).

   In a particle-filter atmospheric data assimilation scheme, tens of thousands, and very soon, millions of observations can influence the analysis only through a few (30-50) free parameters, the particle weights (a "bottleneck"). So, information contained in a huge number of observations, before correcting the forecast fields, should be "summarized" in an extremely small number of particle weights. Obviously, such a tremendous "compression" of observational data is impossible without an immense loss of information. But if we waste valuable observational information, we cannot hope to design an efficient assimilation system.

   In other words, in the SIR filter, local corrections due to local observation-minus-forecast differences are virtually impossible because only the *combined* effect of *all* observations does influence the analysis (through the particle weights), see Eq. (16).

The most severe consequence of the small ensemble size is that the number of degrees of freedom of the meso-scale field we wish to resolve appears to be orders of magnitude larger than the number of adjustable parameters (the particle weights) at the analysis step. Indeed, if we intend, in the analysis, to resolve meteorological details of the size of kilometers, then we have to design an analysis scheme capable of controlling the analysis fields with this granularity. In other words, if in the analysis we admit a feature (say, a cloud system) of the size of 10 km, and observations indicate that we have to correct the forecast in this 10-km area, then the analysis technique should allow this. And this job has to be done simultaneously, in other 10-km "boxes", independently of each other. In a 1000 by 1000 km domain there are 10,000 horizontal "boxes". In the vertical, similarly, if we aim at details, say, 500-m deep, then 10 to 20 vertical "boxes" have to be resolved. Thus, in total, we obtain $10,000 \cdot 20 = 200,000$ 3-D "boxes" or degrees of freedom. But, in the SIR analysis, we have only N (recall, N is the ensemble size) free parameters, with N being about 30–50. Obviously, it is absolutely impossible to control 200,000 degrees of freedom with such a tiny number of adjustable parameters.

Otherwise stated, the SIR analysis equation, Eq. (16), can be viewed as an expansion of the analysis field in the basis functions $X_k^-(j)$, the particles' fields (a kind of *spectral* technique). But the number of the basis functions, N, is so incredibly small as compared to the number of meso-scale details to be resolved, that we have to conclude that the SIR filter is inappropriate with a reasonable (for us) ensemble size. With $N = 50$, only about 50 details over the whole 3-D analysis domain can be resolved, which is far too small for a state-of-the-art and a prospective meso-scale data assimilation system.

The LSIR Ansatz of Van Leeuwen (2003) could be a remedy but it seems to lack mathematical and meteorological justification (see above).

In addition to the above small-ensemble-size issues, the following practical points are worth noting.

1. In 3D and 4D-Var analysis schemes, the very important role of the prior (background) probability distribution (covariances) is well established. To obtain a good analysis, very sophisticated background-error covariance models have been proposed. In particular, control of multivariate covariances, non-separability (vertical by horizontal) of 3-D correlations and other aspects are shown to be of great importance for a successful background-error covariance model. With a particle filter, we do not care about these issues at all, completely relying on the (small) swarm of particles. But particles are selected, to a large extent, at random, so the implied probability distribution is also random (significantly contaminated by sampling noise). Is it good to use this random prior distribution as if it were the true one?

2. As noted by Anderson and Anderson (1999), the SIR filter may fail if the forecast model is *biased*. In this case, all particles can be biased, so that the observations will be not capable of removing this bias at all (because the observation has only "the right" to select from prior particles, which are all biased).

3. The SIR analysis algorithm is so incredibly simple that we may question: why the substantial *simplification* of the analysis technique (from the complicated 3D-Var to the extremely simple SIR) will lead to an improvement? Common practice shows that the converse is almost always the case.

10 Conclusions

The SIR filter is *not* recommended for meso-scale data assimilation because it would imply, for an affordable ensemble size, first, critically inefficient observational usage and second, lack of high-resolution meso-scale fields' analysis corrections.

The LSIR filter is also *not* recommended for meso-scale data assimilation in its present form because of lack of its mathematical and meteorological justification.

11 Acknowledgements

This study has been supported by the Russian Foundation for Basic Research under grant 06-05-08076. I also appreciated valuable discussions with D.B. Kiktev.

12 References

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

Andersson, P., and J.B. Moore, 1979: Optimal Filtering, Prentice-Hall, New Jersey.

Djuric, P.M., 2001: Sequential estimation of signals under model, in "Sequential Monte Carlo Methods in Practice", Eds. A. Doucet, N. Freitas, and N. Gordon.", Springer, 381-400.

Efron, B., and R.J. Tibshirani, 1993: An Introduction to the Bootstrap, Chapman and Hall, London.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics, J. Geophys. Res., 99 (C5), 10143-10162.

Eyink, G.L., and S. Kim., 2006: A maximum entropy method for particle filtering, J. Statist. Phys., 123 (5), 1071-1128.

Farnhead, P., 1998: Sequential Monte Carlo Methods in filter theory, Ph.D. Thesis, Oxford Univ., http://www.maths.lancs.ac.uk/ fearnhea/publications/.

Gordon, N.J., D.J. Salmond, and A.F.M. Smith, 1993: Novel approach to nonlinear/non-Gaussian Bayesian state estimation, IEEE Proceedings-F, 140(2), 107-113.

Houtekamer, P.L., and H.L. Mitchell, 2005: Ensemble Kalman filtering, Quart. J. Roy. Meteorol. Soc., 131, 3269-3289.

Kim., S., G.L. Eyink, J.M. Restrepo, F.J. Alexander, and G. Johnson, 2003: Ensemble filtering for nonlinear dynamics, Mon. Wea. Rev., 131, 2586-2594.

Koller, D., and U. Lerner, 2001: Sampling in factored dynamic system, in "Sequential Monte Carlo Methods in Practice", Eds. A. Doucet, N. Freitas, and N. Gordon.", Springer, 445-464.

Liu, J.S., R. Chen, and T. Logvinenko, 2001: A theoretical framework for sequential importance sampling with resampling, in "Sequential Monte Carlo Methods in Practice", Eds. A. Doucet, N. Freitas, and N. Gordon.", Springer, 225-246.

Marrs, A.D., 2001: In-Situ ellipsometry solution using sequential Monte Carlo, in "Sequential Monte Carlo Methods in Practice", Eds. A. Doucet, N. Freitas, and N. Gordon.", Springer, 465-478.

McGinnity, S., and G.W. Irwin, 2001: Manoeuvring target tracking using a multiple-model bootstrap filter, in "Sequential Monte Carlo Methods in Practice", Eds. A. Doucet, N. Freitas, and N. Gordon.", Springer, 479-498.

Musso, C., N. Oudjane, and F. Le Gland, 2001: Improving regularized particle filters, in "Sequential Monte Carlo Methods in Practice", Eds. A. Doucet, N. Freitas, and N. Gordon.", Springer, 247-272.

Nakano, S., G. Ueno, and T. Higuchi, 2007: Merging particle filter for sequential data assimilation, Nonlin. Proc. Geophys., 14, 395-408.

Pham, D.T., 2001: Stochastic methods for sequential data assimilation in strongly nonlinear systems, Mon. Wea. Rev., 129, 1194-1207.

Pitt, M.K., and N. Shephard, 1999: Filtering via simulation: Auxiliary particle filters, J. Amer. Statist. Assoc., 94(446), 590-599.

Rubin, D.B., 1987: Comment on 'The calculation of posterior distributions by data augmentation', J. Amer. Statist. Ass., 82, 543-546.

Stavropoulos, P., and D.M. Titterington, 2001: Improved particle filters and smoothing, in "Sequential Monte Carlo Methods in Practice", Eds. A. Doucet, N. Freitas, and N. Gordon.", Springer, 295-318.

Van Leeuwen, P.J., 2003: Nonlinear ensemble data assimilation for the Ocean, ECMWF Seminar "Recent developments in data assimilation for atmosphere and ocean".

Xiong, X., I.M. Navon, and B. Uzunoglu, 2006: A note on the particle filter with posterior Gaussian resampling, Tellus, 58A, 456-460.

List of COSMO Newsletters and Technical Reports

(available for download from the COSMO Website: www.cosmo-model.org)

COSMO Newsletters

No. 1: February 2001.

No. 2: February 2002.

No. 3: February 2003.

No. 4: February 2004.

No. 5: April 2005.

No. 6: July 2006; Proceedings from the COSMO General Meeting 2005.

No. 7: July 2007; Proceedings from the COSMO General Meeting 2006.

COSMO Technical Reports

No. 1: Dmitrii Mironov and Matthias Raschendorfer (2001):
Evaluation of Empirical Parameters of the New LM Surface-Layer Parameterization Scheme. Results from Numerical Experiments Including the Soil Moisture Analysis.

No. 2: Reinhold Schrodin and Erdmann Heise (2001):
The Multi-Layer Version of the DWD Soil Model TERRA_LM.

No. 3: Günther Doms (2001):
A Scheme for Monotonic Numerical Diffusion in the LM.

No. 4: Hans-Joachim Herzog, Ursula Schubert, Gerd Vogel, Adelheid Fiedler and Roswitha Kirchner (2002):
LLM ‾ the High-Resolving Nonhydrostatic Simulation Model in the DWD-Project LITFASS.
Part I: Modelling Technique and Simulation Method.

No. 5: Jean-Marie Bettems (2002):
EUCOS Impact Study Using the Limited-Area Non-Hydrostatic NWP Model in Operational Use at MeteoSwiss.

No. 6: Heinz-Werner Bitzer and Jürgen Steppeler (2004):
Documentation of the Z-Coordinate Dynamical Core of LM.

No. 7: Hans-Joachim Herzog, Almut Gassmann (2005):
Lorenz- and Charney-Phillips vertical grid experimentation using a compressible non-hydrostatic toy-model relevant to the fast-mode part of the 'Lokal-Modell'

No. 8: Chiara Marsigli, Andrea Montani, Tiziana Paccagnella, Davide Sacchetti, André Walser, Marco Arpagaus, Thomas Schumann (2005):
Evaluation of the Performance of the COSMO-LEPS System

No. 9: Erdmann Heise, Bodo Ritter, Reinhold Schrodin (2006):
Operational Implementation of the Multilayer Soil Model

No. 10: M.D. Tsyrulnikov (2007):
Is the particle filtering approach appropriate for meso-scale data assimilation ?

COSMO Technical Reports

Issues of the COSMO Technical Reports series are published by the *COnsortium for Small-scale MOdelling* at non-regular intervals. COSMO is a European group for numerical weather prediction with participating meteorological services from Germany (DWD, AWGeophys), Greece (HNMS), Italy (USAM, ARPA-SIM, ARPA Piemonte), Switzerland (MeteoSwiss), Poland (IMGW) and Romania (NMA). The general goal is to develop, improve and maintain a non-hydrostatic limited area modelling system to be used for both operational and research applications by the members of COSMO. This system is initially based on the COSMO-Model (previously known as LM) of DWD with its corresponding data assimilation system.

The Technical Reports are intended

- for scientific contributions and a documentation of research activities,
- to present and discuss results obtained from the model system,
- to present and discuss verification results and interpretation methods,
- for a documentation of technical changes to the model system,
- to give an overview of new components of the model system.

The purpose of these reports is to communicate results, changes and progress related to the LM model system relatively fast within the COSMO consortium, and also to inform other NWP groups on our current research activities. In this way the discussion on a specific topic can be stimulated at an early stage. In order to publish a report very soon after the completion of the manuscript, we have decided to omit a thorough reviewing procedure and only a rough check is done by the editors and a third reviewer. We apologize for typographical and other errors or inconsistencies which may still be present.

At present, the Technical Reports are available for download from the COSMO web site (www.cosmo-model.org). If required, the member meteorological centres can produce hard-copies by their own for distribution within their service. All members of the consortium will be informed about new issues by email.

For any comments and questions, please contact the editors:

Massimo Milelli
Massimo.Milelli@arpa.piemonte.it

Ulrich Schättler
Ulrich.Schaettler@dwd.de