

Sequential parameter estimation for stochastic systems

G. A. Kivman

Alfred Wegener Institute for Polar and Marine Research, Bremerhaven, Germany

Camera-ready Copy for
Nonlinear Processes in Geophysics
Manuscript-No. ???

Sequential parameter estimation for stochastic systems

G. A. Kivman

Alfred Wegener Institute for Polar and Marine Research, Bremerhaven, Germany

Received: ??? – Accepted: ???

Abstract. The quality of the prediction of the dynamical system evolution is determined by the accuracy to which initial conditions and forcing are known. Availability of future observations permits reducing the effects of errors in assessment the external model parameters by means of a filtering algorithm. However, traditional filtering schemes do not take into account uncertainties in specifying the internal model parameters and thus cannot reduce their contribution to the forecast errors. An extension of the Sequential Importance Resampling filter (SIR) is proposed to this aim. The filter is verified against the Ensemble Kalman filter (EnKF) in application to the stochastic Lorenz system. It is shown that the SIR is capable to estimate the system parameters and to predicts the evolution of the system with a remarkably better accuracy than the EnKF. This highlights a severe drawback of any Kalman filtering scheme: due to utilizing only first two statistical moments in the analysis step it is unable to deal with probability density functions badly approximated by the normal distribution.

1 Introduction

Our ability to predict the state of the atmosphere and the ocean highly depends on the accuracy to which initial conditions and forcing functions are known. Various data assimilation techniques have been developed to constrain models with observations in order to reduce the influence of uncertainties in these external model parameters on the forecast skill. In parallel with the forecast errors of this type, there are those caused by uncertainties in internal model parameters. The so-called adjoint method borrowed from the engineering optimal control literature provides a tool to tune the model to available data by adjusting those model parameters. However, this method involves an assumption that the model structure is perfect which is too restrictive for oceanographic and meteorological applications. Accounting for model errors in the parameter estimation problem dramatically increases the dimension of the control space and is affordable

only for low dimensional models (ten Brummelhuis et al., 1993; Eknes and Evensen, 1997, Gong et al., 1998).

Commonly, the physical model is calibrated initially to chose some "optimal" values of the external parameters. Then a scheme of the Kalman filtering is applied for correcting the prediction with available data. At this stage, the model parameters are assumed to be known precisely. However, neglecting uncertainties of this kind leads to overestimating the forecast skill. As a result, the data make a lesser contribution to the analyzed system state and the true trajectory of the system can be quickly lost.

Since only sequential data assimilation is and will be in the nearer future affordable for treating high dimensional transient problems, A question arises of whether it is possible to optimize the internal model parameters \mathbf{m} sequentially. Though they are not dynamical variables, we can easily augment the original dynamical system

$$\frac{d\mathbf{x}}{dt} = \mathbf{M}(t, \mathbf{m}, \mathbf{x}), \quad (1)$$

by the equation

$$\frac{d\mathbf{m}}{dt} = 0. \quad (2)$$

Filtering the extended dynamical system (1), (2) is a challenging task. At first, the system becomes nonlinear even if the original system (1) was linear. At second, the model parameters \mathbf{m} cannot take arbitrary values and usually are bounded below by 0. Consequently, any probability distribution expressing uncertainties in the parameter space is essentially non-Gaussian and even the Gaussian distribution is not absolutely continuous with respect to it. Thus, the system (1), (2) provides a very sensitive test for evaluating of how a filtering scheme deals with nonlinearity and with distributions badly approximated by the normal one.

The aim of this paper is twofold. The first objective is to demonstrate that sequential tuning of the internal parameters is allowable by means of Monte Carlo methods with no additional computational cost. The other goal is to show

that accounting for the whole error statistics leads to notably better forecast. Particle filters provide a tool for that. They are introduced in Sect. 2. Sect. 3 contains an example of filtering the Lorenz system with the well-known Ensemble Kalman filter (EnKF) and with the so-called Sequential Importance Resampling filter (SIR). An extension of the SIR to sequential optimization of the model parameters and comparison of the SIR with an analogous extension of the EnKF is presented in Sect. 4. Sect. 5 contains conclusions.

2 Particle filters

It is well known, that the classical Kalman filter (KF) is optimal in the sense of minimizing the variance only for linear systems and the Gaussian statistics. A linearization of the error covariance evolution used in the Extended Kalman filter (EKF) (Jazwinski, 1973) often turns out to be inadequate. Unbounded growth of the computed error variances due to neglecting nonlinear saturation effects causes the update procedure to become unstable (Evensen, 1992). Miller et al. (1994, 1999) showed poor performance of the EKF in application to the Lorenz system when the data are too sparse or too inaccurate.

To go around the closure problem for the error statistics propagation, Evensen (1994) proposed the Ensemble Kalman filter (EnKF). The heart of the method is Monte Carlo integration of the Fokker-Planck-Kolmogorov (FPK) equation governing the evolution of the probability density function (PDF) that describes the forecast error statistics. In the analysis step, each ensemble member is updated according to the traditional scheme of the KF with the use of the forecast covariance matrix derived from the ensemble statistics. Then, if the data are randomly perturbed, the updated ensemble is shown to have the proper mean and covariance matrix (Burgers et al., 1998).

However, two potential problems for the EnKF are worth to mention. At first, though the EnKF uses the full non-linear dynamics to propagate the forecast error statistics, it mimics the traditional KF in the analysis step and uses only the Gaussian part of the prior PDF. Bennett (1992) pointed out that "one thing is certain: least-squares estimation is very inefficient for highly intermittent processes, ... having probability distributions not well characterized by means and variances". At second, the updated ensemble preserves only two first moments of the posterior. Consequently, the initial condition for the further integration of the FPK does not coincide with the posterior PDF. In the case of a small system noise when the "diffusion" of probabilities is small compared with the "advection", the system does not forget its initial state for a long time and the ensemble becomes non-representative for the forecast error statistics after few analysis steps.

Particle filters provide a tool to solve these problems. As the EnKF, they rely on Monte Carlo integration of the FPK equation. However, instead of updating ensemble members in the analysis step they update their probabilities according to the fitness to observations. Consequently, they do not

involve any model reinitialization that usually injects imbalance in the model state and leads to a shock in the system evolution. Another advantage of these filters is that they make use of the full statistics of the forecast and data errors and thus are truly variance minimizing methods.

The basic algorithm of the particle filtering consists of the following steps:

1. An initial ensemble $\mathbf{x}_n, n = 1, \dots, N$ is drawn from a prior distribution $\rho(\mathbf{x})$.
2. Each ensemble member \mathbf{x}_n evolves according to the dynamical equations.
3. At $t = t_k$ when the data \mathbf{d}_k become available, weights $w_n(t_k)$ expressing "fitness" of ensemble members to the data are computed

$$w_n(t_k) = w_n(t_{k-1})\rho(\mathbf{d}_k|\mathbf{x}_n(t_k)) \quad (3)$$

and normalized so that

$$\sum_{n=1}^N w_n(t_k) = 1.$$

Here $\rho(\mathbf{d}_k|\mathbf{x}_n(t_k))$ expresses the conditional PDF for the data \mathbf{d}_k to be observed when the system state is $\mathbf{x}_n(t_k)$ or, in other words, describes the statistics of the data errors.

4. The final prediction is calculated as the weighted ensemble mean.

This scheme was called as the direct ensemble method in van Leeuwen and Evensen (1996). They found that the vast majority of the ensemble members got negligible weights after few analysis steps and thus only a tiny fraction of the ensemble contributes to the mean. In this case, to obtain a reasonable approximation of the posterior PDF one needed to use an ensemble of about 10^4 members. This drawback of the method is explained by a Kong-Liu-Wong theorem (Kong et al., 1994) which states that the unconditional variance of the importance weights w_n , i.e. with observations treated as random variables, increases in time. Thus, the algorithm becomes more and more unstable.

To go around the degeneracy of the algorithm several approaches have been put forward. One of the most popular schemes is the Sequential Importance Resampling filter (SIR) proposed by Rubin (1988) and applied to filtering the dynamical systems in Gordon et al. (1993). The basic idea of the method is that there is no need in computing further evolution of ensemble members having bad fitness to the data. It is achieved by generating a new ensemble of equally probable members at each analysis step by means of sampling from the old ensemble with replacement. Probabilities for the members to be sampled at $t = t_k$ are assigned to their normalized weights $w_n(t_k)$ calculated by (3) with $w_n(t_{k-1}) = 1$. Smith and Gelfand (1992) have proven that

Table 1. RMS deviation of the filtered solutions with the fixed model parameters from the true trajectory

Filter	EnK	SIR	EnKF	SIR
Ensemble size	250	250	1000	1000
x	7.4	6.5	6.5	5.8
y	8.9	8.1	8.4	7.3
z	8.0	7.8	7.1	6.6

a discrete approximation tends to the posterior PDF when the sample size tends to infinity. This is just opposite to the EnKF which does not produce an approximation to the posterior PDF and preserves only the mean and the covariance.

3 Filtering the Lorenz system

To compare the EnKF and the SIR filter, an identical twin experiment with the Lorenz system

$$\begin{aligned}
 \frac{dx}{dt} &= \gamma(y - x), \\
 \frac{dy}{dt} &= rx - y - xz, \\
 \frac{dz}{dt} &= xy - \beta z,
 \end{aligned} \tag{4}$$

with commonly used parameters $\gamma = 10$, $r = 28$, and $\beta = 8/3$ was performed. Each of Eqns. (4) were perturbed with a white noise having variances 2., 12.13 and 12.13 correspondingly. The reference solution for $t \in [0, 20]$ was computed starting from an initial condition obtained by adding a noise $N(0, 2)$ to the first guess

$$(x_0, y_0, z_0) = (1.508870, -1.531271, 25.46091).$$

The observations for the x - and z - components were generated at each $\delta t = 1$ by adding the $N(0, 2)$ noise to the reference solution.

The experiment design is almost identical to that of Evensen and van Leeuwen (2000). The major differences are the distance between observations which was made twice larger and that only 2 components of the model state were observable. Observability of the whole system state is the case standing far away from that we deal with in meteorological and oceanographic applications. In addition, for more pictorial presentation of results, the assimilation period was chosen to be twice shorter than that in Evensen and van Leeuwen (2000).

Table 1 summarizes results of experiments made with use of the SIR and EnKF for 250 and 1000 ensemble members. As it is seen, performance of the filters improves with increasing the ensemble size. However, this improvement is less than 20% with four-fold increasing the ensemble size and is mostly achieved due to better representation of system oscillations near the attractor within $t \in [7.5, 11.5]$ (compare Fig. 1 and Fig. 2). This reflects the very slow convergence of Monte Carlo methods which is of the order of $O(N^{-1})$, where N is the amount of the ensemble members. Another

important point worth to be mention is that the SIR for the smaller ensemble is almost as effective as the EnKF for the larger ensemble.

The main problem of the SIR with 250 ensemble members is its inability to capture the true trajectory within $t \in [7.5, 11.5]$. The reason for this failure is bad scores of all ensemble members and consequently poor representation of the forecast error statistics. This problem for the SIR is recovered by enlarging the ensemble. One can notice that the EnKF with the smaller ensemble does a better job within this interval. Seemingly, it is a general point. If all ensemble members deviate much from the observations, the EnKF updates the ensemble trajectories and improves their fit to the data, while the SIR changes the ensemble probabilities and do not do anything with fitness of the ensemble members.

The filters for the both ensemble sizes lose the true trajectory at $t=3$ when they assimilate a bad data on the x component coming just after the transition point. Before the analysis step, both filters predict a transition, but the data wrongly tells about the absence of the transition point. The filters accept this information and delay the transition for a while. Then the EnKF predicts the next data at $t=4$ much better than the SIR. However, it propagates the information provided by the data to the unobserved y - component much worse. As a result, the SIR solution for $t \in [4, 5]$ is almost identical to the true trajectory, while the EnKF catches it only at $t=5$. This situation is a stable feature which do not depend on the ensemble size.

There are two more examples of inadequate transmission of the information provided by the data on x and z to the unobserved y . Let us consider the analysis step at $t=19$ for the smaller ensemble. Both filters lost the system trajectory at about $t = 18.5$. With the analysis step, they recover the observed components of the solution. However, the SIR makes a better inference about y and is capable to follow the system trajectory further while the EnKF loses it immediately after the analysis step. The same situation occurs at $t = 7$ for the larger ensemble. In spite large deviations between the forecast and the data, the SIR places the analysis just at the system trajectory. The EnKF updates only the observed components in a proper manner while the y -component is pulled even in a wrong direction.

This does not mean that the EnKF makes the update of the unobserved part of the system state always wrongly when it faces large data misfits. For example, the analysis at $t=12$ was made by the EnKF very precisely. However, keeping in mind that the EnKF does not make use of the whole error statistics one can conclude that the filter has problems in transferring the information from the data to the unobserved part of the system and they cannot be resolved with enlarging the ensemble size.

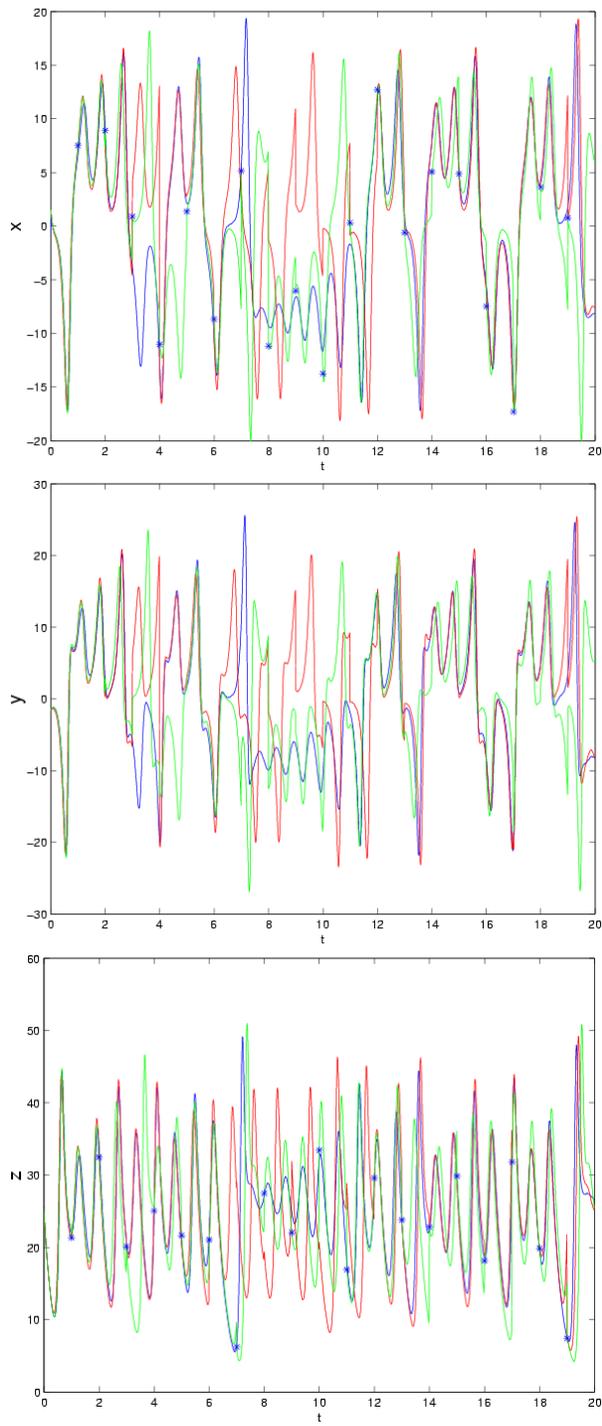


Fig. 1. Filtered solutions for 250 ensemble members with the fixed model parameters: the SIR - red, the EnKF - green. The blue curve is the true trajectory, starts are observations.

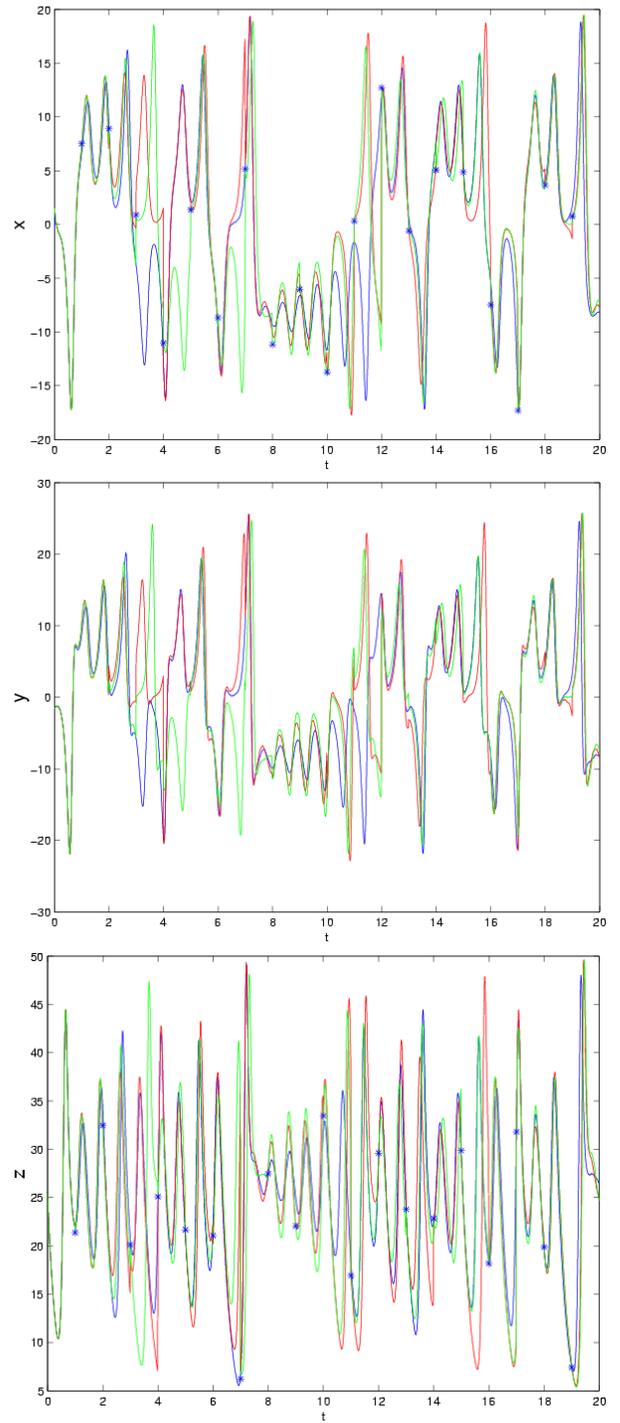


Fig. 2. The same as Fig. 1 for 1000 ensemble members.

Table 2. RMS deviation of the filtered solutions with the adjusted model parameters from the true trajectory

Filter	EnK	SIR	EnKF	SIR
Ensemble size	250	250	1000	1000
x	9.2	7.1	9.7	5.6
y	10.6	8.2	11.0	7.1
z	14.2	7.7	14.7	6.6

4 Sequential combined parameter- and state estimation with the SIR filter

Extension of the SIR and the EnKF for the system (1), (2) is straightforward. To examine their potentialities, the experiment described in Sect. 3 was repeated with the initial ensemble in the parameter space drawn from the homogeneous distribution for $\gamma \in [0, 30.]$ and $r \in [0, 44.8]$.

One can easily verify that the system trajectory is more sensitive to the choice of r . Thus, it is not a surprise that the SIR catches the value of r after the first analysis step (Fig. 4). The next analysis step produces a very good estimate for γ . However, it is impaired with time until the filter collapses in the parameter space. Since the trajectory of the system is tolerant to variations of γ , the quality of the filtered solution (see Table 2) which is of our primary interest still remains almost identical to that presented in Sect. 3.

It is not the case for the EnKF. In the parameter estimation problem, the EnKF has to deal with the PDFs badly approximated by the Gaussian distribution. Such a modification in the problem formulation completely corrupts the filter performance. The solution produced by the EnKF has very little common with the true trajectory (see Fig. 3) and its quality does not depend on the ensemble size (see Table 2).

The failure of the EnKF is caused by its incapability to recover the true values of the model parameters (see Fig. 4). The evolution of estimates for γ resembles a random walk over the parameter subspace. As we can expect and it was the case for the SIR, that this parameter is derived from the data with a lower accuracy compared to that for r . It is astonishing that the EnKF does even a much worse job when estimating the more crucial parameter r . After removing fluctuations from the corresponding curve presented in Fig. 4, one can easily see a pronounced tendency of pushing the estimates of r in the opposite direction with respect to the true value. It is worth noting that the parameter estimates obtained with the EnKF for the both ensembles are indistinguishable. These point out that the EnKF is completely unable to deal with the PDFs defined over constrained sets. An additional evidence for this conclusion is that the EnKF was permanently trying to produce negative analyzed parameter values for some ensemble members and the condition of non-negativeness had to be imposed by force.

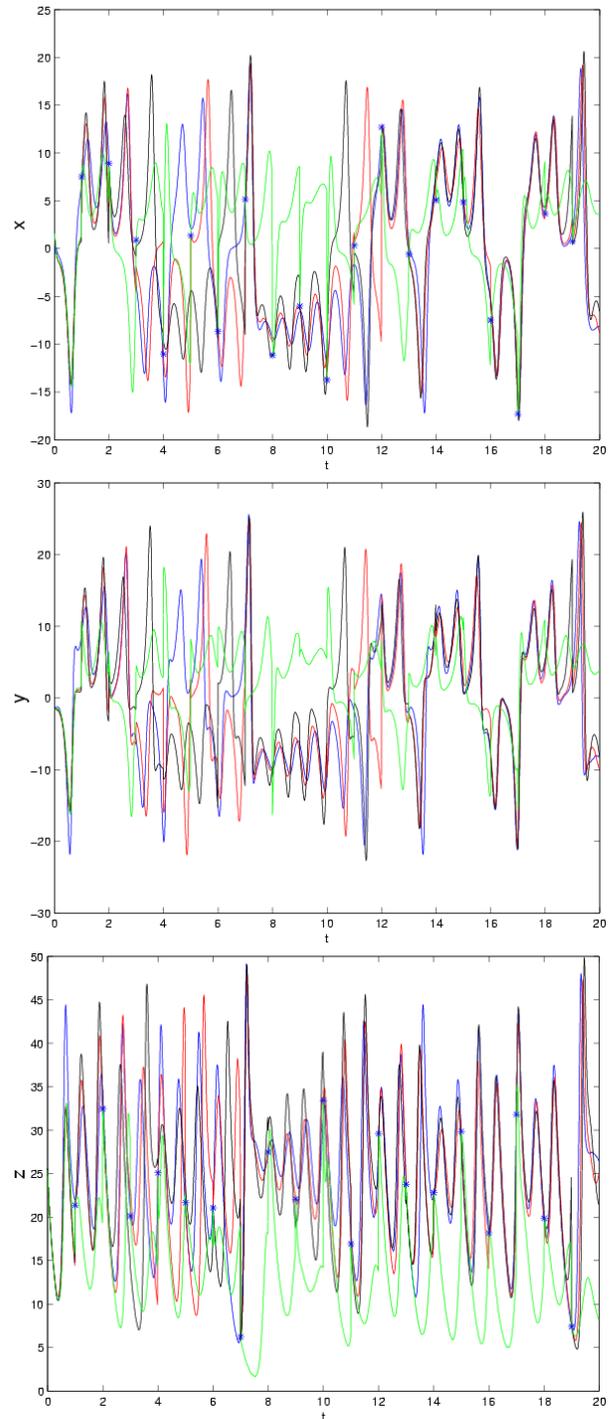


Fig. 3. Filtered solutions with the adjusted model parameters: the SIR for 1000 ensemble members (red) and for 250 ensemble members (black), the EnKF for 1000 ensemble members (green). The blue curve is the true trajectory, starts are observations.

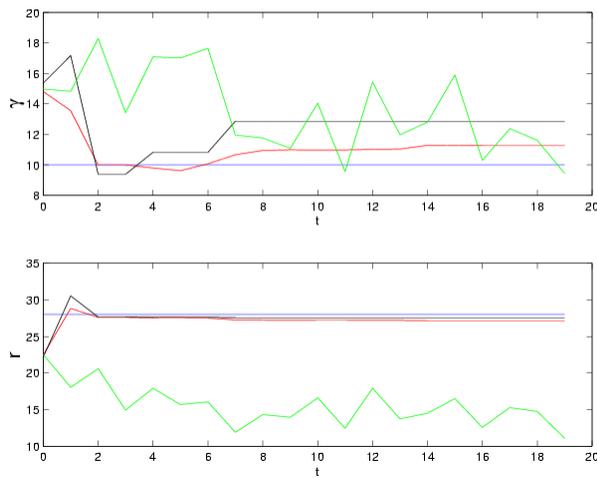


Fig. 4. Evolution of the parameter estimates: the SIR for 1000 ensemble members (red) and for 250 ensemble members (black), the EnKF (green), and the true parameter (blue)

5 Conclusions

Data assimilation for high-dimensional nonlinear ocean and atmospheric models is a challenging task. On the one side, high-dimensionality forces us to use simplified representation of the error statistics and thus to neglect some sources of uncertainties. Developing approaches for low dimensional representation is an active area of research. On the other side, nonlinearity of the models raises a fundamental question of the potentiality of using generalizations of the traditional Kalman filter which is optimal only for the linear systems and for the Gaussian statistics is studied much less.

In a recent paper Verlaan and Heemink (2001) demonstrated better performance the EnKF compared to the EKF for the Lorenz system. That was the reason why the SIR was confronted here only with the EnKF. It was clear *a priori* that the EnKF is subject to two problems. One of them is common for all Kalman filtering schemes in application to non-Gaussian distributions: they do not produce the variance-minimizing estimate in the analysis step. In addition, though the EnKF propagates the error statistics more accurately than the EKF, it initializes the FPK equation with an ensemble that preserves only first two moments of the true analysis error statistics. The SIR is free from these drawbacks and, as the results presented in Sect. 3 reveal, recovers the trajectory of the stochastic Lorenz system with the higher accuracy.

Superiority of the SIR to the EnKF becomes more pronounced when they are applied to the parameter estimation problem. The SIR adjusted the most crucial model parameter very closely to its true value after a couple of analysis steps and the quality of the solution appeared to be independent on whether the model parameters are initially known exactly or with some uncertainty. There is a distinguishing feature of the case considered in Sect. 4 in comparison with numerous applications of the the nonlinear Kalman filters where they showed high skills. Namely, the EnKF faced here a distribution badly approximated by the Gaussian curve. In this

situation it demonstrated total inability to cope with the problem. This point can be of high importance for atmospheric and ocean data assimilation where many state variables such as tracer fields are distributed similarly. In this situation, the Gaussian approximation to the error statistics utilized in the Kalman filter yields totally wrong transmission of the information from the observed variables to the unobserved ones. This failure cannot be avoid by increasing the ensemble size since no convergence exists.

The SIR makes the analysis computationally simpler and, due to utilizing the whole error statistics, much more accurate than the Kalman filter. In addition, it offers more flexibility allowing one to tune poorly known model parameters and easily to consider observations having non-Gaussian error statistics (as it is the case for the tracer fields) and nonlinearly related to the state variables. The main problem of the method is that the solution becomes unstable when the most part of the ensemble members have bad fitness to the data due to undersampling. As it was noticed in Sect. 3, the EnKF performs better in this situation. This weakness of the SIR can be especially pronounced in the parameter estimation problem when all but one of the members die at the resampling step while the lack of the noise in Eqn. (2) prevents the ensemble from regeneration in the parameter space with time. This problem will be studied in the future. A possible solution could be adding a noise to the ensemble if it is nearly to collapse. This procedure makes it possible to restore the ensemble size and even to detect regular temporal oscillations of some model parameters (Losa et al. , 2001).

Acknowledgements. The author thanks his colleagues Jens Schröter, Manfred Wenzel and Svetlana Loza for continuing support and discussion.

References

- Bennett, A. F., *Inverse Methods in Physical Oceanography*. Cambridge Univ. Press, Cambridge, 1992.
- Burgers, G., van Leeuwen, P. J., and Evensen, G., Analysis scheme in the Ensemble Kalman filter. *Mon. Weath. Rev.* 126,1719–1724, 1998
- Gong, J., Wahba, G., Johnson, D. R. and Tribbia, J., Adaptive tuning of numerical weather prediction models: simultaneous estimation of weighting, smoothing and physical parameters. *Mon. Wea. Rev.* 125, 210–231, 1998.
- Gordon, N. J., Salmond, D. J., and Smith, A. F. M., Novel approach to nonlinear/ non-Gaussian Bayesian state estimation. *IEE-Proceedings-F* 140, 107–113, 1993.
- Eknes, M. and Evensen, G., Parameter estimation solving a weak constraint variational formulation for an Ekman model. *J. Geophys. Res.* 102, 12479–12491, 1997.
- Evensen, G., Using the Extended Kalman filter with a multilayer quasi-geostrophic ocean model. *J. Geophys. Res.* 97, 17905–17924, 1992.
- Evensen, G., Sequential data assimilation with a non-linear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics. *J. Geophys. Res.* 99, 10143–10162, 1994.
- Evensen, G. and van Leeuwen, P. J., An Ensemble Kalman Smoother for nonlinear dynamics. *Mon. Weather Res.* 128, 1852–1867, 2000.
- Jazwinski, A. H., *Stochastic Processes and Filtering Theory*. Academic Press, 1973.
- Kong, A., Liu, J. S., and Wong, W. H., Sequential imputations and Bayesian missing data problems. *J. American Stat. Association* 89, 278–288, 1994.

- Losa, S. N., Kivman, G. A., Schröter, J., and Wenzel, M., Sequential weak constraint parameter estimation in an ecosystem model. *J. Mar. Systems* submitted, 2001.
- Miller, R. N., Ghil, M. and Ghautiez, F., Advanced data assimilation in strongly nonlinear dynamical system. *J. Atmos. Sci* 51, 1037–1055, 1994.
- Miller, R. N., Carter, E., and Blue, S., Data assimilation into nonlinear stochastic models. *Tellus* 51, 167–194, 1999.
- Rubin, D. B., Using the SIR algorithm to simulate posterior distributions. in *Bayesian Statistics 3* (Eds J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith), Oxford University Press, 395–402, 1988.
- Smith, A. F. M. and Gelfand, A. E., Bayesian statistics without tears: A sampling-resampling perspective. *The American Statistician* 46, 84–88, 1992.
- ten Brummelhuis, P. G. J., Heemink, A. W. and van den Boogaard, H. F. P., Identification of shallow sea models. *Int. J. Numer. Method Fluids* 17, 637–665, 1993.
- van Leeuwen, P. J. and Evensen, G., Data assimilation and inverse methods in terms of a probabilistic formulation. *Mon. Weath. Rev.* 124, 2898–2913, 1996.
- Verlaan, M. and Heemink, A. W., Nonlinearity in data assimilation applications: a practical method for analysis. *Mon. Weath. Rev.* 129, 1578–1589, 2001.