PARTICAL FILTERING IN STRONLY NON-LINEAR ECOSYSTEM MODELLING

S. Losa, J. Schröter and M. Wenzel
Svetlana.Losa@awi.de

Alfred Wegener Institute for Polar and Marine Research

Abstract

We investigate advantages and specific difficulties in implementation of particle filters when assimilating data into strongly non-linear and non-Gaussian systems. With respect to the last circumstance, a Sequential Importance Resampling (SIR) filter seems to be of a great advantage since the filter updates probability of the particles (according to their agreement with the assimilated data) and thus, allows one to **use** the **full forecast** and **data errors statistics**. Still the most challenging thing in such a filtering is sampling the particles which approximate the continuous probability density function (pdf) evolved according to a stochastic dynamical model. We consider different sampling strategies which depend on our prior knowledge of the system and determine the filter performance cost and quality. All the experiments have been carried out with an ocean biogeochemical model which indeed is an example of a non-Gaussian system.

Data assimilation for stochastic systems

(Prediction)

Dynamical model $\frac{\partial x}{\partial t} = M(p, x, t) + F(t) + \varepsilon$

Predicting, one is uncertain in :
Initial condition $x(t_0)$, model parameters p, external forcing F

Forecast - a PDF - $\rho_f(x(t), p)$ defined on a $X_f \times P$ space, $x(t) \in X_f$, $p \in P$

$$\rho_f^f(x(t), p) = C\rho^f(x(t) \mid x(0), p)\rho(p)\rho_0(x(0))$$

Global definition of PDF over the $X \times P$ space, here $X = \oplus X_f$ is a space of model trajectories:

$$\rho(x, p) = C\rho(p)\rho_0(x(0))\prod_{k=1}^{M}\rho^f(x(k\delta t) \mid x((k-1)\delta t, p)$$

M is a number of time step δt.
Improving the forecast with data assimilation globally – problem of Mdim(Xt)dim(P) dimension

(Analysis)

Improving the forecast with data assimilation globally – problem of $M\dim(X_f)\dim(P)$ dimension

Possible simplification:

1. Neglecting the system errors ε reduces the dimension of the problem by M.
2. Sequential approach

$$\rho_f^a(x(t_1), p \mid d_1) = C\rho_{d_1}(d_1 \mid x(t_1))\rho_f^f(x(t_1), p)$$

Evolving the forecast error statistics ρ_f^f is still a problem of $\dim(X_f)\dim(P)$ dimension

Monte-Carlo methods is to avoid the dimension problem.

(The error of Monte Carlo approximation is of $O(N^{-1/2})$, here N is number of samples)

Ensemble Initialization

Spread of the initial ensemble reflects uncertainties in knowledge of *a prior* system and parameter pdf

$$\rho_0(x(0)) = K^{-1}\sum_{k=1}^{K}\delta(x(0) - x_k(0)),$$

$$\rho(p) = K^{-1}\sum_{k=1}^{K}\delta(p - p_k),$$

δ - Dirac functions

Practically:
An ensemble of K members is generated from an exponential distribution

$$y = \bar{y}\exp(-\frac{y}{\bar{y}})$$

mean of the distribution is assumed to be a first guess.

Data and weighting

The relative weights might be calculated under the assumption of Gaussian

$$\rho_d(d_n \mid x_k(t_n)) \approx \omega^k = C\exp[-0.5(X^k - d)^2 / \sigma^2],$$

or Lorentz data errors

$$\omega^k = C/(1 + (X^k - d)^2 / \sigma^{-2}) \quad \textit{(van Leeuwen, 2004)}$$

where σ^2 is the variance of the observation.

Resampling, practically:
IMSL library, DRNGDA (RNGDA) generates pseudorandom numbers from a general discrete distribution.

The idea – approximating the continuous probability density function (pdf) with an ensemble of δ-functions – particles,- which evolves according to a stochastic dynamical model.

One of the particle filters The Sequential Importance Resampling filter (SIR) has been first introduced by Rubin(1988),
implemented for dynamical systems by Gordon et al. (1993);
for sequential parameter estimation in stochastic systems, introduced by Kivman (2003);
applied for basin scale problem by van Leeuwen(2003, 2004);

References

Brasseur, P., Bahurel, P., Bertino, L., Birol, F., Brankart, J.-M., Ferry, N., Losa, S., Remy, E., Schroeter, J., Skachko, S., Testut, C.-E., Tranchant, B., van Leeuwen, P.-J., Verron, J., 2005. Data assimilation in operational ocean forecasting systems: the MERCATOR and MERSEA developments. *Quarterly Journal of the Royal Meteorological Society*, 131(613), 3561-3582.
Kivman, G.A., 2003. Sequential parameter estimation for stochastic systems. *Nonlinear Process. Geophys.* 10, 253-256.
van Leeuwen, P.J., 2003. A truly variance minimizing filter for large-scale applications, *Mon. Weather Rev.* 131, 2071-2084.
Rubin D.B., 1988. Using the SIR algorithm to simulate posterior distribution,in Bayesian Statistics 3 (Eds. J.M. Bernardo, M.H. Degroof, D.V. Lindleyand, A.F.M. Smith). Oxford Univ. Press.,395-402.

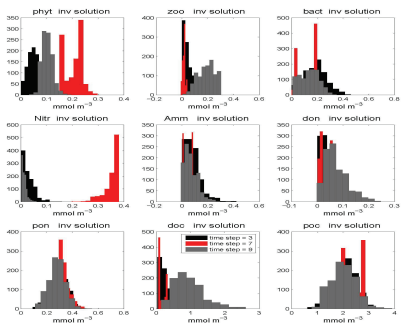
Advantages

When doing particle filtering, one updates probability of the particles according to their agreement with the observed data.

Particle filters make use of the full forecast and data errors statistics; they are truly variance minimizing methods (van Leeuwen, 2003)

Example

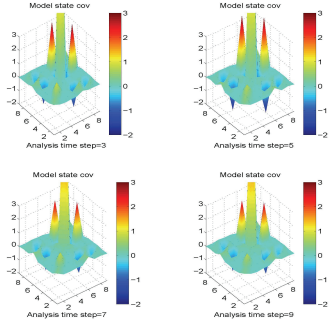
We consider a 9-compartment model describing dynamics of phytoplankton (phy), zooplankton (zoo), bacteria (bact) and cycle of particular and organic matter (poc and dom), similar to what was used in Losa et al (2003). The model was constrained by data of the Bermuda Atlantic Time-series Study, particularly, by measurements of nitrate, chlorophyll, dissolved organic nitrogen and carbon concentrations for the period December 1988 to January 1994. All the data were averaged over the ocean upper mixed layer (UML). The UML thickness were estimated by means of an analysis of BATS temperature profiles for the same period. The UML depth is determined as the depth at which the temperature is $0.5^0C less than that at the surface.



←Figure to the left depicts probability density obtained at different analysis time steps (days 1468, 1547,1580) for 9 components of the ecosystem mode (ensemble size is equal to 1000).

Figure to the right→ shows covariances between the model components calculated after 1468th day of the model integration.

← The pdfs of model states, parameters and variance of model noise obtained after 1547th day of model integration are then used to generate a new ensemble of 200 members to initialize the model (next data assimilation ru

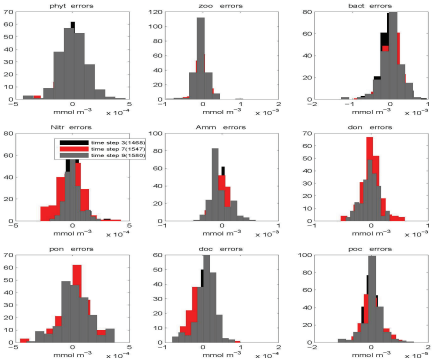


Problems

The SIR filter is known to suffer from degeneration of the ensemble (van Leeuwen, 2003) if either the system noise does not provide sufficient spreading of states which are resampled several times or the ensemble badly approximates the true prior distribution (the distance between the best member and the true state is too big). This problem is even more pronounced in the case of simultaneous state-parameter estimation where regenerating the number of samples in the parameter space is needed.

To avoid this the following procedure is used →

Figure below depicts estimated probability density of model errors, which, as one can see, have appeared to be non Gaussian. The more correct the model errors are accounted for, the better biological model parameters estimates and therefore the model state forecast are (Brasseur et al., 2005)



Model noise generation and jittering model parameters

Levels E of the model noise ε might be considered as additional parameters to be optimized $E \subset P$.

If, at an analysis step, parameter values are resampled (r) many times, a new parameter ensemble can be redrawn (West, 1993) from a smoothed approximation of the *posterior* probability density

$$\rho_r^a(p(t_n)) = K^{-1}\sum_{k=1}^{K}\delta(p(t_n) - p_a^r(t_n))$$

Practically: either from

a uniform distribution within the interval $p \pm \sigma_p$... one has to specify [p – nearest smaller value, p + nearest higher value];
or
a normal distribution with a variance... one has to specify;
...