Variational Particle Approximation for imperfect models

C. Baehr¹,²

Whereas classical data processing techniques work with perfect models, or models with dynamical white noise, geophysical sciences have to deal with imperfect models containing spatially structured errors. For the perfect model cases, in terms of Mean-Field Markovian processes, the optimal filter is known: the Kalman estimator is the answer to the linear Gaussian problem and in the general case Particle approximations are the empirical solutions to the optimal estimator with $O(\frac{1}{\sqrt{N}})$-type of convergences. We will present another way to decompose the Bayes rule, using an one step ahead observation. This method, very expensive with a forward/backward integration, is well adapted to the strong nonlinear

¹C. Baehr, Météo-France - Centre National de Recherches Météorologiques, 42 Avenue de Coriolis, 31057 Toulouse Cedex, France. (christophe.baehr@meteo.fr)

²Météo-France / CNRS, CNRM / GAME
URA1357, Toulouse, France.

²Institut Mathématiques de Toulouse, Université de Toulouse III - Paul Sabatier, Toulouse, France.
or chaotic systems. Then, in order to deal with imperfect model, we suggest in this presentation to learn the (large scale) model errors using a variational correction before the resampling step of the non-linear filtering. This procedure replace the a-priori Markovian transition by a transition kernel conditioned to the observations. This supplementary step may be read as the use of variational particles approximation. For the numerical application, we have chosen to show the impact of our method, first on a simple marked Poisson process with Gaussian observation noises (the time-exponential jumps are considered as model errors) and then on a 2D shallow water experiment in a closed basin, with some falling droplets as model errors.

1. Classical nonlinear filtering

There are many ways to introduce the nonlinear filtering problem. For any time $t < T$, where $T$ is a bounded horizon, considering a random process $X_t$, what could be the information brings by an other process $Y_t$ about the hidden process $X_t$? this question leads to the beautiful theory of the innovation, with the Kushner-Stratonovitch (KS) equation as a flagship. The solution of this infinitesimal generator is a conditional probability law $\hat{\eta}_t(dx_t) = \mathbb{P}(X_t \in dx_t|Y_t = y_t)$. In the general case, to find the solution of the KS equation is still an open result.

Nevertheless, the conditional law $\hat{\eta}_t$ may be decomposed using the Bayes rule, and this decomposition reveals the Markovian structure of the filtering process and give up a sequential algorithm (see Del Moral [2004]). With discrete time step $n.\Delta t$, using a dynamical Markovian transition $M_n$ and a likelihood density $G_n$ according to a sequence of observations, the propagation of the conditional law $\hat{\eta}_n$ has a Markovian representation as a pdf transportation process (also called McKean representation) with $\hat{\eta}_{n+1} = \hat{\eta}_nM_{n+1}S_{n+1,\eta_{n+1}}$.
where $\eta_{n+1}$ is the predictor law given by $\eta_{n+1} = \hat{\eta}_n M_{n+1}$ and $S_{n+1,\eta_{n+1}}$ is the Markovian state selection kernel. To correspond to a filtering process, the selection kernel averaged by $\eta_{n+1}$ has to be equivalent to the Bayes rule. This state selection has not an unique representation as a Markov process. The state resampling $(S_{n,\eta_n}(x, dy) = \psi_n(\eta_n)(dy) = \frac{G_n(\eta_n(dy))}{\eta_n(G_n)}$ coming directly from the Bayes rule) is a possible solution and the most used. But genetic selection kernels $(S_{n,\eta_n}(x, dy) = \varepsilon_n G_n(x) \delta_x(dy) + [1 - \varepsilon_n G_n(x)] |\psi(\eta_n)(dy)\))$ where $\varepsilon_n$ is a tuning parameter), which are equivalent in mean to the Bayes rule, are also possible representation. Anyway, the selection step is a Mean-field process for which a particle approximation is possible. Let be $N > 0$ the number of particles, we denote $(X^i_n)_{1 \leq i \leq N}$ the particle system initiated by $(X^i_0)_{1 \leq i \leq N}$ i.i.d. with $\eta_0$. The transition kernel is given for any $1 \leq i \leq N$ by $M_{n+1}(X^i_n, dX^i_{n+1})$ and the selection uses the particle approximation of the selection kernel $S_{n,\eta_n}$, for instance for a genetic selection, the empirical kernel is $S_{n,\eta_N}(X^i_n, \cdot) = G_n(X^i_n) \delta_{X^i_n}(\cdot) + [1 - G_n(X^i_n)] \psi(\eta_N)(\cdot)$ where $\eta_N \overset{\text{def}}{=} \frac{1}{N} \sum_{j=1}^N \delta_{X^i_j}$ and $\Psi_n(\eta_N)(dx) \overset{\text{def}}{=} \sum_{j=1}^N \frac{G_n(X^i_j)}{\sum_{k=1}^N G_n(X^i_k)} \delta_{X^i_j}(dx)$. The particles set follows the sequence $X^i_n \xrightarrow{\text{Selection}} \hat{X}^i_n \xrightarrow{\text{Prediction}} X^i_{n+1}$. With simple arguments, one can prove the following convergence result: For all $n \geq 0$ and $p \geq 1$, it exists finite constants $C_n^{(p)} < \infty$ such that for any measurable bounded functions $f$, $\sup_{n \geq 0} E(|\eta^N_n(f) - \eta_n(f)|^p)^{\frac{1}{p}} \leq \frac{C_n^{(p)}}{\sqrt{N}} \|f\|$. So we just see that there are several equivalent Markovian selection kernel to adress the filtering problem. They have not the same properties. For instance, it is possible to show numerically (see Bachr and Pannekoucke [2009]) that the behaviour of the different selection kernel is not the same in front of high dimensional dynamics. Now we are going to discuss how to change this state selection step with a model error learning.
2. Reformulation of the filtering Bayes rule

For an observation sequence $y_0, \ldots, y_n$ using the Bayes decomposition, in terms of pdf we have $p(x_n|y_0, \ldots, y_n) \propto p(y_n|x_n) p(x_n|x_{n-1}) p(x_{n-1}|y_0, \ldots, y_{n-1})$. We retrieve in this decomposition the Markovian description of the previous section and its algorithm. But this decomposition is not unique and there is an other way to use the Bayes rule. For instance, we can write $p(x_n|y_0, \ldots, y_n) \propto p(y_n|x_{n-1}) p(x_n|x_{n-1}, y_n) p(x_{n-1}|y_0, \ldots, y_{n-1})$.

What means this new formulation and what could be the corresponding algorithm? This formulation corresponds to a one-step ahead insider. The observation $y_n$ modifying the transition kernel and the result is weighted by the likelihood $p(y_n|x_{n-1})$. Having information one-step ahead, this conditional formulation could be a way to learn missing dynamical pattern. But this insider algorithm seems to be really expensive. The first step is a transition from $x_{n-1}$ to $x_n$ conditioned to the observation $y_n$. The observation acts as a tuning parameter of the prediction model and the good performance of the prediction is highly dependent of the observation quality, even if it means to explore a large amount of possible trajectories. Then there is a selection of the previous state $x_{n-1}$ according to the observation $y_n$, in other words it selects the best trajectories among all the conditioned forecast. These integrations are not trivial and may be quite expensive. Therefore we suggest an other way to proceed to.

3. Change the filtering problem using a conditioned Markov

This section explains an empirical adaptation of the previous result. But it contains an interesting manner to deal with imperfect models. What is the algorithm? First a prediction with an a priori Markovian model. Then a variational correction using the next observation. The combination of the a priori prediction and the variational correction is
the observation conditioned forecast met in the previous section. Then in order to have a
filter, we use a selection kernel using the last observation.

This algorithm with the variational correction is able to learn the smooth structures
of the model error. Indeed, the variational minimization is a correction of the predicted
state $\bar{x}_n$ according to the observation $y_n$. This step gives back a conditioned forecast $\hat{x}_n$
which may be written as $\hat{x}_n = \bar{x}_n + K_n^V(\bar{x}_n)[y_n - H(\bar{x}_n)]$, where $K_n^V(\bar{x}_n)$ is the variational
gain, analogous to the Kalman gain of the linear Gaussian case (see Fleming [1997]).
The second part of the correction summation is the contribution of the innovation process
which is the information held by the observations and not present in the model prediction,
i.e the model errors in the case of a imperfect model. Therefore the algorithm is a three
step sequence : $\hat{x}_{n-1} \xrightarrow{\text{Prediction}} \bar{x}_n \xrightarrow{\text{Correction}} \hat{x}_n \xrightarrow{\text{Selection}} \hat{x}_n$.

Is this still a nonlinear filter? The answer is clearly positive if we consider the Markovian
transition $M_{n,y_n}$ (combination of the kernel $M_n$ and the variational correction) as our
filtering model. But nothing ensure that this process is equivalent to the filtering problem
using the exact transition $M_n$. Here we are thinking about imperfect model, and our
method is an empirical adptation of the exact insider filter.

Nevertheless, the variational correction in the prediction step raises question. The
formulation is valid only for conditional averages. Here we apply the correction to Markov
state, and further to particles, instead of averages. So for a particle $(\bar{X}_n^i)_{1 \leq i \leq N}$ there exists
a compensator process $Z_n^i$ such that $\hat{X}_n^i = \bar{X}_n^i + K_n^V(\bar{X}_n^i)[y_n - H(\bar{X}_n^i)] + Z_n^i$. In the linear
Gaussian problem, it is easy to see that the compensator is a centered Gaussian process
with variance $\hat{P}_nK_nC_n$ ($\hat{P}_n$ is the prediction covariance error, $K_n$ the Kalman gain, $C_n$...
the observation operator). It is interesting to remark that our algorithm for the linear Gaussian case is equivalent to the weighted EnKF developed by Papadakis et al. [2010], but with non-empirical covariance error matrices. Out of the linear Gaussian case, the shape of the compensator process is still an open question.

4. Numerical applications

We propose now two numerical applications. Each one tries to explore the question of imperfect models and both try to show that the add of a variational correction allow the particle approximation to reduce drastically the number of efficient particles.

4.1. The marked Poisson process

The marked Poisson process used in this experiment follows the dynamics equation

\[ X_{n+1} = X_n + G_n \cdot P_n \] with \( X_0 \sim \mathcal{N}(0,1) \), \( G_n \) is a centered Gaussian process (variance \( \sigma^2 = 1 \)) giving the jump amplitude, and \( P_n \) is a Poisson process with parameter \( \lambda = 0.03 \) giving the jump time and modulating the variance of \( G_n \). This process is observed at each time step with Gaussian perturbations with variance 0.5.

First we filter the observations with a classical Kalman filter. To be adapted to a linear Gaussian dynamic, the jump process is represented in the Kalman filter by a Gaussian process with a variance \((\sigma \cdot \lambda)^2\). The results given figure 1 are without surprise, the Kalman is late w.r.t the reference signal and has exponential convergences after every jumps. The second tested filter is a classical particle filter with a genetic selection step. This filter has only 10 particles and we can see on figure 1 its difficulties to follow the reference signal: there are two few proposal particles and the filter is close to the divergence. The next filter is a trajectorial particle filter (filtering on a genealogical tree (see [DelMo04])). Normally
this is the best estimation and it can be used as reference. Again with 10 particles the genealogical filter has trouble to estimate the good signal. The last filter is our variational particle filter (VarPF) with 10 particles. Here the variational correction is given directly by a Kalman correction. Each particle has a compensator process described in the previous section. The variational particle estimation outperforms the other filters. We may find in the behaviour a mixing of the classical Kalman and classical particle filter. But after a jump the filtered process joins the correct level faster than the other filtered signals, and between two jumps we have an optimal estimation like the Kalman may have.

4.2. The 2D-SWE with droplets errors

In this experiment, the dynamics is given by a 2D Shallow Water equation. Numerically we use a finite volume resolution with a $30 \times 30$ grid. To avoid non-observability problems, all the grid points are observed by a direct process and Gaussian noises. Like this the filtering process is only the estimation of the initial conditions, and this is a too easy question. So we add every 75 time-steps a droplet (adding an exponential shape in the center of the grid). The shape of the droplets, the time of the impact and their intensity are not known by the filters. First we use a classical variational filter (3DVAR) with a covariance matrix error empirically given to filter enough the observation perturbations and capture the error model at the same time. Then we use our VarPF with 10 particles. For each variational minimization we use the same covariance matrix used in the 3DVAR in order to have a full comparison. The figure 2 shows one instant, eight time-step after the fall of a droplet. The bottom is our reference system. The 3DVAR is on top and the VarPF in the center of the picture. An animation loop of the simulation shows clearly the ability of the VarPF to learn quickly the error model, faster and better than a 3DVAR. As
an example, the figure 2 shows this distinctly. Only 10 variational particles are necessary to obtain this result even if the state vector has 900 dimensions. The 3DVAR has a smooth and flat response to the error model, only large structures are seen with a weaker intensity. A simple particle filter, where the model error is not implemented (otherwise it would not be a model error), is unable to catch the droplet impact and the evolution of the water height after the event. This is clearly the advantage of the variational particles and the conditioned prediction used in the VarPF.

5. Conclusion

In this presentation we have introduce the variational particle filter. This is an adaptation of the one-step ahead insider filter. But this VarPF is able to learn smooth model errors and requires few variational particles to achieve the filtering. This VarPF is a three steps filter with a prediction, a correction and a selection of Markovian states. In the numerical experiments we have presented, the VarPF seems to be able to learn partially the model errors. This has been done for a simple 1D marked Poisson process as well as a 2D shallow water simulation with droplets as model error. In this second experiment the number of variational particles (10 at the most) is relatively weak with respect to the state vector size (about 900 components). Moreover the VarPF seems to be much more reactive than a simple variational filter in the presence of model errors. To conclude, the VarPF seems to be an interesting alternative to the pure particle filter to the question of the dimensionality or the imperfect models.
References


Figure 1. Marked Poisson process with Gaussian observation noise filtered by four methods. Left above is the result of a classical Kalman filter, bottom left is the filtering by a simple particle with genetic selection. Above right, a trajectorial particle filter with genetic selection. Finally bottom right is our filter with a Kalman correction for any particles and a genetic selection. All the particle approximation use only 10 particles.
Figure 2. Based on a 2D Shallow Water simulation, we add droplets as model errors and observed the height water at all the grid points with a Gaussian noise (state vector with 900 dimensions). The pictures are extracts of the animation loop, eight time step after a droplet fall. Above the result of a classical variational filter (3DVAR), in the center the result of our variational particle filter (VarPF) with 10 particles and the bottom is the reference.