Nonlin. Processes Geophys. Discuss.,
https://doi.org/10.5194/npg-2018-15-AC2, 2018

Nonlinear Processes in Geophysics
Discussions
Open Access
EGU

Interactive comment on "Review article: Comparison of local particle filters and new implementations" *by* Alban Farchi and Marc Bocquet

Alban Farchi and Marc Bocquet

alban.farchi@enpc.fr

Received and published: 24 July 2018

We thank Dr. Stephen G. Penny (Reviewer 2) for his insighful comments and suggestions.

In this article, we describe and compare many different localised PF methods. Contrary to many articles in the PF literature, the algorithmic sections are detailed and many explanations are given about potential numerical choices. The numerical illustrations use several models, not only in one dimension, and an exhaustive exploration of the algorithm parameters is performed. Please keep in mind that this goes well beyond most studies on the subject.

Printer-friendly version

Discussion paper

1 General comments

1.1 Comment 1)

The relative costs between the methods should be calculated and compared, along with RMSE, in each section. Some discussion of costs is made in passing, but no quantitative analyses are offered until the end where a few methods are compared. I suggest extending this to each of the major direct comparisons at the end of the LPFx and LPFy sections.

The numerical complexity of each method is now discussed in Sects. 4.5.2 and 7.5.2. For the BV model with high resolution, the computation times are reported in Table 5. For the low-order models however (standard L96, BV with coarse resolution), we did not to add the "companion plots" suggested by the reviewer for several reasons:

1. The computation time highly depends on many other factors than numerical complexity: implementation, programming language, processors architecture... With low-order models, these factors may be very important and therefore the computation time is often irrelevant.

2. In our configurations, the parameters are not reprensentative of realistic applications: the number of grid points is very small, the number of particles is very high, and the number of spatial dimensions is limited (to 2 in our case).

3. This would add approximately a dozen figures to an article which is already long.

In the PF literature, the only article in which we have seen some discussion about complexity and computation time is the one by Penny and Miyoshi (2016). Finally, we want to highlight the fact that conclusions regarding the computational cost of a method cannot be based on test series with low-order models. From our experience,

the ranking of an algorithm in computational cost looks very different when using the one-dimensional L96 model or the two-dimensional BV model with high resolution.

1.2 Comment 2)

It would be useful to show the results of the EnKF baseline, both in RMSE and computation time.

The EnKF scores were given in the text for the L96 model and shown in Fig. 17 for the BV model. We have added a new figure (Fig. 4) that shows the score of the EnKF with the L96 model. We have also added horizontal baselines in most LPF figures (Figs. 8, 13, 14, 15, 16, 18, 20 and 21).

Since the local PF variants are not outperforming the EnKF baseline, the authors should consider some special case scenarios in which the PF does outperform the EnKF as a motivation for continued development of the local PFs and to show why local PFs may also have advantages over standard EnKFs.

This is not true, several LPFs do outperfrom the EnKF in the standard configuration for the L96 model (see Fig. 20), which is a première to our knowledge. Besides, the point of this article is not to design LPF algorithms that beat the EnKF in every configuration, but rather to improve the design of LPFs. Contrary to many studies on the PF, we have chosen to use the dynamical models in standard configurations, which allows for a fair comparison with the EnKF.

Following your recommandation, we have added a test series in Appendix C, in a configuration built to make the EnKF fail. In this configuration, we use the same strongly nonlinear observation operator as Poterjoy (2016). However, the interpretation of these results is harder, because some legitimate question can be asked:

1. Is this configuration relevant for realistic models?

2. How good are the score of the LPFs? There is no baseline for comparison (the EnKF does not count since we are outside the scope of its assumptions).

1.3 Comment 3)

In general, I find the algorithm names confusing.

The coding system for the LPFs looks complex. But please keep in mind that this system was designed to distinguish more than 20 different algorithms. Other studies that focus on a limited number of methods could use a much simpler coding system.

In fact, the coding system follows one simple principle: capital letters refer to the main algorithmic steps and subscripts are used to differentiate the methods. This is now explained in the caption of Tables 1 and 3.

The first half of the paper uses a complex coding system, while the second half credits the authors who developed the methods.

We find the criticism unfair, because every time we introduce a new method, we cite the authors that inspired our work if any. Both LPF^{x} and LPF^{y} algorithms follow the same convention, with different subscripts to refer to different methods. For the LPF^{y}s, using 'P' and 'RK' as subscripts is a way to distinguish the two different propagation methods and not to credit the author who developed the method. We could have used 'h' and '2' (for 'hybrid' and '2nd order') but obviously this is harder to remember. It serves exactly the same purpose as the subscrits 'e' and 's' (distinction between optimal transport in ensemble space and in state space).

A more consistent and simpler naming convention would be nice from a reader's point

of view, and should be used throughout.

Following your suggestion, the naming convention has been modified. The new system is consistent and as simple as it can be given the fact that it is used for more than 20 algorithms. It is explained in details in the caption of Tables 1 and 3.

In addition, a single table describing every algorithm name, what is does, and what section it can be found in, would also help to add clarity.

Description tables for the LPF^{x} and LPF^{y} algorithms have been added (Tables 2 and 4). Thank you for the suggestion; this is a great addition.

From what I can tell, the S(IR)ˆ xSR_a method appears closest to that of Penny and Miyoshi (2016), as it uses smoothing of weights and adaptive "regularisation jitter" based on the ensemble perturbations, and I think this should be given proper credit, as one of the few LPF methods offered in the geophysical literature that has a combination of good performance and low computational cost.

We did cite Penny and Miyoshi (2016) when presenting the smoothing by weights step in Sect. 4.4.1. As shown by our new results with the high-dimensional BV model, the S(IR)^{x}S$_{PM}$R$_{c}$ (new nomenclature), our generalisation of the LPF of Penny and Miyoshi (2016), does not have a favourable ratio accuracy / computation time.

2 Specific comments

2.1 L28

Also cite Kalnay and Yang regarding the "Running-in-Place" method.

We have added a reference to the RIP as an important precursor method. Note however that, contrary to the MLEF, the 4D-ETKF and the IEnKS, the RIP is not mathematically consistent.

2.2 L39

comma after hybridisation

Done.

2.3 L43

[fewer] particles

Done.

2.4 L60

degree[s] of freedom

Done.

2.5 L63

geophysical system[s]

Done.

2.6 L85

I'm assuming y_k:0 is the set of all observations from time t=0 to time t=k, but perhaps you can state that explicitly.

Yes, this was stated on the same line.

2.7 L91

Of course, there are many different goals in data assimilation. This is a typical goal. My immediate reaction here is that the DA filtering problem consists in estimating pi_k+1|k. This is the goal at least, usually, to make a prediction. Perhaps you can say - "The DA filtering problem consists in estimating pi_k|k and pi_k+1|k with given realizations of y_k:0."

The sentence has been changed to reflect the fact that estimating $\pi_{k|k}$ may not be the only goal of DA.

2.8 L97

particle representatio[n]

Done.

2.9 L111

What do you mean by "pure ensemble transformations"?

"Pure ensemble transformation" means that this is a transformation that act on the ensemble and that should ideally not alter the density. This is explained by the following

sentence. However, we agree that the term "pure" can be confusing and we removed it.

2.10 L170

I'm not sure I understand why this is remarkable. Could you elaborate?

This is remarkable that in w_{k+1}^i, the dependance on $\mathbf{x}_{k+1}^i$ vanishes. We have added a reference to Doucet (2000) for clarification.

2.11 L177

"... to more elaborate algorithms ..."

Done.

2.12 L182

"... models [have led] to weight degeneracy..."

Done.

2.13 L196

" ... it [might seem] surprising that, although MC method[s] have..."

Done.

2.14 L214/217

You switch tenses, first referring to Synder et al. (2008) as a set of authors, and then referring to Snyder et al. (2008) as a paper. Because "et al." means "and others", I prefer the former and recommend changing L217 to: "Snyder et al. (2008) [do] not illustrate..."

Done.

2.15 L223

"...optimal importance proposal [density]..."

Done.

2.16 L227

"... does not [primarily] come from ..."

Done.

2.17 L238

"... [elaborate] models ..."

Done.

2.18 L252

It seems awkward to begin a sentence with a variable name. Perhaps used instead: "The quantity taoˆ 2 would then be defined using"

The beginning of the sentence has been changed. Thank you for the suggestion.

2.19 L266

While I appreciate the implication of calling this a 'discontinuity', there are some complications in defining a concept of continuity on a discrete model grid. Some discussion should be made regarding this point.

Indeed, discontinuity here does not refer to the mathematical notion of continuity. Following your suggestion, we have added some explanation.

2.20 L281

Perhaps you could list some of the past examples of this type via citation.

Done.

2.21 L285

Again, I suggest citing a few example of this type as well.

Done.

Printer-friendly version

Discussion paper

2.22 L332

Not within a circle, but within some general local region. A circle is a common choice.

We changed the subsection. Thank you for helping us clarify this point.

2.23 L367

"and decrease[s] exponentially"

Done.

2.24 L371

The "size" of the blocks using what measure? Number of grid points?

We clarified this point. Thank you for spotting the imprecision.

2.25 L386

I think "hold" should be "held"

Done.

2.26 L390-415

I'm not sure if the point was adequately made that neighboring weights can be made arbitrarily smooth by letting the radius of the taper function (r_s) get large. I.e. as r_s goes to infinity, the global PF solution is recovered.

The asymptotic limit of a LPF^{x} algorithm using smoothing by weights is now discussed in Sect. 4.5.4. When $r \to \infty$, $\mathbf{E}^{r}$ is not necessarily equivalent to the global PF solution (because the resampling is independant at each grid point). When $r \to \infty$ and $r_{s} \to \infty$, $\mathbf{E}^{s}$ is not necessarily equivalent to the global PF solution (again because the resampling is performed independantly at each grid point).

In that sense, I'm not sure why the additional alpha smoothing step is made explicit.

We do not understant your concern about making the "alpha smoothing step" (we guess you mean Eq. (32)) explicit.

2.27 L543

"only [a] big ensemble"

The sentence has been changed.

2.28 L561

"RMSE offers a" to "RMSE offer a"

Done.

2.29 L562

I'm not sure it is settled that the RMSE of the mean is an adequate measure of the PF performance, given that the distribution may not have the mean and mode equal.

We do not entirely understand the point of this remark. One must keep in mind that PFs are suited to compute the mean state and not the mode. Besides, in these weakly

non-Gaussian configurations, mean and mode should not be far from each other.

Further, if we are to adopt a PF solution over an EnKF, then we are acknowledging that the primary data assimilation goal is specifically not mean state estimation, but rather estimation of the state distribution.

We do not agree with this remark. Using a PF does not mean that we are not interested in the mean state. And, again, one must keep in mind that PFs are not suited to "estimat[e] the state distribution". Indeed, with a PF, we primarily have an estimation of

$$\int p(\mathbf{x}) f(\mathbf{x}) \mathrm{d}\mathbf{x}, \tag{1}$$

for any test function f, but no estimation of $p(\mathbf{x})$.

2.30 L572

"yield[s]"

The sentence has been removed.

2.31 L595

I don't understand what this first sentence means. What does it mean to have more information than the truth?

Indeed there was a typo: one should have read "on average more informations than the observations". Sorry about that; it has been corrected.

2.32 P25-25

It would be nice if Figures 4 and 5 were closer to the referencing text. Perhaps you can make that request of the editors.

We will ask for this in the editing process.

2.33 P25

It appears here that you are using a fixed parameter for the 'regularization jitter'. Have you compared this the LPF of Penny and Miyoshi (2016) that set this value adaptively based on the analysis ensemble spread?

The discussion on "adaptative" resampling is located in the "coloured noise" section (5.8). In this section, we developped a method that is an extension of the method by Penny and Miyoshi (2016). We compared our method to that of Penny and Miyoshi (2016) (not shown in this article) and always found better accuracy with our extension. We have added a few sentences in section 5.8.2 about this.

2.34 L651

I'm confused how the higher RMSEs of the S(IR_SU)^ xR algorithm indicates an efficient approach. Could the authors elaborate.

The $S(IR_{su})^{\times}R$ algorithm is the only one that does not use the "adjustment-minimising" property. If it has a higher RMSE, we believe that it means that the "adjustment-minimising" property is efficient. Following your comment, we reformulated the statement.

Printer-friendly version

Discussion paper

What is the RMSE ratio used in Figure 9? Why does the figure caption say "RMSE" while the y-axis says "RMSE ratio"?

The RMSE ratio used in Fig. 10 (new numbering) is detailed in the second sentence of the caption "The scores are displayed in units of the RMSE of ...". The same kind of ratio is used in Figs. 9, 11 and 12.

2.35 L675

I need a reminder at this point - E is the set of ensemble members and X is the set of perturbations around the ensemble mean? Are the xˆ i in (46) the columns of E?

$\mathbf{E}$ is the ensemble matrix (defined in Sect. 2), whose columns are the particles $\mathbf{x}^i$, which is a very common notation in ensemble DA. $\mathbf{X}$ is indeed the set of (normalised) perturbations around the ensemble mean, as defined two lines above. Reminders have been added, thank you for the suggestion.

2.36 L697

change "as following" to "as follows"

Done.

How is equation (48) different than (46)? Could you instead just say it is defined as in (46) with a new formulation for the Gaussian regularization jitter covariance matrix (49)?

Equation (51) is different from Eq. (49) (new numbering) because it uses the local weights w_n^i instead of the global weights w^i that do not exist in LPFs. We explained

in the text why Eqs. (47) to (49) cannot be used (see the second subsection of Sect. 5.8.2)

2.37 L702

Am I interpreting these figures correctly in that the new proposed approaches are all mostly making the RMSE larger relative to the S(IR)ˆ xR (in the small-ensemble size cases of interest)?

Your interpretation is correct. This is discussed in the last subsection of Sect. 5.8.

2.38 L740

The smoothing appears to have significant benefits. Are there any strategies for how this could be applied if an exhaustive optimization of the parameters is not possible (e.g. for a large system)?

As shown by Fig. 12, $\alpha_s = 1$ is optimal in this configuration for the L96 model. We have checked that this is the case in most situations where we used the smoothing (in particular with the BV model). However, we could not find an obvious relationship between the optimally tuned values of r and r_s. Besides, one should keep in mind that, in the "small-ensemble size cases of interest", the benefits of the smoothing are far less impressive than the benefits of OT (this can be seen in Fig. 16).

2.39 L742

Do you have the baseline RMSE values for the EnKF?

See the new Fig. 4. We have also added a baseline to Fig. 13.

2.40 L743

"From these results, we conclude that the smoothing by weights step [of Penny and Miyoshi (2016)] is an efficient way of [reducing] the artificial discontinuities [that were] introduced when concatenating the locally updated particles, especially when combined with the coloured noise regularisation jitter method." I should note that the S(IR)xR_a method appears closest to that presented by Penny and Miyoshi (2016), since their inflation is adaptive and using the terminology here is a regularization jitter scaled by the ensemble anomalies.

Please keep in mind that the work of Penny and Miyoshi (2016) has been cited in Sect. 4.4.1 (where we introduced the method in the first place) and that the $S(IR)^{\times}S_{PM}R_c$ algorithm tested in this section is not the LPF of Penny and Miyoshi (2016) but an improvement thereof, which includes: a more general framework that can be applied to different types of resampling, a tapering function, a smoothing radius and a smoothing strength parameters, coloured noise regularisation.

The other corrections ("reducing" and "that were") have been done.

2.41 L771

The results look very nice with the OT approach. Do you have an analysis of the relative costs of each of the methods as a function of system size, observation count, and ensemble size?

This is detailed in the new section 4.5.2.

2.42 L794

"local LET algorithm" Is that redundant? Perhaps just say "LET algorithm"

"LET" means linear ensemble transform (introduced in Sect. 2.3 with appropriate citations).

2.43 L804

I think it would be appropriate at this point to provide a companion plot that shows the relative cost for each method as well.

Please, see the discussion about general comment 1).

2.44 L809

Perhaps you should put the EnKF baseline on the plot as well.

See the new Fig. 4. We have also added a baseline to Fig. 16.

2.45 L816

"dynamic[s]"

Done.

2.46 L828

"The ETKF requires at least Ne = 12 ensemble members to avoid divergence." This would imply that the number of positive and neutral Lyapunov exponents of the system is 11.

You are right.

2.47 L835

It may not hurt to repeat the definition of each algorithm here.

We now refer to the algorithms' list in the (new) Table 2. Thank you for the suggestion.

2.48 L923

"The SO formalism is elegant." This seems a strange characterization given that the next few sentences describe legitimate problems with the approach.

We wanted to emphasis that the formalism developed in Sects. 7.2.1 and 7.2.2 looks elegant. The next few sentences raise issues that appear when combining the SO formalism with the PF. These issues are not specific to the SO formalism. Following your commment, we mitigated this remark.

2.49 L940

I suggest either staying consistent with the rest of the paper and defining the section using the algorithm name adopted in the paper - LPFˆ y - S(IRP_P)ˆ yR, or renaming the rest of the algorithms in the paper based on the authors that introduced them.

The nomenclature for the algorithms has been changed to follow your suggestions (please, see the discussion about general comment 3)). Thank you for spotting the inconsistency in the naming of the subsections. We corrected this point.

2.50 L968

The terms 'ensemble member' and 'particle' are synonymous - they differentiate the same concept developed in two different fields. The term ensemble does not imply a 2 moment method, so the naming convention shouldn't be used for the purpose stated here.

We now refer to this algorithm as "the second order propagation algorithm".

You are right, the terms "ensemble member" and "particle" designate the same concept but they are used in different contexts. The term "particle" is often used with Bayesian method (or at least with methods based on importance sampling) while the term "ensemble member" is often used with Kalman filters. The naming convention is therefore already commonly being used to distinguish between Bayesian / non-Bayesian methods.

2.51 L970

"one first need[s] to"

Done.

2.52 L971

This computation is expensive for large systems. Is this computed in ensemble space or model space?

We stated in Sect. 7.4.3 that only submatrices of Σ need to be computed. Therefore this computation is not that expensive for large systems.

How to implement Eq. (65) is beyond the scope of this article. For our numerical experiments, we computed it in state space. Note that because of the schur product, there is no obvious formula for Σ in ensemble space.

2.53 L1009

"any distance that need[s] to be computed relative[] to the observation site..."

Done.

2.54 Table 2

The nomenclature table is somewhat helpful, but I'd prefer a full table showing each method for which results are presented, with a description of the method, and the section where it can be found in the text.

The caption of this table has now more details, such that it is more explicit. Following your recommendations, we have added new tables (Table 2 and 4). Thanks, this is a great addition.

2.55 L1013

If the block computing is required to make the algorithms computationally scalable to large systems, then these are the results that should be reported.

This is only a discussion about algorithmic possibilities. The block computation is a way to reduce the computation requirements of LPF^{y}s. It should not be required to make the algorithms computationally scalable to large systems.

2.56 L1073

"size Ne grows[,] the RMSE decreases"

Done.

2.57 L1075

Again, I suggest showing the LETKF baseline RMSEs, as well as the computational costs of each method.

The RMSE values for the LETKF are in the dedicated figures. We have also added a baseline to Figs. 20 and 21. For the computational cost, please see the discussion about general comment 1).

2.58 L1019

"few but [dis]similar LPF algorithms"

Done.

2.59 Figure 21

The better of the white noise and colored noise jitter should be used for each method.

For very small ensemble sizes, the white noise jitter yields lower RMSEs in most test series so far. This is why we used it in this high-dimensional test series.

I have to state again that there should be another case presented in which the LETKF fails and the S()R methods produce superior results.

Please see the discussion about general comment 2).

I very much like the promise of the LPFˆ x OT methods. However, I'd like to see the S(IR)ˆ xSR_a method of Penny and Miyoshi (2016) presented, which should give a nice balance between parallelizable computational costs and accuracy as measured by RMSE - which was the primary goal of the algorithm.

We originally did not select the $S(IR)^{\times}S_{PM}R_c$ for this high-dimensional test series for these reasons:

1. Given the results for the BV model in the coarse resolution configuration, with very small ensemble sizes this algorithm is outperformed by the algorithms using OT resampling.
2. This algorithm is slower than the other algorithm, because in two dimensions, computing the smoothing by weights is numerically expensive.
3. Optimal tuning parameters for this method are harder to find (both because there are more parameters and because the simulations are long).

Following your recommendation, we performed the simulation and reported the results in Fig. 22 and Table 5. The ration accuracy / computation time is not in favor of this method.

Finally, please keep in mind that the $S(IR)^{\times}S_{PM}R_c$ algorithm tested here is not the LPF of Penny and Miyoshi (2016) but our improvement thereof, which includes: a more general framework that can be applied to different types of resampling, a tapering function, a smoothing radius and a smoothing strength parameters, coloured noise regularisation.

Interactive comment on Nonlin. Processes Geophys. Discuss., https://doi.org/10.5194/npg-2018-15, 2018.

Printer-friendly version

Discussion paper