NPG Farchi and Bocquet 2018

General comments:

This manuscript takes nice steps in consolidating the literature in the field, comparing approaches, and offering new ideas in the development of localized Particle Filters. I do recommend that it be accepted for publication after revisions satisfying the following major points:

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.

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

3) In general, I find the algorithm names confusing. The first half of the paper uses a complex coding system, while the second half credits the authors who developed the methods. A more consistent and simpler naming convention would be nice from a reader's point of view, and should be used throughout. 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.

From what I can tell, the S(IR)^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.

Specific comments:

L28:

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

from 2008:

Kalnay, E., S.-C. Yang, 2008: Accelerating the spin-up of Ensemble Kalman Filtering. https://arxiv.org/pdf/0806.0180.pdf

and 2012:

Yang, S.-C., Kalnay, E., and Hunt, B. R.: Handling nonlinearity in Ensemble Kalman Filter: Experiments with the three variable Lorenz model, Mon. Weather Rev., 240, 2628–2646, doi: 10.1175/MWR-D-11-00313.1, 2012a.
https://journals.ametsoc.org/doi/abs/10.1175/MWR-D-11-00313.1

L39:

comma after hybridisation

L43:

[fewer] particles

L60:

degree[s] of freedom

L63:

geophysical system[s]

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.

91:

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+1lk. This is the goal at least, usually, to make a prediction. Perhaps you can say -

"The DA filtering problem consists in estimating pi_klk and pi_k+1lk with given realizations of y_k:0."

L97:

particle representatio[n]

L111:

What do you mean by "pure ensemble transformations"?

L170:

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

L177:

"… to more elaborate algorithms …"

L182:

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

L196:

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

L214/L217:

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…"

L223:

"…optimal importance proposal [density]…"

L227:

"… does not [primarily] come from …"

L238:

"… [elaborate] models …"

L252:

It seems awkward to begin a sentence with a variable name. Perhaps used instead:

"The quantity tao^2 would then be defined using"

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.

L281:

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

L285:

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

L332:

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

L367:

“and decrease[s] exponentially”

L371:

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

L386:

I think “hold” should be “held”

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.

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

L543:

“only [a] big ensemble”

L561:

“RMSE offers a” to “RMSE offer a”

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.

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.

L572:

“yield[s]”

L595:

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

P24-25:

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

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?

L651:

I’m confused how the higher RMSEs of the S(IR_SU)^xR algorithm indicates an efficient approach. Could the authors elaborate. What is the RMSE ratio used in Figure 9? Why does the figure caption say “RMSE” while the y-axis says “RMSE ratio”?

L 675:

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 x^i in (46) the columns of E?

L 697:

change “as following” to “as follows”

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)?

L 702:

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

L 740:

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)?

L 742:

Do you have the baseline RMSE values for the EnKF?

L 743:

“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.

L 771:

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?

L 794:

“local LET algorithm”

Is that redundant? Perhaps just say “LET algorithm”

L 804:

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

L 809:

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

L 816:

“dynamic[s]”

L 828:

“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.

L 835:

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

L 923:

“The SO formalism is elegant.”

This seems a strange characterization given that the next few sentences describe legitimate problems with the approach.

L 940:

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

L 968:

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.

L 970:

“one first need[s] to”

L 971:

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

L 1009:

“any distance that need[s] to be computed relative[] to the observation site…”

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.

L 1013:

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

1073:

“size Ne grows[,] the RMSE decreases”

1075:

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

1019:

“few but [dis]similar LPF algorithms”

Figure 21:

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

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.

I very much like the promise of the LPF^x OT methods. However, I’d like to see the S(IR)^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.