Response to Referees

Andrey A Popov et al.

April 27, 2022

1 General Changes

Another figure has been added to showcase the Lorenz '96 covariance matrix.

The notation concerning probability distributions has been largely simplified.

2 Response to the Editor

> ... He mentions that some of your results are in contradiction with results obtained by other authors in similar circumstances. This must certainly be mentioned in your paper. But it may be very difficult, if not impossible, to explain those contradictions.

We believe we have a simple explanation, namely the type of RMSE metric that is used. See the reply to his comments below. We believe that there is in fact no contradiction between our work and the paper cited in the review, and the results are consistent if the same RMSE formulas are used.

> he finds that your conclusion is too succinct. It is certainly succinct and, if you think you can say more (for instance, as concerns the possible limitations of your approach, or as to which difficulties could be expected in large dimension systems), please do so.

We have significantly expanded the conclusions to include all possible limitations and some insights as to what could be done in the future to further the research into this method and the ETPF in general.

> 2. I do not see the usefulness of putting indices to the symbol π when the latter designates a probability distribution.

We agree and have simplified the notation greatly.

> 3. Why use X^f in Eq. (1) and $\hat{X}^f$ in Eq. (2) ? What is the difference, if any ? (same remark concerning Xa in Eqs 1 and 3)

We have removed $\hat{X}^f$ and kept $\hat{X}^a$. We added a sentence to say that $\hat{X}^a$ converges in distribution to X^a. It exists to highlight the fact that the distribution for a finite ensemble is not exact, and that with perfect inference between the two empirical distributions we will not get perfect inference between the underlying random variables.

> 4. Eq. (4) $\pi(P|X) \to \pi(X|P)$
>
> 5. Eq. (5). Subscript j on rhs is useless (and confusing) (incidentally, I do not understand the presence of the denominator 1 − wa, T wa, in that equation. A brief explanation could be useful).

Both points have been corrected and carefully explained.

> 6. Eq. 7). Second sum runs from j=1 to N^a.

We fixed this value to N^f, as the ensemble of analysis weights is actually applied to the forecast ensemble, thus should have that size. It is a bit counter-intuitive, but this is how all particle filters operate.

> 7. Eq. (9). $X^f \to X^a$ (see also Eq. 14)

This is in fact X^f, as the transport is defined on the forecast ensemble distances.

> 8. Eq. (11), argument of the exponential. $R \to R^{-1}$ 9. Eq. (12). Symbol $\circ$ not defined (defined only on occasion of Eq. 46) 10. L. 357, lead $\to$ led

We agree with all of these, and have made the necessary changes.

3 Response to Referee 1

3.1 Lorenz 63 Test series

> This first test series with the Lorenz 1963 model is using the exact same setup as one of the experiments described by Acevedo et al. (2017), hereafter A17. However, the results reported in the present manuscript (Figs. 4 and 5) are different from the ones reported by A17 (left panel of Fig. 7.1). In particular, I noticed the following differences in score...

We have conducted experiments to determine why there are discrepancies, and came to the conclusion that this is again a matter of metric used. We use the spatio-temporal RMSE

$$\sqrt{\frac{1}{nT}\sum_{\text{time}}\sum_{\text{space}}\text{err}^2}, \tag{1}$$

while A17 uses the time-averaged RMSE

$$\frac{1}{T}\sum_{\text{time}}\sqrt{\frac{1}{n}\sum_{\text{space}}\text{err}^2}, \tag{2}$$

as evidenced by “the resulting time-averaged RMS errors” note in their manuscript. We believe that this is strong motivation for us to include the exact metric we used as a separate identifiable equation so that our results are more reproducible.

As we have shown the reviewer previously, there is no wide consensus between which version of the RMSE to use. Indeed the last author on A17 himself has used the other RMSE metric in other works, further confusing the matter.

In short, we do not believe there are any discrepancies between A17 and our work.

> “ In addition, I would like to come back on the choice of the rejuvenation factor. In general, the optimal rejuvenation factor depends on the ensemble size. For a small range of values of the ensemble size, such as [15, 35] as used by A17, using a constant factor for all values of the ensemble size may be a good approximation. For a larger range of values of the ensemble size, such as [5, 100] as used in the present manuscript, this approximation is less justified. Therefore, I think that the present test series should include a tuning of the rejuvenation factor which depends on the ensemble size, or at least test different values as done by A17. ”

We hope the reviewer agrees that the factor that we use should be close to optimal for some range of ensemble size between 5 and 100. As the ETPF performs worse for the whole range, we do not believe further exploration of this is warranted.

> “ Finally, I would like to mention that I appreciate the efforts that the authors have put into the improvement of their figures. I have one last remark: the dashed lines can not be distinguished from the plain lines in the legend of Figs. 4, 5, 8, and 9. ”

We have modified the lines in the legends to be a bit longer, thus making the distinction more clear.

3.2 Lorenz 1996 test series

> “ the LETKF curve in Fig. 8 does not seem to be correct: the RMSE should be lower than 0.3 with 5 members, close to 0.2 with 10 members, and lower than 0.2 (these scores can be found, for example, in the chapter on the EnKF of Ash et al., 2016, already cited in the manuscript); ”

As in the above, this difference is largely due to our use of a different RMSE measure coupled with all the other differences. As additional note of difference is with our GC implementation. We use the internal parameter θ that is not equal to one to more closely match Gaussian localization. This has been added to the text.

> “ the authors mention that the LETPF does not converge, but Farchi and Bocquet (2018) provide an illustration of the convergence of the LETPF in the exact same setup (red curve in Fig. 16 of their article), with lower RMSE scores as those reported in Fig. 8 of the present manuscript for the LFETPF(G) with 8 and 16 particles.
>
> ...
>
> My intuition is that these differences can be largely explained by the (very restrictive) choice of not tuning the localisation radius and the inflation or rejuvenation factor. ”

We agree with the reviewer on this point. The current study does not aim to optimally tune these factors through an exhaustive parameter search, however, we have taken a close look at the insight in the referenced work, and have done our best to choose parameters that give the LETPF the greatest chance of convergence through a lot of careful hand-tuning. As a result, the experiments have been modified in the text with a smaller localization radius.

> With the non-linear observation operator, the author conclude that the setup is "highly Gaussian". My impression while reading the text is that the purpose of this setup is precisely to be non-Gaussian. If we end up with a Gaussian setup, then this does not provide any added value compared to the first setup. In addition, there is a contradiction with the conclusion of the experiments with the linear observation operator: in a highly Gaussian setup the LETKF outperforms the LFETPF (which is expected and which is not what can be seen in the experiments with the non-linear observation operator).

We agree with the reviewer on this fact. This is purely an error on our part. We have completely revamped the nonlinear localization experiments to account for the facts presented therein.

3.3 Specific Comments

All minor comments have been addressed we hope this time.

> What kind of limitations does prevent the authors from showing P, which is a 40×40 matrix? If they want to, they could easily add a figure showing P, for example in a similar way as Fig. 6.

A figure showing the L96 $\mathcal{P}$ has been added.

3.4 Technical Comments

> In this comment, I mentioned that the discussion on the results and the conclusions are too short. This part of the comment has not been answered. The revised manuscript offers some discussion of the results, but in my opinion this is still not enough. Furthermore, the conclusions of the revised manuscript are even shorter than the original one!

We have expanded the conclusions to include many limitations of the work, and some more discussion on future possibilities. We have also expanded the discussion of the Lorenz '96 approach.

> However, having this discussion right here seems weird. Indeed, at this point, the particle filter has not been introduced, neither has the weight collapse phenomenon. For this reason, I would suggest to move this discussion to the end of the following paragraph

This has been changed as suggested.

> Also please correct the citation to Farchi and Bocquet, 2018.

The bib entry was auto-generated by Google. It seems to have been corrected there. We have corrected it in the manuscript as well.

> Once again, I repeat that "$\mathbf{R}$" should be replaced by "$\mathbb{R}$".

This has finally been addressed.

> Visibly my remark has been misunderstood. In my opinion, there are two possible ways of rigorously treating the additional information...
>
> ...in which case it should be clearly mentioned that P does not have the same meaning in Eqs. (1)-(6) and in Eq. (21);

It is both and neither of the points. The information contained in P is always the same. It is everything that you (the agent performing inference) knows. The difference is what information you choose to ignore. We added a sentence to the description of the standard ETPF that almost everything in P is typically ignored.