A stochastic covariance shrinkage approach to particle rejuvenation in the ensemble transform particle filter – Review report

13th October 2021

In this article, the authors derive a new particle filter algorithm. The algorithm starts by adding some additional members to the ensemble. These additional members are drawn from a normal distribution with a static covariance matrix. The algorithm then uses the ensemble transform algorithm of Reich (2013) to construct the analysis ensemble. The whole idea of this method is to replace the post-analysis regularisation process (called particle rejuvenation in this article) which is usually necessary with particle filtering. The algorithm is finally illustrated using three test series of twin simulations with the 3-variable Lorenz system.

Overall the paper is well written and easy to follow. The presentation of the method is appropriate and understandable. However, I have the impression that several aspects could be improved and that a key methodological aspect is avoided. The presentation of the experiments is correct as well, but a few experiments with a 3-variable system is not enough to make a convincing illustration. In addition, the experimental results are barely discussed and the conclusion is much too short.

1 General comments

1.1 Notation

Throughout the manuscript, the notation is inconsistent. For example, in Eq. (2) the argument if $X|P$ while in Eq. (3) it is $x|y,p$. Using a consistent notation would really make the manuscript clearer and hence help the reader. Furthermore, I strongly recommend to follow the usual conventions of the data assimilation community (which, if I am not mistaken, also coincide with the journal conventions):

- bold face uppercase for matrices (ex $\mathbf{M}$);
- bold face lowercase italic for vectors (ex $\boldsymbol{v}$);
- lowercase italic or greek letters for scalar quantities (ex n or α);

- uppercase italic for sizes (ex N).

1.2 Regularisation or particle rejuvenation

The entire method derived by the authors is designed as a sort of extension of the ETPF of Reich (2013), therefore I am not surprised that the authors adopt the same terminology. Nevertheless, one should keep in mind that "particle rejuvenation" is not a new method invented by Reich and colleagues, it is just a new fancy name for one of the regularisation methods that have been introduced in the 2000s by Musso and colleagues. See, in particular, the chapter "Improving Regularized Particle Filters" by Musso et al. in the book "Sequential Monte Carlo Methods in Practice" by Doucet, Freitas, and Gordon (isbn: 978-0-387-95146-1). This historical perspective does not appear in the manuscript and I think that this is missing.

In addition, I would like to mention that in my opinion, "regularisation" is a better name than "particle rejuvenation", because I think that it better describes what is actually happening (in practice the posterior pdf is regularised). Of course, I acknowledge that the authors should have the right to choose which name they use!

1.3 Particle filter and localisation

The curse of dimensionality, mentioned in the introduction, is one of the main obstacles to the application of PF algorithms to high-dimensional problems (see Snyder et al., 2008, doi: 10.1175/2008MWR2529.1). More recently, a lot of studies have tried to apply localisation techniques in the PF to circumvent the curse of dimensionality (see in particular the review by Farchi and Bocquet, 2018, doi: 10.5194/npg-25-765-2018), which has lead to successful applications of PF algorithms to high-dimensional problems.

In the manuscript, there is no discussion about the scalability of the new method, at a point that a naive reader could think that this new method could be applied as is to high-dimensional problems. This aspect must be clarified in the manuscript. In particular, I think that the author should explain whether the new method should (i) replace localisation or (ii) be used in conjunction with localisation.

1.4 Numerical experiments and conclusions

After reading the article, I am left with the impression that the numerical experiments are very brief. Nowadays, a couple of experiments with the Lorenz 1963 model is not enough to publish an article in a data assimilation journal like NPG. Therefore, I think that Section 5 (with the numerical experiments) has to be extended. At the very least, I think that a test series with the Lorenz 1996 model should be included, which would help illustrate a potential discussion about the scalability of the new method. Of course, implementing a PF algorithm with 40 variables (the classical size of the Lorenz 1996 model) is a challenge without localisation, which is why it is probably more reasonable to start with only 10 or 8 variables in the Lorenz 1996 model.

In addition, the results of the numerical experiments are barely discussed in the manuscript, and the conclusions are very short (only two paragraphs!). I think that this is clearly not enough and that the authors should provide appropriate discussion of the results and conclusions.

2 Specific and technical comments

L. 12-13 "The curse of dimensionality". At this point, the authors could cite Snyder et al. (2008).

L. 17 "by discarding information about the underlying dynamical system". In the EnKF, the information about the dynamical system is taken into account through the use of the dynamical model (for forecast) and the observation operator. Could the authors explain what they mean here?

L. 18 "that lives in $\mathbb{R}^n$". At this point, n is not defined. I would delay this aspect until section 2 where the different spaces are introduced.

LL. 19-20 "into our assumed posterior normal distribution". If both the prior and the likelihood are assumed Gaussian, then the posterior is Gaussian (this is not an assumption).

L. 21 In the reference list, two elements match the key "Popov et al., 2020". This should be corrected.

L. 23 I would suggest the following stylistic transformation "(ETPF) (Reich, 2013)" $\rightarrow$ "(ETPF, Reich, 2013)".

L. 25 "the ensemble limit". To my knowledge, this is not clearly defined in the data assimilation community (even though I agree that this is understandable). I would recommend to explicitly define this term with, for example: "in the limit of an infinite ensemble size".

L. 26 "This means that". The logical connection is incorrect here.

LL. 45-46 "and the supports of the probability densities π_{X^f} and π_{X^a} are subsets of the respective spaces". This could be reformulated because as is, one could understand that it is possible that the support of these pdfs are not subsets of the respective spaces.

LL. 50 and 54 The sum's limits are incorrect in Eq. (2) and (3).

L. 57 "ensemble of weights". I would suggest to name it "weight vector".

L. 58 "Using (3) and (4) empirical estimates of the posterior mean and covariance". This seems weird. I would suggest a reformulation.

L. 59 In equation (5), the authors use the prefactor $N/(N-1)$ to debias the sample covariances. However, for weighted sample the prefactor to debias the sample covariances is $1/(1-\mathbf{w}^\top\mathbf{w})$, which is equal to $N/(N-1)$ only in the case where the weights are uniform $\mathbf{w} = \mathbf{1}_N/N$. Could the authors justify this choice?

L. 61 "The goal of particle filtering (with resampling)". I would rather say that this is the goal of resampling.

L. 62 "the the posterior..." → "the posterior..."

LL. 65-66 "We impose that the empirical mean (5) is preserved by (6)". This is in general not possible with classical resampling algorithms. Of course, this is possible when using a linear ensemble transformation like Eq. (8) but at this point in the manuscript, Eq. (8) is not yet introduced!

L. 71 "$\mathbf{T}^* \in \mathbb{R}^{N^\mathrm{f}\times N^\mathrm{a}}$" At this point, it could be interesting to remind the reader that $\mathbf{T}^*$ has positive coefficients.

L. 73 "$\mathbf{T}^\top\mathbf{1}_{N^\mathrm{f}} = \mathbf{1}_{N^\mathrm{f}}$" → "$\mathbf{T}^\top\mathbf{1}_{N^\mathrm{f}} = \mathbf{1}_{N^\mathrm{a}}$"

L. 78 "$X^\mathrm{a} = \Psi(X^\mathrm{f})$, which has..." → "$X^\mathrm{a} = \Psi(X^\mathrm{f})$ has..."

L. 94 "the factor τ". The authors could mention that τ is usually called the bandwidth.

L. 99 The second line of Eq. (13) is just the transpose of the first line, or did I miss anything?

L. 97-99 It is true that the extra term ensures that the regularisation noise has zero mean. However, the author should mention that this extra term does modify the sample covariance of the noise. The same holds for Eq. (26).

L. 125 "THis" → "This"

LL. 125-126 "This is because we are now incorporating more prior information P in the form of climatological information". From what I understand, such additional information was missing until now. Hence Eq. (2) – and the few following equations – should be written "$\pi_{\hat{X}^\mathrm{f}}(X)$" instead of "$\pi_{\hat{X}^\mathrm{f}}(X|P)$", right?

LL. 134-135 "is assumed to be the sample mean of the dynamic ensemble" I would mention here that this choice is necessary to preserve the mean of the augmented ensemble.

L. 136 "by construction and (13), thus requiring that only the synthetic ensemble anomalies need to be determined". This is hardly understandable. Y would suggest a reformulation.

L. 158 Please define the "$\wedge$" symbol in Eq. (28).

L. 162 "Note that if $\mathcal{P} = \mathbf{\Sigma}_{X^{\mathrm{f}}}$...". How is the division by zero in Eq. (28) handled in this case?

L. 162 "In such a framework the scaling parameter...". I think that a separation is needed here to indicate that this does not apply to $\mathcal{P} = \mathbf{\Sigma}_{X^{\mathrm{f}}}$.

LL. 170-171 I understand that the scaling of $\mathcal{P}$ has no impact on the definition of $\mathbf{\Sigma}_{\mathcal{X}^{\mathrm{f}}}$ defined by Eq. (25), but does it have an impact on γ defined by Eq. (28)?

L. 175 "where the optimal transport matrix $\mathbf{T}^* \in \mathbf{R}^{(N+M)\times N}$ is computed by solving (9)". This is a very concise description and I think that additional description is needed, because this is one of the core elements of the new method. At the very least, it should be mentioned that a posterior weight is needed for all members of the augmented ensemble, and that the formulation of (9) needs to be adjusted to take into account non-uniform prior weights. In addition, note that "$\mathbf{R}$" should be replaced by "$\mathbb{R}$".

LL. 181-182 "In effect we are able to avoid ensemble collapse by enhancing the empirical measure distribution (32) with new prior information". From a theoretical perspective, this has not been proven. This is only illustrated in Section 5 using numerical illustration with a 3-variable model. At this point, the lack of discussion about scalability really hurts (see general comment in section 1.3).

LL. 227 Why not use the time-averaged RMSE, defined as

$$\mathrm{RMSE}(\mathbf{x}^{\mathrm{t}}, \bar{\mathbf{x}}^{\mathrm{a}}) \triangleq \frac{1}{T}\sum_{i=1}^{T}\sqrt{\frac{1}{n}\sum_{j=1}^{n}\left([\mathbf{x}_i^{\mathrm{t}}]_j - [\bar{\mathbf{x}}_i^{\mathrm{a}}]_j\right)^2}, \tag{1}$$

which is the indicator used in most articles in data assimilation?

LL. 230 "with the optimal rejuvenation factor of $\tau = 0.04$". I highly doubt that $\tau = 0.04$ is optimal for all values of the ensemble size N. In order to make a fair comparison, the value of τ should be optimally tuned for each ensemble size N. If not, we give an unfair advantage to any of the method.

LL. 235 "Results in Figure 1 show...". I think that this test series (and the second one as well) is lacking a baseline. For this small 3-variable model, the baseline could be the score obtained with a classical PF (for example the SIR or Bootstrap filter) with a very large ensemble (typically more than 10^3 particles) and with optimally tuned regularisation.

LL. 241-243 "For a low ensemble size... as compared to the ETPF". From what can be seen in Fig. 2, there is a difference between, for example, $\alpha = 1$ and $\alpha = 1.2$ (the latter being more accurate). It is possible that this difference falls into the variability between the 20 independent runs, but this is not clearly explained in the text or in the figure.

LL. 259-260 "We believe that the stochastic covariance shrinkage approach to importance sampling can be used not just for particle rejuvenation in the ETPF, but in other particle filters as well". Let us take the example of the most basic PF, the SIR filter. During the resampling step, some particles (typically those with low weight) are discarded and replaced by other particles (typically those with high weight). If applying the new proposed method, this would unavoidably lead to replacing original particles by synthetic particles, which is probably not something that is desirable. With this small example, I hope that I have convinced the authors that more discussion is needed here regarding the application of the new method to other PFs than the ETPF.

Figs. 1 and 2 Some of the lines cannot be distinguished when the manuscript is printed in black and white. This should be corrected. In addition, I would recommend a log scale for the x-axis and I would recommend to show the grid for clarity.