Review manuscript number npg-2020-10: reviewer comments to the authors

May 12, 2022

1 Reviewer general comments to the authors

This paper is about the use of an hybrid covariance data assimilation scheme within the Canadian Precipitation Analysis. The paper is of very good quality, very well written, nice to read, with a very rigorous presentation of the elements of the study, the observations, the model, the scores used, the results etc. Despite using a methodology that is quite familiar now, an hybrid covariance scheme, the paper is quite innovative as it presents, to my knowledge, one of the first application of an hybrid scheme to precipitation forecasts. The authors provide a very convincing demonstration of the pre-eminence of the hybrid scheme over an Optimal Interpolation scheme only. This demonstration is based on different scores like the normalized root mean square error or the scores derived from a contingency table for binary events like for example the equitable threat score or the false alarm ratio and also with the comparison against ST4 data. In particular they show that the hybrid performs better than the OI only during both winter and summer seasons and in particular confirm previous findings from [Wang et al., 2008] of the effectiveness of the hybrid scheme with sparse observations networks.
All in all I had a hard time finding anything relevant to say about that article just because it is so well written. That said, if I was to say one flaw is that the authors do not emphasize enough the fact that the hybrid scheme not only performs better than the full static case but also better than the full dynamic case. They actually talk about it only in the conclusion. I imagine that the authors are interested in the improvement of their data assimilation system compared to the OI version of it, but they must understand that the fact that the hybrid performs also better than the EnKF for precipitation forecasting can be of great interest for the rest of the community. I would suggest the authors to complete their analysis in both sections 7.1 and 7.2 by commenting further about the full dynamic case, and also to complete the figures 3 and 4 by adding the line of the case $\beta = 1$. I sincerely believe that it would help improving the paper and that it would not require too much work from the authors.

2 Specific comments

Page 3, line 76: "β is comprised between 0 and 1, ensuring that the total background error covariances are conserved". This is true if the matrices $\mathbf{P}_{\mathrm{OI}}^{\mathrm{b}}$ and $\mathbf{P}_{\mathrm{d}}^{\mathrm{b}}$ provide "independent estimations of the true background error covariance matrix", [Ménétrier and Auligné, 2015]. So, I

would be grateful to the authors if they could go a little bit more through that point, and explain why they think that the matrix $\mathbf{P}_{\mathrm{OI}}^{\mathrm{b}}$ they build represents an estimation of the true background error covariance matrix.

Page 4, eq. (4): what do the notations z_i and z_j stand for? Is that the value of the innovations at locations i and j? Please, add the definition of z_i and z_j after eq. (4).

Page 4, line 95: I am not familiar with variographic analysis, in my understanding, you use eq. (4) to fit it on the empirical semivariogram and determine an optimal value of σ_{OI}^2. If I am correct, please can you add a sentence clarifying that point here (even though this is also specified page 5, line 119), otherwise it is unclear why you introduce this function γ here.

Page 4, lines 105-106: I am aware that this comment is obvious but in order to speed-up the computation you could also perform the analysis for each grid-cell in parallel. I guess it would not require a lot of modifications to the existing version of the code. I have no idea how much the computation efficiency is critical in this case though.

Page 7, line 185, eq. (13): that criteria for rejecting observations is baffling to me, I feel like I missed something. Eq. (13) basically means that if the absolute difference between the Box-Cox of the observation and that of CaPA is smaller than a specific threshold then the observation is rejected. While you would like to reject observations that are too "far" from the model to avoid too strong updates. Can the authors correct that point? Or just let me know if I missed something.

Page 7, lines 190-191: if I am not mistaken, I have counted so far 3 quality checks, maybe it could be an idea to summarize them in a table.

Page 7, line 199: "seamless precipitation fields", I do not know here if this is my english that is at fault or my limited knowledge of precipitations, but I do not know what is a "seamless precipitation field", can you precise it between parenthesis maybe, or add a reference if necessary?

Page 9, lines 256-263: based only on the shape on the curves it seems that the hybrid approach brings potentially a dramatic improvement compared to the OI only based approach. Though, a quick calculation shows that the relative reduction of NRMSE of the hybrid approach for the optimal value of β is rather limited with around 3.4%, 2.3%, and 7% reduction of NRMSE, respectively for fig. 2-(a), 2-(b), and 2-(c) (though I must say that in the case of winter 7% is quite good). I would then recommend the authors to go a little bit more through that in that paragraph.

Also, the authors have missed an opportunity here to deepen their analysis and show the benefits one could retrieve from the use of an hybrid scheme, not only compared to the full static case, $\beta = 0$, but also compared to the full dynamic case, $\beta = 1$. Indeed, the authors do not mention that case while at the same time they show that the hybrid performs better than the EnKF only. What I mean is that if the hybrid was performing better than the static case only but no better than the dynamic case it would be of no interest. So, despite the reference case of the authors being $\beta = 0$ I would highly recommend that they treat the case of the standalone EnKF only for the reason aforementioned and that they complete that paragraph accordingly.

Sections 7.1 and 7.2: the authors definitely have to talk more about the case $\beta = 1$. The authors could complete the figures 3 and 4 by adding the curve for $\beta = 1$ and then complete their analysis by emphasizing the fact that the hybrid also improves the results compared to the full dynamic case. I do believe that it would not require too much work from the authors while improving the quality of the paper.

3 Technical corrections

Page 3, line 73: repetition: "the the background field".
Page 4, line 96: $\mathbf{P}^{\mathrm{a}}_{\mathrm{OI}}$, is it an error in the notation? Should not it be $\mathbf{P}^{\mathrm{b}}_{\mathrm{OI}}$?
Page 4, eq. (6): you did not define what is $\mathbf{A}$.
Page 5, line 126: I would recommend not to write "(1) minus..." but "1 minus". The notation (1) is misleading and can make think about the numerotation of an equation.
Page 6, line 155: the acronym SYNOP is not defined, does it stand for synoptic?
Page 9, eq. (16): it seems that there are a few mistakes in the writing of eq. (16), I guess eq. (16) writes:

$$FSS = 1 - \frac{\frac{1}{N_y}\sum_{i=1}^{N_y}\left(f_a(i) - f_o(i)\right)^2}{\frac{1}{N_y}\left[\sum_{i=1}^{N_y} f_a(i)^2 + \sum_{i=1}^{N_y} f_o(i)^2\right]} \tag{1}$$

Page 11, line 296: repetition: "the POD slightly was slightly deteriorated".
Page 14, line 425: repetition: "for the observation density observations"

References

[Ménétrier and Auligné, 2015] Ménétrier, B. and Auligné, T. (2015). Optimized localization and hybridization to filter ensemble-based covariances. Monthly Weather Review, 143(10):3931–3947.

[Wang et al., 2008] Wang, X., Barker, D. M., Snyder, C., and Hamill, T. M. (2008). A Hybrid ETKF – 3DVAR Data Assimilation Scheme for the WRF Model . Part I : Observing System Simulation Experiment. (Lorenc 2003):5116–5131.