Reviewer report for gmd-2023-170 "GHOSH v1.0.0: a novel Gauss-Hermite High-Order Sampling Hybrid filter for computationally efficient data assimilation in geosciences"

This work proposes an ensemble filter (GHOSH) that conducts sampling in such a way that the resulting sample statistics can match the moments of a target distribution up to a specified order. The moment-matching trick is based on Hermite polynomial approximations to the underlying functions, whereas the target distribution is set to be Gaussian. The derived filter is tested in two examples, both indicating that the GHOSH outperforms an existing filter (SEIK) for the experiments conducted in the current work.

The manuscript is clearly written and reasonably organized in general. Below is a list of minor-to-moderate issues spotted in the current manuscript.

Spotted issues

1. Page 1 –
    - Line 2: Consider replacing "one of" by "among" or something similar, since "algorithms" is the subject.
    - Line 20: What does "a higher order of convergence" mean here?
2. Page 2 –
    - Line 42: "Montecarlo" → "Monte Carlo".
    - Line 51 – 52: Rephrase the sentence "the second order approximation is more effective the closer the ensemble members are to each other, thus, the larger the ensemble spread the worse will be the approximation error in the mean computation."
3. Line 58 – 59, Page 3: The "$2r + 1$ ensemble members" requirement does not appear exact. For the unscented transform, one can use either principal

component analysis (PCA) or truncated singular value decomposition (TSVD) to reduce the number of ensemble members, see
https://doi.org/10.1175/2008JAS2681.1
https://doi.org/10.1016/j.physd.2008.12.003

It may be worth discussing the similarities and differences between the ideas used to control the number of ensemble members in the aforementioned works and the current manuscript.

4. Line 109 – 117, Page 5: The discussion on the extension of GHOSH to more generic distributions makes sense. A missing part, however, is that the authors did not explain why they confine themselves to Gaussian distribution in the current work, and what could be the challenges for the generalization of GHOSH to more generic distributions

5. Eq. 38, Page 13: The notation w.r.t the Q_i component is somewhat confusing. I guess it should be $(Q_i^p)^{-1}$, but it looks like Q_i^{p-1}.

6. In the experiments w.r.t the Lorenz96 model, localization does seem used. What is the reason behind this setting?

7. Line 434 – 435, Page 19: Why "it implies that the PCA measures the Pearson correlation"?

8. Line 490, Page 21: If I've understood correctly, the "best" label corresponds to the configuration that leads to the best DA performance. If so, then in Figure 4 one should use one block to represent it, and I don't see the point to use a single row for the representation.

9. Line 661, Page 33: "an higher" $\rightarrow$ "a higher".