1 General response

My overall impression of this work is positive, I think that the authors have done a nice work of outlining a novel data assimilation (DA) framework that is distinguished from a variety of existing ones within the Julia language by focusing on a scalable implementation of particle filtering. I think that this fills a unique gap in the development of DA methodology and shows a lot of potential for research and possibly operational purposes when it is sufficiently mature. However, I do think that there is currently room for improvement in the manuscript including its discussion of the more delicate aspects of particle filtering for DA, and with respect to the framework's numerical validation and proof-of-concept. I outline these points in the major revisions section and I discuss minor items in the section thereafter.

2 Major Revisions

1. Section 3.2. State Model Error. There are many important assumptions about the state model error that I think require additional discussion. It is a critical point that the state process model is forced stochastically, as the types of stochastic filters implemented in this iteration of the framework (boot strap and optimal importance sampling) strongly depend on this assumption to maintain sample diversity. Standard multinomial resampling will lead to ensemble collapse if the empirical filtering measure is a weighted sum of Dirac measures when there is no state process randomness to drive the resampled particles apart. However, many geophysical models make no explicit assumption or use of a stochastic paramterization or of a noise processes. Therefore, I feel like this section requires an expanded discussion of the implementation of the stochastic forcing that is compatible with this framework for typical geophysical models. There is an example use-case with the SPEEDY model which is well-appreciated, but a more detailed discussion of the example use-case and a more general discussion of the implementation of random forcing for other state process models would strongly enhance the impact of the manuscript.

2. Section 3.2. Observation Model.

   > "The number of observations and the indices of the observed variables within a multiple dimensional state space model are needed. The locations of the observation stations can be passed within a simple .txt file. The observations can come from an online integration of the state space model or read in from file/sensor."

   This quote leads me to believe that the observation models which are compatible with this framework are assumed to be sparse direct measurements of the state model process. If this is the case, I would like this to be stated directly and to have a discussion of how indirect nonlinear measurements will be integrated into future iterations of the framework, as that would be a major limitation for use in real-world DA. If I have misinterpreted this statement, I would like to have a clearer discussion of how indirect nonlinear measurements are currently integrated into the model-agnostic framework.

3. Localization. The scope of the package aims to be scalable to high-dimensional operational DA. Please give some word about localization in this framework, as this is generically non-trivial to implement, and more so in a model-agnostic framework. However, without including a means for localization at a primal level in this framework, the scalability of the framework may be intrinsically limited, as localization is utilized in virtually every form of currently operational DA as a means to regularize the filtering problem.

4. Numerical Validation, Section 6.1.

   > "As the tsunami model implemented here is linear and Gaussian one can compare the obtained distributions with a ground truth Kalman filter."

   This demonstration should be included here for the reason stated above, and I don't belive this is currently in the manuscript. Please make a direct empirical validation of the particle filter methods versus the optimal Kalman filter in this context for use as a calibration metric and to understand the relationship between sample size and filter performance versus the optimal estimator.

5. Numerical Validation, Section 7.

    > "As stated in the introduction, one of the key benefits of particle filters is to provide the promise of non-linear and non-Gaussian DA. To highlight this sample distributions of the surface pressure at various observation locations at different time points are shown in Fig. 8. The distributions across the n = 256 particles exhibit heavy tails towards the true surface pressure at the given locations."

    This demonstration of heavy-tails is well-appreciated as this does support the hypothesis that including higher-order information in the estimation of the filtering measure will improve its empirical representation. Nonetheless, the current demonstration I think has missed the opportunity to make a direct validation versus, e.g., a standard flavor of the EnKF as discussed in the surveyed literature. While the current numerical demonstration makes the case that a linear-Gaussian approximation is inaccurate, it doesn't verify that it is inadequate. If the goal is to introduce a framework that addresses intrinsic inadequacies of widely used DA schemes with linear-Gaussian assumptions, a direct demonstration of this point would make a much stronger case for the impact of this framework.

3 Minor Revisions

1. Line 25. This references variational DA in the form of 3D- and 4D-VAR, but it makes no citation to related literature. Please expand this with relevant citations and include some references to more modern ensemble-variational methods. See, e.g., Bannister [2017] and refernces therein.

2. Lines 61 - 74. While differing in the intended scope, the literature review does not discuss DataAssimilationBenchmarks.jl and its two key references [Grudzien and Bocquet, 2022, Grudzien et al., 2022]. Please include these in your literature review.

3. Table 1. The table is not currently consistent with the literature review – please include all referenced software packages in Table 1, including EnsembleKalmanProcesses.jl and DataAssimilationBenchmarks.jl.

References

R. N. Bannister. A review of operational methods of variational and ensemble-variational data assimilation. *Q. J. R. Meteorol. Soc.*, 143(703):607–633, 2017.

C. Grudzien and M. Bocquet. A fast, single-iteration ensemble Kalman smoother for sequential data assimilation. *Geoscientific Model Development*, 15(20):7641–7681, 2022. doi: 10.5194/gmd-15-7641-2022. URL `https://gmd.copernicus.org/articles/15/7641/2022/`.

C. Grudzien, C. Merchant, and S. Sandhu. Dataassimilationbenchmarks.jl: a data assimilation research framework. *Journal of Open Source Software*, 7(79):4129, 2022. doi: 10.21105/joss.04129. URL `https://doi.org/10.21105/joss.04129`.