ParticleDA.jl v.1.0: A distributed particle filtering data assimilation package

Response

July 30, 2023

We would like to thank the reviewers for their helpful comments and feedback; our responses to their main points are outlined below and an updated version of the article with highlighting of the revisions made in response to the comments has been made available.

1 Reviewer 1

1.1 Main comments

> Line 312: My only major comment was triggered by this statement: "It can be seen that the areas of greatest percentage error coincide with areas that lack observation stations." I want to challenge this statement as I don't see such a strong coincidence. Instead the error patterns seem to be dominated by patterns of midlatitude weather systems, which is what you would expect to see if you compared two snapshots of surface pressure in the same model run at different times. In fact, I don't see sufficient evidence that the assimilation is actually working. It probably is, but it's hard to tell unless you compare also with a run without any assimilation. A number of further plots would help to make this section more complete.

Author's response: Please see the updated Figure 8, where we have produced the suggested four sub-figures. We have removed the quoted statement and updated section 7.1.The reviewer is correct in stating that the largest errors in the assimilation run appear to be dominated by mid-latitude patterns. The bottom subplots showcase the benefit of the assimilation. As suggested by the reviewer we have plotted the time averaged error for the mean of the assimilated ensemble and that of an ensemble where no assimilation has occurred. Overall, the assimilated ensemble mean exhibits a lower error.

1.2 Minor comments

Our thanks to the reviewer for pointing out the mentioned typos; these have now been corrected.

1. Line 221: Could you add one or two more sentences to elaborate on the significance of the ESS? Why is this an interesting thing to note?

   Author's response: The estimated sample size (ESS) gives an indication of the level of degeneracy in the particle weights; we have added more information about ESS and its usefulness as a metric in Section 2.2.

2. Section 6.2: Considering, say, the 32 ranks per node case, the biggest test runs on only 4 nodes (128 / 32). Yet the parallel efficiency has already dropped to only around 25%. This is a much bigger drop that I would expect, intuitively. Am I missing something? Scaling tests often run into the hundreds of nodes before encountering such limits.

   Author's response: We agree with reviewer that the drop off seen in the previously included parallel efficiency results as the number of ranks increased was unexpectedly quick and we have spent considerable time improving the parallel efficiency of the package. A major factor in the previous loss in performance was found to be load imbalance across the threads. The problem has been partially remedied by using a dynamic task based thread allocation instead of a static thread model. New scaling results are now included where the model is run on up to 16 nodes on ARCHER2. The new results showcase near optimal weak scaling efficiency for single node cases but with a drop off in performance when going to multi-node runs. Therefore, future work for the package will focus on an improved load imbalancing across nodes.

3. Figure 6, left: Could you clarify which measure of weak scaling parallel efficiency you employ here? I am guessing it is $E(N) = T(1)/T(N)$, where $T(i)$ is the wall time for running on i processors.

Author's response: We are using the following defined weak scaling efficiency: $E(N) = T(2)/T(N)$. This has been clarified in Section 6.2.

4. Line 310: "The standard deviation of the model and observation errors are set to 1 and 10 hPa respectively." — isn't the observation error already stated on line 309? Also what is the "model error" in this case? I thought that the SPEEDY model is integrated without a model error term?

Author's response: In order to clarify the model setup for the SPEEDY case we have now provided additional information in Section 7.

2 Reviewer 2

2.1 Main comments

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 (bootstrap 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 parameterization 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.

Author's response: We agree with the reviewer of the importance of the assumption of the state process model being stochastic for the proposal distributions used here to maintain diversity in the ensemble, and have added text to emphasise this point in Section 2.1. We have also added a paragraph discussing the requirement for the state to evolve stochastically over time when introducing the state space model formulation assumed in Section 2, and noted this may not the case for the usual formulation of geophysical models. For each of the example models considered which are based on solving a system of ordinary or partial differential equations (Sections 5, 6 and 7) we have also included an explicit description of how stochasticity is introduced in the state dynamics.

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.

Author's response: We thank the reviewer for alerting us to this inconsistency in the text; the quoted text was originally written for an earlier version of the package which did indeed assume sparse direct measurements of the state model process at a set of point locations. The package now allows for more general observation models, with filtering using the bootstrap proposal able to be applied to any observation model compatible with Eq. 1 (that is for which the observations at a particular time index depend only on the current state and for which the conditional distribution given the state has a tractable density function), which would allow for example non-linear observation operators. For a filter using the locally optimal proposal, a more restricted linear-Gaussian observation model (Eq. (3) in the paper) is assumed to ensure the local optimal proposal distribution is analytically tractable. While in this case the observations are assumed to depend linearly on the state, they do not necessarily need to correspond to direct (noisy) measurements. We have clarified these points in the text and removed the previous inconsistent description.

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.

Author's response: The authors agree with the reviewer of the importance of spatial localization in scaling particle filter based data assimilation to high-dimensional state space models, specifically spatially extended models with densely distributed point observations, and adding support for this to ParticleDA.jl is one of our priorities for future development of the package, as we now mention in both the introduction and conclusion sections of the paper. With regards to implementing this in a model-agnostic fashion, while we agree this is non-trivial, we believe this should be feasible to do using a similar approach to that currently implemented (and described in Section 3.2 in the revised paper) for the locally optimal proposal for state space models with linear-Gaussian substructure. Specifically, the base interface required to be implemented by models can be extended with functions exposing additional structure that we wish to exploit in a particular class of filters. Specifically for spatially localized particle filters we would need functions to evaluate (i) the spatial distance between the observation point associated with an index of the observation vector and the spatial mesh node associated with an index of the state vector and (ii) the spatial distance between the spatial mesh nodes associated pairs of indices in the state vector. One of the authors has previously worked on implementing spatially-localised particle filtering methods in a model-agnostic data assimilation framework in Python (`https://github.com/thiery-lab/data-assimilation/`) and we plan to use the design there to inform the implementation in ParticleDA.jl.

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 believe 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.

Author's response: This has indeed already been carried out, with the results plotted in Figure 6 which shows the error in filtering estimates of the mean at each observation time *as compared to the true mean computed using a Kalman filter*. Further clarifying information has been added to Section 6.1 and the axis labels in Figure 6 updated to make this clearer.

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.

Author's response: We agree that this is an important point. However, we deem it to be beyond the scope of the work outlined here. We have made explicit reference to the work of (Miyoshi et al., 2014 and Konda and Miyoshi (2019)) in Section 7.1, where a very similar experimental set-up was used with an ensemble Kalman filter to produce non-Gaussian distributions. However, a key difference are the relative ensemble sizes, with 256 particles used in our case to illustrate the heavy tails versus an ensemble of 10,240 members used in Miyoshi et al., 2014 and Konda and Miyoshi (2019).

2.2 Minor Comments

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 references therein.

   Author's response: Additional references have been added.

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.

   Author's response: The additional references have been added.

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.

   Author's response: The table is a summary of the literature review with only Julia packages which are similar in design and purpose to ParticleDA.jl included.