Dear reviewer and NPG community,

Thank you for reviewing our paper. To foster the possibility of having constructive discussions about our work[1], we first reply and suggest action items to the seven "Significant issues" identified by the reviewer as listed under "General comments". In a subsequent post, we would like to take the opportunity to provide point-by-point answers to all your comments, including the "Specific comments" and "Technical corrections".

Before getting to the listed issues, we would like to point out that we clearly stated the *scope of the paper*, including four clearly marked research questions in L26 – L49, which we return to and answer at the end of the paper on L658 – L685. Furthermore, the *contribution of the paper* is described on L79 – L91 in Section 1.2 "Contribution and Outline". As far as we can see, none of the reviewer's general comments seem to discuss or acknowledge the scope of our research, the research questions we aim to answer, or the novel contribution we claim to bring to the scientific community. We are disappointed, and in fact a bit puzzled, that these elements are not discussed in this review.

Rather than reviewing our paper, the reviewer criticizes the established building blocks of our work. Hence, as will be clear from our reply to each of the raised "significant issues", we point out that several of the reviewer comments are outside the scope of our paper on multi-level Monte Carlo and assimilation of sparse ocean data.

Additionally, we point to NPG's aims & scope[2]. The journal is "devoted to breaking the deadlocks often faced by standard approaches in Earth and space sciences", and to apply innovative concepts and methodologies to address the complexity in geoscience systems. We believe that our application of a multi-level ensemble Kalman filter applied to a simplified ocean model fits very well into this. In light of this, we would like to invite the editor to also participate in this discussion.

Yours sincerely,
Florian Beiser
On behalf of all the authors

[1] as encouraged by NPG, see `https://www.nonlinear-processes-in-geophysics.net/peer_review/interactive_review_process.html`

[2] `https://www.nonlinear-processes-in-geophysics.net/about/aims_and_scope.html`

# Replies to General comments

## 1. Assimilation of Unavailable Observations

**Reviewer:** *This paper assimilates unavailable observations of zonal and meridional momentum.*

**Reply:** In our work, we use an identical twin experiment to assess the applicability of the multi-level ensemble Kalman filter (MLEnKF). Contrary to most of the literature on multi-level Monte Carlo, we apply the method to a shallow water equation (SWE) model which is more relevant to oceanographic applications than previous examples that are of a theoretical flavor, see the original presentations Hoel et al. (2016); Chernov et al. (2021). The setup uses sparse observations of a subset of the physical model variables, i.e. the momentum at 50 locations in the domain. This design is in agreement with the experiment's purpose.

In the introduction, L24 – 34, we explain how our work is motivated by search-and-rescue (SAR) operations at sea, and how ensembles of simplified ocean models can be used complementary to available operational circulation forecasts. The in-situ observations mentioned in this regard, can be made by local buoys or drifters (such as Rabault et al. (2022)) released from the SAR vessel. To assimilate such observations into SWE models, they must be mapped from observational space (velocity) to model space (momentum), where the associated observation error accounts for the representation error in this mapping. As a step towards these types of applications, we have in our work shown the applicability of MLEnKF for a relevant model through an identical twin experiment.

**Action item:** We will make the purpose of our experimental setup clearer in the text. We will also improve the description of our motivation in the introduction to better show the relevance of our experiments.

## 2. Inconsistent Localization

**Reviewer:** *$P^f H^T$ localization is applied only to the first part of the Kalman gain, despite the presence of $HP^f H^T$ in the inversed part.*

**Reply:** We use Kalman gain localisation which has been commonly done in data assimilation, see e.g. Chen and Oliver (2010) or Chapter 10.4 in Evensen et al. (2022). In comparison with other localization approaches, this type is natural for our particular situation with sparse data and a large model space (see also our reply to Comment 3).

**Action item:** We will improve text description and add references on this topic, which is established in the literature.

## 3. Challenges in Practical Implementation

**Reviewer:** *Formulation based on the perturbed observation method results in an exceedingly large size matrix of the Kalman gain, posing challenges for practical system implementation and difficulties in parallel computation.*

**Reply:** A characteristic of our case is sparse observational data, meaning that the matrices required in the Kalman gain are not problematically large. Within this regime, the computationally expensive part is in fact the forward calculation of the differential equations of the state (and particularly so at the fine-resolution level). This is why we employ multi-level methods combined with GPU-accelerated models.

It is not the intention that our implementation of the MLEnKF will replace the operational data assimilation systems for integrating all available data. Instead, we study tools and methods capable of assimilating subsets of observational data within ensembles, in this case at multiple levels. We argue that ensembles of (efficient) simplified models has a role to play by complementing the large operational model systems for specialized applications (such as sparse data in SAR operations). See our introduction, L29 – 34. We also write in our abstract L4 – 5, *By applying a multi-level ensemble Kalman filter for assimilating* ***sparse observations*** *of ocean currents....*

Due to this, the reviewer's comment falls outside the scope of our paper. We also point out that the contribution of our work is to make *a new step on the path of making multi-level data assimilation feasible for real-world oceanographic applications* (Abstract, L7 – 8), and we do not hide the fact that there are more steps to be taken in the future for advancing the applicability of MLEnKF further.

**Action item:** We will clarify both in the introduction and in the Kalman updating expressions that our work is within the regime of sparse observational data.

## 4. Numerical Instabilities and Negative Eigenvalues

**Reviewer:** *Negative eigenvalues lead to numerical instabilities when applying eigenvalue decomposition to the inversed matrix in the Kalman gain. The likelihood of implementing the MLMC–based EnKF successfully is questionable.*

**Reply:** Other researchers, before our paper, have recognized the challenges of estimating covariances within the framework of multi level Monte Carlo (MLMC) methods, with no guarantees for getting positive definite covariance matrices (all eigenvalues larger than 0), see e.g., Maurais et al. (2023) and Shivanand (2023) and the references therein. Building on the established understanding, we honestly discuss this challenge in Section 3.3 (L375 – 376):

> *While the empirical measure associated with a classical Monte-Carlo*

*estimator is positive in the sense that all Dirac contributions have a positive sign, this is not the case for multi-level estimators where the differences between the levels can introduce negative values.*

When other researchers in future work aims to further develop a MLMC-based data assimilation scheme that depends on eigenvalue decomposition, we believe that our discussion in Section 3.3, along with that of others, will be a helpful resource for them to find a way to understand this challenge and formulate solutions.

As we discuss in Section 3.3 (L391 – 392), the issue with negative eigenvalues is infrequent (less than 1:10 000 times of the assimilated observations). For sparse observations in particular, we show that the known problem can be handled in a pragmatic way, and that potential errors introduced in the process are negligible.

**Action item:** We will add references to the appropriate literature early in Section 3.3, to help explain that this is a known challenge.

## 5. Unreasonable System Settings

**Reviewer:** *Because of the chosen system settings, such as the absence of stochastic external forcing, filter divergence is likely to occur as seen in decreasing the ensemble spread kept over the assimilation period. Therefore, the experimental period of 10 days was too short to conclude. In addition, there are other numerous issues such as inconsistency between the observation errors used for generating observations and the prescribed observation error covariance matrix.*

**Reply:** While we agree that our experimental setup is not described well enough (as illustrated by some of the reviewer's "Specific comments"), we do not agree with the statement that we cannot make any conclusions.

The aim of the experimental design was to generate ensembles of arbitrary sizes with turbulent and (initially) chaotic behavior, in which drift trajectory forecasts made through pure Monte Carlo experiments would have too high uncertainties to be helpful, but where successful data assimilation improves the forecasts significantly. Since the mentioned decrease in the ensemble spread occurs in both the single- and multi-level experiment (Fig. 6 of our paper), we do not see how this affects the scientific contribution of our paper on the applicability of the MLEnKF. Moreover, since our work is motivated by SAR operations at sea, assimilation and forecast periods of seven and three days, respectively, are already in the upper time-range relevant for this application.

This experiment was first outlined in a paper presenting a GPU-accelerated implicit equal-weight particle filter (Holm et al., 2020), which we refer to in our paper. Here, we should have pointed out for the reader that more details concerning the experiment could be found in that paper, along with a pure

Monte Carlo experiment and data assimilation experiments with too sparse and/or inaccurate observations to give improved forecasts.

On the specifically mentioned inconsistency (related to L457), our intention was to use a standard deviation around 0.1 m/s in the current velocities. To map this value to the observation error covariance matrix, which acts on the momentum, we multiply by the equilibrium depth (230 m) and square the result, which gives us $\mathbf{R} = 529\ \mathbf{I}$. Since this value is arbitrarily chosen, we fixed $\mathbf{R} = 500\ \mathbf{I}$ (both in the observation error and the observation error matrix). We see that our description of this was inaccurate, and it will be clarified when revising the manuscript.

**Action item:** We will add more details about the experimental design and setup, possibly by also including an appendix. We will also carefully go through the specific comments from the reviewer related to this comment.

## 6. Lack of Statistical Tests

**Reviewer:** *The lack of statistical tests in the sensitivity experiments raises questions to significantly identify differences between the single- and multi-level Monte Carlo methods.*

**Reply:** The statistical properties of the MLEnKF has been analysed in theoretical settings by Chernov et al. (2021). Our work aims at exploring the applicability of MLEnKF, focusing on the practical challenges of applying the method to more applied geophysical problems than what has been done before. We keep the single-level EnKF as a reference, and the main result of the paper is and will still be related to the available speed-up of introducing multiple levels. The fact that MLEnKF works at all for this type of model is a novel contribution to both the multi-level and data-assimilation communities.

But going beyond this main result of speed-up, we have, before submitting our paper, conducted tens of replicate runs related to the statistics of the outputs and the sensitivity of the MLMC settings. The results from all these tests have had qualitatively similar results as the reference experiment that we used in the paper. We will however summarize the key findings of these replicate runs in the revised paper.

**Action item:** We will add a discussion around our replicate studies in the simulation study.

## 7. Language and Presentation Issues

**Reviewer:** *The manuscript contains numerous instances of ambiguous expressions, typos, incorrect grammar, and a lack of definitions for words and mathematical symbols. It appears that co-authors may not have conducted a thorough review of the manuscript.'*

**Reply:** We thank the reviewer for pointing out specific grammatical mistakes and suggestions for clarifying the notation. We are prepared to carefully modify and improve the text and notation of the paper.

**Action item:** We will carefully go through the text of the paper and also let a native English person read through and provide comments to us.

# References


Chen, Y. and Oliver, D. S. (2010). Cross-covariances and localization for enkf in multiphase flow data assimilation. *Computational Geosciences*, 14:579–601.

Chernov, A., Hoel, H., Law, K., Nobile, F., and Tempone, R. (2021). Multilevel ensemble Kalman filtering for spatio-temporal processes. *Numerische Mathematik*, 147(1):71–125.

Evensen, G., Vossepoel, F. C., and van Leeuwen, P. J. (2022). *Data assimilation fundamentals: A unified formulation of the state and parameter estimation problem*. Springer Nature.

Hoel, H., Law, K., and Tempone, R. (2016). Multilevel ensemble kalman filtering. *SIAM Journal on Numerical Analysis*, 54(3):1813–1839.

Holm, H. H., Sætra, M. L., and Van Leeuwen, P. J. (2020). Massively parallel implicit equal-weights particle filter for ocean drift trajectory forecasting. *Journal of Computational Physics: X*, 6:100053.

Maurais, A., Alsup, T., Peherstorfer, B., and Marzouk, Y. (2023). Multifidelity covariance estimation in the log-euclidean geometry. *arXiv preprint arXiv:2301.13749*.

Rabault, J., Nose, T., Hope, G., Müller, M., Breivik, Ø., Voermans, J., Hole, L. R., Bohlinger, P., Waseda, T., Kodaira, T., Katsuno, T., Johnson, M., Sutherland, G., Johansson, M., Christensen, K. H., Garbo, A., Jensen, A., Gundersen, O., Marchenko, A., and Babanin, A. (2022). OpenMetBuoy-v2021: An easy-to-build, affordable, customizable, open-source instrument for oceanographic measurements of drift and waves in sea ice and the open ocean. *Geosciences*, 12(3).

Shivanand, S. K. (2023). Covariance estimation using h-statistics in monte carlo and multilevel monte carlo methods. *arXiv preprint arXiv:2311.01336*.