Response to "Review of 'Consistent Point Data Assimilation in Firedrake and Icepack' by Nixon-Hill et al." by Umberto Villa

Reuben W. Nixon Hill, Daniel Shapero, Colin J. Cotter, David A. Ham

February 16, 2024

1 Summary

We would like to thank the reviewer for his careful consideration of our manuscript. A number of the points he raises are very well-made and we have made significant changes to our revised manuscript as a result.

The reviewer's reason for recommending reject is that he feels that the software aspects of the paper are the key interesting component and that these should be the focus of a performance-oriented submission to a software journal such as SoftwareX or TOMS. GMD is, of course, itself a software journal, with a focus on software applicable to geoscientific applications. We therefore don't feel that it's necessary to reorient the paper onto what TOMS or SoftwareX would contain. Conversely, the other reviewer felt that the software aspects were over-detailed but that the message about data assimilation in cryosphere modules was essentially valuable, albeit with reservations about specific matters.

What we can take on board is that the manuscript is insufficiently explicit in what the reader should take away. In short, the core idea of this manuscript is that point evaluation can be fully integrated as a first class, differentiable operation in a symbolic finite element framework such as Firedrake, and that doing so makes it straightforward to assimilate point data by interpolation, as opposed to the (mathematically problematic) extrapolation methods that have frequently been applied. We will make this explanation much more explicit in the abstract and introduction.

2 Response to Specific Comments

1. We will cite hIPPYlib and explain the difference between that work and the approach presented here. We are not the first to assimilate point data and we do not claim to be – oceanographers have assimilated data from drifting buoys into circulation models for years. The key distinction between this and previous work such as hIPPYlib is that the point data is included as fully first class objects in the symbolic layer of the finite element system. This enables data assimilation, which hIPPYlib also supports, but also enables point data to be used almost anywhere in a calculation where any other finite element function could be used. For example, this enables the interpolation to point data to be put in the integrand, as shown in equation 13. The beauty of introducing a fully composable first class operator is that users will certainly find uses for this functionality that we have not thought of.

2. The goal of this experiment is to compare the approach the we propose with what is commonly done in the literature. Many publications have used naive interpolation methods to map discrete observations to a finite element mesh without acknowledging the impact that this arbitrary algorithmic choice can make. We agree that using Gaussian process regression would be superior to naive interpolation, but relatively few publications (at least in the glaciology literature) use it.

3. The reviewer correctly points out several instances where we overstated the statistical rigour of our experiment. We have walked back some of these claims. The purpose of the demonstrations here is to show that the new point data assimilation capabilities of Firedrake work as advertised and that this feature is fully integrated into the rest of the symbolic capabilities of the library. We claim that this new feature is useful whether one is interested in solving problems from the viewpoint of both Bayesian statistics or deterministic inverse problems.

   1. We agree, we overstated our case here and have altered the text accordingly.
   2. This was an oversight on our part and we thank the reviewer for pointing this out. Using the square norm of the gradient is an inverse crime and we have updated the text to reflect this. Nonetheless, the essential point of this exercise is to demonstrate the new point data assimilation capabilities added to Firedrake and that the resulting estimates converge in the limit of a large number of observations. Using the square norm of the second derivative is better practice but does not substantially alter the essential point, so we have kept the simulations as-is.
   3. We have removed discussion of a Bayesian inverse problem.

4. Please see our answer to reviewer Brinkhoff about Sec. 7.1.1.

5. We have expanded on the explanation in the text. We are not using a Kalman filter or the Laplace approximation. Briefly, there are only three parameters to infer and 18 observations, so computing a MAP estimator with the improper uniform prior on the transmissivities is the same as computing the maximum likelihood estimator. We generated 30 independent realisations of the observation set, computed a MAP/MLE from each one, and then fit a normal distribution to the resulting 30 estimates for the transmissivities.

6. The reviewer suggests that scaling experiments are necessary. We suggest that this is not actually valuable in the context of the functionality presented here. The reason is that point evaluation at static points typically accounts for a vanishing proportion of runtime. The circumstances where the performance of point evaluation are likely to become a first order concern are where there are a very large number of particles and these move (hence necessitating frequent updates of the containing cell and local coordinates). This functionality is not presented here and is listed in the future work section. A future paper focussed on, for example, statistics of moving particles would be the appropriate juncture to study the performance of the system.

7.
   1. We have standardised the notation to use J^{point} and J^{field}
   2. This is a well-made point. We have pared back this section significantly.
   3. We have tightened up the formality of the mathematical exposition throughout. We would like to point that the semicolons are not typos but are an indication that the operator in question is linear in the arguments after the semicolon. This notation was introduced at the end of chapter 3 and is an extension of the notation for forms used in, for example, Alnæs et al. (2014). The use of tilde to indicate quantities in local coordinates has been made consistent.

References

Alnæs, M. S., Logg, A., Ølgaard, K. B., Rognes, M. E., and Wells, G. N.: Unified form language: A domain-specific language for weak formulations of partial differential equations, ACM Transactions on Mathematical Software, 40, 9:1–9:37, https://doi.org/10.1145/2566630, 2014.