Review of 'Consistent Point Data Assimilation in Firedrake and Icepack' by Nixon-Hill et al.

Doug Brinkerhoff

July 28, 2023

1 Summary and Main Points

This paper introduces methods in Firedrake for computing cost functions involving pointwise rather than integral likelihoods. The authors demonstrate this capacity on examples from heat conduction, groundwater hydrology, and glaciology.

I generally find the problem that this paper addresses to be an important one, and the PDE-constrained optimization community would be well served by adopting some of the reasoning contained here. I also think that this method is unfortunately muddled by artificial rigor: my chief complaint about this work is that involves more technical detail than is required to adequately address the issue at hand, and in some cases the presentation of that detail contains mistakes, although I do not think that those mistakes necessarily translate into the implementation of the work. I think that this paper can definitely be made suitable for publication in GMD, but requires some significant reframing with respect to how the methods and results are presented.

2 Line-by-line comments

P1 I don't think that 'Partial Differential Equation' should be capitalized.

P2, Para 2 'misfits which' → 'misfits that'

P2, Para 2 I'm not sure what 'finding a first or second derivative of the point evaluation operation' means. I suspect it means that the output of the interpolation operation needs to be amenable to reverse-mode automatic differentiation, but this should be clarified.

Sec. 2 This section is probably a bit too textbook for GMD. I do think it's relevant to remind the reader that finite elements are functions defined everywhere on the domain, and even that they take the form of a weighted sum of basis vectors, but the basic example given makes that point less clear. In the last line of this section, I think it would be better to say 'within the boundaries of the mesh' rather than 'on the mesh'.

Sec. 3 I think that this section significantly overcomplicates the situation. The evaluation of a finite element function at specific locations is just

$$\mathbf{u}_{\mathrm{pts}} = \mathbf{\Phi}_{\mathrm{pts}}\mathbf{w} \tag{1}$$

where $\mathbf{u}_{\mathrm{pts}}$ are the interpolated model predictions, $\mathbf{w}$ are the DoF values (presumably determined from solving a PDE), and $\mathbf{\Phi}_{\mathrm{pts}}$ is a sparse matrix with each desired evaluation location as a row, and each FEM basis function a column. This requires no definitions of vertex-only meshes and and does not require pointwise DG spaces. Indeed, such a formulation doesn't actually work: the spatial integral of a finite valued basis function that is only defined at a single point in $\mathbf{R}^d$ (as in Eq. 7) would be zero: the basis function that accomplishes the intended goal is not pointwise constant but rather the Dirac delta function. The simpler discrete linear operator viewpoint also has a particularly simple adjoint, which is, well, the adjoint (i.e. the transpose). I recognize that the vertex-only mesh may be a convenient way to represent the situation in UFL, but I think that this should be justified from a data structure perspective rather than a mathematical one.

Sec. 4 More or less the same comments as above. I don't understand the need for the complexity here: the Ciarlet definition of a finite element method is useful for defining elements, but Firedrake already abstracts away the evaluation of functions defined over the resulting spaces (see for example the .at(x) functionality).

Eq. 17 It would be more common to write the complete discrete variational problem by prepending the phrase: 'Find $u \in P2CG(\Omega)$ such that'

On P9, regarding pointwise evaluation of DG functions It seems extremely unlikely that real-valued coordinates should fall on mesh boundaries. If for some reason they do, perhaps it would be better to either throw an error or take an average.

P9, last paragraph I again want to emphasize that the interpolation operation reduces to a sparse matrix-vector product.

Eq. 19 what does i mean in this equation? Should these points be elements of a set?

Just after Eq. 20 The regularization isn't a 'guess' (a word which might reasonably be used in this context to describe an initial parameter value for an iterative optimization procedure), but rather encodes assumptions about the function space in which the parameter lives.

Note on nomenclature This is not necessary. Simply define terms as they're used here.

Eq. 21 To be clear, I am strongly supportive of the viewpoint the authors take here, and it is surely better to map to point observations rather than use interpolation to produce an intermediate value. However: I think that this formulation is a straw-man because it ignores the fact that different observations can be weighted differently (or likelihoods can be heteroscedastic, if you prefer). For example, if one used a Gaussian Process to interpolate point data to the computational mesh, and used the associated posterior distribution as the likelihood function (the negative log-likelihood would be the misfit functional in this formulation), then one could also use the posterior covariance as a weight, which would yield large weights (small variances) for locations on the mesh close to the point observations and small weights (large variances) elsewhere. Of course this still involves the introduction of a model controlling observational smoothness – which is probably undesirable – but it is not nearly so awful as suggested here.

Last line of Sec. 6 Have a look at this paper: https://arxiv.org/abs/2303.06871. Several other works in recent years have coupled FEM solvers (including Firedrake) to general AD tools.

Between Eqs. 24 and 25 Again, I don't think this is correct for DG0 on points. The basis functions would need to be Dirac delta functions rather than finite constants. However, Eq. 25 is still correct regardless.

P13 'L-curve' is usually capitalized.

Sec. 7.1.1 The lack of posterior consistency in this case is the result of not actually treating the interpolation in a properly Baysesian way by ignoring the resulting posterior uncertainties! It is certainly true that treating that interpolation operator in a way that ensures that consistency is non-trivial and so the current method is still of great utility because it avoids that necessity (more or less). However, to say that there isn't a way to ensure convergence in the infinite data limit for the interpolated case isn't true.

Eq. 32 I think that there's a units mismatch here: the time derivative is L/T, while a spatial derivative of a velocity is a 1/T. Should those velocities be fluxes?

P18, paragraph 3 I don't understand where the probability density functions come from: this usually requires the use of Bayes' theorem and MCMC or variational inference or something to come up with such distributions. Can you describe in more detail how these were produced with the stated 'ensembles'?

P18, last paragraph Please see comment on Sec. 7.1.1.

Sec. 7.3.1 I don't think it's necessary to justify use of the SSA here.

Sec. 7.3.2 The thickness is not observable from satellite remote sensing, the surface elevation is.

P21, para 3 One challenge here is that velocities themselves are a gridded product and may not represent the 'real' data points (whatever that means for cross-correlation methods). It would be worthwhile to mention this.

Fig. 8 What does the 'm' in the axis label refer to?

P23 L1 What would happen if your test set wasn't randomly sampled? Velocity observations tend to be very close together, often much closer than the characteristic smoothing scale induced by the ice sheet model (typically understood as 6 to 10 ice thicknesses), and this spatial autocorrelation would lead to an underestimate of the regularization parameter because any roughness in θ below that smoothing scale would be averaged by the physics themselves.

P23 Para 2 I don't understand the reasoning here, nor the assumptions about uniform scaling factors. All you are measuring is the RMSE, which could be attributed either to observation errors *or* to model inadequacy.

P23 Para 3 and 4 I don't really understand what this part is trying to say. There are many other mechanisms for doing principled inverse problems than the ones described here (dare I say that adopting Bayesian formalism and the accompanying methods are quite a bit more principled than what is described in this work). In any case, this is better suited for an independent 'Discussion' section, as it is not strictly relevant to the results of the glaciological experiment.

Sec. 8 I don't see how different observations in time is linked to particle in cell methods, except for the fact that they both involve a 0-D object and some notion of time. In general, this section doesn't seem fully considered and the choices of what to describe are kind of arbitrary and not immediately relevant to addressing the shortcomings of the work presented here.