Interactive comment on "The Community Inversion Framework v1.0: a unified system for atmospheric inversion studies" *by* Antoine Berchet et al.

Peter Rayner (Referee)

prayner@unimelb.edu.au

article times natbib amsmath

General Comments

This paper introduces a software system to facilitate ensembles of inverse calculations. The likely focus is atmospheric inversions in which model inputs such as surface fluxes are inferred from combinations of prior information and observations. Such a system

is clearly useful since ensembles based on choices of ingredients (such as models or input covariances) provide a better basis for both the means and uncertainties of posterior estimates. the paper is certainly within scope for GMD since it describes a system which might well be widely used in the future. It is also well written though some final proofreading, especially around pluralisation, will be needed.

I note I am not reviewing the software itself. There is only one major critique of the paper. I think the benefits of the system are slightly over-sold (but what software release doesn't do this). I believe, for example. that CIF doesn't provide a framework for comparing estimates where the target spaces differ such as ENKF with short windows vs long-term mean approaches. That's an intellectual/mathematical task which one could instantiate in software. this would probably involve classes representing PDFs which might be internally represented by ensembles or parameters and functions then defining transformations among these PDFs. That quibble aside there is no doubt the system will help such comparisons.

On the positive side the ability to configure the computational pipeline is a marvellous innovation. Building our own, less ambitious system, we have struggled to make it agnostic about computation.

Perhaps more important than supporting intercomparisons, the tools provided as part of the system will lower the entry barrier for new groups entering the field. This was a surprising spin-off from the TRANSCOM effort 20 years ago. Most systems since then have been too closely tied to a given model for universal use so CIF can provide a major boost to the field. The fact that it is well documented will only help this.

There is one aspect, though, that seems seriously deficient. If I read correctly, methods for calculating uncertainty in the Gaussian posterior are not provided. One important role of a new framework like CIF is to educate, to encourage good practice. That includes calculating and engaging with the uncertainty in the results. Certainly any of the analytic methods under-estimate this term but it is still true that under-sampling of

concentration is a major source of uncertainty for inversions at all scales. I request that the authors at least detail a roadmap for including posterior uncertainty.

Specific Comments

P7 bottom this correlation of errors is a common critique of ensemble methods but I've rarely seen it implemented in a non-ensemble method either and it's perfectly possible to do in ensemble methods, one just adds some extra state variables carrying observational corrections and build the correct autocorrelations into them instead

P8L22 saying "the maximum probability" and "the mode" is tautological I think

P9 congratulations for explicitly writing out the regularisation/reduction step which is so often assumed. It is called pre-conditioning so often that you should add the term for clarity

P9 it's worth commenting that leading eigen-vectors of the Hessian correspond to the low uncertainty parts of the solution which might or might not be what you want

P11 citing the Bousserez and Henze paper is a good advertisement for your approach, you might want to say in the abstract or intro that many new methods are hybrid and this combined framework simplifies implementing such methods

P15 a technical question, how do you handle target variables which aren't needed at various stages, e.g. bias corrections in observations which aren't inputs to the CTM? the pipelining structure doesn't lend itself well to this problem.

P18 I think Eq. 17, do you consider sources as points or area integrals, if so do you need to integrate over the area of the source, especially for nearby observations?

P20 it's worth adding that the amount of intermediate calculation needed is at user discretion; a space-speed trade-off. CMAQ, for example needs almost none, having checkpointed what it needs on the forward run it had to do anyway at the start of the iteration.

References

Cressie, N. and Johannesson, G.: Fixed rank kriging for very large spatial data sets, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 70, 209–226, 2008.

Palmer, P. I., Suntharalingam, P., Jones, D., Jacob, D. J., Streets, D. G., Fu, Q., Vay, S. A., and Sachse, G. W.: Using CO2: CO correlations to improve inverse analyses of carbon fluxes, Journal of Geophysical Research: Atmospheres, 111, doi:10.1029/2005JD006697, 2006.

Rayner, P. J., Michalak, A. M., and Chevallier, F.: Fundamentals of data assimilation applied to biogeochemistry, Atmospheric Chemistry and Physics, 19, 13911–13932, doi:10.5194/acp-19-13911-2019, https://www.atmos-chem-phys.net/19/13911/2019/, 2019.