1 General comments

The authors present new iterative computational methods for performing atmospheric inversions. They treat two cases, in which the mean of the fluxes is specified directly, and in which the mean of the fluxes is unknown and is described by a parametric linear model. Unknown covariance/regularisation parameters can be estimated as part of the iterative approach at relatively low additional computational cost. As a by-product of the method, an approximation of the posterior covariance matrix may be constructed. The methods are described clearly and the authors make a contribution to the state-of-the-art for atmospheric inversions. I support publication after the following concerns are addressed.

I think the authors could devote a bit more space to reviewing other iterative methods and placing their method in that context. For example, the method appears to fall within the broad class of variational methods. A few papers for these methods are cited but it would be useful for the reader to understand a little more about the proposed method in context, and (the subject of my next comment) how it improves upon existing methods.

Connected to the previous comment, I find it hard to evaluate the relative computational benefits of the method. It is clear that the method is faster than the direct inversion. However, is it faster (e.g., takes fewer iterations) than, say, methods proposed in Miller et al. (2020)? How does the computational complexity compare? Relatedly, all the methods tested make use of the fact that the transport matrix, $\mathbf{H}$, was precomputed. That is no limitation of the method, which does not rely on this fact, but it does make it harder to understand the *total* time required for the inversion, which properly includes the precomputation of $\mathbf{H}$. It would be good for the authors to mention this, and perhaps speak a little to the computational burden of calculating $\mathbf{H}$. This is important because the authors suggest that one of primary contributions of the method is its computational efficiency.

The authors present uncertainty quantification for the fluxes but, unlike the reconstructed fluxes, there is no evaluation of whether the approximate covariance matrix is suitable. My concern would be that the low rank approximation could over estimate marginal variances and underestimate correlations (though whether this is true is not obvious to me). A second concern is that the regularisation parameter λ is changing in every iteration (probably more so in earlier iterations), whereas I believe the theoretical results for the posterior covariance in Saibaba et al. (2020) depend on knowing λ (though I may have misunderstood this). These concerns could be addressed empirically by evaluating the suitability of the uncertainties. One simplistic way do this would be to investigate this would be to compare the approximate posterior standard deviations to those from the direct method, which should be available for the six week case study. A second way would be to discuss in slightly more detail what the results of Saibaba et al. (2020) say about this. A third way, perhaps less feasible, would be to consider a scoring rule that takes into account the uncertainty (similar to the reconstruction error for the posterior mean). One such scoring rule could be the posterior log density (which is multivariate Gaussian) evaluated at the truth—higher density would be better. Another scoring rule option could be to consider the calibration of the uncertainty intervals for some quantity of interest such as the total flux (for example, do the 50% prediction intervals for the total flux in each time step contain the total flux 50% of the time?) These ideas are just meant as suggestions, the more general request being to explore whether the uncertainties are suitable.

2 Specific comments

My specific comments are:

- In Section 3.1.2, it would be good for the authors to provide a reference that specifically discusses the use of the discrepancy principle in similar problems.
- On page 14, line 358, it is stated that $\mathbf{X}$ is identical to Miller et al. (2020). But in Miller et al., $\mathbf{X}$ is described as having only one column, an intercept. Have I missed something?

3 Technical corrections

- At several places in-text citations are surrounded in parentheses, e.g. written as (Smith et al., 2020) instead of Smith et al. (2020). Generally I would expect to see the latter when the citation is incorporated into the sentence.

References

Miller, S. M., Saibaba, A. K., Trudeau, M. E., Mountain, M. E., and Andrews, A. E. (2020). Geostatistical inverse modeling with very large datasets: An example from the Orbiting Carbon Observatory 2 (OCO-2) satellite. *Geoscientific Model Development*, 13(3):1771–1785.

Saibaba, A. K., Chung, J., and Petroske, K. (2020). Efficient Krylov subspace methods for uncertainty quantification in large Bayesian linear inverse problems. *Numerical Linear Algebra with Applications*, 27(5):e2325.