Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2019-185-AC1, 2019

Geoscientific Model Development Discussions

Interactive comment on "Geostatistical inverse modeling with very large datasets: an example from the OCO-2 satellite" *by* Scot M. Miller et al.

Scot M. Miller et al.

scot.m.miller@gmail.com

Received and published: 26 December 2019

We would first like to thank Ian Enting for his thoughtful review. His input and suggestions have been extremely helpful, and the manuscript is better for it. Below, we have listed Dr. Enting's comments in bold and discuss the associated changes we have made to the manuscript.

- **This is a synthetic data study. That is an entirely reasonable thing to be doing. However, to say that the analysis is "using CO2 observations from .. OCO-2" (as the authors do in the abstract and in the conclusion) is simply not true.**

  Dr. Enting makes a good point here. We do use synthetic observations, not real

  observations. We have modified the text in the abstract to clearly indicate that these are synthetic observations.

- **A complicating aspect is that the fluxes from carbon-tracker are themselves the product of an inverse calculation and so will have different spatio-temporal correlations than actual fluxes.**

  This is also a fair point. We do not know what actual CO_2 fluxes are, but a flux product like CarbonTracker is likely the closest approximation to actual fluxes that one can find. To this end, we have added a caveat on pg. 11, line 25 (of the original GMDD manuscript); real-world CO_2 fluxes may be different from the CarbonTracker fluxes used here. Hence, the analysis presented in this study is an approximation or prototypical example of the computational challenges one might encounter in an inverse model.

- **I would strongly disagree with the claims of uniqueness of GIM with regard to including other types of information.**

  There are two different things:

  1. **The geostatistical approach of using spatio-temporal correlation structure as a technique for regularising an ill-conditioned inverse problem; and**
  2. **The inclusion of additional information about fluxes (incorporated into GIM through β).**

  There is nothing to prevent the inclusion of additional information into inversion techniques that do not use geostatistical constraints.

  While the specific form of $p(s|\beta)$ that is used leads to linear equations and a direct solution, once direct solutions are replaced by 'variational' approaches, more general forms of $p(s|\beta)$ can be incorporated, either with or

without the use of regularisation by imposing a spatio-temporal correlation structure.

We agree with point #1. Recently, numerous inverse modeling studies that use a classical Bayesian approach have included spatio-temporal correlation structure in the prior covariance matrix (referred to in this manuscript as $\mathbf{Q}$). Hence, while some previous GIM studies have argued this is a unique feature of a GIM, we do not make that argument in the present manuscript. We have modified line 1 on page 4 to make this distinction clearer.

With regard to point #2: we agree that more general forms of $p(s|\beta)$ could be incorporated, but this has not been done in existing atmospheric inverse modeling studies. I.e., there is no atmospheric inverse modeling study that we are aware of that has done that to date. In theory, one could design an inverse model in any number of ways with different distributional assumptions, prior probability densities, hyperparameters, and hyperpriors. However, nearly all atmospheric inverse modeling studies to date use a relatively narrow set of cost functions that have a similar form. In that context, the way in which GIMs have been applied to atmospheric inverse problems and the way in which these studies have formulated the prior probability is unique relative to what has been done to date in classical Bayesian inverse modeling studies using atmospheric trace gas observations.

- **page 11, Line 5. The Miller et al. (2018) study doesn't seem to provide much information about the actual spatio-temporal correlation structure of the OCO-2 data (i.e. the structure of R. More discussion of this would be desirable.**

We have added several sentences to the revised manuscript to clarify the inverse modeling approach used in the manuscript.

The actual error correlation length and correlation time varies depending upon the region in question and the day of the year, depending upon factors like ob-

servation type (e.g., nadir versus glint), atmospheric aerosol concentrations, and variations in surface albedo (e.g., O'Dell et al. 2018). In Miller et al. (2018, 2019), we estimated a spatial error correlation lengthscale between 0km and 3,800km for land nadir and land glint observations (version 9) using a spherical covariance model. We also calculated error correlation times between 0 days and 45 days, also using a spherical model.

In the present manuscript, we used a diagonal structure for $\mathbf{R}$. All existing inverse modeling studies that we are aware of using OCO-2 observations utilize a diagonal structure for $\mathbf{R}$, and we use a similar approach that is prototypcal of existing inverse modeling studies. Also note that the STILT footprints here were run every 2 seconds (approximately every 15km) along the OCO-2 flight track, so the synthetic observations used in this study are spaced further apart than the observations in the full OCO-2 dataset. In some locations, this spacing is comparable to the estimated spatial error correlation length in the OCO-2 observations.

It may be helpful to incorporate off-diagonal elements within $\mathbf{R}$ for future inverse modeling studies – to better account for the information content of the OCO-2 observations and/or to estimate rigorous posterior uncertainties. However, that has not been done in existing OCO-2 studies to date, and doing so would likely necessitate accounting for the highly non-stationary structure of these off-diagonal elements. We discuss the possibility of incorporating off-diagonal elements within $\mathbf{R}$ in the manuscript (e.g., pg. 13 of the GMDD manuscript). The overall focus of this paper is on computational approaches to large inverse problems, and to that end, we felt that developing an approach to describe non-stationary error covariances in $\mathbf{R}$ was beyond the scope of the current manuscript.

References

Miller, S. M., Michalak, A. M., Yadav, V., and Tadić, J. M.: Characterizing biospheric carbon balance using CO_2 observations from the OCO-2 satellite, Atmos. Chem. Phys., 18, 6785–6799, https://doi.org/10.5194/acp-18-6785-2018, 2018.

Miller, S. M., Michalak, A. M.: The impact of improved satellite retrievals on estimates of biospheric carbon balance, Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2019-382, in review, 2019.

O'Dell, C. W., Eldering, A., Wennberg, P. O., Crisp, D., Gunson, M. R., Fisher, B., Frankenberg, C., Kiel, M., Lindqvist, H., Mandrake, L., Merrelli, A., Natraj, V., Nelson, R. R., Osterman, G. B., Payne, V. H., Taylor, T. E., Wunch, D., Drouin, B. J., Oyafuso, F., Chang, A., McDuffie, J., Smyth, M., Baker, D. F., Basu, S., Chevallier, F., Crowell, S. M. R., Feng, L., Palmer, P. I., Dubey, M., García, O. E., Griffith, D. W. T., Hase, F., Iraci, L. T., Kivi, R., Morino, I., Notholt, J., Ohyama, H., Petri, C., Roehl, C. M., Sha, M. K., Strong, K., Sussmann, R., Te, Y., Uchino, O., and Velazco, V. A.: Improved retrievals of carbon dioxide from Orbiting Carbon Observatory-2 with the version 8 ACOS algorithm, Atmos. Meas. Tech., 11, 6539–6576, 10.5194/amt-11-6539-2018, 2018.

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2019-185, 2019.