Atmos. Meas. Tech. Discuss.,
doi:10.5194/amt-2020-451-RC1, 2021

Atmospheric Measurement Techniques Discussions

Interactive comment on "Reduced-Cost Construction of Jacobian Matrices for High-Resolution Inversions of Satellite Observations of Atmospheric Composition" *by* Hannah Nesser et al.

Marc Bocquet (Referee)

marc.bocquet@enpc.fr

Received and published: 15 February 2021

1 Main remarks

This is a well written paper. Nonetheless, there does not seem to be so much added value compared to previous publications on this topic. However, this is a nice illustration of this difficult set of ideas (reduction of the control space) and, which, to me, is very welcome and useful. Among possible improvements, I would list:

- Although it is rather fair as it is, the bibliography and references should be given more attention, be more complete and ultimately improved. Several key references that predate those cited in the manuscript, should be mentioned first.
- A couple of algorithms (in a proper algorithmic environment – typically a pseudo-code) could be provided for both methods described in the manuscript.
- Because this work's objective is the improvement in efficiency and to decrease the inversion's computational cost, the use of parallelism and modern computer architecture should be discussed.

Overall, I believe the manuscript only requires minor revisions but that they should be very carefully addressed.

2 Suggestions and typos:

1. l.14: "be orders of magnitude lower than its coverage suggests": Although I fully agree with the authors, the phrasing seems a bit excessive.
2. l.42: "The solution is generally obtained by minimizing a Bayesian cost function...": Rigorously speaking, there is no such thing as a "Bayesian cost function"; I would suggest: "The solution is generally obtained in a Bayesian framework by minimizing a cost function..." for instance.
3. l.46: "Methods of estimating the error exist (Bousserez and Henze, 2018; Evensen, 2009), but these approaches are computationally expensive, incomplete, and rarely applied in practice.": there are many earlier papers dealing with errors, including posterior errors, with objective estimation. For instance, among our own contributions: Koohkan and Bocquet (2012); Koohkan et al. (2013).

4. l.81-82: "Bocquet et al. (2011) defined a method to find the optimal multiscale grid from an array of all allowable grids, but this requires a large computational investment.": ok, agreed, but solutions had been proposed (and tested with success!) in the companion paper (Bocquet and Wu, 2011), see for instance Section 3.

5. l.84-87: "were subjective and did not consider the information content of the forward model or the observations. Reduced-rank methods (Bousserez and Henze, 2018; Spantini et al., 2015) generate an approximation of the posterior solution at the original dimension n by solving the inversion in the directions of highest information content. Spantini et al. (2015) assumed knowledge of the Jacobian matrix.": Absolutely but so does Bocquet et al. (2011), albeit in the physical space rather in a spectral space.

6. l.78-89: There are other key papers in reduction methods applied to source inverse modelling in atmospheric chemistry that should be mentioned. Those are based on reversible-jump MCMCs. I can think of Lunt et al. (2016); Liu et al. (2017).

7. l.91: "that minimize" $\longrightarrow$ "that minimizes"

8. The introduction is concise but very well written. However, the references chosen in the introduction mostly refer to the authors' works. I am fine with additionally citing your own papers and recent/fresh contributions to the field, but you should at least cite the seminal or key papers for each main idea. For instance: l. 44-45: "This minimum is typically found using a numerical (variational) method, often employing the adjoint of the CTM to compute the cost function gradient (Henze et al., 2007)." : Citing Henze et al. (2007) is fine assuming you do not forget typically earlier works such as Elbern and Schmidt (2001); Quélo et al. (2005) and studies by Greg Carmichael et al.

9. l.101: "of Jacobian matrix construction." $\longrightarrow$ "of the Jacobian matrix construction."

10. l.107: the notation (A lower index) for the prior is very confusing, since in the literature it very often points to the Analysis, i.e. the posterior. I understand that this is the one used by Clive Rodgers (Rodgers, 2000) or in the 1D retrieval community, but this is not the one used by the large majority. Moreover, it is also conflicting with the dedicated notation A for the averaging kernel (which does not refer to the prior but to the posterior). I would strongly suggest to change notation to make the manuscript easier and less confusing to read.

11. l.116: Same issue with $\mathbf{K}$ which is universally used as the Kalman gain matrix (Kalman, 1960), including in the geophysical inverse problems and data assimilation literature.

12. l.120; Equations (2, 3): you forgot the punctuation of the equations. Please check the whole manuscript and its equations.

13. l.172-175: "We will refer to the rate at which the information content accumulates as the number of eigenvectors increases as the information content spectrum.": More simply put, the spectrum is the ordered list of the eigenvalues.

14. l.229-231: I understood the point on clustering. Yet, it seems a bit vague to me. You could be more specific.

15. l.240-242: Again, this passage is not so clear and could be improved, although I guess I roughly understood.

16. Sections 2.4 and 2.5: I believe you could/should add a pseudo-code to each algorithm. The text is rather (though not entirely) clear and adding an algorithm would really help/reassure the reader. Obviously, these are the key sections of

the manuscript, so that it's worth investing time and (manuscript) space in such algorithms.

17. l.261: This was first proposed and proven in Bocquet et al. (2011), section 2.4.
18. l.337-344: The results for the reduced-dimension solution are somehow underwhelming; the resulting DOFS are quite low. Do you have a explanation for this? Or did I miss something?
19. You did not discuss at all the impact of time as you considered a static mesh for the emission. Can you discuss briefly the approximation that such assumption entails?
20. You did not discuss the patterns provided by the eigenvectors (main modes of the DOFS). Is it worth discussing this point?
21. l.425-426: You might want to have a look at solutions proposed in the meteorological data assimilation community to efficiently compute the Jacobian in high dimension, for instance Frolov et al. (2018).
22. I believe you should discuss parallelism of your algorithms and codes. Your paper is targeted at more efficient techniques – which will also depend on how well you are able to exploit parallelism. Please add a thorough discussion on the subject.

References

Bocquet, M., Wu, L., 2011. Bayesian design of control space for optimal assimilation of observations. II: Asymptotics solution. Q. J. R. Meteorol. Soc. 137, 1357–1368. doi:10.1002/qj.841.

Bocquet, M., Wu, L., Chevallier, F., 2011. Bayesian design of control space for optimal assimilation of observations. I: Consistent multiscale formalism. Q. J. R. Meteorol. Soc. 137, 1340–1356. doi:10.1002/qj.837.

Elbern, H., Schmidt, H., 2001. Ozone episode analysis by four-dimensional variational chemistry data assimilation. J. Geophys. Res. 106, 3569–3590.

Frolov, S., Allen, D.R., Bishop, C.H., Langland, R., Hoppel, K.W., Kuhl, D.D., 2018. First application of the local ensemble tangent linear model (LETLM) to a realistic model of the global atmosphere. Mon. Wea. Rev. 146, 2247–2270. doi:10.1175/MWR-D-17-0315.1.

Henze, D.K., Hakami, A., Seinfeld, J.H., 2007. Development of the adjoint of GEOS-Chem. Atmos. Chem. Phys. 7, 2413–2433. doi:10.5194/acp-7-2413-2007.

Koohkan, M.R., Bocquet, M., 2012. Accounting for representativeness errors in the inversion of atmospheric constituent emissions: Application to the retrieval of regional carbon monoxide fluxes. Tellus B 64, 19047. doi:10.3402/tellusb.v64i0.19047.

Koohkan, M.R., Bocquet, M., Roustan, Y., Kim, Y., Seigneur, C., 2013. Estimation of volatile organic compound emissions for Europe using data assimilation. Atmos. Chem. Phys. 13, 5887–5905. doi:10.5194/acp-13-5887-2013.

Liu, Y., Haussaire, J.M., Bocquet, M., Roustan, Y., Saunier, O., Mathieu, A., 2017. Uncertainty quantification of pollutant source retrieval: comparison of Bayesian methods with application to the Chernobyl and Fukushima-Daiichi accidental releases of radionuclides. Q. J. R. Meteorol. Soc. 143, 2886–2901. doi:10.1002/qj.3138.

Lunt, M.F., Rigby, M., Ganesan, A.L., Manning, A.J., 2016. Estimation of trace gas fluxes with objectively determined basis functions using reversible-jump Markov chain Monte Carlo. Geosci. Model Dev. 9, 3213–3229. URL: http://www.geosci-model-dev.net/9/3213/2016/, doi:10.5194/gmd-9-3213-2016.

Quélo, D., Mallet, V., Sportisse, B., 2005. Inverse modeling of NO_x emissions at regional scale over northern France: Preliminary investigation of the second-order sensitivity. J. Geophys. Res. 110, D24310.

Rodgers, C.D., 2000. Inverse methods for atmospheric sounding. volume 2. World Scientific, Series on Atmospheric, Oceanic and Planetary Physics.

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2020-451, 2020.