Atmos. Chem. Phys. Discuss.,
doi:10.5194/acp-2016-800-AC3, 2017

Atmospheric Chemistry and Physics Discussions
Open Access

Interactive comment on "Diagnostic methods for atmospheric inversions of long-lived greenhouse gases" *by* Anna M. Michalak et al.

Anna M. Michalak et al.

michalak@stanford.edu

Received and published: 21 January 2017

Response to Referee #3

We thank the Referee for their constructive input. We have structured our response using the following sequence, per instructions: (1) comments from Referee, (2) author response, (3) changes in manuscript.

COMMENT FROM REFEREE: General comment: The paper attempts to address a challenging topic related to the validation of atmospheric inversions of greenhouse gases. In response to the increasing demand for more robust atmospheric validation tools, the authors review the existing solutions to this problem, using independent data for an indirect validation or using sensitivity experiments with different statistical metrics. The review of methods and the analysis of previous studies is quite extensive

Printer-friendly version

Discussion paper

and provides a valuable overview of the current state of the art for diagnostic methods. The later section aims at evaluating these diagnostics and discusses the usefulness of these approaches related to the problems they try to address. This part of the paper suggests that most of these metrics remain insufficient to evaluate the potential problems affecting inverse flux estimates. The authors fall short of providing suggestions trying to address these limitations, for example by recommending new measurements or methodologies to diagnose and identify them. The two main solutions proposed here are an increase in atmospheric data availability and the increase in spatial resolution to overcome representation errors when evaluating against direct flux measurements. Considering that both options are unlikely to happen in many vast areas across the world, other options should be considered to help the inverse modellers provide more robust results. I would invite the authors to 1. propose clear directions for inverse modellers to address these issues, including methodologies and strategies for measurement campaigns, and 2. suggest new/other statistical metrics to better evaluate inverse results and therefore overcome the limitations of the current metrics in inversion studies. Overall, this paper is a worthwhile contribution reviewing the current diagnostics for inversions but would need to develop this last section to provide more insights to the inversion community. Therefore, I recommend this paper for publication after addressing this problem and the following specifics comments

AUTHOR RESPONSE: Following other authors (e.g., Desroziers et al. 2005 and Talagrand 2014) quoted in our text, we made it clear in Section 4 that diagnostics are ambiguous in a way that is inversely proportional to the amount of information which is input to them. In the revised version, we will refer to the principle of equifinality. The current challenge therefore does not lie in the definition of new metrics, but rather in the increase of the amount of information. We therefore propose (i) to increase the horizontal resolution in order to exploit some of the existing data that can hardly be used now, and (ii) to increase the measurement density. Item (i) is already happening (see, e.g., the comparison between atmospheric inversions and tower/aircraft flux measurements in Lauvaux et al. 2009, Meesters et al. 2012, Broquet et al. 2013),

Printer-friendly version

Discussion paper

while item (ii) is being actively prepared by some companies and by space agencies (e.g., http://www.copernicus.eu/sites/default/files/library/CO2_Report_22Oct2015.pdf). We therefore argue that these prospects are both necessary and achievable in the near future. In the meantime, we also argue that there is much inspiration and much quality to be gained in using the existing data better in most inversion systems by following some of the diagnostics that are presented in our text (we used the expression “crucial tool box” twice).

CHANGES IN MANUSCRIPT: We will develop Section 5 in order to illustrate the fact that the way forward is achievable and already happening.

COMMENT FROM REFEREE: Page 2 - L11: The studies cited here describe component-level surveys of equipment which are isolated in time. The term "monitored" does not reflect the lack of temporal coverage from these methods.

AUTHOR RESPONSE: Agreed.

CHANGES IN MANUSCRIPT: We will mention that providing temporal coverage is an additional challenge.

COMMENT FROM REFEREE: Section 3.3.1: Past studies (e.g. COBRA campaign, or CERES) and more recent ones (e.g. based on EnKF approaches) have tried to use meteorological and GHG data to improve or characterize transport models at continental and regional scales. Similarly, global scale models have also been compared to vertical profiles. The current section is short and would need a more complete list of studies related to transport model evaluation.

AUTHOR RESPONSE: The referee’s point is well taken. We had intended for this section to describe the use of unused atmospheric observations for diagnosing inversion systems as a whole, but we agree that the evaluation of atmospheric transport models is a crucial component thereof.

CHANGES IN MANUSCRIPT: We will add a paragraph focusing specifically on the use of atmospheric observations for evaluating atmospheric transport models.

COMMENT FROM REFEREE: Section 3.4: A description of the most important metrics used with OSSE's would help the readers to understand the possible information that can be recovered from pseudodata experiments. Past studies have also confused the meaning (or interpretation) of these metrics. For example, error reduction analysis may be the most useful metric one could possibly study, but often suffers from over-confidence. Discussions may be useful in this regard, and link to the "grain of salt" compared to a proper evaluation of inversions.

AUTHOR RESPONSE: Agreed. With regard to the error reduction metric in particular, we agree that the reduction of uncertainty is frequently used in OSSEs. However, this metric is primarily used to assess the information content of a particular set of observations, rather than to assess the validity, self-consistency, or robustness of the inversion system itself. This is the reason for which it was not discussed in subsection 3.4.

CHANGES IN MANUSCRIPT: We will add a description of metrics used as part of OSSEs for diagnosing inversions systems.

COMMENT FROM REFEREE: Page 13 - L6-8: Should the readers conclude that these metrics are not addressing the problem? Could the authors provide more insights to explain why these metrics are insufficient? I think most inverse modelers would agree with the statement but examples of shortcomings or reasons for this failure are needed here.

AUTHOR RESPONSE: In our presentation, we have introduced diagnostics as an answer to the needs of "quality control (. . .) (i.e., the evaluation of the flux estimates) [and of] (. . .) quality assurance (i.e., the evaluation of the estimation process that yielded the flux estimates)" for the inversion systems. In this sense, diagnostics address the problem well. If we ask them for more, such as identifying what is going wrong in a system for which some diagnostics show a warning, we may be limited by the amount

Printer-friendly version

Discussion paper

of information actually available within or outside the inverse system under study. This was discussed, in part, in Section 2, which presented some of the unique challenges of developing, applying, and interpreting diagnostics for atmospheric inverse problems.

CHANGES IN MANUSCRIPT: We will recall the purpose of diagnostics at the start of Section 4. We will also formulate the limitations of diagnostics in terms of their capability to infer some property of an inversion system, based on the well- or ill-posedness of that particular inference problem.

COMMENT FROM REFEREE: Page 13 - L15-21: Do we need specific data to implement these methods? The spatial and temporal structures of errors are critical to inversions but the authors should provide more suggestions to address the separation of contributions from prior and transport model errors. This problem is non-trivial and has been studied in other fields in a more systematic fashion. Maybe references from non-GHG assimilation studies may help here.

AUTHOR RESPONSE: To be implemented, these methods do not need other data than the assimilated ones. They have been used in many non GHG assimilation studies, and the example that we give in page 13 L. 17 (Desroziers et al. 2005) actually comes from numerical weather prediction. Recent application examples in this field could be given (e.g., Stewart et al. 2014), but may not help much here. In any case, their capability always depends on the information available from the data. The referee's point about diagnosing the validity of assumptions about prior error vs. transport model errors is well taken, but is already address to the extent possible in Section 3.

Reference: Stewart, L. M., Dance, S. L., Nichols, N. K., Eyre, J. R. and Cameron, J. (2014), Estimating interchannel observation-error correlations for IASI radiance data in the Met Office system†. Q.J.R. Meteorol. Soc., 140: 1236–1244. doi:10.1002/qj.2211

CHANGES IN MANUSCRIPT: None.

COMMENT FROM REFEREE: Page 14 - L3-8: Few studies have tried to address this problem, for example the Global Carbon Project with a more coherent framework to compare inversion results. More generally, the authors could describe how to construct ensembles able to represent inversion errors. Again, possible examples from other communities (e.g. weather prediction systems) may help to find solutions, or at least, avenues that the inversion modelers could take to generate better probabilistic ensembles.

AUTHOR RESPONSE: An ensemble of inversion results represents inversion errors well provided that the corresponding ensemble of inversion set-ups explores the space of uncertainty widely (e.g., the ensemble would not be limited to one particular source of information for its prior fluxes for a given source-sink process) and in a balanced way (e.g., the ensemble would not oversample marginally-different versions of a single transport model at the expense of other transport model types). In practice, this goal is usually hampered by limited resources that favor existing set-ups over the design of systematic explorations of other plausible and defensible set-ups. These statements are general and not limited to the GHG community.

CHANGES IN MANUSCRIPT: We will expand the discussion to better capture this challenge.

Interactive comment on Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-800, 2016.