Atmos. Chem. Phys. Discuss.,
doi:10.5194/acp-2016-800-AC2, 2017

Atmospheric Chemistry and Physics Discussions

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 #2

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: Michalak and colleagues review the recent literature on methods to assess robustness and accuracy of atmospheric inversions of long-lived GHGs. Given the importance of inversions in present biogeochemistry and potentially in future GHG emission reduction verification, such diagnostic methods are of great relevance. After an excellent introduction on the need for diagnostics and the involved challenges, the paper reviews diagnostics applied in the literature. The diagnostics are

put into context in a discussion section. I recommend to publish this work, subject to some comments given below.

AUTHOR RESPONSE: We thank the Referee for their positive assessment of this work.

CHANGES IN MANUSCRIPT: None.

COMMENT FROM REFEREE: When reading through the list of diagnostics, a question that repeatedly came up to me was "How well an inversion actually has to meet these diagnostics to be good enough?" For example, in Sect 3.1.1, how to translate the fit to independent data into a judgement of quality? I realised that it would be asking much to comprehensively answer this question here, and Sect 4 does discuss the limitations of the set of diagnostics. Neverthess, I was wondering whether it would be helpful to put more on that already along the way, to make the paper more practical.

AUTHOR RESPONSE: The Referee's point / question is well taken. The question of the extent to which a given inversion has to satisfy a given metric is application-dependent, and in some cases subjective and perhaps even controversial. This adds to the complexity of applying diagnostics in this particular field.

CHANGES IN MANUSCRIPT: In revising the manuscript, we will add an overview paragraph to each subsection in Section 3 (3.1, 3.2, 3.3, 3.4) providing a clearer context and synthesis across the approaches to be presented in each subsection, and take that opportunity to touch on the question of "how good is good enough" brought up by the referee.

COMMENT FROM REFEREE: I feel it should be mentioned early on that the cited literature can only provide examples, because I'm sure that for most (if not all) diagnostics there are further papers which have also made good use of them, and which in some cases may even deserve credit for actually having introduced them. In this context, the restriction to papers from between 2010 and 2015 does not seem entirely appropriate to me.

AUTHOR RESPONSE: The referee is correct that the manuscript cannot provide a completely comprehensive survey of existing literature. This is always a delicate and subjective balance. For example, Referee #1 actually recommended that we go in the opposite direction, significantly cutting the number of examples presented.

CHANGES IN MANUSCRIPT: In revising the manuscript, we will explicitly state that the referenced manuscripts do not represent a comprehensive set. We will also better articulate our reasoning for the selected level of detail, as also outlined in our response to Referee #1. In terms of the limitation to 2010-2015 and the decision to cite a recent paper vs. an original paper, our goal was primarily to showcase recent applications of specific types of diagnostics, rather than to present a historical view of when specific diagnostic approaches were originally proposed. We do believe that a balance needs to be struck, so in revising the manuscript we will also cite original papers where this would be beneficial, and at a minimum make sure that we do not imply that a recent paper is the original presentation of a given approach when we are in fact simply using it as an example of a contemporary application thereof. Finally, we will augment the existing list of references with some key papers from 2016.

COMMENT FROM REFEREE: I missed explicit mentioning of the "reduction of uncertainty" (1- sigma(Post)/sigma(Pri)), a diagnostic which has been being widely used by many studies, mostly in OSSEs as an alternative to the synthetic inversions explained in Sect 3.4. (In this context, it would be good to mention that the choice of foci and examples is partially subjective according to the working fields of the authors.)

AUTHOR RESPONSE: We agree that the reduction of uncertainty is frequently used in OSSEs and inversion studies. 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. We did, however, briefly discuss this approach in the original Section 3.3.2.

CHANGES IN MANUSCRIPT: We will make the focus of the presented metrics clearer in the first paragraph of Section 3 and subsection 3.4.

COMMENT FROM REFEREE: Specific comments: p 6 l 15-19: Mention already here that the robustness of column data is not yet fully established (as said later in 3.3.2), to avoid an inappropriate message.

AUTHOR RESPONSE / CHANGES IN MANUSCRIPT: Agreed. We will do so.

COMMENT FROM REFEREE: p 6 l 30: Add "global *decadal* atm. growth rates" because this statement is not valid at yearly or shorter time scales any more.

AUTHOR RESPONSE / CHANGES IN MANUSCRIPT: Agreed. We will do so.

COMMENT FROM REFEREE: p 7 l 1-4: The cited study is for N2O - would this also work for CO2 with both sources and sinks? (By the way, I would find it useful to mention which trace gas is being looked at in the individual examples.)

AUTHOR RESPONSE / CHANGES IN MANUSCRIPT: Good point. We will edit to make it clear that this statement is less valid for gases such as CO2. We will also revise throughout to make target gases clearer.

COMMENT FROM REFEREE: p 7 l 5-7: I find that comparisons "across inversions" are misplaced in this paragraph on comparison to "independent estimates", as inversion-inversion comparisons only allow fundamentally weaker conclusions.

AUTHOR RESPONSE / CHANGES IN MANUSCRIPT: Agreed. We will make the change.

COMMENT FROM REFEREE: p 7 l 10: The term "assessment" is so general that it remains unclear what to take from this sentence

AUTHOR RESPONSE / CHANGES IN MANUSCRIPT: We agree that the statement was vague. The cited paper describes the comparison of the seasonal cycle of estimated CH4 mixing ratios (from an inversion constrained by in situ measurements) to that of independent TCCON CH4 columns (both averaged over multiple TCCON locations). The large-scale agreement in this cycle is thought to support the TM5 tropopause height seasonality, because this dynamic contributes to the seasonality of column CH4. This comparison was also made for posterior CH4 columns from an inversion constrained by satellite XCH4 to determine appropriate seasonal bias correction (as explained in the next sentence of the review paper), and the agreement in the seasonal cycles between the observations and the in situ-constrained inversion posterior provides evidence that the phase shift in the satellite-constrained inversion posterior seasonal cycle is not due to a misrepresentation of tropopause height or another large scale seasonal meteorological variable in the transport model. We will clarify this in the revised manuscript.

COMMENT FROM REFEREE: p 8 l 4-6: This is a complicated and unspecific formulation. What about something like "...check whether the flux adjustment by the inversion are still within the specified a-priori probability distribution".

AUTHOR RESPONSE / CHANGES IN MANUSCRIPT: We agree that the statement was vague. A completely objective criterion is difficult to define, however. The example text provided by the referee, for example, would not work, because if one assumes a Gaussian distribution then any value is technically “within” the distribution. We will add a brief discussion of chi-squared statistics etc., but also make it clear that these metrics carry with them assumptions of their own.

COMMENT FROM REFEREE: p 8 l 9-10: Posterior concentration uncertainties can indeed be calculated in theory, but in most larger applications, this is computationally very involved in practice. I feel this should be noted.

AUTHOR RESPONSE / CHANGES IN MANUSCRIPT: Agreed. We will note this in the revision.

COMMENT FROM REFEREE: p 9 l 20+: This has already been said in Sect 3.1.1

AUTHOR RESPONSE / CHANGES IN MANUSCRIPT: The distinction between this portion of Section 3.3.2 and Section 3.1.1 is whether the additional observations are used to evaluate a posteriori fluxes (3.1.1) vs. whether the inversion is conducted multiple times, each time using a different set of observations (3.3.2). We will make this distinction clearer in the revision, and also avoid any redundant discussion.

COMMENT FROM REFEREE: p 9 l 31-32: The sentence "The differences ... data." seems to be incomplete.

AUTHOR RESPONSE / CHANGES IN MANUSCRIPT: We do not believe so. Subject: “The differences in the geographical flux patterns.” Verb: “can be attributed.” How: “through the use . . ..”

COMMENT FROM REFEREE: p 9 l 33: It remains completely unclear what "quantified via ... signal" means.

AUTHOR RESPONSE / CHANGES IN MANUSCRIPT: We agree this was unclear. We mean that calculating the effective number of degrees of freedom provided by a given set of observations gives insight into the information content of those data with respect to fluxes. One can then use this metric to compare different (sub)sets of observations. We will make this clearer in the revision.

COMMENT FROM REFEREE: p 10 l 11-18: This paragraph unspecifically uses the term "sensitivity tests", but I assume it actually refers to synthetic-data tests. It therefore seems to better fit into Sect. 3.4.

AUTHOR RESPONSE / CHANGES IN MANUSCRIPT: We disagree. We are referring to the fact that one can run multiple “real data” inversions, each time using a different subset of available observations. We will make this clearer in the revision.

COMMENT FROM REFEREE: p 10 l 31: add "regional inversions", as this is only relevant there.

AUTHOR RESPONSE / CHANGES IN MANUSCRIPT: Agreed. We will make the change.

COMMENT FROM REFEREE: p 11 l 7-11: This seems to have been said already in the previous paragraph.

AUTHOR RESPONSE / CHANGES IN MANUSCRIPT: Agreed. We will merge the two paragraphs and shorten the discussion where possible.

COMMENT FROM REFEREE: p 11 l 12: add "or data set" after "of a model", as it is not always models that are being used.

AUTHOR RESPONSE / CHANGES IN MANUSCRIPT: Agreed. We will make the change.

COMMENT FROM REFEREE: p 14 l 10-11: The sentence "The ambiguity ... to them" may tentatively be true but due to its awkward formulation it remains unclear what it actually means.

AUTHOR RESPONSE / CHANGES IN MANUSCRIPT: We will replace “ambiguity” by “equifinality”, which better describes what we mean (the same value for a given metric can be obtained by several inversion configurations).

COMMENT FROM REFEREE: p 14 l 29-31: Add e.g. ", used in conjunction with high-precision data". I disagree with the notion that low-quality data will ever be sufficient on their own, even if much larger in number.

AUTHOR RESPONSE / CHANGES IN MANUSCRIPT: We thank the reviewer for pointing this out. We fully agree and will make the change.

COMMENT FROM REFEREE: p 15 l 8: Be specific which diagnostics this sentence is referring to, because otherwise one cannot take any information from this sentence.

AUTHOR RESPONSE / CHANGES IN MANUSCRIPT: We agree that the sentence was too vague. We will replace “(e.g., Candille and Talagrand, 2005)” by “(e.g., the reliability diagram of Talagrand et al., 1999).”

Reference: Talagrand, O., Vautard, R. and Strauss, B. (1999), Evaluation of probabilistic prediction systems. in Proceeding of workshop on predictability, p. 1-25, October 1997. European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, Berkshire RG2 9AX, UK, http://www.ecmwf.int/sites/default/files/elibrary/1997/12555-evaluation-probabilistic-prediction-systems.pdf

COMMENTS FROM REFEREE: Minor comments:

p 4 l 14: I find the specification "aimed at ...and patterns" obvious and thus dispensible

p 5 l 25: I find that "high level groupings of" is unnessecarily confusing and should be deleted.

p 9 l 3-4: replace "an inversion" by "the transport model"

p 11 l 26: Remove "However" as this sentence is not in opposition to the previous sentences.

p 11 l 30: Rather say "can also be used".

AUTHOR RESPONSE / CHANGES IN MANUSCRIPT: We agree with all the minor comments above, and will make edits accordingly.

COMMENTS FROM REFEREE: Typos:

p 3 l 32: "atmosphere"

p 7 l 1: "inform"

p 8 l 26: delete "a comparison of"

p 15 l 13-14: Exchange "artmospheric" and "for"

AUTHOR RESPONSE / CHANGES IN MANUSCRIPT: We agree with all the minor comments above, and will make edits accordingly.

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