Response to Referee #1

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: The manuscript provides an interesting overview of the existing diagnostics to evaluate atmospheric inversions of long-lived tracers. The paper doesn't introduce any novelty in the field, but rather, it establishes a list of the existing tools. It is well written, and there is no obvious "wrong" point to comment on. I was quite pleased with Sections 1 and 2, which are a nice introduction to the topic, for non-specialists. I was unfortunately less convinced by Sections 3 and 4: although they are well written as well, I wonder what kind of reader would actually learn from it. Inverse modeling specialists are already familiar with the concepts that are presented; Non-specialists will get an idea of the diagnostics tools available, but since the paper often doesn't go much beyond listing them, they will have to read the (many) references to actually understand them.

AUTHOR RESPONSE: We thank the referee for recognizing the value and intent of the work. With regard to Sections 3 and 4, our goal is two-fold. Although we agree that inverse modeling specialists will already be familiar with some (or many) of the concepts presented, we doubt that any specialist would be familiar with the full spectrum of approaches presented here, unless they themselves had conducted a full review of the literature. Speaking on behalf of the three specialists who authored this manuscript, although each of us was familiar with many of the approaches we describe, we each also learned about some that were new to us. For non-specialists, we agree that they would need to read additional references in order to get a deeper understanding of a particular approach. We consider this a strength rather than a weakness of the work. In essence, in Section 3 we are providing a guided tour of the literature, which would allow a non-specialist to know exactly where to go for a more in depth presentation of any particular approach.

CHANGES IN MANUSCRIPT: In the revised manuscript, we have more clearly enunciated the dual target audience for this review in Section 1. We have also implemented changes to put the approaches within a clearer context, as described in more detail in other responses below.

COMMENT FROM REFEREE: As an example, in Section 3.1.1 (the first in which some diagnostic tools are actually presented and discussed), in 19 lines, the authors talk about: evaluation inversions against observations left out of the inversions; evaluation inversions against observations from aircraft profiles (and as a one-line example, against vertical concentration profiles); evaluation of satellite observations constrained inversions using in-situ measurements; evaluation of in-situ observations constrained inversions using satellite measurements; evaluation against "all types of independent atmospheric observations". Each of these in less than three lines.

This is not useful to the experienced inverse modelers who are already very familiar with all this. This is not very interesting for newcomers to inverse modeling (it can be summarized in one sentence: "evaluate your results against independent data", the rest is case-specific). Finally, for specialists from other disciplines who would like to get a glimpse at how inverse models are evaluated, it quickly gets boring. Meanwhile, there are important questions that could be discussed here, but that are, in the best case, left to Section 4: comparing observations with their model counterpart is not always trivial (case of satellite observations which may require an important work of data selection, bias correction, and the application of an averaging kernel to the model fields), not always wise (comparing low-resolution model CO2 fields with CO2 observations in an urban environment is not so smart), and not always that useful (the implications of a bias vs. independent observations in the upper stratosphere are not the same than that of a bias in the continental boundary layer). On the other hand, not doing it is sometimes catastrophic (incorrect interpretation of inversions constrained by biased satellite data).

AUTHOR RESPONSE: The referee's point is well taken. As described in the first response above, we believe that there is value to presenting a broad ranging set of references and variations on diagnostics approaches, for both specialists and non-specialists. This goes with our vision of this manuscript as a roadmap to the existing literature. At the same time, we agree with the referee that in some instances the desire to be thorough came at the expenses of synthesis and interpretation.

CHANGES IN MANUSCRIPT: In revising the manuscript, we have examined each sub-section in Section 3 with the goal of keeping detail to the extent it is useful, but at the same time restructuring the discussion in a way that synthesizes information more effectively, as with the example listed by the referee. The main changes come at the start of each subsection (3.1, 3.2, 3.3, 3.4), where we have presented more context for the subset of diagnostics to be presented.

COMMENT FROM REFEREE: Some subsections of Section 3 are better, but overall, the paper would read much nicer with less references, less examples, but more detailed ones (given the pedigree of the authors, I am certain that they can easily find some from their own work, and illustrate them with a few figures). Once again, the key is to define the target readers, and what they should retain: Non specialists don't need to know of tens of examples (they won't remember them all anyway), but they need to understand correctly and completely those that are presented. Specialists might be interested in the many references, but most of them could be moved out of the main text, perhaps to one or several tables (perhaps one for each Section 3.1, 3.2 and 3.3), as it is often done in literature reviews. Before final publication in ACP, I would therefore recommend that the authors consider revising Section 3 and 4, keeping in mind that readers should be able to learn from it without having to read the references and/or the other papers from the special issue.

AUTHOR RESPONSE: We agree with the referee that a review of this sort could fundamentally take one of two forms. The first is, as the referee described, an in depth look at a small selected set of prototypical examples. The second is, as we have done here, a more comprehensive overview of the literature. There are advantages and disadvantages to each. We made the choice to use the second model in part specifically to make it easy for readers to dig deeper if they chose to, by looking up the references included in the manuscript.

CHANGES IN MANUSCRIPT: With the above in mind, we have decided not to restructure the manuscript by limiting discussion to a few prototypical examples and putting the remainder of references in a table. That being said, we take the referee's concerns to heart, and have revised Section 3 to provide more context, synthesis, and interpretation, to the extent possible without substantially increasing the overall length of the manuscript. Section 4 can actually be read without knowing the ten papers referenced in it: these have been included only as examples and options for further reading; hence, these are are now all preceded by either "e.g." or "see also".

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 have added 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 have explicitly stated that the referenced manuscripts do not represent a comprehensive set (see revised Section 3.0). We have also better articulated our reasoning for the selected level of detail, as also outlined in our response to Referee #1 (see revised Section 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 have clarified this in the revised Section 3.0. We do believe that a balance needs to be struck, so in revising the manuscript we have also cited original papers where beneficial, and at a minimum made 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 (see revised Section 3.0 and details in subsequent subsections of Section 3). Finally, we have augmented 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 have made the focus of the presented metrics clearer in the revised Section 3.0, but also added a brief comment about the "reduction in uncertainty" metric in Section 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 have done 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 have done so.

COMMENT FROM REFEREE: p 7 l 1-4: The cited study is for N_2O - would this also work for CO_2 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 have edited to make it clear that this statement is less valid for gases such as CO_2. We have also revised throughout to make target gases clearer where appropriate.

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 have removed the reference to comparisons from one inversion to another.

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 CH_4 mixing ratios (from an inversion constrained by in situ measurements) to that of independent TCCON CH_4 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 CH_4. This comparison was also made for posterior CH_4 columns from an inversion constrained by satellite XCH_4 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 have clarified 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 have made this distinction clearer in the revision, and also avoided 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 have made 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 have made 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 have made 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 have deleted redundant text.

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 have made 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 have replaced “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 have made 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 have replaced “(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 have made 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 have made edits accordingly. The one exception is inform vs. informs; we have kept 'informs' as the subjection is the singular word 'evaluation.'

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 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 have referred 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), 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 have developed Section 5 in order (i) to state the need of providing both uncertainty statistics with the posterior fluxes and some evidence of the statistical consistency of these fluxes with the inversion assumptions, (ii) to develop the need of increased measurement density (we have made the distinction between flux and concentration measurements and have added a brief discussion about possible systematic errors in future measurement types), (iii) to refer to Lauvaux et al. (2009) and Meesters et al. (2012) as existing examples of the possibility to compare high-resolution inversion results with flux measurements.

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 now mention both the spatial (i.e. point) and temporal (i.e. episodic, discrete) aspects of these observations.

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. Sim
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 have added a paragraph focusing specifically on the use of atmospheric observations for evaluating atmospheric transport models in Section 3.1.1, as we felt this was a more natural location for the information presented therein.

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 originally discussed in subsection 3.4.

CHANGES IN MANUSCRIPT: We have added a description of metrics used as part of OSSEs for diagnosing inversions systems. We also added a brief comment about overall error reduction in the revised Section 3.4.

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 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 have recalled the purpose of diagnostics at the start of Section 4. We have also formulated the limitations of diagnostics at the end of the section 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 have expanded the discussion with the words of our response to the comment to better capture this challenge.

Diagnostic methods for atmospheric inversions of long-lived greenhouse gases

Anna M. Michalak[1], Nina A. Randazzo[1,2], Frédéric Chevallier[3]

[1]Department of Global Ecology, Carnegie Institution for Science, Stanford, California, United States
[2]Department of Earth System Science, Stanford University, Stanford, California, United States
[3]Laboratoire des Sciences du Climat et de l'Environnement, Gif-sur-Yvette, France

Correspondence to: Anna M. Michalak (michalak@stanford.edu)

Abstract. The ability to predict the trajectory of climate change requires a clear understanding of the emissions and uptake (a.k.a. surface fluxes) of long-lived greenhouse gases (GHGs). Furthermore, the development of climate policies is driving a need to constrain the budgets of anthropogenic GHG emissions. Inverse problems that couple atmospheric observations of GHG concentrations with an atmospheric chemistry and transport model have increasingly been used to gain insights into surface fluxes. Given the inherent technical challenges associated with their solution, it is imperative that objective approaches exist for the evaluation of such inverse problems. Because direct observation of fluxes at compatible spatiotemporal scales is rarely possible, diagnostics tools must rely on indirect measures. Here we review diagnostics that have been implemented in recent studies, and discuss their use in informing adjustments to model setup. We group the diagnostics along a continuum starting with those that are most closely related to the scientific question being targeted, and ending with those most closely tied to the statistical and computational setup of the inversion. We thus begin with diagnostics based on assessments against independent information (e.g., unused atmospheric observations, large-scale scientific constraints), followed by statistical diagnostics of inversion results, diagnostics based on sensitivity tests and analyses of robustness (e.g., tests focusing on the chemistry and transport model, the atmospheric observations, or the statistical and computational framework), and close with the use of synthetic data experiments (a.k.a. observing system simulation experiments (OSSEs)). We find that existing diagnostics provide a crucial toolbox for evaluating and improving flux estimates, but, not surprisingly, cannot overcome the fundamental challenges associated with limited atmospheric observations or the lack of direct flux measurements at compatible scales. As atmospheric inversions are increasingly expected to contribute to national reporting of GHG emissions, the need for developing and implementing robust and transparent evaluation approaches will only grow.

1 Introduction and the need for diagnostics

The ability to predict the trajectory of climate change requires a clear understanding of the historical and current emissions and uptake (a.k.a. surface fluxes) of long-lived greenhouse gases, and chief among them carbon dioxide (CO_2) and methane (CH_4), over the Earth's land and ocean regions. For the natural components of the global budgets of these gases, understanding historical and contemporary flux patterns is needed for elucidating the biogeochemical processes that control flux variability, and therefore the likely evolution of these fluxes under changing climate scenarios (e.g., Friedlingstein et al., 2014). The ability to constrain the anthropogenic components of greenhouse gas budget estimates, on the other hand, is becoming increasingly central to discussions aimed at setting emissions, or emissions reduction, targets at local to global scales (e.g., Pacala et al., 2010).

Direct monitoring of the fluxes of greenhouse gases is only feasible at a limited number of spatial and temporal scales, however. Point sources of anthropogenic emissions can be measured directly at discrete times for example (e.g., Allen et al., 2015; Subramanian et al., 2015; Zimmerle et al., 2015), while biospheric fluxes over land can be continuously monitored at plot scale (i.e. from a few hectares to a few km^2, depending on sensor height) using the eddy covariance technique (e.g., Baldocchi et al., 2001; Law et al., 2002), and ocean fluxes can also be deduced locally from the difference between the partial pressure of CO_2 measured in seawater and that in the overlying air (e.g., Takahashi et al., 1993, 2002). At the global scale, a network of observation sites tracks the global growth rate of atmospheric concentrations of greenhouse gases, and gives broad insight into the temporal (e.g., seasonal, interannual) and spatial (e.g., hemispheric, latitudinal) signatures of net greenhouse gas emissions (e.g., Tans et al., 1990; Steele et al., 1992).

Deleted: monitored

Deleted: observed

The target applications listed in the first paragraph, however, require an understanding of fluxes at intermediate scales, e.g., from urban to biome to national to continental. Direct observations of fluxes are not feasible at these scales, and gaining an understanding of flux budgets and controlling processes at these scales therefore invariably depends on a process of either "upscaling" small-scale flux observations, or "downscaling" large-scale information provided by atmospheric concentration measurements. Upscaling strategies range from the implementation of mechanistic models calibrated using plot-scale flux observations (e.g., Richardson et al., 2012; Schaefer et al., 2012), to the development of statistical or machine learning approaches for elucidating dominant patterns (e.g., Beer et al., 2010; Jung et al., 2011), to the combination of fine-scale flux measurements with activity data (e.g., fuel consumption for anthropogenic emissions, or burnt area for fire emissions) as the basis of emissions inventories (e.g., van der Werf et al., 2006; Jeong et al., 2014; Lyon et al., 2015). Downscaling strategies, on the other hand, most typically involve the solution of an inverse problem to elucidate spatially and temporally resolved flux information from upwind and downwind observations of atmospheric greenhouse gas abundance (e.g., Enting et al., 2002).

Deleted: .,,

Inverse problems that couple atmospheric observations of greenhouse gas concentrations with an atmospheric chemistry and transport model in order to gain insights into underlying flux patterns have been used since the late 1980s (e.g., Enting and Mansbridge, 1989; 1991). While the observational network has expanded and the statistical and numerical methods have become more sophisticated (e.g., Ciais et al., 2010a; Michalak, 2013; Miller and Michalak, 2017; Houweling et al., 2017), the

Deleted: 2016

Deleted: 2016

underlying principles have remained largely unchanged. Spatiotemporal flux patterns at the Earth's surface lead to spatial and temporal gradients in atmospheric concentrations of greenhouse gases. The inverse problem then amounts to using those gradients to recover information about the flux patterns. From a scientific perspective, an additional goal is often to also gain insight into the enviro-climatic factors driving these patterns (e.g., Gourdji et al., 2012; Fang and Michalak, 2015; Miller et al., 2014, 2016b). Although the principle is simple, the atmospheric inverse problem is ill-conditioned because the diffusive nature of atmospheric transport means that relatively small variations or errors in observed or modelled atmospheric concentrations can correspond to relatively large differences or errors in the inferred flux quantities and patterns. In addition, the atmospheric inverse problem is often under-determined because the sparse observational coverage precludes the possibility of resolving fluxes (spatially and temporally) at all the scales that are of scientific or policy interest, as well as at all the scales to which atmospheric observations are locally sensitive.

Given the high scientific and policy value of accurate greenhouse gas budgets, the growing role of atmospheric inverse problems to obtain these budgets at relevant scales, and the inherent technical challenges associated with the solution of these inverse problems, it is imperative that objective approaches exist for evaluating the scientific value and accuracy of inverse modelling estimates of greenhouse gas fluxes. Here, we review diagnostics that have been implemented in recent studies, and discuss their use in informing adjustments to model setup. We have structured the review in a manner that we hope will be useful to novices and specialists alike. We present a relatively comprehensive survey of recent approaches, in order to provide a detailed representation of the state-of-the art for specialists. At the same time, we have organized the review around high-level categories in order to help guide researchers who are newer to the field and provide an entry point for further inquiry via the cited studies.

Fundamentally, the emphasis of diagnostic tools should be on the scientific value of insights that are based on the solution of an atmospheric inverse problem. This quality control approach (i.e., the evaluation of the flux estimates) also has to be complemented by quality assurance (i.e., the evaluation of the estimation process that yielded the flux estimates). Indeed, the solution of atmospheric inverse problems invariably involves a series of decision points including, but not limited to, (1) the choice of the atmospheric observations to be used, (2) the choice of the atmospheric chemistry and transport model to be implemented, (3) the choice of a statistical framework for defining an objective function that captures the relative contribution of atmospheric observations, the chemistry and transport model, and any prior information in informing flux patterns, and (4) the choice of a numerical framework for the solution of the inverse problem. Each of these choices will have a direct impact on estimates. It is therefore also imperative to have diagnostic tools that can evaluate the self-consistency of the modelling and statistical assumptions specific to the choices made in the setup of the inverse problem. In other words, at a minimum, the ultimate estimates must be consistent with the assumptions inherent to the specific modelling setup that was implemented.

Deleted: ., 2015

Deleted: modeled

Deleted: modeling estimates of greenhouse gas fluxes. The emphasis of such diagnostic tool

Deleted: .

Deleted: modeling

Deleted: modeling

2 Challenges of diagnosing atmospheric inversions

Having established the need for diagnostic tools to assess atmospheric inverse modelling results, the question then becomes one of identifying appropriate diagnostics, metrics, or benchmarks. As discussed in the last section, however, direct observation of greenhouse gas fluxes is not possible at the space and time scales targeted by atmospheric inversions. This is in part because inversion systems for long-lived greenhouse gases are run over time periods ranging from weeks to decades to capture the long dispersion times of tracers in the atmosphere and to capture temporal variability in fluxes. These long timespans are achieved at the expense of relatively coarse horizontal resolutions, ranging from tens of kilometres to one or more degrees, such that the large gap between flux measurements and inverse model scales precludes direct evaluation of inverse modelling results. This gap is filled only rarely by some regional inversions (e.g., Lauvaux et al., 2009, Meesters et al., 2012). This means that there is a basic lack of independent measures of flux to assess inverse modelling estimates.

Deleted: atmospheric

Deleted: price

Diagnostic tools used for assessing inverse modelling estimates must therefore rely on other indirect measures or information about the fluxes to be estimated. Such measures and information should, in principle, be independent from the information used in the solution of the original inverse problem. A natural choice might then be to use additional atmospheric concentration data not assimilated in the original inverse problem, because, as noted earlier, gradients in atmospheric greenhouse gas concentrations are themselves the result of underlying flux patterns. Given the ill conditioned and typically under-determined nature of the atmospheric inverse problem, however, it is often desirable to use as much information (i.e. data) as possible to inform the initial solution of the inverse problem, in order to gain the deepest and most precise insights possible about flux patterns. This goal, however, is at odds with the desire to keep some independent flux-relevant observations for diagnosing the estimates obtained from the inversion. Although this problem is not unique to the solution of atmospheric inverse problems, it is certainly particularly salient in this context. Two examples follow.

Deleted: aimed at gaining an understanding of underlying flux magnitudes and patterns

In some ways, numerical weather forecasting (e.g., Kalnay et al., 2003) bears some resemblance to the flux estimation problem, as they both rely on atmospheric observations and a numerical representation of atmospheric dynamics. In both cases, the ability to diagnose the accuracy and precision of estimates is of high value. Key differences emerge upon closer examination, however. First, the target quantities predicted/estimated in numerical weather prediction, such as temperature, precipitation, and barometric pressure, are ones that can also be measured directly at a large number of locations, via both *in situ* and remote sensing observations, making a comparison to direct benchmarks feasible (e.g., ECMWF, 2016). Although it is technically true that in some cases a scale mismatch still occurs (e.g., a thermometer cannot measure the “average” temperature over a computational grid box), the quantities of interest are less likely to display the strong multi-scale heterogeneity that makes eddy covariance flux observations ill-suited for diagnosing grid-scale inverse-model-derived flux estimates at much coarser spatial resolution. Second, whereas atmospheric inverse problems aim to infer/estimate historical flux distributions that were never observed directly, the accuracy and precision of numerical weather forecast estimates can largely be verified, evaluated, and diagnosed simply by waiting for weather patterns to unfold. This is perhaps best illustrated through the long-standing comparisons of forecast skill among the world's weather forecasting bureaus (Simmons and Hollingsworth, 2002; WMO-LCDNV, 2016).

Another useful example is that of the development of retrieval algorithms for remote sensing observations of atmospheric constituents (e.g., Rodgers, 2000). Let us take as a prototypical example the process of obtaining estimates of column-integrated dry air mole fractions of atmospheric carbon dioxide (X_{CO2}) from the spectrum of reflected sunlight measured by the Orbiting Carbon Observatory (OCO-2) space-borne instrument (e.g., Crisp et al., 2012). In this case, the observations are radiances at specific wavelengths within the spectrum of reflected light, with a focus on specific absorption bands that are observed at high spectral resolution. The forward problem involves the solution of radiative transfer equations. The target variable of primary interest is X_{CO2}. This problem has analogies to the flux estimation problem in that the column-integrated CO_2 concentrations cannot be measured directly per se. A key difference, however, is that a number of validation datasets are available to help diagnose the retrieval algorithm (e.g., Osterman, 2011). These include, among others, observations from ground-based remote sensing instruments (that look up at the sun, rather than down at the Earth, e.g., Wunch et al. (2011)), and targeted campaigns of *in situ* airborne observations that can capture CO_2 concentration variability within a portion of the atmospheric column (e.g., Tadić et al., 2014; Frankenberg et al., 2016). Unlike in the flux estimation problem, there is no direct conflict between using these additional measurements for validation / diagnosis versus using them to directly inform the solution of the inverse problem itself, as there is no clear mechanism by which these additional observations could be routinely incorporated within the core retrieval algorithm, although they can be used for additional empirical bias correction.

Overall, then, while the need for diagnostics to evaluate the scientific validity and statistical self-consistency of flux estimates derived via the solution of atmospheric inverse problems is clear, this need poses very substantial challenges. These include the lack of independent measures of flux at comparable spatiotemporal scales, and the inherent dilemma between using available atmospheric observations for estimation versus validation. These features make the process of developing and implementing diagnostics particularly challenging, and fundamentally different from the challenges observed in other fields that might at first glance appear to be somewhat analogous.

3 Overview of existing diagnostics

Researchers have taken a number of approaches in tackling the challenges associated with the development of diagnostics that are both practical, given the unavoidable limitations in available data, and genuinely informative, in terms of assessing the accuracy and precision of flux estimates. Here we describe existing diagnostics that have been used as part of inverse modelling efforts. We focus primarily on diagnostics that evaluate the validity and self-consistency of the inversion setup, rather than on diagnostics designed to assess the information content of specific data sets. We also discuss how diagnostics are used to inform adjustments to model setup and the trade-offs inherent to alternative possible approaches to model evaluation. We focus primarily on examples from papers published between 2010 and 2016, and on papers that present recent applications of specific diagnostics rather than on the studies where these diagnostics were originally introduced. We do so in order to get a

Deleted: ,

Deleted:

Deleted: these

Deleted: on

Deleted: 2015,

contemporary snapshot of approaches that are currently being used for diagnosing atmospheric inversions. The groupings of diagnostics are ordered here by starting with diagnostics that are most closely related to the actual scientific problem or question being targeted by the inversion, to those that are most closely tied to the statistical and computational setup of the inversion framework itself. More fundamental overriding questions about the types of insights that the range of currently available diagnostics can (or cannot) actually provide are then discussed in Section 4.

3.1 Assessment against independent information

The most natural starting point for assessing the solution of an atmospheric inverse problem is through evaluation against independent information. Although, as discussed in earlier sections, direct observations of surface fluxes are seldom available at compatible scales, at least two additional avenues are available. The first is to evaluate flux estimates against unused atmospheric observations, whether from in situ monitoring or remote sensing. This is accomplished through the solution of the "forward" problem, which translates estimated fluxes into modelled atmospheric concentration fluctuations. The second is to compare estimates against any available large-scale scientific constraints. This approach can be challenging especially when large-scale constraints are themselves uncertain.

3.1.1 Evaluation against unused atmospheric observations

If any atmospheric observations are available that have not been used as a constraint in the solution of the inverse problem, they can be leveraged to evaluate final flux estimates. To do so, final flux estimates are used as an input into the atmospheric chemistry and transport model used as part of the inversion, and predicted concentrations at the times and locations of the additional available atmospheric observations are then compared to the measured concentrations. These additional observations can be of several types, and inform the inversion setup in various ways, given differences in vertical information, spatial coverage, and precision.

Evaluating inversion results constrained by *in situ* observations using independent surface or satellite total column measurements can provide additional information about regional fluxes. The much broader spatial coverage of satellite observations makes it possible to assess flux estimates at large spatial scales, and thus can help to identify large-scale spatial biases that are related to a lack of *in situ* coverage in some regions (e.g., biases in the latitudinal gradient or over land versus ocean) (e.g., Lindqvist et al., 2015). However, it is important to note in the context of these comparisons that the satellite retrievals themselves may have regional biases, as will be discussed later.

Conversely, for inversions constrained by satellite observation of total column concentrations, evaluating results using *in situ* measurements can reveal errors in the column-constrained system's ability to reproduce surface fluxes, which can be related to aspects of the retrieval (such as biases) or to the transport model's representation of boundary layer dynamics (e.g., Locatelli et al., 2015; Cressot et al., 2014).

Comparisons to independent measurements can also be used to isolate transport errors from the other confounding errors. For example, comparing the total column mixing ratios simulated based on posterior flux estimates obtained using surface data to

Deleted: most commonly

Deleted: high level

Deleted: somewhat different ways

Deleted: The evaluation of inversion results against independent aircraft profiles makes it possible to assess both the accuracy of flux estimates and of the atmospheric chemistry and transport model

Moved down [1]: (e.g., Pickett-Heaps et al., 2011; Saeki et al., 2013b

Deleted:). For example, vertical concentration profiles captured by airborne observations provide information about vertical mixing and convection. ... [1]

Deleted: fluxes' or chemistry and transport model's ability to represent surface concentrations independent observations of total column mixing ratios can diagnose a transport model's skill in simulating the seasonality of the tropopause height and of the stratospheric partial column (e.g., Houweling et al., 2014). Performing this type of assessment for multiple inversions constrained by different types of measurements but using the same transport model can provide insight into whether seasonal biases in the inversion are caused by seasonal biases in an observing system or to seasonal biases in the transport model (e.g., Houweling et al., 2014). More generally, vertical transport bias can be assessed by comparing the vertical gradients of posterior vertical profiles to those of observed profiles (e.g., Pickett-Heaps et al., 2011; Saeki et al., 2013b; Liu and Bowman, 2016), because vertical gradients provide information about vertical mixing and convection.

More broadly, evaluation against all types of independent atmospheric observations provides an additional window into the degree to which estimated fluxes capture key features of the atmospheric signal, such as the seasonal cycle, latitudinal gradients, or regional patterns of concentrations (e.g., Zhang et al., 2014; Jiang et al., 2014; Diaz Isaac et al., 2014; Pandey et al., 2016; Liu and Bowman, 2016; Johnson et al., 2016).

3.1.2 Evaluation at aggregated scales against large-scale scientific constraints

The accuracy of inversion-derived flux estimates and the validity of the overall inversion framework can be assessed, at large scales, based on existing understanding of carbon cycle and atmospheric dynamics. This type of evaluation may involve comparisons of the inversion-derived estimates to existing information about flux magnitudes at large scales, about the overall direction of the net flux in a region (i.e. emission vs. uptake), or about flux seasonality. Care must be taken, however, for the approach not to become circular, i.e. for inversion results not to be evaluated by comparing them to assumed features of the very processes that the inversion is trying to inform.

In the simplest case, spatially aggregated posterior fluxes can be assessed based on expert knowledge of the system. For example, methane emissions in regions dominated by natural gas extraction, urbanization, wetlands, or cattle feedlots are expected to substantially outweigh soil methane uptake, and negative estimated emissions in such regions would point to errors in the inversion (e.g., Berchet et al., 2013). Similarly, global decadal atmospheric growth rates and latitudinal gradients of greenhouse gases are well constrained by long-term baseline observations (e.g., Conway et al., 1994), and posterior flux estimates can be evaluated against such large-scale constraints (e.g., Cressot et al., 2014). Evaluation against observed latitudinal gradients provides information not only about global total fluxes, but can also inform the accuracy of the representation of inter-hemispheric transport, although more so for gases with limited uptake at the Earth surface (e.g., Thompson et al., 2014). This comparison is especially helpful when performed using both surface and upper-troposphere or total column concentrations, because this makes it possible to assess how both meridional and vertical transport are represented (e.g., Thompson et al., 2014).

More broadly, inversion-derived fluxes can be compared against independent estimates of fluxes for comparable regions, although the fact that both the inversion-derived and the independent estimates of fluxes are uncertain must be recognized. For example, the fraction of the global CO_2 sink attributable to land versus ocean can be compared between inversions and independent model or mass-balance estimates (e.g., Le Quéré et al., 2015). For specific regions and periods, inversion results

Moved (insertion) [2]

Moved (insertion) [1]

Deleted: Conversely, evaluating inversion results constrained by *in situ* observations using independent satellite total column measurements can provide a different type of information. The much broader spatial coverage of satellite observations makes it possible to assess estimates at large spatial scales, and thus can help to identify large-scale patterns in bias (e.g., changes in bias over a latitudinal gradient or over land versus ocean) (e.g., Chevallier and O'Dell, 2013).

Deleted:

Deleted: informs

Deleted: across

Deleted: (e.g., Rivier et al., 2010; Thompson et al., 2014) or compared to

Deleted:

can also be compared against detailed inventory estimates of fluxes (e.g., Lauvaux et al., 2012; Schuh et al., 2013). A third example (noted already in Section 3.1.1) is the comparison of large-scale seasonal cycles of modelled trace gas concentrations to observations. For inversions constrained by remotely sensed data, checking for consistency in seasonal cycles between observations, estimates from a satellite-data-constrained-inversion, and estimates from an *in-situ*-data-constrained inversion may draw attention to the need for seasonal bias correction in the observations, while also exploring other potential causes of regional or seasonal bias, such as seasonal biases in vertical transport (e.g., Houweling et al., 2014). Lastly, bottom-up studies also provide regional budget estimates at the annual or pluriannual scale that can be compared to inverse modelling results (e.g., Gourdji et al., 2012; Miller et al., 2013, 2014). The comparison may reveal convergence (e.g., Ciais et al., 2010b) or divergence (e.g., Chevallier et al., 2014; Miller et al., 2013, 2014) of the estimates. However, the attribution of any divergence remains subjective, given the uncertainty of the bottom-up estimates themselves (e.g., Chevallier et al., 2014; Reuter et al., 2014; Gourdji et al., 2012).

Finally, large dipoles in estimated fluxes between large regions can point to a lack of observational constraint for certain regions, to overfitting of the observations that do exist, and/or to biases in large-scale transport (e.g., Alexe et al., 2015; Nassar et al., 2011). The presence of flux dipoles can, however, also be representative of real spatial flux patterns, and sensitivity tests focusing on factors such as the coverage of observational constraints can help to evaluate such patterns in posterior fluxes (e.g., Cressot et al., 2014; Rivier et al., 2010) (see also Section 3.3).

3.2 Statistical diagnostics of inversion results

Rather than comparing flux estimates against independent information directly, a second set of strategies focuses instead on assessing whether the prior and posterior flux estimates, uncertainties, and covariances are consistent with the assumptions built into the design of the implemented inversion framework. These strategies thereby focus on statistical self-consistency of the inversion setup, and in this way can point to discrepancies that can signal unreliable results.

The majority of inverse modelling approaches used for greenhouse gas flux estimation leverage a combination of prior information and an observational constraint. Within the mathematical framework of the inversion, the uncertainty and spatiotemporal covariance structure of the prior information (i.e., prior error statistics), as well as the reliability with which the researchers expect to be able to reproduce the atmospheric observations (i.e., model-data-mismatch statistics), are represented through error covariances. These error covariances, the prior information, the observational data, and the chemistry and transport model are then also used to quantify the uncertainty associated with posterior estimates (see e.g., Rayner et al., (2016) for a detailed discussion). This framework provides an opportunity to evaluate the statistical self-consistency of the inversion setup.

For example, under the assumption of Gaussian and unbiased errors and for a given set of assumptions about error correlations, the sum of squared errors follows a chi-squared distribution with a known number of degrees of freedom; for this reason, posterior errors can be used to evaluate or scale assumed prior error variances (e.g., Michalak et al. 2005; Desroziers, 2006; Wu et al., 2013; Lauvaux et al., 2016; Cressot et al., 2014). In some cases, deviations between concentrations modelled based

Deleted: assessment

Deleted: surface

Deleted: fluxes, which can provide information about how well seasonal patterns in atmospheric transport and mixed-layer height are represented in the transport model (e.g.,

Moved up [2]: Houweling et al., 2014).

Deleted: -

Deleted: may also

Deleted: (e.g,

Deleted: these

Deleted: those implied by

Deleted: statistical

Deleted: the

Deleted: in two ways

Deleted: First, posterior flux estimate deviations from prior information can be evaluated within the context of the original

Deleted: prior errors, and can be used to adjust the statistical representation of the prior errors (e.g.,

Moved down [3]: Gourdji et al.,

Deleted: 2012;

Deleted: ; Shiga et al., 2014; Miller et al., 2016). ... [2]

Deleted: modelled and observed concentrations to assess whether the measured concentrations are within the uncertainty bounds implied by posterior flux estimates (e.g., Cressot et al., 2014; Wu et al., 2013). Similarly, deviations between modelled and measured concentrations should have statistics that are consistent with the model-data-mismatch assumptions implemented in the inversion, and this criterion can be used to identify appropriate model-data mismatch statistics (e.g., Gourdji et al., 2012; Shiga et al., 2014; Miller et al., 2016). Both of these approaches can be used to diagnose model overfitting (where the modelled concentrations reproduce observations better than implied by the posterior uncertainties and/or the assumed model-data-mismatch covariances) or a model's inability to reproduce observations. Both of these approaches can also be implemented with independent measurements of trace gases on posterior fluxes and atmospheric observations not included in the original inversion can be used for this purpose (e.g., Chevallier and O'Dell 2013). This approach can also be used to assess how model-data mismatch errors vary seasonally (e.g., Gourdji et al., 2012; Kim et al., 2011). Also, the very high resolution of some regional inversions and the availability of plot-scale flux measurements make it possible to validate the posterior uncertainty of fluxes directly in some cases (e.g., Broquet et al., 2013).

The spatial and temporal autocorrelation of posterior errors can also be used to inform model setup (Diaz Isaac et al., 2014) or to assess the identifiability of underlying fluxes (Yadav et al., 2016).

Other than assessing self-consistency, statistical diagnostics can also be used to quantify the error reduction (or information gain) made possible by the assimilation of atmospheric observations. In this approach, posterior uncertainties are compared to prior uncertainties. In cases where the explicit quantification of posterior flux uncertainties is prohibitively computationally expensive, it can also be approximated through approaches such as the use of a Monte Carlo ensemble of inversions in which model parameters are perturbed for each run (e.g., Chevallier et al., 2007; Cressot et al., 2014; Pandey et al., 2016). More simply, the deviations between atmospheric observations not included in the inversion and modelled concentrations based on posterior vs. prior fluxes can be used as a measure of error reduction (e.g., Liu and Bowman, 2016; Johnson et al., 2016; Lauvaux et al., 2016).

3.3 Sensitivity tests and analysis of robustness

The validity and robustness of inversion-derived estimates can also be assessed through sensitivity tests. These tests involve running additional inversions where one or several components have been altered. The most common of these are changes to the chemistry and transport model used to translate fluxes into atmospheric concentrations, changes to the set of atmospheric observations used to constrain flux estimates, and changes to the implemented statistical or computational framework. Examples of the latter include changes to prior estimates, boundary conditions, and flux spatiotemporal resolutions. Results shed light on the degree to which results are robust to specific implementation choices.

3.3.1 Chemistry and transport model

Recently, as inversions have become more sophisticated, transport model sensitivity tests have become more computationally expensive. As a result, it has become more difficult to assess the impact of model choice on inversion results (e.g., Gurney et al., 2002; Baker et al., 2006). Applications focusing exclusively on synthetic data are covered in Section 3.4, while here we present a few examples that included real observations.

Examining the effect of the choice of a chemistry and transport model can lead to various insights. For example, the transport model used by an inversion may be run using different boundary layer schemes to assess how the representation of vertical mixing affects the interpretation of assimilated data (e.g., Peters et al., 2010). Another aspect is the impact of the spatial resolution of the transport model, and particularly the use of finer grids within mesoscale domains versus the coarser grids typical of global transport models. For example, including a finer-scale nested grid and changing the transport representation

Deleted: ... [3]

Deleted: the accuracy of statistical assumptions in the model may

Deleted: or with other factors

Moved (insertion) [3]

Deleted: A calculation of spatial and temporal correlations of residuals can also provide information on the propagation of errors in the model (Diaz Isaac et al., 2014). In some cases, the

Deleted: makes

Deleted: 2013). Finally, if results from an ensemble of alternative inversions is available, a comparison of the spread of the posterior flux estimates can be compared to the posterior uncertainty of these estimates to determine whether the calculated uncertainties capture the discrepancies between state-of-the-science models (e.g., Peylin et al., 2013).

Deleted:

at these finer scales provides information about the effect of transport representation at scales finer than the grid scale of global transport models (e.g., Rivier et al., 2010). In addition, posterior meridional concentration gradients can be compared across inversions that use different global transport models to assess the effect of interhemispheric transport (e.g., Thompson et al., 2014).

The implementation of more than one transport model in a forward run can also shed light on consistent differences in the ability to represent observed atmospheric concentration signals, seasonal cycles of mixing ratios, or vertical profiles (e.g., Pillai et al., 2012; Diaz Isaac et al., 2014).

3.3.2 Atmospheric observations

Performing inversion sensitivity tests in which only the constraining observational data set is changed between inversions can shed light on the impact of various observations on flux estimates, and therefore on their relative information content with regard to underlying fluxes, and also makes it possible to assess the extent to which conclusions are robust to the choice of observations used to constrain the inversion.

For example, a major effort has been made to quantify the effects of including remotely sensed observations (specifically, satellite retrievals) as an additional constraint beyond *in situ* observations. This is distinct from the applications discussed in Section 3.1.1, where remote sensing observations were not included in the inversions, but were instead used to evaluate inversion-derived flux estimates. Satellite data provide the benefit of broader spatial coverage than *in situ* measurements, potentially informing fluxes in regions not well constrained by current *in situ* networks. However, the informational value and robustness of the information provided by satellite observations is still the subject of ongoing research, and thus their use as constraints in inversions requires special consideration of the impacts of any potential biases. Several studies have included satellite total column or mixing ratio data as an additional constraint on a model otherwise constrained only by *in situ* concentration measurements, to determine whether remotely sensed total column concentrations provide a significant amount of additional information (e.g., Alexe et al., 2015; Houweling et al., 2014; Nassar et al., 2011; Pandey et al., 2016; Saeki et al., 2013a). An inversion constrained only by *in situ* measurements may also be compared to an inversion constrained only by satellite measurements (e.g., Cressot et al., 2014). The spatial distribution and magnitude of fluxes and the source/sink status of particular regions are often the major posterior features compared between inversions constrained by different subsets of available data (e.g., Alexe et al., 2015; Cressot et al., 2014; Houweling et al., 2014; Nassar et al., 2011). The differences in the geographical flux patterns can be attributed through the use of various methods focusing on quantifying the information content and geographical coverage of satellite data. The relative information content of the different observational datasets can be quantified via the degrees of freedom (a metric based on posterior error covariances) provided to the inversion (see e.g., Rodgers 2000), whereby data sets that represent a stronger constraint provide more degrees of freedom (e.g., Nassar et al., 2011). The constraint provided for specific regions by observations with extensive geographical coverage can also be qualitatively analysed by creating visualizations of the sensitivity to fluxes from a certain region (e.g., Nassar et al., 2011). If

Deleted: that are constrained by different subsets of available atmospheric observations

Deleted: using

Deleted: .

Deleted: of

Deleted: signal

satellite retrievals provide a large increase in coverage over a particular region, then this method may help to explain large changes in posterior fluxes in upwind areas.

In addition, the robustness of conclusions about flux distributions derived from satellite observations can be explored by using alternative sets of satellite-derived observations. Studies have checked for agreement in posterior fluxes for inversions run using different satellite instruments and retrieval algorithms (e.g., Alexe et al., 2015; Chevallier et al., 2014; Takagi et al., 2014). The effect of the bias correction scheme used for satellite retrieval post-processing has also been a subject of several sensitivity studies (e.g., Houweling et al., 2014; Alexe et al., 2015; Nassar et al., 2011; Cressot et al., 2014, Basu et al., 2013). Sensitivity tests based on inversions constrained by different subsets of available observations have been used to examine the incremental gain in information obtained by expanding the *in situ* observation network. Such experiments can be used to estimate the uncertainty reduction (see Section 3.2) that could potentially be achieved by assimilating more observations over or downwind from poorly constrained regions, as well as the effects of a more extensive observational network on the estimated spatial and temporal variability of fluxes (e.g., Butler et al., 2010; Saeki et al., 2013b; Kadygrov et al., 2015; Jiang et al., 2014; Peters et al., 2010). They can also be used to determine the value of episodic versus continuous observations (e.g., Peters et al., 2010). These sensitivity tests can also determine whether regions with strong fluxes, such as the “dipoles” discussed in Section 3.1.2, are simply due to a relative lack of constraint for certain regions (e.g., Rivier et al., 2010).

Last, sensitivity tests have also been used to examine the potential role of bias of *in situ* measurements at specific site. In such studies, an offset is added to specific observations, and the results of the control inversion and the inversion with the offset can be compared to determine the effect of potential biases on the posterior flux field (e.g., Peters et al., 2010; Masarie et al., 2011).

3.3.3 Statistical and computational framework

Sensitivity tests can be used to explore the impact of the statistical assumptions and computational framework used in inversions.

For example, the impact of assumptions about the statistical representation of prior errors and model-data mismatch errors can be examined by performing multiple inversions, as can the impact of approaches aimed at optimizing these error statistics (e.g., Bousquet et al., 2011; Cressot et al., 2014; Wu et al., 2013; Ganesan et al., 2014; Berchet et al., 2013). Sensitivity tests may also be run on other statistical parameters such as the assumed correlation length of fluxes (Corazza et al., 2011).

Another key aspect of regional inversions that can be explored through sensitivity tests is the impact of the choice of a dataset used to represent background concentrations of greenhouse gases entering the model domain. This can be done through the implementation of alternative boundary conditions, and/or the exploration of the impact of uncertainty in individual sets of boundary conditions (e.g., Göckede et al., 2010b; Bréon et al., 2015; Schuh et al., 2010; Gourdji et al., 2012).

Similar to the case of boundary conditions, inversions aiming to isolate one component of greenhouse gas budgets (e.g., biospheric CO_2 in the case of CO_2 inversions) must rely on pre-existing estimates of other components of the budget (e.g., fossil fuel CO_2 emissions). The impact of the choice of an estimate can be explored through sensitivity tests (e.g., Peylin et al., 2011; Peters et al., 2010).

The choice of a model or data set to be used as an *a priori* estimate in Bayesian inversions is another source of uncertainty in the inferred fluxes, particularly in areas where the observation constraint is weak. Inversions using alternative inventories or process-based models with different spatial and seasonal flux patterns as priors can be compared in terms of the spatial and temporal distributions of the posterior fluxes to assess the robustness of flux estimates (e.g., Kim et al., 2011; Göckede et al., 2010b; Bergamaschi et al., 2015; Corazza et al., 2011; Peters et al., 2010).

A final example is the use of sensitivity tests to explore the effect of the spatial and temporal aggregation and resolution of the unknown fluxes in the modelling framework. The impact of the choice of flux regions, model grid resolution, model grid nesting, or model time step can all be explored (e.g., Rivier et al., 2010; Göckede et al., 2010a; Kim et al., 2014; Peters et al., 2010).

3.4 Synthetic data experiments

Observing system simulation experiments (OSSEs) are studies in which synthetic observations are constructed at observation times and locations using a prescribed set of fluxes and a chemistry and transport model. These synthetic observations are then used instead of actual observations as data constraints on an inversion. OSSEs are particularly useful for diagnostics because the "true" transport and fluxes are known and can be manipulated. These types of studies constitute a necessary but certainly not sufficient condition for ensuring a good inversion setup, as many complexities of inversions using real observations can only be approximated within a synthetic data experiment context. OSSEs have become a key component of inversion model development, especially as models have become more complex.

Because the "true" fluxes are known in an OSSE, various metrics can be used to assess how well the inversion can recover fluxes. OSSEs can be used to quantify the magnitude and geographical distribution of uncertainty that stems from specific errors or assumptions in the inversion framework, such as transport model errors (e.g., Houweling et al., 2010; Berchet et al., 2015), spatiotemporal flux patterns within regions (e.g., Berchet et al. 2015), biased priors (e.g., Berchet et al. 2015), flux spatiotemporal resolutions (e.g., Wu et al., 2011), or parameter choices within computational data assimilation systems (e.g., Miyazaki et al., 2011, Chatterjee et al. 2012). Posterior flux errors and error covariances can be used to assess the impact of modelling simplifications or data limitations on the accuracy and precision of flux estimation (e.g., Berchet et al., 2015; Gourdji et al., 2010). OSSEs can also be used to understand sources of bias through a simple differencing of posterior and "true" fluxes (e.g., Locatelli et al., 2013; Thompson et al., 2011; Basu et al., 2016; Bloom et al., 2016). Similar tests can be run to determine the effects of observational biases and mistuning of error statistics on the accuracy of posterior estimates (e.g., Baker et al., 2010).

OSSEs can also be used to determine the sensitivity of inversions to transport errors. The model-data mismatch may be compared between an inversion that uses the "true" transport to calculate the sensitivity matrix versus that of an inversion that uses a different transport model (e.g., Chevallier et al., 2010; Houweling et al., 2010; Berchet et al., 2015; Locatelli et al., 2013). Assuming that the difference in performance between these two transport models is comparable to the difference between transport models used in real-data inversions, the inversion with inconsistent transport can be compared to the

Deleted: Taking biospheric CO_2 inversions as an example, assumptions about background CO_2 concentrations, boundary advection of CO_2, and fossil fuel fluxes can potentially have large uncertainties, and quantifying the sensitivity of an inversion to these factors is necessary in order to understand the uncertainty of the posterior fluxes. A study in which a small change in advected or background CO_2 leads to a much larger change in the posterior fluxes in the region of study, as in Göckede et al. (2010b), indicates that a relatively small bias in background CO_2 could produce large errors in the posterior estimates.

Deleted: possible

Deleted: bias or error. The prior flux is

Deleted: influential in inversions over biomes for which the magnitude, spatial distribution, and temporal cycles of fluxes are highly uncertain

Deleted: However, they

Deleted: Nevertheless,

Moved (insertion) [4]

Deleted: OSSEs are often

Moved (insertion) [5]

Moved (insertion) [6]

Deleted:).

inversion with consistent transport to determine how much the inconsistencies in transport affect the inversion. A similar test can be conducted simply by adding transport or chemistry errors to the pseudo-observations for one run of the model (e.g., Gourdji et al., 2010; Baker et al., 2010; Thompson et al., 2011). In addition, the meteorological forcing field may be perturbed independently of the transport model itself, to determine how the underlying meteorological assumptions affect the inversion; this is particularly important because the meteorology is often not optimized for transport runs (as noted by Berchet et al., 2015).

OSSEs are also useful for determining the sensitivity of the inversion to the choice of priors. Within a Bayesian inversion, perturbations of prior fluxes from the "true" fluxes in terms of spatial distribution, temporal distribution, and flux magnitude by region can be used for a synthetic data sensitivity test (e.g., Berchet et al., 2015). This type of study is useful for determining prior-related biases in cases when the bottom-up inventories for a particular trace gas in the model domain are highly uncertain. OSSEs can also provide information about how much information can be obtained from the current observational network. Pseudo-observation sites and types of data (for example, mixing ratios, profiles, column averages, or isotopic signatures from flask samples) can be added or taken away from the inversion to determine how the density and distribution of observations affect the precision and accuracy of the posterior flux field (Villani et al., 2010; Miyazaki et al., 2011; Hungershoefer et al., 2010; Shiga et al., 2013; Basu et al., 2016; Bloom et al., 2016). In addition, the ability of existing monitoring network sites to detect specific types of fluxes or flux patterns can be explored, as well as the impact of various sources of uncertainty on detection (e.g., Shiga et al., 2014; Fang et al., 2014; Miller et al., 2016a). Such experiments can determine how much information about the true flux field is provided by an observational network. The uncertainty reduction from the prior to the posterior estimates (see Sections 3.2 and 3.3.2) provides an overall metric for evaluating the information provided by hypothetical observations (e.g., Chevallier et al., 2010; Baker et al., 2010; Hungershoefer et al., 2010).

Finally, through sensitivity tests, OSSEs can help to determine optimal model resolution and observational averaging for obtaining the most accurate posterior fluxes. This has been done for model temporal resolution and observational temporal averaging (e.g., Gourdji et al., 2010). OSSEs can also be used to test the performance of the optimization of multiscale grids, which can decrease computational costs relative to regularly spaced grids (e.g., Wu et al., 2011).

4 Evaluation of existing diagnostics

We have presented diagnostics as an approach to the needs of quality control and of quality assurance for atmospheric inversion systems. The diagnostics that were presented in Section 3, in many ways, address this question well. The diversity of diagnostics may even give the impression that they can compensate for the lack of direct independent validation measurements described in Section 2, and thereby ensure statistical optimality of inverse modelling systems. Indeed, even uncertain parameters (hyperparameters) of the prior and observation error covariance matrices are optimisable from the assimilated data (e.g., Section 3.3.3). In most cases, however, such an interpretation would be overly optimistic. The diagnostic approaches described in Section 3 provide a crucial toolbox for evaluating and improving flux estimates obtained through the solution of

Moved up [4]: (e.g., Locatelli et al.,

Deleted: Such an experiment may also be done using an ensemble of synthetic data produced using different transport models in order to achieve a representative ensemble of model-data mismatches that may result from model transport inconsistencies in the inversion

Deleted: 2013).

Deleted: Rather than using independent transport models for these tests, one may also use transport models with only the model resolution altered relative to the transport used to obtain the pseudo-observations (e.g., Berchet et al., 2015). This type of test determines how the aggregation of sub-grid-scale transport processes affects the inversion.

Deleted: e.g.,

Moved up [5]: Similar tests can be run to determine the effects of observational biases and mistuning of error statistics on the accuracy of posterior estimates (e.g., Baker et al.,

Moved up [6]: Baker et al.,

Deleted: 2010).

Deleted: inconsistencies between the prior and true fluxes

Deleted: and

Deleted: 2016

Deleted: The diversity of diagnostics that were presented in Section 3 may atmospheric inverse problems. Without diagnostics, it is impossible to assess whether flux estimates are reliable, or to make sense of differences among alternative sets of estimates. At the same time, however, none of the presented approaches overcome the fundamental challenges described in Section 2. As such, the information provided by diagnostic tests must itself be taken with a proverbial "grain of salt," and it is equally important to be aware of the aspects of an inversion that cannot be evaluated using existing diagnostics as it is to assess those that can.

The key information lies in available measurements: diagnostics can only help to reformulate this information by bringing to light the impact of specific assumptions, in the same way that the atmospheric inversion reformulates observed concentrations in terms of surface fluxes, or that a retrieval scheme for an Earth observing system reformulates the measured radiance information into a geophysical quantity. For instance, the principle of objectively tuning error statistics for atmospheric inversions (e.g., Michalak et al., 2004; 2005) ultimately relies on disentangling deviations between prior flux assumptions and observations into components attributable to prior uncertainty versus model-data-mismatch errors. The attribution to these two components of error is based on leveraging differences in their space-time structure, however, and is made easier when the two sources of error have features that are statistically distinct (e.g., Desroziers et al., 2005). Alternatively, some of the statistics may be well known from some other information source and can then play the role of a fixed point to deduce the other ones (e.g., Kuppel et al., 2013). It is important to remember, however, that diagnostics cannot bring original information to the problem, but rather provide a framework for interpreting available information. This is particularly obvious when no real measurements are assimilated (the synthetic data experiments of Section 3.4).

The interpretation of diagnostics is also complicated by the fact that many of them are not independent of the underlying assumptions of the inversion systems themselves (e.g., independence of prior errors from model-data mismatch errors, uncorrelated nature of model-data-mismatch errors, linear observation operator, Gaussian error statistics, etc.). As a result, they may simply express the inadequacy of these assumptions rather than the misspecification of some particular component of the inversion setup. A common example is the inflation of observation error variances to compensate for neglecting observation error correlations, which yields a too small model-data-mismatch (see Section 3.2.2) that cannot be adequately resolved without removing the decorrelation hypothesis (e.g., Chevallier, 2007).

The comparison of inversion results with independent (un-assimilated) concentration measurements (Section 3.1.1) is also partly ambiguous, because an unknown fraction of the misfit is simply caused by the chemistry and transport model that simulates the independent measurements. Similarly, the interpretation of differences between inversion results and flux estimates from bottom-up inventories (Section 3.1.2) may revolve around estimating the uncertainty of the latter (see, e.g., the diverging conclusions of Chevallier et al. (2014) and Reuter et al. (2014) about the quality of the inferred carbon sink of Europe).

Sensitivity tests about some components of the inversion systems, like the chemistry and transport model (see Section 3.3.1), are implemented in an attempt to sample the same error statistics as those specified by the model-data-mismatch and prior error covariance matrices. In practice, however, they may instead reflect different opinions about the error statistics. For instance, intercomparisons of inversion results like those of Transcom (e.g., Gurney et al., 2002, Peylin et al., 2013) form

Deleted:

Deleted: certain statistical assumptions about

Deleted: .,

¨ensembles of convenience¨ rather than statistically-coherent ensembles. They may underestimate the quality of state-of-the-art inversions (because some systems would underperform due to particularly coarse horizontal resolution or due to an out-dated transport simulation configuration) as well as overestimate it (because the few participants cannot sample the whole uncertainty space). To represent inversion uncertainty, inversion intercomparisons should explore 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). However, this goal is usually hampered by limited resources that favour existing set-ups over the design of systematic explorations of other plausible and defensible set-ups.

Overall then, satisfying the diagnostics described in Section 3 is, strictly speaking, neither a sufficient nor a necessary condition for optimality (see also the discussion in Talagrand 2014). The degree of usefulness of diagnostics is proportional to the amount of information that is input to them; conversely, lack of independent information can lead to problems of equifinality, where similar apparent skill is achieving through widely different setups and assumptions. In some cases, the process of identifying and improving weak components of an inverse system itself represents an inference problem that may be ill-posed or under-determined. As a result, the interpretation of diagnostics itself often requires subjective expert knowledge.

Despite their ambiguity, however, the role and diversity of diagnostics has increased over the years, and this is an important and positive development. Indeed, the diagnostics described in Section 3 have proven their practical usefulness in understanding the behaviour of inversion systems, by providing a fresh perspective on inversion results. Moreover, they can reveal, or at least suggest, the presence of hidden flaws in inversion systems by shedding light on the symptoms of these flaws. As such, they form a critical basis for the credibility of the inversion approach to flux estimation. While existing diagnostics tools have limitations, some of which are unavoidable given the challenges described in Section 2, a careful review of the literature makes it clear that the implementation of diagnostics is a necessary step in the "exploration" of an inversion system.

5 Looking ahead

Atmospheric inversions are increasingly expected to contribute to national reporting of greenhouse gas emissions under future international treaties (see the discussions in Ogle et al., (2015) for biogenic emissions, Miller and Michalak (2017) for anthropogenic emissions, and Wu et al., (2016) for urban emissions). The routine run of atmospheric inversion systems will necessitate reinforcing the robustness and the transparency of their process through commonly agreed upon quality insurance and quality control procedures. In practice, this implies systematically providing reliable associated uncertainty statistics together with the posterior fluxes, and some evidence of the statistical consistency of these fluxes with the inversion assumptions. Such norms will have to rely on the systematic implementation of diagnostics of the type discussed here to a large extent, even for emerging applications like the quantification of urban emissions (McKain et al., 2012).

As we have seen in Section 4, many more measurements are needed to decrease diagnostics ambiguities. This requirement primarily relates to concentration measurements rather than flux measurements because scale mismatches usually hamper the

Deleted: ambiguity

Deleted: inversely

Deleted: .

Deleted: 2016

Deleted: concentration

Deleted: Such a

comparison of inversions with the latter (see Section 2). A step in data density may be achieved by hypothetical low cost sensors (Wu et al., 2016) or from future satellite imagers (e.g., Rayner et al., 2014), provided these new data do not suffer from significant systematic errors. Efforts to substantially increase observational coverage are already under way (see, e.g., http://www.climate-kic-centre-hessen.org/miriade.html, or Ciais et al., 2015), but the feasibility of sufficiently limiting systematic errors remains to be demonstrated.

Deleted: but reliable

Deleted:).

Interestingly, a (large) increase in the horizontal resolution of the inversion systems would also make it possible to incorporate direct flux measurements in the diagnostics, even when the targeted scales are coarser (see discussion in Section 2 and Lauvaux et al. (2009) or Meesters et al. (2012)). Inversion systems could also be run at very high resolution for the express purpose of comparing estimates to flux measurements. The validation with accurate flux measurements would avoid some of the ambiguity imposed by the chemistry and transport models on the concentration-based diagnostics.

Deleted: .

This would also open up new directions for diagnostics development. For example, direct comparison to flux observations would make it possible to better assess posterior uncertainties, for instance by building on diagnostics developed in the context of ensemble prediction systems – diagnostics that have not yet been used for atmospheric inversions (e.g., the reliability diagram of Talagrand et al., 1999). These ideas were explored, for example, by Broquet et al. (2013), using aggregates of flux measurements. Among other benefits, the direct validation of the posterior uncertainties would reveal possible departures from normality for flux errors, which may be especially important in the case of systematically positive emissions (e.g., Koohkan et al., 2013). Such diagnostics would certainly help to guide future developments of inversion systems.

Deleted: Candille and

Deleted: , 2005

Deleted:)

Taken together, it is clear that the importance of developing and implementing carefully-designed diagnostics for atmospheric inversions of long-lived greenhouse gases is only going to grow over time.

Deleted: for

Acknowledgments

We acknowledge the support from the International Space Science Institute (ISSI). This publication is an outcome of the ISSI's Working Group on "Carbon Cycle Data Assimilation: How to consistently assimilate multiple data streams". Support for Nina Randazzo was provided by the National Science Foundation under Grant No. 1342076.

References

Alexe, M., Bergamaschi, P., Segers, A., Detmers, R., Butz, A., Hasekamp, O., Guerlet, S., Parker, R., Boesch, H., Frankenberg, C., Scheepmaker, R., Dlugokencky, E., Sweeney, C., Wofsy, S. and Kort, E.: Inverse modelling of CH_4 emissions for 2010-2011 using different satellite retrieval products from GOSAT and SCIAMACHY, Atmos. Chem. Phys., 15, 113-133, doi: 10.5194/acp-15-113-2015, 2015.

Allen, D. T., Pacsi, A. P., Sullivan, D. W., Zavala-Araiza, D., Harrison, M., Keen, K., Fraser, M. P., Hill, A. D., Sawyer, R. F., and Seinfeld, J. H.: Methane Emissions from Process Equipment at Natural Gas Production Sites in the United States: Pneumatic Controllers, Environmental Science & Technology, 49, 633-640, 10.1021/es5040156, 2015.

Baker, D., Law, R., Gurney, K., Rayner, P., Peylin, P., Denning, A., Bousquet, P., Bruhwiler, L., Chen, Y., Ciais, P., Fung, I., Heimann, M., John, J., Maki, T., Maksyutov, S., Masarie, K., Prather, M., Pak, B., Taguchi, S. and Zhu, Z.: TransCom 3 inversion intercomparison: Impact of transport model errors on the interannual variability of regional CO_2 fluxes, 1988-2003, Global Biogeochem. Cycles, 20, doi: 10.1029/2004GB002439, 2006.

Baker, D., Bosch, H., Doney, S., O'Brien, D. and Schimel, D.: Carbon source/sink information provided by column CO_2 measurements from the Orbiting Carbon Observatory, Atmos. Chem. Phys., 10, 4145-4165, doi: 10.5194/acp-10-4145-2010, 2010.

Baldocchi, D., Falge, E., Gu, L. H., Olson, R., Hollinger, D., Running, S., Anthoni, P., Bernhofer, C., Davis, K., Evans, R., Fuentes, J., Goldstein, A., Katul, G., Law, B., Lee, X. H., Malhi, Y., Meyers, T., Munger, W., Oechel, W., U, K. T. P., Pilegaard, K., Schmid, H. P., Valentini, R., Verma, S., Vesala, T., Wilson, K., and Wofsy, S.: FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities, Bulletin of the American Meteorological Society, 82, 2415-2434, 10.1175/1520-0477(2001)082<2415:fantts>2.3.co;2, 2001.

Basu, S., Guerlet, S., Butz, A., Houweling, S., Hasekamp, O., Aben, I., Krummel, P., Steele, P., Langenfelds, R., Torn, M., Biraud, S., Stephens, B., Andrews, A., and Worthy, D.: Global CO_2 fluxes estimated from GOSAT retrievals of total column CO_2, Atmos. Chem. Phys., 13, 8695-8717, doi:10.5194/acp-13-8695-2013, 2013.

Basu, S., Miller, J., and Lehman, S.: Separation of biospheric and fossil fuel fluxes of CO_2 by atmospheric inversion of CO_2 and $^{14}(CO_2)$ measurements: Observation System Simulations, Atmospheric Chemistry and Physics, 16, 5665-5683, 10.5194/acp-16-5665-2016, 2016.

Beer, C., Reichstein, M., Tomelleri, E., Ciais, P., Jung, M., Carvalhais, N., Roedenbeck, C., Arain, M. A., Baldocchi, D., Bonan, G. B., Bondeau, A., Cescatti, A., Lasslop, G., Lindroth, A., Lomas, M., Luyssaert, S., Margolis, H., Oleson, K. W., Roupsard, O., Veenendaal, E., Viovy, N., Williams, C., Woodward, F. I., and Papale, D.: Terrestrial Gross Carbon Dioxide Uptake: Global Distribution and Covariation with Climate, Science, 329, 834-838, 10.1126/science.1184984, 2010.

Berchet, A., Pison, I., Chevallier, F., Bousquet, P., Conil, S., Geever, M., Laurila, T., Lavric, J., Lopez, M., Moncrieff, J., Necki, J., Ramonet, M., Schmidt, M., Steinbacher, M. and Tarniewicz, J.: Towards better error statistics for atmospheric inversions of methane surface fluxes, Atmos. Chem. Phys., 13, 7115-7132, doi: 10.5194/acp-13-7115-2013, 2013.

Berchet, A., Pison, I., Chevallier, F., Bousquet, P., Bonne, J. and Paris, J.: Objectified quantification of uncertainties in Bayesian atmospheric inversions, Geosci. Model Dev., 8, 1525-1546, doi: 10.5194/gmd-8-1525-2015, 2015.

Bergamaschi, P., Corazza, M., Karstens, U., Athanassiadou, M., Thompson, R., Pison, I., Manning, A., Bousquet, P., Segers, A., Vermeulen, A., Janssens-Maenhout, G., Schmidt, M., Ramonet, M., Meinhardt, F., Aalto, T., Haszpra, L., Moncrieff, J., Popa, M., Lowry, D., Steinbacher, M., Jordan, A., O'Doherty, S., Piacentino, S. and Dlugokencky, E.: Top-down estimates of European CH_4 and N_2O emissions based on four different inverse models, Atmos. Chem. Phys., 15, 715-736, doi: 10.5194/acp-15-715-2015, 2015.

Bloom, A., Lauvaux, T., Worden, J., Yadav, V., Duren, R., Sander, S., and Schimel, D.: What are the greenhouse gas observing system requirements for reducing fundamental biogeochemical process uncertainty? Amazon wetland CH_4 emissions as a case study, Atmospheric Chemistry and Physics, 16, 15199-15218, 10.5194/acp-16-15199-2016, 2016.

Bousquet, P., Ringeval, B., Pison, I., Dlugokencky, E., Brunke, E., Carouge, C., Chevallier, F., Fortems-Cheiney, A., Frankenberg, C., Hauglustaine, D., Krummel, P., Langenfelds, R., Ramonet, M., Schmidt, M., Steele, L., Szopa, S., Yver, C., Viovy, N. and Ciais, P.: Source attribution of the changes in atmospheric methane for 2006-2008, Atmos. Chem. Phys., 11, 3689-3700, doi: 10.5194/acp-11-3689-2011, 2011.

Bréon, F., Broquet, G., Puygrenier, V., Chevallier, F., Xueref-Remy, I., Ramonet, M., Dieudonne, E., Lopez, M., Schmidt, M., Perrussel, O. and Ciais, P.: An attempt at estimating Paris area CO_2 emissions from atmospheric concentration measurements, Atmos. Chem. Phys., 15, 1707-1724, doi: 10.5194/acp-15-1707-2015, 2015.

Broquet, G., Chevallier, F., Bréon, F., Kadygrov, N., Alemanno, M., Apadula, F., Hammer, S., Haszpra, L., Meinhardt, F., Morgui, J., Necki, J., Piacentino, S., Ramonet, M., Schmidt, M., Thompson, R., Vermeulen, A., Yver, C. and Ciais, P.: Regional inversion of CO_2 ecosystem fluxes from atmospheric measurements: reliability of the uncertainty estimates. Atmos. Chem. Phys., 13: 9039-9056, doi: 10.5194/acp-13-9039-2013, 2013.

Butler, M., Davis, K., Denning, A. and Kawa, S.: Using continental observations in global atmospheric inversions of CO_2: North American carbon sources and sinks, Tellus Series B-Chemical and Physical Meteorology, 62, 550-572, doi: 10.1111/j.1600-0889.2010.00501.x, 2010.

Chatterjee, A., A.M. Michalak, J.L. Anderson, K.L. Mueller, V. Yadav: Toward reliable ensemble Kalman filter estimates of CO_2 fluxes, *Journal of Geophysical Research – Atmospheres*, 117 (D22), D22306, doi:10.1029/2012JD018176, 2012.

Chevallier, F.: Impact of correlated observation errors on inverted CO_2 surface fluxes from OCO measurements, Geophys. Res. Lett., 34, L24804, doi:10.1029/2007GL030463, 2007.

Chevallier, F., Bréon, F. M., and Rayner, P. J.: Contribution of the Orbiting Carbon Observatory to the estimation of CO_2 sources and sinks: Theoretical study in a variational data assimilation framework, Journal of Geophysical Research, 112, D09307, doi:10.1029/2006JD007375, 2007.Rodgers, 2000;

Chevallier, F., Feng, L., Bosch, H., Palmer, P. and Rayner, P.: On the impact of transport model errors for the estimation of CO_2 surface fluxes from GOSAT observations, Geophys. Res. Lett., 37, doi: 10.1029/2010GL044652, 2010.

Chevallier, F. and O'Dell, C.: Error statistics of Bayesian CO_2 flux inversion schemes as seen from GOSAT, Geophys. Res. Lett., 40, 1252-1256, doi: 10.1002/grl.50228, 2013.

Deleted: Candille, G. and Talagrand, O.: Evaluation of probabilistic prediction systems for a scalar variable. Q.J.R. Meteorol. Soc., 131: 2131–2150. doi: 10.1256/qj.04.71, 2005.

Chevallier, F., Palmer, P., Feng, L., Boesch, H., O'Dell, C. and Bousquet, P.: Toward robust and consistent regional CO_2 flux estimates from in situ and spaceborne measurements of atmospheric CO_2, Geophys. Res. Lett., 41, 1065-1070, doi: 10.1002/2013GL058772, 2014.

Ciais, P., Rayner, P., Chevallier, F., Bousquet, P., Logan, M., Peylin, P., and Ramonet, M.: Atmospheric inversions for estimating CO_2 fluxes: methods and perspectives, Climatic Change, 103, 69-92, 10.1007/s10584-010-9909-3, 2010a.

Ciais, P., Canadell, J., Luyssaert, S., Chevallier, F., Shvidenko, A., Poussi, Z., Jonas, M., Peylin, P., King, A., Schulze, E., Piao, S., Rödenbeck, C., Peters, W. and Bréon, F.: Can we reconcile atmospheric estimates of the Northern terrestrial carbon sink with land-based accounting? Current Opinion in Environmental Sustainability, 2, 225-230, doi: 10.1016/j.cosust.2010.06.008, 2010b.

Ciais, P., D. Crisp, H. van der Gon, R. Engelen, M. Heimann, G. Janssens-Maenhout, P. Rayner, and M. Scholze, 2015: Towards a European Operational Observing System to Monitor Fossil CO_2 emissions - Final Report from the expert group. European Commission, Copernicus Climate Change Service, http://www.copernicus.eu/sites/default/files/library/CO2Report22Oct2015.pdf, 2015.

Conway, T., Tans, P., Waterman, L. and Thoning, K.: Evidence for interannual variability of the carbon cycle from the National Oceanic and Atmospheric Administration/Climate Monitoring and Diagnostics Laboratory Global Air Sampling Network, J. Geophys. Res.-Atmospheres, 99, 22831-22855, doi: 10.1029/94JD01951, 1994.

Corazza, M., Bergamaschi, P., Vermeulen, A., Aalto, T., Haszpra, L., Meinhardt, F., O'Doherty, S., Thompson, R., Moncrieff, J., Popa, E., Steinbacher, M., Jordan, A., Dlugokencky, E., Bruhl, C., Krol, M. and Dentener, F.: Inverse modelling of European N_2O emissions: assimilating observations from different networks,Atmos. Chem. Phys., 11, 2381-2398, doi: 10.5194/acp-11-2381-2011, 2011.

Cressot, C., Chevallier, F., Bousquet, P., Crevoisier, C., Dlugokencky, E., Fortems-Cheiney, A., Frankenberg, C., Parker, R., Pison, I., Scheepmaker, R., Montzka, S., Krummel, P., Steele, L. and Langenfelds, R.: On the consistency between global and regional methane emissions inferred from SCIAMACHY, TANSO-FTS, IASI and surface measurements, Atmos. Chem. Phys., 14, 577-592, doi: 10.5194/acp-14-577-2014, 2014.

Crisp, D., Fisher, B. M., O'Dell, C., Frankenberg, C., Basilio, R., Boesch, H., Brown, L. R., Castano, R., Connor, B., Deutscher, N. M., Eldering, A., Griffith, D., Gunson, M., Kuze, A., Mandrake, L., McDuffie, J., Messerschmidt, J., Miller, C. E., Morino, I., Natraj, V., Notholt, J., O'Brien, D. M., Oyafuso, F., Polonsky, I., Robinson, J., Salawitch, R., Sherlock, V., Smyth, M., Suto, H., Taylor, T. E., Thompson, D. R., Wennberg, P. O., Wunch, D., and Yung, Y. L.: The ACOS CO2 retrieval algorithm - Part II: Global X-CO2 data characterization, Atmospheric Measurement Techniques, 5, 687-707, 10.5194/amt-5-687-2012, 2012.

Desroziers, G., Berre, L., Chapnik, B., and Poli, P.: Diagnosis of observation, background and analysis-error statistics in observation space, Quarterly Journal of the Royal Meteorological Society, 131, 3385–3396, doi: 10.1256/qj.05.108, 2005.

Desroziers, G., Berre, L., Chapnik, B., and Poli, P.: Diagnosis of observation, background and analysis-error statistics in observation space, Quarterly Journal of the Royal Meteorological Society, 131, 3385–3396, doi:10.1256/qj.05.108, 2006.

Díaz Isaac, L. I., T. Lauvaux, K. J. Davis, N. L. Miles, S. J. Richardson, A. R. Jacobson and A. E. Andrews: Model-datacomparison of MCI field campaign atmospheric CO_2 mole fractions. J. Geophys. Res. Atmos., 119, 10,536–10,551, doi:10.1002/2014JD021593, 2014.

ECMWF: Quality of Our Forecasts: http://www.ecmwf.int/en/forecasts/quality-our-forecasts, access: 2016.

Enting, I. G., and Mansbridge, J. V.: Seasonal sources and sinks of atmospheric CO(2) Direct inversion of filtered data, Tellus Series B-Chemical and Physical Meteorology, 41, 10.1111/j.1600-0889.1989.tb00129.x, 1989.

Enting, I. G., and Mansbridge, J. V.: Latitudinal distribution of sources and sinks of co2 - results of an inversion study, Tellus Series B-Chemical and Physical Meteorology, 43, 156-170, 10.1034/j.1600-0889.1991.00010.x, 1991.

Enting, I. G.: Inverse Problems in Atmospheric Constituent Transport, Cambridge University Press, Cambridge, United Kingdom, 392 pp., 2002.

Fang, Y., Michalak, A.M., Shiga, Y.P. and Yadav, V.: Using atmospheric observations to evaluate the spatiotemporal variability of CO_2 fluxes simulated by terrestrial biosphere models, Biogeosci., 11, 6985-6997, doi: 10.5194/bg-11-6985-2014, 2014.

Fang, Y., and Michalak, A. M.: Atmospheric observations inform CO2 flux responses to enviroclimatic drivers, Global Biogeochemical Cycles, 29, 555-566, 10.1002/2014gb005034, 2015.

Frankenberg, C., Kulawik, S. S., Wofsy, S., Chevallier, F., Daube, B., Kort, E. A., O'Dell, C., Olsen, E. T., and Osterman, G.: Using airborne HIAPER Pole-to-Pole Observations (HIPPO) to evaluate model and remote sensing estimates of atmospheric carbon dioxide, Atmospheric Chemistry and Physics Discussion, 10.5194/acp-2015-961, 2016, 2016.

Friedlingstein, P., Meinshausen, M., Arora, V. K., Jones, C. D., Anav, A., Liddicoat, S. K., and Knutti, R.: Uncertainties in CMIP5 Climate Projections due to Carbon Cycle Feedbacks, Journal of Climate, 27, 511-526, 10.1175/jcli-d-12-00579.1, 2014.

Ganesan, A., Rigby, M., Zammit-Mangion, A., Manning, A., Prinn, R., Fraser, P., Harth, C., Kim, K., Krummel, P., Li, S., Muhle, J., O'Doherty, S., Park, S., Salameh, P., Steele, L. and Weiss, R.: Characterization of uncertainties in atmospheric trace gas inversions using hierarchical Bayesian methods, Atmos. Chem. Phys., 14, 3855-3864, doi: 10.5194/acp-14-3855-2014, 2014.

Göckede, M., Michalak, A., Vickers, D., Turner, D. and Law, B.: Atmospheric inverse modeling to constrain regional-scale CO_2 budgets at high spatial and temporal resolution, J. Geophys. Res.-Atmospheres, 115, doi:10.1029/2009JD012257, 2010a.

Göckede, M., Turner, D., Michalak, A., Vickers, D. and Law, B.: Sensitivity of a subregional scale atmospheric inverse CO_2 modeling framework to boundary conditions, J. Geophys. Res.-Atmospheres, 115, doi: 10.1029/2010JD014443, 2010b.

Gourdji, S., Hirsch, A., Mueller, K., Yadav, V., Andrews, A. and Michalak, A.:Regional-scale geostatistical inverse modeling of North American CO_2 fluxes: a synthetic data study, Atmos. Chem. Phys., 10, 6151-6167, doi: 10.5194/acp-10-6151-2010, 2010.

Gourdji, S., Mueller, K., Yadav, V., Huntzinger, D., Andrews, A., Trudeau, M., Petron, G., Nehrkorn, T., Eluszkiewicz, J., Henderson, J., Wen, D., Lin, J., Fischer, M., Sweeney, C. and Michalak, A.: North American CO_2 exchange: inter-comparison of modeled estimates with results from a fine-scale atmospheric inversion, Biogeosci., 9, 457-475, doi: 10.5194/bg-9-457-2012, 2012.

Gurney, K., Law, R., Denning, A., Rayner, P., Baker, D., Bousquet, P., Bruhwiler, L., Chen, Y., Ciais, P., Fan, S., Fung, I., Gloor, M., Heimann, M., Higuchi, K., John, J., Maki, T., Maksyutov, S., Masarie, K., Peylin, P., Prather, M., Pak, B., Randerson, J., Sarmiento, J., Taguchi, S., Takahashi, T. and Yuen, C.: Towards robust regional estimates of CO_2 sources and sinks using atmospheric transport models, Nature, 415, 626-630, doi: 10.1038/415626a, 2002.

Houweling, S., Aben, I., Bréon, F., Chevallier, F., Deutscher, N., Engelen, R., Gerbig, C., Griffith, D., Hungershoefer, K., Macatangay, R., Marshall, J., Notholt, J., Peters, W. and Serrar, S.: The importance of transport model uncertainties for the estimation of CO_2 sources and sinks using satellite measurements, Atmos. Chem. and Phys., 10, 9981-9992, doi: 10.5194/acp-10-9981-2010, 2010.

Houweling, S., Krol, M., Bergamaschi, P., Frankenberg, C., Dlugokencky, E., Morino, I., Notholt, J., Sherlock, V., Wunch, D., Beck, V., Gerbig, C., Chen, H., Kort, E., Rockmann, T. and Aben, I. A multi-year methane inversion using SCIAMACHY, accounting for systematic errors using TCCON measurements, Atmos. Chem. Phys., 14, 3991-4012, doi: 10.5194/acp-14-3991-2014, 2014.

Houweling, S., Bergamaschi, P., Chevallier, F., Heimann, M., Kaminski, T., Krol, M., Michalak, A. M., and Patra, P.: Global inverse modeling of CH_4 sources and sinks: An overview of methods, Atmos. Chem. Phys., 17, 235-256, doi:10.5194/acp-17-235-2017, 2017.

Hungershoefer, K., Bréon, F., Peylin, P., Chevallier, F., Rayner, P., Klonecki, A., Houweling, S. and Marshall, J.: Evaluation of various observing systems for the global monitoring of CO_2 surface fluxes, Atmos. Chem. Phys., 10, 10503-10520, doi: 10.5194/acp-10-10503-2010, 2010.

Jeong, S., Millstein, D., and Fischer, M. L.: Spatially Explicit Methane Emissions from Petroleum Production and the Natural Gas System in California, Environmental Science & Technology, 48, 5982-5990, 10.1021/es4046692, 2014.

Jiang, F., Wang, H., Chen, J., Machida, T., Zhou, L., Ju, W., Matsueda, H. and Sawa, Y.: Carbon balance of China constrained by CONTRAIL aircraft CO_2 measurements, Atmos. Chem. Phys., 14, 10133-10144, doi: 10.5194/acp-14-10133-2014, 2014.

Johnson, M. S., Xi, X., Jeong, S., Yates, E. L., Iraci, L. T., Tanaka, T., Loewenstein, M., Tadic, J. M., Fischer, M. L.: Investigating seasonal methane emissions in Northern California using airborne measurements and inverse modeling, Journal of Geophysical Research: Atmospheres, 121, 22, 13,753–13,767, doi:10.1002/2016JD025157, 2016.

Jung, M., Reichstein, M., Margolis, H. A., Cescatti, A., Richardson, A. D., Arain, M. A., Arneth, A., Bernhofer, C., Bonal, D., Chen, J., Gianelle, D., Gobron, N., Kiely, G., Kutsch, W., Lasslop, G., Law, B. E., Lindroth, A., Merbold, L., Montagnani, L., Moors, E. J., Papale, D., Sottocornola, M., Vaccari, F., and Williams, C.: Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations, Journal of Geophysical Research-Biogeosciences, 116, 10.1029/2010jg001566, 2011.

Deleted: .Discuss.,

Deleted: 2016-572, 2016

Kadygrov, N., Broquet, G., Chevallier, F., Rivier, L., Gerbig, C. and Ciais, P.: On the potential of the ICOS atmospheric CO_2 measurement network for estimating the biogenic CO_2 budget of Europe, Atmos. Chem. Phys., 15, 12765-12787, doi: 10.5194/acp-15-12765-2015, 2015.

Kalnay, E.: Atmospheric Modeling, Data Assimilation and Predictability, Cambridge University Press, Cambridge, United Kingdom, 2003.

Kim, H., Maksyutov, S., Glagolev, M., Machida, T., Patra, P., Sudo, K. and Inoue, G.: Evaluation of methane emissions from West Siberian wetlands based on inverse modeling, Environmental Research Letters, 6, doi: 10.1088/1748-9326/6/3/035201, 2011.

Kim, J., Kim, H. and Cho, C.: The effect of optimization and the nesting domain on carbon flux analyses in Asia using a carbon tracking system based on the ensemble Kalman filter, Asia-Pacific Journal of Atmospheric Sciences, 50, 327-344, doi: 10.1007/s13143-014-0020-y, 2014.

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

Kuppel, S., F. Chevallier and P. Peylin, 2013: Quantifying the model structural error in Carbon Cycle Data Assimilation Systems. Geosci. Model Dev., 6, 45-55, doi:10.5194/gmd-6-45-2013, 2013.

Lauvaux, T., Gioli, B., Sarrat, C., Rayner, P., Ciais, P., Chevallier, F., Noilhan, J., Miglietta, F., Brunet, Y., Ceschia, E., Dolman, H., Elbers, J., Gerbig, C., Hutjes, R., Jarosz, N., Legain, D., and Uliasz, M.: Bridging the gap between atmospheric concentrations and local ecosystem measurements, Geophysical Research Letters, 36, 10.1029/2009GL039574, 2009.

Lauvaux, T., Miles, N., Deng, A., Richardson, S., Cambaliza, M., Davis, K., Gaudet, B., Gurney, K., Huang, J., O'Keefe, D., Song, Y., Karion, A., Oda, T., Patarasuk, R., Razlivanov, I., Sarmiento, D., Shepson, P., Sweeney, C., Turnbull, J., and Wu, K.: High-resolution atmospheric inversion of urban CO2 emissions during the dormant season of the Indianapolis Flux Experiment (INFLUX), Journal of Geophysical Research-Atmospheres, 121, 5213-5236, 10.1002/2015JD024473, 2016.

Lauvaux, T., Schuh, A., Uliasz, M., Richardson, S., Miles, N., Andrews, A., Sweeney, C., Diaz, L., Martins, D., Shepson, P. and Davis, K.: Constraining the CO_2 budget of the corn belt: exploring uncertainties from the assumptions in a mesoscale inverse system, Atmos. Chem. Phys., 12, 337-354, doi: 10.5194/acp-12-337-2012, 2012.

Law, B. E., Falge, E., Gu, L., Baldocchi, D. D., Bakwin, P., Berbigier, P., Davis, K., Dolman, A. J., Falk, M., Fuentes, J. D., Goldstein, A., Granier, A., Grelle, A., Hollinger, D., Janssens, I. A., Jarvis, P., Jensen, N. O., Katul, G., Mahli, Y., Matteucci, G., Meyers, T., Monson, R., Munger, W., Oechel, W., Olson, R., Pilegaard, K., Paw, K. T., Thorgeirsson, H., Valentini, R., Verma, S., Vesala, T., Wilson, K., and Wofsy, S.: Environmental controls over carbon dioxide and water vapor exchange of terrestrial vegetation, Agricultural and Forest Meteorology, 113, 97-120, 10.1016/s0168-1923(02)00104-1, 2002.

Le Quéré, C., Moriarty, R., Andrew, R., Canadell, J., Sitch, S., Korsbakken, J., Friedlingstein, P., Peters, G., Andres, R., Boden, T., Houghton, R., House, J., Keeling, R., Tans, P., Arneth, A., Bakker, D., Barbero, L., Bopp, L., Chang, J., Chevallier, F., Chini, L., Ciais, P., Fader, M., Feely, R., Gkritzalis, T., Harris, I., Hauck, J., Ilyina, T., Jain, A., Kato, E., Kitidis, V., Goldewijk, K., Koven, C., Landschutzer, P., Lauvset, S., Lefevre, N., Lenton, A., Lima, I., Metzl, N., Millero, F., Munro, D., Murata, A., Nabel, J., Nakaoka, S., Nojiri, Y., O'Brien, K., Olsen, A., Ono, T., Perez, F., Pfeil, B., Pierrot, D., Poulter, B., Rehder, G., Rödenbeck, C., Saito, S., Schuster, U., Schwinger, J., Seferian, R., Steinhoff, T., Stocker, B., Sutton, A., Takahashi, T., Tilbrook, B., van der Laan-Luijkx, I., van der Werf, G., van Heuven, S., Vandemark, D., Viovy, N., Wiltshire, A., Zaehle, S. and Zeng, N.: Global Carbon Budget 2015, Earth System Science Data, 7, 349-396, doi: 10.5194/essd-7-349-2015, 2015.

Lindqvist, H., O'Dell, C. W., Basu, S., Boesch, H., Chevallier, F., Deutscher, N., Feng, L., Fisher, B., Hase, F., Inoue, M., Kivi, R., Morino, I., Palmer, P. I., Parker, R., Schneider, M., Sussmann, R., and Yoshida, Y.: Does GOSAT capture the true seasonal cycle of carbon dioxide?, Atmos. Chem. Phys., 15, 13023-13040, doi:10.5194/acp-15-13023-2015, 2015.

Liu, J., and Bowman, K.: A method for independent validation of surface fluxes from atmospheric inversion: Application to CO2, Geophysical Research Letters, 43, 7, 3502–3508, doi:10.1002/2016GL067828, 2016.

Locatelli, R., Bousquet, P., Chevallier, F., Fortems-Cheney, A., Szopa, S., Saunois, M., Agusti-Panareda, A., Bergmann, D., Bian, H., Cameron-Smith, P., Chipperfield, M., Gloor, E., Houweling, S., Kawa, S., Krol, M., Patra, P., Prinn, R., Rigby, M., Saito, R. and Wilson, C.: Impact of transport model errors on the global and regional methane emissions estimated by inverse modelling, Atmos. Chem. Phys., 13, 9917-9937, doi: 10.5194/acp-13-9917-2013, 2013.

Locatelli, R., Bousquet, P., Saunois, M., Chevallier, F. and Cressot, C.: Sensitivity of the recent methane budget to LMDz sub-grid-scale physical parameterizations, Atmos. Chem. Phys., 15, 9765-9780, doi: 10.5194/acp-15-9765-2015, 2015.

Lyon, D. R., Zavala-Araiza, D., Alvarez, R. A., Harriss, R., Palacios, V., Lan, X., Talbot, R., Lavoie, T., Shepson, P., Yacovitch, T. I., Herndon, S. C., Marchese, A. J., Zimmerle, D., Robinson, A. L., and Hamburg, S. P.: Constructing a Spatially Resolved Methane Emission Inventory for the Barnett Shale Region, Environmental Science & Technology, 49, 8147-8157, 10.1021/es506359c, 2015.

Masarie, K. A., Pétron, G., Andrews, A., Bruhwiler, L., Conway, T. J., Jacobson, A. R., Miller, J. B., Tans, P. P., Worthy, D. E. and Peters W.: Impact of CO_2 measurement bias on CarbonTracker surface flux estimates. J. Geophys. Res., 116, D17305, doi:10.1029/2011JD016270, 2011.

McKain, K., Wofsy, S. C., Nehrkorn, T., Eluszkiewicz, J., Ehleringer, J. R., and Stephens, B.: Assessment of ground-based atmospheric observations for verification of greenhouse gas emissions from an urban region, Proceedings of the National Academy of Sciences, 109, 8 8423-8428, doi:10.1073/pnas.1116645109, 2012.

Meesters, A., Tolk, L., Peters, W., Hutjes, R., Vellinga, O., Elbers, J., Vermeulen, A., van der Laan, S., Neubert, R., Meijer, H., and Dolman, A.: Inverse carbon dioxide flux estimates for the Netherlands, Journal of Geophysical Research-Atmospheres, 117, 10.1029/2012JD017797, 2012.

Michalak, A. M.: Atmospheric observations and inverse modeling approaches for identifying geographical sources and sinks of carbon, in: Land Use and the Carbon Cycle: Advances in Integrated Science, Management, and Policy, edited by: Brown, D. G., Robinson, D. T., French, N. H., and Reed, B. C., Cambridge University Press, New York, New York, 144 - 177, 2013.

Michalak, A., Bruhwiler, L., and Tans, P.: A geostatistical approach to surface flux estimation of atmospheric trace gases, Journal of Geophysical Research-Atmospheres, 109, 10.1029/2003JD004422, 2004.

Michalak, A., Hirsch, A., Bruhwiler, L., Gurney, K., Peters, W., and Tans, P.: Maximum likelihood estimation of covariance parameters for Bayesian atmospheric trace gas surface flux inversions, Journal of Geophysical Research-Atmospheres, 110, 10.1029/2005JD005970, 2005.

Miller, S., Wofsy, S., Michalak, A., Kort, E., Andrews, A., Biraud, S., Dlugokencky, E., Eluszkiewicz, J., Fischer, M., Janssens-Maenhout, G., Miller, B., Miller, J., Montzka, S., Nehrkorn, T. and Sweeney, C.: Anthropogenic emissions of methane in the United States, Proceedings of the National Academy of Sciences of the United States of America, 110, 20018-20022, doi: 10.1073/pnas.1314392110, 2013.

Miller, S., Worthy, D., Michalak, A., Wofsy, S., Kort, E., Havice, T., Andrews, A., Dlugokencky, E., Kaplan, J., Levi, P., Tian, H. and Zhang, B.: Observational constraints on the distribution, seasonality, and environmental predictors of North American boreal methane emissions, Global Biogeochem. Cycles, 28, 146-160, doi: 10.1002/2013GB004580, 2014.

Miller, S., Commane, R., Melton, J., Andrews, A., Benmergui, J., Dlugokencky, E., Janssens-Maenhout, G., Michalak, A.M., Sweeney, C. and Worthy, D.: Evaluation of wetland methane emissions across North America using atmospheric data and inverse modeling, Biogeosci., 13, 1329-1339, doi: 10.5194/bg-13-1329-2016, 2016a.

Miller, S., Miller, C., Commane, R., Chang, R., Dinardo, S., Henderson, J., Karion, A., Lindaas, J., Melton, J., Miller, J., Sweeney, C., Wofsy, S., and Michalak, A.: A multiyear estimate of methane fluxes in Alaska from CARVE atmospheric observations, Global Biogeochemical Cycles, 30, 1441-1453, 10.1002/2016GB005419, 2016b.

Miller, S. M. and Michalak, A. M.: Constraining sector-specific CO_2 and CH_4 emissions in the United States, Atmos. Chem. Phys., 17, 3963-3985, doi:10.5194/acp-17-3963-2017, 2017.

Miyazaki, K., Maki, T., Patra, P. and Nakazawa, T.: Assessing the impact of satellite, aircraft, and surface observations on CO_2 flux estimation using an ensemble-based 4-D data assimilation system, J. Geophys. Res.-Atmospheres, 116, doi: 10.1029/2010JD015366, 2011.

Nassar, R., Jones, D., Kulawik, S., Worden, J., Bowman, K., Andres, R., Suntharalingam, P., Chen, J., Brenninkmeijer, C., Schuck, T., Conway, T. and Worthy, D.: Inverse modeling of CO_2 sources and sinks using satellite observations of CO_2 from TES and surface flask measurements, Atmos. Chem. Phys., 11, 6029-6047, doi: 10.5194/acp-11-6029-2011, 2011.

Ogle, S.M., Davis, K.J., Lauvaux, T., Schuh, A., Cooley, D., West, T.O., Heath, L.S., Miles, N.L., Richardson, S.J., Breidt, F.J., Smith, J.E., McCarty, J.L., Gurney, K.R., Tans, P., Denning, A.S. An approach for verifying biogenic greenhouse gas emissions inventories with atmospheric CO_2 concentration data, Environmental Research Letters. 10(3): 034012, 2015.

Deleted: 2016

Deleted: .Discuss.,

Deleted: 2016-643, 2016

Osterman, G., Wennberg, P. O., Gunson, M., and Basilio, R.: OCO (Orbiting Carbon Observatory)-2 Science Validation Plan, National Aeronautics and Space Administration Jet Propulsion LaboratoryCalifornia Institute of Technology, 2011.

Pacala, S. W., Breidenich, C., Brewer, P. G., Fung, I., Gunson, M. R., Heddle, G., Law, B., Marland, G., Paustian, K., Prather, M., Randerson, J. T., Tans, P., Wofsy, S. C., Linn, A. M., Sturdivant, J., Busalacchi, J., Atnonio J., Bierbaum, R. M., Carbone, R., Dabberdt, W. F., Dow, K., Forbes, G. S., Held, I., Lee, A., Pierrehumbert, R. T., Prather, K., Smith, K. R., Snow, J. T., Vonder Haar, T. H., Zeng, X., Meehl, G. A., Elfring, C., Geller, L., Kraucunas, I., McConnell, M., Walser, M., Warden, T., Casola, J., Gaskins, R., Weller, K., Brown, L., Greenway, R., Freeland, S.-A., Purcell, A., Payne, R., and Banskota, S.: Verifying Greenhouse Gas Emissions: Methods to Support International Climate Agreements, National Research Council of the National Academies, Washington, D.C., 2010.

Pandey, S., Houweling, S., Krol, M., Aben, I., Chevallier, F., Dlugokencky, E. J., Gatti, L., Gloor, E., Miller, J. B., Detmers, R., Machida, T., Rockmann, T.: Inverse modeling of GOSAT-retrieved ratios of total column CH_4 and CO_2 for 2009 and 2010, Atmospheric Chemistry and Physics, 16, 5043–5062, doi:10.5194/acp-16-5043-2016, 2016.

Peters, W., Krol, M.C., Van Der Werf, G.R., Houweling, S., Jones, C.D., Hughes, J., Schaefer, K., Masarie, K.A., Jacobson, A.R., Miller, J.B., Cho, C.H., Ramonet, M., Schmidt, M., Ciattaglia, L., Apadula, F., Heltai, D., Meinhardt, F., Di Sarra, A.G., Piacentino, S., Sferlazzo, D., Aalto, T., Hatakka, J., Ström, J., Haszpra, L., Meijer, H.A.J., Van Der Laan, S., Neubert, R.E.M., Jordan, A., Rodó, X., Morguí, J.A., Vermeulen, A.T., Popa, E., Rozanski, K., Zimnoch, M., Manning, A.C., Leuenberger, M., Uglietti, C., Dolman, A.J., Ciais, P., Heimann, M. and Tans, P.P.: Seven years of recent European net terrestrial carbon dioxide exchange constrained by atmospheric observations. Global Change Biology, 16 (4), 1317 – 1337, doi: 10.1111/j.1365-2486.2009.02078.x, 2009.

Peylin, P., Houweling, S., Krol, M., Karstens, U., Rödenbeck, C., Geels, C., Vermeulen, A., Badawy, B., Aulagnier, C., Pregger, T., Delage, F., Pieterse, G., Ciais, P. and Heimann, M.: Importance of fossil fuel emission uncertainties over Europe for CO_2 modeling: model intercomparison, Atmos. Chem. Phys., 11, 6607-6622, doi: 10.5194/acp-11-6607-2011, 2011.

Peylin, P., Law, R. M., Gurney, K. R., Chevallier, F., Jacobson, A. R., Maki, T., Niwa, Y., Patra, P. K., Peters, W., Rayner, P. J., Rödenbeck, C. and Zhang, X.: Global atmospheric carbon budget: results from an ensemble of atmospheric CO_2 inversions, Biogeosci., 10, 6699-6720, doi:10.5194/bg-10-6699-2013, 2013.

Pickett-Heaps, C., Rayner, P., Law, R., Ciais, P., Patra, P., Bousquet, P., Peylin, P., Maksyutov, S., Marshall, J., Rödenbeck, C., Langenfelds, R., Steele, L., Francey, R., Tans, P. and Sweeney, C.: Atmospheric CO_2 inversion validation using vertical profile measurements: Analysis of four independent inversion models, J. Geophys. Res.-Atmospheres, 116, doi: 10.1029/2010JD014887, 2011.

Pillai, D., Gerbic, C., Kretschmer, R., Beck, V., Karstens, U., Neininger, B. and Heimann, M.: Comparing Lagrangian and Eulerian models for CO_2 transport - a step towards Bayesian inverse modeling using WRF/STILT-VPRM, Atmospheric Chem. Phys., 12, 8979-8991, doi: 10.5194/acp-12-8979-2012, 2012.

Rayner, P. J., Utembe, S. R., and Crowell, S.: Constraining regional greenhouse gas emissions using geostationary concentration measurements: a theoretical study, Atmos. Meas. Tech., 7, 3285-3293, doi:10.5194/amt-7-3285-2014, 2014.

Rayner, P.J., Michalak, A.M. and Chevallier, F.C.: Fundamentals of data assimilation, Geosci. Model Dev., Discuss. doi:10.5194/gmd-2016-148, 2016.

Reuter, M., Buchwitz, M., Hilker, M., Heymann, J., Schneising, O., Pillai, D., Bovensmann, H., Burrows, J., Bosch, H., Parker, R., Butz, A., Hasekamp, O., O'Dell, C., Yoshida, Y., Gerbig, C., Nehrkorn, T., Deutscher, N., Warneke, T., Notholt, J., Hase, F., Kivi, R., Sussmann, R., Machida, T., Matsueda, H. and Sawa, Y.: Satellite-inferred European carbon sink larger than expected, Atmos. Chem. Phys., 14, 13739-13753, doi: 10.5194/acp-14-3739-2014, 2014.

Richardson, A. D., Anderson, R. S., Arain, M. A., Barr, A. G., Bohrer, G., Chen, G., Chen, J. M., Ciais, P., Davis, K. J., Desai, A. R., Dietze, M. C., Dragoni, D., Garrity, S. R., Gough, C. M., Grant, R., Hollinger, D. Y., Margolis, H. A., McCaughey, H., Migliavacca, M., Monson, R. K., Munger, J. W., Poulter, B., Raczka, B. M., Ricciuto, D. M., Sahoo, A. K., Schaefer, K., Tian, H., Vargas, R., Verbeeck, H., Xiao, J., and Xue, Y.: Terrestrial biosphere models need better representation of vegetation phenology: results from the North American Carbon Program Site Synthesis, Global Change Biology, 18, 566-584, 10.1111/j.1365-2486.2011.02562.x, 2012.

Rivier, L., Peylin, P., Ciais, P., Gloor, M., Rödenbeck, C., Geels, C., Karstens, U., Bousquet, P., Brandt, J., Heimann, M., Aerocarb Experimentalists: European CO_2 fluxes from atmospheric inversions using regional and global transport models, Climatic Change, 103, 93-115 doi: 10.1007/s10584-010-9908-4, 2010.

Rodgers, C. D.: Inverse Methods for Atmospheric Sounding: Theory and Practice, Series on Atmospheric, Oceanic and Planetary Physics, World Scientific, 238 pp., 2000.

Saeki, T., Maksyutov, S., Saito, M., Valsala, V., Oda, T., Andres, R., Belikov, D., Tans, P., Dlugokencky, E., Yoshida, Y., Morino, I., Uchino, O. and Yokota, T.: Inverse Modeling of CO_2 Fluxes Using GOSAT Data and Multi-Year Ground-Based Observations, Sola, 9, 45-50, doi: 10.2151/sola.2013-011, 2013a.

Saeki, T., Maksyutov, S., Sasakawa, M., Machida, T., Arshinov, M., Tans, P., Conway, T., Saito, M., Valsala, V., Oda, T., Andres, R. and Belikov, D.: Carbon flux estimation for Siberia by inverse modeling constrained by aircraft and tower CO_2 measurements, J. Geophys. Res.-Atmospheres, 118, 1100-1122, doi: 10.1002/jgrd.50127, 2013b.

Schaefer, K., Schwalm, C. R., Williams, C., Arain, M. A., Barr, A., Chen, J. M., Davis, K. J., Dimitrov, D., Hilton, T. W., Hollinger, D. Y., Humphreys, E., Poulter, B., Raczka, B. M., Richardson, A. D., Sahoo, A., Thornton, P., Vargas, R., Verbeeck, H., Anderson, R., Baker, I., Black, T. A., Bolstad, P., Chen, J., Curtis, P. S., Desai, A. R., Dietze, M., Dragoni, D., Gough, C., Grant, R. F., Gu, L., Jain, A., Kucharik, C., Law, B., Liu, S., Lokipitiya, E., Margolis, H. A., Matamala, R., McCaughey, J. H., Monson, R., Munger, J. W., Oechel, W., Peng, C., Price, D. T., Ricciuto, D., Riley, W. J., Roulet, N., Tian, H., Tonitto, C., Torn, M., Weng, E., and Zhou, X.: A model-data comparison of gross primary productivity: Results from the North American Carbon Program site synthesis, Journal of Geophysical Research-Biogeosciences, 117, 10.1029/2012jg001960, 2012.

Schuh, A., Denning, A., Corbin, K., Baker, I., Uliasz, M., Parazoo, N., Andrews, A. and Worthy, D.: A regional high-resolution carbon flux inversion of North America for 2004, Biogeosci., 7, 1625-1644, doi: 10.5194/bg-7-1625-2010, 2010.

Schuh, A., Lauvaux, T., West, T., Denning, A., Davis, K., Miles, N., Richardson, S., Uliasz, M., Lokupitiya, E., Cooley, D., Andrews, A. and Ogle, S.: Evaluating atmospheric CO_2 inversions at multiple scales over a highly inventoried agricultural landscape, Global Change Biology, 19, 1424-1439, doi: 10.1111/gcb.12141, 2013.

Shiga, Y., Michalak, A.M., Kawa, S.R. and Engelen, R.J.: In-situ CO_2 monitoring network evaluation and design: A criterion based on atmospheric CO_2 variability,J. Geophys. Res.-Atmospheres, 118, 2007-2018, doi: 10.1002/jgrd.50168, 2013.

Shiga, Y., Michalak, A., Gourdji, S., Mueller, K. and Yadav, V.: Detecting fossil fuel emissions patterns from subcontinental regions using North American in situ CO_2 measurements, Geophys. Res. Lett., 41, 4381-4388, doi: 10.1002/2014GL059684, 2014.

Simmons, A. J., and Hollingsworth, A.: Some aspects of the improvement in skill of numerical weather prediction, Quarterly Journal of the Royal Meteorological Society, 128, 647-677, 10.1256/003590002321042135, 2002.

Steele, L. P., Dlugokencky, E. J., Lang, P. M., Tans, P. P., Martin, R. C., and Masarie, K. A.: Slowing down of the global accumulation of atmospheric methane during the 1980s, Nature, 358, 313-316, 10.1038/358313a0, 1992.

Subramanian, R., Williams, L. L., Vaughn, T. L., Zimmerle, D., Roscioli, J. R., Herndon, S. C., Yacovitch, T. I., Floerchinger, C., Tkacik, D. S., Mitchell, A. L., Sullivan, M. R., Dallmann, T. R., and Robinson, A. L.: Methane Emissions from Natural Gas Compressor Stations in the Transmission and Storage Sector: Measurements and Comparisons with the EPA Greenhouse Gas Reporting Program Protocol, Environmental Science & Technology, 49, 3252-3261, 10.1021/e35060258, 2015.

Tadić, J. M., Loewenstein, M., Frankenberg, C., Butz, A., Roby, M., Iraci, L. T., Yates, E. L., Gore, W., and Kuze, A.: A Comparison of In Situ Aircraft Measurements of Carbon Dioxide and Methane to GOSAT Data Measured Over Railroad Valley Playa, Nevada, USA, Ieee Transactions on Geoscience and Remote Sensing, 52, 7764-7774, 10.1109/tgrs.2014.2318201, 2014.

Takagi, H., Houweling, S., Andres, R., Belikov, D., Bril, A., Boesch, H., Butz, A., Guerlet, S., Hasekamp, O., Maksyutov, S., Morino, I., Oda, T., O'Dell, C., Oshchepkov, S., Parker, R., Saito, M., Uchino, O., Yokota, T., Yoshida, Y. and Valsala, V.: Influence of differences in current GOSAT X_{CO2} retrievals on surface flux estimation, Geophys. Res. Lett., 41, 2598-2605, doi: 10.1002/2013GL059174, 2014.

Takahashi, T., Olafsson, J., Goddard, J. G., Chipman, D. W., and Sutherland, S. C.: Seasonal-variation of CO2 and nutrients in the high-latitude surface oceans - a comparative-study, Global Biogeochemical Cycles, 7, 843-878, 10.1029/93gb02263, 1993.

Takahashi, T., Sutherland, S. C., Sweeney, C., Poisson, A., Metzl, N., Tilbrook, B., Bates, N., Wanninkhof, R., Feely, R. A., Sabine, C., Olafsson, J., and Nojiri, Y.: Global sea-air CO2 flux based on climatological surface ocean pCO(2), and seasonal biological and temperature effects, Deep-Sea Research Part Ii-Topical Studies in Oceanography, 49, 1601-1622, 10.1016/s0967-0645(02)00003-6, 2002.

Talagrand, O., Vautard, R. and Strauss, B.: 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 , 1999.

Talagrand, O. Errors. A posteriori diagnostics, in Advanced Data Assimilation for Geosciences Lecture Notes of the Les Houches School of Physics: Special Issue, June 2012, Edited by Éric Blayo, Marc Bocquet, Emmanuel Cosme, and Leticia F. Cugliandolo, Oxford University Press, 608 pp., 2014.

Tans, P. P., Fung, I. Y., and Takahashi, T.: Observational constraints on the global atmospheric co2 budget, Science, 247, 1431-1438, 10.1126/science.247.4949.1431, 1990.

Thompson, R., Bousquet, P., Chevallier, F., Rayner, P. and Ciais, P.: Impact of the atmospheric sink and vertical mixing on nitrous oxide fluxes estimated using inversion methods, J. Geophys. Res.-Atmospheres, 116, doi: 10.1029/2011JD015815, 2011.

Thompson, R., Ishijima, K., Saikawa, E., Corazza, M., Karstens, U., Patra, P., Bergamaschi, P., Chevallier, F., Dlugokencky, E., Prinn, R., Weiss, R., O'Doherty, S., Fraser, P., Steele, L., Krummel, P., Vermeulen, A., Tohjima, Y., Jordan, A., Haszpra, L., Steinbacher, M., Van der Laan, S., Aalto, T., Meinhardt, F., Popa, M., Moncrieff, J. and Bousquet, P.: TransCom N_2O model inter-comparison - Part 2: Atmospheric inversion estimates of N_2O emissions, Atmos. Chem. Phys., 14, 6177-6194, doi: 10.5194/acp-14-6177-2014, 2014.

van der Werf, G. R., Randerson, J. T., Giglio, L., Collatz, G. J., Kasibhatla, P. S., and Arellano, A. F., Jr.: Interannual variability in global biomass burning emissions from 1997 to 2004, Atmospheric Chemistry and Physics, 6, 3423-3441, 2006.

Villani, M., Bergamaschi, P., Krol, M., Meirink, J. and Dentener, F.: Inverse modeling of European CH_4 emissions: sensitivity to the observational network, Atmos. Chem. Phys., 10, 1249-1267, doi: 10.5194/acp-10-1249-2010, 2010.

WMO Lead Centre for Deterministic Forecast Verification (WMO-LCDNV): http://apps.ecmwf.int/wmolcdnv/, access: 10 June 2016.

Wu, L., Bocquet, M., Lauvaux, T., Chevallier, F., Rayner, P. and Davis, K.: Optimal representation of source-sink fluxes for mesoscale carbon dioxide inversion with synthetic data, J. Geophys. Res.-Atmospheres, 116, doi: 10.1029/2011JD016198, 2011.

Wu, L., Bocquet, M., Chevallier, F., Lauvaux, T. and Davis, K.: Hyperparameter estimation for uncertainty quantification in mesoscale carbon dioxide inversions, Tellus Series B-Chemical and Physical Meteorology, 65, doi: 10.3402/tellusb.v65i0.20894, 2013.

Wu, L., Broquet, G., Ciais, P., Bellassen, V., Vogel, F., Chevallier, F., Xueref-Remy, I., and Wang, Y.: What would dense atmospheric observation networks bring to the quantification of city CO2 emissions?, Atmospheric Chemistry and Physics, 16, 7743-7771, 10.5194/acp-16-7743-2016, 2016.

Wunch, D., Toon, G. C., Blavier, J.-F. L., Washenfelder, R. A., Notholt, J., Connor, B. J., Griffith, D. W. T., Sherlock, V., and Wennberg, P. O.: The Total Carbon Column Observing Network, Philosophical Transactions of the Royal Society a-Mathematical Physical and Engineering Sciences, 369, 2087-2112, 10.1098/rsta.2010.0240, 2011.

Yadav, V., Michalak, A. M., Ray, J., Shiga, Y. P.: A statistical approach for isolating fossil fuel emissions in atmospheric inverse problems, Journal of Geophysical Research: Atmospheres, 121, 20, 12,490 – 12,504, doi:10.1002/2016JD025642, 2016.

Zhang, H., Chen, B., van der Laan-Luijkx, I., Chen, J., Xu, G., Yan, J., Zhou, L., Fukuyama, Y., Tans, P. and Peters, W.: Net terrestrial CO_2 exchange over China during 2001-2010 estimated with an ensemble data assimilation system for atmospheric CO_2, J. Geophys. Res.-Atmospheres, 119, 3500-3515, doi:10.1002/2013JD021297, 2014.

Zimmerle, D., Williams, L., Vaughn, T., Quinn, C., Subramanian, R., Duggan, G., Willson, B., Opsomer, J., Marchese, A., Martinez, D., and Robinson, A.: Methane Emissions from the Natural Gas Transmission and Storage System in the United States, Environmental Science & Technology, 49, 9374-9383, 10.1021/acs.est.5b01669, 2015.

Page 6: [1] Deleted Anna M. Michalak 5/9/17 3:22:00 PM

). For example, vertical concentration profiles captured by airborne observations provide information about vertical mixing and convection.

For

Page 8: [2] Deleted Anna M. Michalak 5/9/17 3:22:00 PM

; Shiga et al., 2014; Miller et al., 2016).

Second, the posterior uncertainties themselves and the assumptions about model-data-mismatch errors can be assessed by analysing the deviations of modelled concentrations resulting from posterior flux estimates from the concentration observations used to constrain the inversion. The uncertainties associated with modelled concentrations can be calculated based on the posterior uncertainties of the fluxes and the information provided by the chemistry and transport model. These uncertainties can then be evaluated against the actual

Page 9: [3] Deleted Anna M. Michalak 5/9/17 3:22:00 PM

A number of other diagnostics are sometimes used to evaluate the overall statistical setup of inversions. For example, general trends such as the seasonal cycle of model-data mismatch can determine