

Responses to anonymous referee #1:

Main comment:

The paper presents an application of the method developed by Bousserez et al. (2015) for Bayesian posterior uncertainty quantification. The paper is well-written and contains some interesting parts, but a series of simplifications severely limits its value. For instance neglecting correlated model errors for the assimilation of profile retrievals makes the whole discussion about the multi-spectral instrument useless. The other results alone are not enough to populate a paper. Another example is the test about boundary conditions: assuming that their uncertainty results in a single continental offset for the whole period does not look like the real world. More details are given hereafter.

Response:

We would like to thank the anonymous referee for their useful remarks and suggestions that helped improve the manuscript. The revised version of the paper (attached) includes significant modifications and new results that we hope address the referee's comments. In particular, vertical model error correlations for the multi-spectral retrieval experiment are now accounted for, and a more realistic setup (using random noise instead of a single offset) has been adopted to test the sensitivity of the optimization to both boundary and initial conditions. Please see below our detailed responses to the remarks and suggestions. Note that in addition to the new results produced to address the referee's comments, some errors were identified in our previous simulations and have been corrected since (in particular in the boundary condition sensitivity study). Therefore, the entire manuscript has been modified accordingly and in our responses we only point to the modifications directly related to the referee's comments and suggestions.

Detailed comments:

1. p. 19018, l. 14 and elsewhere: why is there an “s” at the end of DOF when the plural is not used? Also note that the DOF is defined again in p. 19023 and 19026.

Response:

DOFs has been replaced by DOF throughout the manuscript. Also, it is now defined only once in the abstract and main text.

2. p. 19021, l. 5-10: the authors suggest that nobody has used Monte Carlo or numerical approximations of the Hessian because of their “prohibitive” cost, but looking at the results shown by, e.g., Meirink et al. (2008) or Cressot et al. (2014) with them, such approaches look straight-forward.

Response:

It is now clarified that previous studies have used Monte-Carlo and inverse Hessian approximations to quantify the information content of the inversion. However, such approaches may be computationally challenging, since for some applications the number of iterations required for convergence (either for optimization or inverse

Hessian estimates) can be prohibitive (e.g., in our case a one-month methane emission optimization requires more than 50 iterations). This is better explained in the revised manuscript. The references mentioned have also been added. For more details, please see modifications in the text of the revised paper (introduction, paragraph 4, in red).

3. p. 19022, l. 6: providing -> provided.

Response:

Has been corrected.

4. p. 19023, l. 16: why is B diagonal? I understand that this conveniently simplifies the algorithm but the authors should explain why it makes physical sense. Why would the diffuse emissions seen in Fig. 1 have uncorrelated prior errors every 50 km? I note that the two references above used a 500 km e-folding correlation length.

Response:

Accurately defining error correlations in bottom-up inventories is a challenging problem due to the sparsity of available flux measurements, and is beyond the scope of our study. Here our primary focus is to understand the relative benefit of different instrumental designs to constrain methane fluxes, so this simplification should not affect significantly our conclusions. However, we now emphasize that our diagonal B assumption is overly optimistic. See text in red in the last paragraph of Section 2.2 for more details.

5. p. 19024, l. 19: Does the 40% relative error apply to grid cell emissions or to the whole domain? Does this number correspond to 1 or 2 sigma? In any case, the authors should clearly indicate the monthly error budget integrated over their domain and give some indication of its realism. This point is particularly important for a study of uncertainty reduction.

Response:

It is now clarified that a 40% error standard deviation is considered for the emissions in each grid-cell. Also, the monthly error budget over all North America (2.9Tg/month) is now indicated and its magnitude compared with previous findings. Please see added text in the revised paper in Section 2.2., last paragraph, in red.

6. p. 19025, l. 1: The authors assimilate profile retrievals. For such a product, model errors are highly correlated between levels and accounting for them is critical (which actually explains why everybody assimilates columns as far as I know).

Response:

The multi-spectral configuration now takes into account model error correlations between vertical levels. Comparisons with in situ data (HIPPO, NOAA flasks measurements) were used to define the model error variances in the boundary layer (BL) and in the free troposphere (FT). Uncorrelated errors were assumed between the BL and the FT, based on the decoupling of the physical processes between these two regions and the in situ comparisons. Error correlations of 1 were assumed within each of those regions. Therefore our results can be seen as representative of a pessimistic

scenario (i.e., lower bounds on the observational constraints). The modifications to the previous setup are now detailed in the revised manuscript in Section 2.3, which has been entirely revised (in particular, see text in red).

7. p. 19025, l. 3-6: the two sentences should be developed to better explain what the authors have used.

Response:

A more detailed description of the multi-spectral retrieval has been included in the revised manuscript (see red text in first paragraph of Section 2.3).

8. p. 19025, l.18: The authors write “This value [8ppb] is consistent with GOSAT column errors reported in Parker et al. (2011).” The reader may guess that the value corresponds to 1 sigma, but in this case the link with Parker et al. is weird. Parker et al. actually write: “from comparisons to TCCON observations we have inferred a single sounding precision for our CH₄ retrievals of 0.4 – 0.8% with estimated biases between -17 ppb and 2 ppb (0.1 to 30.9%)” (their §32). Basically the authors have taken the smallest value in the range for the standard deviation and have neglected the large biases reported by Parker et al.

Response:

We now use a 12 ppb standard error for XCH₄, which is in the middle-range of the errors found in Parker et al. (2011). In our OSSE, biases are not taken into account, since biases can be estimated and removed when performing a real inversion (see, e.g., Wecht et al., 2014). The new setup for the observational errors is detailed in Section 2.3, paragraph 3, in red.

9. p. 19025, l. 22-25: the authors rightly warn the reader against model errors, but such errors are spatially correlated while the authors neglect observation error correlations (p. 19026, l. 10). Also note that retrieval errors themselves are correlated in the real world.

Response:

The reviewer refers to observational error spatial correlations (which include model transport error correlations). Indeed those spatial error correlations are neglected in our study, since accurately estimating them is very challenging and would require extensive comparison with in situ measurements combined with sophisticated localization techniques (due to the small sampling available), which is out of the scope of this study. However, this limitation is now mentioned in the conclusion of the revised paper. Also, note that our study focuses principally on the relative merit of different observational system configurations, whose analysis should not be too sensitive to this simplification.

10. p. 19026, l. 17-18: for a given instrument, the retrieval errors vary with the satellite altitude. How is this dependency accounted for?

Response:

For a given instrument, the altitude of the satellite is fixed. In case the question refers to the variability of the averaging kernel and covariance error profiles for different

locations, the response is that one averaging kernel (and therefore one error profile) is considered for each instrument. Indeed, a larger ensemble of averaging kernels describing a potential range of sensitivities is beyond the scope of this study given the computational cost. However, based on knowledge of thermal IR (e.g., TES) and total column (e.g., GOSAT) retrievals, use of a single averaging kernel is a reasonable approximation as our study is constrained to Northern Hemisphere summertime where the temperature and sunlight conditions provide sufficient signal for the present evaluation, and because our study looks at the relative merits of different observing approaches (see updated description of the averaging kernels in Section 2.2).

11. p. 19027, l. 17-19: this artifact and the accompanying remark suggest that the control vector is not defined appropriately.

Response:

We have now included 3-day inversion results in the revised manuscript. The results from the 3-day, one-week, and one-month inversions essentially show that the observational constraints on the methane fluxes reach a maximum after only 3-day (possibly even less, since previous regional inversion studies based on geostationary measurements have investigated even shorter time-periods, as explained in the first paragraph of Section 3.1 of the revised manuscript). Rather than being an incorrect definition of the control vector, a one-month (or even one-week) time-window simply does not fully exploit the capability of the geostationary measurements in term of the temporal resolution of their constraints on the optimized methane fluxes.

12. p. 19027, l. 28-29: This claim is tied to the realism of the modeling framework and may therefore not be reliable.

Response:

We agree that due to the lack of error correlations in the definition of our prior (B), the DOF we derived is likely too optimistic. We now clarify this in the revised text (see Section 3.1, last paragraph, text in red). However, in the absence of meaningful information to accurately determine prior error correlations, an advantage of our analysis is that it reveals the spatial extend of the constraints pertaining to measurements only, which is useful information and provides an upper-bound on the spatial resolution of the constraints.

13. Section 3.2. What about the initial state of the simulation? How is it accounted for here and what is the impact of a biased initial state? What happens with more realistic error structures (e.g., decoupled errors at the edges both in space and time)?

Response:

The sensitivity studies have been entirely revised. In the revised manuscript (Section 3.3), we now present sensitivity results for both the initial state and the boundary conditions, with realistic random perturbations derived from model-data comparisons. In particular, different perturbations of the initial state are defined for the boundary layer and the free troposphere, with standard deviation of 22 ppb and 46 ppb,

respectively. For the boundary conditions, random perturbations with standard deviation of 16 ppb were used throughout the troposphere. See Section 3.3 of the revised manuscript for more details. Also, a bug was found in our code during the revision of the paper, which explains the very different results obtained in the revised version for the boundary conditions sensitivity experiment.

14. p. 19029, l. 25: the estimate may be mathematically rigorous, but not so realistic. The word "rigorous" is therefore not appropriate.

Response:

The word "rigorous" has been removed. The sentence has been replaced by "For the first time, a grid-scale estimate of the information content of a~high-resolution inversion...".