Responses to Reviewer 2

We thank the reviewer for their comments and questions. Our responses are formatted as follows:

The reviewer's comment/question (numbered) is written in black italic text.

> Our responses are written in normal black text (indented).
>
> The revised text as it appears in the manuscript is written in normal blue text (indented), with relevant changes underlined.

Line numbers refer to the edited manuscript. We have also provided a tracked-changes document, but that has different line numbers.

My first comment (and very important one) here is that this paper has nothing to do with atmospheric measurement techniques and therefore its exact home is GMD and not AMT. Specifically, no "new" measurements are collected or discussed as part of this paper or for that matter no new measurement techniques are also suggested as part of this paper. [Outside the scope of the Journal]

> Thank you for your comment. Our paper concerns the interpretation of observations to infer secondary quantities (e.g., emissions). Other similar papers have been published by AMT in the past (e.g., Varon et al. 2018, Alden et al. 2018).

The authors suggest two new techniques for reducing the cost of computing the Jacobian i.e., reduced rank and reduced dimension methods. First, these are not the only way to reduce the computational size of the problem.

(1) The generally accepted solution to reduce the size of the problem is the one suggested in the paper: "Measuring information content from observation for data assimilation: relative entropy versus Shannon entropy difference" and I would suggest the authors to review this paper. Thus, I would like to see the techniques suggested in this paper in comparison to those mentioned in the paper mentioned above. Note, these issues are nothing new and have been dealt with since 1974. (see paper the information content of remote measurements of atmospheric temperature by satellite infra-red radiometry and optimum radiometer configurations.). Eventually, it is the question of the information content of the observations and not reducing the size of the Jacobian or the information content as expressed through an Averaging Kernel. I would like to see the difference in the answer as received from the method described in Xu's paper in comparison to what is shown in this paper.

> Thank you for your suggestion. Xu (2007) describes the dependence of two measures of information content (the Shannon and relative entropy differences) on optimal reductions in the dimension of the observation vector. We clarified the dependence of the computational cost on the dimension of the state vector. We also added references to Xu et al. (2007) to the introduction and to our discussion of measures of information content.
>
> When $m \gg n$, as for inversions of satellite observations, the Jacobian can be constructed column-wise by conducting $n + 1$ CTM simulations to perturb each of the state vector elements x_i and obtain the corresponding column $\partial \mathbf{y}/\partial x_i$. (L58 – L60)
>
> Several methods have been proposed to decrease the computational cost of high-resolution analytical inversions by optimally reducing the dimension or rank of the observations or state vector. Approaches that reduce the dimension of the observation vector (e.g., Xu, 2007) reduce the computational cost of solving the inversion but not of constructing the Jacobian matrix. Approaches that decrease the dimension of the state vector lower the cost of both computations. (L82 – L85)
>
> The fraction of information content explained by the first i columns of $\mathbf{\Gamma}^*$ is the sum of the i largest eigenvalues divided by the total DOFS (Bousserez and Henze, 2018). The

eigenvalues can also be related to other measures of information content, including the Shannon and relative entropy differences (Rodgers, 2000; Xu, 2007). (L178 – L181)

(2) Please also look at the paper "Stable Signal Recovery from Incomplete and Inaccurate Measurements from Candes, Romberg and Terence Tao" to understand the mathematical theory behind it. For application in atmospheric inversions see: A sparse reconstruction method for the estimation of multi-resolution emission fields via atmospheric inversion

Thanks for your comment. We added a citation to Ray et al. (2015) (*A sparse reconstruction method*).

Other approaches that decreased the dimension of the state vector assumed knowledge of the Jacobian matrix (e.g., Rigby et al., 2011; Thompson and Stohl, 2014; Ray et al., 2015; Lunt et al., 2016; Liu et al., 2017). (L91 – L93)

(3) Following with the previous discussion if you have prior information, then you can aggregate grids where you do not have any chances of encountering methane fluxes without doing a two-step inversion. What is the point of solving for methane fluxes in the deserts of Nevada, Utah and Arizona (see Figure 2 in paper; you have regular grid) unless you expect deserts of Nevada to be big sources of methane emissions? For example, if you do this exercise globally then you would not be solving for methane fluxes in Sahara Desert (no unique information is provided by multitude of observations, even if theoretically a satellite can collect thousands of them). Hence even if the trace of the averaging kernel might show that you can better resolve fluxes in the Sahara Desert solving for these fluxes would be just meaningless implying that you can aggregate your grid.

Thank you for your comment. We have clarified that the averaging kernel sensitivities are low in areas known to have low emissions. The reduced-dimension method therefore functionally considers both the distribution of prior emissions and the observational density to generate a multiscale grid.

$\mathbf{A}$ can be calculated as $\mathbf{A} = \mathbf{I} - \hat{\mathbf{S}}\mathbf{S}_A^{-1}$ or equivalently as

$$\mathbf{A} = \mathbf{S}_A\mathbf{K}^T(\mathbf{K}\mathbf{S}_A\mathbf{K}^T + \mathbf{S}_O)^{-1}\mathbf{K}. \quad (4)$$

Equation (4) expresses the dependence of the averaging kernel matrix on the forward model and both error covariance matrices. The diagonal elements of $\mathbf{A}$ are commonly referred to as the averaging kernel sensitivities. They are highest in highly observed locations with uncertain, high emissions and lowest in poorly observed areas or in regions known to have low emissions. (L131 – L139)

(4) Please also remember that once you go from coarser resolution to finer resolution your posterior variance of the inverse problem is guaranteed to increase. Hence, please explain or mathematically show how does the reduction in posterior variance translate from coarser resolution to finer resolution (not in terms of R i.e., correlation). Can an upper bound be found and does it have spatial structure i.e., what has happened to the error you obtained from the

inversion (second part of equation 2)? Furthermore, what has happened to the trace of the averaging kernel. How has it distributed your trace at finer resolution?

Thank you for your comment. We aren't sure what you mean to ask here because we go from finer resolution to coarser resolution, not vice versa. We believe your question may be answered by the bottom row of Figure 3, which shows the distribution of the trace of the averaging kernel (the averaging kernel sensitivities) in each of the proposed methods. We believe that this may also answer your question about the distribution of the posterior error, since the averaging kernel is a measure of the relative reduction in error from the prior to the posterior: $\mathbf{A} = \mathbf{I} - \hat{\mathbf{S}}\mathbf{S}_{\mathrm{A}}^{-1}$.