We thank the referees for the additional thoughtful comments on the manuscript. Below, we respond separately to each referee. The comments from the referees are in **bold Calibri** font and our responses are in plain Times New Roman font.

Report 1 (Referee #3)

General comments:

As already stated in the review of the first version of the paper, the effort at assessing the errors in the (chemistry-)transport model before going into flux inversions is necessary and the issue remains too rarely treated adequately in inversion papers. It is therefore a very good idea to present a methodology to tackle this issue. The explanations of the methodology and most of the interpretations of the results, regarding mainly biases in the transport, are clear and interesting and the revisions made after the first review have lead to a still clearer description of the work. The reduced number of figures makes it more focussed and easier to read, which I appreciate a lot.

A debate on the terminology has been launched during the first review: using 4D-Var, 3D-Var or variational for the methodology. The authors wish to stick to using 4D-Var and I agree with them that this kind of paper is not the place for deciding on the exact terminology of our domain. Since the description they added makes everything clear mathematically, I think we can drop the debate on the naming of the method.

Specific comments:

****Section 2: data and method:**
- p.8 l.11 and in Eq.5: the way the equation is written is valid for N the number of hourly time steps but actually, there are not N u_i (they vary over 3 days or 1 month, as stated elsewhere) or N Q_i (there is only one as it is constant). Could the equation be written in a more general way so that it would be easy to refer to it directly for any of the experiments?

The equations certainly could be rewritten to make this explicitly clear. However, it would require changing Eqs. 5, 6, and 10 and adding two additional equations to link the different indices. We are reluctant to introduce these changes at this point in the review process since the equations currently in the manuscript are correct. However, if the Editor believes the manuscript would benefit from making these changes, we will do so.

- p.8 l.15 in Eq.6: same remark

See our response above.

- p.9 l.14: is convergence guarranted?

Yes. The 4D-Var scheme works and converges, as shown by our results. However, if there are biases in the observations that are strongly inconsistent with the modeled state, or if the error

covariances are poorly specified, then the code could fail to converge to a global minimum, which would be the case with many assimilation methods.

- p.9 l.22: "the level of noise was estimated to correspond to GOSAT XCH4 uncertainty": do you mean that the point at which the inversions stopped was afterwards found to match the 10 ppb value provided for GOSAT data?

Yes. We calculated the reduced chi-squared for the model fit to the GOSAT observations and ensured that it was about unity. We have added text to make this clear.

- p.9 l.26: "R was assumed to be diagonal": and the diagonal was filled from the 13 ppb value given in the previous sentence? An explicit link between the two sentences would make it clearer.

We have rewritten these sentences to make this clearer.

- p.10 l.1: "50% uncertainty on CH4 emissions in each surface grid box": how much Tg/y does it lead to? Is it realistic?

This is applied at the grid box scale and we have not aggregated these errors to the global scale to get a global error in Tg/yr. However, this 50% error is consistent with other estimates used in the literature. For example, Maasakkers et al. (2019) assumed a 50% error for anthropogenic emissions and estimated a mean error of 58% for wetland emissions at the GEOS-Chem 4º x 5 º grid box scale. For non-wetland, natural emissions they assumed an error of 100%. However, in specifying the total error in each 4º x 5 º grid box they capped the errors at 50%.

Maasakkers et al., Global distribution of methane emissions, emission trends, and OH concentrations and trends inferred from an inversion of GOSAT satellite data for 2010–2015, ACP, https://doi.org/10.5194/acp-19-7859-2019.

- p.11 l.5-6: "to independent observations did not change noticeably": did not change when q increased?

Yes. We have added “larger” to the sentence so that it now reads “for larger values of q above 50 ppb, the fit of optimized CH_4 fields to independent observations did not change noticeably.”

- p.11 l.9: "the WC method was still able to significantly improve": this sentence is not very clear to me: what does the "still" refer to?

This sentence was unclear. We have changed it to state that “As shown in the experiments described in Sect. 2.4, the WC method was able to improve the model and capture the bias in the CH_4 state with q set to 50 ppb.”

- p.11 l.29: "The bias in vertical transport and chemistry": this is for the first two OSSEs, respectively, isn't it?

Yes. We have modified the sentence to read: “In the first and second OSSEs, the bias in vertical transport and chemistry were introduced by turning off convection and chemistry, respectively,...”

- p.12 l.1: "and have the freedom to": this part of the sentence is a bit strange. Do you mean that the bias can be placed anywhere in this configuration?

This was confusing. We mean that the WC scheme has the freedom to determine the location of the bias. We have changed this so that it now states "We configured the WC method to carry out "full state assimilation" (as described in Sect. 2.3) to enable the optimization to determine independently the location and magnitude of the bias in the modeled state."

- p.12 l.6-7: "We also conducted SC 4D-Var assimilation experiment for comparisons with the WC approach in the OSSE with biased surface emissions.": no OSSE with biased surface emissions is referred to before.

Thanks for catching this. This OSSE was removed during the last revision. We have fixed the text.

- p.12 l.22-23: "Short time windows": how long?

We mean less than 2 days, for example. We have added this to the text.

- p.12 l.24-25: "for short temporal correlation length scales": here, all matrices are diagonal, so that there are no correlation lengths used at all, is that right?

Yes, the matrices are all diagonal. The discussion here is about the general implications of choosing short or long forcing windows in the context of the WC scheme.

- p.12 l.31: "above 750 hPa" = above 750 hPa only or above 750 hPa versus the rest of the atmosphere?

We mean on all model levels above 750 hPa. We have modified the text to make this clearer.

- p.12 l.32: does "globally" mean one term for the whole stratosphere?

We mean on all model levels above the surface (and for each grid box). We have modified the text to make this clearer.

****Section 3 Results**

- p.13 l.15: "The state corrections capture the general horizontal and vertical structure of the a priori bias": this seems a bit optimistic when looking at the figure as a whole. Maybe some king of statistical indicator would make it clearer.

The main feature in the a priori bias associated with turning off convection is excessive CH_4 in lower troposphere and a deficit in the upper troposphere, and the state corrections produce a reduction in CH_4 in the lower troposphere and an increase in the upper troposphere. We have modified the text so that it now states that "the state corrections capture the general structure of the a priori bias, which consists of excessive CH_4 in the lower troposphere and a deficit in the upper troposphere."

- p.13 l.25: "fewer GOSAT retrievals": a supplementary table with the number of GOSAT retrievals per region and period would be useful.

We believe that an additional table with the number of GOSAT data in each region in each month is unnecessary as the impact of the varying GOSAT observational coverage on the assimilation is not a focus of the study. This is an interesting issue, but it would be best addressed by a separate analysis.

- p.13 l.26-27: what about the increase above 100 hPa?

A similar response was found in the OSSE with biased surface emissions, which was removed from the revised manuscript, and was commented on previously by the referee. As we noted in our response to that previous comment, we have not examined the consequence of this in the OSSE. We anticipate that the increase in the stratosphere would be inconsequential since it will have a minimal impact on the total column. It is important to remember that the assimilation is trying to match the observed total column.

- p.17 l.8: do the ACE-FTS data and the model always agree on what points are in the stratosphere?

The model is driven by the NASA GMAO reanalysis fields, which we believe simulates the UTLS well when used at its native resolution. However, when these fields are degraded to 4° x 5° (the resolution used here), errors are introduced in the UTLS, which is one component of the model errors that we are examining in this study and in the companion paper.

****Section 4 Discussion of Model Biases**
p.18 l.14: "the asymmetrically larger number of GOSAT measurements in the northern hemisphere." Same remark as above: a table with the number of GOSAT data available in regions discussed in the text would be useful.

See our response above. However, we note that we are assimilating only data over land, so there is a clear hemispheric asymmetry in the amount of data assimilated in the two hemispheres. We have added a reminder in the text that we are assimilating only data over land.

****Section 5 Conclusions**
- p.21 l.29: "artificially introduced biases in emissions, convection,": the part with biases in emissions is not shown any more.

Thank you for catching that error. We have corrected the text.

- p.22 l.28 and 30: what so significantly and significant fraction mean here? Can you quantify?

We have removed this statement.

Technical corrections:

Throughout the text, "a priori" and "a posteriori" are used: shouldn't it be "prior" and "posterior" instead? Note that I am not a native English speaker.

Both “a priori” and “prior” are used in the literature. Indeed, the ACP English guidelines and house standard states that “Common Latin phrases are not italicized (for example, et al., cf., e.g., a priori, in situ, bremsstrahlung, and eigenvalue).” We prefer to keep these phrases in the text.

- p.6 l.4: "10 ppm" -> ppb?

Corrected.

- p.8 l.2: "surface emissions, however, because of": cut sentence in two "surface emissions. However, because of"

Changed.

- p.10 l.9: "the choice of scaling parameter2 -> the choice of the scaling parameter?

Corrected.

- p.10 l.34: "as described in Sect. 2.3": the reference is a bit strange since we are in Sect. 2.3; maybe use sub-subsections or paragraphs?

You are correct. Referring to the section is unnecessary here. We have removed this reference.

- p.11 l.18: "to produce pseudo GOSAT XCH4 measurements" and l.26 "No noise was added to pseudo-observations": from this I understand that the statistics in R are not used to generate the pseudo-obervations. If so, maybe put all the information about the generation of the pseudo-obs together.

Thank you for the suggestion. We have moved the text.

- p.11 l.20: "from one of the four specified sources of model bias": only three are specified above

Changed to “three”.

- p.12 l.30: "results to vertical extent" -> to the vertical extent

Corrected.

- p.13 l.13: "lofted" -> lifted

Corrected.

- p.16 l.10: "a prior" -> a priori (or prior)

We prefer to keep “a priori”.

- p.16 l.24: "leaves a weak positive biases" -> leaves weak positive biases

Corrected.

- p.16 l.25: "Mean a posterior inter-station bias" -> a posteriori (or posterior)

Corrected.

- p.19 l.29: "forth" -> "fourth"

Corrected.

- p.19 l.31: "vertical transports" -> transport

Corrected.

- p.20 l.22: "the dimensionality of inverse problem" -> of the inverse problem

Corrected.

- p.22 l.10: "slowly varying biases, however, a few stations," -> biases. However, a few

Changed.

- p.22 l.27: "5 o, however, the magnitude" -> 5 o. However, the magnitude

Changed.

- p.23 l.2: "model transport, however shorter-lived" -> transport. However,

We prefer to keep the original text.

- p.23 l.6-7: "incorrectly attributing model errors in vertical transport" to emissions?

Changed.

- Figures 1, 3: "units of ppb" -> "ppb"

Changed.

- Figure 9: "set of experiment" -> experiments

Changed

Report 2 (Referee #2)

The authors thoroughly addressed comments by the reviewers and significantly improved the quality of their manuscript.
The overall message is much more straightforward and consistent with the figures presented.
The shortening of some parts makes it much easier to read and only minor revisions remain before the manuscript is ready for publication.
Please find below the list of revisions to apply:

- Sect. 2.3, p. 7: even though the WC formulation is now well explained, I would add a sentence stating that an equivalent formulation would be to include the correction term u in the target vector p

We have added a sentence on Page 8, after Eq. 5, stating that "This approach provides a means of capturing the model errors in the context of the 4D-Var formalism, whereas other approaches may try to account for these errors by including u in p."

- p.9 l.15: For the convergence criterion, could you please indicate how many iterations are necessary for the different set-ups, for WC and SC?

It takes about 20 iterations for the SC assimilation and 30-35 iterations for the WC. We have added this to the text.

- p. 10 l.13: you use L-BFGS-B which does not nicely accommodate very large numbers of unknowns; why not using algorithms common in the community, such as M1QN3 which is designed for large dimensional problems?

For the work presented here, we simply extended the existing GEOS-Chem 4D-Var scheme, and the 4D-Var optimization in GEOS-Chem uses the L-BFGS-B algorithm.

- Fig. 5: it would be more illustrative to have both the prior and posterior bias at the end of the period to see how much is corrected by the WC state inversion.

It is important to remember that the CH_4 lifetime is long and, consequently, any bias in the initial conditions will decay slowly. This was why we conducted this initial condition experiment. In the previous version of the manuscript we had a figure showing the temporal evolution of the initial condition bias. That figure was removed in the revision, but we still summarize (in the last paragraph of Section 3.1) the results that were shown in that figure. Specifically, we point out that the stratosphere (above 200 hPa) is region with the longest timescale for removal of the bias, noting "that by the third month the CH_4 mass had not fully recovered at these levels." Given the 9-year CH_4 lifetime, it is impossible to remove the initial condition bias in less than three months without assimilation, so we do not believe that comparison of the posterior bias with the bias without assimilation would be more informative.