Reply to Reviewer 1. B. Gaubert

Yi Yin et al.
Feb. 26 2018

Paper title: On biases in atmospheric CO inversions assimilating MOPITT satellite retrievals

First of all, we thank the reviewer for his interest in our work and for his thorough and constructive review. Please see detailed replies below to each point being raised.

General Comments:

This paper aims to evaluate the representation of the CO fields obtained from the assimilation of MOPITTv6 total column (XCO) retrievals (one experiment) and several modelling sensitivity tests. The set of parameters from those sensitivity tests includes the use of posterior emission fields, OH fields, and transport through different grid spacing's, boundary layers and convection schemes.

Further research is needed in some studies using data assimilation to investigate, quantify and in fine understand biases from satellite retrievals, model simulation, and potentially from the assimilation methods (observation operator, error characterization, state vector choices). The comparison of analysis/posterior products can provide a different perspective for the nature of those biases, different from instrument validation and model evaluation by itself. In addition, the use of different and independent datasets is usually informative about the quality of the assimilated and/or posterior model fields. However, the paper's objectives, methods and conclusions need stronger clarifications. In other words, the conclusions can be misleading with regards to the scientific methods used. The authors should either review their conclusions or consider another evaluation approach before publication.

As suggested, and as explained in the following, we have carefully revised our text to avoid confusion. We believe that the revised manuscript better delivers our intended message.

From the conclusions/abstracts, it is stated that the "purpose of top-down estimates of CO emissions, in which the model cannot directly correct vertical model biases, it is more robust to assimilate the column than a particular pressure level retrieval, a partial profile [. . .]".

As it is known, assimilating total columns presents some advantages. For instance, total columns can have lower instrument biases. For pragmatic reasons, it makes sense and I agree that if the observation does not give any information about the

vertical distribution, it is more appropriate to not correct for the vertical distribution within the assimilation scheme. However, those conclusions are surprising with regards to the actual work that has been done in the paper. There are no tests that would rigorously test the hypothesis of choosing a method versus another. In addition, some comments are contradictory.

Our argument is not only based on the measurement features that the reviewer highlights here. In data assimilation / atmospheric inverse modelling, the observation errors are defined with respect to the model, and therefore, combine measurement, model, and representativeness errors. Our paper is focused on model errors and suggests large model flaws in the vertical distribution of tracers that need to be accounted for in the data assimilation, but that are not, or cannot be, well characterized in the form of an error covariance matrix.

One important point to consider is that assimilating total columns relies on the vertical profile given by the model, thus, it is just the total column abundance that is shifted. Within that framework, there is no reason to expect improvement of the vertical profile, because it means that the modelled vertical profile is assumed to be perfect (relatively to the satellite observation). Or it can be assumed that it is far to be the case in global models (with coarse resolution in both vertical and horizontal). It is also acknowledged in the paper: "From the CTM perspective, the evaluation against aircraft measurements reveals significant model errors in representing the vertical CO gradient, in particular over the ocean."

Atmospheric inverse modelling can improve the modelled vertical profile if the profile errors stem from surface flux errors. Indeed, model equations are left unchanged while model simulations are adjusted through their boundary conditions. Our point is that the model equation errors are so large that their statistics have to be accounted for in the inversion (both weak- and strong-constraint systems), even though they are not well characterized.

Moreover, those errors persist after optimizing emissions, which means that chemistry and transport are both important. This is also acknowledged in the paper, but in the abstract, it is mentioned that: "Consistent negative prior biases to all types of observations in all sensitivity tests suggest an underestimation of current surface emissions in the Northern hemisphere. In contrast, prior simulations fit the surface air sample observations well in the Southern hemisphere but underestimate CO in the free troposphere and on average in the column."

For instance, Stein et al. (2014) demonstrated that the northern hemisphere spring cannot be attributed to direct CO emission alone. Myazaki et al. (2015) with a simple sensitivity test, a change in the CO + OH reaction rate were able to considerably reduce this bias. These studies suggest that this bias is likely to occur due to a combination of errors from chemistry, deposition, direct and indirect emission processes (vertical distribution, time profile), as well as transport. Far from the sources (in the free troposphere, over the ocean, in the southern hemisphere), it is even more likely that the problem will be due to transport and

chemistry (secondary sources of CO and/or the OH sink). In particular, biases can be shared through the CH4/CO/OH system, as Strode et al. (2015) and Elshorbany et al. (2016) posit. In general, please consider discussing those previously published results with regards to the setup used in this study, which presents some advantages, but also some drawbacks to compare and contrast different possible approaches.

We thank the reviewer for pointing this out. We over-simply summarized the contrast between the NH and SH results in the abstract. Indeed, both emission estimates and CTM errors have contributions to the model-data mismatch. It has been revised as "In the Northern hemisphere, we find consistent negative biases in the prior simulations when compared to surface, aircraft, and satellite data; in contrast, prior simulations fit the surface observations well in the Southern hemisphere but underestimate CO in the free troposphere and on average in the total column."

We also note that the reviewer reinforces our point about the importance of model errors, that may already appear when assimilating the retrieved total column rather than the retrieved profile. We have added more discussion with respect to previous publications in the text when discussing the bias patterns.

In order to improve the study please consider the following:

1. If this paper is aimed to evaluate the impact of assimilating either the profile or the total columns, this can be done using data assimilation experiments together with observation space diagnostics, see El Amraoui et al. (2014). The use of innovation statistics (and data assimilation) diagnostics allows to quantify the bias, while taking into account error variances from both model and observations. In particular, looking at those by Assimilation of total columns and evaluation of profiles (in observation space) or Assimilation of profiles and evaluation of columns (in observation space).

By doing this the diagnostics for the assimilated and non-assimilated observations can be run for each case to verify the consistency between the columns and the profile in the observation space. In a case where assimilation parameters are correct (with regards to those same diagnostics) and underlying assumptions are not too violated, it would give a more robust estimate of the impact on the total columns while assimilating profiles and vice-versa.

Our paper actually shows one side of the diagnostics suggested by the reviewer (assimilation of total columns and evaluation of profiles in observation space). The other side implies assimilating the profiles and therefore assigning error statistics for the model capability to distribute CO in the vertical. Such a choice of a covariance matrix would drive the inversion results to a large extent but would be mostly arbitrary. The situation is different for the study of El Amraoui et al. (2014) that controlled the 3-D CO field directly rather than the emissions. The study of El Amraoui et al. (2014) also concluded that, "for this kind data (MOPITT V3)"—referring to dataset with relatively low degree of freedom for signal (DFS, ~1.5 for vertical profiles and ~1 for the total column)—

"the present method consisting of deducing the profiles from the total columns remains valid when only using the adjoint of the integration operator".

> *2. Other studies that aim to improve the model error representation in chemical data assimilation (e.g., Gaubert et al. 2016; Emili et al. 2016 and reference therein) as well as potentials from strong constraint 4D-Var (e.g. Trémolet 2006) could be discussed and analyzed.*

Gaubert et al. (2016) highlight for model-error accounting as a future direction from their work and we certainly agree. Weak-constraint approaches as in Trémolet (2006) or Emili et al. (2016) represent interesting possibilities for this, but will only be efficient if there is enough information content in the assimilated data to disentangle emission errors from model equation errors. As of today, the response to this question is unclear and necessitates much research.

> *3. The assimilation of compact phase space retrievals (CPSRs) could be considered, which is an alternative approach to profile assimilation (e.g. Mizzi et al. 2016).*

We understand that this suggestion answers our sentence "measurement error correlations are commonly ignored in inversion systems for technical reasons". Mizzi et al. (2016) presented an interesting approach to assimilate profile retrievals with coarse vertical resolution and is referred to in the revised version. It does not solve the problem of assigning model error statistics.

> *4. Thanks to the comparison to HIPPO measurements, the first identification of an upper tropospheric bias last from the MOPITT V4 (Deeter et al. 2010). The identification of the bias and update of the statistics against HIPPO has continued since then (Deeter et al., 2013; Deeter et al., 2017). You can review Martínez-Alonso et al. (2014) for another evaluation of MOPITT profiles and satellite data. Jiang et al. (2013) suggest not to assimilate the profiles in the upper troposphere while Jiang et al. (2017) propose a latitudinal bias correction. There is evidence that errors can arise from the multispectral retrieval for the nighttime oceanic scene (Worden et al., 2010).*

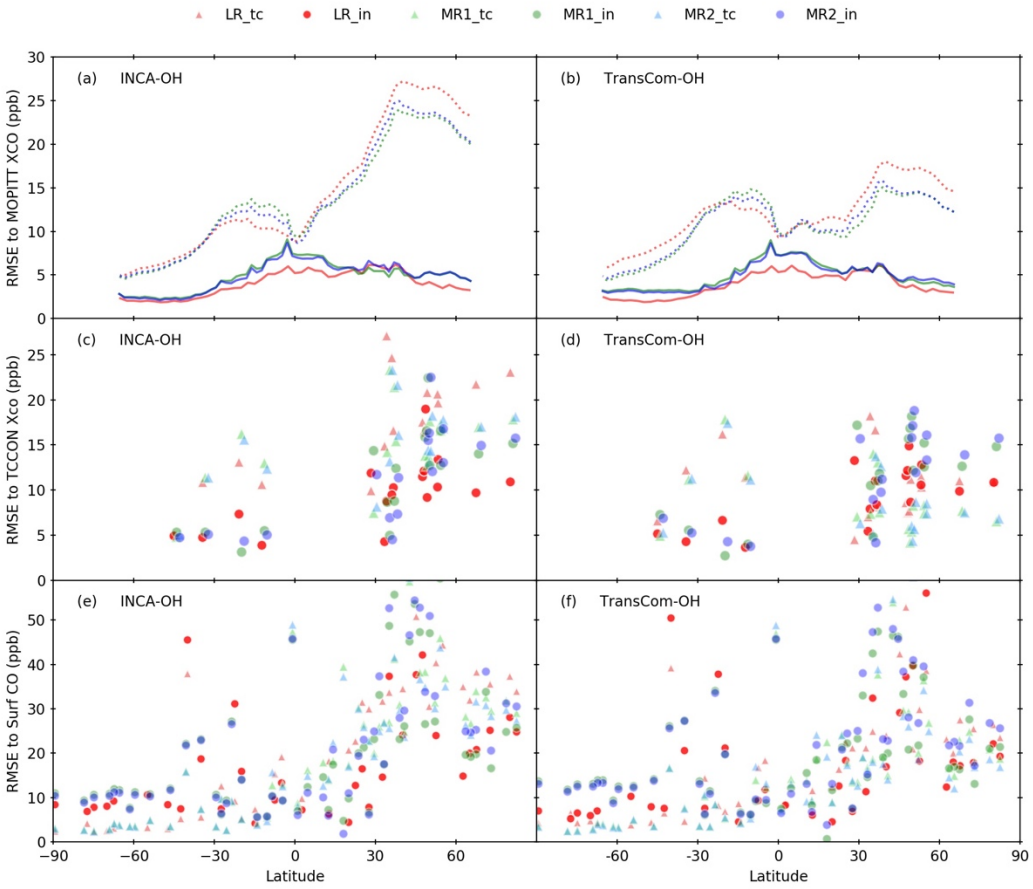
We have included most of the above-mentioned studies and findings in our original manuscript. We thank the reviewer for bringing up the ones that were omitted. We added more discussion regarding the comparison to previous studies in the revision.

> *5. Finally, please consider to include an extensive discussion on the sources of model error and limitation of the setup of the simulations. You can provide more quantitative results. To do so, it is highly recommendable to use conservative measures to compare simulations such as atmospheric burden for CO abundance and RMSE (or absolute errors) for comparison of simulation with observations. On the spatial resolution, one can also discuss the fact that emissions estimates would be different at higher resolution. The change of resolution changes the distribution of emissions and thus chemical regimes. The errors evolution with*

regards to spatial grid spacing can be not linear in chemical transport models. Here, the emissions have been optimized at lower resolution, this leads to serious limitation to conclude about the effect of spatial resolution.

We thank the reviewer for making this suggestion. We extended the discussion much following the advice here and the ones above. Also, following the reviewer's suggestions, we added information on RMSE for comparison of simulation with observations (Fig. S1, please see below); we kept the ones estimating mean biases in the main text, given that the sign of the biases contains useful information.

Figure S1. Annual mean RMSE of the prior and posterior simulations sorted along the latitude. (a) and **(b)** show model RMSE compared to MOPITT X_{CO} using INCA or TransCom OH field respectively. Dashed lines represent the prior simulations, whereas solid lines the posterior simulation; color codes represent model versions as noted in the figure legend. **(c)** and **(d)** show model RMSE compared to independent TCCON X_{CO} measurements. Colors represent model versions with triangles denoting prior simulations and dots denoting posterior simulations. **(e)** and **(f)** show model biases compared to independent ground in-situ measurements from the surface network.



Regarding the setup of spatial resolution, the main point of using the higher resolution is primarily to test its impact on the vertical profiles from 19 levels to 39 levels; the change in the horizontal

resolution is marginal, remaining the same longitudinally and changing from 2.5° to 1.89° latitudinally. We have made this point clearer in the revised text.

> *6. Perhaps, a further discussion on the limitation of the methods is needed, knowing that emission is not the only error source. For instance, there is dry deposition (Stein et al. 2014), time evolution coupling and feedbacks of chemistry (Strode et al., 2015; Gaubert et al. 2016; Elshorbany et al., 2016) and vertical transport, the aggregation of VOC's oxidation in one term (e.g. Jiang et al., 2015 and reference therein), etc.*

We thank the reviewer for making these suggestions. We addressed those points at different parts of the initial manuscript but did not synthesize those into a single section, e.g. we mentioned that "Stein et al., (2014) showed that a higher winter traffic emissions from North America and Europe and a lower dry deposition rate could improve the agreement between simulated and observed CO during winter and spring. (line 652-655)" when discussing possible biases in the prior CO sources in the Northern Hemisphere; we also mentioned that "This could be partly caused by errors in the vertical distribution of tropospheric secondary chemical CO sources, as the surface emissions are generally low but important sources are from chemical oxidation of hydrocarbons and from long-distance transport (Zeng et al., 2015) (line 660-663)" when discussing the Southern Hemisphere model biases. Those points are now better organized in the discussion.

References

Barré, J., et al. (2015), Assessing the impacts of assimilating IASI and MOPITT CO retrievals using CESM-CAM-chem and DART, J. Geophys. Res. Atmos., 120, 10,501–10,529, doi:10.1002/2015JD023467.

Deeter, M. N., et al. (2010), The MOPITT version 4 CO product: Algorithm enhancements, validation, and long-term stability, J. Geophys. Res., 115, D07306, doi:10.1029/2009JD013005.

Deeter, M. N., S. Martínez-Alonso, D. P. Edwards, L. K. Emmons, J. C. Gille, H. M. Worden, J. V. Pittman, B. C. Daube, and S. C. Wofsy (2013), Validation of MOPITT Version 5 thermal-infrared, near-infrared, and multispectral carbon monoxide profile retrievals for 2000–2011, J. Geophys. Res. Atmos., 118, 6710–6725, doi:10.1002/jgrd.50272.

Deeter, M. N., Edwards, D. P., Francis, G. L., Gille, J. C., Martinez-Alonso, S., Worden, H. M., and C. Sweeney (2017), A Climate-scale Satellite Record for Carbon Monoxide: The MOPITT Version 7 Product, Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2017- 71, in review.

El Amraoui, L., Attié, J.-L., Ricaud, P., Lahoz, W. A., Piacentini, A., Peuch, V.-H., Warner, J. X., Abida, R., Barré, J., and R. Zbinden (2014) Tropospheric CO vertical profiles deduced from total columns using data assimilation: methodology and validation, Atmos. Meas. Tech., 7, 3035-3057, doi:10.5194/amt-7-3035-2014.

Elshorbany, Y. F., Duncan, B. N., Strode, S. A., Wang, J. S., and Kouatchou, J.: The description and validation of the computationally Efficient CH4–CO–OH (EC-COHv1.01) chemistry module for 3-D model applications, Geosci. Model Dev., 9, 799-822, doi:10.5194/gmd-9-799-2016, 2016.

Fisher, J. A., Wilson, S. R., Zeng, G., Williams, J. E., Emmons, L. K., Langenfelds, R. L., Krummel, P. B., and Steele, L. P.: Seasonal changes in the tropospheric car- bon monoxide profile over the remote Southern Hemisphere evaluated using multi- model simulations and aircraft observations, Atmos. Chem. Phys., 15, 3217-3239, doi:10.5194/acp-15-3217-2015, 2015.

Jiang, Z., D. B. A. Jones, H. M. Worden, M. N. Deeter, D. K. Henze, J. Worden, K. W. Bowman, C. A. M. Brenninkmeijer, and T. J. Schuck (2013), Impact of model errors in convective transport on CO source estimates inferred from MOPITT CO retrievals, J. Geophys. Res. Atmos., 118, 2073–2083, doi:10.1002/jgrd.50216.

Jiang, Z., Worden, J. R., Worden, H., Deeter, M., Jones, D. B. A., Arellano, A. F., and Henze, D. K. (2017), A 15-year record of CO emissions constrained by MOPITT CO observations, Atmos. Chem. Phys., 17, 4565-4583, doi:10.5194/acp-17-4565-2017.

Mizzi, A. P., ArellanoÂa Jr., A. F., Edwards, D. P., Anderson, J. L., and Pfister, G. G.: Assimilating compact phase space retrievals of atmospheric composition with WRF-Chem/DART: a regional chemical transport/ensemble Kalman filter data assimilation system, Geosci. Model Dev., 9, 965-978, doi:10.5194/gmd-9-965-2016, 2016.

Strode, S. A., B. N. Duncan, E. A. Yegorova, J. Kouatchou, J. R. Ziemke, and A. R. Douglass (2015), Implications of carbon monoxide bias for methane lifetime and atmospheric composition in chemistry climate models, Atmos. Chem. Phys., 15, 11,789–11,805, doi:10.5194/acp-15-11789-2015.

Trémolet, Y.: Accounting for an imperfect model in 4D-Var, Q. J. Roy. Meteor. Soc., 132, 2483–2504, doi:10.1256/qj.05.224, 2006.

Worden, H. M., Deeter, M. N., Edwards, D. P., Gille, J. C., Drummond, J. R., and Nédélec, P. (2010), Observations of near-surface carbon monoxide from space using

MOPITT multispectral retrievals, J. Geophys. Res., 115, D18 314, doi:10.1029/2010JD014242, http://doi.wiley.com/10. 1029/2010JD014242.

Specific Comments:

There are some sentences starting with "It seems" followed by a strong fact statement: "The choice of the prior OH seems to produce the largest differences in the simulated CO concentrations at a global scale." "representing the vertical profile correctly seems to be a grand challenge." This is confusing, please choose between hypothesis or fact.

We have rephrased this expression as "the choice of the prior OH field produces the largest differences in the simulated CO concentration at a global scale across the sensitivity tests we performed in this study", "correctly representing the vertical profile is challenging for either the satellite retrieval or the CTM". We have made similar changes in the rest of the text.

The Section "2.1 MOPITT satellite retrievals of CO total column and vertical profiles" is not clear and contains errors. It needs major revision. Can you merge with section 2.4.4? I don't think the authors should presents the retrieval algorithms. Please provide instead the way they applied the averaging kernels for the columns (equation 1) and the profile (equation 2).

We have revised this section thoroughly. The separation of section 2.4.4 from section 2.1 is made to clarify that only the total columns are used for the inversion (2.1), while the profile is used for evaluation (2.4). We have changed the structure following the reviewer's suggestion and removed the parts resenting the retrieval algorithms. The application of the averaging kernels for the columns and the profile strictly follows those two equations, in which x_{mod} (used here) was modified from x_{true} (used for the MOPITT convention, e.g. see Deeter et al., (2014)) to reflect that the CTM values are used for the corresponding term.

From Yin et al. 2015, it is stated that the MOPITT observations are average on the coarser grid. For a fair comparison, the evaluation should be done at 'higher' resolution, note that high resolution for a simulation at 1.895 by 2.5 degrees is misleading. A comparison of the impact of the vertical grid spacing should be done first, what is the observation error differences between low (2.5◦x3.75◦) and medium (1.895 by 2.5 degrees)? What is the differences in biases?

Indeed, as stated in Yin et al., 2015, all observations located in the same model grid within every 30-minute time steps are averaged as a "super observation" for the assimilation; similarly, the observations are averaged at corresponding model resolutions for the evaluation, not at a fixed coarse resolution. We acknowledge that the expression 'high' resolution is a bit misleading; yet, in the community of global inverse studies, this resolution is relatively high, which was the reason

for the initial choice. To avoid confusion, in the revised text the terms are changed to low resolution (LR) and medium resolution (MR) following the comments.

Concerning the change in the horizontal grid, we do think that it is modest enough (unchanged in longitude, refined from 2.5° to 1.9° in latitude – note that the reviewer's numbers here are incorrect) not to bear significant impact on the results. We have toned down our expression "finer horizontal resolution" to avoid confusion.

Minor Comments

Abstract
Page 1, Line 16: "Carbon monoxide (CO) inverse modelling studies have so far reported significant discrepancies between model concentrations optimized with . . . (MO- PITT) satellite retrievals and surface in-situ measurements."
In my opinion, this first sentence is misleading and is not fair to MOPITT itself. It is common in atmospheric composition that there are discrepancies satellite and surface observations, or between satellite themselves (Kopacz et al. 2010). It usually leads to large improvements in particular for the well-known northern hemisphere spring bias. Which means that the errors are also due to model and coarse resolution. There have been issues in the Southern Hemisphere, but the reasons are not clear. The model errors appear to not be driven by CO emission alone (e.g. Fisher et al., 2015).

The sentence cited here stated the fact that there are discrepancies between MOPITT optimized concentrations and surface observations (as reported by previous studies), but by no means did it imply that this error comes from MOPITT. The reviewer also mentioned here that "*It is common in atmospheric composition that there are discrepancies satellite and surface observations, or between satellite themselves (Kopacz et al. 2010)*". Thus, it stated the motivation of this study to look more closely into the model-data mismatches.

1 Introduction
P2, L76: "with most CTMs showing negative biases to surface and satellite observations in the Northern hemisphere when prescribed with current emission inventories (Naik et al., 2013; Patra et al., 2011; Shindell et al., 2006)." The Patra et al.'s study is about TransCom-CH4 and the paragraph is CTM's biases in Northern Hemisphere spring of CO modelling, please change this reference to Stein et al. (2014).

We referred to Patra et al., 2011 for their study of interhemispheric transport modelling in general, but since we do not elaborate on this further we have removed it.

P2, L85: "MOPITT-based atmospheric inversions were also shown to be biased high when compared to independent in-situ surface observations in the boundary layer (Gaubert et al., 2016; Yin et al., 2015)." Gaubert et al. (2016) showed that assimilating MOPITT improve the CO values at the surface (not the opposite). The cross-validation with FTS in the southern hemisphere suggests that the model has

actually a good prior for wrong reasons, the assimilation improve correlation and suggest an underestimation of biomass burning emissions. Please be more precise, at least indicate that it is in the southern extra-tropical region.

The cited sentence did not mean that assimilating MOPITT degraded its comparison to surface observations, but that the MOPITT reanalysis was biased high when compared to independent surface observations as illustrated in Fig. 5 in Gaubert et al., (2016) (cited below), which is fair in our opinion. It is noted by the authors that "The increase of CO in MOPITT Reanalysis lead to an overestimation with respect to the observed values of around 30 ppb in the SH and less than 10 ppb in the NH". But we agree with the reviewer that more detail regarding these findings could be given in the discussion, which is now added in the revised text.

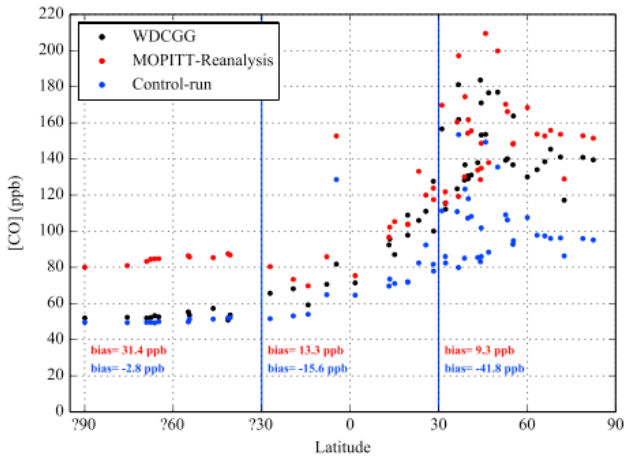


Figure 5. Annual CO average of surface WDCGG observations (black), MOPITT Reanalysis (red), and the Control Run (blue).

P3, L107: "In section 3, we first evaluate the model with MOPITT XCO that is used for assimilation, then [. . .]". Please clarify, only one simulation is done with assimilation of MOPITT XCO, what is the model?

The details regarding the set-ups are given later in section 2. We kept the model description simple in the introduction of the paper structure where different versions are not mentioned yet at this point. Some hints were given in the previous paragraph that "this study evaluates the results of a global MOPITT CO total column assimilation using LMDz-SACS (as described by Yin *et al*. 2015) (line 93-94)", and "We also include a series of sensitivity tests (described in Section 2.2 and 2.3) to discuss model uncertainties (line 101-102)". We have revised the manuscript extensively to better describe our simulations.

P4 L152: change "coverting" to converting

Changed.

P4, L186: "HR2 corresponds to the version called standard physics (SP) in (Locatelli et al., 2015), where more detail regarding these configurations are described." Please rephrase, e.g. "The latter corresponds to a version called

standard physics (SP), presented in Locatelli et al. (2015), where more details about these configurations are described."

Thanks. It is implemented as suggested.

P5, L189: Please rephrase, "as suggested by Radon [. . .]", to "as suggested by the comparison with Radon [. . .]".

This has been corrected.

P5, L191: Please rephrase "For all versions of LMDz-SACS here", For all the different LMDz-SACS configurations presented in this study: [. . .] boundary conditions and horizontal winds [. . .]"

Done.

P5: L214: "One scale factor is applied to the HR results over the globe to conserve the global mass budget to be consistent with the reference version MR." Can you explain and be more explicit? If the 3D fields of Formaldehyde have been simulated using different spatial resolution, why would not you keep a different field? Which budget (CO or HCHO)?

The integrated global burdens of the 3D CH_2O fields produced by the full chemistry model with MR and HR resolutions are slightly different. The former was optimized during the assimilation together with MOPITT X_{CO}. We keep the same monthly global HCHO burden between the two versions to minimize its impact on CO fields to keep the comparison simple. We have revised it as "a scaling factor is applied to the MR results over the globe to keep the monthly CH_2O production in mass consistent between the MR and LR versions".

P6, L269: "It is noted that the MOPITT NIR/TIR retrievals, combining information from both TIR and NIR, have generally higher sensitivity to the lower troposphere compared to TCCON." Please rephrase, add a reference?

We removed this sentence here. Their different vertical sensitivity is discussed later in the result section.

P7 L305: "Individual aircraft profiles are assigned to certain model grid points given their geographic location and pressure levels, and all measurements are then averaged per model grid point at a 30-minute resolution to compare with corresponding model value." What would be the impact of spatial resolution? This comparison is in favor of the coarser resolution, why don't you interpolate to the observations?

Following the reviewer's advice, we tested the interpolation of the model results to the observation's location and time. It does not impact the results.

P7 L327: I do not understand this statement: "the total column of the model state integral is always conserved to compensate uncertainty from vertical resolution

change on the CTM side." Again, the evaluation at lower resolution is in favor of the coarse run.

We rephrased the statement as "When interpolating from the vertical CTM levels to the 10 MOPITT pressure levels to obtain $\boldsymbol{\chi}_{\mathbf{mod}}$, we conserve the pressure-weighted column-mean CO concentration by a scaling factor".

P8 L341: Typo, "of the model version or of the OH field"
P8 L345: change "for both OH" to "for both OH fields".

We corrected both issues.

P8 L350: "It is noted that although we optimize OH together with surface emissions, the system only scales slightly the big-region OH state vectors (in total six big regions over the globe) and thus the inverted surface emissions are sensitive to the prior OH fields." With OH driving the CO sinks (90 %), how would it be possible to not be sensitive to the prior OH field?

Our point was not to overplay the effectiveness of optimizing OH, but we have deleted this sentence.

P8 L350: "with the range showing 1-sigma standard deviation of the mean biases of all model grids, the same applies hereafter if not specified otherwise". Please remind the time period, a large variability is expected for the northern hemisphere spring.

We revised the sentence as "with the uncertainty range showing 1-sigma standard deviation of the monthly mean biases of all model grids over three full years from 2009 to 2010".

P8 L365: "The CO surface sources are conserved in global mass between different resolutions when emitted to the atmosphere, but with the change in resolution, the CO sink via the reaction with OH (associated with different resolutions) may differ." What is the purpose of changing the resolution if you forced to value to be equal to the coarse resolution? Those differences are what can be interested in this study, because you are comparing simulation with different spatial grid spacing.

We keep the same surface and chemical CO sources (CH_2O 3D production) in mass to reduce the features that could introduce differences in the model results. As mentioned earlier, the largest difference between the two resolutions are actually the vertical levels they have, more relevant to the features in the vertical profiles that we are interested in.

P8 L370: "The difference between the HR and MR XCO results (~1.5 ppb) is of a much smaller magnitude than the differences between the prior and the posterior MR simulations (~10.4 ppb); it is also of a smaller magnitude than that induced by the two OH fields in the prior forward simulations when the CO sources are identical (~2.8 ppb for global average with some cancelling effect between the

NH and SH). The differences in modelled XCO between HR1 and HR2 are not significant at a global scale." To remove cancelling effects, could you use RMSE for comparison with measurements, and/or tropospheric abundance (tropospheric mass) for the comparison of different simulations.

We thank the reviewer for this nice suggestion. We added information on RMSE in Fig. S1.

P9 L388: "They also slightly overestimate XCO in SH, resulting in a smaller positive bias 2.5±3.1 ppb using INCA-OH (3.0±2.2 ppb using TransCom-OH)." Please rephrase.

It refers to the comparison to TCCON measurements. It is rephrased as "The posterior simulations also slightly overestimate X_{CO} in SH, producing a small positive bias of 2.5±3.1 ppb when using the INCA-OH field (3.0±2.2 ppb when using the TransCom-OH field)".

P9 L422: "when other setups being the same; here, the 1s standard deviation showing the spread across prior/posterior simulations and between the two HR models." Please do not use the average difference, is it with adjusting emissions and chemistry. You could also do a simulation with full chemistry and updated emissions.

We agree with the reviewer that we should not average the differences between the two cases. We described their differences separately in the revised text. We did not do a full chemistry model as it is not the focus of this study. Here, we simply wanted to show the impact of a change in the vertical resolution on the CO concentration field and whether those uncertainties will change the bias structure.

P10 L430 and Figure 3: What are the different points, is it the latitude band or the time period (Months)? A large seasonal cycle is expected (see https://www.esrl.noaa.gov/gmd/ccgg/globalview/co/co_intro.html).

Each point represents the annual mean quantity for the surface [co] at a given station and for the X_{CO} of the corresponding model grid. The structure of the biases in the seasonal cycle is not discussed in this study, as we wanted to focus on the large scale spatial features for the annual mean values.

P10 L449 and P10 L469: title 3.4.1 MOZAIC measurements over large airports What do you mean by large airports? Did you actually select airports that are larger, in size, in number of flights?

We did not. We revised it as "MOZAIC aircraft measurements".

P10 L455: "are larger in the NH than in the SH, consistent with the different in prior model" Please correct to 'differences' or 'different priors'.

We corrected it.

P11 L511: "Such bias in representing the oceanic vertical profiles suggests error in the vertical distribution of CO source/sink over ocean or in the vertical mixing." Please rephrase, there are errors (plural) from both chemistry, horizontal and vertical transport, ocean is repeated twice.

We revised it as "Such biases in representing the vertical CO profiles over the ocean suggest errors in the vertical distribution of secondary CO sources and its OH sinks, or in the horizontal and vertical mixing".

P13 L600: "The choice of the prior OH seems to produce the largest differences in the simulated CO concentrations at a global scale." It is confusing, please remove 'seems', or you can say "Our study shows that the choice of. . ." Again, do not forgot that having a prescribed OH is a strong approximation (limitation), see for instance (Elshorbany et al.; 2016)

We revised it as "the choice of the prior OH field produces the largest differences in the simulated CO concentration at a global scale across the sensitivity tests we performed in this study."

P13 L615: "The North-to-South gradient in TransCom-OH is also closer to a recent observation-derived near equal N/S OH distribution (Patra et al., 2014)". I think those two studies are related, the TransCom-OH field is from ACTM_0.99, which is designed to have a near equal N/S OH distribution. You can mention this in the introduction, it is supposed to be the best fit to Methyl Chloroform.

We added this information in the method section when introducing the TransCom-OH field. "The TransCom-OH field was developed by combining the semi-empirically calculated tropospheric (following Spivakovsky et al., 2000) and 2-dimensional (2-D) model simulated stratospheric distributions. It is supposed to be the best fit to Methyl Chloroform reduction rate."

Acknowledgement Most of the acknowledgement are not respecting the recommendation, please consider contacting the instrument PI's. For instance, the MOZAIC acknowledgement (if not updated since) should be as follow:
The authors acknowledge the strong support of the European Commission, Airbus, and the airlines (Lufthansa, Air France, Austrian, Air Namibia, Cathay Pacific, Iberia and China Airlines so far) which have carried the MOZAIC or IAGOS equipment and un- dertaken maintenance since 1994. In its last 10 years of operation, MOZAIC has been funded by INSU-CNRS (France), Météo-France, Université Paul Sabatier (Toulouse, France) and Research Center Jülich (FZJ, Jülich, Germany). IAGOS has been additionally funded by the EU projects IAGOS-DS and IAGOS-ERI. The MOZAIC–IAGOS database is supported by AERIS (CNES and INSU-CNRS).

We thank the reviewer for this note. We have revised the acknowledgement section carefully.

Table 1: "List of simulations in this study", -> 'list of simulations done in this study' or 'list of simulations'.

Table 2: "reference" to "references"

Corrected.

Table 3: How are calculated the error bars (more or less)?

We added this information into the table caption. "The uncertainty range represents 1-σ of the annual mean model-data biases across all model grids for Xco or among all stations for surface [co]."