Reply to Reviewer 2

Yi Yin et al.

Feb. 26 2018

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

First, we would like to thank the reviewer for his/her comments that will greatly help clarify our manuscript in the revised version. All comments are accounted for and are individually discussed hereafter.

General Comments:

The submitted paper aims to evaluate the quality of CO emissions inversions and inference of other parameters influencing the CO budget using different model configurations, i.e. horizontal and vertical resolution, OH forcing and physics scheme. While the authors try to provide a detailed evaluation, and provide indications of possible biases on MOPITT retrievals and additionally of model and emissions, the methodology and scientific argumentation employed is not sound to me. It is quite unclear what are the exact goals of the paper and it seems that the focus is too broad. The evaluation is extensive which is appreciable but the discussions try to cover too many topics without robustness and convincing arguments.

We agree with the reviewer that our discussion was a bit scattered, hence we have restructured the discussion to make our argument clearer in the revised text.

This study needs to mature and needs to be supported by additional experiments. I fear I cannot recommend this paper for publication for the following main reasons:

1. *The discussion and especially sections 5.3 and 5.4 are scientifically flawed. The authors drive conclusions without experimenting themselves. I strongly recommend the authors to reconsider their data assimilation experiments and setups before driving such conclusions or consider removing those two sections. I fear that without those two sections the paper will significantly loose substance. Moreover, the sensitivity tests on model parameters are not convincing, a significant increase on model horizontal resolution and using a more detailed chemical scheme would have been more useful to point out intrinsic model deficiencies and uncertainties.*

We respectfully disagree with the reviewer's harsh wording, but we agree that our articulation may not have delivered our message properly and caused some confusion. Therefore, we would like to take the chance to explain our main points, which we believe can be distilled from the experiments performed in this study and are helpful to setting the priories of future global CO inverse studies.

Section 5.3 "**Possible biases in the MOPITT retrievals**" are based on the comparison of posterior model states to other independent observations, including surface measurements, TCCON, and aircraft measurements (HIPPO/MOZAIC). Our study did not aim at evaluating the MOPITT retrievals, yet the fact that the same model state has different signs of biases compared to MOPITT and to other independent measurements reveals some inconsistency.

Section 5.4 "**Assimilating satellite total column vs. vertical profile retrievals**" are addressed "for the purpose of top-down estimates of CO emissions, in which the model cannot directly correct vertical model biases (line 769-770)" based on the vertical bias structures we showed in this study and practices of previous studies in this context:

First, we showed that (1) when updating the surface emissions the overall shape of vertical profiles can only be marginally changed (only if the profile errors stem from surface flux errors), and (2) the posterior model biases to the MOPITT profiles vary along the altitude (with opposite signs of the remaining biases between the near surface levels and the free-troposphere / stratosphere levels). Thinking logically, one would expect that when assimilating one pressure level at a time the inversion would derive different estimates of surface emissions. Indeed, Jiang et al., (2013) has demonstrated that assimilating the MOPITT surface level, the profiles, or the column amounts individually would result in different global CO emission estimates: 125.3, 150.1, and 141.9 Tg/month respectively for the period June-August 2006; this point has been further demonstrated in the follow-up studies in Jiang et al.. (2015, 2017). In such a case where model profile can only be marginally modified via the change of surface emissions, assimilating the full profiles would not bring an obvious advantage compared to assimilating the total column; instead, it poses the issue of addressing observation error correlations among different vertical levels that are difficult to characterize (note that the observation errors here are defined with respect to the inverse model, and thus, include measurement, model, and representativeness errors) and are often ignored.

Second, regarding the MOPITT retrievals, there are known temporal bias drifts in the profile retrievals, but the total column shows the strongest temporal stability as documented by Deeter et al., (2014). We noted in our paper that "the bias drift for the MOPITT V6 TIR/NIR product varies from negative in the lower troposphere (-1.3%yr^{-1} at 800 hPa) to positive in the upper troposphere (1.6%yr^{-1} at 200 hPa) as compared to aircraft measurements; yet, the bias drift in the total column is negligible (Deeter *et al*., 2014) (line 741-745)." In this case, assimilating partial profiles would bear the temporal bias drift of the partially integrated columns. In line with this expectation, Jiang et al., (2017) showed, indeed, different long-term trends in the posterior CO emissions when assimilating the MOPITT total column, the profiles, and the lower profiles (defined as the surface, 900, and 800 hPa pressure levels) (please see Figure 6 of Jiang et al., 2017). Also, correcting the latitudinal bias for each pressure level individually may degrade the overall column consistency, which could explain partly the different results between assimilating the profile and the column, as ideally, they would result in similar budgets and trends.

Regarding reviewer's remarks on our sensitivity test, we tested model configurations that are related to the modelling of the vertical profiles: a doubling of the vertical model resolution, a different convection scheme, and a different OH field. They are intended to show weather model uncertainties associated with these aspects would impact the evaluation results particularly regarding the vertical bias structure, which is well within the scope of this study and helpful for the discussion and conclusion.

> 2. *The quality of the scientific argumentation can be questioned. A lot of references are cited inappropriately. Number of citations do not support statements made in the present paper (see specific comments below). Demonstrations are often approximate and hand-waving. The conditional form is often used when it comes to conclusions (the forms "would" and "could" are widely used). The authors suggest and anticipate from incomplete set of experiments with few references to drive scientific conclusions.*

We thank the reviewer for pointing out some inaccuracy or confusion in our expressions. Since they are mentioned later in this review, we answer them individually below.

> 3. *Last but not least, I am concerned about the methodology itself; statistical method- ology and the significance of the diagnostics.*

We answer this point below where more detail from the reviewer is given.

> *The reliability of the data assimilation algorithm is not discussed as well.*

This paper evaluates the prior and posterior model states with various observations including those being assimilated, which is a measure of its "reliability". The data assimilation algorithm has a 13-year history for atmospheric inversion of CO_2, CO, CH_4, HCHO, refrigerant gases and aerosols at the global scale with 10s of papers documenting its behavior or using its results. We wrote in the introduction that "this study evaluates the results of a global MOPITT CO total column assimilation using LMDz-SACS (as described by Yin *et al*. 2015) against various independent measurements for the period 2009-2011. (line 93-95)". We, therefore, assumed that we could save some space here by referring to Yin et al. (2015) and Pison et al. (2009), two papers that contain further detail and references. We have added more information in the revision.

> *To support the three mentioned points please consider the following specific comments.*

Specific Comments:

Line 66: There are also other references that are using MOPITT and data assimilation to study the temporal distribution and variability of CO, e.g. Inness et al., 2015, Myazaki et al., 2015, Barré et al., 2015.

The sentence is not about MOPITT data assimilation studies in general but about "studying the spatial-temporal distribution and variability of CO sources and sinks at the global scale" with MOPITT. Of the three references suggested here, only the second one fits the topic.

We revised the text to be more inclusive of MOPITT related studies: "The official MOPITT CO product, now reaching version 7, has played a pivotal role in the study of the spatial-temporal distribution and variability of CO sources and sinks (Shindell et al., 2006; Stein et al., 2014; Worden et al., 2013), including applications in global inverse studies (Arellano et al., 2004; Fortems-Cheiney et al., 2011, 2012; Hooghiemstra et al., 2012; Jiang et al., 2017; Kopacz et al., 2010; Yin et al., 2015) and chemical reanalyses (Barré et al., 2015; Gaubert et al., 2016; Inness et al., 2015; Miyazaki et al., 2015)."

Line 75: Please, change plagued by another word. Models are not plagued, they just misrepresent the truth by man-made simplifications.

We changed it to "associated with".

Line 83-84: This is not what Hooghiemstra et al., 2012 are proving. Form the conclusion of the paper it is: "However, in the remote SH (30 – 60Âa˘S), the comparison with MOPITT deteriorates from a 4% negative bias in the a priori to a 10% negative bias in the a posteriori solution, due to an emission decrease suggested by SH surface observations."

We wrote "Assimilating surface in-situ measurements or MOPITT CO retrievals using the same CTM and inversion configuration could result in considerable differences in the posterior surface fluxes (Hooghiemstra *et al.*, 2012)", which, in our opinion, is a fair rephrase of the findings regarding the posterior fluxes of the Hoogiemstra et al., (2012) entitled "Comparing optimized CO emission estimates using MOPITT or NOAA surface network observations". As presented in Figure 9 and Table 1, their posterior global annual emissions assimilating station-only vs. MOPITT-only are 724±75 vs. 806±69 for anthropogenic sources, 704±78 vs. 733±60 for Natural + NMVOC, and 394±60 vs. 304±28 for biomass burning.

Line 86: Gaubert et al., 2016 is not inverting surface emissions as Yin et al., 2015.

We will remove this reference from the sentence.

Lines 84-87: This statement is not well supported by either reference provided. For example, Barré et al., 2015 that is assimilating two types of sounders find opposite conclusions. MOPITT assimilation still underestimates CO at the surface over CONUS. This is probably only true in the southern hemisphere.

Our statement is supported by Yin et al., 2015 (First four columns of Table 2, cited below), in which the mean posterior model biases compared to the surface observations are positive.

Regions	MOPITT Prior bias (ppb)	MOPITT Post bias (ppb)	Surface Prior bias (ppb)	Surface Post bias (ppb)
BONA	−14.1	0.7	−20.8	20.5
USA	−16.7	−3.1	−20.4	20.1
NHSA	−10.0	−3.0		
SHSA	−14.2	0.2		
NHAF	−15.0	−2.0	−20.0	6.8
SHAF	−16.5	0.5	−1.5	13.6
WSEU	−16.1	0.3	−36.7	18.7
ESEU	−16.8	0.4		
BOAS	−17.3	1.0		
MIDE	−16.5	−2.9		
SCAS	−12.0	−0.3		
SEAS	−20.1	−3.6	−30.6	22.6
AUST	−15.5	−1.7	−6.4	15.4
INDO	−3.8	0.2		
OCEAN	−11.9	−0.2	−15.1	9.6

It is also supported by Hooghiemastra et al., (2012) (Fig. 7 copied below). The posterior (MOPITT) concentrations (in red) are in many cases higher than the ground measurements (in black), not limited to the Southern Hemisphere.

D06309 HOOGHIEMSTRA ET AL.: CO INVERSIONS USING MOPITT OR NOAA DATA D06309

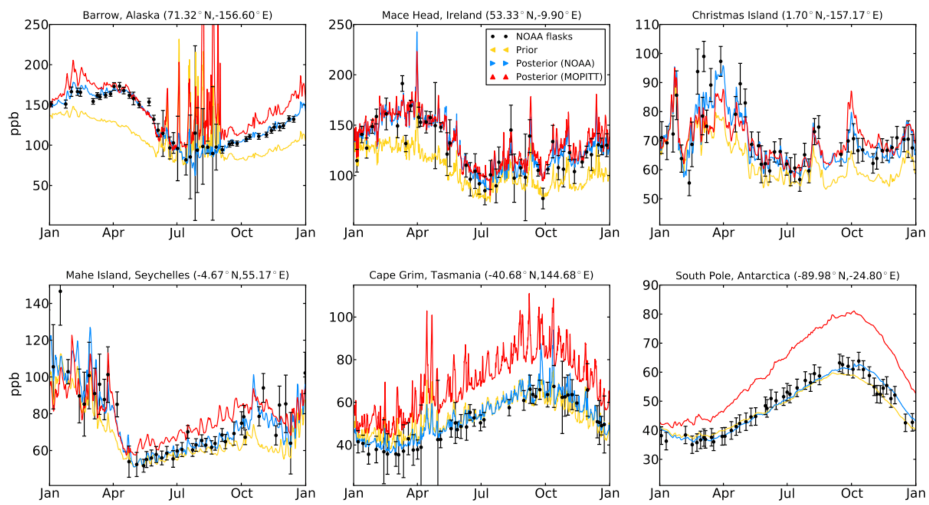


Figure 7. Prior and posterior simulation for 2004 sampled at 6 NOAA stations that were assimilated in the 4D-Var inversion stations-only. The simulation in red used optimized emissions from the MOPITT-only inversion. Black dots represent the NOAA flask observations. Error bars denote the total observation error (including the model representativeness error). The comparison with additional stations is presented in the auxiliary material.

Consistent with the previous point from the reviewer, we are considering here inverse modelling studies, which do not include Barré et al. (2015).

While it is true that Barré et al. (2015) showed "*MOPITT assimilation still underestimates CO at the surface over CONUS* (CONUS means the Continental United States using air quality data from EPA)", the authors also noted that "The suburban monitoring sites could be often located near strong localized pollution sources creating a strong source-to-receptor relationship that is a challenge to duplicate in coarse model resolution [Pfister et al., 2011].We hence **expect the models and data**

assimilation experiments to be significantly negatively biased, due to the difference of representativeness between the surface sites (local measurements) and the model having a 2×2° representative grid box."

Lines 87-90: This statement is not clear at all. Please clarify.

We rephrased the sentence as "Jiang et al., (2013, 2015) further demonstrated that assimilating the MOPITT surface level (or the three near surface levels), the profiles, and the total column amounts would result in different posterior CO emission estimates, suggesting some inconsistency in the information provided to the inversion system on the vertical CO distribution between the CTM and the MOPITT data".

Line 88: differences between what and what?

We mean different posterior emission budgets when assimilating the MOPITT CO profiles or when assimilating the MOPITT surface layer retrieval.

Line 90: While the statement is unclear to me, I do not think this is the Jiang et al., 2015 conclusions.

Jiang et al. (2015) showed that assimilating the MOPITT CO profiles or the near surface layer retrievals results in different surface emissions. We interpret this result in terms of inconsistent information provided to the Bayesian cost function from the surface layers and from the profile data given the same control variable. We have rephrased our expressions to make it clearer.

Line 135: The authors should know what Bayesian means. There is nothing Bayesian in this equation.

We respectfully disagree: the formulation actually comes from the linear Bayesian estimation framework (see, e.g., Rodgers and Connor, 2003, doi:10.1029/2002JD002299). Still, we will rephrase the sentence.

Line 140: Clarify the statement, it sounds as you model a profile from measurements.

The paragraph has been revised, further to this and M. Deeter's comments.

Line 176: 2.5 by 3.75 degrees is now considered as low resolution, change accordingly.

We have changed it to low resolution accordingly.

Line 179: is it another model? I believe you still use LMDz but with a different configuration. Change accordingly.

We changed the sentence as "We also include a higher resolution configuration of LMDZ-SACS with a higher vertical resolution of 39 eta-pressure levels for sensitivity tests, noted as Medium Resolution (MR)."

> *Line 181: Does changing just the latitudinal resolution from 2.5 to 1.89 degrees relevant? It is then mainly a significant increase on the vertical resolution. Why not keeping the same horizontal resolution? Again 1.89 by 3.75 by 39 levels is not considered nowadays as high resolution.*

We changed the notations to medium resolution (MR) and low resolution (LR) respectively. Indeed, the major features we are interested in is the impact of the change in the vertical resolution.

> *Lines 199-200: 2009 to 2011? From what month to what month? It could be almost three years to almost one year though.*

For three full years from Jan. 2009 to Dec. 2011.

> *Lines 316-317: The authors should detail exactly how they apply the observation operator to retrieve Xmod. Have they smoothed the model profile by the averaging kernel, have they considered interpolating partial columns from the model and then convert to log(vmr) to match the MOPITT data? The authors should refer to Barré et al., 2015 section 2.2.4 for the correct approach. I am then uncertain if the method used by the author is the correct one, hence I am doubtful about the validity of the results and discussion about the MOPITT profiles validation in rest of the paper.*

-*"Have they smoothed the model profile by the averaging kernel?"*
Yes. It was noted in our manuscript that "The MOPITT-equivalent model CO vertical profiles are retrieved in a consistent manner as the satellite retrievals using Equation (2) described in section 2.1, in which the $\boldsymbol{\chi_{mod}}$ become the CTM concentrations interpolated at the MOPITT pressure levels to be convolved with corresponding MOPITT prior CO profiles and averaging kernels. (line 315-318)".

-*"have they considered interpolating partial columns from the model and then convert to log(vmr) to match the MOPITT data? "*
Yes. It was noted in our manuscript that "The vector quantities $\boldsymbol{\chi}$ are expressed as log10 of the corresponding volume mixing ratio (line 143-144)".

-*"The authors should refer to Barré et al., 2015 section 2.2.4 for the correct approach. "*
We strictly followed the conversion as documented in the MOPITT retrieval papers and referenced to the sample code provided in User's Guide V4 (Section 7.3 for the profile and Section 7.4 for the total column, page 18-20). We also updated the pressure grids following User's Guide V5, which applies to the V6 products as well. As this information is well documented with the data products and should be followed by default, we do not think it necessary to duplicate such information in the paper, which is not specific to our practice anyhow.

We combined this section (2.4.4) with the section about MOPITT retrieval (2.1) and described our approach as "When comparing column or profile retrievals with the model, we follow the MOPITT Version 4 User Guide (Deeter, 2009) to calculate the model-equivalent to the MOPITT retrievals, expressed in terms of logarithm of the volume mixing ratio, with the proper degree of smoothing and a priori dependence (We note that the association between retrieval levels and atmospheric layers are updated since V5) ".

> *Lines 323-324: Does this mean that you are taking the nearest grid point. If yes, is that appropriate? Since you are doing DA science you should be able to interpolate at the right location.*

Yes, for the surface observations and the aircraft measurements, we take the nearest grid point and time step where/when the observation was made because we consider that the model values are volume averages rather than point values. The link made by the reviewer between the alternative choice and the fact that we are doing DA science remains obscure to us.

> *Lines 325-328: This is unclear to me, what operation the authors are doing here. Are you shifting or scaling the profile in order to keep the same total column value? What is the "uncertainty from vertical resolution change on the CTM"? Please rewrite, develop, explain better.*

Vertical interpolation may change the mass of the profile per unit area, for instance because the surface pressure is not the same in the initial and final vertical grids. We therefore scale the profile to conserve the pressure-weighted column-mean CO concentration. Note that we exclude observations whose surface pressure differs more than 50hPa from the model state.

> *Lines 330-333: It is unclear to me what the authors are doing exactly. Are they averaging monthly model values and then they are comparing with monthly averaged observations? If yes, the entire results of this paper would be flawed. Or are they interpolating model to observation at the right time.*

The models are sampled at the time of the observation with a minimum sampling time step of 30 minutes, as described in the text "Model values are sampled according to the time and location of the measurements at a temporal resolution of 30 minutes and the model associated spatial resolution (line 321-323)". Then, they are averaged per month for each model grid for further analysis, we revised this sentence as "the observed and simulated values are averaged respectively per month for each station or model grid for further analysis (line 330)".

> *Moreover, it seems that the correlations in the rest of the paper are made on monthly averaged biases, reducing the sample for correlation to something small and probably not statistically significant. Looking at the correlations plots I see around 12 to 14 point as a sample size. Would it be statistically more sound to calculate those correlations using the entire sample of observations (not reduced by average biases)? I am then doubtful about the robustness of this score during the further analysis of this paper.*

The analysis was based on monthly averaged values. This choice was made to assess the agreement between model and observation for general spatial patterns other than fine scale variations. Averaging observations and corresponding model outputs by month also helps to weigh the observations evenly, so that the results are less impacted by uneven numbers of available observations in space and time.

If the correlation plot here refers to Fig. 3, each point represents the mean annual bias to the near surface concentration at a certain station (x-axis) and to the corresponding MOPITT X_{CO} that fall within the same model grid (y-axis). Again, this plot is intended to show the spatial covariations of the model biases with respect to the surface observations and to the MOPITT X_{CO}.

Line 358: Please recall what are those big-regions. Cite Yin et al., 2015.

We added this information in the revised manuscript in the method section. "the column-mean OH concentrations are also optimized by scaling factors over six big boxes of the atmosphere: three latitudinal boxes (30–90°S, 0-30°S, 0–30°N) and three longitudinal boxes north of 30°N (North America: 180–45°W; Europe: 45°W–60°E; Asia: 60–180°E)".

Line 370-376: This entire paragraph is confusing to me, please rewrite.

We rephrased it as "The average difference of the simulated X_{CO} between the posterior and the prior simulation using the reference model (LR) is ~10.4 ppb at the global scale, a much larger value than that induced by different model configurations given the same CO sources and sinks (~1.5 ppb between MR and LR) or by the two OH fields (~2.8 ppb between TransCom and INCA at the global scale, with some cancelling effect between the NH and the SH). This comparison suggests that the increment of surface emissions in the posterior estimates compared to the prior is robust with respect to uncertainties in OH distribution, vertical model resolution, and convection / boundary layer mixing scheme that we tested here."

Line 427: Higher sensitivity of what to what?

We rewrite this sentence as "Larger differences of model results between MR1 and MR2 are found in the surface level [co] concentrations than the differences in the simulated X_{CO}, suggesting that the convection scheme has a larger impact on the boundary layer mixing ratio than the convolved total column amount".

Line 432-433: Please see my comment above about the significance and robustness of those correlations. The statistical methodology as it is presented now is not sound to me.

As stated above, the correlations we presented here are spatial correlations between the annual mean model-data bias in the surface [co] at ground stations and in the total column [X_{CO}] at corresponding model grid. Only significant correlations with a confidence level higher than 95%

are shown (e.g. we noted "No significant correlations are found in the posterior simulations over the NH (**Figure 3b**), suggesting no consistent systematic errors of the MOPITT X_{CO} assimilation compared to surface [CO] measurements").

> *Line 443-445: This is again confusing with a "hand-waving" argument displaying only the HR correlations, using the word "likely" and not further investigating the possible error on the vertical error CO profile on the posterior. Additionally, I would not trust a correlation of 0.49 with a sample size of 14 using monthly averaged biases.*

The sample size is the number of available stations. This correlation is statistically significant as stated above. The choice was made to show spatial covariations.

> *Line 469: The word contamination is not appropriately used. For example, there is contamination in data when an instrument is not working correctly and generate a systematic error. Please replace this word.*

We have revised the sentence as "This could be partly explained by strong local emissions in the boundary layer that are measured by MOZAIC observations near airports, but are not represented by our global model".

> *Lines 474-479, lines 506-513, figures 4 and 5: MOZAIC profiles are most likely close to the sources whereas HIPPO measures remote scenes. The bias observed in the posterior for HIPPO profiles are due to an overly long CO lifetime in the simplified chemistry model. The presented data assimilation system infers the surface CO sources but do not directly corrects for CO lifetime error due to an (over) simplified chemical scheme.*

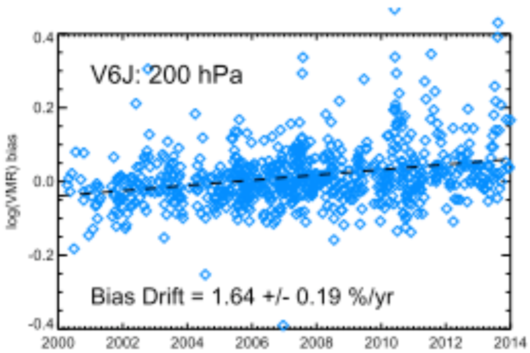
In our system, the average CO lifetime is constrained by the MCF loss rate. However, the reviewer's explanation does not contradict our text "Such bias in representing the oceanic vertical profiles suggests error in the vertical distribution of CO source/sink over ocean or in the vertical mixing. "

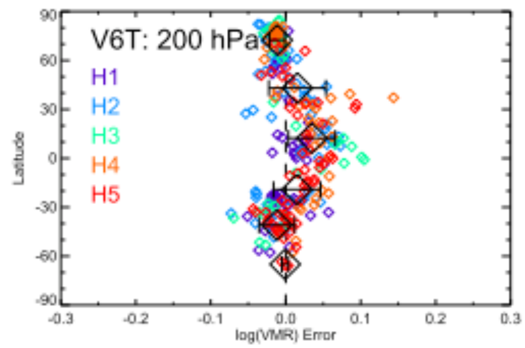
> *Lines 531-532: Please rephrase. "The MOPITT profiles are well reproduced by the model. . .". The model does not reproduce MOPITT profiles.*

The entire sentence is "The MOPITT vertical profiles are well reproduced by the model north of 30°N, but considerable deviations are found for the vertical structure over the tropics and subtropics", which we think is a fair description of Fig. 6d for the posterior model results (dash-dot blue line).

> *Lines 541-544: In Deeter et al., 2014, the MOPITT V6 validation with HIPPO do not see such errors in the upper troposphere. Also, the author should also take into account the spatial sampling of MOZAIC, HIPPO versus MOPITT. The longitudinal distribution of CO, in the tropics can be highly variable.*

Actually, in Deeter et al., (2014), for the years we study here 2009-2011, there are some significant positive biases when compared to NOAA profiles, as shown below on the left, cited from **Fig. 9** of Deeter et al., (2014). The authors also showed the latitudinal distribution of biases with respect to HIPPO for the TIR-only retrieval at 200 hPa, where you can see positive biases in the tropics (as shown below on the right, cited from **Fig. 10** of Deeter et al., (2014)).





It is documented in the text that "V5 TIR-only validation results exhibited a strong latitude dependence in the retrieval bias at 200hPa, with biases of about 25% in the tropics. In contrast, the maximum retrieval bias in log(VMR) at 200 hPa in the V6 results is less than 0.04, which is equivalent to about 10%. **V6 results for other retrieval levels are similar to V5 results, with log(VMR) biases at all levels and all latitudes generally less than 0.05.**" Our analysis (based on V6J) are thus in line with the results of Deeter et al., (2014), but presented in different quantities using a different manner.

Line 565: Which tropical ocean? Rephrase.

We did not look into longitudinal variations. The analysis is for the 30S-30N ocean.

Line 574: "over the ocean"

Corrected

Line 607: What are those big regions? Recall or cite Yin et al., 2015.

We added this information in the methods. Please see reply to Line 358.

Lines 638-643: What is the purpose of this paragraph? It is not clear what the authors are trying to demonstrate. Please clarify, develop, rephrase.

The purpose of this paragraph is to show that "The differences in surface [co] concentration and in the total column X_{CO} at the corresponding model grid are not correlated when using different model configurations (MR vs. LR or MR2 vs. MR1), which suggests that there is not a systematic impact on both the surface [CO] and the X_{CO} comparing one model configuration to another. In other words, it shows that if one model produces a higher X_{CO}, it does not necessarily produce a higher surface [CO] at the same time. As MR consistently produces a slightly higher X_{CO} compared to LR, we could deduce that assimilating the same MOPITT X_{CO} observation using MR would derive a smaller CO emission estimates compared to current LR inverse results.“

Lines 649-652: The syntax of this sentence is not correct.

We rephrased the sentence as “The negative prior simulation biases in the surface [co] and in the Xco are also highly correlated in space (Fig. 3a)”.

Lines 683-695 and section 5.3 in general: The conclusion of "positive biases in the MOPITT retrievals" is flawed here. The authors utilize only one inversion technique from only using total column product. They infer only the surface emissions that is not a direct measured quantity from MOPITT CO retrievals. Depending of a model quality (i.e. resolution, chemistry, horizontal and vertical transport, and so on. . .) inverting the emissions only can lead to good result for the wrong reasons and conversely often having the "correct" emissions and having significant errors in the atmosphere. Data assimilation rely on observation but ALSO on models, you could have the best observation quality, if the model is inaccurate the analysis and the subsequent forecasts would be degraded. Before jumping quick in such important conclusions several things should be tested carefully such as:
Assimilating the CO fields directly with total CO columns and CO profiles
Rerunning the current experiments with a more complete and detailed chemistry

Our remark of “positive biases in the MOPITT retrievals for the total column and at the near surface levels” are made specifically for regions “in the mid- and high latitudes”, where the posterior biases are consistent comparing to both the surface and TCCON observations. The reviewer’s argument does not explain the residual bias to TCCON. Further, a similar bias structure is documented by Jiang et al. (2017). We have added more discussion regarding previous studies documenting MOPITT bias information (e.g. Deeter et al., 2014, Gauber et al., 2016) in the revised manuscript to enrich the discussion.

Line 692-695: Deeter et al., 2014 made the direct comparisons between MOPITT V6T (which is the same as MOPITT V6J over the ocean) and HIPPO measurements: providing a quantification of the MOPITT biases: 1.5% 7.7% at the 200hPa level. How can the authors can explain such discrepancies with those results and figure 4 and 7. The authors compare figure 2 and 4 with figure 7. Again, the representativity of the statistics made here should be considered. HIPPO and MOZAIC cover specific regions whereas MOPITT provide a global picture. Is it reasonable to compare those figures in order to drive conclusions about biases without quantification?

Figure 6 of Deeter et al. 2014 shows +1.5% bias (MOPITT too high) and 7.7% std at 200 hPa for MOPITT vs. HIPPO misfits in terms of log(VMR). This is not inconsistent with what we show in Fig.

7, with zonally-averaged model-MOPITT misfits expressed in terms of VMR about -20 ppb around 200 hPa.

Lines 702-707: This indicate an issue in your CO lifetime (see comments above). I would suggest having an estimate and quantification of your CO lifetime and budgets (e.g. like in Gaubert et al., 2016). This will help you investigating and quantifying what is responsible for the biases in the posterior: MOPITT retrievals, LMDz or the 4DVar.

The CO budgets were analyzed in Yin et al., (2015) in Fig. 8. We added discussion regarding this point in the revised text. Note that the posterior model states agree well with the MOPITT total column as shown in Fig. 2 for the average model-data bias along the latitude, Fig. 8 for the regional statistics, and Table 3 for the zonal statistics.

Lines 709-714: This statment now refers to MOPITT V5T, the rest of the paper is dealing with MOPITT V6J. This is confusing and probably not relevant.

We also described in the text that "The differences between these two instruments (ranging from 0-13%) are significantly larger than that between MOPITT v5 and v6 [Deeter et al., 2014]."

Lines 713-714: The reference to the George et al., 2015 paper is misleading. It makes think the reader that is it a paper about MOPITT biases regarding IASI as a reference. This is not the goal and conclusions of George et al., 2015. Please remove, or rephrase. For a data assimilation comparison between MOPITT and IASI CO profiles, please refer to Barré et al., 2015.

We cited information that is relevant to our discussion from the study of George et al., 2015, which focused on the comparison between MOPITT and IASI CO column, including the influence of their associate prior profile. We changed the word "biases" in line 712 to "differences". We also added references to Barre et al., 2015.

Lines 721-739 and section 5.4 in general: This paragraph is not clear and to my mind drives conclusion without the necessary convincing experiments. The authors "anticipate" that assimilating CO profiles would produce larger biases. Why the authors did not assimilate the profile and then not just "anticipate" but prove this conclusion. I recommend either removing section 5.4 or provide the necessary experiments to support such conclusions.
The authors only support their conclusion by citing papers not accurately that are not using the same model and data assimilation system and experiments. For example, Gaubert al., 2016 do not infer the CO surface emissions.

We showed that 1) updating the surface CO emission would not change significantly the shape of the vertical profile, 2) after fitting the total column there are still negative simulation biases compared to the near surface level MOPITT retrievals. Therefore, it is reasonable to deduce (anticipate) that assimilating those profile retrievals of near surface levels (from surface up to 700 hPa) would derive a larger surface emissions and thus even larger positive biases compared to ground and TCCON observations.

> *Lines 749-754: The authors point out a well-known problem in chemical data assimilation. This can be overcome by using eigenvalue or more generally singular value decomposition to diagonalize R and avoid calculating off-diagonal terms of B (e.g. Migliorini et al., 2008) in a variational framework. Alternatively, approximation and ad-hoc assumption can be made on off diagonal values of B or assuming that R is diagonal and tuning diagonal values of R.*

Our point is not about the technicality of assimilating profiles (apart from the last line where we briefly acknowledge the fact that most inversion systems do not handle full error covariance matrices). Rather, our point is about the fact that we have hardly any basis to guide the assignment of the error covariance matrix from the model equations.

> *Lines 755-760: That is a shame that at the very end of the paper (and few other lines i.e. around line 665) is it stated that the sources and sinks of CO on the model could be responsible of the biases in the posterior analyses. I recommend that this should be reinforced in the entire discussion by having further diagnostics and experiments.*

We invite the reviewer to read our paper again. In many places, model errors are discussed. In the abstract, we wrote "Biases in representing vertical CO profiles are found over the ocean and most significantly in the Southern hemisphere, suggesting errors in the vertical distribution of CO chemical sources/sinks or in the vertical mixing to be improved in future modelling studies"; then, we described in the introduction that "model results are still associated with systematic uncertainties due to a complex interplay of source distributions, transport and chemistry" and "In order to diagnose these inconsistencies that could result from model errors or satellite retrieval biases or both". Other places like lines 443-445, 470, 511-513, 599-666, 669-680 all discussed model errors. We will better organize the discussion to reflect this point.

> *Lines 760-765: I am again not sure about the validity of those statements. There is a difference between assimilating surface network data and assimilating surface re- trieved data. The representativity of this two types of data sets are fundamentally dif- ferent, e.g. coverage, revisit/time-sampling, accuracy, spatial resolution. This is again very speculative, consider removing.*

Our statement is actually a verbatim from (Rayner and O'Brien 2001). We also referred to "Because there is so much less spatial variability associated with column CO_2, even with just a few observations, it will be possible to assess the strength of the Northern Hemisphere carbon sink using a measurement approach that is far less sensitive to model representations of vertical mixing" (Olsen and Randerson, 2004) and to "Because these observations are of the column or partial column abundance, they come close to directly representing a measure of atmospheric CO_2 mass per unit area. As a result, our estimate of NEE are less sensitive to errors in the vertical transport than estimates based solely on surface mixing ratio observations" (Yang et al. 2007). None of these sentences from past papers, either explicitly or implicitly referred to coverage, revisit/time-sampling, accuracy, or spatial resolution.

Line 767: "On either the satellite", syntax error, please rephrase.

We have replaced "either … or" by "both … and".

Lines 776-777: Barré et al., 2015 conducted a study assimilating MOPITT and IASI and compared biases and errors with an extended set of independent observations for validation. Please refer to this paper.

We added a reference to Barre et al. (2015).

References (from the reviewer):

Barré, J., Gaubert, B., Arellano, A. F. J., Worden, H. M., Edwards, D. P., Deeter, M. N., ... Hurtmans, D. (2015).Âa˘Assessing the impacts of assimilating IASI and MOPITT CO retrievals using CESM-CAM-chem and DART.Âa˘Journal of Geophysical Research: Space Physics,Âa˘120(19), 10501-10529. DOI:Âa˘10.1002/2015JD023467

Inness, A., Blechschmidt, A.-M., Bouarar, I., Chabrillat, S., Crepulja, M., Engelen, R. J., Eskes, H., Flemming, J., Gaudel, A., Hendrick, F., Huijnen, V., Jones, L., Kapsomenakis, J., Katragkou, E., Keppens, A., Langerock, B., de Mazière, M., Melas, D., Parrington, M., Peuch, V. H., Razinger, M., Richter, A., Schultz, M. G., Suttie, M., Thouret, V., Vrekoussis, M., Wagner, A., and Zerefos, C.: Data assimilation of satelliteretrieved ozone, carbon monoxide and nitrogen dioxide with ECMWF's Composition-IFS, Atmos. Chem. Phys., 15, 5275– 5303, doi:10.5194/acp-15-5275-2015, 2015.

Âa˘Miyazaki, K.,Âa˘Eskes, H. J., and Sudo, K.: A tropospheric chemistry reanalysis for the years 2005–2012 based on an assimilation of OMI, MLS, TES, and MOPITT satellite data,Âa˘Atmos. Chem. Phys., 15, 8315-8348, doi:10.5194/acp-15-8315-2015, 2015 Migliorini, S., C. Piccolo, and C. Rodgers, 2008: Use of the information content in satellite measurements for an efficient interface to data assimilation.Âa˘Mon. Wea. Rev.,Âa˘136, 2633–2650.

We did not include a reference list here, since they have been included in the original manuscript or introduced by the reviewer.