Authors' response to Reviewer 1

We are thankful to both reviewers for their positive and accurate feedback on our study, and for the improvements they helped us make to this paper.

The responses are organized as follows: the reviewer's comment is in blue, our answers are in black, and the changes proposed for the revised manuscript are in italic (black for modified sentences, grey for unchanged sentences that have been pasted here in order to remind the context).

The manuscript presents a comparison of O3, CO, NOy and H2O from a nudged simulation of the LMDZ-OR-INCA chemistry-climate model with a set of long-running observations of these trace gases from instrumentation carried on-board commercial aircraft. The comparison is made over the upper troposphere / lower stratosphere region from observations made during cruise-level flights using the Interpol-IAGOS software package described in an earlier paper. To account for the strong discontinuities around the tropopause, separate comparisons can be made for the lower stratosphere and the upper troposphere. In addition, the comparison to observations are made for two sensitivity simulations: one with emissions from biomass burning turned off, and one with emissions from lightning turned off.

The observations shows many of the large scale features of the distribution of these trace gases associated with, for example, the differing height of the tropopause between the tropics and the mid-latitudes, the seasonal and regional nature of biomass burning emissions and the impact of monsoon circulations. The comparison with the model shows the model has an ability to reproduce many of these features, with the most significant and widespread bias being for CO in the lower stratosphere. The paper is generally well organized, though the use of particular phrases or words makes comprehension challenging in a few places. I have tried to point these out in the minor comments.

My only significant concern is the treatment of the vertical height coordinate when comparing the aircraft observations to the model. From Section 2.3.1, I can understand how the IAGOS observations are horizontally gridded. And I can understand how the IAGOS data is binned for the version of the data separated into the upper troposphere and the lower stratosphere. With most (or all?) of the data from cruise altitude, what I am missing is how variations in the cruise altitude are treated. At line 197 it is stated 'The climatologies here refer to nearly horizontal maps derived from partial columns in the cruise altitudes.' Are the point measurements assumed to be representative of a certain vertical range to allow a partial column to be calculated? Or are all the different aircraft observations in a particular month kept on their individual altitude points and these are combined to produce a vertically integrated (or vertically averaged) quantity at each grid point? Or

maybe the relatively small variations in altitude are ignored and all data is assumed to be on an average cruise altitude? It is particularly important for the comparison in the lower stratosphere, which shows such strong vertical gradients in CO and ozone.

We thank Reviewer 1 (R1) for pointing out this need of clarification. The altitude of the measurements is taken into account during the interpolation step, by the use of the weighting coefficients (as for the horizontal axes). Once the 3D climatological means are calculated, we derive the vertical averages from the grid cells that meet the sampling criteria.
Thus, the "partial columns" are not necessarily representative of a geographically constant altitude, but the altitude range remains consistent between the model and the gridded observations.

The first phrase of the subsection 2.3.1 has been completed:
"The strategy consists of adapting the IAGOS data to the studied simulation in matter of spatial resolution, following a linear reverse interpolation ***onto the three spatial dimensions****."*

Also, the term "partial columns" has been replaced by "vertical means" or "vertical averages".

Is it sufficient just to classify observations at 'lower stratosphere' without taking into account the distance to the tropopause?

It depends on the accuracy one wants to give to the assessment of the model. Dividing the lower stratosphere into sub-layers would bring further information on the transport processes between troposphere and stratosphere, and could be tested in the next study. The assumption of the LS as a uniform layer is not realistic geophysically, but it still brings relevant information on the model abilities to reproduce the UTLS behaviour.

Minor comments:

Line 2 – The use of 'the latter' in 'The latter is regularly…' does not work here because the preceding sentence does not present two options.
According to the comments from R1 and R2, the formulation has been changed into:
"Evaluating global chemistry models in the upper troposphere - lower stratosphere (UTLS) is an important step toward ***an improved understanding of the chemical composition in this region. This composition*** *is regularly sampled […]"*

Line 15 – is there a word missing in 'as [are] the observed CO peaks due to biomass burning…'
Following the suggestions from R2, the phrase has been reformulated into:

"In the tropics, the upper-tropospheric climatologies are remarkably well simulated for water vapour; they also show well-reproduced CO peaks due to biomass burning in the most convective systems, and the ozone latitudinal variations are correlated between the observations and the model."

Line 139 – Does 'In this study, the LMDZ GCM surface zonal and meridional wind components are nudged…' mean that only the surface winds (lowest model layer) are nudged and the rest of the atmosphere is allowed to freely evolve?
There has been a mistake in the manuscript (which has been corrected): the horizontal winds are nudged at every level, not only at the surface.

Line 168 – 'precising' might be better as 'denoting'
This suggestion has been applied.

Line 179 – what is meant by 'if it is not adjacent to the 2 PVU isosurface'? What comes between a particular grid point and the 2PVU isosurface so that it is not adjacent?
This phrase has been clarified as follows:
"a sampled grid point is considered as upper-tropospheric if its PV is lower than 2 PVU, if it is not the first gridpoint below the 2 PVU isosurface, and […]"

Lines 173 – 195 – A more general comment on Section 2.3.2: There is no discussion of how PV is used to classify the IAGOS measurements. I assume reanalysis data is interpolated in space and time to each IAGOS measurement and PV is calculated with the same exclusion of the transition layer?
Though there exists a IAGOS product which interpolates the ERA-Interim PV onto the aircraft trajectories, it is not used in this study. In our case, the classification is the same between the model gridpoints and the IAGOS-DM gridpoints, as it only accounts for the model daily PV field.
In order to clarify it, we added the following phrase at the end of the paragraph:
"It is worth precising that the same classification applies between the INCA-M and the IAGOS-DM grid points, using the model PV field."

Line 188 – I think 'Consequently, it is worth figuring out that…' would be better as 'Consequently, it is worth keeping in mind that…' or maybe 'worth remembering…'?
The former suggestion has been applied.

Line 219 – I am not sure what is meant by 'In the tropics, the threshold is adapted to the seasons duration by applying a cross product.' A cross product is usually an operation on vectors in linear algebra.
The formulation has been corrected, as follows:
"In the tropics, the threshold is adapted proportionally to the seasons duration."

Line 349 – For 'we chose to draw the mean ratio', in place of 'draw' that can have

different meanings, could I suggest 'plot' or 'display'?
Agreeing with R1, we replaced the verb "draw" by "display".

Lines 353 – 354 – I think it is a bit of a jump to suggest that biomass burning and lighting are realistically distributed because the simulations without these important sources has a poorer comparison with observations.
This phrase has been replaced by:
"*First, the comparison between the different runs shows a better correlation in the reference simulation in the UT, **implying that the impacts from lightning and biomass burning in the reference simulation contribute to a non-negligible part of the geographical similarities between IAGOS-DM and INCA-M.**"*

Lines 354 – 366 – This section jumps around a lot across the different panels of Figure 6 and is difficult to follow. For example,

First, the comparison between the different runs shows a better correlation in the reference simulation in the UT, possibly suggesting that the effects from biomass burning and lightning emissions on ozone production are realistically distributed in space. As expected, no change is observed in the LS for this metric, since the higher amounts of ozone in the LS increase the NOx threshold necessary to trigger a net ozone production (e.g. Hegglin et al., 2006).'

From the second sentence it is not clear what metric is being discussed when it is stated 'no change is observed in the LS for this metric'.
It has been replaced by a more explicit formulation:
"*As expected, no change **in the ozone correlation** is observed in the LS, **[…]**"*

Lines 367 – 368 – 'the shorter and lesser NOy sampling does not lead to strong differences.' might be clearer as 'the shorter period of time and sparser measurements of NOy does not lead to strong differences.'
The formulation suggested by R1 has been substituted to the previous one.

Lines 373 – 375 – I would agree with the statement 'It is likely that the influence of biomass burning on the LS is overestimated because of an excessive exchange between the troposphere and the stratosphere.' The change in bias in CO in the LS is quite surprising.

Line 384 – I can deduce what is meant by 'barycentre' but it is not a correct word. 'mean pressure of the measurements' maybe?
The phrase has been shortened, but made clearer:
"The mean pressures on the right axis have been added in order to identify changes in mean altitude measurements."