## Authors' response to Editor's comments

Dear Dr. Yann Cohen,

I sent the revised version of your manuscript to one of the referees for further review. Unfortunately, this referee did not give a response although he/she had expressed a willing to further review your manuscript before.

I think that in the Author Response you have addressed the issues raised by the referees, and the revised manuscript can be accepted for publication in ACP subject to further minor revision. For example, your Article title seems not to be so concise by using only "climatologies", which cannot be easily judged what it means by a reader who does not know the IAGOS data well. The abstract should have had fewer than 250 words. I suggest that you make necessary revisions according to the Guidelines for authors at the ACP journal webpage (https://www.atmospheric-chemistry-and-physics.net/policies/guidelines\_for\_authors.html).

Yours sincerely, Jianzhong Ma

Dear Editor,

We thank you for these essential comments. The title is now clarified, and the abstract has been shortened down to 250 words, as required by the journal.

1/ The title has been modified as follows:

"Evaluation of  $O_3$ ,  $H_2O$ , CO and  $NO_y$  climatologies simulated by four global models in the upper troposphere—lower stratosphere with the IAGOS measurements"

## to be compared to its previous version:

"Multi-model assessment of climatologies in the upper troposphere—lower stratosphere using the IAGOS data"

With this new version, we clarify the fact that IAGOS are measurements, and we added the chemical species to provide context for the word "climatologies". We keep the latter to make sure the reader does not expect any time series, or trend calculation.

2/ The abstract has been reduced. The new version is quoted below:

"Assessing global models in the upper troposphere (UT) and in the lowermost stratosphere (LS) is an important step towards a better understanding of the chemical composition near the tropopause. For this purpose, the current study focuses on an evaluation of long-term simulations from four chemistry-climate/transport models, based on the In-service Aircraft for a Global Observing System (IAGOS) measurements. Most simulations span over the 1995–2017 period, and follow a common protocol among the models. The assessment focuses on climatological averages of ozone ( $O_3$ ), water vapour ( $H_2O$ ), carbon monoxide (CO), and reactive nitrogen ( $NO_y$ ). In the extra-tropics, the models reproduce the seasonality of ozone,  $H_2O$ , and  $NO_y$  in both UT and LS, but none of them reproduces CO springtime maximum in the UT. Tropospheric tracers (CO, CO, CO) tend to be underestimated, consistently with an overestimation of cross-tropopause exchanges. Most models systematically overestimate ozone in the UT, and nitrogen oxides (CO) background appears as the

main contributor to ozone variability across the models. The partitioning between  $NO_y$  species changes drastically across the models, and acts as a source of uncertainty on  $NO_x$  mixing ratio and on its impacts on atmospheric composition. However, we highlight some well-reproduced geographical variations, as the ITCZ seasonal shifts above Africa, or extratropical ozone ( $H_2O$ ) in the LS (UT) correlated with the observations. These features are encouraging regarding the simulated dynamics in both layers. The current study confirms the importance of separating the UT and the LS with a dynamical tracer for model results evaluation, and for model intercomparisons."

## To be compared with the previous version:

"The evaluation of global chemistry—climate/transport models in the upper troposphere—lower stratosphere (UTLS) is an important step towards a better understanding of the chemical composition near the tropopause, and therefore towards a more accurate assessment of the impact of NO<sub>x</sub> emissions in this region of the atmosphere, notably by subsonic aviation. For this purpose, the current study focuses on an evaluation of long-term simulations from four global models based on in-situ measurements on board passenger aircraft (IAGOS). Most simulations span over the 1995–2017 time period, and follow a common protocol among the models. The assessment focuses on climatological averages of ozone  $(O_3)$ , water vapour  $(H_2O)$ , carbon monoxide (CO), and reactive nitrogen compounds (NO<sub>v</sub>). In the extra-tropics, the models reproduce the seasonality of ozone, water vapour, and NO<sub>v</sub> in both the upper troposphere (UT) and the lowermost stratosphere (LS), but none of them reproduces CO springtime maximum in the UT. The tropospheric tracers (CO and H<sub>2</sub>O) tend to be underestimated by the models, which is consistent with an overestimation of the cross-tropopause exchange. It does not exclude other factors as an underestimation of CO emissions, an underestimation of transport from the surface, or an overestimated CO oxidation by the hydroxyl radical (OH), but the mean distance between observations and tropopause does not vary substantially across the models, and the intermodel differences are not correlated with these chemical biases. Ozone is systematically overestimated in the UT by most models, and NO<sub>x</sub> background appears as the main contributor to ozone variability across the models. The partitioning between NO<sub>v</sub> species changes drastically across the models, and acts as a source of uncertainty on NO<sub>x</sub> mixing ratio and on its impacts on atmospheric composition. However, independently of the mean biases, we highlight some well-reproduced geographical and seasonal distributions, as the ITCZ seasonal shifts above Africa, the upper-tropospheric  $H_2O$  maximum in the Asian summer monsoon, and the extratropical ozone (H<sub>2</sub>O) in the LS (UT) that shows a high correlation with IAGOS (In-service Aircraft for a Global Observing System) observations. These features are encouraging regarding the simulated dynamics in both the troposphere and the stratosphere. The current study confirms the importance of an accurate separation between the UT and LS using a dynamical tracer for model results evaluation but also for model intercomparisons."