Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2020-328-RC1, 2020


Geoscientific Model Development Discussions

# *Interactive comment on* "On a new assessment method for long-term chemistry-climate simulations in the UTLS based on IAGOS data: application to MOCAGE CCMI-REFC1SD simulation" *by* Yann Cohen et al.

**Anonymous Referee #1**

Received and published: 9 December 2020

This work introduces a novel methodology and database based on the IAGOS passenger aircraft observations to evaluate chemistry-climate models in the UTLS region, where there is generally a lack of accurate observations for this purpose. This is a very valuable contribution to the effort of multi-model comparisons and the authors should consider adding their tool to the ESMVal toolbox. However, before I can recommend publication of the manuscript, I would ask the authors for some clarification of their methodology and also to strengthen their case that their rather complex approach is yielding more robust insight into model deficiencies than from more traditional ap-



proaches as outlined in the following.

Major comments

P1L12-14 The sentence 'as a first application. . .' seems confusing at this place. The authors should try to clarify (or attempt a better logical flow of) their description of the evaluation methodology they introduce. First you interpolate IAGOS onto the model grid, then you mask the model at the available IAGOS grid points, then you attempt the evaluation of the model only, right?

The authors have done a commendable job in summarising references in the introduction that use satellite observations for model comparisons using SD simulations and IAGOS particularly. However, there are earlier studies that are also highly relevant to the here proposed evaluation methodology and hence should be mentioned as well. Brunner et al. (2003, 2005), Tilmes et al. (2010), and Hegglin et al. (2010) all use aircraft (and satellite) observations to evaluate CTMs and free-running CCMs in the UTLS, respectively. At least a qualitative discussion of how your findings improve (or whether they yield similar results) compared to the latter two studies in particular is needed, given they use tropopause-based coordinates to evaluate CO and O3 in the upper troposphere and lower stratosphere to evaluate CCMs, an equivalent methodology as proposed here but applied to free-running simulations. While you may argue that your method has the advantage to have near-coincident measurements for the model evaluation, the problem with the SD-simulations could be that certain nudging methodologies may lead to introducing noise (and hence too much mixing) especially visible around the tropopause (see discussion in Orbe et al. 2020). This issue indeed could help your interpretation of your findings (too much/too little ozone in the UT and LS, respectively, and vice-versa for CO).

P4L3 ff. The statement that it is not possible is an exaggeration. As shown by Brunner et al. (2003, 2005) interpolation onto the measurement space can yield in-depth insight into model behaviour and even though the data may not be available from the

multi-model intercomparison data archives, higher resolution that would be needed for this task could be obtained from the modelling centers. In fact, it would be important to prove for a methodology paper as you have presented here that your claim of the gridded IAGOS data being representative of the monthly mean is true. You could do this by sub-sampling your model according to IAGOS measurement locations (including weighting and interpolation on measurement location), then compare the mean of this sub-sampled model field to the IAGOS-based masked model field. Agreement between the two will tell you the accuracy of the assumption underlying your methodology. This extra step in my eyes would be needed to make this methodology paper a sound contribution to model evaluation efforts, and without it the question remains whether differences in the model-measurement comparison arise from comparing apples to oranges as opposed to true model deficiencies.

On a similar note, it would also be good to see what the benefit of the weighted gridding versus a gridding of the observations without the weighting function would be (equivalent to what you have done in Section 4.2 for IAGOS-DM and IAGOS-HR). This is important in order to convince the reader that the weighted gridding process is in fact worth the effort. I would expect that the interpolation/weighting is less needed for the maps on different pressure levels, unlike the result of the tropopause-based evaluation in Section 4.2, but I could be wrong.

P9L22 I am not clear why/how the shorter period for CO measurements compared to O3 matter here? After all, 10 or 20 years may result in not so different climatologies, in fact a 10-year climatology may be more representative than a 20-year one (as for ozone) since you will find quite some trends in the latter. In any case, the determination of Nthres seems rather arbitrary. How did you determine it? Did you make a quantitative analysis by how much the mean would change if more (or less for that matter) than 100 samples were included in the mean?

Minor comments



P10L21 to P11L5-15 Please clarify/write more succinctly here that you use monthly PV fields to distinguish between UT and LS observations. Also, it is not clear to me how the additional screening is applied to the data to get rid of too high (low) values in the UT (LS) respectively (is it applied to the monthly means only or to the input data that are used to generate these monthly means?). This approach seems to introduce a discrepancy between how you treat the model data (for which you were not able to do such a screening) versus the observations, and hence could be a partial explanation of why the model and observations don't agree with each other.

P11L21-23 Again I'm confused. Why is the threshold here 150 and not 100 anymore as mentioned earlier?

Figures 3 and 4, also supplementary figures. I suggest plotting the difference between these fields as well, since it is difficult to compare these plots quantitatively.

Correlations and means in Table 3 and 4 should be calculated for IAGOS-HR as well.

Please double-check your language once more before publication. A few issues are listed below, but this is not a comprehensive list of corrections. . .

P1L5 If a few. . . → Even though a few. . . P1L9 This argument is generally only true for archives of model-intercomparison studies, many institutions save their model fields on a daily or 10-hourly basis. Suggest a corresponding qualifier for this statement. P1L10 mapped at → mapped onto, or do you mean sampled at? P1L10 by MOCAGE CTM. . . of CCMI phase-1 → by the MOCAGE CTM. . . of the CCMI phase-1 P1L14 'Good correlations' → 'Good agreement', you don't calculate correlations. . . P2L4 'only due to . . .' → 'arising from. . .' P2L10 'on the processes' → 'in the processes' P2L12 'the main goal [. . .] lies on the reduction. . .' → 'the main goal [. . .] is the reduction. . .' P2L12 since you indicate the projects of ACCMIP and CCMVal, you should also do so for CCMI, which is IGAC/SPARC. Same on P5 Section 2.2 on CCMI. P2L18/19 How is it achieved? Recommend adding in a short explainer for 'SD' being 'nudged to observed meteorology' or 'specified dynamics' already here. P4L14 that studied → who studied

P4L15 what do you mean by quasi-totality? P6L19 I don't understand what the eccentricity of measurements refers to, please clarify. P6L22 Not clear what this refers to, I assume this approach? P6L25 of MOCAGE simulation → of the MOCAGE simulation P8L15-17 clarify sentence, did you mean . . . has a negligible impact on the distribution of the IAGOS data from the cruise altitude onto the model vertical grid? P9L1 in this vicinity → in its vicinity to clarify that you refer back to the grid point. P9 equation 3: the multiplication sign looks formatted in a funny way, suggest to remove. P9L8 suggest to reorder sentence for clarification to → . . . on the MOCAGE vertical levels considered which span level 28 to 22 and correspond to. . . P9L11 Do you really mean to say that all observations on levels 27 and 28 correspond to ascent or descents of the aircraft? P9L15 This is only true if the observations going into a grid box of your IAGOS-DM is representative for the monthly mean. This will depend on the number of measurements (your N) and the temporal sampling (distributed evenly over the month versus other sampling). P10L12 inconvenient → inconvenience P18L16 be more specific for the general reader → tropopause transition layer P20L27-29 inter-regional averages is not a valid expression. Cross-regional perhaps. . . but maybe better an average calculated over all regions. . . P20L28 as they are similar with zonal averages. . . → as they are similar to the zonal averages. . .

Additional References

Brunner, D., et al. (2003), An evaluation of the performance of chemistry transport models by comparison with scientific aircraft observations. Part 1: Concepts and overall model 2076 performance, Atmos. Chem. Phys., 3, 1609–1631.

Brunner, D., et al. (2005), An evaluation of the performance of chemistry transport models. Part 2: Detailed comparison with two selected campaigns, Atmos. Chem. Phys., 5, 107–129.

Hegglin, M.I., Gettelman, A., Hoor, P., Krichevsky, R., Manney, G.L., Pan, L.L., Son, S.W., Stiller, G., Tilmes, S., Walker, K.A. and Eyring, V., 2010. Multimodel assessment



of the upper troposphere and lower stratosphere: Extratropics. Journal of Geophysical Research: Atmospheres, 115(D3).

Orbe, Clara, David A. Plummer, Darryn W. Waugh, Huang Yang, Patrick Jöckel, Douglas E. Kinnison, Béatrice Josse et al. "Description and Evaluation of the specified-dynamics experiment in the Chemistry-Climate Model Initiative." Atmospheric Chemistry and Physics 20, no. 6 (2020): 3809-3840.

Tilmes, S., et al. (2010), An aircraft based upper troposphere lower stratosphere O3, CO, and H2O climatology for the Northern Hemisphere, J. Geophys. Res., 115, D14303, doi:10.1029/2009JD012731.

---

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2020-328, 2020.