Atmos. Chem. Phys. Discuss.,
doi:10.5194/acp-2016-179-RC2, 2016

Atmospheric Chemistry and Physics Discussions

Interactive comment on "An assessment of the climatological representativeness of IAGOS-CARIBIC trace gas measurements using EMAC model simulations" *by* J. Eckstein et al.

Anonymous Referee #2

Received and published: 19 June 2016

The manuscript "An assessment of the climatological representativeness of IAGOS-CARIBIC trace gas measurements using EMAC model simulations" aims at the characterisation and analysis of representativeness of the IAGOS-CARIBIC trace gas climatology for the upper troposphere/lower stratosphere (UTLS) as obtained from observations onboard commercial aircraft. Representativeness is analysed applying different sampling strategies, including the IAGOS-CARIBIC sampling, to a global scale chemistry transport model and evaluating the obtained populations by means of statistical tests and descriptors. The findings of this study are certainly relevant to the IAGOS-CARIBIC programme as a whole, since they challenge the use of IAGOS-CARIBIC data in the climatological sense. In general, the manuscript is well written and tries to

justify the applied methodologies. However, there are several areas where the choice of methods seems to be rather arbitrary and needs either additional justification or analysis to corroborate the conclusions reached in the manuscript. After these issues are addressed this contribution should be well suited for publication in Atmospheric Chemistry and Physics.

Major comments

Global scale chemistry transport model

There are two major concerns about using the EMAC model as a reference state of the atmosphere. First, the model description in the text is insufficient. It needs to be mentioned how the model was validated against other independent observations. For which species did the model perform well and for which not? Where is the model insufficient to reproduce variability on the scale given by the model resolution? This is especially important since one may suspect that the model will have difficulties reproducing vertical trace gas gradients in the UTLS region. Second, as shown in Figure 1 the model has only 3 levels in the UTLS region and output was only available every 12-hour. Therefore, the model misses large parts of the real variability (see also the CARIBIC comparison). How can it be justified that the model can still be assessed to analyse representativeness?

Sampling strategy

Several choices seem to be arbitrary. I especially don't understand why the temporal domain is not sampled as a whole. Both sampling patterns RANDPATH and RANDLOC only sample 12 and 8 days per month, respectively. It would seem more appropriate to sample daily but on the other hand with a more realistic pattern that resembles that of the CARIBIC flights (i.e., on great arcs between major airports in the northern hemisphere, leaving out transpacific flights, since this region is never covered by CARIBIC). In that case the RANDPATH sampling could be viewed as the maximal achievable sampling pattern by commercial aircraft and RANDLOC could still be seen as sampling the

northern hemispheric UTLS region as a whole.

Selected statistical measures

Again there seem to be arbitrary choices concerning the statistical estimators and tests. If the Komogorov-Smirnov test turned out to be too strict because it requires similarity of the whole distribution, why did you not select other statistical tests that only evaluate one statistical parameter at a time (e.g., Mann-Whitby test for the mean and Levene's or Brown–Forsythe test for variance, all are non-parametric tests suited for atmospheric trace gas observations). Furthermore, the results need to be discussed together with observed seasonality of the trace species as is mentioned by the authors themselves on page 17, line 1, but than dropped without further reasoning 3 lines later. The relative difference does not contain much information in itself and as stated correctly depends on the lifetime of a species.

Minor comments

P1,L11: "formulated above". Not clear from the context where this was formulated

P3,L28ff: Although no details on the measurement techniques are needed here, it would still be interesting to learn something about the overall uncertainties of the measurements and how these compare to the later discussion of representativeness.

P4,L10: Model output every "eleven hours"? Did you mean 12 hours?

P4,L9ff: Additional information on emissions used in EMAC and vertical resolution in the UTLS region would be useful here.

Section 3.1: It should be more prominently mentioned in the first paragraph of this section that you restrict the analysis to the latitude region 35N to 75N. Details follow towards the end of the section and can remain there, but it would be good to make this important detail clear from the beginning. It should also be stated in the abstract.

Table 1: For RANDPATH it is an adjusted Gaussian distribution, as mentioned in the

text.

Table 1 and elsewhere: "Uniform" or "rectangular" distribution should be used instead of "even".

P6,L6f: The good correlation for temperature is not a big surprise, given the strong vertical stratification in the UTLS and the assumably large number of measurements. Since this is one of the few pieces of model validation mentioned, one could add a scatter plot to the supplement.

P6,L9f: It is not clear to me why the limited vertical model resolution is the reason you cannot compare CARIBIC directly to EMAC. The random sampling is still done using vertical interpolation to specific pressure levels. Would't the same argument apply to the random sampling strategy as well and could one not simply drop it and do the analysis of representativeness on discreet model levels instead?

P6,L20f: Why did you choose these cut-off values instead of simply using the standard deviation as a criterion (i.e., redistribute values outside +/- 2 sigma). I don't assume this would change much, but would seem statistically more sound. Alternatively, one could have sampled directly from the observed CARIBIC distribution.

P7,L7ff: I don't agree with the statement that the distribution "is very similar for all datasets". There is a strong offset to higher HrelTP in both random sampling strategies. What is the actual mean HrelTP for all three samples?

p7,L11f: This requires some further justification (see major comment above). Without being aware of the details of Jöckel et al 2015, it seems a bit hard to believe that the model performs equally well for the very different set of species analysed here. There should be additional discussion of the species for which this may not be justified.

p9,l19: How was the mean tau$_s tar calculated? As the mean overall monthly tau_s tar sor as tau_s tar$

Figure2: It would be interesting to add CARIBIC observed tau$_s tar in the figure (where available)$.

Figure3 and others: The y-axis if often titled "variability". It would be useful to give a more concrete title, since the manuscript is dealing with all kinds of variability. This could reduce confusion. In this specific case I assume this is relative standard deviation?

p12,l18: "The differences are small, mostly below an absolute value of 0.15." But this means that the absolute difference between both samples is 1.4 times larger than the value of the reference (or am I mistaken). I am not sure that I would call this small! In general using the log scaled relative difference seems a bit odd and only confuses. Why not use the relative difference as is?

p12,l29f: "A similar analysis has also been performed with data from a random number generator, leading to equivalent results." Are you referring to the RANDLOC sample here?

p13,l13: At least repeat the result of the sensitivity study here. The supplement should not be a paper on its own.

p16,l5: Not clear which correlation is referred to here.

p16,l6: What is an "uncertainty error"? I think the use of representativeness uncertainty would in general work better.

p16,l8-13: This description is completely confusing. I don't understand what is done and why. Please improve the description.

p17,l10ff: Since the discussion on NOx is along the EMAC results, it would be interesting to know how NOx sources in the UTLS are treated in the model. Does the model include a realistic representation of lightning NOx? Has this been analysed in previous studies?

p17,l33f: The representativeness uncertainty of 5lived trace gases is huge considering their atmospheric abundance. It is much larger than their seasonal variability. This aspect needs to be considered in the analysis and discussed along with the results.

p18,l20ff: Finally there is some discussion using species specific thresholds, but again these thresholds are chosen without any justification. They should be related to seasonal variability.

p18,l32: Was it ever shown that R_{rel}"$increases linearly$"?$May be an increasing relationship, b$

Figures: It would be easier to follow the discussion of the figures if sub-panels would be labelled by letters (which is Copernicus style). For example discussion of Figure 8 on page 18.

Figure 1 in supplement: Please explain black line in legend and add fit as additional line to the plot.Indicate which species are behind each point. Is the given fit applied to log(meas) and log(model)? Is it just my impression or does the model actually capture less of the variability for species that have a small relative variability? How could this be explained? I would have expected the opposite.

Technical comments:

P1,L2: It is "representative of" not "representative for".

P5,L6 and elsewhere: "Data" is always plural. Change to "Data were used ..."

Interactive comment on Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-179, 2016.