Supplement to:

An assessment of the climatological representativeness of IAGOS-CARIBIC trace gas measurements using EMAC model simulations

Johannes Eckstein[1], Roland Ruhnke[1], Andreas Zahn[1], Marco Neumaier[1], Ole Kirner[2], and Peter Braesicke[1]

[1] Karlsruhe Institute of Technology, Institute of Meteorology and Climate Research, Herrmann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany
[2] Karlsruhe Institute of Technology, Steinbuch Centre for Computing, Herrmann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany

Correspondence to: Johannes Eckstein (johannes.eckstein@kit.edu)

Abstract. Measurement data from the long-term passenger aircraft project IAGOS-CARIBIC is often used to derive trace gas climatologies. We investigate to what extent such derived climatologies can be assumed to be representative for the true state of the atmosphere. Using the chemistry-climate model EMAC we sample the modelled trace gases along CARIBIC flight tracks. Different trace gases are considered and climatologies relative to the mid-latitude tropopause are calculated. Representativeness can now be assessed by comparing the CARIBIC sampled model data to the true climatological model state. Three statistical methods are applied for this purpose: the Kolomogorov-Smirnov test, and scores based on the variability and relative differences.

Generally, representativeness is expected to decrease with increasing variability and to increase with the number of available samples. Based on this assumption, we investigate the suitability of the different statistical measures for our problem. The Kolmogorov-Smirnov test seems too strict and does not identify any climatology as representative – not even long lived well observed trace gases. In contrast, the variability based scores pass the general requirements for representativeness formulated above. In addition, even the simplest metric (relative differences) seems applicable for investigating representativeness.

Using the relative differences score we investigate the representativeness of a large number of different trace gases. For our final consideration we assume that the EMAC model is a reasonable representation of the real world and that representativeness in the model world can be translated to representativeness for CARIBIC measurements. This assumption is justified by comparing the model variability to the variability of CARIBIC measurements. Finally, we show how the representativeness score can be translated into a number of flights necessary to achieve a certain degree of representativeness.
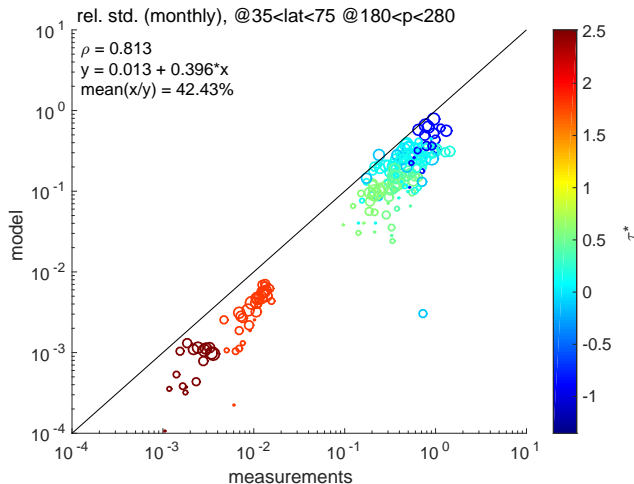
Figure 1. Monthly variability σ_r of MOD$_{\text{CARIBIC}}$ over MEAS$_{\text{CARIBIC}}$. Colorcoding corresponds to the variability τ^* of each species. Data closer to the tropopause is plotted as larger circles.

1 Introduction

This supplement discusses further results of the study of the representativeness of IAGOS-CARIBIC data using the chemistry-climate model EMAC. For abbreviations and methods, please refer to the main text. Four points are discussed here:

Section 2 briefly shows results of the comparison of model and measurement variability. The methods to describe rep-
5 resentativeness developed and tested with model data were also applied to data from a random number generator. This is described in Section 3. Section 4 discusses the sensitivity study of the Kolmogorov-Smirnov test using a subsample of MOD$_{\text{CARIBIC}}$. Section 5 shows how the representativeness uncertainty of MOD$_{\text{CARIBIC}}$ decreases if the pressure range is increased to $10\,\text{hPa} < p < 500\,\text{hPa}$, i.e. how the climatologies produced with data from IAGOS-CARIBIC are dependent on the pressure at which samples are taken.

2 Comparing measurement and model variability
10

In order to compare model and measurement variability, the relative standard deviation $\sigma_r = \sigma/\mu$ (σ being the standard deviation, μ the mean) was calculated for MEAS$_{\text{CARIBIC}}$ (CARIBIC measurements) and MOD$_{\text{CARIBIC}}$ in each month. $\sigma_r^{\text{MOD}_{\text{CARIBIC}}}$ and $\sigma_r^{\text{MEAS}_{\text{CARIBIC}}}$ were calculated in each month. Figure 1 shows the correlation of $\sigma_r^{\text{MOD}_{\text{CARIBIC}}}$ and $\sigma_r^{\text{MEAS}_{\text{CARIBIC}}}$.

As discussed in the main text, $\sigma_r^{\text{MOD}_{\text{CARIBIC}}}$ reaches 40 to 70 % of $\sigma_r^{\text{MEAS}_{\text{CARIBIC}}}$ for all species. The correlation coefficient of the
15 two is 0.81. This shows that the model variability is similar for all species, justifying the use of results from the model datasets for CARIBIC measurements.

3 Calculating representativeness from random numbers

All three methods to investigate representativeness have also been applied to data created with a random number generator. The results of this study are discussed here.

To produce the random numbers, 20 sets of 10^8 numbers were taken from a normal distribution. These 20 sets are referred to as species, well aware of the fact that they are purely artifial. From species to species, the standard deviation σ was set to vary from 10^{-3} to 10^3, values of the exponent again linearly. 20 mean values μ (increasing from 10^4 to 10^8, with a linear increase in the exponent) where distributed randomly onto the 20 species.

3000 samples were taken from each of the 20 species. The sample size increases by 20 for each sample, keeping the sample from before. This way, the relationship of the representativeness score with the sample size is directly accessible. The statistics of each species will be denoted by the index 2, while samples are indexed by 1. For short, this dataset will be named RAND.

The variability τ^* of each species was defined as in Equation 5 of the main text: $\tau^* = \log_{10}(\mu_2/\sigma_2)$, where high values of τ^* stand for low variability. The two requirements set up in Section 3.3 for representativeness in general also have to hold here:

1. Representativeness has to increase with the number of samples.

2. Representativeness has to decrease with increasing variability of the underlying distribution.

With RAND defined in this way, it is possible to test representativeness using the variability analysis following Rohrer and Berresheim (2006) and Kunz et al. (2008) (see Section 4.2) and the relative differences (see Section 4.3). The Kolmogorov-Smirnov test was positive for very few samples (less than fifty numbers, independent of τ^*) and will not be further discussed. Its behaviour with aircraft data was subject of a sensitivity study, the results of which are shown in Sec. 4 of this supplement.

3.1 Variability analysis

The variabiltiy analysis (defined in Section 4.2 and Eq. 3) was applied in a simplified manner. As RAND is independent of time, R_{var} is reduced to just a single value containing the absolute difference of variability of each species of RAND and the sample taken thereof: $R_{var} = |\nu_1 - \nu_2|$, where ν is the mean variability. Figure 2 shows a result. The exact result is a matter of chance, as a random number generator is used. Similar to using $MOD_{CARIBIC}$ and $MOD_{RANDPATH}$, a strong dependance on τ^* and a weak dependance on the number of samples is visible.

Similar to R_{var} when using $MOD_{CARIBIC}$ and $MOD_{RANDPATH}$, the variability analysis using RAND meets the two requirements necessary for describing representativeness, which were described in Section 3.3 and above. This result supports the findng that R_{var} can be used as a statistic for describing representativeness.

3.2 Relative differences

Similar to R_{var}, R_{rel} is reduced to a simple relative difference when using RAND: $R_{rel} = |\mu_1 - \mu_2|/\mu_2$, where μ is the mean. Figure 3 shows the result when applying R_{rel} to RAND. The dependance on τ^* is strong and linear. The result also depends
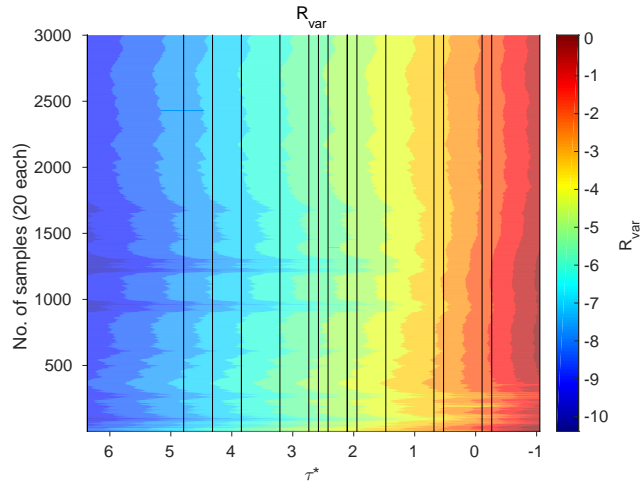
Figure 2. Representativeness score R_{var} applied to RAND. Vertical lines indicate the values of τ^* of each species.
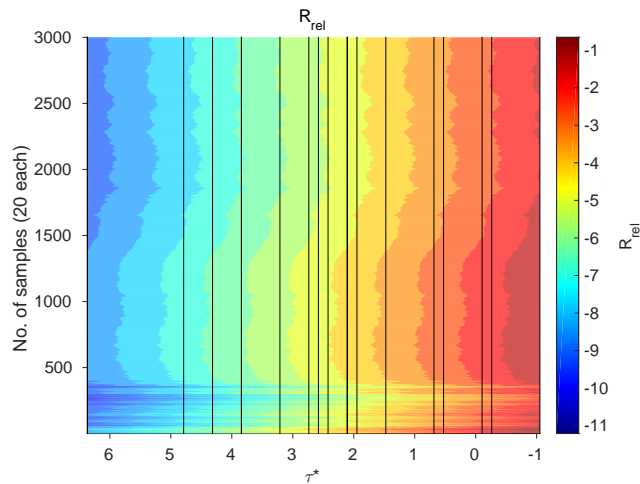
Figure 3. Like Figure 2, but for R_{rel}.

on the number of samples, showing a slow increase with the number of samples. This dependance is sometimes disturbed by better values which are reached by chance when drawing from RAND.

Like for $MOD_{CARIBIC}$ and $MOD_{RANDPATH}$, R_{rel} passes both conditions for a valid description of representativeness: it depends on variability τ^* and on the number of samples. The latter is also being influenced by chance and generally much weaker.

The fact that R_{rel} passes the two conditions for a description of representativeness can be understood with some theoretical considerations. The standard error of the mean is defined by

$$\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}} \tag{1}$$

where $\sigma_{\overline{x}}$, the standard deviation of a sample, can be given by the following equation (N being the number of samples):

$$\sigma_{\overline{x}} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\overline{x}_i - \mu)^2} \tag{2}$$

For $N = 1$, this gives:

$$\sigma_{\overline{x}} = |\overline{x}_i - \mu| \tag{3}$$

Plugging Eq. 3 into Eq. 1 gives:

$$\frac{|\overline{x}_i - \mu|}{\mu} = \frac{\sigma}{\mu\sqrt{n}} = \frac{10^{1/\tau^*}}{\sqrt{n}} \tag{4}$$

and therefore

$$R_{rel} = \log_{10}\left(\frac{|\overline{x} - \mu|}{\mu}\right) = -0.5\log_{10}(n) + \frac{1}{\tau^*} \tag{5}$$

So ideally, R_{rel} should depend inversely on τ^* and directly on the logarithm of the number of values. Figure 3 shows this is approximately true for RAND.

In the case of RAND, R_{rel} can be used to describe representativeness as it passes the two conditions, while R_{var} does not. Theoretical considerations make the finding plausible. RAND can be considered a theoretical abstraction of MOD. The finding here therefore strongly supports that of Sections 5.2 and 5.3, where R_{rel} and R_{var} have also been found to be good descriptors of representativeness when using $MOD_{CARIBIC}$ and $MOD_{RANDPATH}$ or $MOD_{RANDLOC}$. In the main text, we use R_{rel} for final results, as it more suitable to answer the question of representativeness for a climatology.

4 Sensitivity study on the Kolmogorov-Smirnov test

When using $MOD_{CARIBIC}$, $MOD_{RANDPATH}$ or $MOD_{RANDLOC}$, the Kolmogorov-Smirnov test proved not usable, returning all negative results. This indicates that $MOD_{CARIBIC}$ is not representative of $MOD_{RANDPATH}$ in the definition of the Kolmogorov-Smirnov test. This behaviour was tested in a sensitivity study, the results of which are described here.

One of the most frequent destinations within the CARIBIC project is Vancouver, Canada (near $120\,°W$, $45\,°N$, see Figure 4), and only the subset of $MOD_{CARIBIC}$ to this destination is considered in this example to minimize effects stemming of different flight routes. Parts of this reduced dataset were tested with the Kolmogorov-Smirnov test against the whole reduced dataset for all variables. Data was not binned in months, including the whole distribution of datapoints in each height. To produce these partial datasets, each flight was cut into an increasing number of pieces (corresponding to a certain time) and different
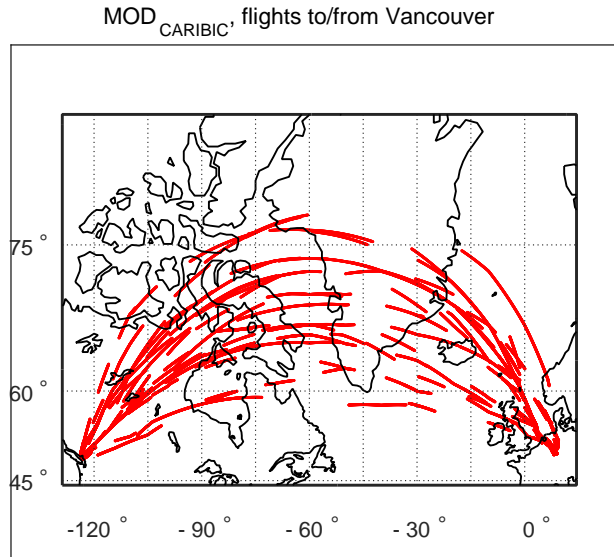
Figure 4. Flightroutes to Vancoucer, Canada, where each flight has been cut into 20 pieces and randomly chosen 30% of those pieces have been plotted. These are tested against the whole data from flights to Vancouver to give one point in Figure 5.

percentages of these pieces were used in testing. Figure 4 exemplifies this method for cutting each flight into 20 pieces and taking 30% of these by showing the corresponding flightpaths.

When applying the Kolmogorov-Smirnov test without binning in months, the result is a profile in HrelTP for each variable. The result can then be diplayed in similar way to Figures 5 and 6. This matrix of height versus species was calculated for each combination of number of pieces and percent of pieces. In each combination, all the profiles of the different variables were averaged to end up with one value betwween 1 and 0 characterizing the result of the test for this combination of number of pieces and percent of pieces. The result can then give an impression of the strictness of the Kolmogorov-Smirnov test.

Figure 5 shows the result of the study. Independent of the number of pieces, the result is positive if all pieces are considered, as the definition of the test prescribes. But only when removing short pieces (shorter than $20\,\mathrm{min}$) is the result also positive for less pieces, even though 70% percent of the data is still needed. When removing whole flights (at the top of the plot), more the 90% of the data has to be taken into account to achieve a positive result of the Kolmogorov-Smirnov test. This result is very similar also for other error probabilities α, taking values of 0.001, 0.01, 0.05 (in the figure), 0.1 and 0.2. The area of failing increases only slightly with the error probability. This showcases the strictness of the test. The Kolmogorov-Smirnov test does not seem suitable to test a dataset measured with aircraft for representativeness of a larger dataset.
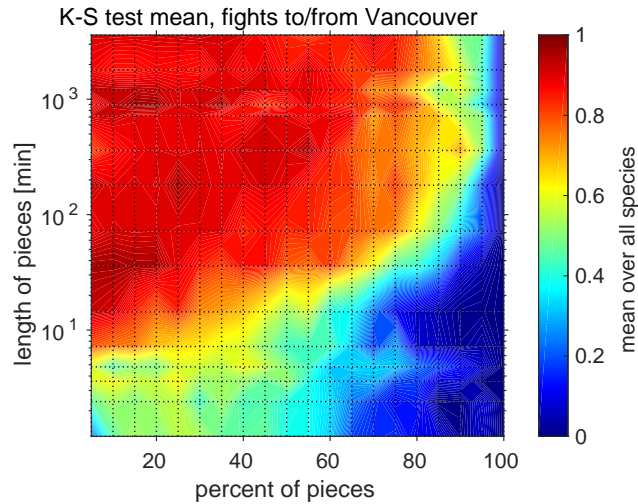
Figure 5. The Kolmogorov-Smirnov test applied to the flights to Vancouver, Canada, of MOD$_{\text{CARIBIC}}$ and subsets of these flights. Dotted lines indicate those lengths in time and those percentages that were tested. 0 stands for a passing the Kolmogorov-Smirnov test, 1 for not passing.

5 Aircraft tropopause pressure bias

By calculating R$_{\text{rel}}$ using MOD$_{\text{CARIBIC}}$ and MOD$_{\text{RANDLOC}}$, an important fact can be illustrated about data collected with instruments on civil aircraft. If data is resorted into heights relative to the tropopause (HrelTP), it still contains data taken at constant pressure altitudes in a limited range. Depending on the pressure at which the data was sampled, it contains infor-
5   mation from different meteorological situations. The height of the tropopause relative to the sample pressure determines the range of values. The effect can be illustrated by calculating R$_{\text{rel}}$ (see Equation 4) using MOD$_{\text{RANDLOC}}$ and MOD$_{\text{CARIBIC}}$ within $10\,\text{hPa} < p < 500\,\text{hPa}$.

Figure 6 shows the results (right hand panel). For comparison, the left hand panel of Figure 6 shows R$_{\text{rel}}$ of the same datasets when setting $180\,\text{hPa} < p < 280\,\text{hPa}$, the range at which CARIBIC measures. On the right, the representativeness
10   uncertainty increases strongly in all heights except just above the tropopause, where MOD$_{\text{CARIBIC}}$ contains most data. Only the long lived species CO_2, N_2O and CH_4 retain their low uncertainties. For the more variable species to the right of the figure, the representativeness uncertainty increases strongly, especially in the troposphere, where the variability increases.

The strong increase in representativeness uncertainty is due to the bias always present in measurement data from commercial aircraft, which can only collect data high above the tropopause when the tropopause is at high pressure and far below when it
15   is at low pressure values. This bias is naturally contained in all data measured at constant pressure and then sorted relative to the tropopause and should be kept in mind when examining climatologies from corresponding platforms.
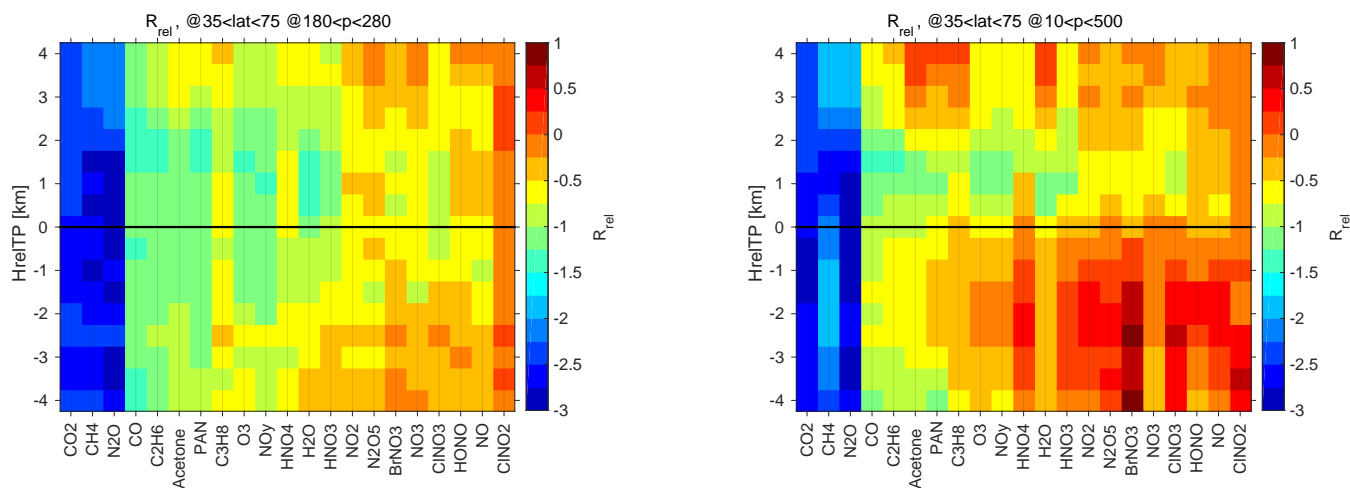
Figure 6. R_{rel} calculated using $MOD_{CARIBIC}$ and $MOD_{RANDLOC}$ with the range of p set to $180\,\text{hPa} < p < 280\,\text{hPa}$ (left) and $10\,\text{hPa} < p < 500\,\text{hPa}$ (right).

Acknowledgements. We thank all the members of the IAGOS-CARIBIC team, especially those who operate the CARIBIC container and Peter van Velthoven of KNMI who provides meteorological support. The collaboration with Lufthansa and Lufthansa Technik and the financial support from the German Ministry for Education and Science (grant 01LK1223C) are gratefully acknowledged. The data analyzed in this paper can be accessed by signing the CARIBIC data protocol to be downloaded at http://www.caribic-atmospheric.com/.

5   This work was partially performed on the computational resource bwUniCluster funded by the Ministry of Science, Research and Arts and the Universities of the State of Baden-Württemberg, Germany, within the framework program bwHPC.

References

Kunz, A., Schiller, C., Rohrer, F., Smit, H., Nedelec, P., and Spelten, N.: Statistical analysis of water vapour and ozone in the UT/LS observed during SPURT and MOZAIC, Atmospheric Chemistry and Physics, 8, 6603–6615, 2008.

Rohrer, F. and Berresheim, H.: Strong correlation between levels of tropospheric hydroxyl radicals and solar ultraviolet radiation, Nature, 442, 184–187, 2006.