Atmos. Chem. Phys. Discuss., 11, C15569–C15576, 2012
www.atmos-chem-phys-discuss.net/11/C15569/2012/

Atmospheric Chemistry and Physics Discussions

Interactive comment on "Impact of sampling frequency in the analysis of tropospheric ozone observations" *by* M. Saunois et al.

M. Saunois et al.

marielle.saunois@lsce.ipsl.fr

Received and published: 22 February 2012

Response to Jennifer Logan:

We thank Jennifer Logan for her comments, which have helped us clarifying the methodology used and improving the manuscript. Below are the responses to her comments that have been quoted [. . .] before each response.

[[. . .] The paper states: "In order to mimic the regular sampling of the soundings, we subsampled the MOZAIC morning dataset using a "regular" sampling method". Indeed, the sondes are flown regularly: every Wednesday at Lindenberg and Prague; Monday, Wednesday and Friday at Hohenpeissenberg, Payerne, and Uccle; and Tuesday and Thursday at De Bilt (to complement nearby Uccle). However, the authors did not sam-

ple the MOZAIC profiles on the dates that the sondes are flown. If they had, they could have quantified the difference that sampling can make with these frequencies, by matching dates, computing monthly and seasonal means, trends etc. They should do this. I suggested it to the lead author when she presented an early version of this work at the AGU meeting in December 2010.]

We perfectly remember this suggestion. Even though not mentioned in the manuscript, we did subsample on the day of launch (Monday, Wednesday and Friday). The maximum difference observed between this subsample and the overall mean reaches up to 2-12% between 500 and 800 hPa. However, it is clear that if we subsample the MOZAIC profiles on the dates of the sondes, we will build only a few different subsamples (3 to 7), which is very poor to quantify the influence of lower sampling frequency on data analysis. The aim of this paper is not to compare nor to reconcile the MOZAIC and sounding data sets but to evaluate the impact of lower sampling frequency on observed ozone variabilities and trends. This justifies the method used in this manuscript. Moreover, sampling MOZAIC on matching dates with sondes assume that the same air masses would be sampled at the two locations. Here we consider weekly sampling (beginning any day of the week) or thrice weekly sampling (i.e every 2-3 days), which is similar to the sondes frequencies and allows us to construct more samples. The methodology has been better explained in the revised version to avoid possible misunderstandings.

[In this paper the MOZAIC data are sampled in an artificial manner, taking every 5th profile. For the months where 2 aircraft were flying in and out of Frankfurt, yielding 4 profiles a day, sampling every 5th profile would sample only a short part of the month (and often consecutive days), quite unlike the weekly sampling by the sondes. Such a high frequency of sampling will likely yields samples that are auto-correlated, rather than independent of each other. With one aircraft a day going to and from Frankfurt, every 5th profile would be every second or third day. The authors seem to have chosen their sampling to build up statistics, but they do not match those in the actual world of existing data. This is unfortunate, and compromises the potential utility of this work. It would have made more sense to sample every 7th day to mimic weekly sampling, and on days 1, 3, 5, 8, 10, 12, etc (2, 4, 6, 9, 11, 13, etc) to mimic those with thrice weekly sampling.]

Unfortunately, there has been a misinterpretation of the text and therefore of the subsampling method. In the manuscript, the explanation of the subsampling is based on an example: a theoretical month documented with 24 profiles (which is rare considering how Frankfurt is highly documented (see Fig. 1). With this theoretical number of profiles, we stated that only 5 subsamples could be created (each one including 4 profiles). For this particular example, every 5th profile is taken. Always sampling every 5th profile would have lead to different sampling frequency, since this would have been depend on the number of profiles available for each month, while, it is clearly stated in the manuscript that we have chosen defined frequencies (weekly and thrice weekly) in agreement to the reality of the soundings. If we consider a month documented with 60 profiles (twice a day, which is closer to the reality of Frankfurt) and we want subsamples of 4 profiles, we can create 60/4 = 15 subsamples (although we took only the first 10) by taking every 15th profile. Considering there were two profiles a day, the first subsample corresponds to one profile of day 1, one of day 8, one of day 16 and one of day 23, i.e one profile a week, as stated in the text. If we want subsamples of 12 profiles using the same month, then we create 60/12 = 5 subsamples by taking every 5th profile over the month. As a result, the first subsample corresponds to one profile of days 1, 3, 6, 8, 11, 13, 16, 18, 21, 23, 26, 28; The second subsample corresponds to one profiles of days 1, 4, 6, 9, 11, 14, 16, 19, 21, 24, 26, 29; The third subsample corresponds to one profiles of days 2, 4, 7, 9, 12, 14, 17, 19, 22, 24, 27, 29; and so on. In this example, we take a profile every 2 or 3 days, which is close to the reality of the thrice-weekly sampling. As a consequence the judgment made on the chosen methodology is unfair and based on a misinterpretation of the text. The frequencies chosen in our study are thus similar to the soundings. The text has been modified in order to avoid such misinterpretation in the future and we use the above example in

the new version of the text, as a consequence Fig 3 has changed.

[The authors use their many sub-samples to discuss “intra-seasonal variability”, but this is artificial, as it does not recognize the temporal sampling by the sondes, which is not biased towards one part of the month. Thus their sampling does not represent a realistic measure of the statistics they seek.]

Following the previous response, we believe that this comment is based on the same misinterpretation. The so-called “regular” sampling avoids taking consecutive days, and therefore does not bias towards one part of the month.

[The authors present seasonal trends for 200 of their sub-samples, and compare them to trends at 6 European sonde stations. They state that: “our study suggests that apparent discrepancies between stations may be attributed to the low sampling frequency, in addition to specific conditions at each station”. (p. 27120.) They should address the actual effect on trends of the sampling by the sonde stations, by taking one aircraft profile on each date that sonde data are available, and omitting sonde data on days there were no aircraft profiles. This of course presupposes that there is no real geophysical variability in ozone between the location of the aircraft and sonde profiles.]

Indeed, the suggested study presupposes that the air masses sampled at the location of the aircraft and sonde profiles are the same at that same day, which might not be realistic. The aim of our paper is not to reconcile the different data sets, in regards to trends, but to assess if there is, and how large, an impact of low frequency sampling on observed seasonal means, inter-annual variability and trends This cannot be discussed using the suggestions of Jennifer Logan but requires the methodology we developed here.

[There are biases between the Brewer Mast sonde and MOZAIC data, particularly in the early years of the MOZAIC record, and these are primarily what give rise to different trends for 1995-2008 between the Frankfurt aircraft data, and the Hohenpeissenberg and Payerne sondes, rather than the sampling frequency. For the period without such biases, 1998-2008, the trends for these two sonde stations, the MOZAIC data, and alpine sites in Europe are very similar, even though the sampling frequencies differ. The trends for Uccle are rather different because of a high bias late in the record. I presented preliminary results of this work at the Second International Workshop on Tropospheric Ozone Changes in Toulouse, France, in April 2011, a workshop attended by several of the authors of this paper. This work has since been completed and submitted for publication (Logan et al., 2011). As noted in our paper, sampling the Frankfurt data on the dates of the Hohenpeissenberg sondes does not remove the bias in the early years of the MOZAIC record.]

We thank Jennifer Logan for pointing this out. After reading the mentioned manuscript, and following your suggestion, we re-compute the linear trends over the period 1998-2008, in order to avoid the first year discrepancy. In the revised manuscript, our discussion is based on this period and we dedicated some time to discuss the result in the light of Logan et al, (2011) results. The discrepancies seen previously in winter disappeared and we find the same agreement between MOZAIC and some sonde sites (e.g. Hohenpeissenberg). However our study shows that different subsamples could lead to different trends over this period, which encompasses the trends derived from the sondes. As a result, we suggest that the sampling could bias the observed trends, however geophysical differences cannot be ruled out.

[I would agree that weekly sampling by the sondes is not optimal for obtaining reliable trends in tropospheric ozone. However, when the trend is large enough, as in the post-Pinatubo period in the 1990s, these trends are readily apparent (Tarasick et al., 2005; Kivi et al., 2007).]

We thank again Jennifer Logan for pointing this out. If the trend is large enough, then the low frequency measurements will be able to capture this trend. Our study is based on the MOZAIC data set, which measurements are available between 1994 and 2009. Unfortunately, the ozone trend over this period is not large, if any. We agree that the period of our study is not the best candidate for such assessment but this is the only

available period. Based on that period our study suggests that the observed trend over the past decade, which is weaker than during the post-Pinatubo period, could be largely influenced by the frequency of sampling.

[Weekly sampling is not an impediment to using many years of sonde data to form climatology, as shown by Logan (1999). The standard error of the monthly means is <7.5% in the lower troposphere for 20 weekly observations in the lower troposphere for extra-tropical stations, and <15% in the tropics. These errors decrease with more observations (Logan, 1999). Such climatologies are useful for evaluating chemistry transport models driven by meteorological products from general circulation models (e.g., Horowitz, 2006; Considine et al., 2008). For evaluation of inter annual variability, a model can be sampled on the days of the soundings, rather than comparing a model monthly mean to a mean formed by 4 or 12 soundings. Of course more frequent measurements are ideal for assessing trends in tropospheric ozone, but that does not preclude careful use of the extant data. Of more concern is the quality of the sonde data, which has had various problems in the past (e.g., Jeannet et al., 2007; Logan et al., 2011). More studies are needed of the consistency of various ozone records in different parts of the world, one of the conclusions of the Workshop in Toulouse.]

We agree here with Jennifer Logan. Our study does not aim to lower the usefulness of the sondes. The sondes provide the longest records of ozone throughout the troposphere and lower stratosphere. That is why it is important to know how reliable they are despite their low frequency sampling, and one could justifiably wonder about the impact of low sampling frequency on the interpretation of ozone distribution and variations. This is what we aim to do in this study. We agree that the best way for a modeler to discuss the IAV simulated in his model, compared to the observed one, is to sample the model on the observation days. However the question of sampling should be taken into account when discussing the IAV based on observations. Our study aims to assess a potential impact of sampling. We believe that our approach is pretty novel, thanks to the MOZAIC program, and as a consequence should be considered for publication and dissemination among the scientific community.

[I could make many comments on this paper, but will restrict myself to a few. The first review has noted the problems with the confusing metrics. As regards the statistics, the authors need to clarify the difference between the standard deviation and the standard error, a term they do not use, but calculate (p. 27116), sigma/(square root (N-1)).]

Following Referee #1's suggestions, we added a table and an appendix to improve the introduction of the metrics used and to clarify the vocabulary used. Also a longer part of the text is dedicated to the presentation of the metrics (Section 3 in the revised manuscript).

[p. 27112. "We do not correct profiles based on the corrections factor provided. The correction factor was scaled to the entire column" It is unclear what this means. Some stations provide data with the integrated profile scaled to independent data for the overhead ozone column (by the so-called correction factor, CF) some do not. Did the authors divide by the CF for the stations that provided the scaled data? If they did, this would be a problem for Payerne, where the mean CF changes from about 1.1 to about 1.0 when the sonde type changed in 2002. Or are they trying to say that they did not use the CFs as a filter for data quality, as is commonly done? The correction factor is not scaled to the entire column, but rather applied to the entire profile.]

In our study we used the same ozone profiles as in Tilmes et al., 2011. At most of the stations, these profiles include already the corrections performed by the data centers. In addition, a column ozone filter is applied to all ozone profiles to reject single profiles with column ozone values of more than 700 DU or of less than 50 DU. In this way, we also filter out unrealistic values of ozone profiles (in partial pressure) at the stratospheric maximum. For the data used here, ignoring profiles corrected by factors outside the range of 0.8 and 1.2, has only a small impact on the averaged profile between 1995 and 2009 (see Figure S1 of Tilmes et al., 2011). This part has been modified in the revised manuscript. The only use of the sondes is made in Section 5

when discussing the effect of low frequency sampling on trends. This comparison is made to highlight that the trends derived from our subsamples are generally similar to those from the sondes.

[p. 27121. The frequency of MOZAIC profiles at Osaka and Tokyo is mostly about 8-18 per month (or about 4-9 days of data), so I do not think these locations are suitable to test the weekly sampling of the Japanese sondes (apart from the fact this should be done with matched dates). There are large latitudinal gradients in ozone over Japan in summer and autumn (Logan, 1999), so the variability sampled by the aircraft will depend on their routes into and out of the airports. I doubt that the high variability is induced primarily by biomass burning, but rather by the dynamics of the summer monsoon which leads to the summer minimum, particularly over southern Japan.]

We thank Jennifer Logan for pointing out this discussion for the Japanese sites. We mentioned the variability of the dynamics as a reason for higher variability seen in Japan compared to Europe or North America. In the revised version we put more weight to the dynamics issues than to the biomass burning emissions. Also some words of caution have been added in the revised manuscript about the small data set available for the Japanese sites. However Boston has even fewer profiles available and no such difference is observed.

Interactive comment on Atmos. Chem. Phys. Discuss., 11, 27107, 2011.