

## Response to the Anonymous Referee #2:

We acknowledge the referee for his useful corrections and suggestions on the study, which have helped us clarify and improve the manuscript. Below are the responses to his comments that have been quoted [...] before each response

*[(1) The analysis does not consider the autocorrelation timescale. This timescale will likely depend on level and time of year. It is likely that weekly ozonesondes produce four statistically independent profiles every month. However, I am rather doubtful that the sampling frequency of 4/month prescribed in this paper also produces independent profiles. As a concrete example, if the autocorrelation timescale is on the order of a week then weekly ozonesonde sampling should produce a reasonable result; the sampling prescribed in this paper will not. Jennifer Logan also brought up this point. This has to be addressed before I can recommend publication. The autocorrelation timescale should be evaluated as a guide as how to sample the MOZAIC data consistently with the ozonesonde sampling. Implementing the above change in the prescribed sampling will likely considerably reduce the number of samples taken from the MOZAIC measurements and may require significantly different analysis throughout the paper.]*

This is a misinterpretation of the text. As we explained previously in our response to Jennifer Logan we actually do weekly and thrice weekly subsampling, as done for the sondes. Our study mimics sonde sampling and then discusses the influence of sampling on the observed seasonal means and variations. The sampling method is such as we avoid picking consecutive days. This has been better explained in the revised version. We agree that it would be interesting to study the correlation timescale of ozone profiles in order to find the best sampling frequency, however it is beyond the scope of the paper.

*[(2) While I am not a statistician, bias in the MOZAIC measurements should also be considered in the analysis. This is particularly true when sampling MOZAIC measurements at less frequented sites. It seems that a more general, valid and interesting approach would be to construct a theoretical distribution of ozone measurements (most likely log-normal) consistent with the MOZAIC measurements at each level and season. Then the question becomes: how many monthly samples (with a given auto-correlation timescale) does one have to take from this distribution to get within a certain accuracy of the distribution mean OR given a number of monthly samples and a given autocorrelation timescale, how closely can one approximate the mean value. Please justify why this approach is not taken.]*

To avoid bias due to low sampling frequency, we have first considered the data set of Frankfurt profiles. This first part allowed us to draw some conclusions, which are then supported when considering less documented airport in the Northern mid latitudes.

Our study is based on the assumption that MOZAIC provides enough measurements so that we have high confidence in the observed MOZAIC means. We have chosen here to use the real data set of profiles in order to assess how different could be a seasonal mean if we consider different subsamples with a sampling frequency similar to those of the sondes. This study, to our knowledge, is the first of the kind to look at the potential impact of sampling. We indeed hope that further work will be done to generalize our findings but this is beyond the scope of this paper.

*[(3) I would agree with the first reviewer that the writing could be improved. The 1st reviewer gives some excellent suggestions for clarification of the terminology. However, I found it was also difficult to follow what the authors were trying to do and where they were going. For example, the analysis*

*methodology is never clearly stated: the paper jumps into “Subsampling methodology”, but never gives an overall outlook as to how this methodology will be utilized in analyzing the measurements. Another example is the fact that the authors really never explain that they are interested in the impact of sampling methodology on ozone trends until section 3.5. In summary, the text is not really “reader-friendly”. It would be helpful if the authors could step back and explain their overall analysis methodology and questions addressed.]*

We thank the reviewer for this suggestion to improve the manuscript. Following the 1<sup>st</sup> reviewer clarification on the terminology has been made. Also in the revised manuscript, we dedicated the Section 3 to the methodology use, which will facilitate the reading of the text. In the introduction, the objectives of the manuscript have also been clarified.

*[(4) The results presented in this paper are particularly useful as a means to evaluate the significance of measured signals. They are less helpful as a means to evaluate models. Ideally modeled ozone profiles should be instantaneously compared against the measured profiles. When comparing instantaneous profiles it does not matter if one really compares in the upper/middle or lower troposphere. However, a significant result that might be taken away from this paper (I write “might” because I have doubts about the analysis methodology – see points (1) and (2) above) is that it is not sufficient to evaluate monthly-modeled profiles against monthly measurements. Or to rephrase: The model output data need to be put on a format comparable with the measurement data (this is the converse of the statement made by the authors on 27110, lines 1-4).]*

We thank the referee for pointing this out. For sure the best way to compare model to observations is to have output at the same dates and interpolate. However this approach is often impractical, especially in the framework of international projects involving an ensemble of models. The topic how observations can be used best to rigorously evaluate numerical models is clearly the next step in the analysis, which should be covered in a subsequent study. We have reworded the paragraph in the introduction and add a word in our discussion on the effect of sampling on the seasonal means (Sect.4.1)

*[(5) Throughout this paper the authors ascribe physical reasons for measured trends and variability (e.g., biomass burning plumes, stratospheric intrusions, changes in emissions) (e.g., see 27115, l. 14; 27116, l. 17; 27112, l. 16; 27124 l. 3; and other locations). While the author’s may be correct in these ascriptions they often give rather vague reasons for their conjectures, suggesting they have not done the analysis necessary to back these conjectures up. Please give references justifying these claims, describe in more detail the model analysis supporting these conjectures or word the statements so it is clear that they are indeed conjectures.]*

Relevant literature has been added in the revised manuscript. Also sentences have been reworded so it is clear when it is conjecture.

*[(6) Nowhere is there a justification in the paper for comparing the specified ozonesonde stations and surface stations with MOZAIC. This comparison is only valid if the same air mass is sampled. While this may be true for the ozonesonde measurements (although it remains to be shown), it seems less likely to be true at the surface sites as these are governed by more local conditions.]*

Actually there is no aim to compare sonde or surface measurements to MOZAIC. However if we assume that different close sites (sondes and MOZAIC) are generally influenced by the same air

masses (which is a reasonable assumption in the free troposphere, at least many studies considered it is, see for example Logan et al., (2011)), then we will expect similar ozone variations including the long-term changes, which indeed agree at least in some cases.

es. Regarding surface measurements, the local events will definitively affect the ozone levels differently from one place to another. As a results looking at the trends derived from surface sites give insight on the range of short term variations in Europe, and of the differences due to local effect. Of course to compare different data set near the surface, a thorough study is needed but far beyond the scope of the paper.

### **Minor Comments.**

*[(1) 27108, l. 10: “uncertainty” – uncertainty in what? The monthly mean?]*

Yes, uncertainty on the monthly means. This has been corrected here and everywhere else in the paper.

*[(2) 27109, l. 20-21: “limited” – please explain what you mean by limited.]*

We meant that ozone sondes were the only sources of regular measurement in the free troposphere. And since 1994, MOZAIC provides also such regular measurements (not dedicated campaign).

*[(3) 27112, ¶ l. 5. Use of correction factor. Why was the correction factor not used to sort out some of the questionable sonde profiles? This has been the standard procedure following analysis by Jennifer Logan. The authors state: “the applied correction factor . . . is small for most of the stations considered”, implying that for some stations this factor is not small. Please justify the analysis procedure used.]*

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.

*[(4) The sonde data and MOZAIC data is used at the surface. Please discuss some of the problems with this: (i) the interference of SO<sub>2</sub> with the measurements; (ii) the fact that low altitude MOZAIC measurements are likely to be biased due to the vicinity of airports and thus may not representative of a larger area.]*

We thank the referee for pointing this out. We added some discussion on the fact that the vicinity of the airport may bias the metrics derived for the lowest levels.

*[(5) It is not clear to me why the data is limited to morning MOZAIC profiles. Certainly the surface measurements better characterize surface ozone variability than MOZAIC. Why aren't the surface measurements used to characterize surface variability? Wouldn't this allow all MOZAIC profiles to be used?]*

Part of the text has been corrected to clarify this point (see Section3). We kept only the morning profiles to avoid the effect of the diurnal cycle in the lowest levels and to keep the same time window as the ozone sonde launches. This limit to the morning profiles is considered in Section

4 (revised version); for the generalisation to other northern hemisphere sites and the case of Windhoek, we use all the available profiles. Of course, the surface measurements would better characterize surface ozone variability than MOZAIC. The objective is not to characterize the surface variability but to assess the effect of sampling; the latter could be done with either one. Indeed we applied the same subsampling method to a surface site and found similar results.

*[(6) 27119, l. 19-20: Another interpretation is that neither data set gives a statistically significant trend (except for the winter MOZAIC measurements). I'm not it is very meaningful to claim two datasets agree with each other if neither dataset indicates a significant result.]*

Our point is neither to reconcile data sets nor to determine a trend over this time period. We suggest here that some subsamples of MOZAIC profiles can lead to trends similar to those derived from the sondes (even though not significant).

*[(7) 27120, l. 2, "It is worth noting. ..." I'm not sure what the authors are trying to say here.]*

We meant that over this specific time period the short term variation of ozone is weak, leading to statically insignificant trends (even with MOZAIC). As a result, a slight change on seasonal means (and maybe sampling) could lead to different trends in sign and/or magnitude. This part has been reworded.

*[(8) 27122, l. 8-9. Please justify this statement.]*

This sentence has been reworded to: The results over this three-year period should be representative of the ozone variability in this region.