Response to all Reviewers and Jennifer Logan:

We would like to thank the two reviews and Jennifer Logan for helpful comments and suggestions that helped to significantly improve the manuscript and the scientific value of the study. We have addressed all specific and general comments by the reviewers and Jennifer Logan and believe that the version of the paper is now acceptable for publication in ACP.

The paper was significantly revised and changes to the text, structure, figures, and to the presentation of material were performed (as outlined below). Therefore, comments with regard to writing expression are often not addressed one-by-one. All suggestions applying to the material from the first version that are still relevant were incorporated into the revised manuscript (see detailed comments below).

In the revised version of the manuscript, besides providing a new climatology, the paper focuses on ozone proberbility distribution functions (PDFs) and includes a detailed discussion on the similarity of ozone PDFs from single stations with others in similarly classified regions. Also differences in ozone PDFs among different regions are discussed. In this way, regional classifications are well justified and shown to be suited for model evaluation purposes.

An application of the climatology is now given for two CAMChem model simulations. We agree with reviewer 2 and Jennifer Logan that an evaluation of two model intercomparison projects needs a much more in-depth analysis than could be performed in one paper. To demonstrate an application of the climatology, we use only two model simulations performed with the same physics and chemistry, but with well-defined differences in the setup. One simulation is performed with offline meteorological fields and the other run with online calculation of winds.

Further, previous studies that are relevant to this study are now recognized in the introduction. The state of science and literature is now reflected in the introduction and throughout the paper. Well-known scientific findings are not repeated and only mentioned if relevant to the new findings.

Summary of changes in the revised version of the paper compared to the earlier version:

- In addition to providing a new climatology for single stations with additional statistics, in the revised version of the manuscript we also focus on the median and the shape of ozone distributions and their similarity in defined regions, which was not performed in previous studies.
- Some regions have been redefined to include only those stations that describe similar ozone characteristics, as discussed in detail in Section 3.
- The Tropics are now separated into three regions, adopting the defined regions by Thompson et al., 2012, in review.

- To consider ozone distributions in comparison to model output or other data, we further provide a dataset for each season and region including all observations at defined pressure levels that were measured between 1995 and 2009.
- In the revised version, we add the comparison of PDFs between ozonesonde measurements and aircraft and surface observations.
- Figures 4 and 5 of the previous manuscript were removed. Instead we compare ozone PDFs and median values over Western Europe, Eastern US and NH Polar East only.
- To keep the paper focused, we chose to not discuss the time evolution of ozonesonde observations. The discussion of ozone changes between 1990-1999 and 2000-09 and Figures 6,7, and 8 are removed.
- Figures 2 and 3 of the revised manuscript are somewhat similar to Figures 8 and 9 in the first version. However, now we focus on the variability in each region and show the Hellinger distance and the spread of the median of ozone PDFs among the different stations. Information about half-width and interannual variability is moved to the Supplement.
- We added a new Figure 4, describing the number of observations per season and year for each measurement site, sorted by regions.
- New Figures: 5,6,7,9, 10, are added, describing ozone PDFs for different regions and altitude intervals.
- In the revised version of the manuscript, we present an application of the new climatology using model simulations from the NCAR CAMChem model.

We address additional specific comments of the two reviewers and Jennifer Logan separately, as listed below:

Response to specific comments that were not addressed above

Response to Reviewer 1 (specific comments).

Page 28748, line 2: The first line of the abstract does not make it clear that this climatology is constructed from ozonesondes only. It refers to 'ozone soundings' but this is ambiguous. Ozone soundings does not exclude measurements from satellite e.g. the Microwave Limb Sounder (MLS).'
Ozone soundings is changed to ozonesonde measurements

You refer here to the compilation of ozone profiles from 1980-1994
but the title of the paper suggests that you are only considering a climatology from 1995 to 2009. Why are the data from 1980 to 1994 also needed?

The sentence has changed and 1980-1994 is removed. However, we also compiled a climatology for the period 1980-94. On the one hand, we want to compare the new climatology with the well established one by Logan et al, 1999. On the other hand, we have added more statistics for both periods.

Page 28748, line 10: By 'compare the variability of ozone distributions within each

region' do you mean compare the inter-station differences in each region or do you mean compare the ozone variability between regions?
We revised the sentence to:
'The Hellinger distance is introduced as a new diagnostic to identify stations that describe a similar shape of ozone distributions and therefore can be grouped into one region.'

Page 28749, line 3: I am surprised that you pick out 'fossil fuel combustion, industrial processes, and biomass burning' as the primary anthropogenic activities that affect ozone. I would have had CFCs and halons at the top of my list, followed by N2O, and then GHGs in general. Or did you mean only tropospheric ozone? If you meant only tropospheric ozone you should say so.
We restructured the Introduction and in introducing background information needed for this paper; stratospheric ozone is not the focus here.

Further detailed comments by the reviewer on the introduction are not specifically addressed since the text changed significantly, which we believe has resolved the problems mentioned by the reviewer.

Page 28752, line 2: You mention the temporal resolution for the validation data from the WDCGG database but say nothing about the temporal resolution of the other validation databases. Why?
The description of all measurements used in this study, are summarized in Section 2 now.

Table 1: This would make far more sense if you listed the number of ozonesonde flights per year rather than the total in the period. Otherwise the values are just not comparable. Some sites appear to have made more flights than others but that's just because they ran over a longer period. I think that the caption should also point out that values are omitted when the station makes less than 12 profiles per season for at least 5 continuous years.
This information is given in Figure S1, in the Supplement, and it was included in that section of the previous version of the manuscript. We have added a Figure 4, showing the time evolution of the number of measurements per station/season and region.

Figure 1 caption: You say that the subtropical stations are shown in black and also that 'Those stations that are not included in selected regions are shown in black'. I don't know how to interpret that. Does that mean that all sub-tropical stations are not in a selected region?
All stations that are not included into defined regions are now shown in grey.

Page 28753, line 10:
±2–3% is ambiguous and is not the correct way to report precision values. Do you mean 2-3% or do you mean 2.5 ±0.5%? The same comment applies elsewhere in this paragraph.
We have revised this paragraph. Earlier studies have done a significant work on the evaluation of different ozonesonde types. Here, we refer to earlier studies and give numbers of precision and accuracy consistent with those.

Page 28753, line 17: What does it mean to 'employ' an ozonesonde? I also don't know how one 'corrects' a station. I'm sorry, but I just have no idea what this sentence is trying to say.
We revised this paragraph. See next comment.

Page 28753, line 22: You refer here to 'factors outside the range of 0.8 and 1.2'. What are these 'factors'? This is the first time they have been mentioned and no description of them is given anywhere. You only say that profiles are ignored if these 'factors' are outside of the range 0.8 to 1.2. If the reader has no clue as to what these factors are, how can this statement make any sense?
We revised this paragraph:
'Ozone profiles provided from the data centers are often scaled to ground-based ozone column measurements, which are strongly influenced by their stratospheric portion. If the scaling factor, called "correction factor", is outside the range of 0.8-1.2, ozone profiles in the troposphere might be biased. Here, we consider all profiles as provided by the data centers, without any additional filtering with regard to correction factors. Ignoring profiles that are corrected by factors outside the range of 0.8-1.2 has only a small impact on the averaged profile between 1995 and 2009 (as demonstrated in the supplement, Figure S2).'

Page 28754, line 1: But how do you interpolate from an altitude grid in meters to a pressure grid in hPa? Somewhere you must be making a coordinate conversion? And surely if you are doing the interpolation in pressure coordinates then you are doing it in ln(P)?
We have clarified this in rewriting this paragraph:

`Each high-resolution ozonesonde profile is converted from geometric height to pressure level using the hydrostatic equation. Ozone data within a layer around each of 26 pre-defined pressure levels was then linearly averaged. Between 1000 hPa and 100 hPa, a layer thickness of 25 hPa centered on the selected pressure level was used. Between 100 hPa and 10 hPa a layer thickness of 2.5 hPa and above 10 hPa a layer thickness of 0.25 hPa was chosen….'

Page 28754, lines 1-4: If you did what you said you did i.e. averaged all available ozone observations between 1980 and 1994, and between 1995 and 2009 at all pressure levels, there is no way that this produces seasonal ozone profiles.
We have clarified this in rewriting this paragraph.

Page 28754, lines 6-7: Can you put a number on 'good agreement' otherwise this is a meaningless, subjective statement. Surely you can quote one number that is indicative of the differences between the two climatologies?
We added 'within +/- 5%'.

Page 28754, line 11: But that deviation could simply be the result of interannual variability i.e. we don't expect seasonal means to be the same from year to year. Or, again,

I don't understand what you are saying.
Page 28754, lines 12-14: But how will knowing the number of profiles that were used in each 15 year average help the modelers using your climatology? If they knew the exact dates and times of the profiles, and where they were located, they could sample the models to match those. But I don't think that that is what you are suggesting. Surely it would make more sense if the uncertainties on your derived climatologies included any uncertainty resulting from temporal and spatial sampling bias in each region and season?
Page 28754, line 16: You say 'sufficient ozone soundings per season' but sufficient in what regard? What is your criterion for sufficiency?
We agree that this paragraph was problematic and have removed it.

Page 28754, line 20: You say 'obtain a sufficient sample size for tests of significance' but tests of significance with regard to what? An averaged ozone profile cannot be flagged as being significant or not significant. It can be flagged as being statistically significantly different from the true profile, or statistically significantly different from the mean profile from another region. What is the benchmark against which you are testing for 'significance'?
The discussion about regions is moved to section 3 and was reworded to clarify the significance.

Page 28755, lines 8-9: You say 'Further, gravity wave activity is most prevalent over the Pacific and Indian Oceans' but it is not clear at all how this connects to the discussion of the variability of tropospheric ozone in the tropics. Nowhere have you discussed (with citations of the relevant literature) exactly how gravity wave activity affects the distribution of ozone in the troposphere.
Thompson et al. (2011) have discussed gravity and Rossby wave signatures in the tropical data used in this study. Also, data from a 2007 campaign over Costa Rica and Panama data, revealed waves that were inferred in lower stratospheric ozone, as discussed by Selkirk et al (JGR, 2010) and Thompson et al. (2010). The discussion about tropospheric ozone in the tropics is moved to Section 3.3.

Page 28755, lines 9-12: I am confused here. Surely the fact that the vertical distribution of ozone and its year-to-year variability over a 10 year period shows significant differences between stations in the tropics is a compelling reason to regionally separate the tropical stations, not a reason to combine them? You would be more justified in combining stations within a region if they showed ozone behavior that is homogeneous in space and time. So the reason that you provide for not regionally separating the tropical stations is, I believe, the very reason why you should. In fact a few lines later you say 'For detailed model evaluation of tropospheric ozone in the tropical region a comparison of single stations is likely to be more meaningful' i.e. the regional heterogeneity of the ozone distribution is a reason NOT to combine all stations in the tropics. Now maybe you don't have enough tropical stations to do this regional separation, but that's a different issue.
We agree with the reviewer and separate the tropical stations into three regions, adopting the regions defined by Thompson et al., 2012 (JGR, in review). These classifications are

based on mean free tropospheric ozone concentrations, tropopause characteristics and an index based on wave activity and associated convection.

Page 28755, line 21: There is one question I had in mind which was not addressed by your brief description here of the Hellinger distance. If two distributions are identical in shape, but do not overlap at all because their medians are very different, is the Hellinger distance 0 or 1? I think that providing some such pedagogical examples would greatly help the reader is gaining an intuitive understanding of what the Hellinger distance represents.

This information has been added: 'The Hellinger distance is a statistical measure for the similarity of two distributions covering values between zero and one, where the H-value is one, if two distributions are completely different, and zero, if two distributions are identical (see Appendix A for more details).'

Page 28756, line 4: It's not clear to me what you mean when you say that the ozone distribution depends on the altitude interval. Do you mean that it depends on the altitude region? Ah, maybe you mean the shape of the PDF and not the shape of the ozone profile itself? Can you please clarify this.

We have changed the paragraph and removed this sentence.

Figure 2: It is not necessary to label ever y panel in the Figure with 'West Europe' since the whole figure is for Western Europe and for Western Europe only. And why are you not showing the ozone PDF around the thermal tropopause i.e. from -1 km to +1 km?
Page 28756, line 22: You refer here to the region within 3 km of the thermal tropopause, but as far as I can tell, you are nowhere considering the region within 1 km of the thermal tropopause.

We have relabeled all the figures and change West Europe to Western Europe. We are not showing ozone PDFs around the thermal TP. The ozone distribution around the TP can strongly vary depending on the precise location of the TP, since ozone describes a strong gradient across the TP. We do not think that model evaluation of ozone distribution in +/- 1km around the TP is very useful, since models cannot precisely calculate the TP, due to their relatively broad vertical resolution. For a more meaningful comparison we consider altitude intervals above and below the TP as provided in this study.

Page 28756, line 25: You refer here to the 'Hellinger distance between different observations' but it is not possible to calculate the Hellinger distance between observations. You can only calculate the Hellinger distance between distributions.

We agree and corrected the wording.

Page 28757, line 6: But MOZAIC doesn't actually collect air samples right? If that is true, then referring to 'aircraft samples from the MOZAIC program' is misleading.

We reworded the entire paragraph on MOZAIC observations in Section 2,

Figure 4: I am confused by the legends in the leftmost panels of this figure. The legend states that soundings at 800 hPa are shown with open squares, soundings at

900 hPa are shown with open triangles, and soundings at 1000 hPa are shown with open diamonds, and yet none of these symbols appear in the top row panel (only filled diamonds) and while the are some open diamonds in the bottom row panel there are also filled diamonds and so I have no idea how to reconcile the information provided in the figure legend with what I actually see in the panels. In the text it says 'Altitude information of surface stations is included in Fig. 4 (right panels, different sizes of diamonds, going from small to large with increasing altitude)'. Sorr y, but at least in my version of Figure 4 I see no variation at all in the size of the symbols. The caption for Figure 4 does not say what the 'r=0.85' and 'r_0=0.99' refer to.

Figures 4 and 5 are removed and replaced by a comparison of ozone distributions for three regions, Eastern US, Western Europe, and NH polar West. Thus, comments that refer to Figures 4 and 5 and the corresponding text no longer apply to the manuscript.

Page 28757, line 28: There was no upper bound on the altitude range for the 3rd altitude interval?
There is not upper bound. Surface observations reach up above 4000m for some stations.

Page 28759, line 16: I don't know what you mean by 'The shape of the ozone distributions'. Do you mean the shape of the ozone PDF, the shape of the ozone profile, the seasonal evolution of the ozone or something else? And whatever it is that you mean, can you please point to a figure where this is shown.
We clarified in the text that we mean: ''The shape of ozone PDFs'. The text points to Figures discussed in the revised version of the manuscript, when applicable.

Page 28759, line 26-29: These sentences confuse a whole lot of things. The temporal variability in surface ozone over China ('large daily variability of ozone') is used to explain the spatial variability in ozone between Japanese stations ('large variability among the stations'). That just makes no sense to me.
We have removed this sentence from the manuscript.

Figure 6 Caption: 'Time evolution of .. from'. And why is the Nor thern Hemisphere Tropics region (incorrectly labeled in the actual figure) now arbitrarily included in the analysis?
To focus the paper, we are no longer discussing the time evolution of ozonesonde observations in the main text any more, nor ozone changes between 1990-1999 and 2000-09. Figures 6,7, and 8 are removed. All comments regarding these figures and the corresponding text do not apply anymore and are addressed point-by-point.

Page 28763, lines 3-5: Surely yours is not the first study to note that there is seasonality in the height of the ozonopause in the tropics with a maximum in DJF and MAM?
Page 28763, lines 20-21: I don't know what you mean by 'The seasonality of tropopause-referenced altitudes'? Do you mean the seasonality of ozone in tropopause-referenced altitudes?

Page 28763, line 28: Again, surely this is not the first paper ever to obser ve that the seasonality in ozone in the lowermost stratosphere depends on the coordinate system considered?
Page 28764, line 11: Just to be clear here, are you talking about trends in the altitude of the thermal tropopause in the subtropics, or are you talking about trends in ozone references to the thermal tropopause? I believe the former but given the earlier opacity of the writing I cannot be sure.
Page 28764, line 14: Surely such speculative statements (and there are many in this paper) could be made more robust, either by extending the analysis to something more than a passive commentar y on the ozone climatology, or to reading and citing existing literature. I know that there are many papers that have documented the hemispheric asymmetr y of the effect of the Brewer-Dobson circulation on ozone.
Page 28764, lines 15-17: This is not the first time that you have made this statement in this paper, and it is cer tainly not the first time it has been obser ved. There is far to much repetition of statements in this paper and the paper would be significantly improved if this repetition was removed and the paper thus shortened.

We have shortened the discussion on ozone seasonality and refer to earlier studies, such as Logan et al., 1999. The text is rewritten to remove the concerns that were mentioned. In the revised version of the manuscript, we present an application of the new climatology using model simulations from the NCAR CAMChem model (see general comments).

All other technical errors are corrected.

Response to Reviewer 2:

General comments:
1) As suggest by the reviewer, we summarized all information relevant to the measurement and processing of the climatology for both single stations and regional aggregates in Section 2. Profiles from all stations within one region are treated equally while performing the regional aggregate, because the sampling frequency is similar for the period 1995-2009.
Uncertainties of ozone measurement of single stations are discussed, referring to earlier performed studies. The variability of regional-aggregates is discussed in detail in Section 4.

2) As discussed in the response to all reviewers, we only focus on two model simulations in the new version of the paper, where daily output is available.

3) We agree with the reviewer and have discussed uncertainties of ozonesonde measurements based on the mentioned studies, refer to Saunois et al., 2011, who discusses the uncertainty of averages and trends in detail, and to a recent study by Logan et al., 2012.

4) We agree with the reviewer and remove all timeseries and the corresponding discussion from the main manuscript. We do give information on the interannual variability of ozone profiles in the supplement.

Specific comments

Abstract

Lines 2-5: The 1980-1994 climatology, i.e. the average over this period, is not discussed in the text, only the timelines are shown in Figs 6 and 7. Therefore, I suggest removing this information from the abstract.

We have removed this information.

Line 14: To my mind, terming the agreement between sondes and other data "excellent" is too positive. See major comment. Please also update the Supplemental Material, Sect. 4.2, Fig. 6.

We have removed this statement in both main text and supplement.

1 Introduction

The introduction is rewritten as outlined in the general response.

Page 28750, line 27: please add DiNunno et al. (2003) (see my reference list at the end of the document)

Page 28751, line 1: please also cite Thouret et al. (2006) and Schnadt Poberaj et al. (2007)

We removed these statements in the introduction and therefore could not include the suggested references.

Page 28755, lines27 to page 28756, line 1: Is there a reason why you compute the Hellinger distance in the UTLS only? Please motivate.

The Hellinger distance was computed for the UTLS and for the troposphere. However, we only showed the results for the UTLS in the previous version of the paper. The new version shows the Hellinger distance for both troposphere and UTLS.

Page 28756, lines 1-3: How do the authors define the tropopause in the high midlatitude and Arctic/Antarctic regions in winter, when identifying the thermal tropopause may not be possible? How are tropopause-referenced ozone profiles obtained under these circumstances?

The tropopause could not be identified for a small percentage of profiles in mid and high latitudes. In those cases, the profile was not included in the climatology.

Page 28756, lines 15-18: The differences are only described. Please indicate potential reasons for the larger spread of median differences over Japan and the US, as well as the large variability of ozone distributions in the lowermost LS.

This is an important point and is addressed in detail in the revised version of the manuscript in Section 3.2. We have also added a new Figure 8 to show differences and we discuss the reasons.

4 Representativeness of regional averages in comparison to independent observations

This section is incorporated into the new Section 4. We have rephrased the text and changed figures and the discussion (see general comments) according to the comments and suggestions of the reviewer. The data sets are now all described in Section 2.

Page 28759, lines 22-24: It is true that the ozonesondes and MOZAIC aircraft data for altitudes between 800 and 400 hPa agree within the variability of both observations over the southeast US and the entire US. However, there seems to be a slight systematic offset with sonde values being higher than MOZAIC values (in agreement of what is seen for sondes/surface measurements). This offset was larger in earlier periods already stated by Thouret et al. (1998) and occurs also in the upper troposphere discussed by Schnadt Poberaj et al. (2009) and may point to systematic differences in the 1990s (not necessarily in the whole period of investigation). This should be mentioned.
We have added this information in the revised version of the manuscript:
'As in the case for the surface observations, ozonesonde measurements are biased high compared to MOZAIC observations over Eastern US, as is also the case for earlier periods (Thouret et al., 1998). Ozone from ozonesonde measurements over US is also biased high at 500 hPa, which might point to some systematic differences between MOZAIC and ozonesonde observations in the 1990s (Schnadt Poberai 2009). '

Page 28760, line 2: "The correlation between ozone soundings and independent observations is smaller ...".: Possibly the lower correlation also has to do with the different sampling frequencies of MOZAIC and ozonesondes. It would be worthwhile to check if the correlation can be increased by just sampling MOZAIC data for those days when ozone soundings were taken.
We are considering ozone distributions and do not discuss the correlation between MOZAIC and ozonesonde data in the revised version of the manuscript.

Page 28760, lines 5-10: I agree that one has to be careful in comparing regional averages of model results to those from observations. This is because many modelers tend to consider individual observational datasets as "reality" and compare their model results to observational datasets in a quantitative way. Hence, I would not restrict this statement to those regions that are under-represented, but would keep it general and state that this is a particular problem for under-represented regions.

We agree with the comment and discuss the spatial variability of surface measurements in comparison to ozonesonde observations. This is especially the case for the surface but less for the free troposphere as stated for example for Western Europe: 'The median difference between surface observations and regional aggregates from ozonesonde data varies between -25 and +25 % at 1000 hPa, and less for higher elevations, as illustrated in Figure 8, left panels.'

6 Vertical profiles for different seasons and regions

General comments on section 6:
The findings from this section are presented in a mostly descriptive manner. In several cases, simple and brief explanations of features of the seasonal cycle and

vertical structure, which are known from the literature, are not given. Indicating these would largely improve the content of this section. In addition, the discussion of the different regions is done in a single paragraph without separation by line breaks or clustering with bullets. This makes the text extremely difficult to read. A much clearer structuring of the text is recommended.

The seasonality of ozone profiles is now briefly discussed in Section 4, focusing on the variability of median values and ozone PDFs between stations within each region. We do show the half-width and interannual variability of profiles in the supplement, as mentioned above. As suggested by the reviewer we have included an explanation about the features or refer to appropriate references. Other individual comments no longer apply due to the restructuring and rewriting of the paper.

Individual comments:
. "A secondary ozone minimum" should be replaced by "An upper tropospheric ozone minimum".
We agree and have changed this.

To me, it looks like that the UT minimum in the NH subtropics (note that in Fig. 9, the graph title is "NH Tropic") also exists to a somewhat lesser extent in SON, while in the tropics, it is visible in DJF and MAM. Please comment.
We have discussed the differences in the UT in both the NH subtropics and the Tropics: 'Median ozone values in the NH Subtropics show a large spread round 300 hPa, with a maximum above 30% in summer. A distinct upper tropospheric ozone minimum occurs in winter, which might be the result of the influence of ozone-poor airmasses transported from the tropical tropopause to the North. Ozone over the Western Pacific and East Indian Ocean region is strongly influenced by deep convection, resulting in a distinct upper tropospheric ozone minimum around 200 hPa …'

Page 28764, lines 3-5 (also page 28765, lines 11/12). The seasonal cycle in the lower stratosphere found over Japan and the United States with second largest values in fall is certainly highly interesting. However, I do not believe that the high values in fall are due to tropospheric-stratospheric exchange processes for the following reasons: 1. the ozone maximum in the lower stratosphere should be in spring, followed by summer, winter, and lowest values in fall (e.g., Schnadt Poberaj et al., 2007, their Fig. 6, MOZAIC data); 2. the seasonal cycle in the upper troposphere, as seen from Fig. 10, neither supports the high LS values in fall, and 3. in the pressure system in the uppermost layers (Fig. 9), the lowest values are found in fall and winter. Possibly, since the regions of interest are situated in the proximity of the subtropical jet stream and associated tropopause breaks, the high fall ozone could result from stratospheric values above the high subtropical tropopause, whereas the lower winter and spring values could result from above the midlatitude tropopause. Please check the seasonal cycle/variability in tropopause altitudes and discuss.
Thanks for the suggestion to look into this. We have investigated this question in detail and conclude that a higher frequency of tropospheric intrusions

(tropospheric/stratospheric exchange) might be responsible for the smaller values in winter and spring in the LMS.

Section 7 Application of the ozone climatology to model studies
This section has changed, so comments do not apply any more.

Page 28768, general comment: Please comment on why the modeled ozone distributions, expressed in terms of Hellinger distance, are so different from the observed ones in the UTLS. To my mind, this should have to do with the altitude-dependence of ozone distributions in the UTLS, as well as model vertical resolutions and problems of the models to place the TP correctly.
Differences between models and observation in the UTLS are interesting and should be discussed in more detail. Since we only want to demonstrate the application of the climatology, but not perform an in depth evaluation of the model runs, we do not add more examples of ozone PDFs. We plan to focus on this part in a follow-up paper.

8 Conclusions
Page 28769, lines 2-4: This result is not surprising. The authors use the same data and only a slightly different averaging period. Please either remove this sentence or motivate properly.
The comparison with the climatology by Logan was performed to support the validity of the new climatology. Also, additional statistics are included in the new climatology period, as was the case for the earlier period.

Page 28769, lines 19-25: Please give brief explanations for the high variability in ozone distributions within the US and Japan region. Similarly, explain why there is large variability below the tropopause, and why distributions are more similar above the tropopause.
This is also discussed in more detail in the revised version of the manuscript.

Page 28769, line 28: "excellent agreement (in both shape and median values)". As mentioned before, there is still a small possibly systematic positive offset of the sonde versus the MOZAIC data due to problems in 1995-1998. Hence please rephrase to something like "agreement within the range of uncertainty". The shape of the MOZAIC data has not been discussed in the manuscript. Please adapt the text accordingly.
We removed this comparison and focus on comparison of median values and ozone PDFs between ozonesonde observations and surface and aircraft data. We have mentioned the offset between MOZAIC and ozonesonde data between 1995-98.

Page 28770, lines 24-25: Probably only true for large regions. In smaller regions like Europe or Japan, averaging model results over the whole region instead of interpolating to the sonde sites and then aggregating, probably does not result in significant differences due to the constraints in horizontal resolution of many global models and numerical diffusion of some.

We added a discussion about the spatial variability based on surface measurements for Western Europe and Eastern US.
Page 28771, line 4: briefly indicate potential reasons for the modeled overestimation of lower stratospheric ozone
We also have removed this part and will discuss this in a different paper.

Appendix A
All the information of Appendix A is now included into Section 2. We corrected all mistakes that were listed by the reviewer.

Appendix B and Supplemental Material
We corrected all errors pointed out by the reviewer.

Supplemental Material
Fig. 1, upper left panel (Resolute (1995-2007): The SON total number (91.7) should be an integer.
The number of available profiles can vary with altitude. We have performed an average over all altitude levels, which can result in a fractional number.
Fig. 1 : "The average of all available profiles are shown in dashed lines ()." It is not necessary to show the dashed lines, because they are hardly visible anyway. The right part showing the percentage difference between the corrected and all profiles is sufficient.
We have left those in, just in case people would like to explore the differences.

Fig. 3: Which colors represent which seasons? Information missing.
We have added the color information.

Jennifer Logan:

Various comments from Jennifer Logan are addressed in the response to all reviewers. More specific comments are addressed below:

We followed the suggestion by Jennifer Logan and focus on one topic of the paper and explore ozone distributions for different regions (as outlined above). New statistics using the Hellinger distance are introduced in this study that have not been brought into this type of scientific application before.

We agree with Jennifer Logan that the grouping was not performed carefully in the earlier version of the paper. Therefore, a detailed analysis is performed in the revised version of the paper to justify the grouping of newly defined regions. In particular we focus on Japan and discuss differences of ozone distributions for different stations. Indeed, Sapporo shows different characteristics from Kagoshima and Tateno, and is not included in this region anymore. Also Naha is more tropical than Kagoshima and Tateno and is not included. So we limit the Japan region to two stations. The TP height over stations in Japan (and other regions) is discussed in connection to differences in ozone. As mentioned above, tropical stations are now separated and grouped into three independent regions.

While exploring differences in ozone distribution between different stations, longitudinal differences in ozone PDFs in mid-latitudes are discussed that were not discussed in the same way in earlier studies. The findings point to the fact that zonal averages as provided in earlier climatologies do not allow model evaluation with regard to the longitudinal variation of ozone distributions in northern mid- and high latitudes. Of course, other papers have pointed to the fact that zonal averages are problematic. We try to offer useful new information about regional aggregates.

As suggested by Jennifer Logan, the comparison between ozonesondes and surface and aircraft observations is also updated in the new version of the paper and the comparison of PDFs is added as well as a discussion about the spatial variability.

In one way, Jennifer Logan is right in mentioning that we have duplicated her earlier work in producing an ozone climatology as part of this paper albeit for a later period. This was done deliberately, to provide a climatology in a similar format, to make it easier for modelers to make use of the new climatology. However, this is only one part of the study. In addition to providing a monthly averaged ozone climatology, including mean, standard deviation, and the number of profiles entering the average, we provide information about median, the half-width of the distribution (calculated as (75th percentile - 25th percentile)/2), and the more information about interannual variability, defined here as the range of the 5th and 95th percentile of the annual median ozone value. We also add information to derive ozone distributions for regional-aggregates. These are new features of the climatology that should be useful to evaluate models, as presented in the paper.