The paper presents a global ozone climatology based on ozonesonde data for the period 1995 to 2009. This climatology, available for evaluation of global chemistry and chemistry-climate models, is principally relevant for the global modeler community and deserves publication in ACP. However, I have serious reservations about the manuscript quality in its current state: In many places, the text is copiously written and individual sections are not well motivated and structured, examples of which are given in the specific comments. There are several major and a vast number of minor issues throughout the manuscript. In sections 6 and 7, I have comments to almost every sentence. Summarizing, I do recommend publication only if the manuscript receives a most careful and thorough revision.

Technical comment

The numbers and descriptions within the figures seem somewhat small and are hardly readable in the draft form of the manuscript. However, since I do not know how much enlarged the figures will be in the final print, I leave the decision to the editor if the figure descriptions need to be enlarged.

Major comments

1) A clear and concise overview of the methodology, i.e. the averaging procedure should be given at some point in the manuscript, e.g. in section 2. It is not always clear from the text if medians or (arithmetic) means are evaluated. It is also not clearly elaborated how the regional aggregates are computed: are temporal averages first computed for the individual stations and are these then aggregated over the regions? The order of the procedure may be important: for example, the stations over Japan are characterized by different seasonal cycles depending on their latitudes. While Sapporo is characterized by a midlatitude seasonal cycle, the more southern stations are more subtropical in nature. A simple aggregate without preceding temporal averaging of individual station data would lead to biasing toward the seasonal cycle of the station with the most data. In the figures, please also always clearly indicate whether medians, means or seasonal arithmetic means of medians are shown (not clear in Figs. 4 and 5). Parts of the confusion concerning the figures may arise from the fact that the relevant information on medians/averages is sometimes hidden in the middle of a lengthy figure caption.
2) The authors use HTAP and CCMVal2 model data as an example of model validation, discussed in section 7 and Figs. 11 and 12. I do see a problem in the way the data are compared: the HTAP model data are stored as monthly means. The HTAP values in Fig. 11 are means and medians of the 22 model monthly means. Similarly, for the CCMVal2 data, climatological mean monthly averages are calculated. In contrast, the sonde climatological values are apparently represented by monthly averaged medians. As the authors stress themselves in Appendix B, tropospheric ozone distributions are often not Gaussian in nature. Hence, comparing modeled means with measurement medians/monthly averages of medians may not always be suitable. For this reason, Figs. 11 and 12 should additionally include the climatological monthly arithmetic mean of the sonde data. In case of HTAP, for which no median information is available, this information should be used to evaluate biases and correlations. For the CCMVal2 data, on the basis of the 10-day instantaneous values, it would also be possible to compute the climatological median values and compare these with observational data.
3) General uncertainties of the ozonesonde measurements must be more thoroughly discussed. I am afraid that modelers will use this climatology as sole data set for model evaluation and assume that it represents the true state without realizing that

the data have (partly considerable) uncertainties and errors, particularly in the 1980s, but also in the 1990s. For example, the Hohenpeissenberg BM sonde behaved differently in the JOSIE 1996 laboratory experiments when compared against a UV photometer (Smit et al., 1998) and in situ in comparison with MOZAIC (also a UV photometer) (e.g., Thouret et al., 1998). As Thouret et al. (1998) shows, the Hohenpeissenberg sonde as well as other sondes of that time showed significantly higher ozone than MOZAIC. A much better agreement between instruments is found for the period since about the year 2000. Hence, while the differences between datasets may only be a few ppb, they may result in different long-term trends when applying different instruments (e.g., Logan et al., 2011, submitted to JGR, http://www.people.fas.harvard.edu/~logan/-jal.papers.html). It is highly relevant to communicate these problems to the modeling community.

4) The manuscript presents an ozone climatology. Additionally, the underlying time series are also presented, as well long-term differences between two climatological averages (section 5). I recommend either removing the complete section or removing the long-term trends and shifting the timelines to the supplement for the following reasons: the tropospheric discussion of the timelines is redundant and can be also discussed using the interannual variability information in Fig. 9. As mentioned above, evaluating long-term trends is rather complex due to measurement uncertainties (e.g., Logan et al., 2011). If the authors wish to publish these results, I recommend doing this in a separate manuscript that includes a thorough analysis of the uncertainties as done in Logan et al.

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.

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.

1 Introduction

Page 28749, lines 27-29: Please substitute to: " The distribution of ozone in the stratosphere is largely controlled …"

Page 28749, line 25 to page 28750, line 5: the whole paragraph is not well-motivated and structured. If the authors try express that climate change may affect stratospheric ozone, then please motivate accordingly early in the paragraph and list the arguments thereafter.

Page 28749, lines 4/5: Why? Can you put it this generally? Please clarify.

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)

Page 28751, line 27: it should be "Hellinger distance"

Page 28751, line 15 to end of section: Too detailed at this point. I recommend to keep these paragraphs short and to give a clear and concise overview on what the

manuscript is about, shifting the detailed information (e.g., info on datasets used for comparison, as well as which percentiles are used) to the respective sections.

2 Selected Ozonesondes and Regions

Page 28752, first paragraph: It would make the text much easier to read if the links concerning the ozonesonde datasets were removed from the text and moved into a table. This table would list all datasets used in the study and should also include references for the datasets. It could also include the links to the datasets used for evaluating the representativeness of the sonde data (WDCGG, CASTNET, EMEP, MOZAIC, section 1).

Page 28753, lines 7-16: this paragraph is somewhat lengthy, please try and phrase it more concisely. You should mention that the bias of the different sondes was evaluated against a UV photometer.

Page 28753, lines 18/19 vs. lines 21-23: Have you applied the CF range criterion or not? Please clarify.

Page 28754, lines 1-4: Please indicate the methodology more precisely. Have you calculated arithmetic means and/or medians from the seasonal samples?

Page 28754, line 26: Comment on the "Japan" region: From the model evaluation viewpoint it makes sense to merge the Japanese sonde data. However, there are significant differences in the seasonal cycle at Sapporo, which shows midlatitude characteristics (high values in spring and summer) and the more southern stations, which are more subtropical in character. Through the merging process, typical features of individual stations may be lost. Please comment.

Page 28755, lines 4-16: Reading this paragraph is difficult: First, it is stated that the tropical stations are combined into one group. But next a list of arguments is given, why the data from the different stations should actually better not be grouped. It is not clear from this list "why a grouping of the tropical region into sub-regions is therefore difficult". Isn't it rather that there are simply not enough stations to reasonably combine the data into sub-regions? Please clarify in the text.

Page 28755, line 7: "Walker circulations" -> "Walker circulation"

3 Variability of ozone within different regions

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

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?

Page 28756, line 9: … exposes differences in the mean or in the median?

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.

4 Representativeness of regional averages in comparison to independent observations

Page 28757, lines 1-3: The sentence is difficult to read. Please rephrase.

Page 28757, lines 7, 15, and 18: MOZAIC, EMEP, WDCGG, CASTNE, see comment for page 28752, first paragraph.

Page 28757, lines 11/12: Please explain which ozonesonde data you used to compare MOZAIC data in Frankfurt with. There is no ozonesonde station nearby Frankfurt.

Page 28759, lines 3-5: As the authors mention, there is a slight offset between MOZAIC and ozonesonde data in the middle to upper troposphere with MOZAIC showing smaller values than the ozonesonde data. This is due to the fact that differences between sondes and MOZAIC were much larger in 1995-1998, as mentioned by several authors, e.g. Thouret et al. (1998), Schnadt Poberaj et al. (2009), and Logan et al. (2011). Taking the longer averaging period, these problems are largely reduced. You should mention this in the text.

Page 28759, line 6: Which are the two ozonesonde stations in the southeast US, please mention and refer to your Table 1.

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.

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.

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.

Page 28760, lines 8-9: What is the specific methodology/what are the features of the particular model evaluation that the authors perform? Please state here already.

Page 28760, lines 11-16: Cumbersome phrasing. I recommend changing the sentence to something like this: "To illustrate the interannual variability, timelines of seasonal mean ozone depending on region and altitude levels are illustrated in Figs. 6 and 7." The last part of the original sentence on the median and half-width of the distribution is already explained in the figure caption and would only be redundant in the text.

Page 28760, lines 17-27: It is not clear why this introduction is necessary in this section. It does not relate to what is discussed later on. Please shift to the introduction or remove completely. Line 21: Please either cite all the work done recently concerning influence of stratospheric ozone on tropospheric trends, i.e. "(Tarasick et al., 2005; Thouret et al., 2006; Ordonez et al., 2007; Terao et al., 2008)" or just cite "(Hess and Zbinden, 2011; and references therein)".

Page 28761, line 8: "… consistent with earlier findings …" The ozone seasonal cycle in the troposphere and stratosphere is well-known and has been discussed in

numerous publications before. Thus, please change to "… consistent with the literature…"

Page 28761, lines 19/20: Why is mid-latitude ozone larger than at high latitudes? Why is this so? Please briefly explain.

Page, 28761, lines 24-29 and page 28762, lines 1-20: This paragraph should be removed. See major comment 4).

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.

Individual comments:
Page 28762, line 21: Remove first part of sentence "For model and data comparisons". This is clear anyway from the motivation of the paper.

Page 28762, lines 24-25: The seasonal cycle over Canada is also characterized by a summer minimum. So it should be "In high northern latitudes and Canada, …". Please give an appropriate explanation why there is a summer minimum at the surface at high latitudes.

Page 28762, Lines 25-27 to page 28763, line 1: Please be more specific: in which altitude regions are the profiles similar?

Page 28763, lines 1-6: How are the tropical characteristics over the US and Japan defined? Please also indicate why there is a seasonal cycle in the ozonopause in the tropics. "A secondary ozone minimum" should be replaced by "An upper tropospheric ozone minimum". 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.

Page 28763, lines 7-11: First the authors state that structure and zonal variability of UT ozone are not shown (this is not possible anyway, because the authors have grouped the tropical stations into one group). Still the zonal variability is described in the next sentence. These consecutive statements are very confusing, because after the first sentence the reader expects that the tropical UT ozone variability is not discussed in the present paper. Hence, because it is described elsewhere and the present analysis does not provide information on the issue, I recommend removing the description on the longitudinal variations in tropical UT ozone.

Page 28763, lines 26-28: In the p –system, air masses cannot be clearly assigned to either upper troposphere or lowermost stratosphere; averages usually represent mixtures of ozone from both altitude regions. Hence, I suggest substituting "seasonality in the lowermost stratosphere" by "seasonality in the tropopause region".

Page 28763, line 25 to page 28764, lines 1-3, and Fig. 10: The seasonal cycles described in the indicated paragraph are not correctly assigned to the altitude ranges above the tropopause (< 2 km above the TP, > 2 km above the TP). Reading from Fig. 10 and lines 25/26, within the lowest two kilometers above the tropopause, the

lowest mixing ratios occur in DJF/SON and higher ones in MAM/JJA. This is the same in the whole NH except for the tropics. Above these two kilometers, the seasonal cycles are similar at high northern and middle latitudes with values from low to high in the order SON, JJA, DJF, MAM, whereas over the US and Japan, the seasonal cycle differs. Please correct.

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.

Page 28764, line 22: Please refer to section 7: "For model evaluation (section 7), …"

Page 28764, lines 22-24: Sentence is confusing. There is no trend information in Figs. 9 and 10. Indicated are the half-width of the ozone distribution and interannual variability of the annual median ozone values.

Page 28764, lines 25-29: Please be more precise. In the lower to middle troposphere, according to Fig. 9, the width of the ozone PDF is rather constant. Only in the upper troposphere, the width starts increasing maximizing at the tropopause.

Page 28765, lines 4-8: In the first half of the sentence, the ozone distribution in the tropopause-referenced altitudes is discussed ("Considering the ozone distribution with regard to tropopause-referenced altitudes, …"), whereas in the second half the effects in pressure coordinates are addressed. This is confusing. Please separate this sentence into two. Additionally, the variability of ozone in the pressure system cannot be clearly assigned to either upper troposphere or lower stratosphere. Thus, please substitute "lowermost stratosphere" by "UTLS".

Page 28765, lines 12-15: Why is the half-width of the distribution in the Tropics larger than in tropopause-referenced altitudes than in pressure altitudes?

Page 28765, lines 15-18: Where does this large interannual variability in the SH midlatitudes come from? Please comment.

Section 7 Application of the ozone climatology to model studies

Page 28766, line 7 + Fig. 11; page 28767, lines 5-7 + Fig. 12: Apparently, model seasonal means are compared against averages of medians from the observations. See major comment 2). Please clarify and/or correct.

Page 28766, lines 16: Any idea why the models tend to simulate the spring peak too early?

Page 28766, line 28: please describe "10-day instantaneous model results" in a little more detail including the information that model data are from 0 UTC on the 1st, 11th, 21st of a month.

Page 28767, lines 9-12: It is not clear to me why the models would overestimate ozone in the lower stratosphere if the TP is shifted upward in the models. Please explain. Since the overestimation still occurs in tropopause referenced altitudes, should it not rather be linked to shortcomings in the representation of stratospheric transport, as indicated in the next sentence?

Page 28767, lines 26-28: The sentence is very confusing: Contrary to what is written, the half-widths of the CCMVal2 distribution are not displayed in Fig. 12. Additionally, it is not clear whether medians of observational and model data are compared or arithmetic means or both. See also major comment 2).

Page 28768, lines 6/7: I would rather put it like this: "Above the TP, the variability of model performance in terms of median and H-distance is smaller than below the tropopause". The middle and right panel of Fig. 13 actually look pretty much the same for most models.

Page 28768, lines 10-12: Please indicate which model this is (see also comment concerning Fig. 13).

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.

8 Conclusions

Page 28769, lines 1 and 2: What are the conclusions drawn from the investigation of the timelines?

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.

Page 28769, lines 4-9: This is the first time that the use of timelines is motivated properly. A similar motivation would be very helpful to introduce section 5.

Page 28769, line 8: It is important to keep in mind potential uncertainties of the sonde data particularly in the earlier period 1990-1999, which may call in question the numbers derived for long-term changes "(5-10%) in high northern latitudes". As discussed in the major comments, I am afraid that modelers not aware of data uncertainties. Hence I suggest to remove these numbers from the conclusions.

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.

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.

Page 28770, lines 12-15: please cf. with comment on page 28764, lines 3-5, Sect. 6.

Page 28770, line 17: "has not been done to date". This is not exactly true. Thouret et al. (2006) and Schnadt Poberaj et al. (2007) provide climatological seasonal cycles for different regions in the Northern Hemisphere, however based on MOZAIC data only. Hence, you should constrain this statement to ozonesonde data.

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.

Page 28771, line 4: briefly indicate potential reasons for the modeled overestimation of lower stratospheric ozone

Appendix A

Page 28771, line 13: "The climatology". You do not provide the climatology, which would be averaged values over a certain number of years, but the monthly mean/median values on which the climatology is based. Please rephrase.

Page 28771, line 13: "hpa" -> "hPa"

Page 28771, line 14: Substitute to "we provide monthly averaged mean and median ozone profiles" and then remove "(mean and median)" in lines 17/18

Page 28771, lines 19/20: "… for the two periods considered." is not clear.

Page 28771, lines 22-24: Please cf. comment in Sect. 3, page 28756, lines 1-3.

Appendix B

Page 28772, 15-22: This motivation for using the Hellinger distance may be correct, but is somewhat out of place in the context of ozone distributions. I recommend removing this sentence. To my mind, the fact that ozone distributions are often not Gaussian in the troposphere and the reference to Fig. 15 is sufficient.

Page 28772, line 26: "shape" -> "shapes"

Page 28773, lines 10-13: For which figures do you use variable bin sizes to calculate the CDF, for which not? The two sentences seem inconsistent. Please specify.

Page 28774, line 18: "Fig. 15, …" remove comma

Page 28774, lines 22-24 vs. line 27: " … against the percentage difference of the medians of the distributions …" vs. "… if the differences in the mean are small …" Mean or median?

Figures

Figs. 9 and 10: The first sentence of the figure caption in Fig. 9 is confusing "Altitude distributions (medians) … averaged for different seasons …". Have you displayed medians from seasonal samples or is there some additional averaging implied? Please clarify.

Fig. 13: Please indicate which models (HTAP or CCMVal2) have been used. Additionally, please add a legend assigning the colors to individual models, as done e.g., in Eyring et al. (2010).

Supplemental Material

Fig. 1, upper left panel (Resolute (1995-2007): The SON total number (91.7) should be an integer.

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.

Fig. 3: Which colors represent which seasons? Information missing.

References

DiNunno, B., et al., Central/eastern North Pacific photochemical precursor distributions for fall/spring seasons as defined by airborne field studies, JGR, 108, 8237, doi:10.1029/2001JD001044, 2003.

Logan, J.A. et al., Changes in ozone over Europe since 1990: analysis of ozone measurements from sondes, regular aircraft (MOZAIC) and alpine surface sites, submitted to JGR, 2011.

Schnadt Poberaj, C., et al., A UT/LS ozone climatology of the nineteen seventies deduced from the GASP aircraft measurement program, Atmos. Chem. Phys., 7, 5917-5936, 2007.

Schnadt Poberaj, C., et al., Long-term changes in UT/LS ozone between the late 1970s and 1990s deduced from the GASP and MOZAIC aircraft programs and from ozonesondes, Atmos. Chem. Phys., 9, 5343-5369, 2009.