Atmos. Chem. Phys. Discuss., 8, S7107–S7119, 2008
www.atmos-chem-phys-discuss.net/8/S7107/2008/

Atmospheric Chemistry and Physics Discussions

Interactive comment on "Validation of ozone measurements from the Atmospheric Chemistry Experiment (ACE)" *by* E. Dupuy et al.

E. Dupuy et al.

Received and published: 16 September 2008

We thank Anonymous Referee #2 for his/her valuable comments, which we feel have helped improve the manuscript. We hope to have replied to each comment to the satisfaction of the Anonymous Referee. We have made efforts to implement almost all the suggested changes, including possible additions to the manuscript, while trying not to add to its length. The responses are detailed below, with the original comments indicated in italics.

This is part 1 of our response to Anonymous Referee #2 addressing the General and Major Comments. Because of the maximum length allowed, the Minor Comments from Referee #2 will be addressed in a separate author comment.

ACPD

8, S7107–S7119, 2008

Interactive Comment

GENERAL COMMENTS

This paper is an extensive collection of separate comparisons of ACE-FTS and ACE-MAESTRO O_3 profiles with various satellite, ground-based and airborne instruments. It is well written and well structured, appropriate references are given. Care is taken to specify processor version numbers for the correlative instruments and their estimated uncertainties. Most comparisons have been done with the same methods and with similar collocation criteria, and are presented in figures with similar layout. The conclusion of this paper will be very important for future use of ACE O_3 profiles.

The paper is very lengthy. Especially the descriptive parts of sections 5 and 6 can be reduced considerably, by collecting similar information for each study for instance in tables and giving details only when necessary, see further details in the specific comments.

In spite of the wealth of collocation data available in this paper, and some almost hidden nice remarks on statistical significance, there is no attempt made to calculate or to discuss the significance of the comparisons. The reported uncertainties for the correlative O_3 profiles are not used in an assessment of the uncertainties in the calculated differences. It should be discussed and calculated how the different vertical smoothing methods, the different collocation criteria, and the different characteristics of the correlative profiles affect the significance of the average difference profiles.

In the end the individual average relative difference profiles for the comparison studies are collected in two figures and a description of the figures is given as main conclusion *of the paper. There is no discussion on statistical significance or relative importance of the different comparisons. A comparison with a few thousand collocations of a well validated instrument should have a different impact on the quantitative conclusion than a comparison where the number of collocations or the level of validation is much less.*

While alternative approaches could be taken, we have chosen to view this paper as a comprehensive document to be used as a reference for quality assessment of the current ACE ozone products, both ACE-FTS version 2.2 Ozone Update and ACE-MAESTRO version 1.2 ozone. Therefore, this paper groups comparison analyses for both ACE instruments and includes comparisons with a broad range of data products. In Sections 5 and 6, we have attempted to include only the information necessary to describe each data set and the comparisons with ACE.

The significance of the results is illustrated by the de-biased standard deviations and the standard errors of the mean differences, shown for all statistical comparisons. We have added discussion of the de-biased standard deviations of the mean differences, and of the standard errors of the mean, in these sections. Whenever possible, we have taken the reported uncertainties of the correlative instruments into account in this discussion. We have used Figures 46 and 47 (Figures 45 and 46 in the revised manuscript) to summarize the statistical comparisons of VMR profiles between ACE and the correlative measurements. The main conclusion drawn from these figures is that there is very good consistency in the comparison results, especially for ACE-FTS. As can be seen from Figure 46 (now Figure 45), the results at stratospheric altitudes (and to a lesser extent for mesospheric altitudes) are extremely consistent, regardless of the characteristics of the comparison measurements, the number of coincidences, and the level of validation of the correlative data. The same can be said about ACE-MAESTRO for which there remains, however, a significant difference between the sunrise and sunset observations. We feel that the consistency of the results, despite the variety of comparison data (obtained in various spectral domains with different obser-

ACPD

8, S7107–S7119, 2008

vation geometries and techniques) constitutes in itself an assessment of the quality of the ACE-FTS version 2.2 Ozone Update and ACE-MAESTRO version 1.2 ozone data products.

SPECIFIC COMMENTS

Major comments:

Section 2: I miss a description of the ACE profiles in terms of sensitivity, retrieval uncertainties, and averaging kernels. This is needed to understand the observed differences.

The retrieval algorithms of both ACE instruments use methods other than optimal estimation. Therefore, averaging kernels are not readily available and the covariance matrices are not stored during the operational processing. Furthermore, a detailed error budget including systematic errors is not available for the ACE-FTS version 2.2 data products but will be produced for the next data release (v3.0). This has been clarified in the text. Currently, only spectral fitting errors are available for ACE-FTS and ACE-MAESTRO. A sentence was added in both sections (2.1 and 2.2) to give their typical values. Without estimates of the systematic errors of the ACE instruments, rigorous precision validation is not possible. Hence, the main focus of this paper is bias determination for the current ACE ozone products. This has also been clarified in the text.

For example, most difference profiles are given between 5 and 60 km. However the FTS measurement starts at cloud-top, as stated in p.2520,l.20. What is the effect of this lower limit on the difference profiles of the lowest altitudes?

The lower limit of the difference profiles is generally 8-10 km, extending in a few cases to 5 km. The presence or absence of clouds (as well as the satellite beta angle) influences the lower altitude limit of the ACE-FTS retrieved ozone profiles. Therefore,

ACPD

8, S7107–S7119, 2008

Interactive Comment

there is a sharp decrease in the number of coincidences (indicated in each figure) from ~15 km downwards. The scatter of the individual relative differences at these lower altitudes increases accordingly and the statistical significance of the results is lower (for detailed validation at the lowest altitude levels see Hegglin et al. (2008)). This has been clarified in the revised manuscript.

Are a priori profiles used in the retrieval? What is their expected influence, especially around the lowermost and uppermost retrieval altitudes?

The ACE-FTS fitting algorithm does not rely on optimal estimation and does not impose constraints derived from a priori information. The retrievals are thus insensitive to the first-guess profiles over the vertical ranges analysed. Section 2.1 description was emended to clarify the description of the retrieval process.

p.2520,l.19: The altitude coverage of the measurements extends from the cloud-tops to 100-150 km. How are clouds expected to affect the retrieval? There is no mention of cloud or cloud-height dependence in the validation results. Is it not studied or is it expected to be unimportant. Please explain.

The suntracker used by the ACE instruments cannot operate when thick clouds are present in the field of view. Therefore, the measurements effectively stop when thick clouds are encountered and no retrievals can be done. The reported VMR profiles, therefore, do not extend below the cloud top level, and the validation results are expected to be insensitive to clouds. This information was added in Section 2.

I miss a direct comparison between FTS and MAESTRO profiles.

This comparison was made by Kar et al. (2007) for all data acquired by the instruments in 2004. We have redone this analysis with all data from the comparison period used in this paper (Feb. 2004 - Aug. 2006) and no significant difference was found in the

ACPD

8, S7107–S7119, 2008

Interactive Comment

results. We expanded the paragraph summarizing the results of Kar et al. (2007) in Section 2.2 to include this information.

Section 3: It is not clear why these specific time and distance coincidence criteria are used. I would expect a discussion here on the variability of ozone at different altitudes in time and space, mostly determined by dynamical processes. This discussion would result in the preferred time and distance criteria. Relaxing the time difference criterion to 24 hours, for e.g. the comparison with ozonesondes, while leaving the distance criterion at 800km seems strange in this respect.

Coincidence criteria can vary widely between different validation studies. The coincidence criteria for our study have been chosen to ensure a sufficient number of coincidences in all comparisons while trying to limit the scatter resulting from relaxed coincidence criteria. They are consistent with criteria used for other satellite solar occultation instruments that measure at high latitudes (e.g. Randall et al. (2003)) and we found that the mean relative differences we obtained were not sensitive to these criteria. This result is similar to that noted in a recent paper on TES validation with ozonesondes by Nassar et al. (2008) who stated that the choice of coincidence criteria impacts the variability determined much more than the bias. Furthermore, the distances between coincident observations are comparable to the typical ground-track distance (300-600 km) covered by a single measurement (occultation or limb-scan) of a satellite instrument.

p.2525,l.11-13: Figure 1 is not a good example. This figure shows that the two comparison instruments do not have a latitudinal bias with respect to eachother, not that the differences in the ozone values are independent of latitude. A good example would be a figure with average difference profiles for the five latitude bands showing that they are not significantly different.

This figure has been deleted. We have revised the description and discussion of the

ACPD

8, S7107–S7119, 2008

Interactive Comment

sensitivity of the comparison results to different criteria in our revised manuscript. However, to avoid increasing the length of the paper further, we have not included any figures showing results of these sensitivity studies for the different comparison datasets.

p.2525,l.22-24: 'In such cases ...': This is not good practice for validation. Allowing multiple observations from the comparison instrument per individual ACE value only increases the spread in the result. It doesn't change the significance of the average difference. Calculations of the uncertainty in the mean, used in the figures, are wrong. The standard deviations should be divided by the square root of the number of INDEPENDENT pairs MINUS one.

We agree with Anonymous Referee #2 that multiple coincidences will have an impact on the standard deviation of the ensemble while the mean differences will be little changed. Since we could not conduct a complete precision validation study, we have chosen to keep all coincident events in the calculations. The statement on the uncertainty of the mean given above is incorrect. The factor $N-1$ is found in the expression of the standard deviation since the mean value is calculated from the same sample. However, calculating the uncertainty of the mean (or standard error) requires further division by $\sqrt{N}$ but not by $\sqrt{N-1}$ (Randall et al., 2003; von Clarmann, 2006, Equation 31). There is no requirement that the pairs be independent. For clarification, we have added the equations of the standard deviation (of the bias-corrected differences) and of the standard errors in the methodology section of the revised manuscript.

p.2526,l.7-11: 'Day/night differences ...': What is the expected effect on the average difference?

As noted in the text, the diurnal variations in the ozone concentrations are not expected to have a significant impact at stratospheric altitudes (see, e.g., Schneider, N. et al. (2005)) while, above 50 km, the nighttime increase in the ozone abundance is typically 30% around 52 km and ∼60% at 60 km. We believe that the tight (2 h) time

ACPD

8, S7107–S7119, 2008

Interactive Comment

difference criterion that we have chosen serves to minimize the impact of true diurnal effects on the calculated mean differences. The good consistency of the mean relative differences, even at mesospheric heights, is also a good indication that the day/night differences are not the main factor in the observed biases. A more detailed investigation of these mesospheric differences is on-going.

Section 4: Different methods of smoothing have been used for different comparisons. The idea behind this is not made clear. Why not do all the comparisons with the same vertical smoothing? What is the influence of the different methods on the comparison results? It is not only the smoothing that is important when comparing two profiles with different vertical resolution, also the sensitivity to different altitudes is different for different instruments. This is expressed in the averaging kernels. How do the different smoothing methods affect the significance of the result?

The smoothing methods used in this work depend on the vertical resolution of the comparison dataset. For comparison measurements with lower vertical resolution than ACE (FTIRs, MWRs), averaging kernels calculated during the comparison instrument retrieval process were used to smooth the ACE-FTS and ACE-MAESTRO profiles according to the method of Rodgers and Connor (2003). For comparison measurements with higher vertical resolution than ACE, we could not use the Rodgers and Connor (2003) method because no averaging kernels are available for the ACE measurements. In this case, different methods were used. For most of the in situ and high-resolution profile comparisons, the ACE-FTS profiles were smoothed using triangular convolution functions, to take account of the field of view of the instrument and the altitude spacing of the measurements, and the ACE-MAESTRO profiles were smoothed using a Gaussian filter, a method consistent with the previous work of Kar et al. (2007). The column integration approach, preferred by one of our collaborators, was used for the ozonesonde and lidar comparisons described in Section 6.6. This technique gave results that were consistent with those obtained using the triangular convolution function

ACPD

8, S7107–S7119, 2008

Interactive Comment

and Gaussian filter smoothing methods.

p.2526,l.17-19: 'Tests with other interpolation methods ...:': Please show this and give more details.

Tests of different interpolation methods were done for several comparison datasets. The mean relative differences obtained with these interpolation methods (linear, cubic spline and quadratic) were very consistent with maximum differences of a few percent at the lowest altitudes. This information has been included in the revised manuscript. However, for brevity, a new figure has not been added.

p.2524,l.20-23:'Analysis of the variation ... consistent systematic biases': This is an important finding and should be made reproducible. The study should be described: which correlative instruments are used, which geometric parameters were studied?

As mentioned above, we have revised our discussion of the sensitivity of the results to the temporal and geolocation criteria. The sentence referred to above has been moved to a later paragraph. These studies were done for all satellite instruments and ozonesondes.

p.2529: Again no scientific reason is given for using different methods to calculate relative differences. How does it affect the significance of the results? The method used for GOMOS is again different than what is described here. Should be added in this section.

For the relative difference calculations, we have attempted to present results using a consistent methodology. In this approach, we have used different denominators for the ozonesondes and satellite comparisons based on the assumption that the in situ high-resolution ozonesonde measurements are a good reference for the comparisons, while satellite-borne measurements are affected by larger uncertainties and a more

ACPD

8, S7107–S7119, 2008

logical reference is the average of both instruments' VMRs (as described in Randall et al. (2003)). This is now explained in the revised manuscript. The exceptions (ground-based instruments, GOMOS and ASUR) are explained by the fact that the analyses were performed by different research groups. The description of the calculation method applied to GOMOS is detailed in the relevant section. We have added a sentence in Section 4 to direct the reader to it.

Sections 5 and 6: In general these sections are too long. Much of the information is of similar form in each subsection. It would be better to arrange the information important for judging the significance of the comparison results in a table. Similar information per correlative instrument is, for instance, software version, estimated uncertainties based on validation with what instruments, vertical resolution and valid altitude range, references to retrieval and validation papers, selection criteria for filtering and for coincidence. Other specific information per instrument relevant for the conclusions can be left in the text.

As stated earlier, we have attempted to include only the necessary information for discussing each dataset and the comparisons with ACE in the subsections of Sections 5 and 6. We would prefer to keep all of this information in paragraph form rather than include some of it in a table. However, the software version for each satellite dataset has been included in Table 1.

The description of the difference profiles should be made more uniform. First the terminology used for the relative difference profiles is different per study: 'fractional differences', 'relative differences', 'percent differences'. The terminology in the description of the difference profiles is also very diverse: 'agreement is better than 'x%', 'within %', 'within +x to +y%', 'within x-y%'. Also the altitude ranges over which the average differences are reported are different, and it is not clearly stated why. The word 'typical' is used often where in fact an approximate average value is meant, just give the average *value.*

The text was modified in an effort to homogenize the terminology used to describe the difference profiles and discuss the mean relative differences. Average values have been used in the revised text. In Table 7, the typical values have been replaced with average values of the mean relative differences. The altitude ranges reported in Table 7 are those where the mean relative differences were within $\pm$10%, with the exception of comparisons with SMR and the Eureka lidar as noted in footnote (b) of Table 7.

I recommend to use a more integrated approach, and not describe all the difference profiles separately in detail. In the end, the reader would like to have a feeling on how well ACE profiles compare with the 'truth'. You compare with many different data sets, most of them are validated, and have an estimated bias with respect to this 'truth'. In reality also these other instruments are all compared to each other. This means that many of the mentioned bias estimates are related to each other. Please use a rigorous method to derive bias estimates for the average relative difference with respect to the 'truth' at each altitude, considering of course that you have only limited knowledge of the 'truth'.

As stated previously, the strength of this work resides in the use of one or more comparison datasets from nearly 20 different instruments or sets of instruments to provide statistical comparisons with the ACE ozone data products. Considering the consistency of the results illustrated by Figures 46 and 47 (45 and 46 in the revised manuscript), we feel that the biases of the ACE-FTS version 2.2 Ozone Update and ACE-MAESTRO version 1.2 ozone data products have been correctly characterized. Table 7 gives our best estimate of how the ACE products measure the atmospheric ozone abundance. Combining the wealth of validation results obtained for the correlative datasets, in order to derive a “true” profile to be compared with, is a considerable task and is beyond the scope of this paper. However, Section 6.9 shows how this can be done with a smaller number of comparison instruments. The biases derived in Section 6.9 are consistent

ACPD

8, S7107–S7119, 2008

Interactive Comment

with the overall results of the paper. Finally, standard errors are shown and discussed in the paper and are the quantity needed to judge the significance of the estimated biases.

Within this analysis also the standard deviation of the differences should be discussed. It is a measure for the combined random retrieval errors in both data sets combined with natural variability within time and space, associated with the looseness of coincidence criteria, and with errors associated with representativeness for the true atmospheric profiles, as indicated by observation operators. The large standard deviation in Figure 13 is a good example where the reader would like to know if this is expected.

As indicated above, we have included a discussion of the standard deviations and standard errors of the mean differences in the revised version of the paper. Note that we have followed the approach of von Clarmann (2006) by using “de-biased” standard deviations, a better measure of the combined precision of the instruments compared.

References

Hegglin, M. I., Boone, C. D., Manney, G. L., Shepherd, T. G., Walker, K. A., Bernath, P. F., Daffer, W. H., Hoor, P., and Schiller, C.: Validation of ACE-FTS satellite data in the upper troposphere/lower stratosphere (UTLS) using non-coincident measurements, Atmos. Chem. Phys., 8, 1483-1499, 2008.

Kar, J., McElroy, C. T., Drummond, J. R., et al.: Initial comparison of Ozone and NO_2 profiles from ACE-MAESTRO with Balloon and Satellite Data, J. Geophys. Res., 112, D16301, doi:10.1029/2006JD008242, 2007.

Nassar, R., Logan, J. A., Worden, H. M, et al.: Validation of Tropospheric Emission Spectrometer (TES) nadir ozone profiles using ozonesonde measurements, J. Geophys. Res., 113, D15S17, doi:10.1029/2007JD008819, 2008.

Randall, C. E., Rusch, D. W., Bevilacqua, R. M., et al.: Validation of POAM III ozone:

ACPD

8, S7107–S7119, 2008

Interactive Comment

comparison with ozonesonde and satellite data, J. Geophys. Res., 108(D12), 4367, doi:10.1029/2002JD002944, 2003.
Rodgers, C. D., and Connor, B. J.: Intercomparison of remote sounding instruments, J. Geophys. Res., 108(D3), 4116, doi:10.1029/2002JD002299, 2003.
Schneider, N., Selsis, F., Urban, J., et al.: Seasonal and diurnal ozone variations: observations and modeling, J. Atmos. Chem., 50, 25–47, 2005.
von Clarmann, T.: Validation of remotely sensed profiles of atmospheric state variables: strategies and terminology, Atmos. Chem. Phys., 6, 4311ñ-4320, 2006.

Interactive comment on Atmos. Chem. Phys. Discuss., 8, 2513, 2008.

ACPD
8, S7107–S7119, 2008