Atmos. Chem. Phys. Discuss., 8, S2099–S2112, 2008
www.atmos-chem-phys-discuss.net/8/S2099/2008/

Atmospheric Chemistry and Physics Discussions

ACPD

8, S2099–S2112, 2008

Interactive Comment

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

Anonymous Referee #2

Received and published: 28 April 2008

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

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

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.

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

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

ACPD

8, S2099–S2112, 2008

Interactive Comment

of cloud or cloud-height dependence in the validation results. Is it not studied or is it expected to be unimportant. Please explain.

I miss a direct comparison between FTS and MAESTRO profiles.

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.

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.

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.

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

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 comaprison results? It is not only the smoothing that is important when comparing two profiles

ACPD

8, S2099–S2112, 2008

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?

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

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?

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.

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.

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

ACPD

8, S2099–S2112, 2008

Interactive Comment

value.

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 eachother. This means that many of the mentioned bias estimates are related to eachother. 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’.

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.

Minor comments:

p.2521,l.2: ’pressure and temperature profiles are used to calculate’: The VMR profiles are calculated from the spectra, not from temperature and pressure profiles, please reformulate and summarise the retrieval method for VMRs.

p.2523,l.4-11: The conclusions for MAESTRO profiles from Kar et al., 2007, are almost the same as the conclusions in this paper. What is the added value of the studies in the current paper?

p.2523,l.16: ’good agreement’: from Kar et al., 2007? Please quantify ’good’.

p.2524,l.12:’time differences were calculated in Universal Time (UT)’: time differences are in hours or minutes, it is irrelevant which coordinate system is used when talking

ACPD

8, S2099–S2112, 2008

Interactive Comment

about differences.

p.2525,l.5-11: 'To test the sensitivity ... during these comparisons': Was this done for all correlative data or a subset?

p.2525,l.16-17: 'Careful examination ...': This is a vague sentence, should be clarified. It is not clear how one can examine 'time series' (i.e. relative differences as a function of time) as a function of distance. What is the 'observation geometry' quantitatively?

p.2527,eq.(1): please make clear in the formula what depends on what, e.g:

$$x_S(z_i) = \frac{\sum_{j=1}^{n_{hr}} w_j(z_j^{hr} - z_i) \cdot x_{hr}(z_j^{hr})}{\sum_{j=1}^{n_{hr}} w_j(z_j^{hr} - z_i)}$$

p.2528,l.1-2: What are partial columns reported on a grid? The partial column is the column between two altitudes. Are the ACE profiles reported in partial columns for the study in 6.6? If partial columns are used for comparisons, than how does it affect the quantitative overall results?

p.2528,l.12-18: Very good, but the filtering is mentioned only in a few cases. Please check and report for all correlative products which filtering is done.

p.2528,l.22-24: The filtering by 'visual examination' is not mentioned in any of the comparisons, please specify why profiles have been rejected.

p.2530,l.4: should be: 'relative differences CAN become negative'

p.2531,l.4: 'below the tropopause': The tropopause is usually not at 18km, please rephrase.

p.2531,l.14: Change 'The agreement' to 'The average agreement'.

p.2534,l.23: change 'in all cases' to 'for all latitudes'

p.2534,l.25: change '9-41 km' to '23-41 km' p.2535,l.6: change 'the statistics of the analysis' to 'the number of coincidences'

p.2536,l.27-p.2537,l.3: 'There is no noticeable ... based on these comparisons.': You already have the error bars, giving you the information on significance (note the remarks made earlier). If it is not significant (please calculate!), it is enough to say: 'There is no significant difference between the ACE-FTS SR and SS comparisons.'

Section 5.2.1. Odin/OSIRIS: Is v3.0 not validated? What are the expected differences with the older version(s)?

p.2539,l.3: 'Since the comparisons': This is not a good reason, It could theoretically be possible that all occultation instruments have the same bias between sunrise and sunset measurements. Furthermore, Figure 6 shows a bias between SR and SS measurements above $\sim$40 km, which is not explicitly noted in the text.

p.2539,l.13: change '$+4$ to $+11\%$' to '0 to $+11\%$'

p.2539,l.14: shoud read '... largest RELATIVE differences ...'

p.2539,l.18: remove '(except $\ldots + 9\%$)': the error bar is large enough.

p.2539,l.23-25: Most coincidences with OSIRIS are within an hour (table 1): what would be the expected difference? Can't you check this statement with narrower coincidence criteria?

p.2540,l.11: change 'above 40 km' to 'between 40 and 60 km'

p.2542,l.1: change 'relative differences' to 'Average relative differences'

p.2543,l.16: figure 14 only goes up to 64 km.

p.2544,l.2: change '22 and 42 km' to '19 and 42 km'

Section 5.4.1 Envisat/GOMOS: Version IPF 5.00 is used, what is the expected difference to version 6.0a, which was validated?

p.2545,l.16-26: This method should be described in section 4. The reason given to use

ACPD

8, S2099–S2112, 2008

Interactive Comment

this method is a good one, so why are not all comparisons done with this method? For 1240 coincidences, is the result different from when the mean instead of the median was used?

p.2545,l.25: give formula for w_i

p.2546,l.18: change 'more important' to 'larger'

p.2546,l.20: change 'significantly' to 'strongly'

p.2546,l.4-25: Here a photochemical correction is used to account for the time difference between the GOMOS and ACE-FTS measurements. Why is this not also done for the other high-altitude comparisons?

p.2546,l.22-25: 'difficult to draw conclusions': If we a are talking about 1240 coincidences, it is very well possible to draw conclusions about the average difference between GOMOS and ACE-FTS above 60km, especially when estimating the uncertainty introduced by the model.

p.2546,l.26-p.2547,l.2: this paragraph is better moved to before the paragraph starting at p.2546,l.16. What does the 10.5km width mean? Is this not expected from the GOMOS resolution? If not, is there another explanation?

p.2548,l.27: 'which cannot be accounted for by the combined systematic uncertainty estimates': why not? First, there still is no FTS uncertainty estimate, this is to be derived in this paper, and second the uncertainty estimate of MIPAS was 10%, as stated in l.16.

p.2549: There is no information on uncertainty estimate / validation results for reduced-resolution mission ESA product.

p.2549,l.26-27: 'Differences are within $\pm 8-10\%$': why a range?

p.2550: most striking features in Figure 20 are the discontinuities in the average profile gradient of the IMK profile. Please explain.

ACPD

8, S2099–S2112, 2008

Section 5.4.3 Envisat/SCIAMACHY: Version 1.63 of the IUP profiles are used. What is the expected difference to the validated version 1.62?

p.2552,l.24-26: does this suggest a solar zenith dependent bias in the SCIAMACHY profiles?

p.2555,l.14: 'constant solar zenith angle of 78°': how is that possible, do you mean 'constant zenith viewing angle'?

p.2556,l.10: change 'by up to' to 'down to'

p.2556,l.18: 'very good agreement around the ozone VMR peak': not at all.

p.2556,l.19-22: Of course not: the mean of the individual differences is at each altitude exactly equal to the difference of the means:

$$\frac{1}{N}\sum_{i=1}^{N}(x_A(i) - x_B(i)) = \frac{1}{N}\sum_{i=1}^{N}x_A(i) - \frac{1}{N}\sum_{i=1}^{N}x_B(i)$$

I guess that there is an error in the calculations, underlying this figure.

Section 6.2 (FIRS-2): The comparison is between profiles which are more than 24 hours apart, this only makes sense if it has been checked if approximately the same air mass has been measured.

p.2559,l.14-19: I don't understand this explanation. I don't see any evidence of an altitude shift in one of the other comparisons, so why here? ACE-FTS retrievals are starting at cloud-top, as stated in section 2, so why is there an ACE-FTS profile below 18 km if there are clouds?

p.2560,l.20-25: Very good to use PV to confirm measuring the same air mass, in fact it can also be used to get better coincidence criteria. See also the remarks to Section 3 and Section 6.2.

p.2560,l.27-p.2561l.3: No numbers should be given here. From Figrue 29 and 30 it is

ACPD

8, S2099–S2112, 2008

Interactive Comment

clear that the ACE profiles are within the error bars of the smoothed SPIRALE profiles. This would be clearer if you would have plotted error bars over the difference profiles.

p.2561,l.25-27: Why average ACE profiles for one ozonesonde measurement? This only increases the spread. There should be one profile which has the best coincidence.

p.2562,l.6: change 'the differences' to 'the average relative differences'

p.2562,l.16-19: I don't understand this statement. I don't see any significant difference in figure 32 between the SR (top,middle) and SS (bottom,middle) absolute difference profiles.

p.2562,l.19-21: I don't understand the message. It is true that the VMR values are small below 15 km and that therefore the relative differences can be high while the absolute differences are not. However, both the relative and the absolute differences seem to be significant from the plot.

p.2563,l.10: why are only stations included with at least three coincidences?

p.2565,l.4: change 'ACE-MAESTRO are comparable' to 'ACE-MAESTRO with lidars (bottom panel) are comparable'

p.2566,l.1-2: change 'The agreement between ... is within -12 to $+4\%$ ...': to 'The average relative difference between ... is between -10 and 0% ...'.

p.2566,l.12: change 'lidars and show' to 'lidars but show'

p. 2567,l26-29: The degrees of freedom are not used, since the whole useful FTIR altitude range is integrated into one partial column to be compared to that from ACE. Why not compare smaller altitude ranges?

p.2568: the correlation coefficients should be calculated with similar coincidence criteria for all stations

p.2568,l.20: change 'qualitative' to 'quantitative'

ACPD

8, S2099–S2112, 2008

Interactive Comment

p.2570,l.22-24: a link to Table 6 would be in place here.

p.2572,l.11: remove 'remarkably' and 'very'

p.2572,l.14: change 'relative difference' to 'average relative difference'

p.2573,l.2: remove 'very'

p.2573,l.11-21: The expected bias in the comparisons due to diurnal variation should be quantified, also a test should be performed with narrower coincidence criteria, and possibly using photochemical corrections to account for remaining time differences.

p.2574,l.15-17: Not clear from Figure 47. It seems that the deviation between SR and SS average difference profiles starts around 35 km.

p.2574,l.26: I would expect that an offset of a few kilometers would be more apparent in the comparisons.

p.2575,l.17-18: Any ideas why the SR/SS bias is not apparent in the SMR comparisons?

p.2575,l.25-26: Why is it unlikely to account fully for this bias?

p.2576,l.3; l.21; l.24: remove 'very'

p.2577,l.11-16: These studies were either not mentioned at all, or only briefly mentioned, or only shown as one example. That is far too meagre to put it in the conclusions.

Table 1:

The latitude coverage given for a few of the coincidence data sets are confusing:

* SAGE II has observations in two small latitude bands: 75-80S and 75-80N. Table 1 suggests that there are also coincendences between 75S and 75N.

* SAGE III has observations in two latitude bands: 45-80N and 25-60S. Table 1 sug-

ACPD

8, S2099–S2112, 2008

Interactive Comment

gests coincidences between 25S and 45N.

* POAM III has observations in two latitude bands: 55-71N and 63-88S. Table 1 suggests coincidences between 63S and 55N.

* The approximate number of events do not always match the number of events in the text, why not give the actual number?

* The number of events should be split in SR and SS

* The column distance should not have the mean and standard deviation, but instead the distance within which 50% or 68% of the coincidences are falling.

Table 2, caption: change 'emphasized' to 'italic'

Table 6: The version numbers of the satellite data used in the MWR comparisons are not always the same as those used for the ACE comparisons: GOMOS (6.0a in section 5.4.1), MIPAS (only 4.62 in secton 5.4.2), Lidar (none mentioned in section 6.6). Explain in caption what a range in the number of coincident pairs means.

Table 7:

* The number of events for SAGE II versus FTS SR should probably 199 instead of 99.

* This table should have an extra column for version number and an extra column for bias of the correlative instrument

* The column 'typical difference' should be replaced by 'average relative difference'.

* The whole table will probably change when the more rigorous statistical significance has been performed.

Figure 1, caption: 'symbols and colors correspond to those in the top panel': symbols are '+' in the bottom panel, 'o' in the top panel. Colors and symbols are not explained. Color seems to be indicating the time??

Figures with difference profiles:

ACPD

8, S2099–S2112, 2008

* An extra line should be added indicating the uncertainty estimate of the correlative instrument.

* The error bars should be the standard deviation divided by the square root of the number of INDEPENDENT pairs MINUS one.

Figure 12: Plot should go up to 55 km, according to the text no measurements above 55 were taken.

Figure 17: It is not clear from the caption whether the difference profiles use the convoluted and photochemically corrected profile (black dashed) or only the photchemically corrected one (blue).

Figure 29: Error bars are missing in the middle and right plot. Error bars in the left plot are never smaller than the width of the symbol, so this can be left out of the caption.

Figure 36 and 37 miss information on uncertainty. This can either be given in the caption or by plotting for example only those points for which the 2σ values are smaller than the absolute value.

Figure 42: It seems that Kiruna and Thule are giving deviating results. Please discuss in text.

TECHNICAL CORRECTIONS:

p.2560, l.26: change 'on the order' to 'of the order'

p.2560, l.27: change 'convolution functions' to 'triangular convolution function'

p.2569,l.1: remove: 'coincident or'

Figure 16: The middle and right figure are exactly the same except for the x-axis.

Figure 17: Change 'extrapolated' and 'ext' to, e.g, 'corrected' and 'corr'

Figure 31: 'solid black' is actually 'solid blue'

ACPD

8, S2099–S2112, 2008

Interactive Comment

Figure 40: The dotted lines are confusing. Suggestion: make one solid line for $y = 0$, a dashed line for the average value, and two dotted lines indicating the standard deviation. Leave out the lines at 10% intervals.

Figure 41: The colors/symbols in the lower right panel are wrong.

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

ACPD

8, S2099–S2112, 2008

Interactive Comment