

Interactive comment on “Comparisons between SCIAMACHY and ground-based FTIR data for total columns of CO, CH₄, CO₂ and N₂O” by B. Dils et al.

B. Dils et al.

Received and published: 23 August 2005

We would like to thank the referee for the review and useful comments and believe that the paper will be considerably improved by them.

General comments

It is very important to perform a proper validation of trace gas measurements from satellites before using these measurements for scientific studies. This is particularly true for greenhouse gas observations, which have only recently become available, and for which highly demanding precision requirements are set. The present study is a first attempt to validate measurements of CO, CH₄, CO₂, and N₂O from the NIR channels of SCIAMACHY by using a network of ground-based FTIR instruments. FTIRs can

Full Screen / Esc

Print Version

Interactive Discussion

Discussion Paper

measure the total column of the respective trace gases, and thus constitute one of the few possibilities to directly validate the (total-column) satellite data. Therefore, a study as presented here is an excellent means to determine the quality of the satellite measurements. There are, however, a number of problems with the present paper, which should be addressed before it can be published. Firstly, using FTIR for global validation of satellite measurements of greenhouse gases is not an easy task. 1. The FTIRs are often located on high mountains, thus sample a considerably smaller part of the atmospheric column than the co-located satellite measurements.

The problem has been overcome to some extent by re-scaling all column data to zero altitude. This cannot be done in a perfect way, because the auxiliary data needed to do this precisely, especially the vertical distribution of the species, are missing. But at least, it makes the data better comparable. This was explained in the paper.

2. The FTIR network is not uniform in the sense that measurements at the different stations are done in a different way, possibly leading to systematic biases between the stations.

All stations belong to the NDSC, and therefore comply with NDSC standards. Regular efforts are carried out in the frame of the network to make the data at different stations mutually consistent; these include intercomparison campaigns, retrieval algorithm comparisons, data exchanges, etc.

3. The FTIR network gives a poor global and even latitudinal coverage, so that statements on the global quality of the satellite data are hard to make.

This is true. Unfortunately, there exist no independent data sets that do have really global coverage, apart from MOPITT data, but for CO only, and global model data, like TM3 and TM5. Comparisons with MOPITT data have been carried out to some extent by the algorithm developer groups, (Gloudemans, A. M. S., Schrijver, H., Straume, A. G., Aben, I., Maurellis, A. N., Buchwitz, M., de Beek, R., Frankenberg, C., Wagner, T., and Meirink, J. F.: CH₄ and CO total columns from SCIAMACHY: Comparisons with

[Full Screen / Esc](#)[Print Version](#)[Interactive Discussion](#)[Discussion Paper](#)

TM3 and MOPITT, in Proceedings of Second Workshop on the Atmospheric Chemistry Validation of ENVISAT (ACVE-2), ESA/ESRIN, Frascati, Italy, 3-7 May 2004, ESA SP-562 (on CD), 2004.) -they suffer from the fact that MOPITT data have lower sensitivity close to the surface. Efforts are being made to use the FTIR data for the validation of the above global model data, to get more confidence in the use of the models to perform a validation on the global scale.

4. The authors acknowledge these issues, and do some efforts to overcome them. Still, I feel that more work should be done to address the first two points.

While it is obvious that the availability of a very large set of (low altitude) FTIR ground based stations, covering almost the entire globe, would certainly be ideal, the current reality is different. That said we feel that with the stations as available to us now, significant and sounded conclusions can and are drawn. We did not shy from pointing out the limitations of the datasets used and a conscious reader will take note of them when assessing our results.

5. A second shortcoming of this study is that the data sets from the three satellite retrieval algorithms cover only part (and different parts) of the target year 2003. This makes intercomparison and discrimination between the algorithms very difficult.

Again, as above, while the datasets were in some respects limited, useful conclusions could be drawn from them. The goal of this paper is not to assess which algorithm performs best but to simply point out the areas of improvement for each algorithm individually. It is true that a detailed algorithm inter-comparison would need a more homogeneous approach (datasets covering the same time period, subject to the same cloud detection algorithms etc.), and that this may be a further validation step once the individual algorithms have become mature.

6. Thirdly, the statistical analysis methods used in the paper should be presented and explained more clearly (e.g. equation 4, see specific comments).

[Full Screen / Esc](#)[Print Version](#)[Interactive Discussion](#)[Discussion Paper](#)

This modification has been made in the revised version of the paper.

7. Finally, many of the differences that are found between the satellite data and the FTIRs are washed away as being non-significant. For example: figure 8 shows the biases in CH₄ between the satellite retrievals and the FTIRs, and these are concluded to be 'very small' (P2693, L7). However, if these biases were related to the overall global variation in CH₄ concentration, the conclusion would actually be that they are very large.

Here the referee has a point and modifications have been made in the revised version of the paper that among other points states that the biases are indeed significant. Furthermore, in stead of listing the standard deviations on the biases we now list the standard errors in order to better asses the significance of these biases.

8. Summarizing: on the basis of this study, I am not convinced that 'the products are useful for qualitative geophysical studies on a global scale' (P2679, L15-16), although I'm not completely sure what this statement means. If the analysis is improved along the lines above, there is a good chance that some more specific conclusions can be drawn.

The paragraph with this statement has been completely rephrased to better point out the limitations (which have certainly become clear in this article) of the SCIAMACHY data products. However, in the case of CH₄, we observed no latitudinal dependence of the bias, nor any systematic deviations from the seasonal variability. Thus, while it is true that the precision target (1 percent) is far from achieved, coarse qualitative studies which identify regional differences in CH₄ concentration only, require far less stringent criteria on the scatter and are therefore certainly feasible. The same applies to CO for which the data user, when taking into account the issues raised in this paper, can use the data for relatively coarse qualitative studies such as the detection of large biomass burning events.

Specific comments 9. P2682, L9-10: *The inter-hemispheric gradient in CH₄ columns*

[Full Screen / Esc](#)[Print Version](#)[Interactive Discussion](#)[Discussion Paper](#)

should not be compared with the gradient in surface concentrations. Also: could the authors give a reference of models which predict a lower gradient?

The referee is right that the inter-hemispheric gradient in CH₄ columns cannot be compared with the gradient in surface concentrations; we had done so in the absence of comparable data. To overcome the problem, we have since then used model (TM5) data to verify the interhemispheric gradient in the FTIR CH₄ columns, and we have observed a very good agreement. This new information is now integrated in the paper.

10. P2683, L4-7: Is it not better to perform the polynomial fit on the individual data points? When first a daily average is taken, a day with only one measurement will receive the same weight in the fitting as a day with many measurements (which has a more precise daily average).

Yes, but since for several stations only daily averaged data was available, all datasets have been averaged first for consistency's sake. The impact of the use of daily averages on the end results was verified and turned out to be negligible.

11. P2683, eq. (1): I would prefer the use of a different notation, not using 'words' but symbols (see also later comments).

This modification has been made in the revised version of the paper.

12. P2683, L28 / P2684, L1: How have these accuracies been estimated?

This is based on individual (per station) evaluations of the data accuracy taking into account the systematic error sources (mainly spectroscopy, some retrieval model parameters, ...) as well as on the results of comparisons between the retrievals at different stations (algorithm intercomparisons).

13. P2684, L6: The target molecules (except for CO) are well-mixed; their concentration is more or less constant in the entire troposphere. Therefore, the total column scales with surface pressure in first order.

[Full Screen / Esc](#)[Print Version](#)[Interactive Discussion](#)[Discussion Paper](#)

We agree. But since we did not dispose of the surface pressure data at all stations, we have used the geometric altitude of the stations for scaling.

14. P2684, eq. (2): Why have the authors used a scale height of 7.4 km here, whereas in table 3 a scale height of 8 km is mentioned in relation to the IMAP retrievals?

The 8 km scale height was internally used by the IMAP research group, in their cloud filtering algorithm only. All data in this article have been subjected to a 7.4 km scale height.

15. P2684, L17: I guess this 'auxiliary information' refers to surface pressure. Of course, it would be good to have surface pressure measurements at the FTIR stations. In this way, variability in sea-level pressure could be taken into account. However, for the validation of total column measurements (as opposed to column-averaged mixing ratios) from satellite it does not help a lot, since no surface pressure measurements for the satellite pixels are available. The present approach, normalizing to zero altitude, is thus fine, especially because the altitude of the FTIR stations and the mean altitude of the SCIAMACHY pixels are very accurately known. Still, the normalization could be done in a much more sophisticated way than equation (2), by using sea-level pressure and temperature profile information from meteorological models. Such refinements may not be needed for CO, but are probably necessary for the well-mixed greenhouse gases, which need to be measured with at least 1percent precision.

The auxiliary information also refers to VMR and Temperature profiles

16. P2684, L18: How do you know that the normalisation procedure induces possible errors 'up to 15percent'?

By evaluating the differences in the scaling factor when making different assumptions as to the VMR profile of the target gas, e.g., constant in the whole troposphere and zero above, or constant throughout the whole atmosphere etc.

17. P2685, L16-21: Why are data not available for the whole year for all algorithms?

[Full Screen / Esc](#)[Print Version](#)[Interactive Discussion](#)[Discussion Paper](#)

The large data gaps (in some cases only 4 months are available) make it even more difficult to draw any conclusions concerning e.g. the ability of the satellite measurements to capture seasonal variations. The authors are encouraged to try and get more complete data sets.

Such efforts are underway with the upcoming new versions of the retrieval algorithms but again we feel that, while the datasets were in some respects limited, useful conclusions could be drawn from them.

18. P2686, L13-15: Do I understand correctly that CO₂ is available from the IMAP algorithm? If so, why is it not evaluated?

It was available but given the extremely limited dataset of both the IMAP CO₂ and FTIR CO₂ available, no meaningful conclusion could be drawn from it. It was therefore not included in the article.

19. P2686, L24-26: What is meant with this sentence?

All WFM-DOAS (and IMLM) pixels with solar zenith angles >85 (80) deg are filtered out. Such high *sza*'s are typically encountered near the poles, and therefore, the number of coincident observations near the poles gets reduced.

20. P2687, L5-7: I suppose the 'appropriate correlative measurements' refer simply to surface pressure measurements? Are these really not available from the FTIR stations?

At present local surface pressure data are not standard included in the data files that have been submitted by the ground-based FTIR teams in the Envisat CAL/Val database and that we are using. We are currently in the process of collecting these data from the data providers.

21. P2688, L4: In two recent papers, Sussmann et al. (2005a and b) compared averaging kernels (AK) of WFM-DOAS and the FTIR at Zugspitze. While for CO the A_{ks} were quite similar, this was not the case for CH₄. Have the authors estimated the

[Full Screen / Esc](#)[Print Version](#)[Interactive Discussion](#)[Discussion Paper](#)

impact of these differences between the averaging kernels?

In both his studies, Sussmann et al. state that the averaging kernels of the ground based FTIR CO and CH₄ are similarly shaped as the SCIAMACHY kernels. In any case, according to these papers, the impact of these small differences in averaging kernels is negligible with respect to other error-inducing factors such as the scaling of the column data to zero altitude and the differences in averaging kernels between individual stations of the NDSC network.

22. *P2688, eq. (3): Continuing on my previous remark on notation, SCIA might be re-placed by xSCIA i,j , where j is the measurement counter on a particular day i .*

This modification has been made in the revised version of the paper

23. *P2689, L2: It would be helpful to mention these scaling factors explicitly in the paper.*

This modification has been made in the revised version of the paper

24. *P2689, eq. (4): This equation is hard to read, mainly because of the notation. What does this equation mean? If the authors' purpose is to 'evaluate the scatter in SCIAMACHY measurements themselves', why do they not follow the same procedure as for the FTIR, namely calculate the standard deviation of the SCIAMACHY data with respect to a polynomial fit through these data?*

The equation is rewritten in the form as suggested by the referee. It was our purpose to calculate the scatter around the 'real' seasonal cycle (in casu that given by the FTIR measurements scaled by the bias) instead of a potentially erroneous one affected by the SCIAMACHY retrieval algorithm.

25. *Moreover, equation (4) refers to scatter in daily-averaged measurements. In practice, this daily averaging is spatial averaging over a grid box around the FTIR station. The scatter in individual SCIAMACHY measurements will be larger.*

[Full Screen / Esc](#)[Print Version](#)[Interactive Discussion](#)[Discussion Paper](#)

This is now mentioned in the revised version of the paper.

26. P2689, L15-17: *Please give references for the desired target precisions.*

The following references have been added: ESA Study 15247/01/NL/MM 'The potential of spaceborne remote sensing to contribute to the quantification of anthropogenic emissions in the frame of the Kyoto Protocol', F.M. Bréon, Ph. Peylin, 2003. The changing atmosphere, An integrated global atmospheric chemistry observation theme for the IGOS partnership (IGACO), ESA SP-1282, Report GAW n°159 (WMO TD n° 1235), September 2004.

27. P2692, L2-4: *Could the authors spend some words on explaining the incredibly low P-values for all cases, even if, from R, there appears to be absolutely no correlation?*

This is mainly due to the large amount of data samples, which, even with a low correlation coefficient, make the hypothesis of absolutely no correlation extremely unlikely.

28. P2692, L12-23: *Do the authors have any clue as to the possible reason of this intriguing dependence of the bias on the total column?*

Due to the inherent complexity of the algorithms and the various sets of parameters used, the reasons could be manifold, a combination of factors and above all (for the moment) unclear.

29. P2693, L7: *As noted in the general comments, biases of the order of 5-10percent are large for CH4. Therefore, the characterization 'very small' should be removed.*

This has been modified accordingly in the revised version of the paper.

30. P2693, L8-10: *The fact that the FTIRs exhibit a variation of 3.3percent with respect to the polynomial fit, implies that SCIAMACHY data cannot be validated beyond this 3.3percent using the present validation method. This is a serious problem considering that a target precision of around 1percent is desired, and should be mentioned more clearly.*

[Full Screen / Esc](#)[Print Version](#)[Interactive Discussion](#)[Discussion Paper](#)

Some extreme outlier datapoints (which had previously gone undetected) have since been removed from the FTIR g-b dataset. This drastically lowers the erroneous 3.3percent scatter to 1.6percent. Once this 1.6percent mark is attained, other validation methods will indeed become necessary to check whether the day to day variability (which is included in the FTIR's 1.6percent) is accurately represented by the SCIAMACHY data products. In this article, no product came close to reaching this level. That said, a line is added in the final version of the article that comments this aspect.

31. P2693, L12-13: *Is 'closely followed' the correct terminology for a $(5.03 \text{ minus } 3.6)/3.6 = 40 \text{ percent difference}$? Is $(6.56 \text{ minus } 5.03)/5.03 = 30 \text{ percent 'only slightly larger'}$? What does the 'factor 2.3' refer to?*

This has been modified accordingly in the revised version of the paper.

32. P2693, L21: *Again, the differences between the retrieval algorithms are by no means 'minimal'.*

This has been modified in the revised version of the paper.

33. P2694, L12: *Do the authors mean 'statistically significant', or 'considerable'?*

Statistically significant

34. *Same question for lines 18/19.*

Considerable. This has been changed to 'considerable' in the text to avoid confusion.

35. P2694, L16: *Why are the biases for N₂O 'not significant'? They are at least large: of the same order as the global variation.*

Because the standard error on the bias is larger than the bias itself. Therefore the bias is not significant as the 'real' bias could well be 0.

36. P2694, L23-25: *There seems to be a clear seasonality in the bias of N₂O in Lauder, with a maximum in January and a minimum in June/July. Can the authors comment on*

[Full Screen / Esc](#)[Print Version](#)[Interactive Discussion](#)[Discussion Paper](#)

this?

This has been modified in the revised version of the paper.

37. *P2695, L3-7: I am not convinced by this study that the variability of in particular CH₄, CO₂, and N₂O can be detected by SCIAMACHY.*

While it is not possible to detect small emission variations, a coarse qualitative study of the large/small emission regions or the detection of biomass burning events is certainly possible for CH₄ and CO, since this requires far less stringent criteria than those listed in table 4. That said, this line has been completely reworded.

38. *P2695, L23-24: I do not believe that statements on the precision of the CH₄ measurements can be made on the basis of this study. Similarly for the other gases.*

see above

39. *P2696, L11-16: One would expect that spatial averaging reduces the scatter in the measurements. Thus, there should be less variability in the large-grid data than in the small-grid data.*

This is what is mentioned in the article, namely that higher variability caused by changes in column values at different locations within the sampling grid is offset by the larger dataset.

40. *P2697, L1-5: This is a weak point of the study. Of course, the conclusion of an intercomparison like this may be that the FTIR data do not allow to discriminate between the respective satellite retrieval algorithms. But the authors should at least try to get data sets that can be compared (cover the same time period, etc.). Otherwise they might as well focus on the evaluation of one single retrieval algorithm. Besides, I do think that some differences in the performance of the algorithms were detected. For example: the bias in CO dependent on CO itself was found for the WFM-DOAS and IMAP algorithms but not for the IMLM algorithm.*

Full Screen / Esc

Print Version

Interactive Discussion

Discussion Paper

Again, it was not our intention to do a detailed intercomparison to assess which algorithm works best. If this were the case a more uniform dataset would indeed have been essential.

41. *Table 3: 'Over land (altitude > 0)': does this suggest that land is characterized by an altitude > 0?*

Yes, the altitude is the parameters that we used to (in this case) make the discrimination land-sea.

42. *Table 3: 'weighted variance': what kind of weighting is meant here?*

Variance of the fit residual is simply the RMS of the residual structure while the weighted variance is somewhat different: The IMAP group performs a weighted least squares fit and weight it with the estimated pixel errors. If these are correct, the remaining residual structure is expected to be 1. Thus, the 0.1-10 criteria of the weighted variance of the fit residual to filter strange fits.

43. *References Sussmann, R. and Buchwitz, M., Validation of ENVISAT/SCIAMACHY columnar CO by FTIR profile retrievals at the Ground-Truthing Station Zugspitze, ACPD, 5, 557-572, 2005a. Sussmann, R., Stremme, W., Buchwitz, M. and de Beek, R., Validation of ENVISAT/SCIAMACHY columnar methane by solar FTIR spectrometry at the Ground-Truthing Station Zugspitze, ACPD, 5, 2269-2295, 2005b.*

These have been added in the final version of the manuscript

44. *Technical corrections P2683, L24: 'hereabove' -> 'above' P2691, L21: 'more positive' -> 'higher'*

While 5percent is more positive than -30percent, one could argue that the 5percent bias is in fact lower, hence 'more positive' in stead of 'higher'

P2692, L15/16: 'lesser underestimation': is this correct English? P2696, L26: 'associated to' -> 'associated with' Table 1: 'Coordinates' -> 'coordinates' Table 4:

[Full Screen / Esc](#)[Print Version](#)[Interactive Discussion](#)[Discussion Paper](#)

'SCIALACHY' -> 'SCIAMACHY' Figure 9: '... incremented by 50percent' -> '... incremented by 25percent'

All technical corrections, except the second ('more positive' -> 'higher) have been made in the revised version of the paper.

Interactive comment on Atmos. Chem. Phys. Discuss., 5, 2677, 2005.

Full Screen / Esc

Print Version

Interactive Discussion

Discussion Paper