The article compares the stratospheric vertical NO_2 distribution determined from the SCIAMACHY limb observations to those derived from the occultation instruments SAGE II, HALOE and ACE-FTS for collocated observations of the years 2003/04 (or 2004/05). Differences between the SZA of the limb and occultation measurements are accounted for by a photochemical correction of the occultation profiles.

The paper is of good quality, written good and the comparison of a 2 year dataset with a large number of collocations is able to provide a good validation.
However there is one important drawback:
Due to the strong annual cycle and function of latitude the comparison of just the global mean NO_2 profile is not very meaningful. Moreover it also has the disadvantage of possibly leading to paradox results (see comment below).
Therefore the comparison of the datasets should better be performed for data that is grouped into latitude bins and also sorted by seasons. This will provide the reader with the full information on the comparisons and also give the correct conclusions on the agreement.
Besides this main issue, there are smaller points, which should be improved (see general comments).
After correcting for these and including more meaningful plots and values for the comparison and validation, the article should be considered for publication after major revision.

General comments

1) One main quantity for the comparison and the conclusions of the article is the profile of the mean relative difference for all collocated profiles.
Stratospheric NO_2 has a large latitudinal and seasonal dependence. By reducing the comparison to the global and annual mean, the differences in the observation of these effects will cancel out.
Imagine you have two instruments measuring NO_2 profiles with 50 % difference for the northern hemisphere and -50% for the southern (and/or 50 % difference for the winter and -50 % difference for the summer season). By taking the global mean of all measurements, the negative and positive differences cancel out, and the mean relative difference is 0%, although it is 50% (or -50%) for every individual comparison ! On the other hand, if you have two instruments with relative differences between 0 and +20% for every individual comparison, the global mean difference would be approx. 10%, indicating a worse agreement than in the first example, although it is better for every collocation !
Thus, the mean relative difference does not say anything about the quality of the agreement for a species that variable like NO_2.
Although for SCIAMACHY and SAGE 2 (or HALOE or ACE-FTS) the observed NO_2 as function of latitude and season will not be exactly opposite like in the above example, it will also not agree perfectly. Therefore any positive and negative differences are weighted against each other, and this necessarily occurs also for the comparison of SCIAMACHY to SAGE 2.
One can also conclude that from the results shown in Figs. 8,9,10 and Table 1: relative differences between SAGE II and SCIAMACHY are positive for the tropics and negative for the mid latitudes and high latitudes. In the global mean this cancels out and the resulting difference is – for a wide altitude range – smaller than in any of the regions alone !
Surely, the authors agree that subtracting the differences for the tropics from the differences for the mid latitudes is not a valid approach, but exactly that happens when the global mean is calculated. This occurs also for the comparison to HALOE, with mainly negative differences for high latitudes and positive differences for mid latitudes and tropics.

Thus, the conclusion that the global mean of the profiles agree within 10 % is very misleading and wrong. Correct would be to say that – after cancelling out all positive and negative differences against each other – there still remains a discrepancy of larger than 10%.
Moreover, for the comparison of vertical profiles the difference is further reduced for the global mean, since the peak altitude of the profile also is a function of latitude and season. So, positive and negative differences due to different peak altitudes are also cancelled out against each other !
I think considering these arguments it becomes obvious that comparing the global means is not a good approach for validation purposes, at least not for NO_2 profiles (and especially when discrepancies for different latitude regions or seasons occur in opposite directions).
Therefore I suggest and recommend avoiding to use this global mean relative difference and not to draw any conclusions regarding the agreement of the observations based on it.
A more reasonable comparison would be a histogram, revealing how many individual collocated comparisons agree (as function of altitude) within 10%, how many within 20%, how many within -10%, within -20 %, and so on.
Also the standard deviation given for the mean profile and the mean relative difference is a much more conclusive and valid value for the agreement. The plotted standard deviation gives the range of values for which 66% of the profiles agree (if it is calculated in this way). This is for the reader a much more useful value than the global and annual mean where all kind of averaging out can occur (and does).
Or another suggestion: correlation plots of the NO_2 measured by limb and occultation for certain altitude levels, grouped as function of season and latitude bins. This would also give the full, detailed and correct information on the agreement.
If the authors want to keep the plots for the global mean difference, then these (or similar statistics and plots) need to be included (and discussed) additionally. And in that case, in the discussion of the relative difference for the global mean, especially when conclusions regarding the agreement are drawn based on this value, it needs to be pointed out more clearly that this value is used (see also detailed comments).
However, I hope the authors agree with my comment and follow the suggestion, to make with a little more effort statistics and figures that will provide the reader with better information on the observations and allow a more correct conclusion on the agreements and simply skip the plots for the global mean.
Please don't reply that the same wrong approach has been performed in other articles on comparison of NO_2 profiles derived from satellite observations.
This approach is valid for species like CO_2 that are more less globally constant, but not for NO_2 which increases by factors as function of latitude and varies by a few orders of magnitude due to seasonal effects. It seems that this issue has been overlooked for the review of earlier articles, but that can not be an argument to continue this wrong approach.

2) It is good that the authors divided the comparison into latitude bins. However, I think it is necessary grouping the data also as function of season due to the argument outlined above. Just like for the mean for all latitudes, this approach of averaging all data can again wrongly reduce the discrepancy and lead to erroneous results. Moreover, valuable information is lost or masked by this annual mean value. The information on the agreement for the different seasons is very important and should not be skipped, since much can be learnt from it.
Histograms and/or correlation plots as described above for latitude bins and seasons would give the information that is needed for such a validation study and should be easy to prepare from the dataset.
Also, the values for the southern hemisphere should be shown. At least for SAGE 2 there seem to be enough collocations. Showing the comparison only for the northern hemisphere raises the guess that agreements are worse for the southern hemisphere.

3) Some of the figures need to be improved (if the authors want to keep them):
For figures 10 b), d) and f) the lines for the relative differences are covered by the legends, such that sometimes the maximum is not visible.

4) The description of the retrieval approach can be improved.

5) For some quantities a description how they are calculated is missing. For example the standard deviation of the relative difference (the plots say: "rel. diff. + st. dev.") could be the relative difference for the standard deviation of the mean profile (which would be wrong) or, really the standard deviation of the relative differences (of all collocations).

6) Not all figures are introduced in the text with a proper description. Also, the reference to the figures is sometimes unclear.

7) "better than" , "considerably better " , "worsens " , "still better than" :
I think the article would improve if these qualitative slogans when describing the agreement are skipped. When the correct and meaningful numbers are provided, the authors and the readers may assess the agreement much better.

Detailed comments

Abstract

It should be mentioned already here that the studied NO_2 retrieval for SCIAMACHY is based on the "Information operator approach".

Line 8/9: "… retrieved from limb measurements … from the SCIAMACHY instrument … "
– better: from SCIAMACHY limb measurements

Line 15: "… for each year and instrument."
– the sentence begins with "For each instrument, …" so the second time instrument can be skipped

Line 16: "…found to be better than 10% …"
Better: "Agree within … "
Also it needs to be said that the values given here are for the mean of all measurements. However as outlined in the general comments, this value is very misleading and should be skipped. If the authors want to use this value they need to point out clearly that it is a global average for all seasons and should give the values for the more valid standard deviation also.

Chapter 1 (Introduction)

P4755, line 7/8 : N_2O is the major source for stratospheric NO_2, but this does not make it the major cause for ozone depletion ! The next sentence is (more) correct.
However, the term "ozone depleting gas" should be replaced by the more commonly used "ozone depleting substance".

P4755, line 22: HALOE is missing in front of corresponding bracket

Chapter 2

P4757, line 2: please make clear that the signal to noise ratio given here is for the spectra, not the NO_2 results.

P4757, line 17: NO2 → NO_2

P4757, line 28 : "While not part of the NO_2 retrieval … ": please motivate why the information about the SCODA cloud product is given here.

P4758, line 8 : " … spectral information from all spectral points …" : skip one spectral
Same sentence: "The data vector y … makes use of a technique similar to DOAS".
– How can a data vector make use of a technique ? Please rephrase.
And what is the difference to DOAS ?

P4758, line 23 : "… to get rid of … " : please rephrase

P4759 line 6 : after "implementation" there should be a ":" .
Also, it is "Ring spectra" not "ring"

P4762, line 4: "… at high latitudes, it is about 3.0x10^9 molec/cm^-3 at about 20 km altitude"
These values should be given more accurate: the peak is clearly above 20 km, the maximum value is approx. 2.8x10^9 molec/cm^-3
Maybe more important: NO_2 at high latitudes is strongly variable due to seasonal effects. The value given here may be correct for summer but not for the other seasons.
Also, this example for 77.5 °N can not be generalized, since number densities of NO_2 vary strongly within the region of the high latitudes (60 to 90°) also for one season.

P4762, line 6: " … figures mentioned previously, … " → figures 1 and 2

P4762, line 12: " … and show similar values as above."
– what is meant here ? please specify and rephrase

P4762, line 25: "in this case, …" – the meaning of this sentence is not clear, or can be understood only with much effort for the reader, please rephrase

P 4763, line 11: "… near the measurement tangent heights." – where are they ?

P4763, line14: below this 15 km – please rephrase

P4763, line 19: "In most cases, the altitude range sensitive to NO_2 … "
– an altitude range is sensitive for NO_2 ? You probably mean the instrument.

P4766, line 21. "To keep the collocations and as altitudes below 20 km are already difficult to analyze due to other errors described here …"
– better rephrase and specify

Chapter 3

P4767, line 15: HALOE, SAGE II, ACE-FTS: These abbreviations are introduced in chapter 1 already.

P4770, line 2: SCIATRAN → SCIAMACHY

P4771, line5: “Thus, seasonal differences might dominate in these comparisons.”
No, it is vice versa. When you perform the comparison for one season then there are no seasonal differences for this comparison. Instead, when comparing the annual mean, then this value is dominated by the averaging out of the seasonal differences.
So, this argument is not valid for excluding the results for the first half of year 2005.

P4772, line 5: “ … it worsens for the tropics in this case for all altitudes.”
– Please explain how this is possible. Does this indicate an error in the model for the diurnal effect error or its application on the comparison ?
I think this point is very interesting and should be studied and explained.
Also, remember that you are looking at seasonal means. Thus it is possible that the agreement for the mean gets worse, while the agreement for every season or every collocation in fact improves with the diurnal effect correction (compare the two examples in the general comment) ! It seems it is not unlikely that this is what happens here, but without the full seasonal and latitudinal resolved comparison one can not say this surely. However, the possibility of such an effect shows, how dangerous conclusions based on the mean profile for all seasons can be !

P 4772, line 23: “The collocation pairs feature lower absolute values in the other two regions and larger differences as compared to high altitudes.”
– how large are the differences ?

P 4774, lines 9-12: please rewrite and specify

P4774, line 20: something wrong in this sentence. I think “MRD values” after Table 1 needs to be skipped.

Chapter 4

Conclusions should be rewritten considering the comments above.

Figures

Figure 10 b), d) and f):
For these plots the legend needs to be outside of the axisplot, since the lines are not visible for all interesting altitudes, sometimes even the maximum of the rel. differences is covered by the legend.

Figure 11 and 18: for plots b), c) and d), e) x-axes are overplotted by each other.