

## ***Interactive comment on “Global and long-term comparison of SCIAMACHY limb ozone profiles with correlative satellite data (2002–2008)” by S. Mieruch et al.***

**Anonymous Referee #3**

Received and published: 26 September 2011

The paper discusses comparison and validation of SCIAMACHY satellite measurements of ozone profiles in the stratosphere using independent satellite measurements. The validation is based on using both co-located measurements and zonally averaged profiles. The long-long term comparison relies on time series analysis including trend fitting. Statistical significance tests are applied to draw conclusions. The validation of the satellite measurements is very important topic and a crucial step before data can be used for science. The proposed method to study possible drifts in the long time series by comparing the trends with correlative data is interesting. The topic is very relevant to AMT. The manuscript is generally well written and it is easy to read. However, there

C1664

are some weaknesses in the manuscript which should be taken into account before publishing the work.

My main concern is related to the methodologies presented:

1) When validating data it is important to take the error estimates of the data into account. This is particularly important when discussing the significance of the results. In the manuscript the error estimates (random & systematic) of SCIAMACHY data are not discussed at all. I would also like to see some discussion about the error estimates vs natural variability.

2) The resolution of the measurements vary from instrument to instrument and this should be taken into account in the comparisons or at least commented if it does not play major role. The resolution of the solar occultation measurements seems to be much better than SCIAMACHY.

3) The statistical significance tests should be performed using extra attention in order to avoid false conclusions. In general, I find Bayesian statistical approach, with proper error characterization, more solid background for making conclusions than various statistical tests (but this is only my personal view). It is in particular important to distinguish what we are looking for: in the data validation we are typically more interested about finding systematic biases than ‘statistically significant results’. As an example, bias at one altitude might not be statistically significant, but when the same behavior can be seen at several altitudes the systematic behavior is clear and may indicate some instrument/retrieval issues. Would it be possible to discuss the purpose of the significance tests in this light? In addition, the error estimates should be taken into account when applying significance tests.

4) I find the conclusions that statistically significant trend is observed at 2002-05, but not at time period 2004-08 a bit confusing as the time periods overlap with 2 years.

Minor comments:

C1665

P4870 -L19 and elsewhere: I don't think that for modern computers it is time consuming to find co-located measurements. In my opinion, the work could be motivated by obtaining better statistics when using zonally averaged comparisons. At high latitudes during eg. ozone hole conditions this is more complicated and representativeness might play role. This could also be commented.

P4874 – L6. Difference in local time: what does it mean for ozone in the stratosphere? Is it better to compare night-day?

P4876 – L14. The sentence is unclear.

P4881 eq (12) – this seems to be variability and not errors. The error estimates of the data are not included as far as I can see.

Fig 6: random errors or variability?

P4882 – L8. Here m deg of freedom, but above month. Please, use other letter here for consistency.

The impact of solar cycle should be discussed in the context of observed/not observed trends.

Figures 9-11 are rather unclear to interpret. Without the significance tests I would draw conclusions based on (connected) areas that are similar (or smoothly varying) which would indicate systematic behavior. The significance tests don't seem to agree with these areas and that is somewhat puzzling.

I would like to encourage the authors to include the other Envisat instruments (GO-MOS and MIPAS) to the comparison since they have the exactly same time period of measurements. This would probably improve the interpretation of the results.

---

Interactive comment on Atmos. Meas. Tech. Discuss., 4, 4867, 2011.