

Review Kramarova et al, ACPD 13, 2013.

General comments

Kramarova et al present a thorough validation analysis of the SBUV V8.6 monthly zonal mean data set, using various satellite and ground-based instruments as reference. The data set, preprocessing, and comparison methodology are well described in the paper. The authors made a comprehensive analysis of the spatial dependence of bias and drift of each SBUV(/2) instrument. Such information is valuable and required prior to merging the individual SBUV data records.

When the comments below are addressed I would recommend publication.

Specific comments

- p2251, par1: Add a note on the required detection level for the O3 trend. This would help evaluate whether the SBUV instruments are sufficiently stable to detect such a trend. Right now, at the end of the paper, an SBUV instrument is called stable when it is drifting less than 1% per year. Which is quite a lot in my opinion with respect to the actually expected O3 trends.
- p2553, par3: Add the values of the 21 SBUV pressure levels. They are likely mentioned in (Bhartia et al., 2012), but it would certainly be useful to mention them here as well.
- p2554, par1: Add a reference with the background and motivation of the SBUV data screening.
- p2554, par1: Clarify what happens if the mean latitude of O3 profiles is not within 1 degree of the band centre. I assume the monthly zonal bin is discarded.
- p2551, par1: Has the bias between ascending/descending O3 profiles in a (latitude, month) bin been investigated?
- Sec 2.2: Add **bias and precision** estimates for each of the reference instruments, sometimes none or only one is mentioned. Given the importance of the study of SBUV stability in this work, it is imperative to mention **drift** estimates for the reference instruments as well.
- Sec 2.2.2: (Livesey et al., JGR 2003) recommend a) not to use UARS MLS V5 data after mid-1998 for trend analyses, and b) warn to be cautious with data from mid-1997 to mid-1998. MLS data from this period is used in this work, at least a comment on the possible instability of MLS data is in place.
- Sec 2.2.3: I assume the vertical oscillations seen in Tropical UTLS Aura-MLS V3.3 data are not too relevant for your work, given that ozone is integrated over 250-16hPa. Is that correct?
- Sec 2.3: Add a short description of the ozonesonde data used to validate the tropospheric column (Sec. 4).

- Sec. 2.3.2: Please add a (more recent) reference to a study of the bias and drift of 6 NDACC lidars (4 of which used in your work): Nair et al., 2012, AMT, doi:10.5194/amt-5-1301-2012.
- p2558, l15: The resolution of lidars worsens with altitude, it reaches 3 km at 45 km altitude.
- P2558, l18: Mention that due to the 10 % error screening fewer measurements are available at bottom and top (above ~5 hPa) of the lidar profile. This information is useful in Sec 2.4.3.
- p2560, l1: SAGE-II and lidar data are used over long periods (>15 years) when considering the validation over all SBUV instruments. But not when single SBUV validation is done. Please clarify this, as the statement can be misleading.
- Sec 2.4.2: The vertical resolution of the microwave instruments is worse than that of SBUV. Did you quantify the impact of comparing SBUV-MWR partial columns finer than the MWR-resolution?
- p2561, l28: SBUV profiles are weighted with distance from the correlative profiles. Is this spatial distance, or is a temporal component included as well? If yes, how? Do you have an idea of the horizontal smoothing error contribution?
- Sec 2.4.3: Clarify the space/time collocation criteria. Especially the spatial window is not clear to me. What I understood:
 - SAGE II: SBUV within ($\pm 1^\circ$ lat, $\pm 14^\circ$ lon, same day)
 - MWR: SBUV within ± 1.5 h at Mauna Loa and same day 9AM-5PM at Lauder
 - other: SBUV within ± 12 h
- Sec 2.4.4: The bias is calculated as mean of absolute differences relative to a fixed x_a (I assume monthly zonal values?). While it could be calculated as mean of percent differences, where every term is relative to \widehat{X}_{ext} . Do you expect large differences? And is the a-priori reference more suitable in that case?
- Sec 2.4.4: Eq. (2) represents the (biased) standard deviation of the absolute difference $\widehat{X}_{sbuv} - \widehat{X}_{ext}$, **not the standard deviation for the relative bias b** as defined in Eq. (1). **In general, I found it sometimes difficult to follow which standard deviation is referred to:** is it sample standard deviation σ (from Eq.2), or bias standard deviation $\frac{\sigma}{\sqrt{N}}$ (with σ from Eq.2)? I assumed all results/figures refer to the standard deviation of the bias. In that case I would replace Eq.2 by the expression for bias standard deviation and mention in the text that this estimator will be used throughout the rest of the text.
- Sec 2.4.4: How do you calculate the standard deviation of the relative differences? Is it $\frac{\sigma}{x_a}$ (with σ from Eq.2)? In that case, the standard deviation of the relative bias would become $\frac{\sigma}{x_a \sqrt{N}}$?
- Sec 2.4.4: The explanation of the drift calculation should be clearer. Do you deseasonalize the SBUV and EXT timeseries separately, compute differences, and then regress? Or do you compute the differences, deseasonalize SBUV-EXT and then regress? I understood the 1st method, but the text is not entirely clear on that (the next phrase mentions that “the anomalies (i.e. SBUV-EXT) are deseasonalized”).
- p2563, l18-20: Did you check whether the fit residuals are Gaussian?

- Sec 3.1, par2: The seasonal signature (1st four lines) is also discussed in Sec 3.1, par 8. Maybe move these lines to the end of the section?
- Sec 3.1, par2: The “shorter overlap” sounded strange at the first reading, since SAGE and UARS-MLS have a similar overlap as Aura-MLS for individual SBUV (>~5 years). The poorer spatial and temporal sampling of SAGE and UARS-MLS leads to larger (>2x) standard deviations in the differences (evident from Fig. A1), which would make it more difficult to discover a seasonal cycle in the differences. Error bars are larger, but I did not really find that Fig. A9-10 are inconsistent with Fig. A8.
- Sec 3.1, par2: In addition, I found it hard to see the seasonal cycle from Fig. 3, with its large temporal scale. Could you add a reference to the more useful figures A8-10?
- p2565, l22-24: Fig. 3 suggests that the negative N11 bias for 4-2.5 hPa could be is mainly built up after ~1997, in the descending phase of the orbit. Is that correct?
- Sec 3.1, par7: Are the larger std. dev. for UARS MLS and SAGE II larger due to a smaller sample size? (see also comment on *Sec6, par2*).
- Sec 3.1, par8: Partly mentioned in par2.
- Sec 3.2: Add references to the relevant timeseries and standard deviation plots in the Supplement.
- Sec 3.2.2, par4: The vertical structure of the bias standard deviations is explained by the lower lidar precision above 2-5hPa, and the fact that fewer profiles enter the bias calculation due to the 10% lidar precision screening.
- Sec 4, par3: The qualitative structure of SBUV (Fig.10) and sonde (Fig.11) is quite similar. Could you add a comment on that?
- Sec 4, par3: There is no real description of the ozonesonde data, nor of the preprocessing + collocation criteria. These should be added in Sec 2.3, 2.4.1 and 2.4.3.
- Sec 5.1, par2: It is difficult to quantify the drift and its error bar from Fig 12.
- Sec 5.1, par3: Drifts are larger relative to SAGE II, but these are insignificant due to the larger error bars. Please clarify this in the text.
- Sec 5.2, par2: Is the “mean drift” the mean of regression at individual stations, or the regression of mean time series?
- Sec 5.2, par2: Do the (2 sigma?) error bars in Fig 13-14 represent std. dev. of mean drift or the std. dev. of the drift sample?
- Sec 5.2: The magnitude and vertical structure N16-N18 drift results are in good agreement for the Aura-MLS and microwave comparisons (to slightly lesser extent for lidars). Could you add a comment on that?
- Sec 6: Could you add a comment on the latitude dependence of the bias and drifts for some instruments?
- Sec 6, par2: Are you discussing standard deviations of the differences, or the standard deviations of the bias? If it is the latter, is the larger bias std. dev. for SAGE and UARS-MLS mainly due to a smaller sample size? (see also comment on *Sec 3.1, par7*)

- Sec 6, par6: How stable does the instrument need to be in order to detect the expected stratospheric O3 trends?
- References: Please update (if any) the status of the papers in preparation/discussion/review.
- Fig 1: Add short phrase in caption that SAGE and MLS instruments are shown as well.
- Fig 12: I assume that drifts estimates are shown for each 5° latitude band, while the error bars are only shown for 5 particular bands. Is that correct? Please clarify this in the caption. Upon 1st reading one could think that the drift results are for 20° lat bands.
- Fig 12: In general, it is difficult to read magnitude of drift and error from the plot, since so many lines are superimposed. Is there a possibility to improve this figure, e.g. by slightly offsetting (in X) the error markers for the different SBUV instruments?
- Fig 12: Replace “Percent” → “Percent per year” in label on Y-axis.
- Fig 13-14: Are error bars 1 sigma or 2 sigma? Mention this in caption. See also comment *Sec 5.2, par2*.
- Fig 14: Replace “Drift, %” → “Drift, % per year” in label on X-axis.
- Appendix: Figures have different labels (A.xx) than the references in text (S.xx), please fix this.
- Appendix: Add y-axis label for Figs 2-4 and 8-13. Replace “Drift, %” → “Drift, % per year” for Fig 15-17.
- Appendix, Fig 6: “Larger deviations were detected for the upper layer due to the reduced number of lidar observations.” See comment *Sec 3.2.2, par4*: the increased std. dev. at the top is due to lower precision of lidar and the (subsequently) reduced number of observations.
- Appendix, Fig14: Not discussed in the text. If this is not planned, it should be dropped from the Appendix.

Technical corrections

p2250

- l2 : We present **the** validation
- l8 : on board **the** UARS and Aura satellites
- l17 : We also estimate **the drift of** the SBUV
- l21 : creating **a** merged SBUV data set
- l24: and **start** to fall (2x begin close to each other)
- l26: in the **middle of** the century, between

p2251

- l2 : observations, **very well** calibrated
- l19-21: **The data are consistent with that from SBUV/2 instruments on NOAAs 17, 18, and 19, even though the calibration is still being finalized.**

p2252

- l10: the newly processed v8.6 SBUV **monthly zonal mean** profiles
- l24: **The** Brion-Daumont-Malicet ozone cross sections
- l26: were used **instead** of

p2253

- l2: based on ~~the~~ Aura MLS and ozonesonde
- l4-5: and **were used to reprocess ozone data for all instruments**
- l6: absolute radiance **calibrations** for
- l7: **The calibration** for N4
- l24: Define PMF, only occurrence in text.
- l25-26: rather than **for** day-to-day

p2554

- l1: all **level 2** ozone profiles
- l16: through **the end of** 1992
- l21: We found **that** the main source
- l22-23: which **is** the error due to **vertical variability in the ozone density** that the SBUV

p2555

- l18: In this section we **a** provide brief description of each independent **dataset**.
- l21: Add **ERBS** as satellite platform (mentioned for MLS and SBUV instruments as well).

p2557

- l13: remove **~** in front of about

p2558

- l11: 34.4°N → **34° N** (same precision as for other stations)

p2559

- l19: except the bottom layer **where layer 0 and 1 were** combined

p2560

- l12: Analysis of **the** corresponding averaging
- l13: **four** independent pieces

p2562

- l7: “Appropriate coincidence criteria in both time and space are very important for validation.” → But what are space criteria? (see *Specific Comment*)

p2563

- l2: number of coincident **mzm** profiles
- l4: standard deviation for the **relative** bias is estimated using

p2564

- l1: mean biases and standard deviations **in the middle and upper stratosphere** (this would clarify distinction with Sec 4)
- l26: we do not **recommend** to

p2566

- l17: different label as caption in Supplement
- l25: Fig. A.1 shows **1-3 %** rather than 1-5 %

p2567

- l25: Mauna Loa biases are **negative** above 4 hPa

p2568

- l8: refer to Fig 5 in supplement

p2570

- l10: at all locations points **to a** systematic error
- l11: causing Umkehr instruments **to** underestimate ozone amounts

p2571

- I15: The biases for the 250–25 hPa layer are **negative, from 0 to –2%**, outside **of** the tropics.
- I24-26: This example demonstrates the increase in the precision of SBUV measurements in the tropics when **the vertical resolution is downgraded by combining the layers up to 16 hPa.**

p2572

- I27: Three **of** SBUV instruments

p2573

- I7: as **function** of latitude
- I8: which corresponds to **a confidence level of 95%.**
- I23-24: **–2.5 % yr^{–1} to +2 % yr^{–1}.**

p2574

- I1-2: where drifts **are larger.**
- I10: and find **consistent** results

p2575

- I6: after the **upgrade of the lidar** in 2001. (in Sec 3.2.2 you mention *upgrade* not repair)
- I18: **The slightly** larger drifts for N18 are **insignificant, most likely** due to the shorter overlap periods.

p2576

- I4: We validate SBUV monthly zonal mean profiles **from NOAAs 4, 7, 9, 11, 14, 16, 17 and 18** against independent

p2577

- I16-17: **Larger and significant** drifts (more than $\pm 1\% \text{yr}^{-1}$) are detected for the ascending

p2594

- I3: and **Payerne** (47° N)

Appendix, p2

- I5: This points **to** a good correspondence

Appendix, p3

- I2: The **dispersion** of differences

Appendix, p8

- I6: in both **extratropics**, with a 6-month