Response to reviewer #1:

We thank referee 1 for a thorough review. The manuscript is certainly improved in many respects thanks to the referee's comments.

"As mentioned further below, I feel that some more caution might be required before excluding the solar cycle."

Please see response to the comment "mentioned further below".

"Overall, the text seems a bit wordy and lengthy, but I have no simple solution how the text could be shortened."

The text was shortened in several spots thanks to suggestions from reviewers 2 and 3. We have eliminated the following sentences regarding solar variability:

Per 100 units of 10.7 cm radio solar flux, the change in ozone is (0.88±0.79)%, in agreement with the ~1% found by Randel and Wu (2007) in a pocket of statistically significant solar signal near 25 km using SAGE II data at low latitudes. Similarly, a statistically significant 1% response from solar minimum to solar maximum was also found by Soukharev and Hood (2006) in a narrow vertical range near 28 km at low latitudes and later confirmed with improved regression modelling (Hood et al., 2010).

"*One important thing that is missing in the paper is a discussion how the uncertainty margins / confidence intervals were obtained. This is important, because only (roughly) correct uncertainty margin allow correct assessment of the significance of trends and other results. Did the authors account for AR1 auto-correlation in the residuals? Autocorrelations in the residuals usually mean that the uncertainty estimates from standard linear regression are too small. When auto-correlations in the residuals are accounted for, the uncertainty margins usually increase (e.g. Weatherhead et al., 2000). What about longer auto-correlations in the residuals (see also Vyushin et al. 2010)?"*

The 95% confidence interval is obtained using the MATLAB regress routine. Specifically, it is the difference between the coefficient estimate and its lower confidence bound. The confidence interval is assumed to be symmetric about the fitting coefficient. We did not account for the AR1 auto-correlation in the residuals in the original paper. In the revised manuscript, the uncertainty margin takes into account the AR1 auto-correlation in the residuals using:

$$\varepsilon'_{c_1} = \varepsilon_{c_1}\sqrt{\frac{1+\phi}{1-\phi}}$$

where ε'_{c_1} is the autocorrelation-corrected uncertainty on the linear-trend fitting coefficient. Longer auto-correlations are not significant in the merged dataset except at 18.5 km but still there is a significant decrease in the two month autocorrelation to 0.54, which is smaller than the expected AR2 auto-correlation of $0.777^2 = 0.604$ based on the AR1 auto-correlation coefficient of 0.777. The OSIRIS dataset has AR2 autocorrelations of ≤0.36, with this maximum value again at 18.5 km. The SAGE 2 dataset has AR2 autocorrelations of ≤0.29 with a maximum value at 19.5 km. Because the AR2 auto-correlation in the residuals is weak, we did not consider longer auto-correlations in the residuals. The autocorrelation correction of the linear trend uncertainties is calculated after the final model is determined at each altitude, i.e., it is not calculated prior to that point. We have added a Section 2.3 in which we now write:

The uncertainty margin takes into account the AR1 auto-correlation in the residuals using:

$$\varepsilon'_{c_1} = \varepsilon_{c_1}\sqrt{\frac{1+\phi}{1-\phi}} \quad (6)$$

where ε'_{c_1} and ε_{c_1} are the uncertainty on a fitting coefficient with and without accounting for autocorrelation of the residuals, respectively, and ϕ is the autocorrelation between time series shifted by 1 month. The scaling of the fitting coefficient uncertainties to account for auto-correlation is calculated after the final model is determined at each altitude, i.e. it is not calculated prior to that point.

"*Also, the abstract and several places in the text should include more cautioning words about the significance of the trends. Figure 4 is especially revealing in that aspect: The large negative trend e.g. at 18.5 km over the entire data period very much depends on the high SAGE II values around 1990 (near solar maximum, see also the higher values around 2000 to 2004, the next solar maximum). The continued negative trend (e.g. at 18.5 km) during the 2001 to 2012 OSIRIS period also depends very much on the few high values around 2002 to 2004 (near solar max) in that record. I think this needs to be recognized in a few places in the text, and also in the conclusions and in the abstract.*"

We agree that the high values around 2002 to 2004 could have a strong effect on the trend. But high values in the 2000 to 2004 time frame would lead to a positive trend from SAGE II and the merged dataset because these years lie in the latter half of both records. The SAGE II values around 1990 are not so high, particularly at 18.5 km where the trend magnitude is largest. For example, the SAGE II ozone number density in August 1999 is actually higher. Also, the peak in the observed ozone anomaly is highest in 1989. The solar cycle effect is minor relative to the contribution by ENSO (p16681 lines 12-15) as 1988-89 and 1999-2000 where major La Nina events. The large changes in ozone anomaly over short periods at these times are indicative of shorter term variability (ENSO) , not variations due to the solar cycle. The reviewer is trying to eyeball a solar signal by identifying two ozone maxima that are 11 years apart but are mostly a response to a separate phenomenon. The expected peak periods would be 1991 and 2002 if these high values were predominantly related to the solar cycle (assuming no lag). We doubt that the higher values near 1990 have a strong influence on the trend. Also, the inclusion of the solar cycle as a basis function still leads to a significantly negative trend at 18.5 km.

In the conclusion, we now write:

The difficulty in fully capturing the interannual variation of the seasonal amplitude in the 2002-2004 time frame could affect the trend determined for the OSIRIS time period (2001-2012) and to a lesser extent, the trend in the merged dataset.

Also, in section 3, we now write:

Understanding what controls the year-to-year variability of the seasonal cycle (Ploeger et al., 2012; Witte et al., 2008) could lead to an improved regression model and have an impact on estimated trends and their uncertainties.

In the abstract, we now write:

Analysis of the merged dataset (1984-2012) shows a statistically significant negative trend at all altitudes in the 18-25 km range including a trend of (-4.6±2.6)% /decade at 19.5 km where the relative standard error is a minimum. Underlying strong fluctuations in ozone anomaly due to El Nino Southern Oscillation, the altitude-dependent Quasi-Biennial Oscillation, and tropopause height need to be taken into account to properly determine the trend. Improvements are suggested for future regression modelling efforts which could reduce trend uncertainties and biases in trend magnitudes.

<u>Detailed comments</u>

Introduction: I think somewhere in the introduction it should be mentioned why lower stratospheric ozone trends in the tropics are interesting. Connection with modelled changes in Brewer Dobson circulation. Randel and Thompson 2011 paper.

We now write:

Trends in the tropical lower stratosphere are of interest given modelled changes in the Brewer-Dobson circulation (e.g. Bunzel and Schmidt, 2013), and with the significant negative trend in this region observed over the last quarter century using a combination of SAGE II and ozonesonde data (Randel and Thompson, 2011).

Bunzel, F., Schmidt, H.: The Brewer-Dobson circulation in a changing climate: Impact of the model configuration, J. Atmos. Sci., 70, 1437-1455, 2013.

pg 16664, line 19, pg 16665 line 17, pg 16669 line 22: It would be good to give sources / URLs for the used data sets.

The URLs are now given:

SAGE II v7.0 ozone profile data (available at https://eosweb.larc.nasa.gov/cgi-bin/searchTool.cgi?Dataset=SAGE2_AEROSOL_O3_NO2_H2O_BINARY_V7.0) …

and

(available at ftp://odin-osiris.usask.ca/Level2/daily)

The website for the ENSO data is given in the References section:

www.esrl.noaa.gov/psd/enso/mei/table.html

Why has the filtering a large effect? Is not the main point that the SAGE data are heavily contaminated and not usable in periods/ regions with large aerosol loading?

The filtering removes contaminated points which would otherwise lead to a biased ozone anomalies and trend. A significant fraction of the SAGE II monthly data points are filtered near the tropopause.

pg 16666 line 6: Maybe add reference Adams et al. 2013?

Done.

"anomaly period" should probably be "overlap period"?

Corrected. Thanks.

pg 16675 line 6: It would be good to give a reference for the used elimination / regression procedure.

There is no reference for the exact procedure. We believe the description in this paper is nearly sufficient to allow the reader to repeat this work. We added one sentence at line 26 of p16675:

After adjusting the lag, the stepwise elimination procedure is repeated using the criteria listed above.

"1/(1/0.5 ± 12/27)" for the expected periods. The use of years in the first term and months in the second term confused me. Maybe better to use "1/(12/6 ± 12/27)"

Done. Thanks.

pg 16670, lines 7 to 14: One thing that is not mentioned, and might be important, are possible correlations / similarities between tropopause pressure and ENSO. I would expect that there is something, and the authors should check and comment on that. (relevant also for a few other places in the text).

This is mentioned in line 14 of p16670. We have added a comma before "and a slight anti-correlation (r=-0.2) with ENSO (…)" to help clarify that we are providing the correlation between tropopause pressure and three other predictor variables (aerosol extinction, QBOa, and ENSO).

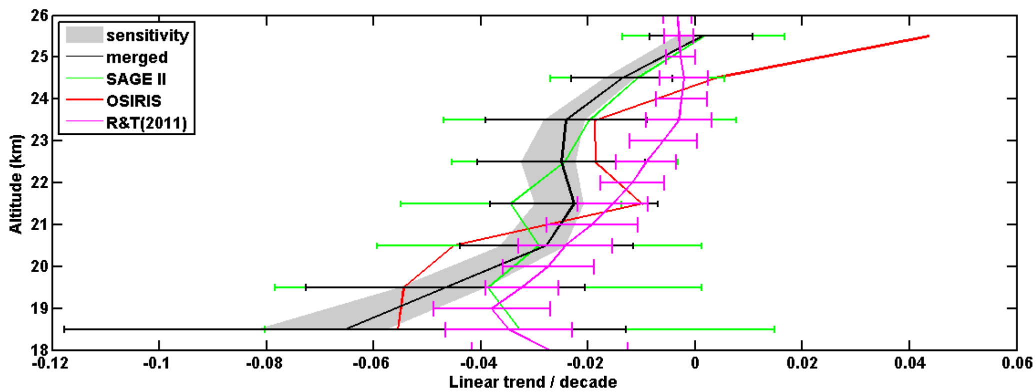
pg 16678, line 6: As mentioned above, a discussion of the resulting uncertainties, auto-correlation in the residuals, etc. is missing here.

We have inserted a paragraph regarding this matter in a new section titled "Discussion":

Regarding the autocorrelation of the merged dataset given our choice of a monthly time step, we find that there is no obvious altitude dependence to the autocorrelation, except for higher autocorrelation at 18.5 km (ϕ=0.777), whereas the other altitudes are in the ϕ=0.56±0.05 range for AR1. For a two month lag, ϕ is less than 0.32 at all altitudes, except again at 18.5 km. If the inter-sensor bias between SAGE II and OSIRIS at 18.5 km is not equal to the bias during the overlap period (e.g. due to instrument degradation), a high autocorrelation will result in the merged dataset. At 18.5 km, the merged dataset has a much higher AR1 auto-correlation than for the individual SAGE II and OSIRIS datasets (0.26 and 0.65, respectively). The auto-correlation in the merged dataset is not much larger than in the individual datasets at higher altitudes. The autocorrelation statistics suggest that a 2-month sampling increment should be considered in future work.

pg 16679, lines 14 to 24: I think the comparison to the most relevant paper (Randel and Thompson 2011) is missing here! In fact, I think the trend result from Fig. 8c of Randel and Thompson (2011) should also be plotted in Figure 3. Also the results for QBO and ENSO from Table 2 should be compared / put in the context of Randel and Thompson (2011).

We compare with Randel and Wu (2007) and Forster et al. (2007) because we are comparing/validating our SAGE II trends with published SAGE II trends for the same time period. Randel and Thompson (2011) use the 1984-2009 time period and a much wider latitude band. We compare with merged trends with Randel and Thompson (2011) in the conclusion. We agree that the trend result from Randel and Thompson (2011) would make an interesting addition to our original Figure 3, so here it is:

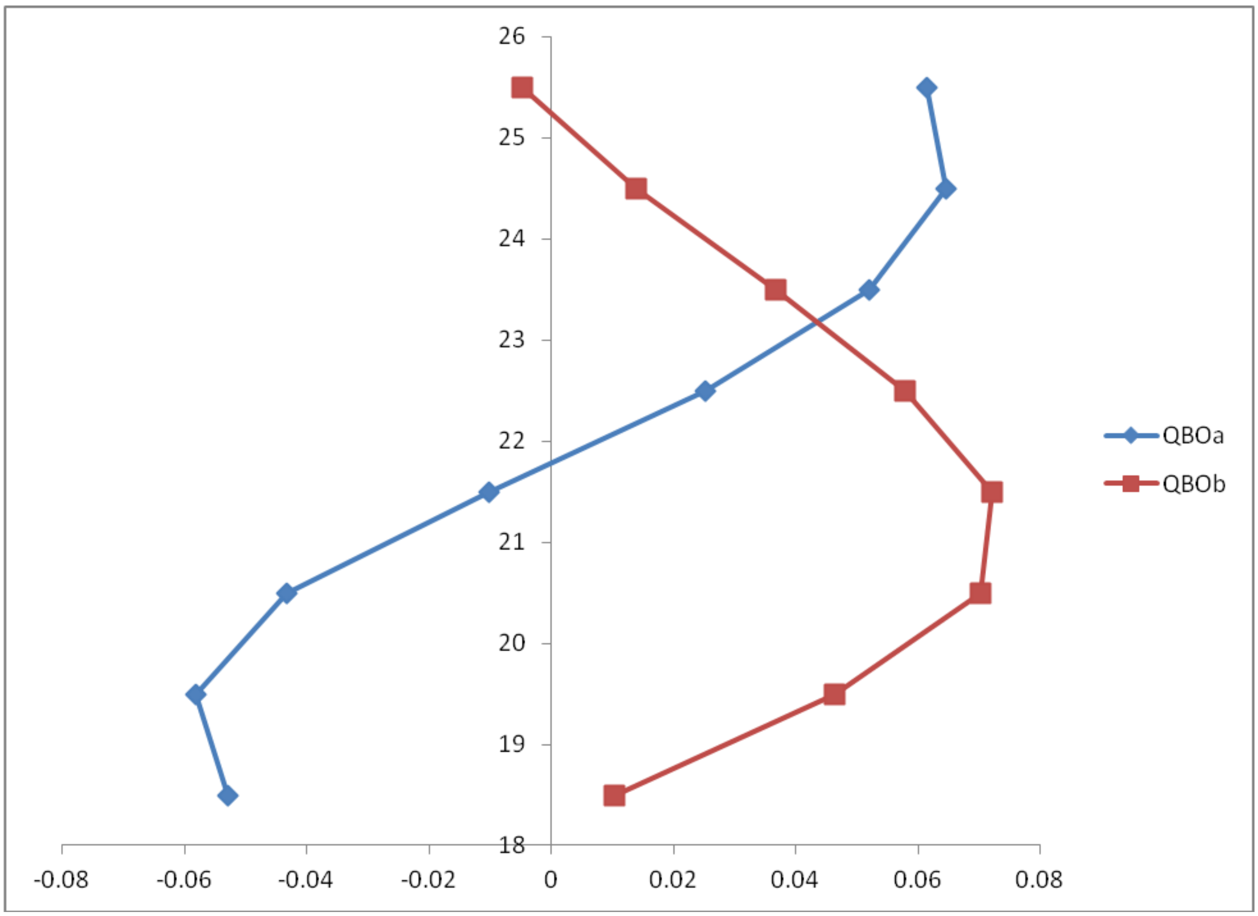


We now move test from the Conclusions to Sect. 3.1 and add a sentence:

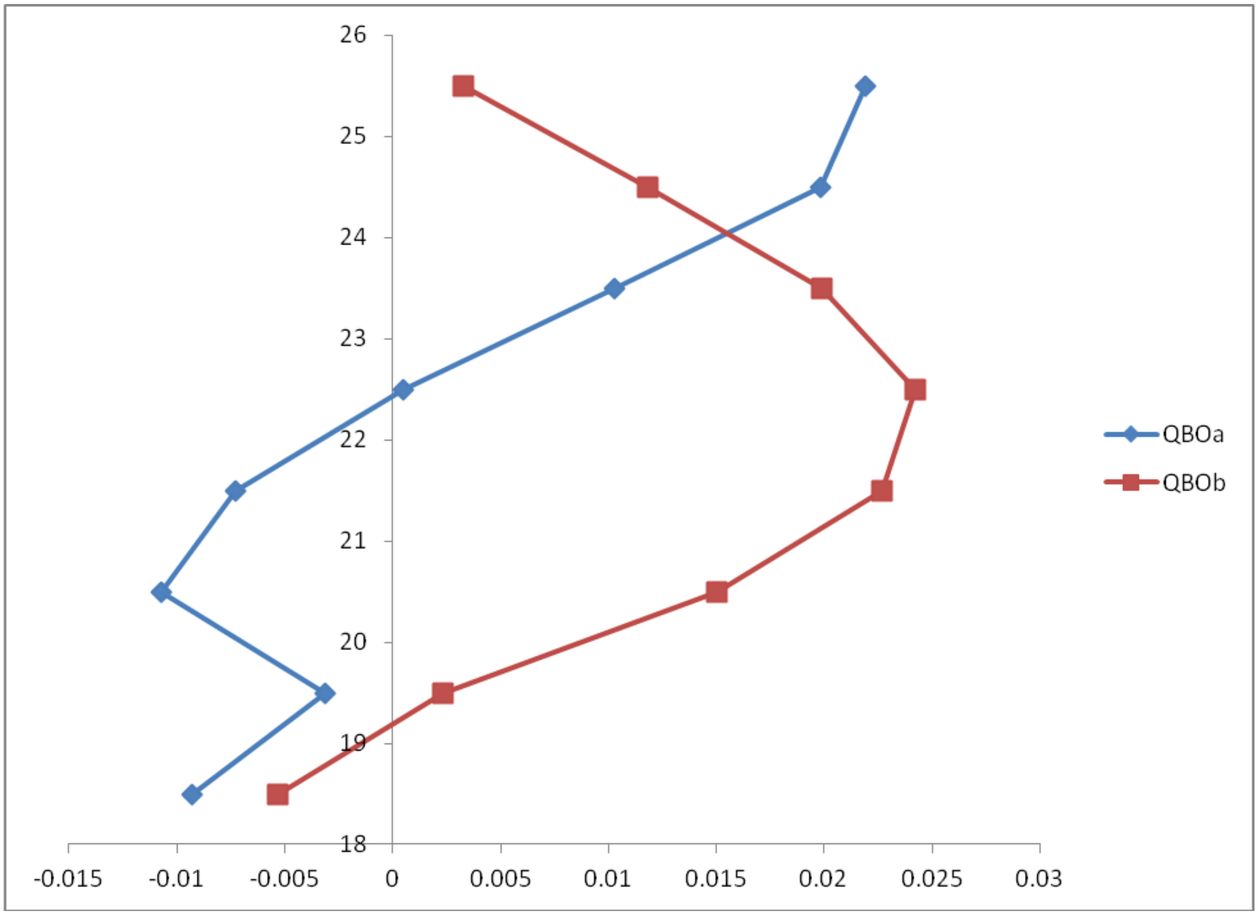
Comparing with the only other linear-trend study in the tropical lower stratosphere using SAGE II merged with more recent data, Randel and Thompson (2011) found statistically significant negative trends at 22.5 and 20.5 km (but not statistically significant at 24.5 km) in a 20°N-20°S band in the 1984 to 2009 period using SAGE II plus ozonesondes. Our results are quite similar to those of Randel and Thompson (2011) with a statistically significant negative trend (1984-2009) in the 17.5 to 24.5 km range but not statistically significant at 25.5 km (Fig. 4). Our merged trend (1984-2012) is only larger in a statistically significant way at 22.5-24.5 km than theirs, whereas our SAGE II trend (1984-2005) is not larger than their (merged) trend at any altitude.

The ozone response profile to ENSO is nearly identical in shape, sign, and magnitude to that shown in Randel and Thompson (2011). The ozone response profile to QBO is different than Randel and

Thompson (2011) because our approach is different. If we use their two orthogonal components of the QBO, we obtain the negative extremum at 19.5 km for one component and a maximum at 21.5 km in the other component. There are two similar plots below. The first is the ozone response to each QBO component (same quantity as Randel and Thompson [2011]) but for our latitude band (4.5°S to 4.5°N). The second figure shows the 20°S to 20°N latitude band, which Randel and Thompson (2011) used for SAGE II data, but strangely, for ozonesondes they used a different latitude band (10°S to 10°N). Our regression software assumes a constant latitude band width over the merged data record and we made no effort to change our software in this regard. The first figure (4.5°S to 4.5°N) shows that the maxima and minima occur at the same altitudes as Randel and Thompson (2011), but with larger amplitudes because of our narrower (equatorial) latitude band. This is expected since the QBO is stronger at the equator than at higher latitudes. The second figure (20°S to 20°N) shows that the extrema are shifted higher by 1 km, which is also expected from the general morphology of the QBO signal, but the amplitudes are slightly smaller given our wider band relative to the ozonesonde data. These figures were generated applying our best regression model at 25.5 km, plus a linear term.



4.5°N-4.5°S



20°N-20°S

In Sect. 3.2, we now write:

At 18.5 km, the ozone response to ENSO is (-5.8±4.2)%, a value that agrees very well with a previous estimate (Randel and Thompson, 2011).

and in Section 3.3, we now write:

The response to QBO is consistent with Randel and Thompson (2011) if one accounts for the different widths of our respective latitude bands.

Section 3 in general: I think it would be useful to have more structure in this section, e.g. by having sub-sections on linear trend, ENSO, QBO, tropopause pressure and solar cycle. The same might be true for section 2.2.

Taking the reviewer's excellent advice, we have added sub-sections on linear trend, ENSO, QBO, tropopause pressure and solar cycle as well as a discussion section.

Figure 2: The axis labels and numbers should be made larger. The lines should be thicker.

We increased the font size of the titles of the axes and made the time series line thicker.

Figure 3: The axis labels and numbers should be made larger. All lines should be made thicker. (In my printout they are hardly visible). The trend result from Randel and Thompson (2011) should be added to the plot.

Done.

Figure 4: I think it would be good to also plot a line for the derived linear trend - since the linear trend is one of the key topics of the paper. I also feel that it is important to point out the few high data points around 1990 and from 2001 to 2004, all near solar maxima, that are really key drivers for the long-term declining trend.

We have plotted the trend line at all three selected altitudes. As discussed above, the data points in 1990 are not high. The points in 2001 are also not high. The points in 2002 to 2004 are high and this is discussed in the text but one could equally point to the low values in 2010 due to the El Niño event.

Caption of Figure 4: El Chichon erupted in 1982/1983, that should be mentioned. The biggest eruption was Pinatubo in 1991, which is in the Figure, but is missing completely in the caption.

We now write:

…because of the residual aerosol layer from El Chichon (peak altitude of 19.5 km in 1985) which erupted in April 1982 and then Nevado del Ruiz which erupted on 13 November 1985 (Yue et al., 1991) and whose peak altitude was 22.5 km in early 1987. The eruption of Mt. Pinatubo, which occurred in June 1991, is marked in the upper panels.