**Referee #1**

Major issues

1) the first point raised by the referee in essence comes down to the question if the uncertainty in proxies as defined in our paper can be reduced, possibly by weighting the proxies in a weighted multivariate regression analysis.

First of all, this is a valid and highly relevant question. As pointed out in the paper, the topic of uncertainties in variables used in multivariate regression analyses of ozone has not drawn much attention. Our paper is thus a first attempt to try to quantify the impact of proxy uncertainties on the regression analysis. We are quite certain that there are possibilities to improve on our analysis, and we hope that our paper provides some incentive for others to have a closer look.

With regard to filtering the data ensemble, there are many textbooks providing methodologies how to optimally select the independent variables in your regression. There are rules of thumbs (make sure to have considerable fewer independent variables than degrees of freedom), pre-processing analyses (check the cross-correlation of the independent variables); check the probability distribution of the independent variable, as the presence of outliers (think about the year 2002, which we will address later) can have deteriorating effects on the regression), post-processing analyses (check to what extent residuals are randomly distributed; look at the p-values of the regression coefficients) and in-processing analyses (out-of-sample testing)

In addition, although we have chosen a probabilistic approach to estimate the impact of "errors in variables" on the regression, there are other methods to address this "errors in variables". However, these methods require well characterized variable errors. And, as shown in our paper, for most variables errors were unavailable or difficult to properly characterize.

Because of a lack of properly characterized variable errors, some methods for estimating the optimal set of independent variables also become "tricky". A frequently used approach is to select the set of variables that results in the best post-regression statistics (random residuals, smallest residuals, best correlation). However, such a selection method does not take variable errors into account (this becomes particularly problematic if errors are systematic and/or non-linear).

We thus have reservations to the post-processing selection of an optimal set of variables – even though we do apply that method in our paper as well, *i.e.* our results indicate that several variables do not contribute to the regression.

So, is a regression that explains a small fraction of the observed ozone variations not as "good" as a regression that explains most of the observed ozone variations? Well, not necessarily: if the variations in the independent variables are dominated by errors than much of the "goodness" of the regression may be artificial. The answer to this question still comes down to the question on how well one understands the variable errors.

Based on our analysis, one could argue that the understanding of variable errors is not very large: results can be all over the place depending the estimated variable errors. Worse yet, even the choice of optimal time period for determining ozone is not well motivated, adding an additional layer of uncertainty.

On the other hand, our analysis does provide some guidelines on what and what not to include. Do use the EP-flux, preferably including Aug+Sep. Include Sep+half Oct for ozone, but nut much more. Do not use the EESC. Do use the PWLT and the SAM. Exclude volcanic ash and the combinedQBO-solar flux index.

Furthermore, the length of the period over which the trends are calculated is also quite important.

With that knowledge now established with this paper, one can start thinking about how to improve the process, or alternatively, to start thinking about other methods than a multivariate regression on total ozone to determine stratospheric ozone recovery. However, we argue that both aspects beyond the scope of our current paper, which we view as a starting point for future research.

2) What about the solar-QBO index.

First of all, in section 2.3 we state that we use the 30 hPa winds - consistent with Kuttippurath et al. [2013] and Haigh and Roscoe [2006] and Roscoe and Haigh [2007]. Reason for only testing the combined solar-QBO index is that we want to remain consistent with Kuttippurath et al. [2013].

The referee has a number of questions about the use of the combined solar-QBO index. It is suggested that maybe both parameters should be included in the regression analysis separately, and provide uncertainty estimates of both (including uncertainties in the QBO phase).

However, as explained in Holton and Tan [1990], the QBO and solar effects cannot be considered separately. Whereas the solar influence modifies tropical stratospheric ozone and dynamics, the transport of the solar signal to higher/polar latitudes depends on the phase of the QBO. Or in other words: the QBO modulates the solar signal [Labitzke, 2004]. It is thus necessary to somehow combine both proxies when studying solar and QBO effects on stratospheric ozone, for which Haigh and Rosco [2006] and Roscoe and Haigh [2007] provide a methodology. In addition, Labitzke [2004], as well as Roscoe and Haigh [2007], show that there is little influence of the phase of the QBO and the solar cycle on Antarctic vortex dynamics during Antarctic spring (contrary to Antarctic summer and autumn as well as the Arctic regions, where there are clear QBO and solar effects).

In summary, it is well established that the QBO and solar effects on stratospheric ozone should not be separated, and there were already indications that their effect on Antarctic springtime ozone is small. Whether then to include them in the regression altogether is could be discussed, but since our analysis is based on a study that does include the combined solar-QBO index, we decided for the purpose of comparing our results to start with the same proxies as previous studies.

Nevertheless, our findings confirm the lack of QBO-solar signal in Antarctic springtime ozone, which we interpret as the Antarctic vortex being too strong for the fairly modest dynamical signals to penetrate.

*We included a summary of this discussion in the section on the combined solar-QBO index and in the conclusions.*

Note that a similar argument applies for stratospheric volcanic ash: literature exists suggesting that the influence of volcanic ash on Antarctic ozone is small or absent, also something opposite to stratospheric ozone elsewhere, and thus that there appears no need to include volcanic ash in studying Antarctic springtime ozone. Similarly, we interpret that as the influence of volcanic ash – either via dynamics/QBO or chemistry – is too small to affect the very strong Antarctic vortex and changes via chemistry are small compared to the very large effect of ozone depleting substances on Antarctic springtime ozone. Consistence of our findings on volcanic ash with other published results was already included in the paper.

3) Sensitivity of trend analysis on including 2002.

We performed a test of the ensemble without including the year 2002. Indeed, trends in ozone without including 2002 are larger, but not by much (mean trend difference +2.9 %; 2-sigma spread in trend differences ranging from -12.5 to 17.9 %).

In absolute numbers:

> without the year 2002, the 2-sigma spread in post-break trends is:
>
> 1.91 to 4.67 DU/year with a mean of 3.29 DU/year
>
> with the year 2002, the 2-sigma spread in post- break trends:
>
> 1.80 to 4.12 DU/year with a mean of 2.96 DU/year

Note that for trends calculated based on ozone itself – without "correcting" the ozone record based on the regression results - the inclusion of 2002 does matter, and trends are significantly different:

> with 2002 the post-break trend in ozone for all 8 ozone scenarios varies from 0.52 to 2.09 DU/year
>
> without 2002, the post-break trend in ozone for all 8 ozone scenarios varies between 2.35 and 2.96 DU/year.

Hence, the numbers above actually indicate that the regression is very effective in removing the 2002 anomaly.

For volcanic years (1983, 1984, 1992, 1993), this matters less. Even for the incorrected trends in ozone the differences are small.

> With all volcanic years, the pre-break ozone trend ranges from -4.72 to -6.82 DU/year

Without the volcanic years, the pre-break ozone trend ranges from -4.62 to -6.68 DU/year

In relative terms, the differences are less than 4% for pre-break ozone trends with and without inclusion of volcanic years.

*We added a remark about the sensitivity to 2002 in the discussion of the trends, as well as about the sensitivity to inclusion of "volcanic" years.*

**Minor comments**

Minor comments are addressed accordingly. Below only follow comments that require a more detailed response.

- Abstract: see general comments to the editor

- PWLT vs LINT. See discussion of our incorrect implementation of the PWLT in the regression, which is not trivial. After proper implementation the distinction between PWLT and LINT vanishes.

- Solar-QBO effects. See discussion above. We added a summary of the discussion above to the paper.

- What is the proper solar index? With the question about F10.7 vs. Lyman-α the referee confirms that it is very unclear what the proper solar proxy is. However, as we argue above, it is crucial to combine both QBO and solar index into one new index. Haigh and Rosco [2006] and Roscoe and Haigh [2007] provide a methodology for doing so, using particular QBO and solar proxies, so we used them here as well. As noted by other referees, the final goal of the paper is not to discuss the optimal set of proxies but trend uncertainty. Yet as this comment and several others show, it is not simple to discuss one without discussing the other.

- Figure 2 has been updated, now including a legend indicating the color corresponding to each QBO-solar scenario combination as described in the figure caption.

- Section 2.8, Vernier et al. [2011] reference added. Also added Solomon et al. [2011] rather than Trickl et al. [2013], as we argue Solomon et al. [2011] is more appropriate for discussing global changes in stratospheric aerosol.

- As explained above, EESC based trends we calculate from applying an Ordinary Linear Regression (OLR) to the EESC pre-break and post-break shape multiplied with the regression coefficient. In particular the pre-break EESC shape includes a levelling of the EESC near the break year (late 1990s). Hence, the linear fit error for the pre-break period will automatically be larger due to the non-linearity of the EESC shape. This turns out to be less of an issue for the post-break trend (whose trends is smaller and thus less affected by the levelling off than the pre-break trend.

- Table 1 trend uncertainties (now table 2): we calculated EESC-based trends by come and ORL to the pre-break and post-break EESC multiplied with the regression coefficient. Since we calculate EESC-based ozone trends using an OLR for both the pre-break and post-break period separately, there is no relation between pre-break and post-break EESC-based trend errors (see previous bullet). See also the previous bullet. We do see that the different EESC scenarios have different errors (2-sigma) for 5.5, 4 and 2.5 years Age of Air, respectively
    - Pre-break: -5.3 ± 0.1, -5.8 ± 0.4, -5.9 ± 0.7

- Post-break: 1.01 ± 0.12, 2.07 ± 0.03, 2.97 ± 0.16

The discrepancy between our EESC-based trend errors and those of Kuttippurath et al. [2013] suggest their error calculation differs from ours, but their paper does not discuss how their trend errors are calculated. Note that the regression based error in the EESC fit in our regression is much smaller than the error of the ORL fit to the pre-break and post-break EESC change.
Nevertheless, as we argue in our paper, a trend error calculation should include the residuals, *i.e.* the variations in ozone unexplained by the regression. The difference in PWLT trend errors – which does take the regression residuals into account - and the EESC trend errors therefore suggest that the EESC-based trend errors are overconfident.

- Figure 4: figure caption should indeed be “1979-BREAK” and “BREAK-2012” rather than “1979-1999” and “2000-2012”.

- A check was performed on the consistent use and description of BREAK periods in document.

- Auto-correlation: the main point of adding these references is that ozone time series generally are auto-correlated. If so, and in particular when calculating trends and trend uncertainties, one has to make sure that time series for which the trend is calculated does not show much – if any – auto-correlation. Fortunately this is the case, but otherwise the trend uncertainties should be corrected for auto-correlation (thereby increasing trend uncertainties, thus decreasing confidence in detection of recovery, which is thus relevant for this paper).

- Red/Blue outlines in figures 5 + 6. These lines indeed indicate the sum of all probability distributions of the scenarios. Reason to add them is that in the end it is these outlines on which the trend significance is based (even though we do make some refinements in the paper). Hence, we prefer to leave them in.

- Although we agree that aerosols have little effect, we feel that some discussion is needed as this finding is relatively new – there were a number of publications in the 1990s suggesting a significant influence of volcanic aerosol on Antarctic springtime ozone and only recently some new research (including this paper) combined with longer time series of ozone suggests no influence of volcanic aerosol on Antarctic ozone. Hence, we prefer to keep the figure as it presents a still rather new result that contrasts a number of older papers.

- Figure 6, lower panel: caption modified, it indeed shows the distribution of the regression value for the aerosol variable, including the distributions for the three different EESC scenarios. This was not properly explained.

- Section 3.6: optimal regression. As outlined in detailed above, we argue that even though this paper does not aim at providing the best set of regression parameters, choices in regression parameters do affect trend significance and the distribution of statistics from the regressions are related to the use of particular combinations of regressors. Some of the detailed questions and remarks by all referees also confirm a need of some understanding of the relevance of each variable in the regression. We nevertheless condensed section 3 in total, but a final word on what is and what is not an important variable in the regression and what appears to be better choices for time periods over which to average is really needed.

- Page 18516, lines 13-22, discussion of why not to use EESC. We agree, we removed most of the section but included the remark by the referee, which nicely summarizes the issue.

- Figure 9, table 4. Is now table 7, the table is modified to include percentage of significant changes also for all break years but ending either in 2010 or 2012, which are consistent with the red bars in figure 9. We also referred to table 6 in the caption of figure 9. The panels in figure 9 are swapped.

- With regard to tables 4 and 5 with significant trends for the ozone and EPflux scenarios: after proper implementation of the PWLT regression results are more consistent. For ozone including September and at least part of October results in the better trend significance (so the period should not be too short), while avoiding making the period too long. For the EPflux this means including September and August. This is explained in the text, including a warning that still a proper physical justification is lacking (for example, to include mainly full calendar months is quite arbitrary when you think of it).

- Figure 4: added the 2000-2012 trends and 2-sigma errors as presented in table 1 (now table 2). Added a brief discussion of them to section 3.3, nothing that maybe the uncertainties in the 1979-1999 ozone trends are larger than estimated from a single regression, but that 1979-1999 trends are nevertheless all statistically significant.

**References**

HOLTON, J., and H.T. Tan (1980), The influence of the equatorial Quasi-Biennial Oscillation on the global circulation at 50 mb*., J. Atmos. Sci. 37*, 2200–2208.