Referee #2

Major concerns

1) Try to better distinguish trend significance and explanatory power.

See also our answer to Referee #1 and point (3) below as well as our general comments to the editor. Although indeed we focus primarily on trend significance, the results of our study do have consequences for the optimal set of independent variables used for the regression. Furthermore, the two – significance and explanatory power – are not independent: the better the explanatory power of your independent parameters, the more likely it is your trends will become statistically significant as a better fit will remove some variations that otherwise would be considered noise in the trend calculation.

With the major revisions of the document we put less emphasis on the regressor selection and more emphasis of the trend significance. The main message should be clearer.

2) Distinguish between trend significance and trend consistency with expected recovery

It is currently expected that the Antarctic ozone hole will start to recover – or HAS to. When looking at trend estimates from various studies it is clear that stratospheric ozone over Antarctic has slowly started to increase. However, until now this increase in not considered statistically significant – a requirement for the recovery of ozone as defined by WMO – even though some studies do claim a statistically significant increase. However, as we show, that may be a too confident statement, as some errors in the methodology for determining recovery haven not been taken into account. The WMO ozone assessment report 2014 is very careful in its wording on this, noting exactly what is described above: Antarctic ozone appears to be increasing, but the increase is not statistically significant – nor does it already has to be based on modelling studies.

We realize that the distinction between an increase and a statistically significant increase may lead to some confusion, but we want to avoid that it is concluded that – because of lack of a statistically significant increase – the ozone hole is not recovering. In particular because of the fact that we do not yet necessarily expect to already see a statistically significant increase [Newman et al., 2007; WMO, 2010; Hassler et al., 2011], we wanted to make sure that people understand that despite the lack of statistically significant increase what we see is still consistent with expections.

We have added a brief discussion of the current insights vs what we expect, also in the light of the changed results after implementation of the proper PWLT regression, and we now suggest that based on our results it can be expected that more confidence in recovery based on multivariate regressions can be expected before 2020.

3) Contradictory statements with regard to the EESC: EESC being a better fit parameter vs avoid EESC in the fitting

This is related to the question of how to define what the best set of independent variables is, and how to determine independent variable errors.

There are various ways to determine the best independent variables. One which is often used is to look at the post-regression statistics and define the best fit model based on the best correlations and smallest residuals.

However, that may be misleading, as it does not consider the independent variable errors.

Alternatively, one can study the independent variable errors – as done in this paper – and conclude that there are structural uncertainties in variables (here: EESC).

That the EESC fit results in the best fit may be related to the fit models used: the EESC implicitly considers the possibility of saturation of ozone depletion around the EESC peak. A piece-wise linear fit does not, so the change before and after the break is rather abrupt, leading to potential fit issues around the maximum.

Obviously there would be other ways to for fitting linear fits (for example a 3-section linear fit rather than the 2-section fit applied here). This is part of 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.

We tried to clarify this in several parts of the revised paper.

Minor comments

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

- Abstract: shortened in line with the request by the referee.

- Added a table with the url's of the data sources

- Description of how trends are calculated from the EESC fit is added (see extended discussion in answer to referee #1)

- Solar flux, QBO and solar flux – QBO index. We moved sections 2.3 and 2.4, discussing the solar flux and QBO separately to the supplementary information. A sentences was added at the end of the section discussing the combined solar flux – QBO index noting that uncertainties in the individual solar flux and QBO index are considerably smaller (see SI) than the uncertainty of the combined solar flux – QBO index.

- Reference to scenarios "above". There is a little misunderstanding, as the scenarios referred here are the volcanic aerosol scenarios, not the ozone scenarios. Text was modified for clarificiation.

- Brief description of the MSR dataset added to the now section 2.7 (prev. 2.9).

- A table was added with online data sources (upgraded from the supplementary information)

- Trend errors in the Kuttippurath et al. [2013] study vs. our estimates. See discussion in response to referee #1. We have no proper explanation, other than Kuttippurath et al. [2013] apparently has a different method for determining trend errors. We simply took the EESC scenario multiplied with the regression value, and then applied an Ordinary Linear Regression to the pre-break and post-break periods separately. For the post-break period the trend errors are comparable, but for the pre-break period they are not. Kuttippurath et al. [2013] do not provide a description of how they derive at their trend errors.

- Figure 4: caption states 1979-1999 and 2000-2012. This is incorrect, also noted by referee #1. The figure was adjusted accordingly.

- Mean trends (EESC, PWLT) and 95% CI values are added to section 3.3

- Supplementary figures are not needed any more (no LINT trends after proper impelementation of PWLT in regression).

- Section with reference to Chehade et al. [2014] was deleted for various reasons.

- The discussion section was completely rewritten (for reasons outlined in the general remarks to the editor).

- See discussion on persistent features in ozone and EP flux scenarios. We hypothesize that most likely the 21-30 period is so short that variability in vortex dynamics start to play a role, as well as exposure to solar radiation. A bit like just using one day of each year to discuss long term trends in ozone. A somewhat different shape of the vortex or how long the parts of the vortex have been exposed to solar radiation may considerably affect then can result in considerable differences in vortex mean ozone. (not added to the revised document, as this is merely an untested hypothesis).

- Figure 9: grey bars were removed.