General comments

Incorrect implementation of PWLT fit

We discovered that we had incorrectly implemented the PWLT fit in the regression (implementation is not trivial). We apologize for this. Our misunderstanding is partly related to a lack of detailed description in most relevant papers on how to implement the PWLT in the regression. This has a number of consequences for our paper:

What has changed:

- For a few scenario combinations the post-break trends up to 2010 now become statistically significant, better in line with Kuttippurath et al. [2013].

- The number of ensemble-wide significant trends for the period up to 2010 now is of the order of 30-60% (was 0-20%). That means that still it still cannot be argued that the recovery of Antarctic ozone is statistically significant.

- A majority of post-break trends up to 2012 are statistically highly significant – consistent with, and evidence of, the notion that the longer the post-break period, the better the statistical significance.

- PWLT and EESC trend distributions are more similar, which is actually what one would expect and should have triggered some suspicion on our side.

- There is a bit more consistency on which ozone and EP-flux scenarios results in the best explanatory power (but still with caveats)

- Separate description of PWLT and LINT trends (thus with/without connecting the separate piece wise trends) is no longer included.

What has remained the same:

- The EESC regression results have not changed

- Our advice thus remains to avoid using the EESC as it is unclear what is its best description

- How to define the best ozone and EP flux time periods over which to average remains unclear

- It is still argued that this may lead to introduction of non-chemical variations in the ozone record

- We still find that a longer post-break period results in a larger number of statistically significant trends

- Similarly, a longer post-break period still does not necessarilly always result in improved statistics

- Volcanic ash, the QBO and the solar flux still do not improve the regressions

How this affects conclusions:

- results now better show that a multi-variate regression help in reducing deterministic non-chemistry variations in average ozone.

- there is general tendency for more significant trends with increasing time period. This is in a way similar to what we already concluded, but the number of significant trends to start with (from) is larger, suggesting we are closer to detection of recovery. This bodes well for the near future and expectations are that a high confidence of the recovery occurring may be reached before 2020.

- Nevertheless, one remains to be careful with the interpretation of regression results, as it remains unclear what the best ozone and regressor records are for this type of study (area and in particular the time period for which to average are ill defined).

- the abstract and conclusions were modified accordingly.

How to advance

We are open to considering the revised version of the paper as a resubmission (thus re-entering the discussion phase), but with the same referees.

Reason is that there have been major revisions and modification of the text, the tables and the figures, even though what is presented in the figures and tables has remained the same.

Similarly, the setup of the paper has also remained the same, but the discussion of the (optimal) regression model has been reduced in favor of more emphasis on the statistical analysis and trends.

However, we leave it up to the editor to decide on how to proceed exactly and are willing to accept other options (the incorrect implementation of the PWLT trend was our mistake).