Review of ACP(D) paper 10.5194/acp_2017-1157

“Symptoms of total recovery inside the Antarctic vortex during Austral spring”

by Pazmino et al. [2017]

This paper discusses recovery of springtime stratospheric ozone within the Antarctic stratospheric polar vortex based on total ozone column measurements. Or recovery of the so-called Ozone Hole.

The method applied is a multi-variate regression, used to estimate and separate the influence of various atmospheric processes that are known to affect Antarctic springtime stratospheric ozone depletion.

This is a well-known and widely used methodology in previous papers looking into this topic. This paper thus also builds on previous work and publications on the topic that have appeared in scientific literature especially in the last decade.

As outlined below, I think this paper is publishable but some important revisions are needed.

There is one recommendation [1] that the authors I hope are willing to consider but would be a considerable amount of work, and which I could live with if not part of the revision given that suggestions for revision [2] are incorporated.

Full review

Because of the legacy of the topic, it appears useful to consider what is new in this paper compared to what has already been published.

- Longer time period (1979 – 2015)
- A new proxy for the regression model
- A new/different method to estimate the vortex edge (needed for calculation of the annual average amount of springtime Antarctic ozone depletion)
- An piece-wise time trend based on a combination of a linear function and a polygon (second order; quadratic)
- Discussion of results for two periods (September average and 15 September – 15 October average). The latter is not a commonly used time period. The choices made in the paper will be discussed later on.
- Analysis of alternative Antarctic Ozone Hole metrics (area with total ozone columns < 150 DU and < 125 DU as compared to the standard 220 DU Ozone Hole area).

It is also useful to consider what is more or less new with regard to the findings of the paper

- Most proxies used in the regression do not reduce trend uncertainties. Piece-wise trends and heat fluxes alone (with or without the new GRADS proxy) explain more than 90% of the long term variability. Hence, based on this paper it could be argued that most proxies could be discarded, which is consistent with previous work.
- The longer time period considered leads to higher statistical significances of the post-peak trends in Antarctic springtime stratospheric ozone (from 2001 onwards; as expected based on previous papers).
- Higher statistically significant trends for the September period compared to 15 Sep – 15 Oct (consistent with previous findings)

Major comment.

This paper relies on a limited set of ozone records (Sep average, 15 Sep – 15 Oct average; area 220/150/125 DU), and a limited set of proxies used in the multi-variate regression. In two recent papers [de Laat et al., 2015, 2017; 2016JD025723], we explore the uncertainty ranges associated with the choices that can be made with regard to the time period over which the ozone parameter is calculated, and uncertainties associated with proxies as used in multi-variate regressions.

A paper like this, and also most previous papers on the subject, thereby only consider a few options in a much larger parameter space of options. This has the risk that it limits the view and interpretation (tunnel vision). The few time series that are looked at are then seen as the truth, every wiggle becomes meaningful, and too much attention is given to the formal statistical significances, whereas structural uncertainties are important as well.

For example, we have shown that rather arbitrary choices with regard to the proxies used in the regression have a strong impact on the formal statistical trend errors. We therefore argued that structural uncertainties are much larger than the formal statistical trend errors, which is important for confident statements about whether recovery has started or not. The same applies for the time period over which the ozone metric of choice is calculated. We see considerable differences in trends and trend uncertainties.

Furthermore, we also argued in de Laat et al. [2017] for the use of the ozone mass deficit rather than average ozone or area as the preferred metric to study long term changes in springtime Antarctic stratospheric ozone depletion. The motivation was that the OMD suffers less from what is discussed above (arbitrary choices) then average ozone and area.

This paper does not address these issues, nor are results put in the context of this work. The paper does show and confirm that most proxies in these multivariate regressions are not really useful. Confirms that statistical significances of post-peak trends become better because of a longer record (but which has to, given the mathematical nature of linear regressions). Confirms that there are differences in trends between September and 15 Sep – 15 Oct. And confirms that there are uncertainties associated with several parameters that need to be defined in advance (vortex position, vortex stability).

But there is no real discussion about why these are the appropriate choices. The GRADS proxy helps in improving the explained variability. But is that the justification? Smaller residuals? If so, I'm sure even better proxies can be constructed. Furthermore, the GRADS proxy is detrended. Why? If the GRADS proxy truly represents a physical process, why isn't GRADS allowed to also change on longer timescales (note that this is a point of contention in recent literature: is recovery fully attributable to ODSs or are there other long term changes in atmospheric dynamics that also play a role?). The same is true for use of the parabolic trend. It is not the standard approach in regression studies (all studies cited use PWLT), but the effect using two linear trends or one parabolic and one linear trends is not discussed (as far as I could see). It could also be argued based on figure 1 that none of the vortex edge definitions really captures only the vortex core. All still capture some high ozone columns around the vortex edge, which likely introduces variability in the ozone record not related to inner-vortex ozone depletion.

Consider that the standard 220 DU value used for OMD and even area fall well inside the 600 K vortex edge.

This is an exhaustive list of issues, which is exactly the point we want to make here: the issues raised in recent literature about arbitrariness of choices that are made, and the corresponding risk of tunnel vision.

Note that this is also why in de Laat et al. [2017] it is proposed to step away from the whole regression business.

This paper does show that ozone variability is mostly governed by depletion (ODSs) and heat fluxes or vortex (in)stability. How to properly account for the heat fluxes or vortex (in)stability is, however, not really clear, and this paper introduces yet another approach. In de Laat et al. [2017] it is instead proposed to simply remove the years that are characterized by a more unstable vortex from the record. Such years can be easily identified, but how they affect ozone depletion is much more complex, and appears to depend for example on the exact timing of vortex disturbances [de Laat and van Weele, 2011; 10.1038/srep00038]. This paper provides some additional ammunition for the proposal to step away from the regression methods.

The presence of this exhaustive list of issues and questions would be less of a problem if the paper introduced new concepts or new ideas, but the paper mostly builds on previous work and confirms what other papers have also concluded.

The new concepts and ideas that are introduced in the paper do not help in clarifying in what has recently emerged: the sometimes large structural uncertainties in this particular field of research, and arbitrariness with which analyses are performed. If anything, they only confirm the existence of large structural uncertainties and the arbitrariness.

So, how do I think this paper could be improved?

[1] One possibility would be to include additional analyses cover more of the parameter space. The paper already also uses average ozone and area, so a mass deficit could be included as well (see Fig. 5 of de Laat et al. [2017]).

The use of different area definitions based on different ozone thresholds could also be expanded – like looking at changes in the probability distributions of total ozone (a bit like Yang et al. [2008; 10.1029/2007JD009675], but much more extensive). However, that would require a considerably amount of additional work.

I could live without such an analysis if:

[2] regardless, results should be discussed within the context of recent publications and criticism of existing methods of Antarctic stratospheric ozone recovery detection. This is currently lacking, as also reflected in the conclusions section, which is more of a summary than a conclusion.

The challenge here will be to discuss it in such a way that that discussion does not undermine the findings of the paper.

So, what should be discussed are what I consider the most important findings of the paper:

- Most proxies in the MVR do not contribute much (if anything) to reduce trend uncertainties (small explanatory power)
- September yields a higher statistical trend significance than 15 Sep-15 Oct.
- Range of trend values and trend significance levels are indicative (or not) for structural uncertainties and systematic errors (this needs to be further supported)

In addition, I think the following should be included in a revision:

- report 2000-2010 and 2000-2012 trends & statistics for comparison with the 2000-2015 trends (and significances). This is helpful for comparison with results from older previous papers using MVR methods but somewhat different proxies.

- Use of “area” for 150 DU of 125 DU is an interesting more or less novel approach. Results show that such small TCOs did not occur until the late 1980sand early 1990s, indicative that these parameters are more sensitive for more severe ozone depletion. This also means that these parameters should return back to zero values earlier than the TOC columns return to 1980 levels. This method/analysis could be expanded more, by using the 150 or 125 DU also as vortex edge proxies (average ozone within area), and for Ozone Mass Deficit calculations (which traditionally is based on the 220 DU level but that is somewhat arbitrary). Possibly also report 175 and 200 DU results.

  (in all honesty, I think the analysis of long term changes in probability distributions could be a topic of a completely separate paper)

Minor comments

Page 1, line 25-26, and line 29 (and correspondingly tables 2 & 3), in particular the range of trend values that are reported.

How should this range be interpreted? Could this be considered representative of the structural uncertainty?

Page 2, line 14-15, the explanation of why October ozone behaves differently from September ozone.

October ozone is partly governed by different processes than September ozone. First of all, catalytic photochemical ozone destruction ceases in October. Rather, there is regeneration of ozone due to photolysis of O2 and oxidation of CH4 and carbon monoxide [Grooss et al., 2011; 10.5194/acp-11-12217-2011]. Furthermore, there is continued downward transport of ozone rich outer-vortex air into the vortex from the upper stratosphere down to the lower stratosphere [de Laat and van Weele, 2011; doi:10.1038/srep00038]. And there is vortex dynamics, as the authors correctly note. Together, these processes to a large extent determine October Antarctic inner vortex ozone.

Page 4, line 15. It is stated that a 5-day smoothing is applied to the EL of the maximum PV gradient. However, as far I know Nash et al. [1996] does not call for a 5-day smoothing. If that is right, then what is the justification of the 5-day smoothing?

Page 5, line 6-7. Correlations. Sometimes the paper uses R, sometimes R^2. Be consistent, preferably using R^2 and only refer to R if the correlation is negative (still providing R^2).
See also: page 5 - line 16, Page 10, line 17, and make sure to check throughout the paper.

Page 5, figure 4. The differences between SAT and MSR2 are fairly straight forward to explain. Up until 1993, both rely solely on TOMS. From 1993-1995, MSR2 relies on SBUV, and thanks to the data assimilation gaps are filled. From 1996 onwards, MSR-2 also uses GOME (1996 to 2005), SCIAMACHY (2002-2012), OMI (2004-), and GOME2 (2007-). Furthermore, MSR-2 uses ground-based total column data to account for inter-instrument differences. As a result, the estimated average MSR-2 total ozone column bias has been estimated at 1% [van der A et al., 2015; amt-8-3021-2015].

Add to line 21 the following “whole vortex. The data assimilation of MSR2 to some extent does fill gaps when ozone measurements are limited.”

Add after line 25. These differences are caused by MSR-2 starting to use multiple satellite total ozone column records after 1996, the procedures in MSR-2 to account for inter-instrument differences, and the data assimilation methodology that allows for filling gaps [van der A et al., 2015].

Page 6, line 25. It is stated that both PWLT and a combined parabolic trend – linear trend is generally used. The latter is not true, all papers cited only rely on a PWLT. The parabolic trend is a new concept introduced in this paper. As such, it should be explained later in the paper what the differences are that associated with both PWTs (the PWLT appear no to be used in the paper at all).

Grammar, typos.

Page 1, line 27. Replace “lower than” with “smaller than”
Page 2, line 4: change to “interannual variability of ozone as a function of the 11 year”
Page 2, line 8. I assume what is meant is “for the period over which the ozone record is calculated and for …”
Page 2, line 12. “ozone content is deepest”, I think what is meant here is “where ozone depletion is largest” or “where the ozone deficit is largest”.
Page 2, line 19. “update of the ozone”
Page 2, line 22. “full development of Polar ozone depletion”. I think what is meant here is “the period of fastest catalytic photochemical ozone destruction”
Page 3, line 35. Include reference to de Laat et al. [2017; 10.1002/2016JD025723] as a paper that also uses MSR2.
Page 4, line 15. Change to “This limit is subsequently smoothed temporally with”
Page 4, line 17. Start with “The Nash criterion”
Page 4, line 29. Change to “On this particular day, the region …”
Page 4, line 32. Change to “consist of”
Page 4, line 35. Change to “using the new classification.”
Page 4, line 36. Change to “The standard classification estimates a 40 DU and 20 DU larger ozone mean …”
Page 5, line 3. Change to “for the SAT data series … … based on the single …”
Page 5, line 4. Change to “Error bars represent the two sigma …”
Page 5, line 7. Change to “at the 2σ level”
Page 5, line 12. Change to “is preferred since it takes …”
Page 5, line 35. Change to “The ODS contribution to long-term trends in ozone is represented by piece-wise linear trend …”
Page 6, line 15. Start new paragraph after “period”
Page 6, line 21. Change to “with a p-value”
Page 7, line 14-17. Rephrase line “Despite … Weber et al. 2017)”. I assume you want to note that although September shows large variability in total ozone, it is still a commonly used month for recovery detection.
Page 7, line 18. Remove “are highlighted”, change “conclude that” to “identify ”
Page 7, line 18. Change “on October” to “for October”
Page 7, line 20-21. Delete “In our study … previous section.”
Page 7, line 25. Change to “the year 2000 was characterized by …”
Page 7, line 26. Change to “September, and yields a relatively high …”
Page 8, line 7. Add reference to Chipperfield et al. [2017; doi:10.1038/nature23681], who amongst others discuss the differences in pre-post peak ozone recovery rates.
Page 9, line 30. Change to “at 550K where the trend after …”
Page 9, line 37-38. Change to “Trends estimate for the second period show slightly”
Page 9, lines 40-41. Please rephrase, I don’t fully understand what is meant here.
Page 10, line 1. Change to “higher than 3, the threshold value …”
Page 10, line 21. Change to “ The ozone hole is also frequently defined as …”
Captions of figure 11 + 12: OMIT → OMI