Reply to A. T. J. de Laat review of manuscript acp-2017-1157

Symptoms of total ozone recovery inside the Antarctic vortex during Austral spring

Andrea Pazmino on behalf of all co-authors

We thank A. T. J. de Laat for the important and helpful review of our manuscript. Many interesting suggestions were incorporated to the new version of the manuscript. Please find our answers (in red) in three different sections: Comments to full review (1), Reply to major comments (2) and Reply to minor comments (3)

1. Comments to 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)

We appreciate your time and your general comment about our work, which allowed us to improve the paper. Since total ozone data is now available for MSR-2 until the end of October 2017, we decided to extend our study to the year 2017 using SAT and MSR-2 data. Due to this extension, all figures of the manuscript have been revised, except Figure 1 where white dot marks where added to highlight the region considered inside the vortex by the 400 K-600 K classification range. In addition we have noticed that the figure 12 of the original manuscript about the time shift of low values was not very clear. A new figure, Figure 13, has been produced in order to better illustrate the time delay in appearance of low total ozone values within the vortex. Similar conclusions as in the original version of the manuscript were provided. Furthermore the word “Multiple” was added to the title to highlight the fact that different signs of recovery were obtained in this work, e.g. (1) Significant positive trends of total ozone since 2001 in September and for the first time in the period of maximum ozone depletion (15Sept-15Oct) using MLR analysis on average ozone inside the vortex and Ozone Mass Deficit, (2) Decrease of occurrences of very low ozone values within the vortex and (3) increased delay of occurrence of low total ozone levels in the September 1st – October 15th period.

We generally agree with your appreciation of what is new and what is less new in our paper. Regarding the former, as you mention, one of the novelty of this work is to consider several different isentropic levels in the range 400 K – 600 K to make the classification based on the well-known Nash Criterion, in order to better constrain the ozone value inside the vortex. We think also that the addition of the GRAD proxy, based on physical considerations, provides a better agreement between observation and regressed values. The study of the very low ozone values within the vortex, based on different thresholds, provides also interesting indices towards ozone recovery.

We agree that it is also important to highlight the results confirming previous ones. Some of recent works using MLR have been already mentioned in the paper (Chipperfield et al., 2017; Weber et al, 2017) and also using other methods (Solomon et al., 2016).
It is true that most proxies in our MLR analysis do not significantly reduce trend uncertainties and piece-wise trends added to heat flux can explain more than 80% of vortex variance, but it is interesting to evaluate the contribution of proxies that are commonly used.
Longer time series generally results in higher statistical significance but due to higher ozone interannual variability in the last decade, each year can count in the trend analysis, considering the still relatively short ozone records since 2001.

2. Reply to Major comments

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.
Our work builds on previous studies and especially on recommendations made in de Laat et al. (2015) to optimize the multi-linear regressions. One of the purposes of this paper is to reproduce the variation of ozone inside the vortex during the last decade, especially from 2010 where increased variability is observed. This is how we came up with the GRAD proxy related to vortex stability during the studied month/period. This proxy is linked to the potentially mixing between inside and outside vortex regions during the period.
Further, in order to take into account the rounding off of the ozone loss due to saturation since the 1990s, which is especially visible by the end of September/beginning of October, we included a polynomial function to the linear functions used to evaluate long-term trends.

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.
We have considered different scenarios (2 ozone datasets, 3 different proxies' combinations, different criteria for vortex limit). Our MLR analysis could reproduce very well ozone in the last decade and we show that the GRAD proxy, based on physical explanation, improves the agreement between observation and regressed values. A robust estimation of structural uncertainties requests a "big-data" treatment as in de Laat et al. (2015). This was done already and is out of the scope of our study. However, a comparison of the maximum trend difference between the scenarios considered in the study and the retrieved trend uncertainties provides some evaluation of the structural uncertainty of our analysis.

The following paragraph was added in the conclusions of the marked-up version of the paper (page 14, line 26 to line 33):
"The structural uncertainties of the MLR analysis linked to the selection of proxies were not fully analysed in this work, as in De Laat et al. (2015). The main sensitivity tests concerned the baroclinicity of the vortex and the impact of its stability during the studied periods. Trend differences in the various scenarios analysed provide some quantification of related uncertainties and are lower than the statistical trend uncertainties. Further, the large determination coefficients obtained for both periods analysed give confidence in the retrieved trends. The Heat Flux proxy that provides the largest explanatory power in the various fits is a well-known driver of vortex temperature conditions that are the primary causes of polar ozone depletion in periods of high ODS levels. The influence of the GRAD proxy in recent years highlights the importance of the vortex stability for the containment of the ozone hole during the period of maximum depletion"

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.
As mentioned previously, the motivation of our study was to try and understand the causes of larger ozone variability in the last decade, especially in 2010 and 2012. This is why we chose to base our study on total ozone record. Regarding areas, the use of several thresholds allows us to follow the temporal evolution of areas with low ozone and find possible signs of recovery.

Further, we agree that OMD is a good metric to study the long-term changes. We have thus incorporated this metric using the 220DU threshold in our MLR analyses. Results related to OMD are included in the new Section 5.3 and in Section 6 where we explore the evolution of low ozone values.
The following figure (Figure 11) has been added to the revised version of the paper (Sect. 5.3)

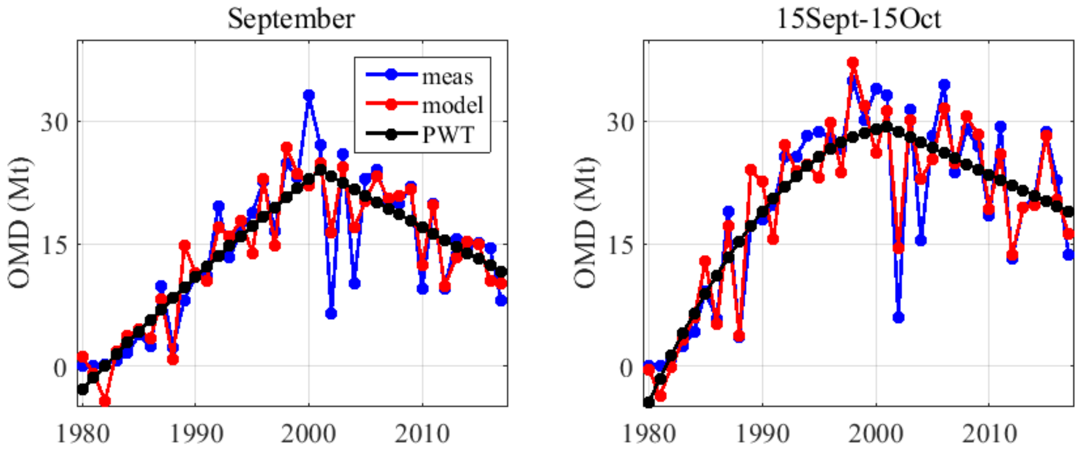


Figure 11: OMD (in Mt) computed from total columns of MSR-2 dataset lower than 220 DU and south of 60°S for September (left panel) and 15Sept-15Oct (right panel). Regressed values by MLR analysis using GRAD, HF and PWT are also shown as well as the fitted PWT proxy.

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.
We have taken into account the legacy of previous works to choose the classical proxies that were used to explain the ozone variability; particularly the work of big-data performed by de Laat et al. (2015). Besides, it is well known that the heat flux (HF) is an important proxy to explain ozone variability. It impacts the evolution of temperature inside the vortex and the build-up of Polar Stratospheric Clouds. But it does not provide an estimation of the vortex permeability and diffusion processes during the period for which the analysis is done. The choice of the GRAD and HF proxies is thus based on physical considerations and not on statistical ones. They allow us to better follow the evolution of the polar vortex on the 1980 – 2017 period. The article was modified in order to highlight those different points in Sect. 5.2.3 as shown in the tracked change version of the manuscript.

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?).
Both Heat Flux and GRAD proxies were detrended in order to avoid interference to the trend proxy that would be difficult to quantify. Such a treatment is commonly applied to proxy data in MLR analysis. Besides, as shown in Figure 6, not detrending the GRAD proxy would mainly influence the 1980 – 2000 period while the main emphasis of our study is on the recovery period from 2001.
In the paper it is mentioned that our estimation of trend is not necessary due only to ODS.

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).

We agree with you that a specific discussion was not sufficiently developed on the effect of using a parabolic function in addition to two linear functions instead of two linear functions only for the evaluation of long-term trends. The goal of this additional function is to explain the behaviour of ozone chemical destruction and the effect of saturation of the ozone loss. This was mentioned in the marked-up version of the manuscript on Page 7, line 26 to line 28:

"In this work our Modified PWLT model (PWT) uses an additional function in order to take into account the slower growth of ODS near the turnaround year and the ozone loss saturation effect within the Antarctic polar vortex in October (Yang et al., 2008)."

A new Sect. 5.2.4 (PWT vs PWLT) was added to better explain the differences between trends retrieved with our PWT model and the more classical PWLT method.
A figure was added in the supplementary material in order to show that the PWT provides a better representation of long-term ozone evolution within the vortex, especially for the 15Spet.-15Oct. period. Figures S1 and S3 display total ozone inside the vortex and OMD for September and 15Sept-15Oct using the MSR-2 data. The corresponding PWLT and modified PWT regressed values are also shown.

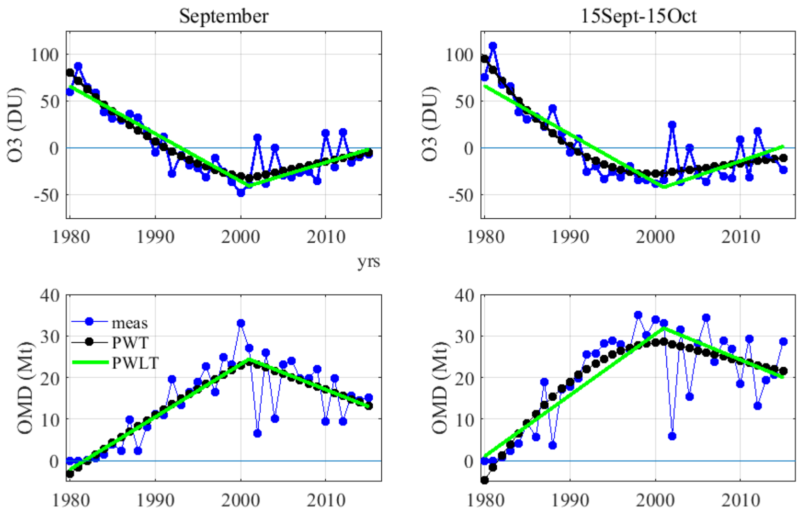


Figure S1 and S3. Top panels: average ozone inside the vortex using the 400 K-600 K range classification for September (left) and 15Sept-15Oct (right). Bottom panels: OMD for both periods. Fitted PWT (black line) and PWLT (green line) are represented in each panel.

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.
Our objective was not to capture with our classification the inner vortex only since ozone destruction can occur in the vortex edge in September. The idea was to better constrain the vortex without using any arbitrary ozone-based threshold. As seen in the Fig. 1 of the manuscript, the combination of the different iso-pv lines enables a better selection of the low ozone area. On the particular day shown in the figure, the area selected for the computation of the ozone average is limited by the 400 K vortex line near South America and by the 600 K line on the opposite side. White dot marks were added in the figure in order to better highlight the region selected by the 400 K-600 K range classification method.

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.
In our work, we used another approach and tried to reproduce the ozone variability for all years of the studied period. As mentioned previously the objective was to try and explain ozone variability in the last decade

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.
We hope that the many arguments we have developed previously help explain the contributions of our paper. Despite structural uncertainties in the MLR technique, it is widely used in ozone and climate studies and the level of agreement with observations obtained with our model gives us some confidence in our results. Further, your review allowed us to substantially improve the article and better explain our approach

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 OMD was included in the MLR analysis using for a threshold of 220 DU, south of 60°S. For simplicity, the OMD is computed for the periods of our study, e.g. September and 15Sept-15Oct. The OMD was also evaluated for different thresholds in order to compare with our evaluation of ozone hole areas with low ozone values.

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.
We have added two additional thresholds (175 DU and 200 DU) to refine the study. A choice of alternative thresholds as a function of probability distributions of total ozone could be done in a future 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.
This issue was revised in the paper. For example in the introduction section (page 2, line 18 to line 22)
"**The limitation of MLR analysis is that only formal statistical error of trend is estimated and structural uncertainties linked to the single and arbitrary combination of proxies is not taken into account. De Laat et al. (2017) inferred trend values from daily Ozone Mass Deficit (OMD) computed from a multi-sensor reanalysis dataset without using any model but filtering the anomalous years with low polar stratospheric cloud (PSC) volume. The authors found positive and highly significant trend of OMD since 2000.**"

The conclusions were also modified accordingly (page 14, line 26 to line 30).
"The structural uncertainties of the MLR analysis linked to the selection of proxies were not fully analysed in this work, as in De Laat et al. (2015). The main sensitivity tests concerned the baroclinicity of the vortex and the impact of its stability during the studied periods. Trend differences in the various scenarios analysed provide some quantification of related uncertainties and are lower than the statistical trend uncertainties. Further, the large determination coefficients obtained for both periods analysed give confidence in the retrieved trends."

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.

Those two points are more emphasized in the text and especially in the conclusions. For example in page 13, line 27 to line 37 about the different proxies:

"While the HF combined with GRAD proxies reproduce quite well the interannual variability of ozone, other proxies such as Aerosols, QBO, SF and AAO present smaller explanatory power and contribute less to reduce trend uncertainties."

- Range of trend values and trend significance levels are indicative (or not) for structural uncertainties and systematic errors (this needs to be further supported)

See our answer above

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.

A sensitivity test was made on the length of the datasets after 2001 varying the end year from 2010 to 2017 using PWT and PWLT proxies. The PWLT proxy has shown higher positive trends for the recovery period for both months compared to the trends based on PWT MLR analysis. PWLT presents significant trends but for PWT the significance of trends for the 15Sept-15Oct depends on the end year. As expected, error bars are smaller depending on the length of the dataset.

The 2001-2010 and 2001-2012 results were compared to previous works using PWLT proxies since PWT was never used before. A significant trend of 4.15 DU yr^{-1} is found in September, higher than 3.3 DU yr^{-1} reported in de Laat et al (2015) for Sept-Nov period and turnaround year in 2000. The R^2 coefficient is higher than 0.87 for all cases. Since trend proxies and ozone period are different, a comparison with previous results for the 2001-2010 and 2001-2012 was no introduced in the paper. A comparison with Solomon et al results for September was however included.

- 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.

Additional thresholds were considered as suggested. Our study is now based on 125, 150, 175, 200 and 220DU (Fig. 12). Such thresholds were also included in the analysis of OMD evolution.

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

It is a very interesting suggestion and we will consider it for a future work.

3. Reply to 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?
Since a small number of cases were considered in our study to evaluate structural uncertainties, this range can be interpreted in a qualitative way. A discussion in the Conclusion section about the comparison of extreme trend values found for the different cases treated in the paper at each period and the formal statistical error bars was performed. In September, the range of extreme cases is within the 2 sigma statistical error bars, while it is higher for the 15Sept-15Oct period.

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.
We are aware of the processes governing the ozone levels in October (see previous publications of our team Godin et al., J. Geophys. Res., 106(D1), 1311-1330, 2001 and Pazmiño et al, Atmos. Chem. Phys., 8, 5339–5352, doi:10.5194/acp-8-5339-2008, 2008 which analysed some of them). In our study, we focused on the baroclinicity of the vortex linked to vortex dynamics. We agree that it is not the only process affecting total ozone levels. The sentence was thus changed as follows (page 2, line 26 to line 28):
"The baroclinicity of the polar vortex in October and its displacement from the geographic pole can also contributes to the variability of the total ozone series averaged during the month of October."

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?
A justification of this smoothing has been added (page 4, line 34 to line 35):
"This limit is subsequently smoothed temporally with a moving average of 5 days to reduce the noise in the vortex edge data series."

Page 5, line 6-7. Correlations. Sometimes the paper uses R, sometimes R2. Be consistent, preferably using R2 and only refer to R if the correlation is negative (still providing R2).
See also: page 5 - line 16, Page 10, line 17, and make sure to check throughout the paper.
We decided not to follow your suggestion. We prefer to use a correlation coefficient R when comparing data records and a determination coefficient R^2 when comparing measurements with regression model retrievals.

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].
Please note that figure where changed since SBUV data was replaced by MSR-2 considering the spatial coverage of TOMS/OMI.

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].
This part was changed (page 6, line 1 to line 19)

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)

You are right, previous papers use only PWLT. A new section, the Section 5.2.4 (PWT vs PWLT) was added to the revised version of the paper.

Grammar, typos.

Page 1, line 27. Replace "lower than" with "smaller than" done

Page 2, line 4: change to "interannual variability of ozone as a function of the 11 year" done

Page 2, line 8. I assume what is meant is "for the period over which the ozone record is calculated and for …" Yes. Done

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". The sentence was changed accordingly

Page 2, line 19. "update of the ozone" done

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" The sentence was modified accordingly

Page 3, line 35. Include reference to de Laat et al. [2017; 10.1002/2016JD025723] as a paper that also uses MSR2. done

Page 4, line 15. Change to "This limit is subsequently smoothed temporally with" done

Page 4, line 17. Start with "The Nash criterion" done

Page 4, line 29. Change to "On this particular day, the region …" done

Page 4, line 32. Change to "consist of" done

Page 4, line 35. Change to "using the new classification." done

Page 4, line 36. Change to "The standard classification estimates a 40 DU and 20 DU larger ozone mean …" done

Page 5, line 3. Change to "for the SAT data series … … based on the single …" done

Page 5, line 4. Change to "Error bars represent the two sigma …" done

Page 5, line 7. Change to "at the 2σ level" done

Page 5, line 12. Change to "is preferred since it takes …" done

Page 5, line 35. Change to "The ODS contribution to long-term trends in ozone is represented by piece-wise linear trend …" done

Page 6, line 15. Start new paragraph after "period" done

Page 6, line 21. Change to "with a p-value" done

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.
Yes you are right. The sentence was removed and replaced in Page 8, line 22 to line 23 by
“Although pronounced decrease in total ozone is observed in September, recent works have used ozone records obtained during this month to detect the ozone recovery (Solomon et al., 2016; Chipperfield et al., 2017; Weber et al., 2017).”

Page 7, line 18. Remove “are highlighted”, change “conclude that” to “identify ” done

Page 7, line 18. Change “on October” to “for October” done

Page 7, line 20-21. Delete “In our study … previous section.” done

Page 7, line 25. Change to “the year 2000 was characterized by …” done

Page 7, line 26. Change to “September, and yields a relatively high …” done

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. done

Page 9, line 30. Change to “at 550K where the trend after …” this sentence was removed

Page 9, line 37-38. Change to “Trends estimate for the second period show slightly” This sentence was removed

Page 9, lines 40-41. Please rephrase, I don’t fully understand what is meant here. The sentence was modified (page 10, line 18 to line 23)
“*Despite the good agreement between regressed values and measurements especially for the period 15Sept-15Oct and for the range classification method (400 K-600 K), it is not possible to attribute ozone significant increase to ODS decrease. In addition, the ratio between trends before and after 2001 is larger than 3-which could be due to the effect of desaturation of the ozone loss.”*

Page 10, line 1. Change to “higher than 3, the threshold value …” sentence modified, see previous point (page 9 of original version)

Page 10, line 21. Change to “ The ozone hole is also frequently defined as …” We prefer to write “generally” instead of “also frequently”(page 9 of original version)

Captions of figure 11 + 12: OMIT ⇒ OMI
Previous Fig. 11 and 12 were modified. The Fig 11 is the new Fig. 12 and Fig 12 is the new Fig. 13 in the revised version and only MSR-2 data was used.

Reply to Anonymous Referee #2 review of manuscript acp-2017-1157

Symptoms of total ozone recovery inside the Antarctic vortex during Austral spring

Andrea Pazmino on behalf of all co-authors

We thank Anonymous Referee #2 for the time devoted to evaluate our work. Your valuable comments have helped us to improve our manuscript. Since MSR-2 total ozone data have become available until the end of October 2017, we decided to extend our study to the year 2017 using SAT and MSR-2 data. Due to this extension, all figures of the manuscript have been revised, except Figure 1 where white cross marks where added to highlight the region considered inside the vortex by the 400 K-600 K classification range. In addition we have noticed that the figure 12 of the original manuscript about the time shift of low values was not very clear. A new figure, Figure 13, has been produced in order to better illustrate the time shift in appearance of low total ozone values within the vortex. Similar conclusions as in the original version of the manuscript were provided. Furthermore the word "Multiple" was added to the title to highlights that different signs of recovery were obtained in this work, , e.g. (1) Significant positive trends of total ozone since 2001 in September and for the first time in the period of maximum ozone depletion (15Sept-15Oct) using MLR analysis on average ozone inside the vortex and Ozone Mass Deficit, (2) Decrease of occurrences of very low ozone values within the vortex and (3) increased delay of occurrence of low total ozone levels in the September 1st – October 15th period.

Please find our answers to your comments (in red):

I concur with much of what the other reviewer articulated, in particular these points from DeLaat's review:
1. "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."

2. "This paper does not address these issues, nor are results put in the context of this work."

3. "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."

I especially agree with DeLaat's concerns about the 'structural uncertainties' in this regression analysis, so please address all issues described in his review. In addition, there are other issues below related to ozone data sets that need to be addressed in a revised manuscript. If revisions are made that address both DeLaat's and my review, this paper could be published in ACP.

Please see our answer to de Laat's review to the different points specified above.

Specific topics of Concern

The composite satellite total ozone time series, referred to as SAT. The merging of satellite data sets into a single record is something to be done very carefully. Instrument measurements have bias and drift, and combining data sets in order to extract small trends (i.e., ozone recovery) requires a great deal of care and a good deal of knowledge about each instrument's characteristics and sampling pattern (i.e., coverage). I see no evidence here that any such considerations were used when combining the data sets. In fact in Figure 4, the difference between the assimilated ozone time series (MSR) and the SAT shows big jumps! There is a large trend from 1990-2005. Does this represent an unphysical trend (i.e., changes in the observing system) in the assimilation, or is this coming from how the individual data sets in the SAT were merged? Have you tried your trend analyses on the 5 merged ozone data sets referenced in Weber et al. [2017]? Without any discussion or justification of how the data sets were merged in this study, I don't see how the trend results presented here (and especially their uncertainties!) can be taken seriously.

Since MSR-2 data are based on the assimilated satellite ozone time series already corrected from offset, trends and variations of solar zenith angle and temperature in the stratosphere, we would like to consider in addition in our work a satellite datasets commonly used in ozone studies (including recovery) with similar kind of instrument (TOMS and OMI) and similar retrieval; without applying any correction. Since SBUV data are sparse, we have decided in the revised version to fill the 1993 – 1995 gap years with MSR-2 data, but taking into account the same spatial coverage as that of TOMS and OMI instruments (see new Fig. 4). Finally, due to the important differences observed in September, particularly the unexplained trend in the 1990-2005 period that you mention, we decided not to include SAT datasets for trend retrieval in that month. Discussions about this difference as well as the issue of spatial coverage of the SAT data is discussed in the Sect. 4 of the marked-up manuscript. The following paragraph was added in page 6, line 1 to line 19

"MSR-2 total ozone data series inside the vortex are compared to SAT series as shown in Fig. 4, which displays the relative difference between MSR-2 and SAT for the 400K-600K range classification. Differences of about ±0.5% are observed in the 1980s. Small differences are expected during this period since only TOMS data are used in both data sets until 1993. In the 1993-1995 period discrepancies between both curves are only due to the differences in the selection of MSR-2 data for the SAT record in order to have similar spatial coverage as the data from the other instruments incorporated in the SAT time series. These differences varying between -1 and 0.5 % represent an estimation of the impact of reduced spatial coverage in SAT dataset on the averaged total ozone level in September. The 15Sept-15Oct period presents negligible differences. The addition of GOME (1996-2005) in MSR-2 assimilation could explain the discrepancies with the SAT dataset that considers only TOMS-EP. From 2001, differences are larger and generally positive, reaching ~5% in September and ~3% in 15Sept-15Oct. period. These increased differences are especially visible during the period where data from instruments on board the ENVISAT platform (e.g. SCIAMACHY) are assimilated in the MSR-2 record. Overall, values in September present a mean bias of 1.3 % (dash blue line in Fig. 4), and in 15Sept-15Oct a smaller bias value of 0.5 % (dash red line in Fig. 4). Temporal evolution of the differences, e.g. negative trend in the 1980s and positive trend in the 2000s, can have an impact on the long-term ozone trends retrieved from both records. In general, differences between SAT and MSR-2 records are caused by MSR-2 starting to use multiple satellite total ozone columns 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)."

Regarding the 5 merged ozone data sets referenced in Weber et al. [2017], they correspond to zonal averages and cannot be used for total ozone classification as a function of equivalent latitude as it is done in our study.

The 'range method' is not clearly explained. I understand that you are using it to see the sensitivity of the calculated trends to the definition used for the area of depletion, and I get that you calculate different areas depending on which isentropic level is used, but exactly how are you deciding which levels to use? Are you averaging over all the 400-600K level results? Only some of them? Do you choose the same range for each year? The details of this methodology were not made clear. It's interesting that in the end you conclude that the 475K results are as good as the other definitions. Is this because this is an altitude where there is some of the most severe depletion? An explanation for this result should be offered.

The range method was better explained in Sect. 3.2 (Methodology for classification), in Page 5, line 4 to line 8.

"*The total ozone column may thus not represent the ozone behaviour inside the vortex. In order to consider possible vortex baroclinicity, another approach is used, where vortex classification at different isentropic levels is considered at the same time. For this second approach, the range of selected isentropic levels is chosen in the altitude region of maximum ozone depletion: from 400 K to 600 K with a step of 25 K. The same 9 isentropic levels considered for 400 K-600 K range classification are applied each year.*"

The range classification considers selected isentropic levels between 400 K and 600 K with a step of 25K. Then the same 9 isentropic levels are used each year for the classification. While this new classification provide a better constraint of low ozone values within the vortex, differences in trend results are not significant at 2 sigma levels. We suggest in the revised version that the reason could be the good correlation between the different data sets (R>0.98) using the different methods. Sentences in the Conclusions (page 13, line 15 to line 24) were changed as follow:

"For the classification of total ozone measurements inside the vortex, the classical Nash et al. (1996) method is used. In order to evaluate the impact of vortex baroclinicity on trend analysis, classifications using a single isentropic levels (475 K, 550 K) and a range of levels (400 K – 600 K) are tested. Systematic differences are found between the various total ozone time series. However the inter-annual variability is similar with correlation coefficients ranging from 0.98 to 0.99 in both studied periods. While larger trend values are generally found with the 475 K classification, the differences with trends related to the 400 K – 600 K range classification are not significant at 2σ level."

The satellite instruments used (all UV sensors) do not see to the south pole in early September. The analysis calculated results for the polar region for the entire month of September, but measurements cannot be made at the highest latitudes in early September. Thus the 'September average' will be more strongly weighted by lower latitudes and later September dates. Please describe how the satellites' sampling of the polar area varies over September and what this does to the 'September averaged' quantity. This may impact the meaning of the trend results as they will include more of the late September, higher dynamical variability measurements.

We agree with your arguments. We have excluded SAT datasets for September also for this reason. In addition we have included a sentence on UV sensors sampling in September where measurements are not available for regions poleward of 77°S in the beginning of September, 82°S mid-September and 89°S at the end of the month in the Sect. 2.1 of the marked-up manuscript (page 3, line 31 to line 33)

"*Since TOMS and OMI UV sensors do not receive enough UV light in early September, originating from regions not illuminated by the Sun (from 77°S to 82.5°S up to mid-September), these regions were not considered to compute the total ozone mean value in MSR-2 data.*"

The impact of spatial coverage differences between SAT and MSR-2 was discuseed in Sect. 4 (page 6, line 6 to line 10).

"*In the 1993-1995 period discrepancies between both curves are only due to the differences in the selection of MSR-2 data for the SAT record in order to have similar spatial coverage as the data from the other instruments incorporated in the SAT time series. These differences varying between -1 and 0.5 % represent an estimation of the impact of reduced spatial coverage in SAT dataset on the averaged total ozone level in September. The 15Sept-15Oct period presents negligible differences.*"

Multiple ~~S~~symptoms of total ozone recovery inside the Antarctic vortex during Austral spring

Andrea Pazmino[1], Sophie Godin-Beekmann[1], Alain Hauchecorne[1], Chantal Claud[2], Sergey Khaykin[1], Florence Goutail[1], Elian Wolfram[3], Jacobo Salvador[3,4,5], Eduardo Quel[3]

[1]LATMOS, UVSQ Univ. Paris Saclay, UPMC Univ. Paris 06, CNRS, Guyancourt, France
[2]LMD, CNRS, Ecole Polytechnique, Palaiseau, France
[3]CEILAP-UNIDEF (MINDEF-CONICET), UMI-IFAECI-CNRS-3351, Villa Martelli, Argentina
[4]Universidad Tecnológica Nacional, Facultad Regional Bs. As. (UTN-FRBA), Ciudad Autónoma de Bs. As., Argentina
[5]Universidad Nacional de la Patagonia Austral, Unidad Académica Río Gallegos (UNPA-UARG) and CIT-CONICET Río Gallegos, Argentina

Correspondence to: Andrea Pazmino (andrea.pazmino@latmos.ipsl.fr)

Abstract. The long-term evolution of total ozone column inside the Antarctic polar vortex is investigated over the 1980-~~2016~~ 2017 period. Trend analyses are performed using a multilinear regression (MLR) model based on various proxies for the evaluation of ozone interannual variability (heat flux, Quasi-Biennial Oscillation, solar flux, Antarctic Oscillation and aerosols). Annual total ozone column corresponding to the mean monthly values inside the vortex in September and during the period of maximum ozone depletion from September 15th to October 15th are used. Total ozone columns from ~~combined SBUV, TOMS and OMI satellite datasets and~~ the Multi-Sensor Reanalysis (MSR-2) dataset and from a combined record based on TOMS and OMI satellite datasets with gaps filled by MSR-2 (1993 – 1995) are considered in the study. Ozone trends are computed by a piecewise trend ~~model~~ proxy (PWT) that includes two linear functions before and after the turnaround year in 2001 and a parabolic function to account for the saturation of the polar ozone destruction. In order to evaluate average total ozone within the vortex, two classification methods are used, based on the potential vorticity gradient as a function of equivalent latitude. The first standard one considers this gradient at a single isentropic level (475K or 550K), while the second one uses a range of isentropic levels between 400K and 600K. The regression model includes a new proxy (GRAD) linked to the gradient of potential vorticity as a function of equivalent latitude and ~~that represents~~ representing the stability of the vortex during the studied month period. The determination coefficient (R^2) between observations and modeled values increases by ~0.05 when this proxy is included in the MLR model. ~~The higher~~ Highest R^2 (~~0.93~~0.92-0.95) and ~~the~~ minimum residuals are ~~observed~~ obtained for the second classification method for both datasets and months periods. Trends in September over the 2001 – 2017 period are statistically significant at 2 sigma level over ~~2001-2016 period~~ with values ranging between ~~1.85~~ 1.84±1.03 and ~~2.67~~ 2.83±1.48 DU yr^{-1} depending on the methods and ~~data sets~~ considered proxies. This result confirms the recent studies of Antarctic ozone healing during that month. Trends ~~after~~ from 2001 are 2 to 3 times ~~lower~~ smaller than before the turnaround year as expected from the response to the slowly ozone-depleting substances decrease in Polar regions.

For the first time, significant trends are found for the period of maximum ozone depletion. Estimated trends ~~in~~ for the 15Sept-15Oct period over 2001 – 2017 ~~are smaller than in September. They~~ vary from ~~1.15~~ 1.21±0.83 to ~~1.78~~ 1.96±0.99 DU yr^{-1} and are ~~hardly~~ significant at 2σ level.

MLR analysis is also applied to the Ozone Mass Deficit (OMD) metric for both periods, considering a threshold at 220 DU and total ozone columns south of 60°S. Significant trend values are observed for all cases and periods. A decrease of OMD of 0.86±0.36 Mt yr^{-1} and 0.65±0.33 Mt yr^{-1} since 2001 are observed in September and 15Sept-15Oct, respectively.

Ozone recovery is also confirmed by a steady decrease of the relative area of total ozone values lower than ~~150~~ 175 DU within the vortex in the 15Sept-15Oct period since 2010 and a delay in the occurrence of ozone levels below 125 DU since 2005. ~~Comparison of the evolution of the ozone hole area in September and October shows a decrease in September, confirming the later formation of the ozone hole during that month.~~

1 Introduction

The evolution of total ozone content (TOC) in Antarctica during Austral spring is deeply linked to the important stratospheric ozone decline that was highlighted for the first time by Chubachi et al., 1984 and Farman et al., 1985. Nowadays the photochemical and microphysical processes leading to the massive and seasonal destruction of ozone in Polar Regions are well understood. The latest Ozone Assessment Reports (WMO, 2007, 2011, 2014) have confirmed the stabilization of ozone loss in Antarctica since 2000. The challenge now is to assess the impact of the observed reduction in the concentration of ozone depleting substances (evaluated in the polar regions to ~10 % in 2013 from the peak values in 2000, WMO, 2014) on the amplitude of the ozone destruction every year. During the last decade, several studies have been carried out to quantify a possible increase in total ozone column in the Antarctic polar vortex in spring directly linked to this decrease in the polar stratosphere. Most analyses use multi-parameter linear regression (MLR) models with different proxies to represent the interannual variability of ozone as a function of the 11 year solar cycle, the quasi-biennial oscillation (QBO), volcanic aerosols or eddy heat flux (Salby et al., 2012; Kuttipurath et al., 2013; De Laat et al., 2015). These studies generally show a significant increase of TOC since 2000 for September-November average period but they differ on the proxies used for the quantification of ozone inter-annual variability. De Laat et al. (2015) used a “big data” ensemble approach to calculate trends. Several scenarios were considered ~~for the period of ozone data set~~ for the period over which the ozone record is calculated and for the different proxy records. They found that the significance of trends could vary from negligible to 100 % significant at 2 sigma levels depending on the scenario considered. They have also determined the optimal proxy records and ozone record scenarios to obtain the best regression. The limitation of MLR analysis is that only formal statistical error of trend is estimated and structural uncertainties linked to the single and arbitrary combination of proxies is not taken into account. De Laat et al. (2017) inferred trend values from daily Ozone Mass Deficit (OMD) computed from a multi-sensor reanalysis dataset without using any model but filtering the anomalous years with low polar stratospheric cloud (PSC) volume. The authors found positive and highly significant trend of OMD since 2000.

Solomon et al. (2016) evaluated ozone trend using a Specified Dynamics version of Whole Atmosphere Community Climate Model (SD-WACCM). The authors have shown a significant healing in September but not in October where ozone depletion is largest ~~content is deepest~~ during the first two weeks. They also explain the difficulty of estimating trend in October by the large variability of ozone linked to temperature variations and transport. The baroclinicity of the polar vortex in October and its displacement from ~~in relation to~~ the geographic pole can ~~explain~~ also contributes to the variability of the total ozone series averaged during the month of October.

The direct link observed between ~~the~~ positive trends of total ozone within the polar vortex and the reduction of ozone-depleting substances (ODS) remains thus an open question, given the natural variability of the Antarctic vortex and the possible contribution of greenhouse gases (GHGs) to the trends (Chipperfield et al., 2017).

The purpose of the present paper is to provide an update of the ozone evolution inside the Antarctic vortex during the last decades taking into account the vortex baroclinicity. The main aim is to determine the different contributions to ozone inter-annual variability and to estimate the post 2001 total ozone trend and related significance for different periods: September, which corresponds to ~~the full development of Polar ozone depletion~~ the period of fastest development of catalytic photochemical ozone destruction and mid-September to mid-October when the maximum ozone loss is reached.

This ~~study~~ paper is organized as follows. Ozone datasets from satellites and multi-sensor reanalysis are presented in Sect. 2 and the description of the method used for total ozone column classification ~~classifications~~ inside the vortex in Sect. 3. ~~Influence~~ The influence of vortex baroclinicity on total ozone column inside the vortex is assessed in Sect. 4 by using a new classification method ~~and standard one at~~ compared to standard ones based on a single isentropic level. Ozone trends before and after ~~a~~ the turnaround year calculated using multi-regression model for September and mid-September to mid-October are presented and discussed in Sect. 5. Results on trends using OMD records as a metric are also presented. The temporal evolution of the ~~number~~ amount of very low total ozone values inside the vortex is evaluated in ~~Section~~ Sect.6. Conclusions are finally presented in Sect. 7.

2 Total ozone column data series

Total ozone global fields from satellite observations (TOMS~~,~~ and OMI ~~and SBUV~~) and multi-sensor reanalysis (MSR) are used in this study to cross-check trend estimation before and after a turnaround year over the 1980-~~2016~~ 2017 period.

2.1 Space-borne observations

Total ozone columns data series of NASA's Total Ozone Mapping Spectrometer (TOMS) instrument onboard Nimbus-7 (N7) and Earth Probe (EP) between 1980 and 2004 are used. The instrument is a single monochromator that was designed for near-nadir measurements of the total ozone column (e.g. McPeters et al., 1998). TOMS measures the backscattering of solar radiation by the Earth's atmosphere in six 1 nm-bands of ultraviolet wavelength between 306 nm to 380 nm, more or less absorbed by ozone. Total ozone column is inferred from the ratio of two wavelengths, 317.5 nm strongly absorbed by ozone and 331.2 nm weakly absorbed (Bhartia and Wellemeyer, 2002). Level 3 gridded TV8 data of 1.0° (lat) x 1.25° (lon) of total ozone columns of TOMS were used in this work and are available from the Goddard Earth Sciences Distributed Information and Services Center (GES DISC) in simple ASCII format in the NASA anonymous ftp site (ftp://toms.gsfc.nasa.gov/pub/*satellite*/data/ozone/)

Ozone total column observations of Ozone Monitoring Instrument (OMI) onboard Aura satellite are also used to continue TOMS measurements from 2005 to 2017 ~~2016~~. The OMI instrument is a nadir viewing hyperspectral imaging in a push-broom mode. OMI measures the solar backscatter radiation in the complete spectrum of the ultraviolet/visible wavelength range (270 nm - 500 nm) with 0.5 nm spectral resolution (Levelt et al., 2006). Total ozone column used in this work was retrieved using TV8 algorithm, hereafter referred to as OMIT in order to maintain continuity with TOMS data record (McPeters et al., 2008). Level 3 daily gridded data of OMIT with better spatial resolution (1.0° x 1.0°) than TOMS is used. Data are also available on NASA anonymous ftp site.

The total ozone column data series was combined by using specific satellite data over the following periods: TOMS-N7 (1980-1992), TOMS-EP (1996-2004) and ~~OMIT~~ OMI (2005-~~2016~~ 2017). Note that data of 1993-1995 are sparse or missing for the September-October period. In order to complete the data series, total ozone columns of multi-sensor reanalysis 2 (see Sect. 2.2) ~~observations of Solar Backscatter Ultraviolet Radiometer (SBUV) onboard NOAA-9 (SBUV-N9)~~ were used for those years. ~~SBUV is a nadir viewing double grating monochromator instrument that measures backscattering of solar radiation by the Earth's atmosphere at 12 wavelengths from 250 nm to 340 nm (Frederick et al., 1986). Daily files of individual measurements (level 2) of version 8.6 (V8.6) are used in this work and are available on NASA anonymous ftp site (ftp://toms.gsfc.nasa.gov/pub/sbuv/NOAA09/). An overview of the V8.6 SBUV data record is presented in McPeters et al., 2013.~~ Since TOMS and OMI UV sensors do not receive enough UV light in early September, originating from regions not illuminated by the Sun (from 77°S to 82.5°S up to mid-September), these regions were not considered to compute the total ozone mean value in MSR-2 data.

TOMS/~~OMIT~~OMI and MSR-2 ~~SBUV~~ data series have previously been used in different scientific studies of ozone recovery in the Southern polar region (Salby et al., 2012; Kuttipurath et al. 2013; Solomon et al., 2016~~, Weber et al., 2017~~). ~~The bias induced by using SBUV data on 1993-1995 is discussed in Sect. 4.~~ Hereafter the 1980-~~2016~~ 2017 composite satellite total ozone series will be called SAT.

2.2 Multi-Sensor reanalysis

Ozone Multi-Sensor Reanalysis version 2 (MSR-2) provides global assimilated ozone fields for the period 1980-~~2015~~ 2017 based on 14 satellite data sets (van der A et al., 2015). ~~Ozone data of 2016 were not yet available at the time of writing this paper and are therefore not included in the study.~~ The 14 polar orbiting satellites measuring in the near-ultraviolet Huggins band were corrected to construct a merged satellite data series that are assimilated within the chemistry-transport assimilation model TM3-DAM to obtain ~~MRS-2~~ MSR-2 data (see van der A et al., 2010 for a detailed description and van der A et al., 2015 for last improvements of the assimilation model). Corrections of offset, trends and variations of solar zenith angle and temperature in the stratosphere were computed in satellite data sets by comparisons with individual ground-based Dobson and Brewer measurements from World Ozone and Ultraviolet Data Center (WOUDC). Those corrections are specified in van der A et al. (2015), table 2.

Daily gridded forecast ozone data of MSR-2 at 12 UTC and spatial resolution of 0.5° x 0.5° were used in this work and they are available from the Tropospheric Emission Monitoring Internet Service (TEMIS) of KNMI/ESA (http://www.temis.nl/protocols/o3field/data/msr2/).

Different studies on trends in the South Hemisphere have used ~~MRS-2~~ MSR-2 data (Kuttipurath et al., 2013, de Laat et al., 2015, 2017). Hereafter the 1980-~~2015~~ 2017 ozone series will be called MSR-2.

3 Data classification within the vortex

In order to consider total ozone columns only within the polar vortex, the data classification is performed by evaluating the vortex's position at different isentropic levels from May 1st to December 31, each year. Two classification methods are then applied in order to evaluate the impact of baroclinicity of the vortex on the averaged total ozone columns in both studied depletion periods. The first one is based on a single isentropic level, while the second one considers a range of isentropic levels.

3.1 Vortex position

For each day of the studied periods, the vortex position is determined by using a 2-D quasi-conservative coordinate system (equivalent latitude/potential temperature) described by McIntyre and Palmer (1984) where the pole in equivalent latitude (EL) corresponds to the position of maximum potential vorticity (PV). This conservative system is computed from PV field simulated by the Modélisation Isentrope du transport Mésoéchelle de l'Ozone Stratosphérique par Advection (MIMOSA) PV advection model (Hauchecorne et al., 2002). The model was forced by ERA-Interim (Dee et al., 2011) meteorological data (2.5°x 2.5°) of European Centre for Medium-Range Weather Forecasts (ECMWF). Daily advected PV fields (1°×1°) on the 30°S–90°S latitude band at 12 UTC are used to calculate EL on the isentropic level range between 400K and 600K with a step of 25K.

Following Nash et al. (1996), PV is evaluated as a function of EL and three particular regions are identified: inside the vortex, characterized by high PV values, at the vortex edge, corresponding to high PV gradients and outside the vortex (or surf zone) with small PV values. The limit of the vortex corresponds to the EL of maximum PV gradient, weighted by the wind module. ~~Then this limit is~~ This limit is subsequently smoothed temporally with a moving average of 5 days to reduce the noise in the vortex edge data series.

3.2 Methodology for classification

The Nash criterion was already used in several studies to distinguish measurements (ozone profiles and total columns) inside and outside the vortex in the Southern Hemisphere (Godin et al., 2001; Bodeker et al., 2002; Pazmino et al., 2005, 2008; Kuttipurath et al., 2013, 2015). In the case of total columns, measurements were considered inside the vortex when their corresponding EL was larger than the EL of the vortex limit at a specific isentropic level (e.g. 550 K, Bodeker et al., 2002; Pazmino et al., 2005). However this “standard” method does not take into account the baroclinicity of the vortex. It can result in the classification of total ozone columns inside the vortex while partial columns below or above the selected isentropic level are outside the vortex. The total ozone column may thus not represent the ozone behaviour inside the vortex. In order to consider possible vortex baroclinicity, another approach is used, where vortex classification at different isentropic levels is considered at the same time. For this second approach, the range of selected isentropic levels is chosen in the altitude region of maximum ozone depletion: from 400 K to 600 K with a step of 25 K. The same 9 isentropic levels considered for 400 K-600 K range classification are applied each year.

In order to illustrate the impact of vortex baroclinicity on the classification of total ozone column inside the vortex, Fig. 1 shows MSR-2 total ozone fields on October 7, 2012, with the vortex position computed at different isentropic levels superimposed. The vortex position curves are represented by black to light grey colours. On this particular day, the region classified inside the vortex ~~considering~~ in the 400 K-600 K range is limited by the vortex position at 400 K (black line) towards Antarctic West coast and Palmer Peninsula and ~~by~~ at 600 K (light grey line) towards the Antarctic East coast. The white dot marks in the Fig. 1 show the limit of the region considered in this new classification In the case of standard classification using a single level at 475 K or 550 K, the region estimated as inside the vortex ~~consists in~~ consists of an area with total ozone columns higher than 400 DU. These areas are not considered in the classification using several isentropic levels between 400 K and 600 K isentropic levels. Regions where total ozone columns are lower than 220 DU are taken into account by the ~~different~~ classifications at all the isentropic levels. A daily mean total ozone column of 213.4 DU was computed inside the vortex using this new classification method. ~~Standard classification estimate~~ The standard classification estimates a 40 DU and 20 DU ~~higher~~ larger ozone ~~mean~~ average values at 475 K and 550 K respectively on that day.

4 Vortex baroclinicity

Both methods of classification described in the previous section were applied to satellite composite total ozone data series SAT and ~~MRS-2~~ MSR-2 reanalysis at each grid point. For each year, daily mean total ozone amount inside the vortex was averaged over two periods: the whole month of September, and the period of maximum ozone depletion between September 15th and October 15th (15Sept-15Oct). Figure 2 shows the evolution of total ozone average inside the vortex for the 15Sept-15Oct period between 1980 and ~~2016~~ 2017 for the ~~SAT~~ MSR-2 data series computed with the standard classification method based on the single isentropic level (475 K and 550 K) and with the second method using the 400 K – 600 K range of isentropic levels. Error bars represent the two sigma standard error (2σ). Similar interannual total ozone variability is observed for the time series obtained by the different methods. The correlation coefficients between the range method and the standard one at 475 K and 550 K are ~~0.976~~ 0.98 and ~~0.995~~ 0.99 respectively. Despite these good correlations, the data series are significantly different at the 2σ level. Higher ozone values are found with the standard method, especially for the 475 K level, ~~with~~ which shows a mean difference with the TOC ozone time series based on the range method of ~15 % over the whole analysis period. Three years stand out in the comparison: 1995, 1999 and 2011, during which the inside vortex region was systematically larger at 475 K compared to higher isentropic levels during the period. Similar results are observed for September (not shown). In this work, the second method is ~~preferably used~~ preferred since it takes into account the ozone loss at different isentropic levels, which strongly impacts the total column.

~~SAT~~ MSR-2 total ozone time series obtained in September and 15Sep-15Oct with the range classification method are displayed in Figure 3. September presents ~~~7~~ 8% higher ozone mean values than the 15Sept-15Oct period. Similar interannual variability is observed between the two periods as shown by the correlation coefficient of ~~0.97~~ 0.98. The last ~~three~~ four years present very similar ozone values around ~~200~~ 205 DU in September while in 15Sept-15Oct period they show ~~shows~~ larger variability.

MSR-2 total ozone data series inside the vortex ~~were~~ are compared to SAT series~~. Figure 4 illustrates~~ as shown in Fig. 4, which displays the relative difference between MSR-2 and SAT ~~considering~~ for the 400 K-600 K range classification. Differences of about ±0.5 % are observed in the 1980s. Small differences are expected during this period since only TOMS data are used in both data sets until 1993. ~~During the first decade of 1990s, a large variability fluctuating from -3% to 5% is found linked to SBUV averaged total ozone column. In fact, SBUV level 2 measurements do not allow a homogeneous sampling of the whole vortex as MSR-2.~~ In the 1993-1995 period discrepancies between both curves are only due to the differences in the selection of MSR-2 data for the SAT record in order to have similar spatial coverage as the data from the other instruments incorporated in the SAT time series. These differences varying between -1 and 0.5 % represent an estimation of the impact of reduced spatial coverage in SAT dataset on the averaged total ozone level in September. The 15Sept-15Oct period presents negligible differences. The addition of GOME (1996-2005) in MSR-2 assimilation could explain the discrepancies with the SAT dataset that considers only TOMS-EP. From 2001, differences are larger and generally positive, reaching ~5% in September and ~3% in 15Sept-15Oct. period. These increased differences are especially visible during the period where data from instruments on board the ENVISAT platform (e.g. SCIAMACHY) are assimilated in the MSR-2 record. Overall, values in September present a mean bias of ~~1.4~~ 1.3 % (dash blue line in Fig. 4), and in 15Sept-15Oct a smaller bias value of ~~0.6~~ 0.5 % (dash red line in Fig. 4). Temporal evolution of the differences, e.g. negative trend in the 1980s and positive trend in the 2000s, can have an impact on the long-term ozone trends retrieved from both records. In general, differences between SAT and MSR-2 records are caused by MSR-2 starting to use multiple satellite total ozone columns 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).

Despite the differences between SAT and MSR-2 datasets, a purpose of this work was to analyse in the same way satellite data such as those included in the SAT record without any correction or adjustment and the MSR-2 record, which accounts for inter-instrument differences using ground-based total column data. Due to the larger differences observed between both data sets in September especially in the 1995 – 2010 period, which may have an impact on trend analysis, it was decided to retrieve trends from the SAT dataset in the 15Sept-15Oct only.

In the next section, ozone data series ~~using~~ based on the different classification methods are used to evaluate the impact of vortex baroclinicity on ozone trends inside the vortex for both studied periods.

5 Trend analysis

5.1 Method

In order to evaluate ozone recovery in Antarctica, estimation of trends before and after 2001 were calculated using a multi-regression model (Nair et al., 2013) updated from the AMOUNTS (Adaptative Model for Unambiguous Trend Survey) model (Hauchecorne et al., 1991; Kerzenmacher et al., 2006). Different common explanatory variables such as eddy heat flux (HF), solar flux (SF), Quasi-Biennial Oscillation (QBO), Aerosols (Aer), Antarctic Oscillation (AAO) are used to explain total ozone variability over the 1980-~~2016~~ 2017 period. These proxies were widely applied in different trend studies (e.g. de Laat et al., 2015 and references herein). The ODSs contribution ~~on~~ to long-term trend ~~of~~ in ozone ~~can be stated~~ is represented by piece wise ~~linear~~ trend functions (PW~~L~~T). ~~Total~~ The total ozone variability (Y) can be expressed following Eq. (1):

$$Y(t) = \mathrm{K} + C_{HF}\mathrm{HF(t)} + C_{SF}\mathrm{SF(t)} + C_{QBO30}\mathrm{QBO30(t)} + C_{QBO10}\mathrm{QBO10(t)} + C_{Aer}\mathrm{Aer(t)} + C_{AAO}\mathrm{AAO(t)} + C_{GRAD}\mathrm{GRAD(t)} + \mathrm{PWT(t)} + \in \mathrm{(t)}, \quad (1)$$

where t is the time in year from 1980 to 2016 2017, K is a constant, C_{proxy} are the regression coefficients of the respective proxies mentioned above and ∈ (t) is the total ozone residuals. Table 1 shows the respective information on for each proxy: source, specific characteristics and time window where proxy values are averaged to represent the respective year value. QBO effect on ozone variability is estimated using two proxies at 30hPa (QBO30) and 10hPa (QBO10), which are out of phase by $\sim\frac{\pi}{2}$ (Steinbrecht et al., 2003). The HF proxy corresponds to the average over the August-September period of the 45-day mean Heat Flux in the 45°S-75°S latitude range at 70hPa from MERRA-2 analyses. The time window of August-September is selected for computing the mean HF, following de Laat et al. (2015) recommendation to obtain the best regression results. For the Aer term, a merged proxy of monthly aerosol optical depth (AOD) is computed from updated Sato et al., (1993) dataset for the 1980-1990 period and from four satellite data series (SAGE II, OSIRIS, CALIOP and OMPS) for the 1991-2016 2017 period. AOD datasets are averaged over the 40°S-65°S zonal region in the 15-30 km altitude range. Updated Sato et al. data are obtained from NASA monthly AOD at 550nm. The satellite AOD data over 1991-2016 2017 period were computed at 532 nm. The Sato et al. data set was converted to 532 nm according to Khaykin et al. (2017). The merged AOD proxy was obtained by normalizing the Sato et al. time series to the SAGE II data in December 1991. The regression code uses the AOD values in April before the complete formation of the vortex in order to avoid possible contamination of aerosols satellite data by Polar Stratospheric Clouds. The April AOD proxy is represented by a bold black line in Fig. 5 together with Sato et al. (1993) and satellites datasets for the 1991-2016 2017 period.

A new GRAD(t) proxy was developed in order to take into account the stability of the vortex during the studied period. This proxy corresponds to the maximum gradient of PV as a function of EL at 550 K during both studied periods (e.g. September and 15Sept-15Oct). It is calculated from ERA-Interim data. GRAD and HF proxies are detrended by removing a 3rd order polynomial fit to minimize correlation with PWLT proxies. Figure 6 displays GRAD and HF proxies before and after removing trends. An anti-correlation of 0.45 ~0.55 between these two proxies is observed with a p-value <0.01, but the addition of GRAD proxy provides a much better agreement between measurements and model, especially during the last decade. The contribution of the GRAD(t) proxy to the improvement of the MLR results is discussed in Sect. 5.2.3.

For the long-term trends, two piecewise linear trend ($PWLT(t) = C_{t1}\text{t1}(t) + C_{t2}\text{t2}(t)$) functions calculated before and after the turnaround year are usually used to estimate the change of slope in the long-term evolution of ozone linked due to ODS (e.g. Reinsel et al., 2002; Kuttipurath et al., 2013, de Laat et al., 2015). In this work our a Modified PWLT model (PWT) is used uses an additional function in order to take into account the slower growth of ODS near the turnaround year and the ozone loss saturation effect inside the within the Antarctic polar vortex in October (Yang et al., 2008). The PWT model is represented by Eq. (2):

$$PWT(t) = C_{t11}\text{t11}(t) + C_{t12}\text{t12}(t) + C_{t2}\text{t2}(t) \quad (2)$$

where C_{t11} and C_{t2} are the coefficients of the linear functions and C_{t12} of the parabolic function. First The first period is represented by a linear time proxy t11 and a parabolic time proxy t12. The second period is expressed only by a linear time proxy t2. The proxies are stated computed as follow:

$$\text{t11} = \begin{cases} t & 0 < t \le T_0 \\ T_0 & T_0 < t \le T_{end} \end{cases} \quad (3)$$

$$\text{t12} = \begin{cases} \left(t - \left(\frac{T_0+1}{2}\right)\right)^2 & 0 < t \le T_0 \\ \left(\frac{T_0-1}{2}\right)^2 & T_0 < t \le T_{end} \end{cases} \quad (4)$$

$$\text{t2} = \begin{cases} 0 & 0 < t \le T_0 \\ t - T_0 & T_0 < t \le T_{end} \end{cases} \quad (5)$$

T_0 corresponds to the turnaround year in the considered period. In this work, 2001 was selected as the turnaround year when equivalent effective stratospheric chlorine (EESC) maximizes for a mean age-of-air of 5.5 yr (Newman et al., 2007). The corresponding value for T_0 is 22. T_{end} corresponds to the number of years considered in the study (37 38 for 1980-2016 2017). The minimum of the parabolic time proxy t12 is set to the middle of the period before the turnaround year so that the slope of the proxy is zero on that year. In this case the coefficient of t11 (C_{t11}) can be considered as the linear trend before 2001. After 2001, t11 and t12 are constant and then the linear trend is given by t2 coefficient (C_{t2}) the C_{t2} coefficient. Figure 7 represents the evolution of the three piece-wise proxy anomalies normalised by the corresponding standard deviation. The improvement using PWT instead of PWLT is discussed in Sect. 5.2.4.

5.2 Results Trend results for the averaged total ozone column records

The multi-regression model described in previous section was applied to SAT and MSR-2 total ozone anomalies series (anomalies time series computed as monthly total ozone – mean total ozone) of for the September and 15Sept.-15Oct. periods and to SAT for the 15Sept-15Oct period only. Times series of total ozone data corresponding to the different classification methods described in Sect. 4 were also used to evaluate the impact of vortex baroclinicity on total ozone trends.

5.2.1 September

A rapid decrease of ozone levels occurs within the polar vortex in Antarctica from the last two weeks of August to the end of September when the necessary sunlight to start the ozone catalytic destruction cycles is present again above austral polar regions. Important differences of in total ozone levels are found inside the vortex between the first and second half of September. Very low values are especially observed during the last week. Despite total ozone variability during September, most recent studies were performed to detect the first signs of ozone recovery linked to ODS decrease since 2000 using multi-linear regression model on ground-based and satellite datasets or chemistry-climate / chemistry-transport models simulations Although pronounced decrease in total ozone is observed in September, recent works have used ozone records obtained during this month to detect the ozone recovery (Solomon et al., 2016; Chipperfield et al., 2017; Weber et al., 2017). Those papers used studies use data or simulations poleward of ~60°S and conclude that identify first signs of Antarctic ozone recovery are highlighted in Antarctica for for September but not yet on for October due to the larger dynamical variability during that month. In this work, results from our multi-regression model will be are evaluated and compared to those previous works for the September period. In our study, trend estimation was performed before and after 2001 as described in the previous section. Figure 8 illustrates the results of the regressions model described in section 5.1 for the SAT MSR-2 total ozone data series inside the vortex using the 400 K-600 K range classification. The top panel represents the deseasonalized total ozone observations as well as the regressed ozone values. The model results reproduce quite well the interannual variability of measurements except in 2002 when the vortex split in two parts in late-September due to a major sudden stratospheric warming (e.g., Allen et al., 2003). Likewise, the year 2000 characterised was characterized by a large ozone hole area in September, and yields presents a relatively high value of residual of ~2030 DU on that year. Contributions of the different proxies are shown in the second to fourth panels of Fig. 7. Fitted HF and GRAD were added (black line in second panel of Fig. 8) due to the correlation between both proxies. The model term linked to the HF+GRAD fitted proxy represents the second largest contribution to total ozone interannual variability (~10 % of the total variance) after the PWT proxy which contributes to about 80 % of the total variability. Other proxies contributions to interannual ozone variability (third panel of Fig. 8) represent only 1 % of total ozone variability. Aerosol contribution of proxy contributes by respectively ~9 6 DU in 1992 and ~3 DU in 1983 due linked to Pinatubo in 1992 and ~6 DU due to and El Chichon eruptions. significantly helps HF+GRAD fitted proxy to explain ozone variation on those years. Negligible impact is seen in other years. Fitted QBO (QBO30hPa + QBO10hPa) explains ±5 DU ozone variability. The contributions of SF and AAO proxies ~~is~~ are negligible.

The model explains ~~93~~ 92 % of the ozone variability as deduced from the determination coefficient R^2. The estimated total ozone trends before and after 2001 are ~~-5.55±0.71 DU yr^{-1} (-26.7±3.4 % decade^{-1})~~ -5.31±0.67 DU yr^{-1} (-25.2±3.2 % decade^{-1}) and ~~1.87±1.18 DU yr^{-1} (9±5.7 % decade^{-1})~~ 1.84±1.03 DU yr^{-1} (8.8±4.9 % decade^{-1}), respectively. Both trends are significant (i.e. statistically different from zero) at 2σ. The 1980-2000 period presents higher depletion rate compared to Weber et al., 2017 (from -12 to -19% per decade depending on dataset) and comparable rate for ~~2001-2016~~ the period of recovery (8-10 % decade^{-1}). Comparable values of trends are found when the 475 K classification level is used (~~-21.3±3.6 % decade^{-1}~~ (-21±3.2 % decade^{-1} and ~~11±5.9% decade^{-1}~~ 10.1±5 % decade^{-1}). The 400 K-600 K classification allows us to obtain the best agreement between observations and regressed values (higher R^2) and lower χ ($\sqrt{\sum_i (obs_i - mod_i)^2/(n-m)}$) of residuals. Those results are represented in Table 2 ~~for SAT and~~ for MSR-2 total ozone datasets inside the vortex and for the three classifications analysed in this study. ~~Similar results are observed between both data series. MSR-2 presents slightly higher trend values after 2001. Note that 2016 was not available in MSR-2 data. SAT dataset is shortened to 2015 to evaluate sensibility of our model to the year 2016. Similar results are found compared to MSR-2 but with slightly higher values of trends after 2001, still significant at 2σ.~~ Despite trend values after 2001 for the 475 K classification are ~~systematically~~ higher by about ~~35~~ 28 % than for the 400 K-600 K classification range, trend results between both classifications are not significantly different at 2σ level, suggesting a limited effect of vortex baroclinicity on trend estimation using MLR analysis. The different results in Table 2 generally present a ratio between trends before and after 2001 ~~smaller than~~ close to 3, similar to that of ODS trends before and after the peak (Chipperfield et al., 2017). ~~Thus~~ This indicates that the ozone recovery trend could be due to ODS decrease. Nonetheless this trend cannot be reliably associated to chemical processes only and other processes could also play a role.

Computed trends over the 2001-~~2016~~ 2017 September period obtained with our model range from ~~1.85 to 2.67~~ 1.84 to 2.36 DU yr^{-1} for all cases studied. They are all significant at 2σ level. Solomon et ~~al.,~~ al. (2016) found significant total ozone trend of 2.5±1.7 DU yr^{-1} in September from SBUV and ozonesonde observations and similar results from the chemistry+dynamics+volcanoes (Chem-Dyn-Vol) simulation (2.8±1.6 DU yr^{-1}) using the Whole-Atmosphere Community Climate Model (WACCM). Estimated total ozone trend when only chemistry is considered in the model (Chem-Only) correspond to only half of the final trend (1.3±0.1 DU yr^{-1}).

A simulation test was done to evaluate the pertinence of using other proxies than PWT, HF and GRAD since only these fitted proxies present significant regression coefficient values at 95 % confidence interval. Results are represented in Table 2. Slightly lower determination factor R^2 is computed if only PWT, HF and GRAD are considered for September and comparable residual and trends~~. MSR-2 data series have presented alike results~~. This results suggest that the others proxies provide marginal improvement to the MLR analysis.

5.2.2 September 15 to October 15

In order to confirm healing of the Antarctic ozone hole, it is important to evaluate trends for the period where lowest total ozone values are observed inside the vortex e.g. between September 15th and October 15th. The same analysis as for September is thus performed. Figure 9 illustrates the ~~regressions model~~ results of regressions model for total ozone of ~~SAT~~ MSR-2 data series inside the vortex using the 400 K-600 K classification. ~~The~~ It shows that the interannual variability of measurements is better represented by the model than in September~~, with~~. For that period, the determination coefficient R^2 ~~of~~ is 0.95 (see also Table 3). As for the September regression, the sum of fitted HF and GRAD proxies (black line ~~on~~ in second panel) represents ~~together~~ the second largest contribution to total ozone interannual variability (~13 % of the total variance) after the PWT proxy (~80 %) and the last decade of measurement is correctly reproduced by the model. Significant trends of ~~-5.86±0.6 DU yr^{-1} (-30.2±3.2 % decade^{-1})~~ -5.81±0.6 DU yr^{-1} (-29.8±3 % decade^{-1}) and ~~1.27±1 DU yr^{-1} (6.8±5.6 % decade^{-1})~~

1.42±0.92 DU yr^{-1} (7.3±4.7 % $decade^{-1}$) are estimated before and after 2001. Similar results are observed if a single level classification is used with higher trend values after 2001 for 475 K. All trend results are comparable within ±2σ. ~~MSR-2 presents~~ Results based on the SAT record are similar ~~results~~ with slightly higher trend values after 2001. Note that the addition in the MLR analysis of the 2 most recent years (2016-2017), which were characterized by weak ozone holes, changed the significance of the 2001-2017 trend from hardly significant to significant more than 2σ. Results obtained in the 1980-2017 period by the MLR analysis show thus for the first time a significant recovery in the 15Sept-15Oct period.~~except for classification at 550 K where trend after 2001 is significant at only 1.8σ~~. Solar Flux, QBO, Antarctic Oscillation and Aerosol (third panel of Fig. 9) explain ~1 % of the total variance. QBO explains ±3 DU interannual variability and Aerosol signal amount to ~~~7~~ 6 DU and ~~~5~~ 3 DU linked to Pinatubo in 1992 and El Chichon in 1983. SF contribution varies from 4.5 DU during the maximum (except for the last solar cycle, ~1 DU) to ~~-2.4~~ -2.2 DU during the minimum. AAO represents negligible contribution. Same test as for September was performed where proxies of SF, QBO, AAO and Aerosol were removed from the linear regression. Results are presented in Table 3. Negligible difference in trends, R^2 and residuals are observed if those proxies are considered or not in the MLR analysis. In addition, lower chi-values are found for smaller number of fitted parameters, which is the case for the regression using PWT, HF and GRAD only.

The different cases shown in Table 3 present significant trends at 2σ over the 1980-2000 and the 2001-2017 periods. ~~Trends estimated on second period present slightly significant values at 2σ except for MSR-2 at 550 K. Ratios of computed trends before and after 2001 are higher than 3 in both total ozone data series.~~ Computed trend with 400 K-600 K range classification is comparable to the Chem-Only trend calculated by WACCM in Solomon et al.~~, 2016~~ (2016). Despite the good agreement between regressed values and measurements especially for ~~Whatever ozone depletion could be considered as well constrained by~~ the period 15Sept-15Oct ~~(15Sep-15Oct)~~ and ~~by~~ for the range classification method (400 K-600 K), ~~no direct relationship to healing could be deduced~~ it is not possible to attribute ozone significant increase to ODS decrease. In addition, the ratio between trends before and after 2001 ~~(before/after 2001)~~ is ~~higher~~ larger than 3 ~~threshol d value to consider a recovery due to ODS decline~~ which could be due to the effect of desaturation of the ozone loss.

5.2.3 Impact of GRAD proxy on trend estimation

The HF proxy represents the cumulative effect of wave activity on vortex stability (e.g. a high HF corresponds to a warmer vortex) that seems insufficient to represent total ozone variability over the last decade, especially in 2010 and 2012. The GRAD proxy was developed in order to consider also the vortex stability during ~~the~~ both studied periods ~~(e.g. September and 15Sept-15Oct.)~~. ~~Figure 10 highlights the improvement in the regressed values by using the GRAD proxy~~. Since Aerosols, QBO, SF and AAO represent lower contribution to ozone variability, trend analyses using HF, PWT proxies only and including or not the GRAD proxy are performed in order to highlight the impact of this parameter. Figure 10 shows residuals of MLR analysis with and without GRAD on MSR-2 data inside the vortex for the 400 K-600 K classification range for September and 15Sept-15Oct periods.. ~~It shows the residuals of the trend simulation with and without GRAD as explanatory variable in the model. SAT measurements inside the vortex applying 400 K-600 K classification were used for this study. Results for September and 15Sept-15Oct periods are shown in the top and bottom panels of Fig. 10 respectively. As shown in Table 1, the residuals' variance (χ) is significantly reduced with the addition of the GRAD proxy.~~ The residual anomalies are significantly reduced after 2002 when GRAD is used, especially in the 15Sept-15Oct period. The second panels of Figures 8 and 9, show that in some years HF and GRAD proxies are in phase as during 2009-2014 when GRAD intensifies the HF contribution to ozone variability. This improvement is especially visible for the years 2010 and 2012. When both proxies are anticorrelated, as in ~~2015 and 2016~~ 2005-2008, the improvement linked to the GRAD proxy is also observed. ~~As a further illustration,~~ Table 2 and 3 show the results of the regressions excluding GRAD proxy for September and 15Sept-15Oct respectively. The determination coefficient is generally reduced by ~~~0.05~~ 0.07 and the χ values are ~~higher by 20 % to 35 %~~ 25 % to 50 % higher. Trend values are mostly similar but the error bars are reduced when GRAD is used as explanatory variable, especially after 2001. Computed trends before and after 2001 show lower and higher trends in absolute values respectively but within ±2σ of initial trend results including the GRAD proxy. Trends over the 2001-2016 2017 period estimated without the GRAD proxy are still significant at 2σ in September but only at 1.8σ for the and 15Sept-15Oct for both datasets.

5.2.4 PWT vs PWLT

In order to evaluate the improvement of an additional parabolic function to the linear functions of the piece-wise trend proxy, the classical piece-wise linear trends (PWLT) is applied in the MLR analysis of MSR-2 datasets. Figure S1 shows average total ozone anomalies of MSR-2 inside the vortex (400 K-600 K range classification method) in September and 15Sept-15Oct and the retrieved trends using both the PWLT and PWT methods. In the case of the 15Sept-15Oct period, the PWT model provide a better representation of long-term ozone evolution compared to PWLT, as it better captures ozone loss saturation during the 1990s. The trends error bars are also smaller using PWT before and after 2001. In addition, a better agreement between measurements and model values is observed with a higher R^2 and lower residuals. The 2001-2017 trend error bars are ~60 % higher if PWLT is used and the trend value itself is nearly double. In the case of September, a slight improvement in R^2, residuals and error bars is obtained with PWT. The 2001-2017 trend value with PWLT is 40 % larger.

5.3 Results using OMD metric

OMD has been used in previous studies to evaluate ozone loss and ozone recovery (e.g. de Laat et al., 2017). This metric has the advantage to be independent of the vortex position. Total ozone MSR-2 data were used to compute the average daily OMD on September and 15Sept-15Oct periods. The total ozone columns are referenced to the 220 DU threshold value and the corresponding mass deficit of the partial column (220 DU – total ozone column) is computed at each grid point (e. g., Bodeker and Scourfield, 1995). Only total ozone columns south of 60°S and lower than 220 DU are considered and the daily OMD correspond to the sum of OMD at each pixel multiplied by the cosine of the latitude and the square of the Earth's radius. Table 4 shows the MLR analysis of OMD using different sets of proxies as for ozone average in Table 2 and 3. The contributions of Aerosols, AAO, QBO and SF do not shown an impact on MLR analysis where similar R^2, χ, trend and error bars values are obtained. On the other hand, the inclusion of GRAD results in higher R^2 and lower residuals in both periods. The effect of GRAD is more important in 15Sept-15Oct as shown for averaged total ozone.

For the different cases and periods shown in Table 4, the OMD trend values are significant at 2σ. The MLR analysis using GRAD, HF and PWT proxies provides trends of -1.29±0.24 Mt yr^{-1} and 0.86±0.36 Mt yr^{-1} in September and -1.61±0.22 Mt yr^{-1} and 0.65±0.33 Mt yr^{-1} in 15Sept-15Oct. De Laat el al. (2017) found a similar trend for the recovery period of 0.77 Mt yr^{-1} for the averaged OMD between the 220 and 280 day of year. Figure 11 displays the comparison between the OMD records and results of MLR analysis for the September and 15Sept-15Oct period, together with the trend components of the model. The effect of ozone loss saturation is particularly visible in the 15Sept-15Oct period. There are some years that are not totally explained by the model, e.g. 2002 and 2004 for both periods and 2000 for September. The contributions of GRAD, HF and GRAD+HF are shown in upper panels of Fig. S2 where GRAD intensifies HF contribution in 2010 and 2012, while both proxies are anti-correlated in 2005-2008 as observed for the total ozone analysis. The residuals with and without GRAD are shown in the bottom panels of Fig. S2. The improvement linked to the use of GRAD proxy is particularly visible in the last decade.

As for total ozone, MLR analysis using PWLT was performed for comparison with the PWT model. Figure S3 shows the OMD records together with PWLT and PWT components of the regression model for the both periods. Similar agreement is obtained for September but the regression results in higher residuals for 15Sept-15Oct using PWLT (not shown). Major difference is observed in the period 2001 - 2017 with a large trend value of -0.91±0.41 Mt yr^{-1}, corresponding to an increase of 40 % in absolute value.

6. Temporal evolution of low total ozone values inside the vortex

The ozone hole is generally defined as the region with total ozone columns lower than 220 DU. This standard value was used in different studies to evaluate the ozone depletion from the Ozone Hole Area (OHA) (e.g. Newman et al., 2006; Solomon et al., 2016) or the Ozone Mass Deficit (OMD) (e.g. de Laat and van Weele, 2011, de Laat et al., 2017) metrics. ~~In this work, the relative daily area inside the vortex with~~ In order to evaluate how the ozone hole is influenced by very low ozone values, the surface relative to the vortex area occupied by ozone values lower than different threshold levels is computed for each day and ~~integrated~~ averaged over different periods (September, 15Sept-15Oct and October). The top panel of ~~Figure 11~~ Fig. 12 shows the evolution of these average relative areas ~~with respect to the vortex's area of these low values relative to three~~ for five different thresholds: 220 DU, 200 DU, 175 DU, 150 DU and 125 DU, for the 15Sept-15Oct period. MSR-2 ~~and TOMS/OMIT~~ datasets ~~are~~ is used for this analysis and vortex areas ~~is~~ are estimated ~~with~~ using the 400 K-600 K range classification, the results of previous sections having shown that the range classification better constrains the ozone hole area. ~~Relative areas at 220 DU of both satellite datasets are similar. They~~ Results show increasing ~~amounts~~ areas during the 1980s, a stabilisation in the 1990s and a higher inter-annual variability since 2001. ~~The evolution of relative areas of MSR-2 with respect to 125DU and 150DU thresholds presents lower values than TOMS-EP/OMIT data. Nevertheless comparable inter-annual variability is observed.~~ In contrast to the 220 DU threshold case,~~relative areas~~, the evolution of relative areas ~~to 150DU and 125DU~~ corresponding to lower thresholds shows a delayed increase from the ~~mid-~~beginning of the1980s ~~and~~ to ~~mid-~~ the early 1990s ~~respectively~~, reaching a maximum in ~~both~~ all cases in 2000. After 2000, a larger interannual variability is generally observed and from 2006 ~~and~~ a steady decrease is seen for thresholds lower than 200 DU. ~~from 2006, with very large inter-annual variability from 2000.~~ ~~In the three thresholds cases, several~~ In all cases, several anomalous years are observed with important reduction of ozone depletion: 1988, 1991, 2002, 2004, 2010 and 2012. Note that these years correspond to a high contribution of HF+GRAD proxies to the regressed ozone values (Fig. 10, second panel). If we exclude these anomalous years, the 220 DU relative area ~~corresponds to~~ remains fairly stable at about 90% of the total vortex in average since 1990. In the most recent years, ~~125 DU and 150 DU~~ relative areas for 125 DU and 150 DU thresholds decrease to less than 10 % and 30 % respectively ~~of the vortex area~~ from their peak value of 21 % and 57 % reached in ~~of~~ 2000. If such trend persists, the frequency of very low ozone values (e.g. below 125 DU) is expected to become negligible in the coming decade.

In addition, OMD were computed for the same thresholds (bottom panel of Fig; 10). The evolutions of OMD present similar behaviour as the relative area but in this case, OMD at 220 DU threshold shows a visible decrease since 2000. Nowadays OMD of threshold lower than 150 DU presents very small values lower than 0.2 Mt.

Solomon et al. (2016) have ~~shown a slow~~ highlighted for the first time a delay in the formation of the ozone hole after 2000. This shift can be explained by the slower ozone loss rates after ~~polar sunrise~~ sun appearance over the Pole, due to ODS decrease in the polar stratosphere. In this work, ~~a~~ such possible time shift was investigated by ~~comparing relative area inside the vortex lower than 220 DU in September, 15Sept-15Oct and October (Fig. 12). Higher relative area is observed in the 1980s in October, followed by 15Sept-15Oct and much smaller values for September. At the end of 1980s and 1990s, 15Sept-15Oct presents the higher relative area and in 2000s it shows similar area as in September. Since 2010, October areas are similar to those in 15Sept-15Oct, while September areas show increasingly smaller values. This divergence could be explained by a delay in the full formation of the ozone hole in September~~ computing the first day when ozone levels below certain thresholds occur inside the vortex (using the 400 K-600 K classification range), from September 1^{st} to October 15^{th} (Fig. 13). The same thresholds values as for Fig. 12 were used. In order to avoid influence of spurious values, the number of 1°x1° grid cells with total ozone columns below the various thresholds in the first day (or start day) has to be higher than 10. For each curve, day values equal to 244/245 correspond to years when ozone levels below the corresponding threshold have appeared at or before the beginning of September. For the 220 DU threshold, the dark blue curve shows that this is the case since 1983. For the 200 DU threshold, lower ozone values appear before the beginning of September after the mid

1980s. For the other thresholds, we observe a decrease, with some variability, of the start day during the 1980s, and the 1990s for the two lowest thresholds and an increase after 2000 - 2005. This increase is most visible on the 125 DU threshold curve and to some extent also on the 150 DU threshold curve. In 2016, ozone levels below 150 DU have appeared in the beginning of September, typically as for ozone holes at the end of the 1990s but levels below 125 DU still appear later. No values for a particular year in the threshold curves indicate that total ozone levels were above that threshold during the whole period considered. This is the case for the two lower thresholds before 1985 and for the 125 DU threshold in 2002, 2004 and 2017.

7 Conclusions

Two satellite based data series MSR-2 and SAT (TOMS/OMI with gaps in 1993-1995 filled by MSR-2) datasets have been used to evaluate total ozone trends within the Southern polar vortex over the 1980-2016 2017 period. A multi-regression model is applied to ozone values averaged over the September month and the 15 September to 15 October period in order to compute long-term trends before and after the maximum of EESC ODS peak in the polar stratosphere that occurred in around 2001 (Newman et al., 2007). The 15 Sept. – 15 Oct. time range corresponds to the period of maximum ozone depletion. It is not commonly used in previous works. Proxies and time windows for averaging them are selected following de Laat et al. (2015) work.

Different methods to classify For the classification of total ozone measurements inside the vortex , the classical Nash et al. (1996) method is used.. are used in In order to evaluate the impact of vortex baroclinicity on trend analysis, classifications using a single isentropic levels (475 K, 550 K) and a range of levels (400 K – 600 K) are tested. Systematic differences are found between the various. The total ozone time series. using single level (475 K or 550 K) for classification show systematic differences compared to 400 K-600 K classification range but their However the inter-annual variability is similar with correlation coefficients ranging from 0.98 to 0.99 in both studied periods. R>0.97 for both cases. Trends estimation results in higher absolute values using the Nash et al. criterion at 475K level compared to 550K level or the more restrictive condition based on the 400 K-600 K altitude range. However the corresponding computed trends are not significantly different at While larger trend values are generally found with the 475 K classification, the differences with trends related to the 400 K – 600 K range classification are not significant at 2σ level.

The use of combined piece-wise linear and parabolic functions for the trend proxies (PWT) in the 1980 – 2000 and 2001 – 2017 periods provides a good representation of the total ozone long term behaviour inside the vortex (after removal of interannual variability), especially for the 15Sept-15Oct period, probably in relation with the effect of ozone loss saturation. The classical PWLT used in previous studies seems to overestimate the trends during the recovery period.

A new proxy (GRAD) representing the vortex stability over the both studied periods (September and 15Sept–15Oct), the GRAD proxy, is used is included in the multilinear regressions regression. This proxy improves the representation of total ozone inter-annual variability by the regressed values especially over the last decade, with. It results in ~0.05 higher value for the R^2 determination coefficient, and lower fitted residuals and smaller trend uncertainties for the different classification methods and datasets. The contribution of GRAD is particularly important since 2010. For the various time series and classification criteria used in this study, In general, the best agreement between observations and regressed values is found for the 15Sept-15Oct period with higher R^2 coefficients.. While the HF combined with GRAD proxies reproduce quite well the interannual variability of ozone, other proxies such as Aerosols, QBO, SF and AAO present smaller explanatory power and contribute less to reduce trend uncertainties.

In the period of increasing ODS (1980-2000), the MLR analysis shows negative and significant trends for both studied periods, similar to values found in previous studies (e.g. Kuttippurath et al., 2013 and de Laat et al., 2015). The 15Sept-15Oct period presents slightly higher negative trends in absolute value than the month of September.

In the 2001-2017 period, positive trends are obtained for all scenarios. The largest trends and highest significance are found for the September period, with a trend value of 1.84 ± 1 DU for the MSR-2 total ozone record using the 400 K-600 K range classification method. For the 15Sept-15Oct period, a lower trend of 1.42 ± 0.92 DU is obtained using the same record. Better fit and smaller residuals are obtained for that period. Differences with trend results from the other SAT data set evaluated in the study are not statistically significant.

The ratio between trends before and after 2001 varies according to the studied period. Only September trends present a ratio of ~3 as expected for an ozone response to ODS evolution. However, as for other trend studies based on MLR fit to observations, it is not possible from this analysis to fully attribute the retrieved trends to ODS evolution only.

~~Comparable negative trends before 2001 ranging from -4.9 DU.yr^{-1} to -5.6 DU yr^{-1} significant at 2σ levels are obtained for both periods (Sept and 15Sept-15Oct), and for the different data series and vortex classifications. Significant positive trends after 2001 of 2.2 DU.yr^{-1} in average are obtained for the September time series for all data sets, classification methods and regression analyses. The ratio between trends before and after 2001 is smaller than 3, in line with the slower decrease of ODS amounts in the stratosphere after the turnaround year. Despite this, it is not possible to ensure that the estimated trends are only due to ODS evolution. Regarding the 15Sept-15Oct period when ozone depletion is larger, retrieved trends are smaller with an average of 1.4 DU.yr^{-1} for all studied cases. In contrast to the September trends, the ratio before and after 2001 is generally larger than 3. The trends are slightly significant at 2 sigma level except for the classification at 550K. The best trend model applied on the data using the 400K-600 K range classification method shows increasing trends of respectively 1.27 ±1.0 DU.yr-1 and 1.18±1.12 DU.yr-1 for the SAT and MSR-2 records.~~

The evolution of ozone mass deficit was also analysed using MSR-2 data. MLR analysis on this metric confirms the findings obtained for total ozone columns, e.g. a general improvement of the fits with the GRAD proxy and the main explanatory power provided by the GRAD, HF and PWT proxies. The 2001-2017 OMD trends are higher in absolute value for September (-0.86±0.36 Mt.yr^{-1}) than for 15Sept-15Oct (-0.65±0.33 Mt.yr^{-1}). They are significant at 2σ level in both cases. These results are in general good agreement with those obtained in De Laat et al. (2017). Similar reductions of 53 % and 35 % of OMD are computed for September and 15Sept-15Oct respectively, which seem overestimated considering the ODS decrease during the 2001- 2017 period.

The structural uncertainties of the MLR analysis linked to the selection of proxies were not fully analysed in this work, as in De Laat et al. (2015). The main sensitivity tests concerned the baroclinicity of the vortex and the impact of its stability during the studied periods. Trend differences in the various scenarios analysed provide some quantification of related uncertainties and are lower than the statistical trend uncertainties. Further, the large determination coefficients obtained for both periods analysed give confidence in the retrieved trends. The Heat Flux proxy that provides the largest explanatory power in the various fits is a well-known driver of vortex temperature conditions that are the primary causes of polar ozone depletion in periods of high ODS levels. The influence of the GRAD proxy in recent years highlights the importance of the vortex stability for the containment of the ozone hole during the period of maximum depletion.

Polar ozone recovery was also evaluated by examining the temporal evolution of ~~low~~ relative areas occupied by ozone ~~values~~ levels below various thresholds within the vortex. ~~Areas with values lower than 125, 150 and 220 DU relative to the vortex area were computed using both datasets and the 400K-600K classification.~~ Very small total ozone columns (<150 DU) did not occur inside the vortex before the late 1980s and early 1990s. For the 125 DU, ~~and~~ 150 DU and 175 DU thresholds, relative areas ~~show~~ display a steady decrease since the beginning of the 21st century, while for the 200 DU and 220 DU thresholds, the relative area's evolution is ~~more~~ quite stable. All ~~three~~ relative area curves are marked by increased variability since 2000. Relative areas related to the lowest thresholds show a more rapid decrease, which further points towards polar ozone recovery. OMD records based on the same thresholds show a similar behaviour.

~~The delay in the formation of the ozone hole can be evaluated by comparing the 220 DU relative areas for the September, 15-Sept-15Oct, and October periods. October (September) areas showed the largest (lowest) values until 1987. After that~~

year, the evolution of the curves was inverted and the largest relative areas were generally observed in September (88 % in average of the polar vortex occupied by total ozone columns lower than 220 DU) until the first decade of 2000s. Since 2010, October relative areas are similar to those of the 15Sept-15Oct period, with still more than 75 % of the vortex characterized by total ozone levels less than 220 DU in most of the years. These results confirm Solomon et al. (2016) findings on the delay in the formation of the Antarctic ozone hole in September.

In summary, this work shows clear signs of ozone recovery in September with significant trend values at 2σ and smaller but still significant increasing trends in the 15Sept-15Oct period, when ozone depletion is highest. Recovery in October is confirmed by the very low ozone values (below 150 DU) within the vortex since 2010. Comparison of ozone hole areas in September and October confirm the later formation of the ozone hole in September.

In summary, this work present clear symptoms of polar ozone recovery. Recovery is found for the month of September and for the first time for the period of maximum ozone depletion, e.g. from September 15 to October 15. For both studied periods, recovery is deduced from the significant positive trends in total ozone, significant negative trends of ozone mass deficit and from the steady decrease of the occurrence of low ozone values within the polar vortex. As ODS continue to decrease in the next years, it is likely that ozone recovery in the Polar vortex in spring will become more evident.

Data availability. The source of the different total ozone column datasets and classical proxy time series utilised in this work are publically available from the websites given in the text and in Table 1. The satellite data used to build the Aerosol proxy are available at https://eosweb.larc.nasa.gov/project/calipso/calipso_table for CALIPSO; https://eosweb.larc.nasa.gov/project/sage2/sage2_table for SAGE II, http://odin-osiris.usask.ca/ for OSIRIS; http://mls.jpl.nasa.gov/products/h2o_product.php for MLS; and https://ozoneaq.gsfc.nasa.gov/data/omps/ for OMPS V1 LP. Other data as equivalent latitude and GRAD proxy are available upon request.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. The authors thank NASA/GSFC and TEMIS for total ozone column data of TOMS/SBUV/OMI-TOMS and MSR-2, respectively. They are grateful to Cathy Boone of ESPRI data centre of Institut Pierre Simone Laplace (IPSL) to provide ERA interim data. This work was supported by the Dynozpol/LEFE project funded by the French Institut National des Sciences de l'Univers (INSU) of the Centre National de la Recherche Scientifique (CNRS).

References

Allen, D., Bevilacqua, R., Nedoluha, G., Randall, C., and Manney, G.: Unusual stratospheric transport and mixing during 2002 Antarctic winter, Geophys. Res. Lett., 30, 1599, doi: 10.1029/2003GL017117, 2003.

Bhartia, P. K., and Wellemeyer C.: TOMS-V8 total O3 algorithm, in OMI Algorithm Theoretical Basis Document , vol. II, OMI Ozone Products, edited by P. K. Bhartia, pp. 15–31, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA., 2002.

Bodeker, G. E. and Scourfield, M. W. J.: Planetary waves in total ozone and their relation to Antarctic ozone depletion, Geophys. Res. Lett., 22(21), 2949–2952, 1995.

Bodeker, G. E., Struthers, H., and Connor, B. J.: Dynamical containment of Antarctic ozone depletion, Geophys. Res. Lett., 29(7), 1098, doi:10.1029/2001GL014206, 2002.

Chipperfield, M, Bekki, S., Dhomse, S, Harris, N. R. P., Hassler, B., Hossaini, R, Steinbrecht, W., Thiéblemont, R., and Weber M.: Detecting recovery of stratospheric ozone layer, Nature, 549, 211–218, doi:10.1038/nature23681, 2017.

Chubachi, S.: A special observation at Syowa station, Antarctica from February 1982 to January 1983, in: Atmospheric Ozone, edited by: Zerefos, C. and A. Ghazi, Springer Netherlands, 285-289, doi:0.1007/978-94-009-5313-0_58, 1985.

Farman, J. C., Gardiner, B. G., and Shanklin, J. D.: Large losses of total ozone in Antarctica reveal seasonal ClOx/NOx interaction, Nature, 315, 207–210, doi:10.1038/315207a0, 1985.

De Laat, A. T. J., van Weele, M., and van der A, R. J.: Onset of Stratospheric Ozone Recovery in the Antarctic Ozone Hole in Assimilated Daily Total Ozone Columns, J. Geophys. Res., 122, 11880–11899, https://doi.org/10.1002/2016JD025723, 2017.

de Laat, A. T. J., van der A, R. J., and van Weele, M.: Tracing the second stage of ozone recovery in the Antarctic ozone-hole with a "big data" approach to multivariate regressions, Atmos. Chem. Phys., 15, 79-97, doi:10.5194/acp-15-79-2015, 2015.

de Laat, A. T. J. and van Weele, M.: The 2010 Antarctic ozone hole: Observed reduction in ozone destruction by minor sudden stratospheric warmings, Sci. Rep., 1, 38, doi:10.1038/srep00038, 2011.

Dee, D.P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhle.r, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q. J. R. Meteorol. Soc., 137, 553-597, doi:10.1002/qj.828, 2011.

Frederick, J. E., Cebula, R. P., and Heath, D. F.: Instrument characterization for the detection of long-term changes in stratospheric ozone: An analysis of the SBUV/2 radiometer, J. Atmos. Oceanic Technol., 3, 472–480, doi:10.1002/jgrd.50503, 1986.

Godin, S., Bergeret, V., Bekki, S., David, C., and Mégie, G.: Study of the interannual ozone loss and the permeability of the Antarctic polar vortex from aerosols and ozone lidar measurements in Dumont d'Urville (66.4°S, 140°E), J. Geophys. Res., 106(D1), 1311-1330, doi:10.1029/2000JD900459, 2001.

Hauchecorne, A., Chanin, M.-L., and Keckhut, P.: Climatology and trends of the middle atmospheric temperature (33-87 km) as seen by Rayleigh lidar over the south of France. J. Geophys. Res., 96, 15297–15309, doi:10.1029/91JD01213,1991.

Hauchecorne, A., Godin, S., Marchand, M., Heese, B., and Souprayen, C.: Estimation of the Transport of Chemical Constituents from the Polar Vortex to Middle Latitudes in the Lower Stratosphere using the High-Resolution Advection Model MIMOSA and Effective Diffusivity, J. Geophys. Res., 107(D20), 8289, doi:10.1029/2001JD000491, 2002.

Kerzenmacher, T. E., Keckhut, P., Hauchecorne, A., and Chanin, M.-L.: Methodological uncertainties in multi-regression analyses of middle-atmospheric data series, J. Environ. Monit., V8, 7, 682-690, doi: 10.1039/B603750J, 2006.

Khaykin, S.M., Godin-Beekmann, S., Keckhut, P., Hauchecorne, A., Jumelet, J., Vernier, J.-P., Bourassa, A., Degenstein, D.A., Reiger, L.A., Bingen, C., Vanhellemont, F., Robert, C., DeLand, M., and Bhartia, P.K., Variability and evolution of mid-latitude stratospheric aerosol budget from 22 years of ground-based lidar and satellite observations, Atmos. Chem. Phys., 17, 1829–1845, doi:105194/acp-17-1829-2017, 2017.

Kuttippurath, J., Goutail, F., Pommereau, J.-P., Lefèvre, F., Roscoe, H. K., Pazmino, A., Feng, W., Chipperfield, M. P., and Godin-Beekmann, S.: Estimation of Antarctic ozone loss from ground-based total column measurements, Atmos. Chem. Phys., 10, 6569–6581, doi:10.5194/acp-10-6569-2010, 2010.

Kuttippurath, J., Lefèvre, F., Pommereau, J.-P., Roscoe, H.K., Goutail, F., Pazmino, A., and Shanklin, J.D.: Antarctic ozone loss in 1979–2010: first sign of ozone recovery, Atmos. Chem. Phys., 13, doi:10.5194/acp-13-1625-2013, 1625-1635, 2013.

Kuttippurath, J., Godin-Beekmann, S., Lefèvre, F., Santee, M. L., Froidevaux, L., and Hauchecorne, A.: Variability in Antarctic ozone loss in the last decade (2004–2013): High-resolution simulations compared to Aura MLS observations. Atmos. Chem. Phys. 15, 10385–10397, doi:10.5194/acp-15-10385-2015, 2015.

Lefèvre, F., Brasseur, G. P., Folkins, I., Smith, A. K., and Simon, P.: Chemistry of the 1991–92 stratospheric winter: three dimensional model simulations, J. Geophys. Res., 99, 8183–8195, doi:10.1029/93JD03476, 1994.

Levelt, P. F., van den Oord, G. H. J., Dobber, M. R., Mälkki, A., Visser, H., de Vries, J., Stammes, P., Lundell, J. O. V. and Saari, H.: The Ozone Monitoring Instrument, IEEE Trans. Geosci. Remlote Sens., vol. 44, 5, 1093-1101, May 2006.

McIntyre, M. and Palmer, T.: The 'surf zone' in the stratosphere, J. Atmos. Terr. Phys., 46, 825–849, doi:10.1016/0021-9169(84)90063-1, 1984.

McPeters R.D, Krueger, A.J., Bhartia, P.K., Herman, J.R., Wellemeyer, C. G., Seftor, C. J., Jaross, G., Torres, O., Moy, L., Labow, G., Byerly, W., Taylor, S. L., Swissler, T., and Cebula, R. P.: Earth Probe Total Ozone Mapping Spectrometer (TOMS) Data Products User's Guide, NASA Reference Publication 1998-206895, NASA, Washington DC, 1998.

McPeters, R., Kroon, M., Labow, G., Brinksma, E., Balis, D., Petropavlovskikh, I., Veefkind, J. P., Bhartia, P. K., and Levelt, P. F.: Validation of the Aura Ozone Monitoring Instrument total column ozone product, J. Geophys. Res., 113, D15S14, doi:10.1029/2007JD008802, 2008.

McPeters, R. D., Bhartia, P. K., Haffner, D., Labow, G. J., and Flynn L.: The version 8.6 SBUV ozone data record: An overview, J. Geophys. Res. Atmos., 118, 8032–8039, doi:10.1002/jgrd.50597, 2013.

Nair, P.J., Godin-Beekmann S., Kuttippurath J., Ancellet G., Goutail F., Pazmino A., Froidevaux L., Zawodny J.M., Evans R.D., Wang H.J., J. Anderson, J., and Pastel, M.: Ozone trends derived from the total column and vertical profiles at a northern mid-latitude station, Atmospheric Chemistry and Physics, 13, 20, 10373-1038, doi:10.5194/acp-13-10373-2013, 2013.

Nash, E. R., Newman, P. A., Rosenfield, J. E., and Schoeberl, M. R.: An objective determination of the polar vortex using Ertel's potential vorticity, J. Geophys. Res., 101, 9471–9478, doi:10.1029/96JD00066, 1996.

Newman, P. A., Nash, E. R., Kawa, S. R, Montzka, S. A., and Schauffler, S. M.: When will the Antarctic ozone hole recover?, Geophys. Res. Lett., 33, L12814, doi:10.1029/2005GL025232, 2006.

Newman, P. A., Daniel, J. S., Waugh, D. W., and Nash, E. R.: A new formulation of equivalent effective stratospheric chlorine (EESC), Atmos. Chem. Phys., 7, 4537-4552, doi:10.5194/acp-7-4537-2007, 2007.

Pazmino A. F., Godin-Beekmann, S., Ginzburg, M., Bekki, S., Hauchecorne, A., Piacentini, R., and Quel, E.: Impact of Antarctic polar vortex occurrences on total ozone and UVB radiation at southern Argentinean and Antarctic stations during 1997–2003 period, J. Geophys. Res., 110, D03103, doi:10.1029/2004JD005304, 2005.

Pazmino, A. F., Godin-Beekmann, S., Luccini, E. A. , Piacentini, R. D., Quel, E. J., and Hauchecorne, A.: Increased UV radiation due to polar ozone chemical depletion and vortex occurrences at southern sub-polar latitudes in the period [1997–2005], Atmos. Chem. Phys., 8, 5339–5352, doi:10.5194/acp-8-5339-2008, 2008.

Reinsel, G. C., Weatherhead, E. C., Tiao, G. C., Miller, A. J;, Nagatani, R. M., Wuebbles, D. J., and Flynn, L. E.: On detection of turnaround and recovery in trend for ozone, J. Geophys. Res., 107(D10), 4078, doi:10.1029/2001JD000500, 2002.

Salby, M.L., Titova, E.A., and Deschamps, L.: Changes of the Antarctic ozone hole: Controlling mechanisms, seasonal predictability, and evolution, J. Geophys. Res., 117, D10111, doi:10.1029/2011JD016285, 2012.

Solomon, S., Ivy, D. J., Kinnison, D., Mills, M. J., Neely, R. R., and Schmidt, A.: Emergence of healing in the Antarctic ozone layer, Science, doi:10.1126/science.aae0061, 2016.

Steinbrecht, W., Hasslet, B., Claude, H., Winkler, P., and Storlarski, R. S.: Global distribution of total ozone and lower stratospheric temperature variations, Atmos. Chem. Phys., 3, 1421-1438, doi:10.5194/acp-3-1421-2003, 2003.

van der A, R. J., Allaart, M. A. F., and Eskes, H. J.: Extended and refined multi sensor reanalysis of total ozone for the period 1970–2012, Atmos. Meas. Tech., 8, 3021-3035, https://doi.org/10.5194/amt-8-3021-2015, 2015.

~~Weber, M., Coldewey-Egbers, M., Fioletov, V. E., Frith, S. M, Wild, J. D., Burrows, J. P., Long, C. S., and Loyola., D.: Total ozone trends from 1979 to 2016 derived from five merged observational datasets - the emergence into ozone recovery, Atmos. Chem. Phys. Discuss, https://doi.org/10.5194/acp-2017-853, 2017.~~

Weber, M., Coldewey-Egbers, M., Fioletov, V. E., Frith, S. M., Wild, J. D., Burrows, J. P., Long, C. S., and Loyola, D.: Total ozone trends from 1979 to 2016 derived from five merged observational datasets – the emergence into ozone recovery, Atmos. Chem. Phys., 18, 2097-2117, https://doi.org/10.5194/acp-18-2097-2018, 2018.

WMO (World Meteorological Organization): Scientific assessment of ozone depletion: 2006, Global Ozone Research and Monitoring Project-Report 50, Geneva, Switzerland, 2007.

WMO (World Meteorological Organisation): Scientific assessment of ozone depletion: 2010, Global Ozone Research and Monitoring Project, Report 52, 516 pp., Geneva, Switzerland, 2011.

WMO (World Meteorological Organization): Scientific Assessment of Ozone Depletion: 2014, Global Ozone Research and Monitoring Project, Report No. 55, 416 pp., Geneva, Switzerland, 2014.

Yang, E.-S., Cunnold, D. M., Newchurch, M. J., Salawitch, R. J.; McCormick, M. P., Russell, J. M., Zawodny, J. M., and Oltmans, S. J.: First stage of Antarctic ozone recovery, J. Geophys. Res., 113, D20308, doi:10.1029/2007JD009675, 2008.

Table 1: Information of proxies (source, characteristics and time window for the mean yearly value).

Proxy	Source	Characteristics	Time window
HF	NASA/Goddard Space Flight Center https://acd-ext.gsfc.nasa.gov/Data_services/met/ann_data.html	45-Day Mean Heat Flux between 45°S-75°S at 70hPa from MERRA 2	Aug.-Sept.
SF	Dominion Radio Astrophysical Observatory (National Research Council Canada) ftp://ftp.geolab.nrcan.gc.ca/data/solar_flux/monthly_averages/solflux_monthly_average.txt	Monthly averages of Solar Flux at 10.7cm wavelength	Sept.
QBO	Institute of Meteorology (Freie Universität Berlin) http://www.geo.fu-berlin.de/en/met/ag/strat/produkte/qbo	Monthly mean Quasi-Biennial Oscillation at 30 and 10hPa	Sept.
Aer	1980-1990: NASA/Goddard Space Flight Center https://data.giss.nasa.gov/modelforce/strataer/	AOD@550nm, 15-30km, 40°S-65°S zonal mean.	April
	Jan. 1991 – ~~Nov. 2016~~ April 2017 composite data series	AOD@532nm merged satellite time series of SAGE II, OSIRIS, CALIOP and OMPS following method described in Khaykin et al. (2017) 15-30km, 40°S-65°S zonal mean	
AAO	NOAA/National Weather Service ftp://ftp.cpc.ncep.noaa.gov/cwlinks/	Daily AAO index	Same as O3
GRAD		Daily maximum of PV slope at 550K computed from ERA-Interim data	Same as O3

Table 2: Coefficient of determination R^2, trends ± 2σ in DU yr^{-1} before and after the turnaround year 2001 derived from multi-regression model using as input ~~SAT (1980-2016~~ 2017~~) and~~ MSR-2 (1980-~~2015~~ 2017) total ozone anomalies inside the vortex for September using three classification methods described in Sect. 3.2. The residual is represented in DU by $\chi = \sqrt{\sum_i (obs_i - mod_i)^2/(n-m)}$ where obs_i and mod_i correspond to observations and model monthly mean, n the number of years and m the number of parameters fitted as in Weber et al. (2017)

	~~Composite satellite data (SAT)~~			~~Multi-Sensor Reanalysis (MSR-2)~~		
	~~400 K-600 K~~	~~475 K~~	~~550 K~~	~~400 K-600 K~~	~~475 K~~	~~550 K~~
~~R^2~~	~~0.93~~	~~0.89~~	~~0.91~~	~~0.92~~	~~0.90~~	~~0.91~~
~~Trend before 2001~~	~~-5.55±0.71~~	~~-4.99±0.83~~	~~-5.35±0.75~~	~~-5.36±0.71~~	~~-4.99±0.78~~	~~-5.27±0.75~~
~~Trend after 2001~~	~~1.87±1.18~~	~~2.57±1.38~~	~~1.85±1.24~~	~~1.97±1.25~~	~~2.67±1.38~~	~~2.02±1.32~~
~~χ~~	~~11.44~~	~~13.37~~	~~12.00~~	~~11.33~~	~~12.49~~	~~12.02~~
~~Without SF, QBO, AAO and Aerosols~~						
~~R^2~~	~~0.92~~	~~0.88~~	~~0.89~~	~~0.91~~	~~0.89~~	~~0.89~~
~~Trend before 2001~~	~~-5.57±0.67~~	~~-5.07±0.76~~	~~-5.34±0.74~~	~~-5.34±0.67~~	~~-5.05±0.72~~	~~-5.23±0.75~~
~~Trend after 2001~~	~~1.96±1.07~~	~~2.51±1.21~~	~~2.01±1.17~~	~~2.07±1.16~~	~~2.67±1.25~~	~~2.15±1.31~~
~~χ~~	~~11.10~~	~~12.58~~	~~12.21~~	~~11.06~~	~~11.90~~	~~12.46~~
~~Without GRAD~~						
~~R^2~~	~~0.88~~	~~0.83~~	~~0.86~~	~~0.87~~	~~0.83~~	~~0.86~~
~~Trend before 2001~~	~~-5.44±0.91~~	~~-4.87±1.02~~	~~-5.24±0.93~~	~~-5.23±0.89~~	~~-4.84±0.99~~	~~-5.15±0.91~~
~~Trend after 2001~~	~~2.27±1.48~~	~~2.99±1.67~~	~~2.24±1.52~~	~~2.16±1.57~~	~~2.88±1.74~~	~~2.21±1.61~~
~~χ~~	~~14.55~~	~~16.40~~	~~14.92~~	~~14.31~~	~~15.83~~	~~14.63~~

	Multi-Sensor Reanalysis (MSR-2)		
	400 K-600 K	475 K	550 K
R^2	0.92	0.90	0.92
Trend before 2001	-5.31±0.67	-4.90±0.74	-5.23±0.68
Trend after 2001	1.84±1.03	2.36±1.16	1.92±1.07
χ	10.74	12.02	11.12
Only with GRAD, HF and PWT			
R^2	0.91	0.89	0.89
Trend before 2001	-5.32±0.64	-5.00±0.71	-5.21±0.70
Trend after 2001	1.91±0.94	2.26±1.04	2.00±1.04
χ	10.61	11.82	11.71
Only with HF and PWT			
R^2	0.84	0.77	0.82
Trend before 2001	-5.34±0.84	-4.79±0.97	-5.23±0.89
Trend after 2001	2.04±1.24	2.83±1.48	2.13±1.31
χ	14.04	16.03	14.81

Table 3: Idem Table 2 for Sept15-Oct15 period. SAT dataset is also presented.

	Composite satellite data (SAT)			Multi-Sensor Reanalysis (MSR-2)		
	400K-600K	475K	550K	400K-600K	475K	550K
R^2	0.95	0.94	0.94	0.95	0.94	0.94
Trend before 2001	-5.86±0.60	-5.56±0.68	-5.60±0.66	-5.84±0.63	-5.61±0.70	-5.64±0.71
Trend after 2001	1.27±1.00	1.78±1.14	1.21±1.01	1.18±1.12	1.72±1.24	1.15±1.26
χ	9.68	10.95	10.59	10.07	11.11	11.31
Without SF, QBO, AAO and Aerosols						
R^2	0.94	0.93	0.93	0.94	0.93	0.92
Trend before 2001	-5.88±0.58	-5.67±0.66	-5.60±0.64	-5.84±0.60	-5.71±0.65	-5.63±0.69
Trend after 2001	1.06±0.93	1.37±1.05	1.05±1.03	1.08±1.04	1.42±1.13	1.07±1.19
χ	9.62	10.94	10.67	9.87	10.77	11.36
Without GRAD						
R^2	0.91	0.89	0.90	0.91	0.89	0.90
Trend before 2001	-5.73±0.80	-5.42±0.88	-5.48±0.83	-5.67±0.82	-5.43±0.90	-5.48±0.86
Trend after 2001	1.56±1.33	2.08±1.46	1.48±1.37	1.32±1.47	1.88±1.61	1.29±1.53
χ	12.96	14.21	13.34	13.21	14.45	13.78

	Multi-Sensor Reanalysis (MSR-2)			Composite satellite data (SAT)		
	400K-600K	475K	550K	400K-600K	475K	550K
R^2	0.95	0.94	0.94	0.96	0.94	0.94
Trend before 2001	-5.81±0.60	-5.55±0.66	-5.63±0.77	-5.86±0.57	-5.57±0.64	-5.64±0.65
Trend after 2001	1.42±0.92	1.73±1.01	1.58±1.02	1.70±0.87	1.96±0.99	1.79±0.99
χ	9.67	10.65	10.77	9.21	10.39	10.46
Only with GRAD, HF and PWT						
R^2	0.94	0.93	0.93	0.95	0.93	0.93
Trend before 2001	-5.86±0.56	-5.71±0.64	-5.67±0.64	-5.93±0.56	-5.75±0.66	-5.70±0.63
Trend after 2001	1.21±0.83	1.42±0.95	1.35±0.94	1.40±0.83	1.56±0.97	1.47±0.93
χ	9.35	10.68	10.65	9.35	10.84	10.55
Only with HF and PWT						
R^2	0.87	0.82	0.86	0.88	0.83	0.87
Trend before 2001	-5.89±0.84	-5.74±0.98	-5.70±0.86	-5.96±0.82	-5.78±1.00	-5.72±0.84
Trend after 2001	1.45±1.24	1.70±1.45	1.57±1.27	1.63±1.21	1.82±1.47	1.68±1.24
χ	14.06	16.40	14.39	13.71	16.18	14.03

Table 4: Coefficient of determination R^2, trends ± 2σ in Mt yr^{-1} before and after the turnaround year 2001 derived from multi-regression model using OMD dataset (MSR-2 total ozone columns and threshold of 220 DU, see the text) for September and 15Sept-15Oct over 1980-2017 period. The residual is represented in DU by χ as explained in Tab. 2.

	September	15Sept-15Oct
R^2	0.85	0.91
Trend before 2001	1.28±0.25	1.59±0.24
Trend after 2001	-0.78±0.39	-0.68±0.37
χ	4.04	3.85
	Only GRAD, HF and PWT	
R^2	0.82	0.90
Trend before 2001	1.29±0.24	1.61±0.22
Trend after 2001	-0.86±0.36	-0.65±0.33
χ	4.04	3.68
	Only HF and PWT	
R^2	0.78	0.85
Trend before 2001	1.29±0.26	1.61±0.27
Trend after 2001	-0.88±0.38	-0.70±0.39
χ	4.37	4.44

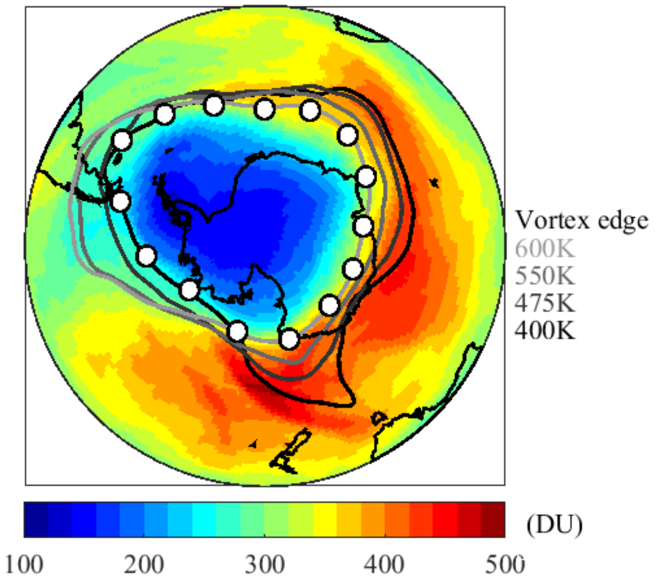


Figure 1: Total ozone (DU) from MSR-2 on October 7, 2012 at 12 UT. Vortex edge position at different isentropic levels are auditioned to the map and represented by black to light grey lines. White dot marks identify the region considered inside the vortex using the 400 K -600 K range classification.

Supprimé:

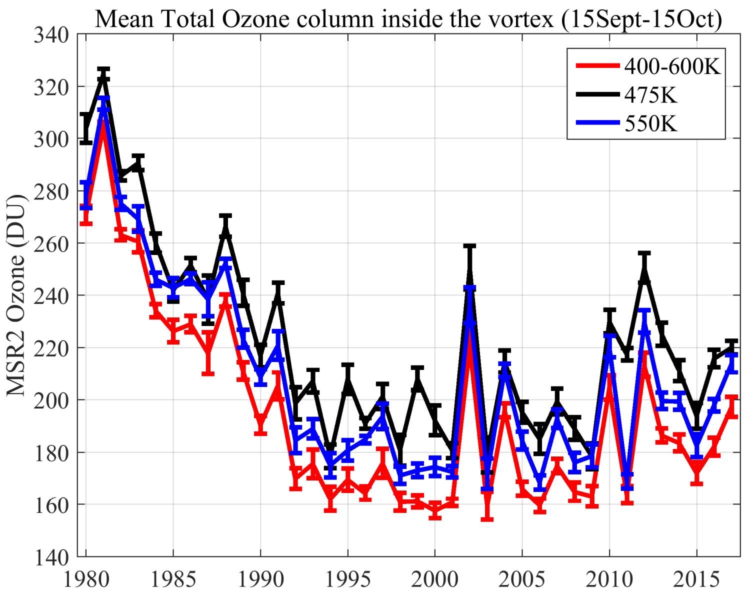
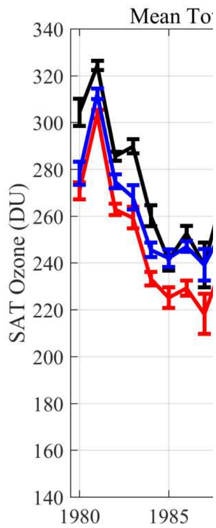


Figure 2: Evolution of total ozone of ~~satellite composite (SAT) measurements~~ MSR-2 dataset inside the vortex averaged each year on 15Sept–15Oct period for different classifications: standard method at 475 K and 550 K represented by black and blue lines, respectively and method considering the 400 K-600 K altitude range by red line. Error bars represent twice the standard error.

Supprimé:



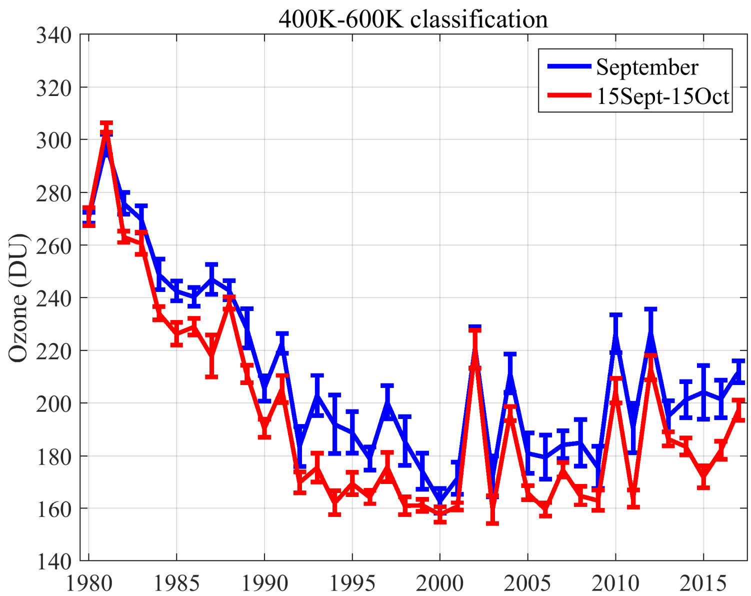
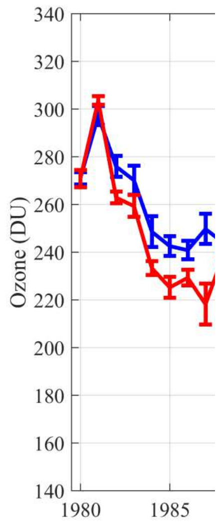


Figure 3: As in Fig. 2 but only for 400K-600K classification on different periods: September and mid-September to mid-October. Error bars represent 2σ.

Supprimé:



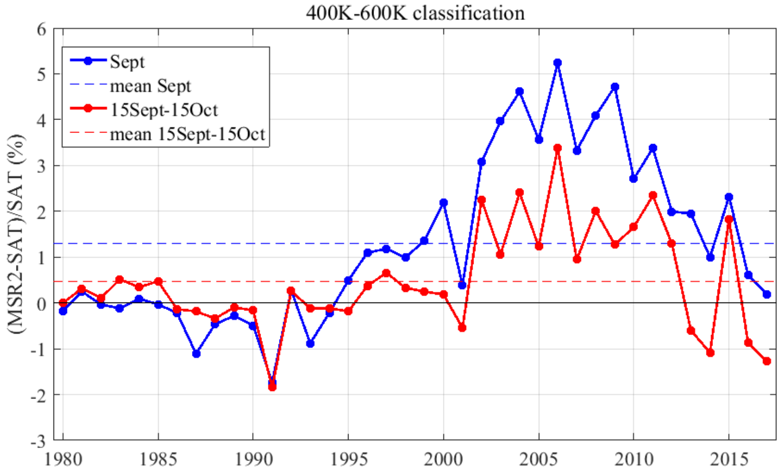
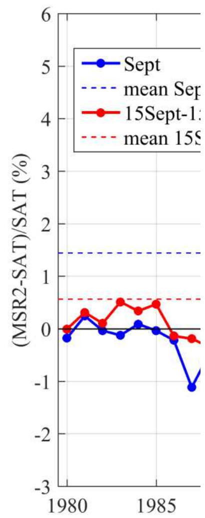


Figure 4: Relative difference between MSR2 and SAT mean total ozone inside the vortex for September (blue curve) and 15Sept-15Oct (red curve) periods. Horizontal dash lines correspond to the mean bias between data series.

Supprimé:



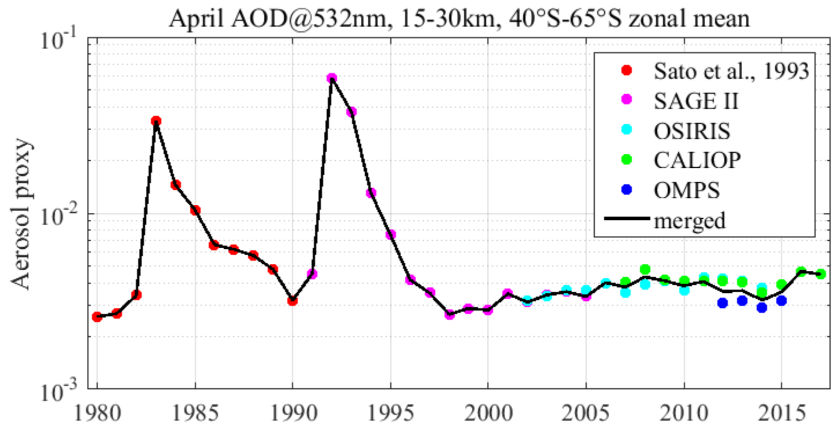
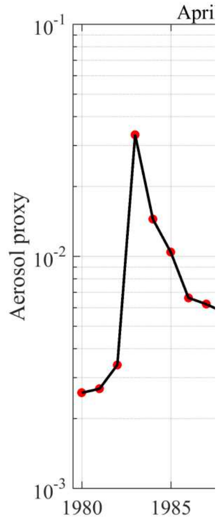


Figure 5: Time series of April monthly mean AOD at 532nm within 40°S-65°S and 15-30km of normalised Sato et al., 1993 dataset (see main text) and from satellites (SAGE II, OSIRIS, CALIOP, OMPS). The corresponding merged data is represented by the bold line.



Supprimé:

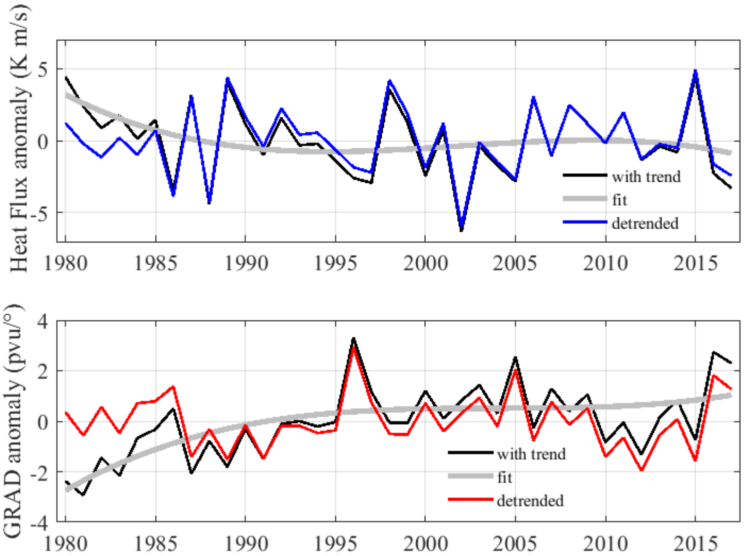
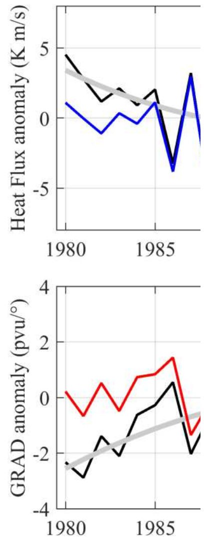


Figure 6: Heat Flux (top panel) and Gradient - GRAD (bottom panel) anomalies for the 15Sept-15Oct period: before removing a polynomial fit of 3rd order (black line), fit (grey line) and after removing the fit (blue/red line).

Supprimé:



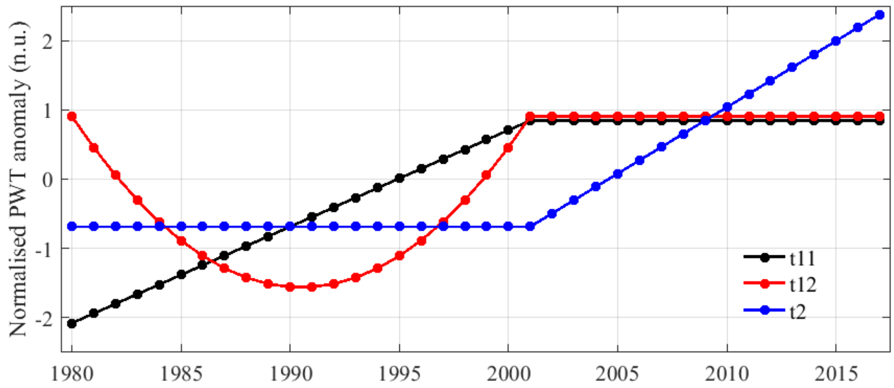
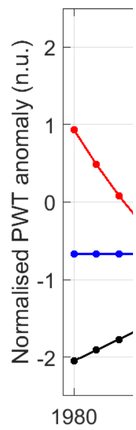
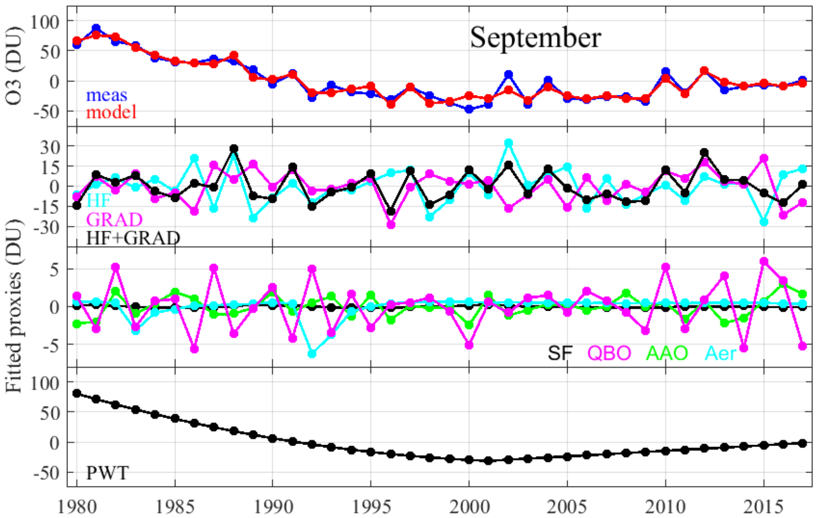


Figure 7: Anomalies of the linear functions before and after 2001 (t11 and t2, respectively) and parabolic function (t12) that correspond to the PWT proxy (see Eq. 2 to 5). Each proxy anomaly is normalised by the corresponding standard deviation.



Supprimé:



Supprimé:

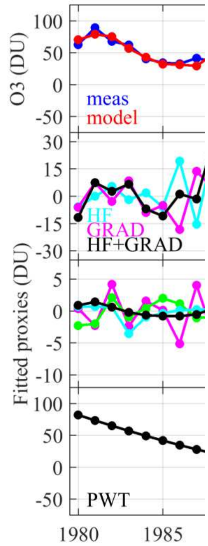
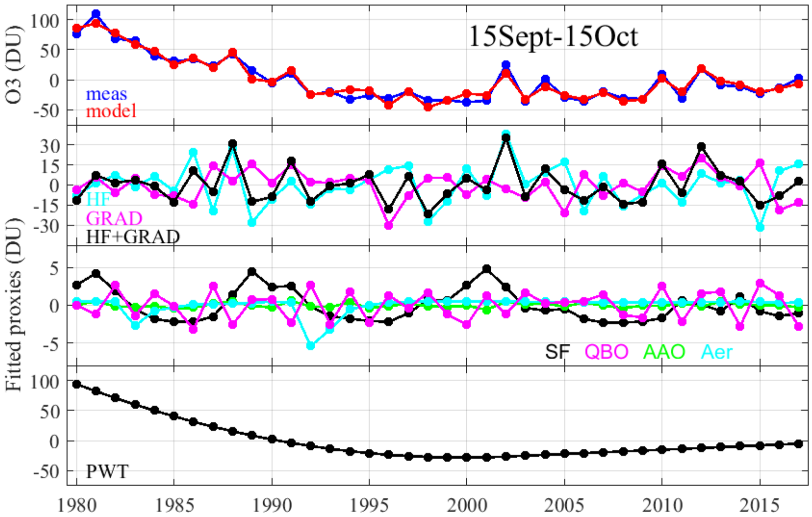


Figure 8: Deseasonalised total ozone inside the vortex of ~~SAT~~ MSR-2 series (meas) and regression model (model) for September using 400 K-600 K classification (top panel). Contributions of proxies are also shown: Heat Flux - HF, gradient - GRAD and the combination of both HF+GRAD (second panel); solar flux - SF, QBO (QBO at 30hPa + QBO at 50hPa), Antarctic Oscillation - AAO and Aerosol - Aer (third panel); and ~~PWLT~~ PWT (bottom panel). Ozone anomalies and contributions of proxies are given in DU.



Supprimé:

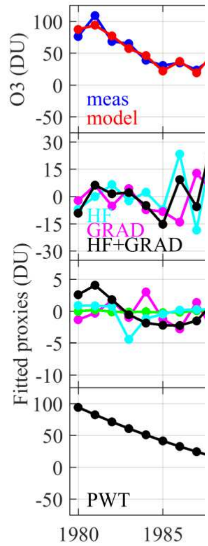


Figure 9: As in Fig. 7 for 15Sept-15Oct.

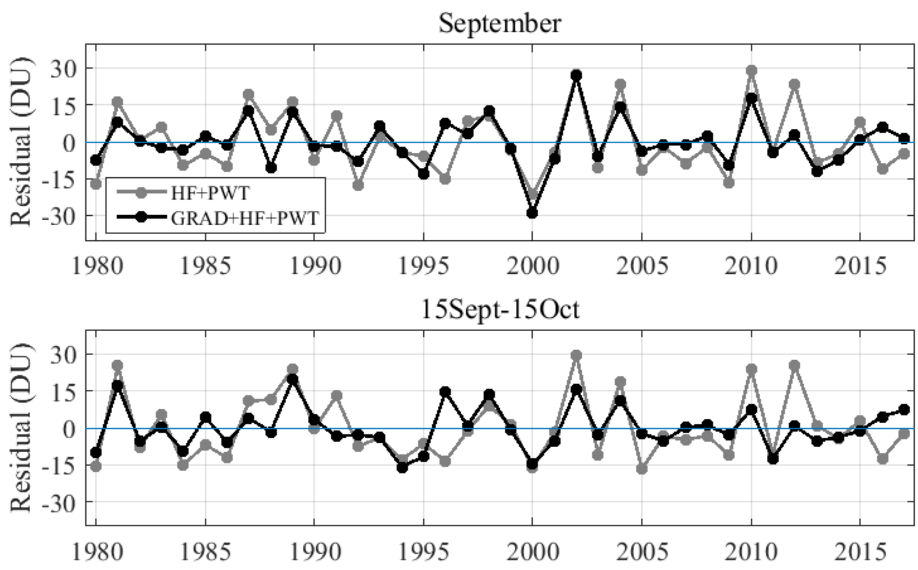
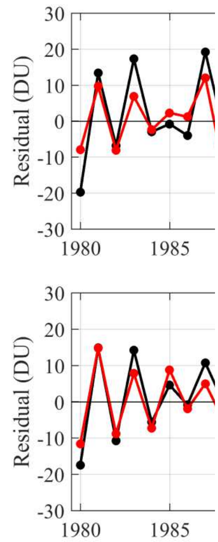


Figure 10: Residual (in DU) with and without contribution of GRAD proxy for September (top panel) and 15Sept-15Oct period (bottom panel)

Supprimé:



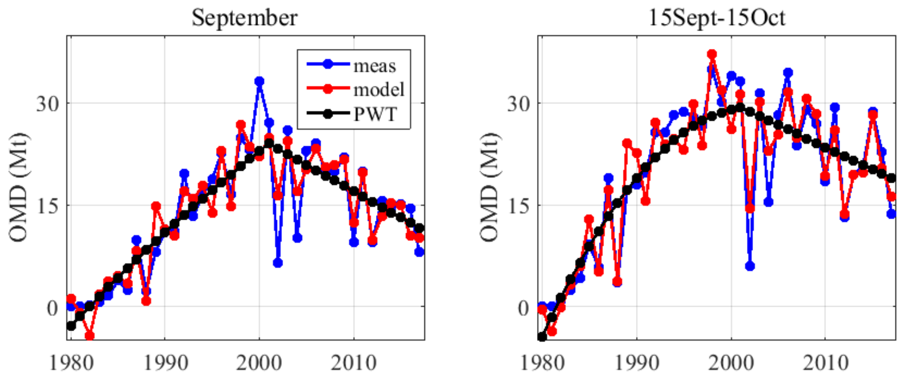


Figure 11: OMD (in Mt) computed from total columns of MSR-2 dataset lower than 220 DU and south of 60°S for September (left panel) and 15Sept-15Oct (right panel). Regressed values by MLR analysis using GRAD, HF and PWT are also shown as well as the fitted PWT proxy.

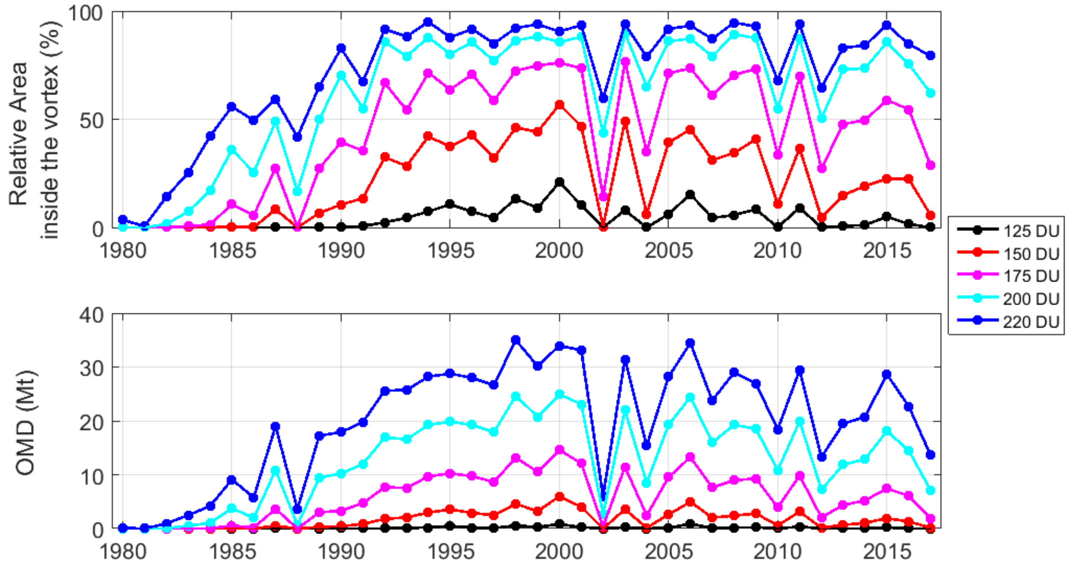
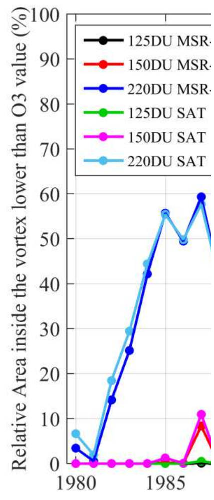


Figure ~~11~~ 12: Relative area inside the vortex (in %) with values lower than ~~3~~ 5 level thresholds (125, 150, 175, 200 and 220 DU) computed from MSR-2 ~~and TOMS/OMIT~~ datasets using 400 K-600 K classification on 15Sept-15Oct period (top panel). OMD (in Mt) time series computed from MSR-2 total ozone data for the same 5 thresholds and time period are displayed in the bottom panel.



Supprimé:

Figure 12: Relative area inside the vortex (in %) lower than 220 DU computed from TOMS/OMIT dataset using 400K-600K classification on September, 15Sept-15Oct and October.

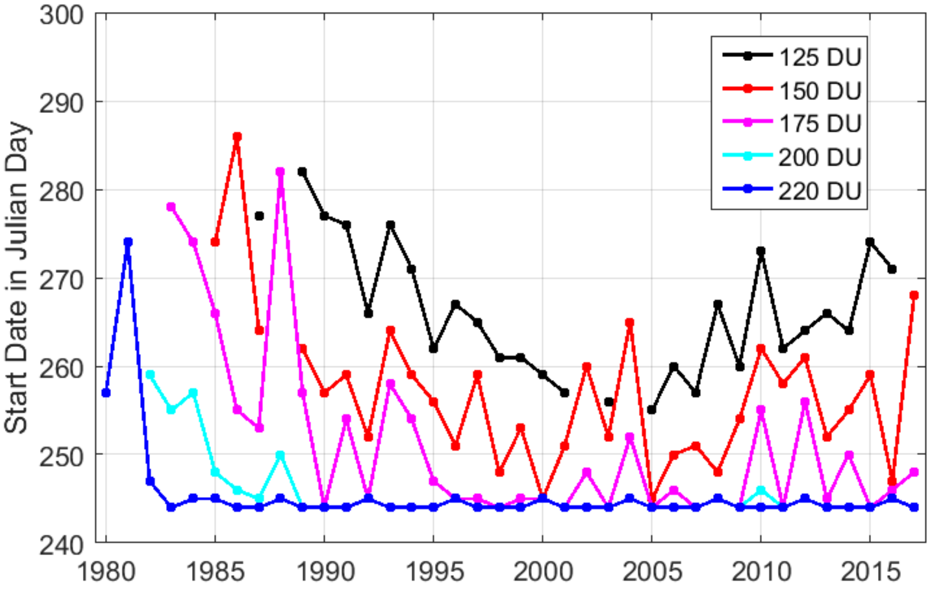
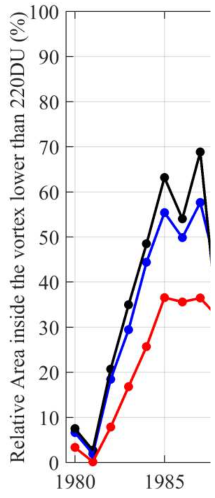


Figure 13: Start day of occurrences of total ozone levels lower than different thresholds (125, 150, 175, 200 and 220 DU) computed from the MSR-2 dataset using the 400 K-600 K classification between September 1st and October 15th.



Supprimé: