Response

We thank the referees for their additional comments, which are addressed below.

Furthermore, after taking a close look at the paper with the modification made, we made some additional changes to the discussion (section 4) for the part between lines 684 and 711. Parts of this section were moved to either section 3.1 or the discussion section, and the remainder of the part between lines 684 and 711 has been adjusted for readability and consistency.

Referee #1

We have included/modified the suggested technical corrections.

Referee #2

- We modified the discussion and conclusions so that it should be clear now where the "30-60%" originates from (table 7) and how to interpret those numbers.

- We rephrased the last sentence of the Abstract, explaining what we mean when we talk about "choices":

  "… uncertainties in parameters and independent variables and choices in defining the optimal time period and area for calculating the ozone record and the independent variables …"

  Note that it is important to realize that both uncertainties in parameters and choices in time period and area over which to average hamper detection of recovery.

- We modified the discussion, giving less prominence to the notion that for some certain regressors extending the period does not always lead to higher significance. This discussion was originally added to show that all uncertainties discussed in the paper leave (quite) some room for interpretation and that not all results conform to prior expectations. It is therefore tempting to focus only on favorable results confirming prior expectations while neglecting unfavorable results, so we wanted to make sure that the unfavorable results (although minor) are mentioned.

- We added a sentence noting that Solomon et al. [2005] show that there appear to be some volcanic effects in height resolved ozone during the ozone hole season, but that these effects are too small in magnitude and vertical extend to leave a detectable impact in total ozone. When looking superficially there still appears to be some change in total ozone related to volcanic activity, but Poberaj [2011] show that these are more likely related to (coincidental) dynamical effects than to volcanic aerosols. Our study as well as those of Knibbe et al. [2014] confirm the lack of volcanic signal in total ozone. We made sure that in the discussion our results only apply to total ozone, not height resolved ozone.

- Adding the finding that we do find positive post-break trends in ozone (1.66 to 4.74 DU/year; 95% CI) is an important finding. It is one of the requirements for detection of ozone recovery, and despite uncertainties with regard to trend significance ALL trends in ozone are positive. There was no a priori reason to assume that had to be the case, and it is consistent with various other studies noting that although trends may not necessarily be statistically significant, they tend to be positive. No changes made.

- See also second comment. We added the explanation in the Abstract of what hampers detection (uncertainties in data AND choices for area and time period over which to average). That way the Abstract and Conclusions are consistent.

- Line 767-769: deleted the sentence. It is now implicitly included in the previous section.

- Added a sentence to both the discussion and conclusions section about the relation between statistical significance and length of the record noting that our results suggest that with continued extension of the total ozone record and using multi-variate regressions detection of Antarctic ozone hole recovery may be reached before 2020.

This is what our findings imply (see table 7). Currently (2010 or 2012) we are around 50% statistical significance. Our results also indicate that for each year added the significance increases by approximately 10 % (order of magnitude). Although just a ball-park estimate, it suggests that by 2020 you may expect to have passed the 95% confidence level.

- Caption changed. Table 5 shows the results for the time series ending in 2010 for all break years (97, 98, 99). Also adjusted the description of table 5 in the discussion section.

Referee #3

General comment & discussion on trend error calculation

Thanks to the extensive response of referee #3 we now finally understand where the differences stem from with regard to the trend error calculations in Kuttippurath et al. [2013] and ours. In essence there are different aspects relevant for the trend errors.

(A) The EESC regression parameter (C5 in referee #3 comment) has an associated error
(B) Linear regression of EESC multiplied with regression parameter (C5) pre-break and post-break have their own associated errors
(C) Residual of the multi-variate regression

In Kuttippurath et al. [2013] equations (2) and (3) of the #R3 report are used for calculating pre-break and post-break ozone trends. Equations (4) and (5) of the #R3 report are used to calculate the pre-break and post-break ozone trend errors.

In particular, the ozone trend error is calculated as:

$$\frac{dO_3}{dt}_{\substack{error \\ time\ period}} = EESC_{error} \times \frac{dEESC}{dt}_{time\ period}$$

For trend error calculations the pre-break and post-break linear trends in the EESC must be calculated. These linear trends have their own errors because of the fact that the shape of the pre-break and post-break EESC curve are not completely linear.

In our paper we calculate trend errors based solely on this last step, which results in fairly small trend errors.

If we calculate the trend error based on the method in Kuttippurath et al. [2013] we find order of magnitude similar values similar to those of kuttippurath et al. [2013], and indeed the post-break trend error is proportional to pre-break trend error via the ratio of the pre-break and post-break trends in EESC. See table values in bold.

Both methods describe two different error sources and should be taken into account together. For the case discussed here the error associated with the EESC regression dominates the error from the linear regression on both pre-break and post-break trends. See table values in bold + italics

	Kuttippurath et al. [2013]		This study	
Period	EESC	PWLT	EESC	PWLT
1979-1999	-4.50 ± 0.65	-5.02 ± 1.11	-5.39 ± 0.22 **-5.39 ± 0.97** ***-5.39 ± 0.99***	-5.66 ± 1.03
2000-2010	1.11 ± 0.16	2.91 ± 2.73	1.04 ± 0.12 **1.04 ± 0.19** ***1.04 ± 0.22***	3.30 ± 2.85

However, after further consideration of the method applied in Kuttippurath et al. [2013] we argue that the method applied there is mathematically not justified and leads to conceptual problems. Underlying reason is that the error associated with the EESC regression is only valid for the entire EESC time series, not parts of it, as outlined below.

Consider the following hypothetical situation with the following EESC shape: the pre-break increase is as the current EESC but the post-break changes is flat (zero). According to equation (5) of the #R3 report and the equation here above, the associated ozone trend error would then be zero regardless of the regression parameter (C5).

This is counter-intuitive: in case of similar noise levels, one would expect that the smaller the trend, the larger the trend error. Or at least a non-zero trend error.

The EESC regression error applies to the entire record and should be considered as such. Because of the particular shape of the EESC (increase-decrease) the regression error will be largely determined by corresponding long term changes in ozone (decrease-increase), not the year-to-year changes in ozone. However, for the two separate periods one would expect the trend errors to also be dependent on the year-to-year variations in ozone (the sign change in EESC which dominates the fit with the total ozone record does not matter anymore in case of the two separate periods). Hence, the post-regression separation in two periods should be considered very carefully.

This is highlighted by the very different trend errors of the PWLT fit (see table above). Whereas the EESC-associated error is considerably smaller for the post-BREAK period compared to the pre-BREAK period, the PWLT trend errors show exactly the opposite: the pre-BREAK trend error is much smaller than the post-break trend error. The latter is more intuitive: the pre-BREAK period is longer and the trend is larger, which should results in smaller trend errors.

In summary, we argue that we think the calculation of trend uncertainty in Kuttippurath et al. [2013] is not justified, and that as far as we are concerned it is unclear how to actually do that (other than *a posteriori* applying a sort of piece-wise linear trend calculation, which then should be included in the regression to start with rather than be applied *a posteriori*). Trend errors as derived from the method applied in Kuttippurath et al. [2013] are counter intuitive: one would expect larger trend errors for the post-BREAK period, not smaller. The PWLT errors do are consistent with this notion. If the goal is to determine pre-break and post-break trends and errors, then the fit parameter should consist of two separate parts, as is done when including the PWLT in the regression.

We have modified sections 3.1 and the discussion in section 4, as well as table 3 according to the discussion above. Note that we have not referred to this in the in the abstract or the conclusions, as that would put too much emphasis on it and there were already sufficient arguments to argue that use of EESC in the regression is not preferred.

Specific comments

Line 28: included

Line 34-35, changed to “the onset of recovery of the Antarctic Ozone Hole”

Line 70-71, changed to “Both studies do not consider the effect of deterministic variations in ozone on estimating when the onset and complete recovery of the Antarctic ozone hole recovery occur.”

Line 202-203. This is a tricky point. Salby et al. [2012] indeed only identify 2002 as an outlier. However, Weber et al [2003, 2011] and updates in the upcoming WMO 2014 ozone assessment report indicate that in both hemispheres there are outlier years that ‘skew’ the statistics.

We suggest to modify the text and make it more specific and use somewhat less strong wording.

“The Arctic and Antarctic may behave very similar [Weber et al., 2003; 2011] or much less similar [Salby et al., 2012]. This is because the notion of hemispheric similarities in how the EP flux affects ozone depletion so far may be biased by a few outlier years (2002 and 2006 for the SH, 1996, 2010 and 2011 for the NH).”

Line 515-517. Section between lines 511-517 was completely changed.

“Note that there is slight difference in the 1979-1999 trends for the period ending in 2010 and 2012 because of the difference in total record length, which results in slightly different regression coefficients. However, calculating trend errors for the EESC-based pre-BREAK and post-BREAK trends in ozone using the EESC regression error as done in Kuttippurath et al. (2013) is not justified. The trend errors depend on the actual trend values themselves (see table 3): the EESC-fit based post-break trend error is much smaller than the pre-BREAK trend. In the hypothetical case of no (zero) trend the trend error would also be zero, which would be physically unrealistic. The PWLT on the other hand shows opposite differences in trend errors: the post-BREAK trend error is much larger than the pre-BREAK trend error, conform expectations.”

Line 560-562. Correct, should be table 3. We did an additional check on all table and references to them in the paper as there had been a change in table numbering between versions.

Line 563-565. See discussion above about trend errors.

Line 626. Figure 3 should be figure 4 (middle panel). The lower panel of figure 6 also shows a tri-modal distribution. Text adjusted accordingly.

Line 638. Text adjusted accordingly.

Line 652. Text adjusted accordingly.

Line 653-654. See discussion above about different approaches on trend error calculations. Answer: the EESC fit uncertainty will be largely determined by the long ter changes in ozone (decrease-increase), not the year-to-year variations. However, when separating the decreasing and increasing part of the EESC, the year-to-year variations should be considered as the sign change in long term ozone changes do not matter anymore.

Line 707-711. Reading the comment, we think that the referee actually refers to the previous section (696-706). This section was replaced by the next statement, which is less pessimistic and does more justice to our analysis. It shows that there is a balance: a gain in confidence by removing some deterministic variations in ozone, but with some adverse effects as the multi-variate regression introduces new uncertainties. The section then ends on a positive note (lines 707-711):

"Furthermore, this implies that with continued extension of the total ozone record detection of Antarctic ozone recovery may be reached before 2020 using multi-variate regressions. Note that although in total the number if statistically significant trends increases with record length, this is not necessarily always the case (compare Table 5 and Table 6 – thus BREAK-2010 vs. 1998-2012 trends).

On the other hand, the trend significance level if significant trends is generally between 2σ and 3σ (not shown), indicating that a considerable amount of variability is not accounted for in the regression. Our analysis also hows that detection of the 2nd stage of ozone recovery based on just one arbitrary selected (set of) regressor – ozone record combination(s) does not reflect the structural uncertainties present in the underlying data.

Nevertheless, the appearance of larger groups of statistically significant results occurring for longer time series and a certain persistence among ozone scenarios and EP flux scenarios shows that these type of analyses are capable of removing deterministic variations in average ozone, and that with increasing length of the post-break period more statistically significant results can be expected."

Line 724-727: see discussion above. Now that we full understand what was done in Kuttippurath et al. (2013) we argue that that method does not take the fit residuals into account.

Line 735. The larger range of trends in the PWLT compared to the EESC fits is indicative of the limited use of the EESC – its post-peak trend is largely determine by the pre-defined EESC shape and thus does not do justice to all uncertainties. This is what we refer to as a lack of flexibility when using the EESC to determine post-BREAK ozone trends.

Line 774. Suggestion to change to "are expected", rather than "may" or "will", because we expect that future updates of this analysis provide better clues about the onset of recovery.

Table 5. Table caption modified (ending 2010).

Line 977. Typical maximum wind speed numbers added (20-30 m/s)

Line 982. Typo corrected.

Line 1000-1026.

Figures 6-7. Suggestion included.

Tracing the second stage of Antarctic ozone hole recovery with a "big data" approach to ~~multi-~~multivariate regressions

A.T.J. de Laat, R.J. van der A, M. van Weele

Royal Netherlands Meteorological Institute, De Bilt, The Netherlands.

Abstract

This study presents a sensitivity analysis of ~~multi-~~multivariate regressions of recent springtime Antarctic vortex ozone trends using a “big data” ensemble approach.

Our results indicate that the poleward heat flux (Eliassen-Palm Flux) and the effective chlorine loading explain, respectively, most of the short-term and long-term variability in different Antarctic springtime total ozone records. The inclusion in the regression of stratospheric volcanic aerosols, solar variability and the Quasi-Biennial Oscillation is shown to increase rather than to decrease the overall uncertainty in the attribution of Antarctic springtime ozone because of large uncertainties in their respective records.

Calculating the trend significance for the ozone record from the late 1990s onwards solely based on the fit of the effective chlorine loading is not recommended, as this does not take fit residuals into account resulting in too narrow uncertainty intervals, while the fixed temporal change of the effective chlorine loading does not allow for any flexibility in the trends.

When taking fit residuals into account into a piecewise linear trend fit, we find that approximately 30-60% of the regressions in the full ensemble result in a statistically significant positive springtime ozone trend over Antarctica from the late 1990s ~~onwards~~to 2010 or 2012. Analysis of choices and uncertainties in time series show that, depending on choices in time series and parameters, the fraction of statistically significant trends in parts of the ensemble can range from negligible to a complete 100%. We also find that, consistent with expectations, the number of statistically significant trends increases with increasing record length.

Although our results indicate that the use ~~multi-~~multivariate regressions is a valid approach for assessing the state of Antarctic ozone hole recovery, and it can be expected with increasing record length results will move towards more confidence in recovery, uncertainties in parameters and independent variables and choices in ~~uncertainties in choices~~ defining the optimal time period and area for calculating the ozone record and the independent variables currently do not yet support formal identification of the onset of recovery of the Antarctic Ozone Hole.

1. Introduction

An important question in 21st century ozone research is whether the ozone layer is starting to recover as a result of the measures taken to reduce emissions of Ozone Depleting Substances (ODS) as agreed on in the Montreal Protocol [UNEP, 2012] and its subsequent amendments and adjustments.

The World Meteorological Organization has defined three different stages of ozone recovery [WMO, 2007]. The first stage consists of a slowing of ozone depletion, identified as the occurrence of a statistically significant reduction in the rate of decline in ozone due to changing stratospheric halogens. The second stage revolves around the onset of ozone increase (turnaround), identified as *the occurrence of statistically significant increases in ozone - above a previous minimum value - that can be attributed to declining stratospheric halogens*. Note that what is meant by “statistically significant” is not specified. Finally, the third stage is the full recovery of ozone from ODSs, identified as when the ozone layer is no longer affected by ODSs, or alternatively, once stratospheric ozone levels have returned to pre-1980 values.

The first stage of ozone recovery has already been identified in observations to have occurred roughly in the late 1990s [WMO 2007, 2011]. The third stage is not expected to occur until somewhere halfway the 21st century or later [WMO, 2011]. The spatial distribution of total ozone after the third stage probably differs somewhat from the pre-1980 distribution due to climate change – in particular changes in the stratospheric chemical composition and temperature structure [Bekki et al., 2011, and references therein].

As far as the second stage of ozone recovery is concerned, it has recently been argued that a statistically significant increase in ozone beyond a minimum and attributable to decreases in ODSs can be identified for the Antarctic ozone hole [Salby et al., 2011, 2012; Kuttippurath et al., 2013; Knibbe et al, 2014]. To some extent this is surprising as it has long been thought that identification of the second stage of ozone recovery could only be expected after 2020 [*e.g.* Newman et al., 2006; Eyring et al., 2007]. Those estimates were based on (model) simulations of ozone from which it is calculated when the ozone trend from a certain starting year onwards would qualify for "statistically significant", or in other words, would emerge from the year-to-year natural variations in ozone ("noise"). Such methods implicitly assume Both studies do not consider the effect of deterministic variations in ozone on estimating when the onset and complete recovery of the Antarctic ozone hole recovery occur. that ozone variations around the trend are not deterministic (random).

However, it has also long been established that many stratospheric ozone variations are in fact deterministic. Various processes have been identified that affect stratospheric ozone variability in the Southern Hemisphere on an inter-annualinterannual basis, like volcanic aerosols [Telford et al., 2009], the Southern Annular Mode (SAM) [Thompson and Wallace, 2000; Jiang et al., 2008], the poleward heat flux or Eliassen-Palm flux (EP flux) [Randel et al., 2002], solar variability [Soukharev and Hood, 2006], and the Quasi Biennial Oscillation (QBO) [Jiang et al., 2008]. If the physics and chemistry are sufficiently understood, it might be possible to filter out part of the ozone variations from the ozone records by means of a multi-variaetey regression, resulting in a smoother ozone record for which trend significance might be reached earlier. This approach, in essence, forms the basis of the suggested identification of the second stage of ozone recovery reported by Salby et al. [2011, 2012], Kuttippurath et al. [2013] and Knibbe et al [2014].

However, none of these studies did systematically consider the uncertainties in the proxies that were selected for the regressions. In addition, no motivation or discussion was provided for the choice of a specific ozone record, *e.g.* a consideration of taking annual, seasonal, and/or monthly means of total ozone, and the integration over a chosen spatial domain.

Hence, we want to address the following question in this study: Is the suggested detection of the second stage of ozone recovery robust when uncertainties in the regression parameters and for different selected ozone records are taken into account? This question is approached here with combined multiple scenario – Monte Carlo ensemble simulations using the same regression methodology as presented in Kuttippurath et al. [2013] but by inclusion of various uncertainties leading to a large ensemble of different regressions. We analyze this “big data” ensemble for robustness of the individual regressions.

Kuttippurath et al. [2013] considered different Antarctic vortex definitions and thus different vortex ozone records. They found that regression results were not very sensitive to the Antarctic vortex definition. Hence, we decided to use September-November Antarctic vortex core (poleward of 70ºS) average total ozone column based on the Multi Sensor Reanalysis (MSR; van der A et al. [2010]), also because from a practical point of view this definition does not require additional information about the location of the vortex edge. The selected regressors are the SAM, solar flux, QBO, EP flux, stratospheric volcanic aerosols and the Equivalent Effective Stratospheric Chlorine (EESC), similar to

Kuttippurath et al. [2013]. The EESC can be used to estimate ozone trends. Kuttippurath et al. [2013] also calculated Piece Wise Linear Trends (PWLT) for estimating ozone trends as alternative for the EESC-based ozone trends, an approach we will follow here as well.

In this paper, we extend the analysis by introducing both several differing scenarios for the ozone record and regressor records of the EP flux, volcanic aerosols, and EESC. Monte Carlo variations were applied to the regressor records of the solar flux, QBO, SAM by adding random variations. While we focus on parameter uncertainties in this study, additional uncertainties do exist, for example with respect to possible time lags between regressors and the ozone record. The resulting ensemble of regression results provides a big data pool of about 23 million different regressions that is analyzed in terms of probability distributions of the explanatory power of the regressions (R^2), the ozone trends and corresponding ozone trend uncertainties, and the regression coefficient values quantifying the dependence of ozone on a particular regressor. We also investigate if some way of optimization is possible for the chosen scenarios, and we discuss the likelihood of detection of the second stage of ozone recovery within the context of all uncertainties presented. Note that the uncertainties discussed here differ from formal errors that come with a standard ~~multi-~~multivariate regression. Also note that we implicitly assume that the relation between the independent variables and ozone is linear, even though the relation may very well be non-linear. The latter will to some extent be considered in our study and is part of the discussion of the results, but the issue of non-linearity of the regressor-ozone relation is not addressed in detail, in particular because, as will be shown, for many regressors the non-linearity of its relation with ozone is insufficiently characterized, or even unknown.

This paper is organized as follows. Section 2 describes the observational datasets used and the ozone and regressor scenarios or Monte Carlo simulations performed. Section 3 discusses the probability distributions of the explanatory power of the regressions, trends and regression values, including how the distributions depend on scenarios or Monte Carlo results. Section 4 discusses the question of detection of the second stage of ozone recovery, and in section 5 everything is wrapped up and some conclusions are drawn.

2. Multivariate regression parameter uncertainties

Online data sources of the ozone observation records and applied regressors can be found in Table 1.

2.1 Method

A common method for analyzing total ozone records is the use of a ~~multi-~~multivariate linear regression, a method that we will use in this paper as well. The goal of the method is to attribute both ~~inter-annual~~interannual as well as decadal variations in the ozone record to processes that are expected or known to affect the total ozone record (Kuttippurath et al. [2013], and references therein). In the regression, the total ozone variability (Y) as a function of time (t) is expressed as

$$
\begin{aligned}
Y(t) &= K && \text{(Constant)} \\
&+ C_1 HF(t) && \text{(Poleward Heat Flux or Eliassen-Palm (EP) flux)} \\
&+ C_2 SAM(t) && \text{(Southern Annual Mode index)} \\
&+ C_3 (SF \times QBO)(t) && \text{(Solar Flux × QBO index)} \\
&+ C_4 Aer(t) && \text{(Stratospheric Aerosol optical depth)} \\
&+ C_5 Trend(t) && \text{(Total ozone trend)} \\
&+ \varepsilon(t) && \text{(Total ozone residual)}
\end{aligned}
$$

In which t is the time from 1979 to 2010 or 2012, K is a constant and regression coefficients C_1 to C_5 are the regression coefficients for the respective proxies. The ozone trend (C_5) can be related to the time-dependent equivalent effective stratospheric chlorine loading (EESC) or a piecewise linear trend (PWLT) before and after a predefined break year. The PWLT regressions are calculated by including two linear terms in the regression: the first term is a linear fit for the entire time window, the second term is a linear term only for the years after a chosen break year [Kuttippurath et al., 2013].

The analysis of regression results will focus on two parameters that have previously been used in papers investigating Antarctic ozone recovery [Yang et al., 2008; Salby et al., 2011, 2012; Kuttippurath et al., 2013; Knibbe et al., 2014]: the serial correlation R between the regression-based ‘reconstructed’ ozone record and the observations, and the post-break trends and trend significance. Since the focus of our paper is to investigate trend significance, not specifically what parameters can best explain Antarctic ozone, we will only look in some detail at the usefulness of certain regressors. However, our analysis does provide indications of what are more and less useful regressors.

In sections 2.2 to 2.7 the uncertainty in each of the proxies that is used as a regressor is discussed. These uncertainty ranges determine the spread in the ensemble that is used in the “big data” analysis. A summary of the regressor uncertainties and how they are incorporated in this study can be found in Table 2. The solar flux and QBO are combined into one proxy as discussed in section 2.3.

2.2 Poleward heat flux (EP flux)

Figure 1 shows the poleward heat flux, here represented by the (vertical) EP flux [Andrews et al., 1987] at the 70-hPa level and averaged poleward of 40ºS for the combined months of August and September, as well as the average EP flux available for a given year for a variety of datasets. Note that the datasets do not all completely overlap in time. Before 2000 there are considerable differences between the datasets. After 2000 these differences are smaller, which to some extent is traced to the lack of ERA40 data beyond 2001 and lack of JRA data beyond 2004. The lower panel shows the relative differences between the five datasets and their mean. The standard deviation of all data is 7.65%, but from 2000 onwards only 2.67%.

Another source of uncertainty in the use of the EP flux as proxy is the choice of the time window over which the average EP flux is calculated. This choice depends on what is thought to be the relationship between variations in EP flux and ozone depletion. The basic theory states that the poleward movement of stratospheric air is proportional to the strength of the residual mean stratospheric circulation (Brewer-Dobson circulation), which in turn is driven by the poleward eddy heat flux. The poleward eddy heat flux is expressed by the upward component of the Eliassen-Palm flux that measures the upward transport of momentum by planetary waves [Andrews et al., 1987; Salby et al., 2012, and references therein]. Planetary wave activity in the Northern Hemisphere affects Arctic Polar vortex stability and thus Arctic ozone depletion. However, to what extent this is similar in the Southern Hemisphere is still a topic of debate. The Arctic and Antarctic may behave ~~either~~ very similar~~ly~~ [Weber et al., 2003; 2011] or ~~not~~ much less similar [Salby et al., 2012]. This is because the notion of hemispheric similarities in how the EP flux affects ozone depletion so far ~~is~~ may be ~~heavily~~ biased ~~on~~ by ~~only one~~a few outlier years (2002 and 2006 for the SH, -1996, 2010 and 2011 for the NH).

Current research efforts try to gain a better understanding of the physical and photochemical mechanisms by which the heat flux and planetary wave action affects Antarctic stratospheric ozone. A recently proposed mechanism [de Laat and van Weele, 2011] involves a pre-conditioning of Antarctic inner stratospheric vortex air whereby stratospheric temperatures affect PSC formation which in turn affects the buildup of a halogen reservoir that later during Austral spring changes the rate of catalytic ozone destruction. This preconditioning mechanism explains some years with anomalous ozone depletion, but not all. For example, during Austral winter 2013 the Antarctic vortex remained largely undisturbed – opposite to 2010 and 2012, see de Laat and van Weele [2011] and Klekociuck et al. [2011], thus allowing for widespread PSC formation and pre-conditioning the inner vortex air for efficient ozone depletion. However, from the start of Austral spring 2013 (halfway August) onwards the Antarctic stratospheric vortex got disturbed by planetary wave activity. As a result, the amount of springtime ozone depletion remained below what has been experienced during previous years with similar preconditioning. This suggests that there are multiple pathways as well as complicated interactions between chemistry and physics that can lead to reduced Antarctic springtime ozone depletion. Hence, it is unclear which regressor or regressors could act as proxies for these complex processes.

A further complicating factor is the disintegration of the Antarctic vortex, which is again controlled by planetary wave activity [Kramarova et al., 2014]. The stability of the vortex determines how long the ozone depleted inner-vortex air remains intact after photochemical ozone depletion ceases during Austral spring. Variations in the duration of Antarctic vortex stability introduce variations in the Antarctic total ozone record which are not related to variations in photochemistry.

We attempt to reflect these issues in our uncertainty range for the proxy used to account for the EP flux variations in multivariate regressions. Salby et al. [2011, 2012] and Kuttippurath et al. [2013] use the August – September mean EP flux poleward of 40ºS and at the 70-hPa level, the baseline also used in this study. Weber et al. [2011] uses the 100-hPa poleward heat flux rather than the 70-hPa heat flux and the average over the region between 45ºS and 75ºS rather than between 40ºS and 90ºS. They further show that there is no particular favorable wintertime month or period from the perspective of Antarctic springtime ozone depletion over which to average the EP flux. Hence there is a certain arbitrariness involved in selecting the optimum EP flux averaging period and region.

For our study we define eight different EP flux scenarios, using different periods, latitudes and heights (see Table 2), all based on the ECMWF ERA Interim dataset. Performing the same exercise as in Figure 1 for these eight scenarios, the standard deviation of the EP flux time series is 21.5%. This is considerably larger than the variability among the same EP fluxes of the different reanalysis datasets discussed above. Thus, the uncertainty in EP flux estimates largely originates in using different periods, latitudes and/or heights for which the EP flux is calculated, rather than in the use of different reanalysis datasets to calculate the same EP flux.

2.3 The mixed solar-QBO index

In Kuttippurath et al. [2013] the effects of solar variability and QBO variability are combined into one proxy. As explained in Holton and Tan [1990], in studying high-latitude variability and trends the QBO and solar effects cannot be considered separately. Whereas the solar influence modifies tropical stratospheric ozone and dynamics, the transport of the solar signal to higher/polar latitudes depends on the phase of the QBO. As a result, solar effects on winter polar Antarctic stratospheric temperatures also depend on the phase of the QBO [Labitzke, 2004]. If the QBO is westerly (easterly), stratospheric temperatures vary in phase (out of phase) with solar activity. It has been proposed by Haigh and Roscoe [2006] and Roscoe and Haigh [2007] to combine the QBO and solar activity into a new regression index that takes this effect into account:

$$Solar-QBO \;\; index = \left(Solar - S_m\right)\times\left(QBO - Q_m\right)$$

In which S_m is the mean of the solar flux and Q_m the midpoint of the QBO range. However, as Roscoe and Haigh [2007] note, this new index is rather sensitive to the choice of S_m and Q_m, in particular as the index is by construction the product of two anomaly fields, and thus sensitive to sign changes. In addition, the choice of S_m and Q_m is also arbitrary. Roscoe and Haigh [2007] solve this by selecting averages for which the best total ozone column regression results are obtained. However, the best regression results may not necessarily mean that the regressor is the best representation of the underlying physical mechanism, in particular as regression results also depend on other proxies and in principle there can be a cancellation of errors from different proxies in the regression. Thus, the sensitivity of the combined solar-QBO index on the calculation method of the anomalies must be further investigated.

Figure 2 shows the resulting solar flux – QBO index time series, given various assumptions in its calculation. Clearly there is a considerable variability in the index values. The lower plot shows that the variability for every single anomaly varies by ± 200%. This is rather large compared to the estimated uncertainties in both individual solar flux and QBO proxies. Hence, using a combined solar flux – QBO proxy introduces a considerable amount of additional uncertainty. For the uncertainty range in our regressions we construct 100 Monte Carlo time series in which for each single solar-flux QBO index value random Gaussian noise is added with an amplitude of 200% of the index value.

Note that the uncertainties in the individual QBO and solar flux proxies are much smaller than the uncertainty in the combined solar flux – QBO index which is relevant for high-latitude trends (see supplementary information for a separate discussion of the solar flux and QBO index).

2.4 Southern Annular Mode

The Southern Annular Mode (SAM) is a widely used index that reflects the zonal symmetry of the tropospheric circulation in the Southern Hemisphere. The symmetry of the Southern Hemisphere circulation has long been identified as an important mode of variability of the Southern Hemisphere climate. A positive index is characterized by anomalously high surface pressure at mid-latitudes and anomalously low surface pressure at latitudes closer to Antarctica.

The SAM used in this study is derived from the National Oceanic and Atmospheric Administration (NOAA). It is based on Empirical Orthogonal Functions (EOF) applied to the monthly mean National Centers for Environmental Prediction and National Center for Atmospheric Research (NCEP/NCAR) reanalysis [Kalnay et al., 1996] 700-hPa height anomalies poleward of 20°S for the Southern Hemisphere, with the seasonal cycle being removed. The monthly SAM index is constructed by projecting the daily and monthly mean 700-hPa height anomalies onto the leading EOF mode. Both time series are normalized by the standard deviation of the monthly index (1979-2000 base time period). Since the leading pattern of SAM is obtained using the monthly mean height anomaly dataset, the index corresponding to each loading pattern becomes one when it is normalized by the standard deviation of the monthly index.

However, there is no unique SAM index due to the existence of different meteorological datasets and different methods to quantify the symmetry of the Southern Hemisphere circulation. Kuttippurath et al. [2013] use the AntArctic Oscillation (AAO) index, which is in fact a certain choice of SAM index. A study by Ho et al. [2012] provides a comprehensive analysis of eight different SAM indices. Their analysis shows that the correlation (R^2) between the indices varies between 0.45 and 0.96 for seasonal values and 0.73 and 0.96 for monthly values. This corresponds with random (Gaussian) variations between 20-100% (root-mean-square value). For most of the indices the correlation is better than 0.75. As a point of reference, adding random Gaussian noise of 50% to a time series of a parameter and calculating its correlation with the original time series results to a correlation (R^2) of almost 0.8.

For the uncertainty analysis we construct 100 Monte Carlo time series in which for each single SAM index value Gaussian noise is added with – to be on the conservative side - an amplitude of 100% of the index value.

2.5 EESC loading

Uncertainties in the estimates of the EESC loading originate from two factors: the mean age-of-air, which reflects how fast stratospheric halogen concentrations decline due to transport velocity of halogen poor tropospheric air from the tropical stratosphere to the polar stratosphere, and the so-called ‘fractional release’, the rate with which Ozone Depleting Substances (ODSs) release chlorine and bromine in the stratosphere. ODSs typically have not yet been dissociated when they enter the stratosphere at the tropical tropopause, and thus have fractional release values of zero. After transiting through the upper stratosphere, the ODSs in an air parcel get fully dissociated due to their exposure to energetic radiation and the fractional release values get close to 1.0 [Newman et al., 2007].

To complicate matters, the mean age-of-air in the stratosphere is not a constant but varies with latitude, height and season [Stiller et al., 2008]. On average, the age-of-air increases with height, *i.e.* it takes longer for tropospheric air to travel higher in the stratosphere, and the age-of-air also increases towards the poles because of the time it takes for air to travel from the tropical "source" region to higher latitudes. In the Antarctic vortex regions there is a strong seasonal dependence of the age-of-air due to the isolation of inner vortex air during Austral winter and spring, while upper stratospheric and mesospheric air slowly descends in the Antarctic vortex. The descending air is particularly "old" air and causes strong vertical gradients in the age-of air in the wintertime polar vortex. Stiller et al. [2008; their figure 7] show that the age-of-air almost triples going up from 15 km (θ = 400 K; age-of-air ~ 4 years), to 20 km (θ = 400 K; age-of air ~ 7 years), to 25 km (θ = 600 K; age-of-air ~ 9 years), to finally 30 km (θ = 750 K; age-of-air ~ 11 years). How to account for this variability in a regression is unclear, but it is unlikely that one age-of-air value can be attributed to the total ozone column.

Moreover, ozone variability in the Antarctic vortex is determined by different processes at different altitudes. Halogen related ozone depletion typically maximizes between 15-20 km altitude (~ 100-50 hPa, US Standard atmosphere 1976; θ = 400-500 K), whereas the effect of vortex stability on ozone depletion is seen predominantly between 20-30 km altitude (~50-10 hPa; θ = 500-750 K) [de Laat and van Weele, 2011]. Thus, total ozone columns observations which are vertically integrated amounts of ozone are being affected by different processes at different altitudes.

The age-of-air may also not be constant over the time period over which ozone trends are determined. Due to a changing climate the stratospheric circulation may speed up [*e.g.* Engel et al., 2009; Bunzel and Schmidt, 2013], causing a decrease in the age-of-air with increased warming, which obviously then depends on the exact warming. This introduces yet another uncertainty for the periods from 1979 to 2010 or 2012 that are considered in this study.

The age-of-air uncertainties do not manifest themselves as a random process, which would make it useful for applying a Monte Carlo method, but as a structural uncertainty, *i.e.* the entire EESC shape would change for different parameter settings. Such uncertainty could be captured by applying a parametric bootstrap rather than a Monte Carlo approach. However, such parametric approach would also not suffice because we use total column observations and we know that ozone at different altitudes would be affected by different parameter values.

A pragmatic approach with regard to the sensitivity of the regression to EESC values is testing the robustness of the regression results as a function of the assumed EESC time evolution. For the uncertainty analysis we assume three different EESC scenarios with an age-of-air of 2.5, 4 and 5.5 years and a half-width of, respectively, 1.25, 2 and 2.75 years. Largest differences between the three scenarios are in their post-peak trend in EESC (see later on in Figure 3).

2.6 Volcanic aerosol.

Aerosols from sufficiently strong volcanic eruptions can reach the stratosphere and affect stratospheric ozone chemistry. In particular strong eruptions occurring in the tropics can have long lasting effects on stratospheric ozone. Aerosols reaching the tropical stratosphere are slowly transported towards middle and high latitudes. It can take up to a decade before the stratosphere is cleared from volcanic aerosols [Vernier et al. 2011; Solomon et al., 2011]. Volcanic eruptions at middle and high latitudes have much shorter lasting effects. These aerosols enter in the descending branch of the stratospheric circulation and will be relatively quickly removed from the stratosphere.

The short-term effect of stratospheric volcanic aerosols is heating of the stratospheric layer which affects stratospheric ozone in the tropical belt. The dominant long-term effect of stratospheric volcanic aerosols on global and polar ozone is however the increase in aerosol surface area density and subsequent heterogeneous ozone loss. Model simulations of volcanic aerosol effects on stratospheric ozone suggest that in particular under cold conditions (high latitude, wintertime, lower stratosphere) total ozone columns can be reduced by up to 10-15 % [Rozanov et al., 2002]. During other seasons, total ozone column depletion by volcanic aerosols is of the order of a few percent.

Since 1979 two major tropical volcanic eruptions have affected stratospheric ozone: El Chichón, Mexico, in 1982, and Pinatubo, Philippines, in 1991. Although the total amount of stratospheric aerosols by both eruptions has been characterized relatively well, there appear to be considerable uncertainties associated with the time evolution of the aerosol amounts in the Southern Hemisphere. A brief and incomplete survey of a latitudinal volcanic aerosol radiative forcing data [Ammann et al., 2003] and a global volcanic aerosol proxy record [Crowley and Unterman, 2012] as well as the standard volcanic aerosol index used in Kuttippurath et al. [2013] – aerosol optical depth, Sato et al. [1993] and updates, available via NASA GISS – all show that there are large differences between the El Chichón aerosol peak relative to the Pinatubo peak. Large differences are seen in global, hemispheric and Southern Hemisphere (Antarctic) aerosol amounts as well as differences in the exact timing of the peak aerosols [Sato et al., 1993; Ammann et al., 2003]; Crowley and Unterman, 2012]. The El Chichón aerosol peak relative to the Pinatubo peak for high Antarctic latitudes can be similar [Ammann et al., 2003], about three times smaller [Sato et al., 1993] to (globally) eight times smaller [Crowley and Unterman, 2012]. The Pinatubo peak aerosol in the Southern Hemisphere was about half the size of the global-mean Pinatubo peak [Ammann et al., 2003].

Kuttippurath et al. [2013] shift the Southern Hemisphere aerosol data by six months to account for the transport of aerosols. Although they report that the six month shift results in the best statistics, the analysis presented in the previous paragraph shows that the effect of the shift is relevant for the shape of the volcanic aerosol changes, but does not introduce variations as large as the other variations in volcanic aerosol indices. Given that a time shift is included in the 6 volcanic aerosol scenarios defined above, we do not add additional time shifts in the aerosol record.

We define six volcanic aerosol scenarios that reflect the uncertainty in the volcanic stratospheric aerosol records. Base scenario is the scenario used in Kuttippurath et al. [2013] which in turn uses the NASA GISS stratospheric aerosol record. A second scenario is with the Pinatubo aerosol curve scaled so that the maximum matches the El Chichón aerosol peak, the Pinatubo curve maximum is 2.5 times the El Chichón peak, and the Pinatubo curve maximum is five times the El Chichón peak. The uncertainty in timing of the Southern Hemispheric aerosol peak is considered by a shift of the El Chichón peak one year back compared to the Pinatubo peak and a shift of the Pinatubo peak one year back compared to El Chichón peak.

2.7 Ozone Scenarios.

It is *a priori* unclear what would be the most appropriate ozone scenario to use in the regression. Both Salby et al. [2011, 2012] and Kuttippurath et al. [2013] use the September-November three-month averaged total ozone record. However, as discussed in the introduction, different processes affect ozone during different time periods. Studies in the literature use very different time periods for averaging ozone to investigate Antarctic ozone trends. We define eight different ozone scenarios to reflect the ozone records used in literature (see also de Laat and van Weele [2011]), using the MSR dataset [van der A et al., 2010]. The MSR is a 30-year total O_3 column assimilation dataset for 1979–2008 based on a total of eleven satellite instruments measuring total O_3 columns – including SCIAMACHY - that were operating during various periods within these 30 years. For the period 2009-2012 the MSR dataset was extended with assimilated SCIAMACHY and GOME-2 total ozone column data. Apart from the September-November three-month averaged total ozone record we also use averages of total ozone over the month of September, the month of October, the two-month period September-October, a very long period (19 July – 1 December), a very short 10-day period (21 – 30 September), the period 7 September – 13 October, and a year-dependent "worst" 30-day period (30-day average with the largest Ozone Mass Deficit).

2.8 Other uncertainties

Kuttippurath et al. [2013] address two other important uncertainties for the determination of the ozone trend. First, the area over which the ozone record is defined (Inside Vortex, Equivalent Latitude 65ºS-90ºS, and Vortex Core). The area is important for the absolute amounts of ozone depletion but Kuttippurath et al. [2013] show it is much less relevant for the differences in trend. That is, the uncertainties in the estimated linear trend dominate the uncertainties due to different areas over which the ozone anomalies are calculated. A second uncertainty on their ozone trend derives from the use of different ozone datasets (ground-based, TOMS/OMI and MSR). Also here the uncertainties in the estimated linear trend dominate the uncertainties due to the different data sets. Hence, we do not include these uncertainties in our analysis.

In addition, there are many studies trying to identify the moment where ODSs stop increasing and/or where ozone stops decreasing. The maximum ODSs appears somewhere between 1997 and 2000 (Newman et al., 2007), depending on geographical location and height. However, due to saturation effects – there are more than sufficient ODS present to destroy all Antarctic ozone – the moment where ozone starts to be affected by decreasing ODSs may actually be later (Kuttippurath et al., 2013; Kramarova et al., 2014).

The moment of a structural break in ozone based on observations indicates an early break around 1997 (Newchurch et al., 2003; Yang et al., 2008). However, some processes affecting stratospheric ozone vary on long time scales – solar effects and volcanic eruptions come to mind – which may affect the observations-based analysis of break points (Dameris et al., 2006). Note that we confirm this break year of 1997 based on a applying a break-point analysis algorithm to the MSR ozone record (not shown). Hence, we decided to use three different break years that have been identified and/or are most commonly used: 1997, 1998 and 1999.

2.9 Selected uncertainties ranges and ozone record scenarios.

Figure 3 shows the baseline regressor time series and the scenarios for ozone, the EP flux, EESC loading and volcanic aerosols. A total of 100 different solar flux – QBO index and SAM index time series are used to span their uncertainty range (not shown in Figure 3). All scenarios and Monte Carlo results combined provide 11.5 million different choices for the regressions (100×100×8×8×6×3; see Table 2). Ozone trends are calculated based on the EESC loading or using a piecewise linear trend (PWLT) analysis. For the PWLT ensembles the three different EESC scenarios are irrelevant. Instead, the sensitivity of the regressions is tested using three different break years (1997, 1998 and 1999). In total we analyze approximately 23 million ~~000~~ different trends using the EESC and PWLT scenarios.

Note that by basing our analysis on both different ozone and EP flux scenarios certain time-lag relations are taken into account. It should also be noted that the use of such a wide range of scenarios indicates that much remains unclear about what best describes Antarctic ozone depletion and the time-lag relations between ozone and explanatory variables.

3 Scenario analysis

3.1 Reproducing Kuttippurath et al., [2013].

First a ~~multi-~~multivariate regression is performed similar to Kuttippurath et al. [2013] in which the MSR dataset is used within the Vortex core (70º-90ºS). The results are summarized in their Figure 5 and Table 4 which are duplicated here in Table 3 along with the results from a ~~multi-~~multivariate regression based on the same variables as used in Kuttippurath et al. [2013].

Our results reproduce the results from Kuttippurath et al. [2013], although there are minor differences in the absolute numbers, most likely related to differences in EP fluxes [Jayanarayanan Kuttippurath, *personal communication*, September 2013]. The trends for the periods 1979-1999 and for 2000-2010 are of comparable magnitude in both studies, as well as the PWLT significance levels for the period 1979-1999 and the EESC trends for both 1979-1999 and 2000-2010. The magnitude of the recovery for 2000-2010 based on the PWLT is slightly larger, but also in our analysis the post 2010 linear trend in ozone is significant beyond 2σ. For the correlation of the regression model with the ozone record we obtain a value of 0.87 (R^2) comparable to the 0.90 (R^2) reported in Kuttippurath et al., [2013]. Thus, the results are sufficiently similar to proceed with studying the effects of the uncertainties in regressors and ozone record scenarios on the regression results. Note that there is slight difference in the 1979-1999 trends for the period ending in 2010 and 2012 because of the difference in total record length, which results in slightly different regression coefficients. ~~Note that we calculate the pre-break and post-break EESC-based trends by applying linear regressions to the EESC curve multiplied with the EESC regression coefficient for the pre-break and post-break time~~

~~periods. As a result, EESC-based trend errors are related to the non-linearity of the EESC curve, and the trend errors differ for both the pre-break and post-break time periods. Our EESC-based trend errors differ from those in Kuttippurath et al (2013), which lacks a description of how EESC-based trend errors are calculated.~~

However, calculating trend errors for the EESC-based pre-BREAK and post-BREAK trends in ozone using the EESC regression error as done in Kuttippurath et al. (2013) is not justified. The trend errors depend on the actual trend values themselves (Table 3): the EESC-fit based post-break trend error is much smaller than the pre-BREAK trend. In the hypothetical case of no (zero) trend the trend error would also be zero, which would be physically unrealistic. The PWLT on the other hand shows hand shows opposite differences in trend errors: the post-BREAK trend error is much larger than the pre-BREAK trend error, conform expectations.

3.2 Ozone record and regressor correlations.

Before analyzing the ensemble of regression results it is important to investigate the correlations between the different regressors. If correlations between regressors are too large, they cannot be considered to be independent, and it should be decided which one to omit from the analysis, as the regression otherwise cannot separate which variability is related to which regressor. Furthermore, it is *a priori* useful to understand how regressors correlate with the ozone record, as a small correlation implies that a regressor can only explain a limited amount of ozone variability.

Table 4 shows the mean correlation between the different regressors and their 2σ spread based on the ozone record and regressor selections and/or Monte Carlo results (SAM, SF×QBO index). The EP flux correlates positively with the EESC and negatively with the SAM and, to a lesser extent, also with the SF×QBO index. The other regressors do not show significant cross-correlations. Only for a few individual ozone record scenarios, regressor selections and Monte Carlo results cross-correlations are found to exceed 0.5.

The uncertainty in the correlations with the ozone records ranges between about 10% and 20% for each of the regressors. Small cross-correlations between the regressors however do not provide a justification for *a priori* omitting one of the regressors.

3.3 Trends.

Figure 4 shows the probability distributions of the ozone trends for 1979-Y_B and Y_B-2012 periods, in which Y_B is the break year which can either be 1997, 1998 or 1999. For the 1979-Y_B period the mean EESC trend is -5.56 DU/year (-4.00 to -7.06; 95% CI) and the mean PWLT trend is -6.40 DU/year (-4.22 to -7.18; 95% CI). For the Y_B-2012 period the mean EESC trend is +1.97 DU/year (+0.84 to +3.32 DU/year; 95% CI), and the mean PWLT trend is +3.18 DU/year (+1.66 to +4.74; 95% CI).

For the 1979-Y_B period the distributions of EESC and PWLT trends (top panel) are rather similar, although the PWLT correlations show a larger peak towards high correlations compared to the EESC correlations (bottom panel). However, for the Y_B-2012 trends the probability distributions are very different (middle panel). The EESC trends show a tri-modal distribution, because only three different EESC curves were used. These three EESC curves differ predominantly in their post-1997 EESC trends (see also Figure 3). In addition, the tri-modal EESC trend probability distribution for Y_B-2012 (middle panel) shows that in the linear regression the EESC fit is determined by the 1979-Y_B period more than by the Y_B-2012 period, as the pre-break trend distribution does not show the same tri-modal shape. This is not surprising because the trends for the 1979-Y_B period are larger and cover a longer period than for Y_B-2012.

The correlations distributions (lower panel) are similar for the lowest and highest correlations for both the EESC and PWLT regressions, but in the bulk of the distribution the PWLT results in systematically higher correlations than the EESC regressions.

The upper two panels of Figure 4 also include the 1979-1999 and 2000-2012 PWLT trends and 2σ errors as reported in Table 32. The uncertainty range of the 2000-2012 PWLT trend in Table 32 and the range in Figure 4 are quite similar. However, the uncertainty range of the 1979-1999 PWLT trend in Table 32 is considerably smaller. This shows that uncertainties in the 1979-1999 ozone trends are larger than estimated by a single regression estimated even though all 1979-1999 trends are statistically significant.

The auto-correlation of the ozone residuals is small (one-year lag values are approximately zero), indicating that the auto-correlation present in the ozone record (*e.g.* Fioletov and Shepherd, 2003; Vyushin et al., 2007) is related to some of the processes described by the regression parameters and are removed by the multi- multivariate regression. Auto-correlation thus does not have to be taken into account in the trend significance calculation.

3.4 Regression model performance: sensitivity to the independent variables

Sensitivities of the PWLT-based and EESC-based regressions to the ozone and EP flux scenarios are shown in Figure 5. PWLT-based regressions show that the PWLT distribution peak at high correlations is a consistent feature of different ozone records (*Sep-Nov, Sep-Oct, Sep, 7 Sep - 13 Oct, worst 30 days*). Similarly, use of several different EP fluxes also aligns with the PWLT correlation distribution peak, in particular the EP flux scenarios that include both the August and September months. For ozone, correlations get smaller for, respectively, the longest period (19 July – 1 December), shortest period (21-30 September) and October averages.

Figure 6 shows the probability distribution of volcanic aerosols for both the PWLT and EESC regressions. Volcanic aerosols have little impact on the explanatory power of the regression results, as already indicated by lack of correlation of this regressor with the ozone record. The PWLT regression coefficient values show that the effect of volcanic aerosols on total ozone can be either positive or negative, largely depending on the assumed amount of Pinatubo aerosols relative to El Chichón aerosols, although the distribution predominantly suggests positive regression values. The EESC regressions show a similar sign dependence of total ozone on volcanic aerosol, but with no clear sign of the regression value. None of other parameters (EP~~-FLUX-~~ flux scenario, Ozone scenario) have a sign-dependent effect on the aerosol regression coefficient value for both the EESC and PWLT scenarios. The strong sensitivity of the volcanic aerosol regression value – including sign changes – to either aerosol or EESC scenario indicates that including volcanic aerosols is not very important for the multivariate regression and better should be excluded altogether from ~~multi-~~multivariate regressions due to insufficient information in the Antarctic total ozone record to constrain the total ozone – volcanic aerosol relation.

For the solar flux – QBO index (Figure 7, panel A) we find no clear dependence of regression coefficient values on any of the scenarios or parameters. The probability distributions for both the EESC and PWLT regressions are very similar. Hence, like for volcanic aerosols, the solar-QBO parameter better should be excluded altogether from ~~multi-~~multivariate regressions because the Antarctic ozone record also contains insufficient information to constrain the ozone – solar-QBO relation

The SAM regression coefficient values show a continuous random distribution while the overall dependence is predominantly negative (Figure 7, panel B). A positive phase of the SAM correlates with more ozone depletion than a negative phase of the SAM. This is a well-known two-way effect: tropospheric circulation changes affect Antarctic stratospheric ozone on the short term, while the long term changes in Antarctic ozone have affected the tropospheric circulation in the Southern Hemisphere [Kirtman et al., 2013; IPCC AR5, Ch. 11, section 11.3.2.4.2 and references therein].~~.~~

~~-~~For the EP~~-FLUX~~ flux the regressions show a positive dependence (Figure 7,~~.~~ panel C) and a similar distribution for both the EESC and PWLT regression.

3.5 Optimal regressor and ozone record scenarios

Based on the analysis of the entire ensemble presented here it might be possible to choose an optimal set of regressors as well as an optimal ozone record scenario for Antarctic ozone trend analysis. Volcanic aerosols (Figure 6), the QBO and the solar cycle

(Figure 7) are shown to have little effect on the regression and thus better should be excluded. For the EP flux, it appears that including the months August and September leads to a better fit (higher correlations; Tables 5~~4~~ and 6~~5~~, Figure 5). For ozone, results suggest that there is no clear optimal time-window over which to calculated average ozone, but it appears that the period should not be too short, not be too long, and should include September and preferably the first half of October (Tables 5~~4~~ and 6~~5~~).

In addition, the use of three different EESC scenarios results in tri-modal distribution features in several parameters (Figures 4~~3~~, middle panel and 6, lower panel), suggesting that care has to be taken with in particular the ozone trend values attributed to changes in EESC~~'~~s. Furthermore, the post-break trends are particularly sensitive to the choice in EESC scenario (Figure 3). It could therefore be argued that using a PWLT for post-break trend estimates is preferred over using the EESC-based post-break trend as its distribution better reflects structural uncertainties in the regression and takes the regression residuals into account for calculation of trend uncertainties.

Figure 8 illustrates what the best single regressions in the entire ensemble for all three regression models separately look like. The best EESC-regression correlation ($R^2 = 0.95$) was found for a case with Sep -Nov ozone, Jul-Aug EP flux and an EESC with an age-of-air of 4 years. For the best PWLT-regression correlation ($R^2 = 0.96$) these were the same with 1997 as optimal break year. Reason for the high explanatory power is that in all three cases the specific SAM anomalies align with strong ozone peaks whereas the solar flux – QBO index variations coincidentally align with the smaller ozone anomalies.

4. Discussion: Second stage of ozone recovery and trend significance.

Given the broad range of outcomes for the different types of regressions and regressors, an important question is not only if ozone has started to increase after the late 1990s, but also whether the trend is statistically significant and can be attributed to declining stratospheric halogens, which is required by WMO for the second stage of ozone recovery to be formally identified. Because the EESC curve-shape is prescribed, there is no degree of freedom allowing for different pre-break and post-break trends in the EESC regression. As discussed in section 2, it is not clear *a priori* which EESC scenario is the optimal choice or if it is even appropriate to use just a single EESC scenario. Furthermore, as already discussed in section 3.1, the method for estimating the trend uncertainty of the post-BREAK trend for the EESC fit in Kuttippurath et al. [2013] is not justified, and ~~Hence,~~ how to assign an ~~overall~~ uncertainty to the EESC curve remains an open question. Therefore, a ~~better~~ more appropriate approach would be to investigate whether the PWLT post-break trends are statistically significant as they use the ozone fit residuals for their significance calculation.

Figure 9 shows the probability distribution of correlations (R^2) of the PWLT regression models *vs.* ozone for the entire Monte Carlo dataset, as well as the fraction of post-break PWLT trend estimates that are statistically significant (2σ) for both the periods ending in 2010 and 2012. This figure is comparable to Figure 4 (lower panel) and Figure 5, but with larger correlation bins for visualization purposes. Results indicate that trends only become statistically significant beyond a certain explanatory power of the regression model. This is not surprising: only when ozone residuals after removing the regression results are sufficiently small can the post-break trend become statistically significant.

This automatically requires a high correlation between the ozone record and the selected regression model. The analysis here shows that statistically significant trends require a correlation (R^2) of at least approximately 0.60. Furthermore, only for high ozone-regression model correlations ($R^2 > 0.80$) the majority of trends become statistically significant. In addition, the level of significance for the statistically significant trends is in most cases still less than 3σ (not shown), indicating that a considerable amount of variability is not accounted for in the regression.

In section 3.5 the results of the ensemble were analyzed to determine optimal scenarios in terms of explanatory power (R^2). However, the second stage of ozone recovery requires also a statistically significant post-break year trend. We therefore analyzed the percentage of statistically significant post-break trends in the ensemble for the PWLT-based regressions. We focus on the ozone record and EP flux scenarios as the uncertainties associated with these two parameters are the most important ones, as discussed before. Table 5 shows the percentage of regressions for each combination of ozone record and EP flux scenarios that is statistically significant for the ozone records ending in 2010. There are large differences in the fraction of statistically significant PWLT-based trends, ranging from less than 0.1% (*21-30 September* average ozone) to a complete 100% significance (*September-October* and *October* ozone, *45S-75S Aug-Sep EP flux*). Table 6 shows the same results as Table 5, but only for the break year 1997 and the period ending in 2012. In this case there is a large number of ozone record - EP flux scenario combinations with statistically significant trends. If we would consider only the EP fluxes that include the August and September months, then with the exception of the *21-30 September* time window nearly all trends are statistically significant.

Excluding the year 2002 from the regressions has a significant impact on the post-break ordinary linear ozone trends without applying a multi-multivariate regression. However, it hardly has any effect on the post-break ordinary linear trends when including the ordinary linear trend in the multi-multivariate regressions (not shown), indicating that the multi-multivariate regression effectively removes the anomalous year 2002. Excluding volcanic years from the regression had no significant effect on both the ozone trends before and after the regression, consistent with our finding that there appears to be little (direct) impact of volcanic aerosols on Antarctic springtime total ozone. Note that Solomon et al. [2005] showed that volcanic effects can be seen in Antarctic ozone profiles during the ozone hole season – in particular in the UTLS region, but that the magnitude and vertical extent of the effects are too small to be detectable in total ozone column variations.

Table 7 shows the number of significant trends as function of the length of the period over which the trend is calculated. The number of significant trends varies between approximately 30-60%, and that the number of significant trends further depends on the choice of break year, with an overall increase in the number of statistically significant trends increasing steadily with increasing length of the period over with trends are calculated. . This is not surprising as the regression trend error decreases with increasing number of points for which the trends are calculated (Supplementary Information equation S2). Furthermore, this implies that with continued extension of the total ozone record detection of Antarctic ozone recovery may be reached before 2020 using multi-multivariate regressions. Note that although in total the number oif statistically significant trends increases with record length, this is not necessarily always the case: (For example by comparinge Table 5 and Table 6 – thus for it is concludeshowsd that for some scenario combinations the number of significant trends is larger for a (shorter) period (BREAK-2010) than for the full period vs. 1998-2012 trends).

Excluding the year 2002 from the regressions has a significant impact on the post-break ozone trends themselves. However, it hardly has any effect on the post-break trends from the regressions (not shown), indicating effective removal of the anomalous year 2002 from the results. Excluding volcanic years from the regression had no significant effect on both the ozone trends before and after the regression, consistent with our finding that there appears to be little (direct) impact of volcanic aerosols on Antarctic springtime ozone.

It is tempting to interpret, based on some selections of our results, that the significance is sufficient for identification of the second stage of ozone recovery by 2012. However, comparing Table 5 and Table 6 – thus 2000-2010 trends vs 1998-2012 trends – shows that the longer period not always results in increased statistical significance. In particular, the need to average ozone over longer periods of time may introduce long term changes in average ozone that are not related to photochemical ozone destruction. Furthermore, the tTrend significanceothe is generally between 2σ and 3σ (not shown), indicating that a considerable amount of variability is not accounted for in the regression. In addition, our analysis shows sthat detection of the 2nd stage of ozone recovery based on just one more or less arbitrary selected (set of) regressor – ozone record combination(s) does not reflect the structural uncertainties present in the underlying data.

Nevertheless, the appearance of larger groups of statistically significant results occurring for longer time series and a certain persistence among ozone scenarios and EP flux scenarios~~,~~ shows that ~~these type of~~ multivariate regression, preferably using piecewise linear analyses before and after a predefined break year, are capable of removing deterministic variations in average ozone, and that with increasing length of the post-break period more and robust statistically significant results can be expected.

5. Conclusions

The primary goal of this study was to investigate whether or not the 2nd stage of ozone recovery – a statistical increase in ozone attributable to ozone depleting substances – can be detected, given uncertainties in underlying data. A detailed sensitivity analysis of widely used ~~multi-~~multivariate regression analysis of total ozone columns was presented focusing on Antarctic springtime ozone. By combining regressor scenarios and Monte Carlo simulations for various ozone record scenarios, a total of approximately 23 million different multivariate regressions were performed.

Our analysis shows that detection of the 2nd stage of ozone recovery based on one more or less arbitrary selected (set of) regressor – ozone record combination(s) does not reflect the structural uncertainties present in the underlying data.

Use of the post-break trends based on fitting the EESC to the total ozone record is not recommended, as these trends are solely based on the pre-defined EESC shape, ~~and~~ do not allow flexibility in the trend calculation while it is unclear how to assign uncertainties to the EESC-regression-based trends in total ozone. ~~Because the resulting EESC fit based trend uncertainties do not take the ozone fit residuals into account the EESC scenarios result in overconfident ozone trend uncertainties, neglecting structural uncertainties and sensitivity to the chosen scenario.~~

Our analysis further shows that the EP flux and the SAM effects are capable to explain significant parts of Antarctic ozone variations and the removal of these effects improves the analysis of recovery~~,~~, in contrast to the inclusion in the regressions of volcanic aerosols and the combined QBO/Solar flux index.

Consistent with expectations, we find a robust gradual small increase in Antarctic ozone since the late 1990s that can be attributed to decreases in ODS for selected combinations of regressors, although the magnitude of the increase is rather uncertain (~~+~~1.66 t– ~~+~~4.74 DU/year; 95% CI). The trend significance shows a clear dependence on the length of the period over which the trend is calculated. The number of statistically significant trends in our ensemble varies between approximately 30-60%, depending on the length of the period, with an average of approximately 50%.

The limited information present in the Antarctic ozone record for volcanic aerosols (essentially two isolated peaks) is consistent with Knibbe et al. [2014], who found little evidence for volcanic effects on total ozone throughout the Southern Hemisphere. Furthermore, Poberaj et al. [2011] also reported little impact of volcanic aerosols from the Pinatubo eruption on Southern Hemispheric ozone, attributing it to dynamical conditions favoring more poleward transport of ozone from the tropics and mid-latitudes than usual, thereby "*overcompensating the chemical ozone loss …and reduce the overall strength of the volcanic ozone signal*".

The lack of correlation between Antarctic ozone and the solar-flux/QBO combined index was also found by Knibbe et al. [2014] for both Antarctic (and Arctic) ozone trends. This lack of QBO-solar signal in Antarctic springtime ozone – also e.g. reported in both Labitzke [2004] and Roscoe and Haigh [2007] - may be related to the dominance in absolute values of the ozone change of ozone depletion and vortex dynamics over potential indirect solar influences on Antarctic springtime ozone.

From our analysis it remains unclear what the appropriate time window would be over which to average the ozone record and the EP flux. Results indicate that the best regression occur for ozone averaged over a time window that includes the ozone hole season – typically September and part of October. On the other hand, the time window should also not extend far beyond the ozone hole season as more and more non-photochemical ozone variations are introduced in the averaged ozone with a longer averaging period. Similarly, for the EP flux we find that including both the August and September months result in the best regressions. However, the choice for using complete calendar months is rather arbitrary, and better choices may exist which is here left for future research.

The lack of a proper definition of appropriate time windows drives our recommendation that care has to be taken with drawing firm conclusions about Antarctic ozone recovery based on ~~multi-~~multivariate regression of Antarctic vortex average ozone. Given uncertainties in parameters and independent variables and choices in defining the optimal time period and area for calculating the ozone record and the independent variables ~~I~~it is tempting to discard those results ~~in the full ensemble~~ that do not confirm our expectations, but without proper justification of what constitutes the best set of independent explanatory variables – for example physically compelling arguments - there is the danger of working towards an expected answer.

~~Another last finding is that~~Finally, ~~a longer post-break period does not necessarily always results in more significant trends, which provides yet another indication to remain careful with drawing too firm conclusions from multivariate regressions. On the other hand~~despite these uncertainties, our results indicate it can be expected that with extending the ozone record and using a ~~multi-~~multivariate regression method to remove well-selected non-ODS influences from the total ozone record – the second stage of recovery of the Antarctic ozone-hole ~~will~~ may be detectable before 2020. Future updates of the analysis in this paper by extension of the present-day ozone records are expected ~~will rather soon~~ provide indications whether this moment approaches fast or not.

Acknowledgements

The authors wish to thank the following authors of chapter 3 of the 2014 WMO ozone assessment report – Sophie Godin Beekmann, Martin Dameris, Peter Braesicke, Martin Chipperfield, Markus Rex and Michelle Santee – as well as John Pyle, Ted Shepherd and in particular Paul Newman for encouraging us to write this paper.

References

Van der A et al. (2010), Multi sensor reanalysis of total ozone Atm. Chem. Phys., 10, 11277-11294, doi:10.5194/acp-10-11277-2010.

Ammann et al. (2003), A monthly and latitudinally varying volcanic forcing dataset in simulations of 20th century climate, Geophys. Res. Lett., 30, 1657, doi:10.1029/2003GL016875.

Andrews et al. (1987), Middle Atmosphere Dynamics, Academic Press, Orlando, Florida, 489 pp.

Anet et al. (2013), Impact of a potential 21st century “grand solar minimum” on surface temperatures and stratospheric ozone, Geophys. Res. Lett., 40, 4420–4425, doi:10.1002/grl.50806.

Bekki et al., (2011), Future Ozone and its impact on surface UV, Chapter 3 Scientific Assessment of Ozone Depletion: 2010, Global Ozone Research and Monitoring Project - Report No. 52, 516 pp., Geneva, Switzerland.

Bunzel and Schmidt (2013), The Brewer–Dobson Circulation in a Changing Climate: Impact of the Model Configuration. J. Atmos. Sci., 70, 1437–1455. doi: http://dx.doi.org/10.1175/JAS-D-12-0215.1

Crowley and Unterman (2012), Technical details concerning development of a 1200-yr proxy index for global volcanism, Earth Syst. Sci. Data Discuss., 5, 1-28, doi:10.5194/essdd-5-1-2012.

Dameris, M., S. Matthes, R. Deckert, V. Grewe, *and* M. Ponater (2006), Solar cycle effect delays onset of ozone recovery, Geophys. Res. Lett., 33, *L03806*, *doi:*10.1029/2005GL024741.

De Laat and van Weele, (2011) The 2010 Antarctic ozone hole: Observed reduction in ozone destruction by minor sudden stratospheric warmings, *Sci. Rep.*, *1*, *38*, doi: 10.1038/srep00038.

Dudok de Wit et al. (2009), Finding the best proxies for the solar UV irradiance, Geophys. Res. Lett., 36, L10107, doi:10.1029/2009GL037825.

Ebdon (1960): Notes on the wind flow at 50 mb in tropical and sub-tropical regions in January 1957 and January 1958. Q. J. Roy. Met. Soc., 86, 540-542.

Engel et al. (2009), Age of stratospheric air unchanged within uncertainties over the past 30 years, Nature Geoscience 2, 28 – 31, doi:10.1038/ngeo388.

Eyring, et al. (2007), Multimodel projections of stratospheric ozone in the 21st century, J. Geophys. Res., 112, D16303, doi:10.1029/2006JD008332.

Fioletov, V. E., *and* T. G. Shepherd (2003), Seasonal persistence of midlatitude total ozone anomalies, Geophys. Res. Lett., *30*, *1417*, *doi:*10.1029/2002GL016739.

Graystone (1959), Meteorological office discussion on tropical meteorology. Met. Magazine, 88, 117.

Haigh, J.D., The impact of solar variability on climate (1996), Science, Vol. 272, no. 5264 pp. 981-984, DOI: 10.1126/science.272.5264.981

Haigh and Roscoe (2006), Solar influences on polar modes of variability, Meteorologische Zeitschrift, Volk.15, pp. 371-378

Ho et al. (2012) The Southern Annular Mode: a comparison of indices, Hydrol. Earth Syst. Sci., 16, 967-982, doi:10.5194/hess-16-967-2012.

Hood, L. (1997), The solar cycle variation of total ozone: Dynamical forcing on the lower stratosphere, *Journal of Geophysical Research*, *102*(D1), 1355-1370.

Huang and Massie (1997), Effect of volcanic particles on the O2 and O3 photolysis rates and their impact on ozone in the tropical stratosphere, J. Geophys. Res., 102(D1), 1239–1249, doi:10.1029/96JD02967.

Jiang et al. (2008): Interannual Variability and Trends of Extratropical Ozone. Part II: Southern Hemisphere. J. Atmos. Sci., 65, 3030–3041, doi: 10.1175/2008JAS2793.1

Kalnay, E., et al. (1996). The NCEP/NCAR 40-Year Reanalysis Project, *Bull. Am. Met. Soc*, **77** (3): 437–471.

Kirtman et al. (2013), Near-term Climate Change: Projections and Predictability. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Stocker et al. (EDS.), Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, *in press*.

Klekociuk et al. (2011), The Antarctic ozone during 2010, *Australian Meteorological and Oceanographic Journal, 61 (4)*, 253-267.

Knibbe, J. S., et al. (2014), Spatial regression analysis on 32 years of total column ozone data, Atmos. Chem. Phys., 14, 8461-8482, doi:10.5194/acp-14-8461-2014,.

Kramarova et al. (2014), Measuring the Antarctic ozone hole with the new Ozone Mapping and Profiler Suite (OMPS), Atmos. Chem. Phys., 14, 2353-2361, doi:10.5194/acp-14-2353-2014.

Kuttippurath et al. (2010), Estimation of Antarctic ozone loss from ground-based total column measurements, *Atmos. Chem. Phys., 10*, doi: 10.5194/acp-10-6569-2010, 6569-6581.

Kuttippurath et al., Antarctic ozone loss in 1979–2010 (2013) first sign of ozone recovery, *Atmos. Chem. Phys., 13*, doi: 10.5194/acp-13-1625-2013, 1625-1635,.

Labitzke (2004), On the signal of the 11-Year sunspot cycle in the Stratosphere over the Antarctic and its modulation by the Quasi-Biennial Oscillation (QBO), Meteorologische Zeitschrift, Vol. 13, No. 4, 263-270.

McCormack, J. P., & Hood, L. L. (1996). Apparent solar cycle variations of upper stratospheric ozone and temperature: Latitude and seasonal dependences. Journal of Geophysical Research: Atmospheres, 101(D15), 20933-20944.

Naujokat (1986), An update of the observed quasi-biennial oscillation of the stratospheric winds over the tropics. J. Atmos. Sci., 43, 1873-1877.

Newman et al. (2006), When will the Antarctic ozone hole recover? Geophys. Res. Lett., 33, L12814, doi:10.1029/2005GL025232.

Newman et al. (2007), A new formulation of equivalent effective stratospheric chlorine (EESC), Atmos. Chem. Phys., 7, 4537-4552, doi:10.5194/acp-7-4537-2007.

Poberaj et al. (2011): Missing Stratospheric Ozone Decrease at Southern Hemisphere Middle Latitudes after Mt. Pinatubo: A Dynamical Perspective. J. Atmos. Sci., 68, 1922–1945. doi: http://dx.doi.org/10.1175/JAS-D-10-05004.1

Randel et al. (2002), Changes in column ozone correlated with the stratospheric EP flux, J. Meteor. Soc. Jap., 80, 4b, 849-862.

Rozanov et al. (2002), Climate/chemistry effects of the Pinatubo volcanic eruption simulated by the UIUC stratosphere/troposphere GCM with interactive photochemistry, J. Geophys. Res., 107(D21), 4594, doi:10.1029/2001JD000974.

Salby et al. (2011), Rebound of Antarctic ozone, *Geophys. Res. Lett., 38*, L09702, doi: 10.1029/2011GL047266.

Salby et al. (2012), Changes of the Antarctic ozone hole: Controlling mechanisms, seasonal predictability, and evolution, *J. Geophys. Res., 117*, D10111, doi: 10.1029/2011JD016285.

Sato et al. (1993), Stratospheric aerosol optical depths, 1850-1990. J. Geophys. Res., 98, 22987-22994, doi:10.1029/93JD02553.

Solomon, S. et al. (2005), Four decades of ozonesonde measurements over Antarctica, J. Geophys. Res., 110, D21311, doi:10.1029/2005JD005917.

Solomon, S., et al. (2011), The persistently variable “background” stratospheric aerosol layer and global climate change. Science, Vol. 333, no. 6044, pp. 866-870, doi: 10.1126/science.1206027

Soukharev and Hood (2006), Solar cycle variation of stratospheric ozone: Multiple regression analysis of long-term satellite data sets and comparisons with models, J. Geophys. Res., 111, D20314, doi:10.1029/2006JD007107.

Stiller et al. (2008), Global distribution of mean age of stratospheric air from MIPAS SF6 measurements, Atmos. Chem. Phys., 8, 677-695, doi:10.5194/acp-8-677-2008.

Telford et al. (2009), Reassessment of causes of ozone column variability following the eruption of Mount Pinatubo using a nudged CCM, Atmos. Chem. Phys., 9, 4251-4260, doi:10.5194/acp-9-4251-2009, 2009.

Thompson, D.W.J. and J.M. Wallace (2000): Annular Modes in the Extratropical Circulation, Part I: Month-to-Month Variability, J. Climate, 13, 1000–1016.

United Nations Environment Programme (2012), The Montreal Protocol on Substances that Deplete the Ozone Layer, Nairobi, Kenya, ISBN 978-9966-20-009-9.

Vernier, J.-P., et al. (2011), Major influence of tropical volcanic eruptions on the stratospheric aerosol layer during the last decade, Geophys. Res. Lett., 38, L12807, doi:10.1029/2011GL047563.

Vyushin, D., V. E. Fioletov, and T. G. Shepherd *(*2007*)*, Impact of long-range correlations on trend detection in total ozone, J. Geophys. Res., 112, *D14307*, *doi:*10.1029/2006JD008168

Weber et al. (2003), Dynamical control of NH and SH winter/spring total ozone from GOME observations in 1995-2002, *Geophys. Res. Lett*., *30*, 1853, doi: 10.1029/2002GL016799.

Weber et al. (2011), The Brewer-Dobson circulation and total ozone from seasonal to decadal time scales, *Atmos. Chem. Phys., 11*, 11221-11235, doi: 10.5194/acp-11-11221-2011.

World Meteorological Organization (2007), Scientific Assessment of Ozone Depletion: 2006, Global Ozone Research and Monitoring Project - Report No. 50, 572 pp., Geneva, Switzerland.

World Meteorological Organization (2011), Scientific Assessment of Ozone Depletion: 2010, Global Ozone Research and Monitoring Project - Report No. 52, 516 pp., Geneva, Switzerland.

Yang, E.-S., D. M. Cunnold, M. J. Newchurch, R. J. Salawitch, M. P. McCorm~~i~~ack, J. M. Russell III, J. M. Zawodny, and S. J. Oltmans *(*2008*)*, First stage of Antarctic ozone recovery, J. Geophys. Res., 113, D20308, doi:10.1029/2007JD009675.

Tables

EP flux http://www.awi.de/en/research/research_divisions/climate_science/atmospheric_circulations_old/projects/candidoz/ep_flux_data/
QBO http://www.geo.fu-berlin.de/met/ag/strat/produkte/qbo/
Solar flux ftp://ftp.geolab.nrcan.gc.ca/data/solar_flux/monthly_averages/solflux_monthly_average.txt
SAM ftp://ftp.cpc.ncep.noaa.gov/cwlinks/
EESC http://acdb-ext.gsfc.nasa.gov/Data_services/automailer/
Volcanic aerosol http://data.giss.nasa.gov/modelforce/strataer/
Assimilated total ozone http://www.temis.nl/protocols/O3global.html

Table 1. Data sources

regressor	**variations**
Average EP flux - 8 scenarios	- 70 hPa, 40ºS-90ºS, Aug-Sep (baseline) - 70 hPa, 40ºS-90ºS, Jul-Aug - 70 hPa, 40ºS-90ºS, Jul-Sep - 70 hPa, 40ºS-90ºS, Jul - 70 hPa, 40ºS-90ºS, Aug - 70 hPa, 40ºS-90ºS, Sep - 70 hPa, 45ºS-75ºS, Aug-Sep - 100 hPa, 40ºS-90ºS, Aug-Sep
Solar flux – QBO index - 100 Monte Carlo series	- Random variations in Solar flux – QBO anomalies - 200% Gaussian noise variations on single solar flux – QBO anomalies

SAM index - 100 Monte Carlo series	- 100% random error on annual mean SAM index values
EESC loading - 3 scenarios	- EESC shapes based on different age of air of 2.5, 4.0 and 5.5 years
Volcanic aerosol - 6 scenarios	- Baseline Volcanic Aerosol index (NASA GISS) - Pinatubo peak scaled to El Chichón peak - Pinatubo peak 2.5 times the El Chichón peak - Pinatubo peak 5 times the El Chichón peak - El Chichón peak shifted one year back compared to Pinatubo peak - Pinatubo peak shifted one year back compared to El Chichón peak
Ozone record - 8 scenarios	- Sep-Oct-Nov average ozone (baseline) - Sep-Oct average ozone - Sep average ozone - Oct average ozone - 7 Sep – 13 Oct average ozone - Very short 21-30 Sep average ozone - Very long 19 Jul – 1 Dec average ozone - “Worst” 30-day average ozone.

Table 2. Summary of the uncertainties for the proxies discussed in section 2.1 to 2.9 and their inclusion in the regression analysis in this study.

	Kuttippurath et al. [2013]		This study	
Period	EESC	PWLT	EESC	PWLT
1979-1999	-4.50 ± 0.65	-5.02 ± 1.11	-5.39 ± 0.97~~22~~	-5.66 ± 1.03
2000-2010	1.11 ± 0.16	2.91 ± 2.73	1.04 ± 0.19~~2~~	3.30 ± 2.85
1979-1999			-5.26 ± 1.02 ~~0.21~~	-5.75 ± 1.00
2000-2012			1.09 ± 0.23~~10~~	3.28 ± 2.49

Table 3. EESC-based Antarctic vortex core ozone trends and their 2σ trend uncertainties (DU/year) derived from ~~multi-~~multivariate linear regression. The trends in ozone based on EESC regression are calculated by an Ordinary Linear Regression based of the pre-defined change in EESC multiplied with the EESC regression coefficient for the time period under consideration [*cf.* Kuttippurath et al., 2013]. The EESC trend is in pptv/year, the EESC regression coefficient is in DU/pptv, hence the trend in ozone is in DU/year, allowing direct comparison with the PWLT ozone trends (also in DU/year)

	EP- flux	EESC	AEROSOL	SF×QBO
SAM	-0.31 ± 0.27	-0.03 ± 0.17	-0.09 ± 0.19	-0.09 ± 0.29
SF×QBO	0.08 ± 0.28	0.07 ± 0.42	-0.02 ± 0.19	
AEROSOL	0.05 ± 0.17	0.03 ± 0.30		
EESC	0.25 ± 0.17			

Table 4. Cross correlations and their 2σ variance between explanatory variables.

Ozone \ EP flux	Aug - Sep	Jul - Aug	Jul - Sep	Jul	Aug	Sep	45°S-75°S	100 hPa
Sep - Nov	27.7	16.9	43.5	2.3	18.2	2.6	84.9	70.7
Sep - Oct	**98.5**	80.7	**99.7**	37.5	71.5	71.2	**100.0**	**100.0**
Sep	34.8	23.1	41.9	5.0	23.0	15.0	60.8	60.4
Oct	**99.4**	72.5	**99.2**	35.2	63.7	77.6	**100.0**	**99.9**
21 - 30 Sep	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	1.0	1.9
7 Sep – 13 Oct	54.3	19.5	55.2	5.3	17.8	24.4	92.1	90.7
Worst 30 days	87.3	52.9	94.2	18.8	42.2	53.6	**99.6**	**99.1**
19 Jul – 1 Dec	30.1	21.8	36.3	4.6	27.4	5.3	78.6	68.2

Table 5. Percentage of statistically significant regressions for each combination of ozone and EP flux scenarios, as defined in section 2, based on the PWLT regression model for ~~the all~~ break years 1997, 1998, and 1999, and all ending in 2010. Each ensemble consists of results of 180,000 single regressions (6 volcanic aerosol scenarios, 100 SAM and 100 QBO-solar index Monte Carlo runs, 3 break years). Numbers in bold are statistically significant > 95%.

Ozone \ EP flux	Aug - Sep	Jul - Aug	Jul - Sep	Jul	Aug	Sep	45°S-75°S	100 hPa
Sep - Nov	**99.9**	10.7	92.6	0.1	36.3	29.5	**100.0**	**100.0**
Sep - Oct	**100.0**	52.2	**100.0**	4.2	73.4	**100.0**	**100.0**	**100.0**
Sep	**100.0**	40.3	**99.5**	2.0	67.5	96.4	**100.0**	**100.0**
Oct	**100.0**	12.1	**98.0**	0.6	27.2	**98.1**	**100.0**	**100.0**
21 - 30 Sep	0.1	<0.1	<0.1	<0.1	<0.1	<0.1	2.0	20.9
7 Sep – 13 Oct	**100.0**	10.0	**97.7**	0.8	19.6	**98.4**	**100.0**	**100.0**
Worst 30 days	**100.0**	20.0	**99.4**	1.2	29.6	**99.7**	**100.0**	**100.0**
19 Jul – 1 Dec	**99.9**	25.6	**95.3**	1.5	66.1	56.5	**100.0**	**100.0**

Table 6. As table 5 but for the break year 1997 and the period ending in 2012.

Start year	End year	Length (years)	significant trends
2000	2010	11	34.3%
1999	2010	12	47.8%
1998	2010	13	59.5%
all	2010		47.3%
2000	2012	13	39.0%
1999	2012	14	52.7%
1998	2012	15	60.7%
all	2012		50.5%

Table 7. Fraction of statistically significant trends (%) in all regression results for different break years, period lengths and different types of trend calculations. The start year and end year refer to the time period for which trends are calculated. The "all" start years refers to the statistics for all three start years scenarios combined.

Supplementary information

QBO.

The Quasi-Biennial Oscillation (QBO) of the winds in the equatorial stratosphere has been discovered in the 1950s through the establishment of a global, regularly measuring radiosonde network [Graystone, 1959; Ebdon, 1960]). The Free University of Berlin has compiled a long-term record from 1953 onwards of daily wind observations of selected stations near the equator. From these daily measurements monthly mean zonal components were calculated for pressure levels of 70, 50, 40, 30, 20, 15, and 10 hPa. For the period after 1979 only measurements from Singapore are used. The QBO data set is supposed to be representative of the equatorial belt since various studies have shown that longitudinal differences in the phase of the QBO are small [Hood, 1997]. It should be noted, however, that some uncertainties arose at higher levels during the early years from the scarcity of observations. More information on the original data and their evaluation can be found in Naujokat [1986].

As proxy for the regressions we will use the 40-hPa QBO index, also used in Kuttippurath et al. [2013]. Salby et al. [2011, 2012] chose to use 30-hPa winds instead. Typical maximum zonal wind speeds are 20-30 m/s. ~~-~~The relevance~~y~~ of the choice of QBO index will be evaluated later. Information on the uncertainties in the monthly QBO data is not available. One indirect method to estimate the uncertainties is by examining QBO index variability close to the maximum and minimum of the QBO cycles, where the QBO index values remains more or less constant for some months. Assuming that during the maximum or minimum in the QBO phase variations from month to month are indicative of uncertainties in the QBO, we come up with estimated uncertainties of around 1.5-2.0 m/s in the zonal mean wind speeds.

Solar flux

Variations in incoming solar radiation – in particular the shorter ultraviolet wavelengths – have an effect on stratospheric ozone [Haigh, 1996; McKormack and Hood, 1996; Soukharev and Hood, 2006; Anet et al., 2013]. A standard proxy for variations in incoming solar radiation in ozone regression studies is to use the monthly mean 10.7 cm radio flux, as also used in Kuttippurath et al. [2013]. This data set was obtained via NOAA/NESDIS/NGDC/STP.

However, there are other solar activity proxies available. Ideally, in absence of true UV spectral measurements, one would like to use a proxy that is representative for solar activity at those wavelengths where stratospheric ozone formation occurs, which is of roughly between 200 and 300 nm. Dudok de Wit et al. [2009] tried to identify the best proxy for solar UV irradiance, and concluded that proxies derived from a certain wavelength range best represent the irradiance variations in that wavelength band. Thus, the 10.7-cm radio flux might not fully represent solar UV variability. Using the results from Dudok and de Wit et al. [2009] to analyze a set of seven solar activity proxies dating back to at least 1979 based on the solar2000 model and obtained from NOAA/NESDIS/NGDC/STP (F10.7, Lyman-alpha, E10.7, and the solar constant S), we will assume in our regressions that the uncertainty range associated with the solar proxy is approximately 15% of the root-mean-square of the anomaly values.

Why do standard errors of an ordinary linear regression relative to the regression slope not depend on the actual regression itself?

This analysis is based on the "Data Analysis Toolkit" document (chapter 10), written by Prof. James Kircher, Professor of Earth and Planetary Science at the University of California, Berkley and emeritus Goldman Distinguished Professor for the Physical Sciences.

http://seismo.berkeley.edu/~kirchner/

The standard error of the regression slope **b** of an ordinary linear regression of two variables **x** and **y,** and the regression slope **b** itself can be written as:

$$s_b = \frac{b}{\sqrt{n-2}}\sqrt{\frac{1}{r^2}-1} \qquad and \qquad b = r\frac{S_y}{S_x} \tag{S1}$$

In which $\mathbf{s_b}$ is the standard error of the regression slope, **n** the number of data points of the variables **x** and **y, r** is the Pearson correlation coefficient between the variables **x** and **y**, and $S_{x,y}$ is the standard deviation of the variables **x** and **y**.

For a statistically significant trend one generally defines that the trends (slopes) should exceed two times the standard error. Or, in other words, the standard error of the regression slope divided by the regression slope itself should be less than 0.5

The standard error of the regression slope relative to the regression slope itself – which directly relates to statistical significance of the trend - becomes, based on the equation above:

$$s_b / b = \frac{1}{\sqrt{n-2}}\sqrt{\frac{1}{r^2}-1} \qquad \text{(S2)}$$

which only depends on the correlation between the variables **x** and **y** and the number of data points of variable **x** and **y** (record length).

Figures

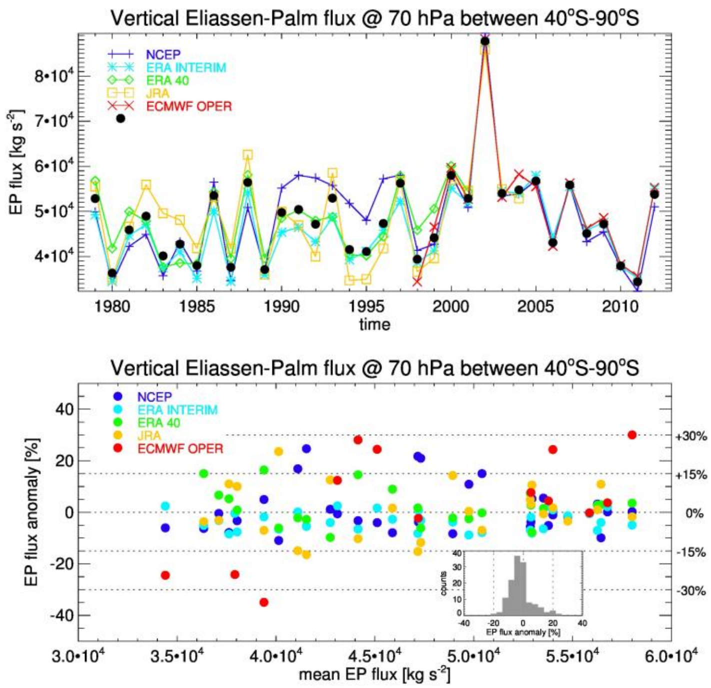


Figure 1. Vertical Eliassen-Palm (EP; kg/s^2) flux at 70 hPa between 40ºS and 90ºS for five different meteorological datasets for the period 1979-2012 averaged for the two month period August-September: NCEP reanalysis 1979-2012, ECMWF ERA INTERIM 1979-2012, ECMWF ERA 40 1979-2001, Japan Reanalysis 1979-2005, ECMWF operational analysis 1998-2012. Top panel shows the EP flux as function of time, including the mean EP flux for each year based on all datasets. Bottom panel shows the EP flux anomalies (%) of a given year as function of the mean EP flux (black dots in the upper panel) for all meteorological datasets available for that year. The insert shows the probability distribution of the relative anomalies. Data are obtained from the EP flux data website of the Alfred Wegener Institute (AWI) for Polar and Marine Research in Bremerhaven, Germany.

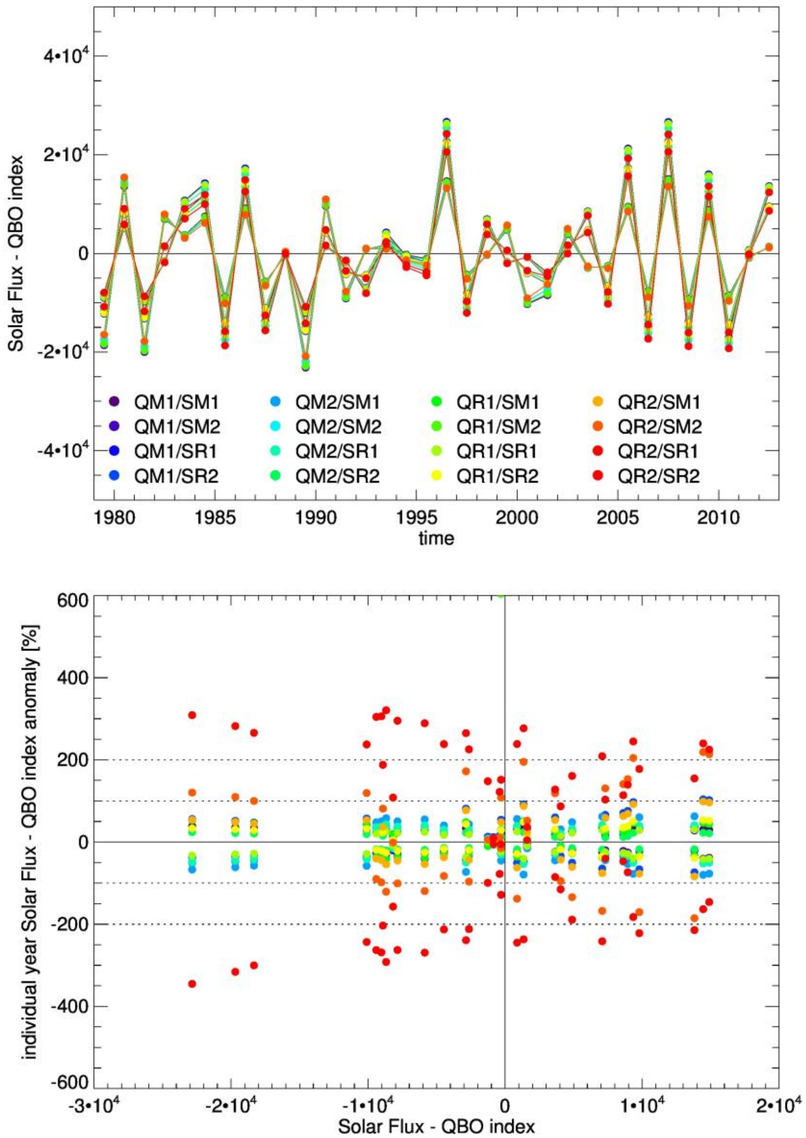
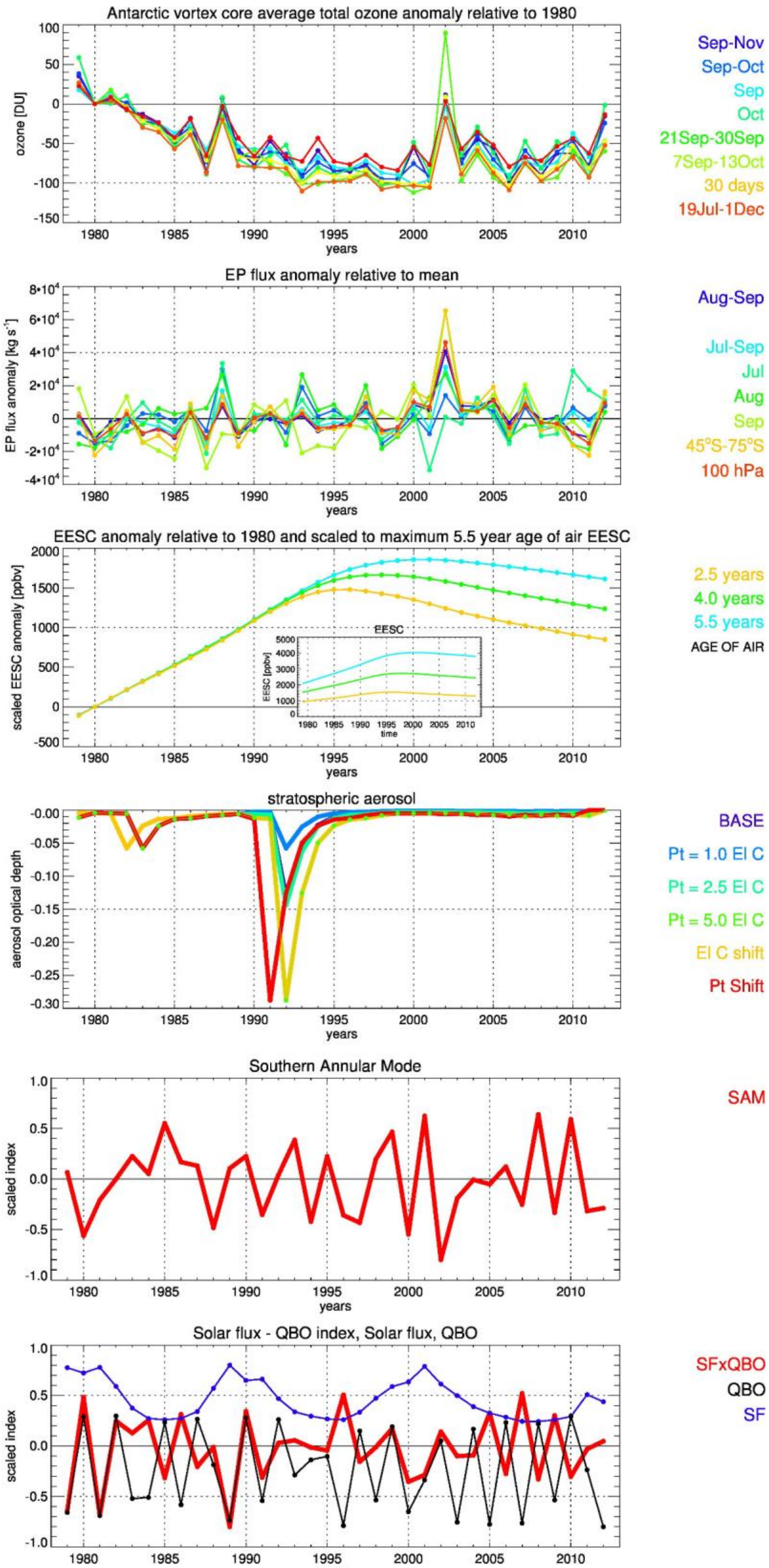


Figure 2. Time series of the combined Solar flux- QBO index (arbitrary units) (upper plot) and the index anomalies relative to the average of different possibilities to derive at the index. The Solar flux ("S") and QBO ("Q") anomalies were calculated based both on the average ("M") as well as the range of Solar flux and QBO values ("R", see section 2.5 for the explanations of the "range"), and for both the entire record of Solar flux and QBO values (1947-2012 and 1953-2012, respectively; "1") as well as for the period 1979-2012 ("2"), resulting in a total of 16 combinations. The different colors denote the different combinations.



Antarctic vortex core average total ozone anomaly relative to 1980
ozone [DU]
years
Sep-Nov
Sep-Oct
Sep
Oct
21Sep-30Sep
7Sep-13Oct
30 days
19Jul-1Dec
EP flux anomaly relative to mean
EP flux anomaly [kg s^-1]
years
Aug-Sep
Jul-Sep
Jul
Aug
Sep
45°S-75°S
100 hPa
EESC anomaly relative to 1980 and scaled to maximum 5.5 year age of air EESC
scaled EESC anomaly [ppbv]
years
EESC
EESC [ppbv]
time
2.5 years
4.0 years
5.5 years
AGE OF AIR
stratospheric aerosol
aerosol optical depth
years
BASE
Pt = 1.0 El C
Pt = 2.5 El C
Pt = 5.0 El C
El C shift
Pt Shift
Southern Annular Mode
scaled index
years
SAM
Solar flux - QBO index, Solar flux, QBO
scaled index
years
SFxQBO
QBO
SF

Figure 3. Time series of regressors for the period 1979-2012. For ozone, EP flux, EESC and stratospheric aerosol all scenarios as defined in section 2 are included (indicated by the different colors). For the SAM and the solar flux - QBO index only the baseline time series is shown, and both indices – being unitless to start with - are scaled for proper comparison. Ozone values are in DU, EP fluxes are in kg/s, EESC values are in ppbv and stratospheric aerosol is in optical depth.

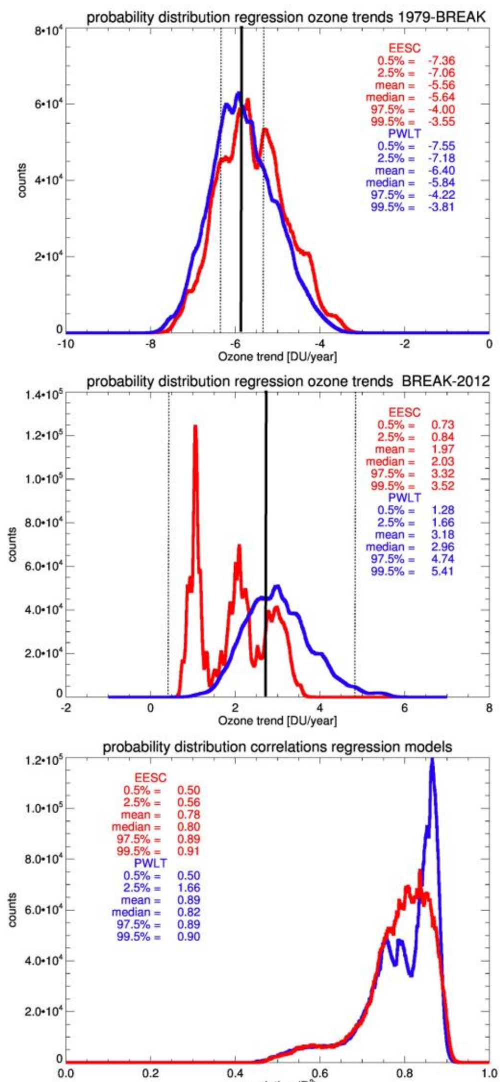


Figure 4. Probability distribution of ozone trends for the period 1979-break (upper plot) and break-2012 (middle plot) as well as time correlations (R^2) for the regression models

and the ozone record scenarios (lower plot). The colors indicate the distributions for the two different long-term ozone regressions (EESC, PWLT). Indicated in the figure are also the 0.5%-2.5%-mean-median-97.5%-99.5% probability values of trends and correlations. The vertical black lines in the upper two panels indicate the trend (solid) and 2σ errors (dotted) of the PWLT regression results of table 2 for the period 2000-2012.

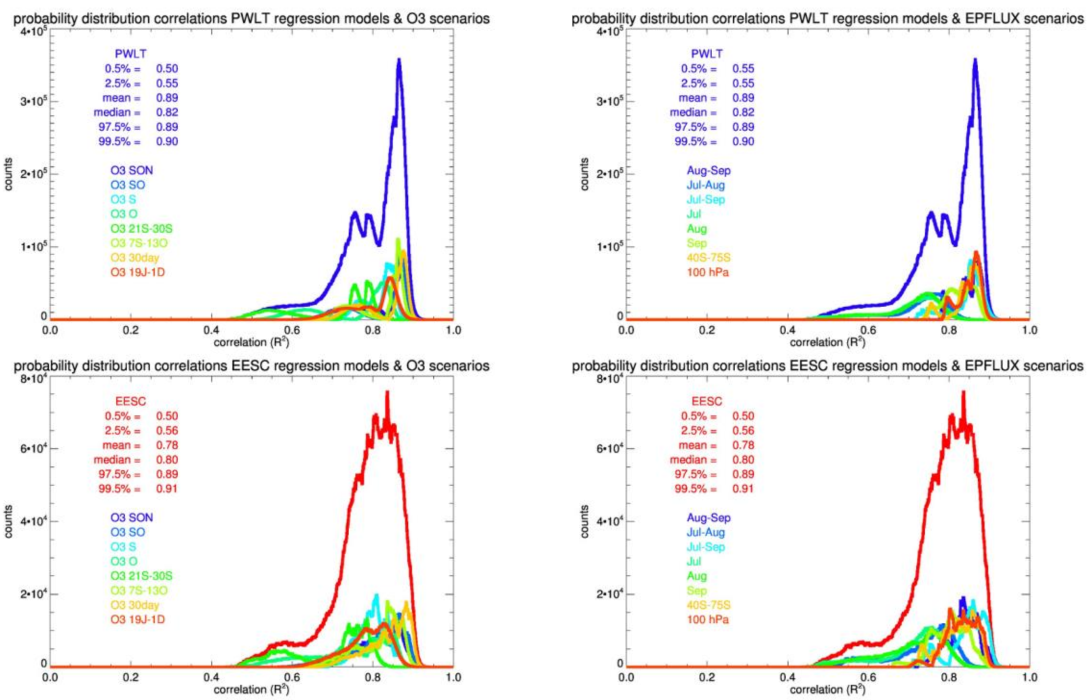


Figure 5. Probability distribution of regression model – ozone scenario correlations as Figure 4, lower plot, for the PWLT and EESC regression model and sensitivity to the different ozone scenarios and different EP flux scenarios, indicated by the different colors. The blue and red outlines show the sum of all scenarios combined.

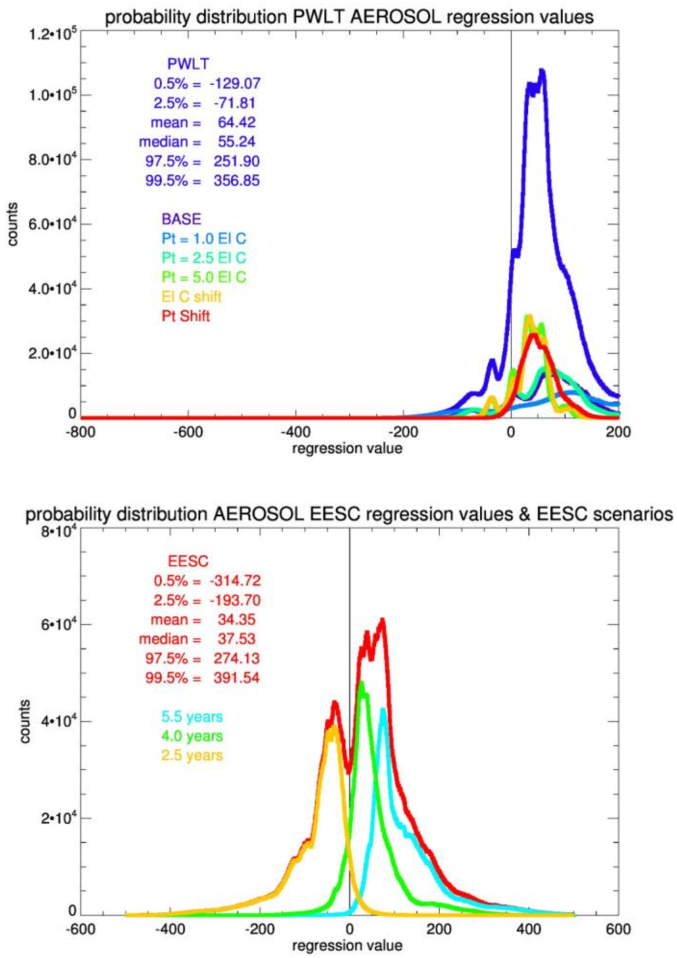


Figure 6. Upper panel: probability distribution of aerosol scenario regression coefficient values of all PWLT regression results. Indicated in the figure are also the 0.5%-2.5%-mean-median-97.5%-99.5% probability values of trends and correlations. Included are also the distributions for the different stratospheric aerosol scenarios, indicated by the different colors. Lower panel: probability distribution of the aerosol regression coefficient values of the EESC regression model results. Included are also the distributions for the three different EESC age of air scenarios, indicated by the different colors. The blue and red outlines show the sum of all scenarios combined.

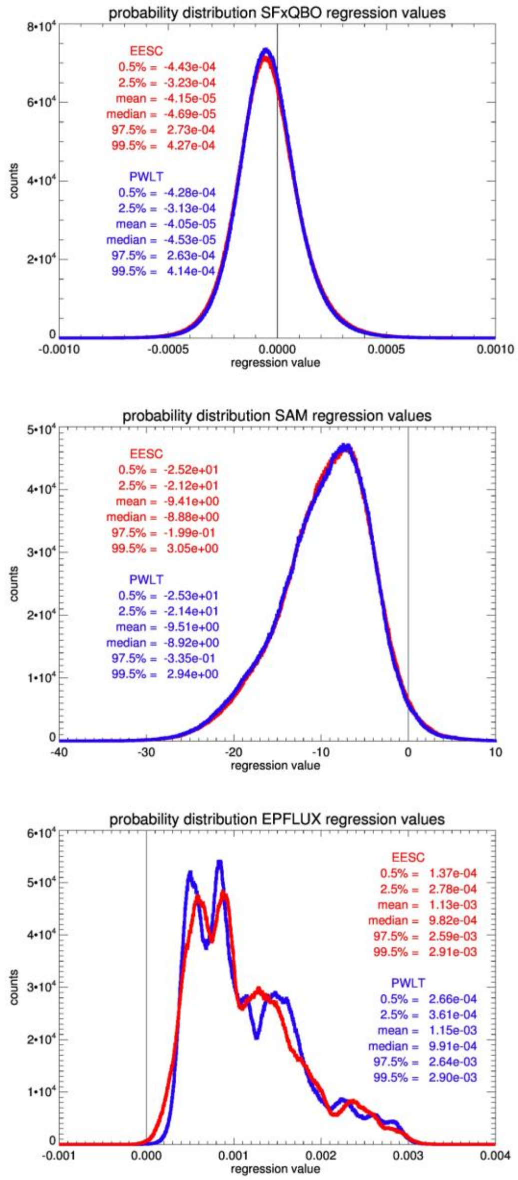


Figure 7. Panel A: probability distribution of the solar flux – QBO index regression coefficient values of all EESC and PWLT regression model results. Panel B: probability distribution of the SAM index regression coefficient values of all EESC and PWLT regression model results. Panel C: probability distribution of the EP flux regression coefficient values of all EESC and PWLT regression model results. Indicated in the figure are also the 0.5%-2.5%-mean-median-97.5%-99.5% probability values of trends and correlations.

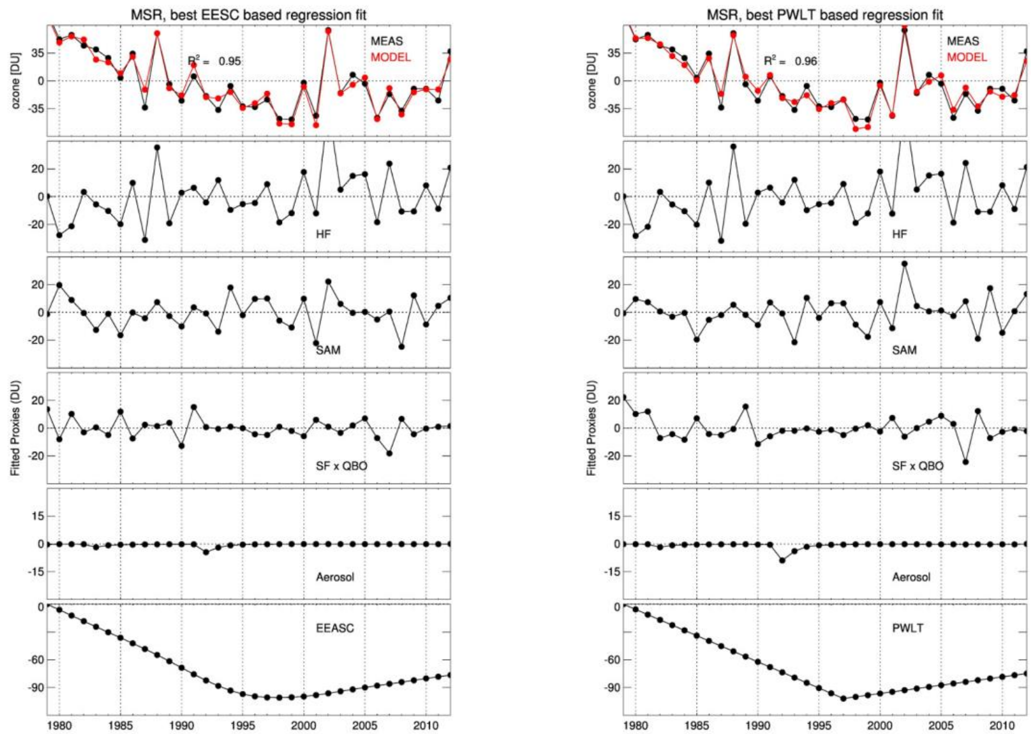


Figure 8. Optimal regression model result for the EESC and PWLT and regressions (upper panels, red line) as well as the corresponding ozone record scenario (upper panel, black line). The ozone variations attributable to each are also shown. Ozone and ozone anomalies are given in DU.

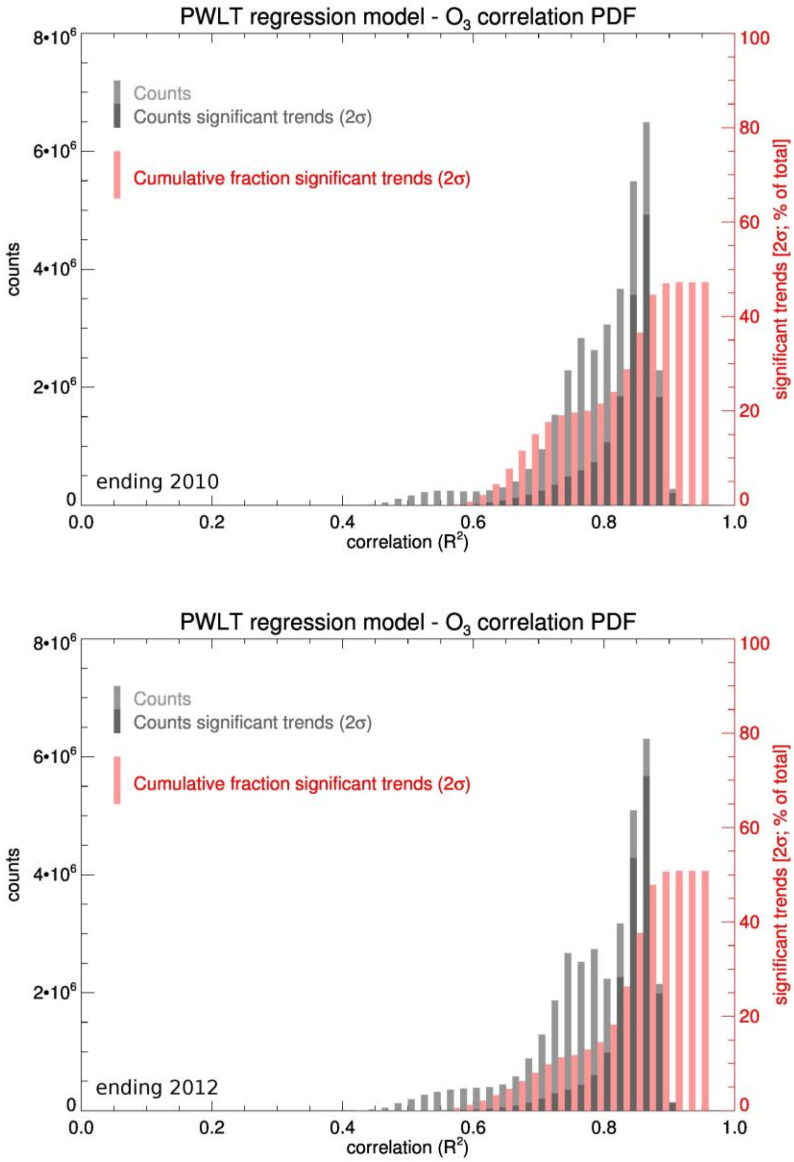


Figure 9. The probability distribution of regression model – ozone record scenario correlations (R^2) as shown in Figure 5 for the PWLT regressions and the cumulative fraction of statistically significant (2σ) ozone trends for each correlation interval (red, right axis). The upper panel shows the distribution for the regressions ending in 2010, the lower panel for the regressions ending in 2012. See also Table 7.