Anonymous Referee #1

General comments:

Based on simulations by the UM-UKCA chemistry climate model, the authors investigate past and current trends in total column ozone, the years when the expected increases in total ozone might become significant, and the years when total ozone might return to 1980s levels. These questions are relevant for the expected recovery of the ozone layer, and for checking the success of the international Montreal Protocol protecting the ozone layer. The used data and methods appear solid. The paper is generally clear, concise and well written. What I am missing, however, is a more thorough comparison with observed trends, and with existing literature. I am also missing a more detailed explanation, how uncertainties were estimated. I think all this can be fixed in a revised version. After addressing my comments below, the paper should be acceptable for publication in ACP.

We thank the referee for their positive and detailed comments. Our detailed response is given below in bold

Major comments

1.) I think it is absolutely necessary to compare the simulated trends and their uncertainties with trends from observations. In my Fig. 1 below, I have overlaid observed trends from Fig. 7 of Weber et al. (2017) onto the simulated trends from Fig. 2 of the current manuscript. Note that the trends from Weber et al. are in % per decade, so at higher latitudes they tend to appear smaller compared to the DU per year trends of the current manuscript. Nevertheless, the observed trends appear to be much smaller at mid-latitudes and in both hemispheres. The uncertainties, on the other hand seem to be quite comparable. A comparison like that would be extremely valuable. I urge the authors to add such a comparison to their paper. Preferably this would be in an additional Figure, comparing simulated and observed trends (e.g. from Weber et al. 2017), as well as their uncertainties, and using the same units (DU per time or % per time). If this comparison confirms the impression from my Fig. 1 below, and the observed trends at mid-latitudes are indeed much smaller, this would be an important finding. Such an apparent lack of significant ozone increases at mid-latitudes would question our expectations for ozone recovery (see also Ball et al. 2017).

We have added reference to recent studies (e.g. Pawson et al., 2014; Weber et al., 2018) throughout the manuscript. Discrepancies between observed and modelled trends in the mid latitudes is indeed an important concern, as highlighted by the reviewer, and one we explore in the revised version of the manuscript. Comparison of modelled and observed trends is, of

course, very complicated. The reviewer has overlaid our modelled trends (in DU/yr) with the trends calculated by Weber et al. (2018) from the NASA dataset (in %/decade). One should note the difference in calculated trends between the datasets presented in Fig 7 of Weber et al. Trends in the NASA dataset are generally smaller and less statistically significant than those of, for example, the WOUDC and GSG datasets. Further, Weber et al. state that the trends they calculate are roughly half of those presented by Pawson et al. in the latest ozone assessment (although still within the uncertainties of these trends), and attribute this to low total column ozone values at the end of their datasets. These low ozone values are stated to be within the expected interanual variability, and therefore part of the system noise rather than reflect a change in trend/processes. While our figure 1 shows that the model in general accurately reflects the observed interannual variability (e.g. by comparison with the Bodeker dataset), there is no reason to expect a free running model to reflect individual anomalies in the noise (particularly when using ensembles where these features would be averaged out between high and low ensemble members when calculating the mean trend) Therefore our trends are likely greater than those of Weber et al. as we do not model these low ozone years at the end of the record. We feel that while this does not reflect some shortcoming of the model, or some discrepancy of total column ozone trends, it could equally well result from the inability of a model projection to accurately predict values of total column ozone for individual years even if the projected trend and interannual variability are well captured.

Converting our modelled trends into %/decade, we get the following trends between 60S-60N:

latitude	-60	-50	-40	-30	-20	-10	0	10	20	30	40	50
DU/yr	0.67	0.45	0.30	0.18	0.12	0.09	0.10	0.17	0.27	0.42	0.56	0.62
%/dec	1.98	1.39	1.01	0.65	0.45	0.34	0.37	0.62	0.94	1.35	1.63	1.68

These trends are in general larger than those presented by Weber et al., but consistent with those of Chehade et al. (2012) and Pawson et al. (2014) within trend uncertainties of both studies. This leads us to conclude, as Weber et al. also concluded, that the smaller trends presented by their study in comparison to earlier observations studies and now to our model study is the result of recent years with low annual means compared to previous years that result from interannual variability. While it is possible that some process that is not captured by the model is responsible for the discrepancy between our modelled trends and those of Weber et al. (most likely transport related, e.g. Ball et al., 2017), there are not enough years of observation to state definitively that this discrepancy does not relate to interannual variability and low column ozone values at the end of the observation record not present in the study of Pawson et al.

This discussion has been added to the manuscript in section 5 (discussion of regional trends) and section 7 (discussion and conclusions). Please also note that direct comparison of our model data with the results of Weber et al is further complicated by differences in the MLR analysis performed in the two studies and the need when comparing modelled and observed quantities to compare collocated data points. As such, we feel it is beyond the scope of this paper (and probably misleading) to produce a figure with both datasets on and to undertake the analysis this properly requires. We do however take the reviewer's point and, as requested, significant changes have been made to the manuscript to discuss the comparison of our data with the studies of Weber et al., and Pawson et al.

2.) A similar comment applies to the recovery detection years, where the current results need to be put more into the context of existing literature. My Figs. 2 and 3 below, for example, compare recovery trend magnitudes and detection years from this study (Figs. 2 and 3) with those from Figure 3 of Weatherhead et al. (2000). Weatherhead et al., from their 2 D model, find trends that are only about half the size of trends in the current study, and also find detection times that are about twice as long as in the current study. In my Figs. 2 and 3 below, I have scaled the Weatherhead et al. results to account for that (see also Eq. 2 of Weatherhead et al. 2000). The comparison in my Fig. 3 indicates that the expected detection years in the current study are generally earlier than in the Weatherhead et al. study, particularly in the tropics, but also at Northern mid-latitudes. Clearly the magnitude of the expected trend plays a large role, especially when trends go to zero (tropics). I think this needs to be brought out much clearer in the current manuscript.

Total column ozone trends present here are around a factor of 2 larger than those of Weatherhead et al. (2000). This most likely is a result of the inclusion of CO_2 induced cooling of the stratosphere in our simulations (and the resulting increases in O_3), while the simulation performed by Weatherhead et al. using the GSFC 2-D model "assumes no direct temperature changes due to greenhouse gas emissions." Many studies (e.g. from Haigh and Pyle, 1982 onwards and including recent studies by Fleming et al., 2011; WMO, 2014; Butler et al., 2016, Keeble et al., 2017) have shown that CO_2 induced cooling of the stratosphere contributes significantly to projected future O_3 increases at a range of latitudes. Accounting for this difference, we feel that the key finding is consistent between the two studies – significant recovery trends will first be identified in the mid latitudes due to the relatively large trends and small interannual variability in these regions. There is an offset in when trends are expected to become significant, with our modelled projected detection years earlier than those of Weatherhead et al., particularly in the tropics. This is unsurprising, as tropical column ozone recovery is dependent on declining CFCs, increasing GHG and changing BDC speeds (see e.g. Eyring et al., 2013; Meul et al., 2016; Keeble et al., 2017), with increases to the BDC offsetting

ozone recovery in the lower stratosphere and resulting in smaller recovery trends. However, the major cause for this difference likely comes from the large number of data points when using the ensemble of 7 simulations to determine significance. We calculate the trend magnitude from January 2000 to month x for each ensemble member, and then increment x by one until trends have been calculated for all months from 2000 to 2080. Trend significance is calculated using data from all 7 ensemble members so that it is comparable with trend significances calculated for figure 2. Then an array of months is constructed for trends which are positive and significant at the 95% confidence level. Initial recovery is defined as the first time a significant trend can be identified (excluding the first 36 months (i.e. up to January 2003) to prevent identifying significant trends when too few data points are present). The error bars on our figure 3 represent the 95% confidence interval (~2sigma) of the range of these 7 values. In this way we explore the impacts of the unexplained noise term on trend detection. In comparison, Weatherhead et al. calculate the trend and trend uncertainty for modelled column ozone values for the years 2000-2020 from a single model integration, and then use these terms in their equation 2 to calculate the number of months required. As a result there are likely large differences in the trend significance term, and hence detection dates. It should be noted that no individual ensemble member shows significant trends at any latitude by 2017 when trend significance is determined using data from only one ensemble member.

The discussion of figure 3 has been modified to reflect these points (see section 5: regional trends), and at the end of the section the following text has been added comparing our results with the Weatherhead results:

"Once the difference in projected recovery trends is accounted for, these findings are consistent with those of Weatherhead et al. (2000), who also identified the midlatitudes as the best location to identify early signs of ozone recovery. However, there is an offset in when trends are expected to become significant between the two studies, with projected detection years modelled in this study generally occurring earlier than those of Weatherhead et al. (2000), particularly in the tropics. This discrepancy in the tropics is unsurprising, as recovery of the tropical ozone column is dependent on the competing influences of declining CFCs, decreasing stratospheric temperatures and changing BDC speeds (see e.g. Eyring et al., 2013; Meul et al., 2016; Keeble et al., 2017), with increases to the BDC offsetting ozone recovery in the lower stratosphere and resulting in smaller column ozone recovery trends. There is poor agreement in modelled projections of future BDC speeds, and as a result projections of tropical column ozone differ significantly between models (e.g. WMO, 2011)."

3.) What also needs to be discussed more is one of the main messages of Fig. 3 of the current manuscript. Essentially, this figure says that by 2017 significant ozone recovery should have been detected between 15_ and 50_ latitude and in both hemispheres. In reality, however, I don't think that is the case, e.g. Weber et al. 2017. So what is going wrong? Is the model too optimistic? Are the uncertainty bars too small? Are the observations too bad? Is the atmosphere not doing what it is supposed to? I think these questions need to be discussed more, and could really be key points of the paper. Just pointing out the large sample size of the simulations (e.g. page 8 lines 16, 17) is not enough. Certainly, to be meaningful, these results need to be translated into something that is observable in the real world.

The early detection of trends in figure 3 is attributable to the large number of points available when using and ensemble of integrations. For figure 3, trend significance is calculated using the trend uncertainty obtained when all ensemble members are used but the trend magnitude is calculated for each ensemble member individually, to provide a range of dates that trend significance can be identified. This range of dates reflects the impact of unaccounted for noise on the trend magnitude, while maintaining significance thresholds which are consistent between figures 2 and 3. It should be noted that no individual ensemble member shows statistically significant recovery by 2017, and only when considering data from all the ensembles together when calculating trends uncertainties can significant trends be identified in the mid latitudes. This can most easily be seen when considering the trends shown for figure 2. Trend uncertainties are calculated using the equation from Weatherhead et al. (1998), in which the standard deviation of the uncertainty in the linear trend is calculated by:

$$\boldsymbol{\sigma_{trend}} = \frac{\boldsymbol{\sigma_{data}}}{\boldsymbol{n^{3/2}}}\sqrt{\frac{\mathbf{1}+\boldsymbol{\varphi}}{\mathbf{1}-\boldsymbol{\varphi}}}$$

When using all ensemble members together, n=17x12x7=1428, while for each individual ensemble member n=17x12=204, while the $\boldsymbol{\sigma_{data}}$ term is roughly comparable between each ensemble member and when the full ensemble is used. As a result, trend significance can be determined earlier than when using observations/single member CCM integrations.

4.) In this general context, I am surprised about the small uncertainty of the detection years in the tropics in the authors' Fig. 3. Since the uncertainty of the trends in the authors' Fig. 2 includes zero, no trend is a possiblity, and detection of a significant trend would take forever. Why is that not reflected in the small tropical error bars in Fig. 3? Compare also the (much more realistic) large spread between the blue and red data points in the tropics in Fig. 4, or the late tropical detection years in Weatherhead et al. (2000).

In our response to the points above we have detailed how we calculate trend uncertainty and identify significant recovery trends. The small uncertainty in trend detection results primarily from the use of a large ensemble of model runs, which reduces the trend uncertainty resulting from interannual variability not accounted for in the MLR. As discussed above, we use data from all 7 ensemble members to calculate trend uncertainties, and the spread in dates obtained for figure 3 results only from the difference in trend magnitude between each ensemble member, which is not significantly different in the tropics as interannual variability is low here in comparison to the high latitudes. Interpretation of these trends in the tropics is hard as the change in column ozone resulting from CFC changes is small in comparison to interannual variability. Further, recovery trends in the tropics are also driven by CO2 induced cooling of the upper stratosphere and changes to the BDC (see, e.g., Keeble et al., 2017 who analyse partial column ozone changes in this ensemble). These factors contribute at least as strongly to column ozone trends as chlorine catalysed ozone depletion.

5.) Generally, I am missing clear explanations, how the error bars where obtained in Figs. 2 to 4. See my detailed comments below for specifics

We have endeavoured to add more detail about how error bars are calculated for all figures. Please see our responses to the detailed comments below.

6.) Since the authors have not really presented much information about point (i) the slowing of past ozone decline and the date of minimum column ozone, I suggest to delete this specific point, especially in abstract and conclusions. I agree with the authors statements in Section 4, especially page 5, lines 19, 20: the date of minimum ozone is a poor metric and therefore point (i) should not really be given much attention and should not be mentioned in abstract and conclusions.

We feel that this point should remain in the abstract and conclusions as, firstly, identification of the minimum values is the earliest indication that ozone has stopped declining, and secondly that the timing of minimum values is a further metric by which to assess the models performance. There is good agreement between the simulations presented here and the timing of the minimum ozone values in the Bodeker dataset. These points are addressed in section 4, although briefly because, as the review states, this metric is a poor one for determining anything quantitatively about recovery, and as a result we feel should also be present in the abstract and conclusion.

Detailed comments

Page 1, lines 10, 11: See my major comment 6, above.

Please see discussion above

Page 1, lines 12 to 14: I find this sentence weird and confusing. Of course, all kinds of mistakes can be made. Maybe just drop this sentence, move the "(e.g. solar cycle, QBO, ENSO)" after "natural cycles" in the following sentence, and start that sentence with something like "Our investigations point to the need . . ."

This sentence is intended to reflect the important point that while a recovery trend may be deemed statistically significant for some set of years, when additional years are included the recovery trend may become non-significant. From our data presented in Figure 3 this effect is seen for the recovery trend at almost all latitudes, but is clearest at 60-70N where the trend from 2000-2005 is statistically significant at the 95% confidence level, but the trend from 2000-2010 is not. It is not until ~2014 that a statistically significant trend is modelled that remains significant no matter how many additional years are considered. This effect is arising from interannual variability. Separating initial and robust recovery in this way is important as while it can be calculated easily using model projections, it cannot be known for observed trends as we do not know what the future observed values will be. Therefore, we feel it is an important point to remain in the abstract, but have amended the sentence to prevent confusion. It now reads:

"It is important to note that, while a statistically significant recovery trend could be calculated at a particular point of time, additional years of observations may lead to a reduced significance of trends due to either a decrease in the magnitude of the trend or an increase in interannual variability. This highlights the need to ensure that the impact of natural cycles (e.g. solar cycle, QBO, ENSO) on total ozone is correctly described in statistical models, especially in the tropics where chemical depletion of the column is small."

Page 1, line 17: See my major comment 3, above.

More detail has been added to the manuscript regarding the calculation of trends and their uncertainties in this study and the implications of the results presented here. Please see our response to comment 3 above for detail.

Page 1, line 18: What do you mean by "sizeable"? I think what you really mean is something like the ratio of trend to natural cycle variability, or trend to unexplained variability. Please reword, clarify.

This sentence has been amended to read "The influence of the natural cycles on trend determination is least at latitudes where the trends are sizeable and the ratio of trend magnitude to interannual variability is large."

Page 1, line 22: This is a good statement, but it is in conflict with the small tropical uncertainty bars in Figure 3. See also my major comments 4 and 5.

This sentence has been amended to better reflect the findings of this study. It now reads: "Significant trends cannot be identified by 2017 at the highest latitudes, due to the large interannual variability in the data, nor in the tropics, due to the small trend magnitude, although it is projected that significant trends may be identified in these regions soon thereafter."

Page 1, line 25, "were shown to"; page 5, line 16, "seen to be"; page 7, line 9, "are found to": I suggest to drop such unnecessary wordings, possibly also in other places.

These phrases have been removed from the text

Page 2, line 5: Drop "the difference"?

This 'difference' is here to make clear that interannual variability varies with latitude, and we feel it should remain.

Page 2, line 9: Drop "Solomon et al. 2016". I don't think that paper says much about changing BDC / ozone transports.

This reference has been removed from the discussion of BDC/transport changes.

Page 2, line 20: I agree that the linear assumption is "somewhat simplistic". However, so many studies, including complex CCM studies, have shown that, in the end, the whole system behaves remarkably linear, and that the linear assumption does work very well. So maybe replace "somewhat simplistic" by "surprisingly robust"?

'somewhat simplistic' has been removed so that the sentence now reads "These statistical approaches nearly all work by relying on the assumption of a linear relationship between a proxy variable and its impact on total ozone."

Page 2, line 33: Maybe drop that line, and reword the previous sentence? See my major comment 6.

Please see discussion relevant to major comment 6.

Page 2, line 34: I would add “after accounting for natural variability” after “values”. In fact you say and show later that accounting for these cycles is important. Obviously, unaccounted for ups and downs are prone to misinterpretation. So here, and in other places, little text and attention should be given to those “raw” results.

This has been added to the manuscript

Page 2, lines 38, 39: I would drop “as proxies for atmospheric observations”.

This phrase has been removed

Page 3, line 24: I wonder about the sea-surface and ice conditions. For two reasons: In Fig. 1, the red line appears to be much smoother after 2000, and much more variable from 1960 to 2000. Are your runs using observed sea surface conditions before 2000, and some climatology after 2000? Do missing real surface conditions have something to do with the mismatch between your simulated trends and observed trends e.g. from Weber et al. (2017), see also my supplemented Fig. 1. I think you should clarify this, and also make some statements about the importance of sea surface conditions for these ozone trends. I think there is past work by Braesicke and others on the influence of sea surface conditions on the stratosphere, and probably a lot more to be cited here – ask John Pyle.

SST and sea ice fields for the entire length of each simulation used in this study are taken from an integration preformed with another coupled ocean model (HadGEM2-ES), as discussed in section 2 (p3, l25). As a result, there is no abrupt change in this lower boundary condition, which would result from switching from observed to modelled SSTs. However, the result is that the modelled SSTs differ from observed SSTs, and as a result the timings of ENSO events also differs. While this will affect interannual variability for the raw model values, the MLR analysis includes an ENSO term and as a result this shouldn’t influence the trends derived from the ozone residuals in a significant way. As a result, it is unlikely that differences between the trends of Weber et al. and those presented here result from differences in SSTs.

The fact that the red line appears to be much smoother after 2000, and much more variable from 1960 to 2000 most likely results from the prescription of stratospheric aerosols in the model. Historic stratospheric aerosol loadings are prescribed from 1979 through 2004

(following the SPARC dataset) and outside of these times background aerosol loadings are prescribed. As a result, historic volcanic eruptions (most notably El Chichon in '83 and Pinatubo in '91) are included. While there is a term in the MLR regression for stratospheric aerosol loadings, the effects on ozone of volcanic eruptions is dependent on stratospheric chlorine levels (as we discuss in section 7, p8 l14 and also discussed by Weber et al.). Therefore, the impacts of these eruptions are likely not completely removed, resulting in the relative difference in perceived interannual variability before and after 2000.

Braesicke and Pyle (2004) has been added as a reference to paragraph 1 of the introduction where SSTs are discussed.

Page 3, lines 35, 36: This equation needs a lot more explanation. Are you using monthly means, or what? Are the data deseasonalized?

We use deseasonalized monthly mean data. This point has been added to the manuscript in the paragraph following the first equation.

If the TO3e;l;i is to be meaningful ozone, all the predictors have to be normalized to mean 0, or to 0 under "normal" conditions. This should be stated.

Here TO3e;l;I does not correspond to any particular modelled data point, but instead corresponds to the intercept term of the MLR

What does the subscript i mean in TO3e;l;i? Calendar month?

$TO3_i$ is a constant value of total column ozone, corresponding to the y intercept term of each MLR. To avoid confusion, this term has been replaced by $i_{e,l}$

Certainly the ax need to depend on latitude and be axl. **AND** Is the regression applied to all ensemble runs at once (providing axl), or individually to each run (providing axe;l)?

The MLR analysis is applied to each model run separately, resulting in the 7 light red lines in figure 1. Subscripts for the alpha terms have been added to the regression model

How do you deal with autocorrelation in the Ne;l:t? Autocorrelation can be substantial, e.g. 0:6 in the tropics, (see Plate 3 of Weatherhead et al. 2000). This reduces effective sample size and increases error bars, and needs to be accounted for. In the same direction: How independent are the ensemble runs? In the model world they may be independent, but compared to the real world, they are not really independent samples drawn from a large population.

Autocorrelations is accounted for when calculating confidence intervals for trends using the methodology described Weatherhead et al., 1998, such that:

$$\mathbf{std.dev_{trend} = std.dev_{noise}/n^{3/2}.sqrt((1+phi)/(1-phi))}$$

The 95% confidence interval is then calculated as 2* std.dev$_{\mathbf{trend}}$

This point has been added to the text at the end of section 3.

Regarding the ensemble, it is not accurate to assume that each model integration is independent. Each integration uses the same boundary forcings, differing only in their initial conditions. As such, while this provides some information on projection uncertainty, it likely does not fully capture it, and as a result uncertainty estimates may be a minimum.

Page 4, lines 12 to 14: I would move that sentence much closer to the Equation. I think it is important to understand what is fitted.

This sentence has been moved towards the start of the paragraph, so that it now reads:
"in which the α values are the coefficients returned from the MLR for each explanatory variable (denoted by the superscript) and vary between latitude range and ensemble member. The explanatory variables included in the MLR are the QBO, solar cycle, ENSO, volcanic aerosols and ESC. The subscripts *e*, *l* and *t* indicate that the alpha value or explanatory variable differs with ensemble member, latitude and/or time respectively. For the QBO, two terms are included, QBO_{50} and QBO_{30}, which correspond to equatorial westerly winds at 50 hPa and 30 hPa respectively. Two QBO terms are included to account for the phase shift in the total column ozone response with respect to QBO changes at different altitudes. The solar cycle is represented by the top of atmosphere solar flux, represented in the MLR as $solar_t$. ENSO effects on column ozone are represented by $ENSO_t$, the detrended sea surface temperature anomalies in the NINO3.4 region. Volcanic aerosols are included as hemispheric aerosol optical depths, and so are different for the northern and southern hemispheres to account for the lack of interhemispheric transport of aerosols emitted into the stratosphere from high latitude eruptions. The final term included in the MLR, *ESC*, represents stratospheric chlorine concentrations. This term is equal to the ESC concentration at 30 km for each latitude bin to account for the time taken for ODS to be transported to higher latitudes. Any month to month

variation not accounted for by the explanatory variables in the MLR is represented by the noise term $N_{e,l,t}$. The *Solar$_t$*, *ENSO$_t$* and *aerosol$_t$* terms are all prescribed forcings in the model and do not vary between ensemble members. In this study we use deseasonalised monthly mean total column data, and so there is no need for a seasonal cycle term in the MLR model."

Page 4, line 17: I am missing an explanation how the trends in Fig. 2 are obtained. Presumably for the MLR trends, you fit straight lines to the RO3e;l;t and obtain the trend uncertainty from the fit residuals / remaining noise? Again: Are all ensembles fitted at the same time, or do you fit each ensemble separatately?

Trends for the raw model data and MLR data are obtained by fitting independent linear trends to the modelled total column ozone values (TCO3 on LHS of first equation) and the ozone residuals produced when accounting for natural cycles (RO3 on LHS of second equation). All ensembles are fitted at the same time. This information has been added to the manuscript at the end of section 3 by adding the following paragraph:

MLR analysis and ozone residuals are produced for each individual ensemble member of the raw model data, resulting in seven RO3 time series. For both the raw model data and the ozone residuals, decline and recovery trends are calculated using a independent linear trend fit for the periods 1980-1997 and 2000-2017 respectively. When calculating trends for both the raw model data and ozone residuals, a single linear fit is produced using data from all seven ensemble members rather than producing a fit for each individual ensemble. In order to make any robust conclusions about the statistical significance of modelled trends, some measure of the trend uncertainty is required. Here we calculate trend uncertainties following the methodology of Weatherhead et al. (1998), in which the standard deviation of the uncertainty in the linear trend is calculated by:

$$\boldsymbol{\sigma_{trend} = \frac{\sigma_{data}}{n^{3/2}}\sqrt{\frac{1+\varphi}{1-\varphi}}}$$

where σ_{data} is the standard deviation of the time series in question (either raw model data or RO3), *n* is the number of months in the time series and *φ* is the autocorrelation coefficient for a 1 month lag. As discussed by Weatherhead et al. (1998) autocorrelation can be substantial for monthly mean time series, particularly in low latitudes, and failure to account for it results in an underrepresentation of trend uncertainty. As for the trend calculations, trend uncertainties are calculated using data from all seven ensemble members together, and as such for the 17 year periods considered for the decline and recovery phases *n* is very large (*n*=17x12x7=1428).

Don't the ax need subscripts e; l? How is autocorrelation in the fit residuals dealt with? How do you obtain the raw model trend in Fig. 2? By simply fitting straight lines to the TO3e;l;t on the left side of the first equation? Or piecewise linear trends? Please add text here or later, and answer these questions.

Alpha terms do need e,l subscripts, and these have been added to both equations. Autocorrelation is dealt with following Weatherhead et al. (1998), as detailed above. Independent linear trends are used to obtain trends from the raw model data. This information ha been included in the manuscript at the end of section 3 in the text provided for the comment above.

Page 5, lines 18, 19: See major comment 6. Same paragraph: I think you should also add some arguments based on your Fig. 1, e.g. the large uncertainty range for minimum ozone from 1992 to almost 2010, with little difference between blue and red curves.

The following sentence has been added to the last paragraph of section 4:

"Even outside the high latitudes, where interannual variability in total column ozone values is largest, identification of the year of minimum ozone is uncertain, with each of the seven residual ozone time series having minimum values at different times between 1992-2005 (light red lines, Figure 1)."

Page 5, line 35: 95% confidence intervals – obtained how and from what? Please explain.

Information on how the trend uncertainties are calculated has been added to the end of section 3, and in section 5 we stipulate that the 95% confidence intervals = $2\sigma_{trend}$

Page 5 lines 36, 37: "heterogeneous . . . vortex". Not only that. The Brewer Dobson Circulation also "transports" the large ozone trends from the upper stratosphere polewards and downwards into the lower stratosphere. There, near the ozone maximum, they make a big difference for column ozone (whereas otherwise upper stratospheric ozone does not contribute a lot to the total column). Please reword, or add some text.

This sentence has been expanded to include reference to the effects of BDC transport of ozone anomalies, so that it now reads:

"During the decline phase, ozone trends from both datasets are greatest at high latitudes due to the heterogeneous activation of chlorine on PSCs within the polar vortex and the transport of midlatitude polar ozone depletion signals to high latitudes by the Brewer-Dobson Circulation."

Page 5, lines 38, 39: Please add "declining" or "1980 to 1997" before "trends". Otherwise this is misleading and might be mistaken with the increasing 2000 to 2017 trends.

To avoid confusion, this sentence has been changed to:
"Negative trends in the raw column ozone data from 1980-1997 are significant at all latitudes…"

At some point, you might also want to point out that by picking 1980 to 1997 and 2000 to 2017, you have picked two one-and-a-half solar cycle long periods. This would maximize solar cycle effects on the trends (e.g. solar max at one end, solar mind at the other end). So some of your results might include large solar cycle effects – but still the comparison of raw and MLR trends in Fig. 2 does not look too bad. You do have a corresponding discussion on page 6, lines 9 to 18. However that discussion reads a bit awkward, and, to me, puts too much focus on the "raw trends", which obviously are influenced by the solar cycle and obviously should not be used. Maybe reword that discussion.

This point has been raised when discussing, in section 5, the difference in trend magnitude for the raw and ozone residual trends in figure 2. This discussion has been reworded to clarify the point above and avoid confusion.

Page 6, line 2: maybe add "and Pinatubo aerosol effects"

This sentence has been changed to read:
"This is the result of the eruption of Mt Pinatubo and the pronounced solar minima during the 1990s, both of which resulted in lower column ozone values and so a greater trend from 1980."

Page 6, line 10: add "and autocorrelation" after "variance" and "(Weatherhead et al. 2000)" after "data".

These have been added to the text

Page 6, line 19: Replace "month" by "year"? Also in other places in this paragraph?

The first instance of month has been replaced by year, as this is what is plotted in Fig 3. However, the script used to calculate trend significance here iterates over each successive month of ozone residual data, and so use of months within the text of the manuscript is more accurate than year. The first two sentences of this paragraph have been altered to reflect this.

Page 6, lines 19 to 30: How did you obtain the error bars in Fig. 3? From comparing results of the different runs? Is that realistic? See my major comments 4, 5.

Please see our response to the major comments above for details on how error bars and uncertainties are calculated for figure 3.

Page 7, line 24: add “calendar” after “to”, and replace “certain months” by “e.g., in September”

This sentence has been changes to:
“However, the signature of this recovery is very sensitive to calendar month, and earlier signs of recovery may be identified in certain months (e.g. September; Solomon et al., 2016)”

Page 7, line 35: same as major comment 6.

Please see our response to major comment 6.

Page 8, lines 2 to 5: same as page 1, lines 12 to 14.

As with our response to the earlier comment, this sentence has be changes to avoid confusions. It now reads:
“However, an important caveat is that while a statistically significant recovery trend could be calculated at a particular point of time, additional years of observations may lead to a reduced significance of trend due to either a decrease in the magnitude of the trend or an increase in interannual variability. This highlights the need to ensure that the impacts of natural cycles (e.g. solar cycle, QBO, ENSO) on total ozone are correctly described in the MLR.”

Page 8, lines 12 to 19: What about transport variations? You are not talking / accounting for them at all. See also my comment above about sea surface conditions. I think you should add something here, and also discuss the differences to the observations more, e.g. citing Weber et al. 2017 and Ball et al. 2017. See also my comment above about sample size, and my major comments 2 and 3.

Transport changes of course play a key role in short and long term ozone recovery trends. Better comparison with the observed trends of Weber et al. (2018) and Pawson et al. (2014) has

been added to the manuscript, as well as discussion of the transport changes discussed by Ball et al. (2018). We conclude that any transport changes which have affected the observed column and profile recovery are not captured by the ensemble mean and instead represent a feature of interannual variability and not some fundamental shift in stratospheric circulation driven by, for example, increased GHG concentrations.

Page 8, line 20: Same as Page 1, line 18.

This sentence has been clarified by changing it to:
"The influence of the natural cycles on trend determination is least at latitudes where the trends are sizeable and the ratio of trend magnitude to interannual variability is large."

Page 8, line 23: To me, it is worrying that precisely there Weber et al. 2017 find small and non-significant increases (see also Ball et al. 2017). I think you need to comment more on that, and I think this difference could be a key message from this study. See also my major comment 1.

This is a difficult point to evaluate as different datasets studied by Weber et al. give significant trends in different locations. In general, our results are consistent with the findings of fig 7 in Weber et al., who identify significant positive trends in the northern midlatitudes in some datasets (particularly WOUDC and GSG), although do not see such robust recovery in southern midlatitudes. This likely results from reduced trend uncertainty in our study due to the large number of datapoints provided by the ensemble members, and also the larger trend magnitudes modelled than identified by Weber et al., which they attribute to particularly low column zoone values at the end of the record consistent with interannual variability. Please see our response to comment 1 for further information.

Page 15, Figure 2: In the legend in the Figure. Please replace "Model trend" by "simple trend" or "raw trend". That would be clearer, and the "MLR" trends are "model" trends as well. In the caption, please explain how the error bars where obtained.

Names in the figure legend have been changed to match the figure caption and main text. Extra text has been added to the figure captions for figure 2-4 which provide information on how the error bars were calculated.

Page 16, Figure 3: In the caption, please explain how the error bars where obtained. See also my major comment 4.

Extra text has been added to the figure captions for figure 2-4 which provide information on how the error bars were calculated.

Page 17, Figure 4: Why was that not done for the MLR / residual total ozone as well? Should that not be shown? In the caption, please explain how the error bars where obtained.

The aim of the final section of the paper is to highlight that at some latitudes the absolute ozone column abundance may not recover to its 1980 values, due in part to the fact that the 1980s were years during a solar maximum, and in part to other factors which overwhelm the recovery trend expected by decreasing CFCs (e.g. increasing BDC speeds resulting from increased GHG concentrations). Further, due to the high interannual variability in the Arctic, it may be possible for years late in the 21st century to have very low column ozone abundancies due to the high natural variability in these regions (see e.g. Bednarz et al., 2016). For these reasons it was felt that discussion of the raw modelled column ozone abundancies was more pertinent than the ozone residuals calculated elsewhere in the manuscript. Extra text has been added to the figure captions for figure 2-4 which provide information on how the error bars were calculated.

General comments:

The expected stratospheric ozone recovery from the effect of halogenated ozone depleting substances (ODSs) has received much attention in recent years. Yet detecting the recovery of the ozone layer is complex due to a number of factors, including internal and external variability, that obscure the emerging signal associated with the slow decline in ODSs levels. The manuscript addresses this issue by investigating three stages of ozone recovery. To this end, the authors use total column ozone (TCO) changes based on experiments of the UM-UKCA and multiple linear regression (MLR) analysis. Although models are not perfect (e.g. often show significant disagreement compared to observations), they are a valuable mean to explore ozone changes due to specific factors (i.e. ODSs levels). Overall, the manuscript addresses relevant issues with regard to the evolution of the stratospheric ozone layer and uses appropriate data and methods. The text is technically well written. I have minor specific comments (detailed below), which I hope will help the authors improve the paper. In general, I suggest more detailed description and evaluation, additional comparison with ozone measurements, and further discussion on existing literature. Therefore, the manuscript is recommended for publication after the specific and technical comments are addressed.

Specific comments:

a. In the Introduction section, the authors clearly set out the stratospheric ozone depletion in the last decades associated with man-made emissions of ODSs. Due to international efforts banning the use of these substances, the ozone layer is expected to recover and the study aims to explore different stages. However, significant work has been done on detection and attribution of ozone recovery, hence it would be appropriate (and helpful for the broader audience) to briefly introduce key findings, remaining issues, and link it with the novelty of this work. Moreover, this will help relate and put into context the main findings here later in the manuscript.

Reference to the recent findings of Chehade et al. (2012), Pawson et al. (2014) and Weber et al. (2018) has been added to the introduction section in order to further establish the novelty of this work.

b. In the Model configuration and simulations section (page 3, lines 25–26), the authors explain that the simulations used were performed in support of the CCMI activity, and that are described in more detail in Bednarz et al. (2016) and Keeble et al. (2017). Bednarz et al. (2016) described that the simulations included a future climatological solar cycle since 2009 based on the observed cycle 23, which is not consistent with the description given in the manuscript (page 3, lines 16–17). Please clarify.

This is a mistake here – the details of the solar cycle should be those provided by Bednarz et al. (2016) – historic observed solar forcings are applied until 2009, after which cycle 23 is repeated until the end of the simulation. The manuscript has been corrected to account for this. The MLR and results presented in this study are not affected by this error as we use the correct top of atmosphere solar flux prescribed in the model for this analysis.

c. In the Removing natural cycles section, the text describes a MLR analysis to identify the impacts of natural variability on TCO. Since the results of this study heavily rely on the MLR analysis, I think this section requires more detailed description of the statistical method. In particular, the TO3i and Ne,l,t terms need better description (i.e. "… some constant value" and "Any noise…"). Also, an evaluation of the MLR analysis is important – i.e. How good is it? How much of the model raw data is captured by the MLR and how much "noise" is left? –. The manuscript already includes some references on MLR analysis that may help.

Further detail has been added to section 3 describing the MLR model and its terms. The TO3i term (now changes to I following advice from reviewer 1) corresponds to the intercept term of the MLR. N corresponds to the month to month variations not accounted for by the other explanatory variables. We feel that the MLR does a good job accounting for the natural cycles we seek to remove, as shown by comparing the red and blue lines in figure 1 and discussed in the manuscript in section 4.

d. For the Modelled global column ozone and minimum values section, it may be appropriate a statement about the choice of not including the polar regions in this analysis (Figure 1), since, in other sections and figures these regions are included and also discussed in the last paragraph here (page 5, lines 18–25). In fact, the latter paragraph argues that minimum column ozone values are a poor indicator of ozone recovery by giving examples based on polar regions. Is there any particular reason for not using the latest version (3.3) of the Bodeker Scientific database? The latest version, in addition to include some improvements on the methodology, could be expanded until 2016 in the inset of Fig. 1. Also it would be nice to include the Bodeker Scientific database in the acknowledgements, as recommended on the website.

For figure 1 we present only data from 60S-60N as i) the interannual variability at high latitudes, particularly the Arctic, is very large and identification of longterm changes is more difficult, and ii) high latitude polar ozone depletion is strongly seasonal, and this feature dominates monthly mean time series as the seasonal cycle changes with changing stratospheric

ozone depletion. We have clarified in the discussion about minimum column values that we only consider values between 60S-60N.

An earlier version of the Bodeker Scientific database (v2.8) was used as this dataset includes monthly mean values, as are presented from our model results. We are reticent to use the latest version as it does not, at present, include monthly mean data, and these would have to be calculated from the daily data provided. This requires a number of decisions which would need to be made (e.g. what spatial and temporal coverage is required for a monthly mean datapoint) which are not trivial and may not match those reached by Bodeker Scientific themselves. In this case, any representation of monthly mean time series we produce may not match the final monthly mean dataset provided by Bodeker Scientific in the future, which may lead to confusion. Bodeker Scientific has been added to the acknowledgements.

e. Regional trends section. This section includes very interesting results. However, modelled results in Fig. 2, both “raw” and residual data, could be compared to observed trends. In turn, this may lead to some evaluation/discussion and to put into context these results with existing literature. Nevertheless, there is some discussion (outlook) on TCO trends between 2000–2017 in the Discussion and Conclusions section (page 8, lines 15–19). Error bars representing the 95% confidence interval may need a line or two detailing how these are estimated and whether they account for autocorrelation. Are these confidence intervals calculated in the same way for all analyses?

Additional information on how trend uncertainties are calculated has been added to the manuscript, alongside comparison of our modelled trends with observed trends calculated by Chehade et al. (2012), Pawson et al. (2014) and Weber et al. (2018). Please see our response to review 1 for further information.

f. Return to historic values section. Figure 4 shows that TCO values in the tropics (<30º) reach the “1980 last recovery” between ~2060s–2070s. However, the main text (page 7, lines 12–14) explains that “…, it is the only region in which total column ozone abundances are not greater than their 1980s values by the end of the simulation,…”. Please clarify. Also, is there any particular reason for not showing(using) “ozone residuals” on Fig. 4 as in previous analyses? I understand the study aims to explore ozone recovery addressing natural cycles.

The aim of the final section of the paper is to highlight that at some latitudes the absolute ozone column abundance may not recover to its 1980 values, due in part to the fact that the 1980s were years during a solar maximum, and in part to other factors which overwhelm the recovery trend expected by decreasing CFCs (e.g. increasing BDC speeds resulting from increased GHG

concentrations). Further, due to the high interannual variability in the Arctic, it may be possible for years late in the 21st century to have very low column ozone abundancies due to the high natural variability in these regions (see e.g. Bednarz et al., 2016). For these reasons it was felt that discussion of the raw modelled column ozone abundancies was more pertinent than the ozone residuals calculated elsewhere in the manuscript. Extra text has been added to the figure captions for figure 2-4 which provide information on how the error bars were calculated.

Technical comments:

Page 1, lines 9–10. “This approach…”. The approach or method has not really been introduced. I suggest rephrasing this sentence (e.g. Here internal atmospheric variability… is accounted for by…).

This sentence has been reworded to read “The impacts of modelled internal atmospheric variability are accounted for by applying a multiple linear regression model to modelled total column ozone values, and ozone trend analysis is performed on the resulting ozone residuals.”

Page 1, line 28. Substitute “ODSs” for “ODS” for consistency throughout the text.

Substituted

Page 1, line 34. Randel and Wu (1995) did not explore the effects of Mt Pinatubo eruption on stratospheric ozone.

This reference should be Randel et al., 1995 (Randel, W. J., Wu, F., Russell, J. M., Waters, J. W., and Froidevaux, L.: Ozone and temperature changes in the stratosphere following the eruption of Mount Pinatubo, J. Geophys. Res., 100, 16753–16764, doi:10.1029/95JD01001, 1995.) and has been corrected in the text and reference list.

Page 1, lines 37–38. References order.

References have been reordered.

Page 1, line 38. Delete “,” between “other” and “non-chlorinated”.

Deleted

Page 2, line 2. References order.

References have been reordered.

Page 2, lines 13–15. "Good agreement…" This sentence is a bit confusing, rephrasing maybe?

This sentence has been rewritten to read:

"If good agreement is found between the model and observations when all processes are included, then evidence of ozone recovery due to decreasing stratospheric halogen loadings can be identified by excluding other processes. For example, Solomon et al. (2016) found evidence for healing of the Antarctic ozone layer in September when polar halogen chemistry is included but interannual dynamical variability and volcanic factors are excluded."

Page 2, lines 22–23. "…, data from fully coupled chemistry-climate model…" is a bit misleading since you use imposed SSTs. I would clarify "fully coupled" (chemistry and radiation schemes?).

To avoid confusion, the word fully has been removed so that the sentence now reads "To explore future ozone trends and recovery, data from coupled chemistry-climate model (CCM) simulations are required."

Page 2, line 26. Spell out "SSTs".

SSTs have been defined here before the first use of SST and later in section 2 just "SSTs" is used.

Page 2, line 37. Substitute "-" from "–".

Corrected

Page3, line 23. Could use just "SSTs", as it was introduced before.

See comment above

Page 5, line 33. Substitute "… DU year-1…" for "… TCO (DU year-1)…"? I am aware that "for the column ozone" is mentioned later in the sentence, though it is somehow confusing.

This sentence has been reworded to read "Figure 2 shows total column ozone trends (in DU year^{-1}) obtained from the raw data from the UM-UKCA simulation and the ozone residuals for the decline (1980-1997) and recovery (2000-2017) phases, averaged over 10° latitude bands."

Page 7, lines 9–10. "… 1980s values for the first time (light red)…" should be "blue".

We apologise for the confusion – this figure was replotted a number of times. The text now matches the colours in the figure.

Page 7, line 21. Typo: "airmasses".

Corrected

Page 7, line 26. I would substitute "expected" for "projected" (e.g. acknowledging these are modelled results, which are model and scenario dependent).

This change has been made

Page 8, line 15. Typo: "… Unlike a recent analyses…"

This has been corrected to "In contrast to a recent analysis of total ozone measurements…"

Page 8, line 26. Typo: "… The tropics have too small a trend…"

This sentence has been altered to read:
"The magnitude of the column ozone recovery trend in the tropics is too small in comparison with the natural variability resulting from the solar cycle and the QBO to identify significant trends."

Page 15, line 2; and Figure 2, legend. Please follow same consistency in the naming, both for the figure and the main text.

All figures have bee reproduced so that figure legends, figure caption and the main text use consistent naming conventions.

On ozone trend detection: using coupled chemistry-climate simulations to investigate early signs of total column ozone recovery.

James Keeble[1], Hannah Brown[1], N. Luke Abraham[1,2], Neil R. P. Harris[3], and John A. Pyle[1,2]

[1]University of Cambridge, Department of Chemistry, Cambridge, UK
[2]National Centre for Atmospheric Science, Cambridge, UK
[3]Centre for Environmental and Agricultural Informatics, Cranfield University, Cranfield, UK

Correspondence to: James Keeble (james.keeble@atm.ch.cam.ac.uk)

Abstract. Total column ozone values from an ensemble of UM-UKCA model ~~**runs**~~simulations are examined to investigate different definitions of progress on the road to ozone recovery. ~~**This approach takes into account the**~~ The impacts of modelled internal atmospheric variability ~~**of the**~~are accounted for by applying a multiple linear regression model ~~**in assessing the statistical significance of each definition**~~to modelled total column ozone values, and ozone trend analysis is performed on the resulting ozone residuals. Three definitions of recovery are investigated: (i) a slowed rate of decline and the date of minimum column ozone; (ii) the identification of significant positive trends; and (iii) a return to historic values. A return to past thresholds is the last state to be achieved. ~~**However, while recovery may appear to be robust at a particular point of time, additional years of observations may lead to a reduced significance of trends due to natural variability (e.g. solar cycle, QBO, ENSO). This points to the need to ensure that the impact of**~~ Minimum column ozone values, averaged from 60°S-60°N, occur between 1990 and 1995 for each ensemble member, driven in part by the solar minimum conditions during the 1990s. When natural cycles ~~**on total ozone is correctly described in statistical models, especially**~~are accounted for, identification of the year of minimum ozone in the ~~**tropics where chemical depletion**~~resulting ozone residuals is uncertain, with minimum values for each ensemble member occurring at different times between 1992 and 2000. As a result of this large variability, identification of the date of ~~**the column is small.**~~minimum ozone constitutes a poor measure of ozone recovery. Trends for the 2000-2017 period are positive at most latitudes and are statistically significant in the mid-latitudes in both hemispheres when natural cycles are accounted for. This significance results largely from the large sample size of the multi-member ensemble. ~~**The influence of the natural cycles on trend determination is least at latitudes where the trends are sizeable. Thus, while ozone recovery can be identified in certain months over Antarctica, the mid-latitudes are the best place to identify early recovery as the trends are large compared to the variability. Over the Arctic, total column ozone is too variable for a signal to be easily detected: this arises both from the large dynamical interannual variability and from the large changes in chemical ozone loss from year to year. In the tropics, trends are too small compared to the natural variability to identify any statistical significance.**~~ Significant trends cannot be identified by 2017 at the highest latitudes, due to the large interannual variability in the data, nor in the tropics, due to the small trend magnitude, although it is projected that significant trends may be identified in these regions soon thereafter. While significant positive trends in total column ozone could be identified at all latitudes by ~2030, column ozone values which are lower than the 1980 annual mean can occur in the mid-latitudes until ~2050, and in the tropics and high latitudes deep into the second half of the 21st century.

1 Introduction

2017 ~~marks~~marked the 30th anniversary of the Montreal Protocol, which was implemented to protect the stratospheric ozone layer from the harmful effects of ozone depleting substances (ODS). These gases, mostly inert in the troposphere, ~~were~~

~~shown to~~ breakdown when they reached the stratosphere, with the subsequent products then leading to chemical ozone depletion (e.g. Molina and Rowland, 1974; Stolarski and Cicerone, 1974; Rowland and Molina, 1975). Controls introduced under the Montreal Protocol and its subsequent amendments first slowed the rate of accumulation of these halogenated ~~ODSs~~ODS in the atmosphere, and since the late 1990s their atmospheric concentrations have begun to decline (Newman et al., 2006; Mäder et al., 2010; WMO 2011; 2014). A reduction in equivalent stratospheric chlorine (ESC; Eyring et al., 2007) concentrations should lead to an increase in atmospheric ozone as the strength of the halogen catalyzed ozone destruction cycles declines. However, detecting recovery of the stratospheric ozone layer is complicated by a number of additional factors which affect the year to year variability of total column ozone values. These factors include volcanic eruptions, such as the eruption of Mt. Pinatubo in 1991 (e.g. Randel ~~and Wu,~~et al., 1995; Telford et al., 2009), changes in the solar cycle (e.g. Brasseur, 1993; Van Loon and Labitzke, 2000; Austin et al., 2007; Calisesi and Matthes, 2007) and variability in ozone resulting from a range of factors affecting dynamical variability, including the quasi-biennial oscillation (QBO; e.g. Hollandsworth et al., 1995; Baldwin et al., 2001; Leblanc and McDermid, 2001) and variations in sea surface temperatures, particularly those related to the El Niño-Southern Oscillation (ENSO; e.g. Braesicke and Pyle, 2004; Manzini, 2009; Randel et al., ~~2009; Manzini,~~ 2009). In addition, long term total column ozone trends are driven in part by emissions of other~~,~~ non-chlorinated anthropogenic species, such as CO_2, CH_4 and N_2O, which affect stratospheric ozone concentrations by altering stratospheric temperatures and dynamics (Haigh and Pyle, 1982; ~~Plumb, 1996;~~ Avallone and Prather, 1996; Plumb, 1996; Eyring et al., 2010; 2013; Iglesias-Suarez et al., 2016), and in the case of CH_4 and N_2O by acting as source gases for reactive HO_x and NO_x species (Chipperfield and Feng, 2003; Ravishankara et al., 2009; Revell et al., 2012; Meul et al., 2014). Identification of significant trends is also made problematic by the difference in year-to-year variability in total column ozone values in different regions. For example, high northern latitudes exhibit very large interannual variability in winter and spring, while variability in the ~~southern hemisphere~~Southern Hemisphere is comparatively smaller. Furthermore, there is a dynamical response to changes in chemical ozone depletion in the stratosphere, which may enhance/impede future recovery by altering the transport of ozone (e.g. McLandress et al., 2011; Braesicke et al., 2013; Keeble et al., 2014~~; Solomon et al., 2016).~~). In comparison, the chemical ozone depletion signal in the tropics is small and total column ozone variability is dominated by features such as the solar cycle, QBO and ENSO. As a result of all of these factors, identifying robust recovery of total column ozone and ascribing that recovery to a decline in stratospheric halogen species is a complex issue.

~~Recovery~~For past trends, recovery of the stratospheric ozone layer could be detected using two different methodologies~~. For past trends~~: process oriented studies and statistical analysis of datasets. For the first, observations can be compared with a detailed chemistry-transport model which includes all known processes. ~~Good~~If good agreement is found between the model and observations when all processes are included, ~~but not~~then evidence of ozone recovery due to decreasing stratospheric halogen loadings can be identified by excluding other processes. For example, Solomon et al. (2016) found evidence for healing of the Antarctic ozone layer in September when polar halogen chemistry is included ~~and~~but interannual dynamical variability ~~is~~and volcanic factors are excluded~~, could constitute a sign of ozone recovery (e.g. Solomon et al., 2016). Alternatively~~. For the second method, a statistical approach can be followed in which data are used to detect significant change between time periods. The impact of confounding changes (e.g., QBO, solar cycle, etc.) can be quantified using multiple linear regression and removed from the statistical analysis of the data in order to provide a better estimate of long-term trends (e.g. Staehelin et al., 2001; Reinsel et al., 2005; WMO, 2007; Harris et al., 2015; Chipperfield et al., 2017). These statistical approaches nearly all work by relying on the ~~somewhat simplistic assumption of a linear relationship between a proxy variable and its impact on total ozone.~~assumption of a linear relationship between a proxy variable and its impact on total ozone. Using this method, a number of recent studies has started to explore if observed total column ozone and ozone profile values show signs of recovery (e.g. Pawson et al., 2014; Harris et al., 2015; Steinbrecht et al., 2017; Ball et al., 2018; Weber et al., 2018). These studies have indicated that statistically significant recovery of column ozone values can be identified in some datasets at some latitudes, but that this is not true for all datasets (e.g. Weber et al., 2018). As recovery trends are calculated over relatively short time frames (<20 years), identification of trend magnitude and trend significance from observations can be affected by high or low values at the beginning or end of the observational record (compare, for example, the trends derived by Pawson et al. (2014) with those of Weber et al. (2018)).

To explore future ozone trends and recovery, data from ~~fully~~ coupled chemistry-climate model (CCM) simulations are required. Each CCM simulation constitutes a possible future evolution of stratospheric ozone. In order to sample the effect of internal atmospheric variability on ozone and to derive an estimate of uncertainty of future trends, multiple ensemble members can be run in which the initial conditions of each simulation are modified but the same forcings are prescribed (e.g. GHG (greenhouse gas) evolution, ~~SSTs,~~ aerosol loadings). Greater confidence can be assigned to significance of the mean trend as the number of ensemble members increases. Multiple ensemble members also give information about the possible range of future trends~~.~~ and as a result are not as sensitive to high or low values at the beginning or end of the record of any individual ensemble member, in contrast to single member simulations and observational records. Thus, using an ensemble of future projections from a single CCM can provide additional insight into the detection of different phases of ozone recovery.

In this study, we use results from a chemistry-climate model coupled with statistical approaches to explore different definitions of ozone recovery (see Reinsel et al., 2005; Weatherhead and Andersen, 2006; Chipperfield et al., 2017). In particular we define three stages of total column ozone recovery:

1. A reduced rate of decline in ozone and the date of minimum ozone.
2. Statistically significant increases in column ozone values, after accounting for natural variability, that can be ascribed to reductions in ESC.
3. Return of total column ozone values to some specified past value (typically 1980 or 1960).

Identifying when and if each of these stages has occurred at different latitudes~~—~~, and being able to assess the confidence with which this can be done~~–~~, is fundamental to determining the success of the Montreal Protocol. For this work we use the ozone fields calculated in an ensemble of UM-UKCA transient simulations, which are described ~~briefly~~ in section 2~~, as proxies for atmospheric observations~~. We carry out a statistical analysis of the model results, as outlined in section 3, to identify when each of these stages of recovery occurs for different latitude ranges. These results are presented in section 4, 5 and 6 and implications are discussed in section 7.

2 Model configuration and simulations

An ensemble of transient simulations was performed using version 7.3 of the HadGEM3-A configuration of the Met Office's Unified Model (Hewitt et al., 2011) coupled with the United Kingdom Chemistry and Aerosol scheme (hereafter referred to as UM-UKCA). This configuration of the model has a horizontal resolution of 2.5° latitude by 3.75° longitude, ~~and~~with 60 vertical levels following a hybrid sigma-geometric height coordinate with a model top at 84 km. The chemical scheme used in this configuration of the model is an expansion of the scheme presented in Morgenstern et al. (2009) in which halogen source gases are considered explicitly, resulting in an additional 9 species, 17 bimolecular and 9 photolytic reactions. Stratospheric aerosol concentrations are prescribed using a climatology based on observations (from SPARC, 2006; described by Eyring et al., 2008) for the historical part of the run, after which background concentrations of stratospheric aerosol loadings are prescribed. HadGEM3-A includes an internally generated quasi-biennial oscillation (QBO), which in this configuration of the model has a period of ~27 months while the magnitude of modelled easterly(westerly) equatorial zonal wind speed is ~25 m s^{-1} (10m s^{-1}), both aspects in good agreement with observed zonal winds at Singapore (e.g. Lee and Smith, 2003). The configuration of the model used for this study includes the effects of the 11-year solar cycle in both the radiation and photolysis schemes. The top of atmosphere solar flux follows historical observations from 1960 to ~~2012~~2009, after which a ~~climatological~~repeating solar cycle is imposed which is an ~~average of~~amplitude equivalent to the ~~four solar cycles preceding 2012.~~observed cycle 23 (as detailed in Bednarz et al., 2016).

The transient simulations were performed following the experimental design of the WCRP/SPARC CCMI REF-C2 experiment (Eyring et al., 2013), which adopts the RCP6.0 scenario for future GHG and ODS emissions. Two of these ensemble members were run from 1960 to 2099 and an additional five were run from 1980 to 2080. All ensemble members have identical time-dependent boundary conditions, but differ in their atmospheric initial conditions, thereby providing an estimate of internal atmospheric variability. The simulations were performed in an atmosphere-only configuration, and each ensemble member uses prescribed sea surface temperatures ~~(SSTs)~~ and sea ice fields taken from a parent coupled atmosphere-ocean HadGEM2-ES ~~integration~~simulation as lower boundary conditions. The simulations used for this study are described in more detail in Bednarz et al. (2016) and Keeble et al. (2017), and were performed in support of phase 1 of the Chemistry–Climate Model Initiative (CCMI; Morgenstern et al., 2017).

3 Removing natural cycles

Identifying an increase in total column ozone resulting from reductions in stratospheric chlorine requires removing the effects of natural processes, such as volcanic eruptions, the QBO, ENSO and solar cycle, from the modelled total column ozone data, as these cycles may impose short terms trends in the data which are wrongly interpreted as signs of recovery. In order to identify the impacts of these natural processes on modelled total column ozone we create a statistical model using multiple linear regression (MLR) analysis. This process assumes that the modelled total column ozone values, TO3, can be reproduced by combining some constant value of total column ozone ~~$TO3_i$~~, i, which corresponds to the intercept term of the MLR, with a number of explanatory, or predictor, variables. This statistical model can be expressed as:

$$TO3_{e,l,t} = \cancel{TO3_{e,l,t} + \alpha^{QBO_{50}}} i_{e,l} + \alpha_{e,l}^{QBO_{50}}.QBO_{50_{e,t}} + \cancel{\alpha^{QBO_{30}}}\alpha_{e,l}^{QBO_{30}}.QBO_{30_{e,t}} + \cancel{\alpha^{solar}}\alpha_{e,l}^{solar}.solar_t + \cancel{\alpha^{ENSO}}\alpha_{e,l}^{ENSO}.ENSO_t + \cancel{\alpha^{aerosol}}\alpha_{e,l}^{aerosol}.aerosol_t + \cancel{\alpha^{ESC}}\alpha_{e,l}^{ESC}.ESC_{e,l,t} + N_{e,l,t}$$

in which the α values are the coefficients returned from the MLR for each explanatory variable (denoted by the superscript) and vary between latitude range and ensemble member. The explanatory variables included in the MLR are the QBO, solar cycle, ENSO, volcanic aerosols and ESC. The subscripts *e*, *l* and *t* indicate that the alpha value or explanatory variable differs with ensemble member, latitude and/or time respectively. For the QBO, two terms are included, QBO_{50} and QBO_{30}, which correspond to equatorial westerly winds at 50 hPa and 30 hPa respectively. Two QBO terms are included to account for the phase shift in the total column ozone response with respect to QBO changes at different altitudes. The solar cycle is represented by the top of atmosphere solar flux, represented in the MLR as $solar_t$. ENSO effects on column ozone are represented by $ENSO_t$, the detrended sea surface temperature anomalies in the NINO3.4 region. Volcanic aerosols are included as hemispheric aerosol optical depths, and so are different for the northern and ~~southern hemispheres~~Southern Hemispheres to account for the lack of interhemispheric transport of aerosols emitted into the stratosphere from high latitude eruptions. The final term included in the MLR, *ESC*, represents stratospheric chlorine concentrations. This term is equal to the ESC concentration at 30 km for each latitude bin to account for the time taken for ODS to be transported to higher latitudes. Any ~~noise which is~~month to month variation not ~~explained~~accounted for by the explanatory variables in the MLR is represented by the noise term ~~N.~~$N_{e,l,t}$~~.~~ The~~ subscripts e, l and t indicate that the explanatory variable differs with ensemble member, latitude and/or time respectively.~~ $Solar_t$, $ENSO_t$ and $aerosol_t$ terms are all prescribed forcings in the model and do not vary between ensemble members. In this study we use deseasonalised monthly mean total column data, and so there is no need for a seasonal cycle term in the MLR model.

This statistical model can then be used to remove the component of total column ozone variations related to the QBO, solar cycle and volcanic aerosol changes from the raw model data to leave a set of ozone residuals, RO3, which retain the long-term trend and any interannual variability not explained by the MLR:

$$RO3_{e,l,t} = TO3_{e,l,t} - \left(\cancel{\alpha^{QBO_{50}}}\alpha_{e,l}^{QBO_{50}}.QBO_{50_{e,t}} + \cancel{\alpha^{QBO_{30}}}\alpha_{e,l}^{QBO_{30}}.QBO_{30_{e,t}} + \cancel{\alpha^{solar}}\alpha_{e,l}^{solar}.solar_t + \cancel{\alpha^{ENSO}}\alpha_{e,l}^{ENSO}.ENSO_t + \cancel{\alpha^{aerosol}}\alpha_{e,l}^{aerosol}.aerosol_t\right)$$

MLR analyses and ozone residuals are produced for each individual ensemble member of the raw model data, resulting in seven RO3 time series. For both the raw model data and the ozone residuals, decline and recovery trends are calculated using independent linear trend fits for the periods 1980-1997 and 2000-2017 respectively. When calculating trends for both the raw model data and ozone residuals, a single linear fit is produced using data from all seven ensemble members rather than producing a fit for each individual ensemble. In order to make any robust conclusions about the statistical significance of modelled trends, some measure of the trend uncertainty is required. Here we calculate trend uncertainties following the methodology of Weatherhead et al. (1998), in which the standard deviation of the uncertainty in the linear trend is calculated by:

$$\sigma_{trend} = \frac{\sigma_{data}}{n^{3/2}}\sqrt{\frac{1+\varphi}{1-\varphi}}$$

where σ_{data} is the standard deviation of the time series in question (either raw model data or RO3), n is the number of months in the time series and φ is the autocorrelation coefficient for a 1 month lag. As discussed by Weatherhead et al. (1998) autocorrelation can be substantial for monthly mean time series, particularly in low latitudes, and failure to account for it results in an underrepresentation of trend uncertainty. As for the trend calculations, trend uncertainties are calculated using data from all seven ensemble members together, and as such for the 17 year periods considered for the decline and recovery phases n is very large (n=17x12x7=1428).

4 Modelled global column ozone and minimum values

Figure 1 shows deseasonalised monthly mean total column ozone anomalies relative to 1980 values, averaged over 60°S-60°N, from 1960 to 2100 for each individual ensemble member (light blue lines) and the ensemble mean (dark blue line). A sharp decrease in total column ozone is modelled from 1980 to the late 1990s, consistent with increased ESC loadings resulting from the use and emission of ODS. From the late 1990s until ~2070 column ozone values gradually increase, exceeding their 1980s values by ~2030, and their 1960s values by ~2050. Beyond 2070 total column ozone values remain relatively constant until the end of the century. Superimposed on these long term trends is the effect of the solar cycle, which imprints a distinctive 11-year oscillation in the data. Alongside the modelled total column ozone anomalies are shown values from version 2.8 of the Bodeker Scientific total-column ozone dataset in black (Bodeker et al., 2005). There is generally good agreement between modelled total column ozone anomaly values and the Bodeker dataset; decadal total column ozone changes, the response of column ozone to the solar cycle and the magnitude of interannual variability are all well captured by the model ensemble throughout the time period during which the observations and model data overlap.

Also shown in Figure 1 are the ozone residuals calculated when the effects of natural cycles are removed, as detailed in section 3 (red lines). This dataset follows the long term trends of the raw UM-UKCA data, but the cyclic short term trends in column ozone values have been removed. Most ~~obviously~~obvious from Figure 1 is the removal of the 11-year solar cycle signal, leading to a much smoother, monotonically increasing trend from 2000 to 2060 compared to the raw model data.

As discussed above, the first signs of detectable ozone recovery would be identified as a reduced rate of decline in column ozone and the date of minimum ozone. Modelled total column ozone values generally decrease from 1980 to the late 1990s (blue line, Figure 1), consistent with the increase in ESC amounts. However, this decrease is not constant; rapid decline is modelled from 1980 to 1985 and from 1990 to 1995, while between these periods total column ozone abundances are relatively constant, or even increase (see inset in Figure 1). This feature is also seen in the Bodeker dataset, and predominantly results from the impact of the solar cycle on stratospheric ozone concentrations. As top of atmosphere solar flux decreases from solar maximum to solar minimum, rapid decline of total column ozone occurs as this effect combines with the impacts of increasing ESC. Conversely, as top of atmosphere solar flux increases, enhanced stratospheric ozone production temporarily offsets the chemical ozone destruction resulting from increased ESC concentrations. This is confirmed by analyzing the ozone residuals (red line Figure 1), which show a much smoother decline from 1980 to the late 1990s, and highlights the importance of understanding the drivers of short term trends in raw total column ozone values when trying to assess longer term trends.

As well as influencing the trajectory of declining column ozone abundances, natural cycles also affect the timing and magnitude of minimum total column ozone values. In the raw model data, the minimum total column ozone values averaged over 60°S-60°N are reached between 1992-1994, depending on the ensemble member, which is several years before the peak loading of ESC in 1997 (e.g. Mäder et al., 2010; WMO 2011; 2014). This offset in timing between peak ESC and total column ozone ~~minima~~minimum results from the impact of the solar cycle, as discussed above, and the eruption of Mt. Pinatubo on total column ozone. The early 1990s was a time of low top of atmosphere solar flux, while the eruption of Mt. Pinatubo increased stratospheric sulphate surface area density, both reducing total column ozone abundances. When the effects of these natural cycles are removed, ozone residuals (red line Figure 1) are larger than modelled total column ozone values ~~are seen to be greater~~ throughout the early 1990s.

Although this work indicates minimum column ozone values occurred in the 1990s, this is a poor metric for making robust conclusions about ozone recovery. Firstly, the ozone minimum may occur because there is no more capacity for increased chemical depletion despite increasing ESC. This is the case over Antarctica during springtime during the 1990s, where near complete destruction of polar lower stratospheric ozone occurs and any additional increase in ESC would have a negligible effect. Secondly, minimum column ozone values are very sensitive to dynamical conditions. For example, Bednarz et al. (2016) have shown that even under much lower stratospheric halogen loadings significant ozone depletion can occur in the Arctic lower stratosphere during conditions which favour a cold, stable polar vortex. Even outside the high latitudes, where interannual variability in total column ozone values is largest, identification of the year of minimum ozone is uncertain, with each of the seven residual ozone time series having minimum values at different times between 1992-2000 (light red lines, Figure 1).

5 Regional trends

Decline and subsequent recovery of total column ozone is often calculated using ~~piecewise~~ linear trends ~~in~~for two periods either side of an inflection time (e.g. Newchurch et al., 2003; Reinsel et al., 2005; Jones et al., 2009; Nair et al., 2013; Chehade et al., 2014). Previous studies have identified 1997 as the inflection time for long-term total column ozone observed trends (e.g. Harris et al., 2008), and as a result we define the decline phase as 1980-1997 with the recovery phase defined from 2000-2017. Here we calculate independent linear trends for both the decline and recovery phases firstly using the raw total column ozone data from the UM-UKCA model (discussed below) and then using model data in which the effects of the natural processes discussed above have been removed using the statistical model introduced in Section 3.

Figure 2 shows ~~DU year^{-1} trends for the decline (1980-1997) and recovery (2000-2017) phases for the~~total column ozone trends (in DU year^{-1}) obtained from both the raw data from the UM-UKCA simulation and the ozone residuals ~~when natural cycles are accounted for~~for the decline (1980-1997) and recovery (2000-2017) phases, averaged over 10° latitude bands. Error bars associated with each trend represent the 95% confidence intervals~~.~~ ($2\sigma_{trend}$), calculated as described in section 3. During the decline phase, ozone trends ~~from~~for both ~~datasets~~the raw model data and ozone residuals are greatest at high latitudes due to the heterogeneous activation of chlorine on PSCs within the polar vortex~~.~~ and the transport of mid-latitude ozone depletion signals to high latitudes by the BDC. The uncertainty associated with ~~these~~the trends is also largest at high latitudes, due to the higher year-to-year variability in chemical and dynamical processes at high latitudes compared with the tropics. ~~Trends~~Negative trends in the raw column ozone data from 1980-1997 are significant at all latitudes, although when natural cycles are removed the trends from 10°S-10°N are not significant. At all latitudes there is a more negative trend in the raw UM-UKCA data compared with the dataset in which the natural cycles have been removed. This is the result of the eruption of Mt Pinatubo and the pronounced solar ~~minima~~minimum during the 1990s, both of which resulted in lower column ozone values and so a greater trend from 1980. This can be clearly seen in Figure 1 by comparing the blue and red lines. However, the close agreement in trends between the raw modelled values and the ozone residuals calculated when natural cycles are accounted for indicates that natural cycles have had only a small contribution to the trends during the period 1980-1997 (consistent with the findings of Gillett et al., 2011). Trends for the decline phase, calculated for both the raw model data and ozone residuals, agree within the uncertainty estimates with those obtained from observation datasets by Weber et al. (2018) and those obtained from CCMVAL2 models (e.g. Pawson et al., 2014).

When considering the recovery phase, positive trends are modelled at all latitudes from 2000-2017~~. These~~ for both the raw model data and ozone residuals. For the raw model values, these trends are only significant at the 95% confidence interval in the ~~southern hemisphere~~Southern Hemisphere between 80°S-50°S. However, when the effects of natural cycles are removed from the data, significant positive trends can be identified in the ~~southern hemisphere~~Southern Hemisphere between 80°S-30°S, and ~~for the first time~~also in the ~~northern hemisphere~~Northern Hemisphere from 20°N~~-~~-70°N. ~~For both datasets significant~~Significant trends cannot be identified by 2017 at the highest latitudes, in either dataset, due to the large ~~year to year~~interannual variability in the data~~.~~, nor in the tropics due to the small trend magnitude. As for the decline phase, while accounting for nature cycles in the ozone residuals reduces the trend uncertainties, it does not significantly affect the trend magnitudes, indicating that natural cycles do not significantly contribute to recent increases in column ozone values.

Trends for the recovery phase, calculated for both the raw model data and ozone residuals, are consistent with those calculated for observational datasets by Chehade et al. (2014), Pawson et al. (2014) and Weber et al. (2018), and agree with all three studies within trend uncertainty estimates. In general, while in agreement with each of these studies, trends presented here are larger than those presented by Weber et al. (2018), and are closer in magnitude to the findings of Chehade et al. (2014) and Pawson et al. (2014). The modelled trends presented here indicate, as Weber et al. (2018) conclude, that the differences between the latest trends calculated from observations by Weber et al. (2018) and the earlier trends estimated by Pawson et al. (2014) result from lower than average column ozone values at the end of the observational record, which are part of natural interannual variability and likely do not represent some fundamental shift in the trajectory of ozone recovery.

Identification of significant trends depends on the gradient of the trend, the number of data points (in this case the number of modelled monthly means) and the variance and autocorrelation of the data~~.~~ (e.g. Weatherhead et al. 2000). Analyzing the near global (60°S-60°N) raw total column ozone data (blue lines, figure 1), year 2000 is a solar maximum year and so total column ozone values are relatively high compared to the following few years. It is not until ~11 years later, during the next solar maximum, that trends become positive. Trend analysis on data between 2000 and 2015 could indicate that there is a significant positive trend, which could in turn lead to the conclusion that significant recovery of the ozone layer had begun. However, as further years are considered, from 2015 to 2020, total column ozone values start to decline as the solar cycle moves towards a solar ~~minima~~minimum, and the magnitude of the recovery trend is reduced while the variance in the residuals increases. Now, trends calculated from 2000 to 2020 are no longer statistically significant. As a result, when assessing recovery trends it is necessary to use datasets in which the effects of natural cycles have been accounted for, such as the ozone residuals calculated in section 3. However, as the MLR analysis performed on the raw data does not capture the full modelled interannual variability (hence the occurrence of the noise term, $N_{e,l,t}$), trend magnitudes calculated using ozone residuals from any individual ensemble member are similarly affected by anomalously low or high values at the start or end of the time series. The use of multiple member ensemble simulations in this study mitigates this effect.

~~Figure 3 shows the month~~The impacts of the noise term on trend magnitudes for each of the individual ensemble members is explored in more detail in figure 3, which shows the year after 2000 at which trend significance can be identified in the ozone residuals for either the first time (blue), or final time (red). ~~Here trends are identified as significantly different from zero when they have a p-value < 0.05. Starting from 2000, the first month~~The year a trend becomes significant can be calculated for each ensemble member by identifying the first month after January 2000 in which trends become significant (initial recovery) and the month after which they remain significant (robust recovery~~) can be calculated for each ensemble member. Distinguishing between these initial significance and robust~~). Here trend significance is calculated using the trend uncertainty obtained when all ensemble members are used, as discussed above, but the trend magnitude is calculated for each ensemble member individually so as to provide a range of dates that trend significance can be identified. This range of dates reflects the impact of unaccounted for noise on the trend magnitude, while maintaining significance thresholds which are consistent between figures 2 and 3. Error bars on figure 3 represent the 95% confidence intervals, calculated as twice the standard deviation of the seven values obtained for the year trend significance is identified (one for each ensemble member).

Distinguishing between initial and robust recovery significance dates is necessary since, as discussed above, trends can be significant after a number of months and then become non-significant as more data is added ~~which increases~~so that the variance in the data can increase or ~~decreases~~ the magnitude of the trend can decrease. The first instance of detection of significant trends can be considered as false recovery if it does not coincide with the time after which trends never become non-significant. ~~This is most clearly exhibited in the raw model data when considering the solar cycle, as discussed above, but is also true for the ozone residuals. However, when natural cycles are removed, trend estimates become significant earlier due to the reduced variance in the data, and both initial and robust recovery occur closer together. Note that,~~Note that if the MLR described in section 3 accurately represented all drivers of interannual variability (i.e. the N term was zero), there would be no distinction between initial and robust recovery.

~~Mid~~ For the ensemble of ozone residuals presented here, mid-latitude trends become significant earlier than those of the tropics or high latitudes. This is due to the high degree of interannual variability at high latitudes, particularly in the Arctic, and the small magnitude of the trends in the tropics. Therefore, it is likely that both initial and robust recovery will first be observed in the mid~~–~~-latitudes. In addition, both measures of recovery occur at similar times, minimizing the risk of identifying false recovery. Correct identification of robust recovery is important when considering observations of total column ozone and highlights the need to treat detection of significant recovery for the first time with caution as additional months/years of observational data may reduce the statistical significance of any observed trends. It should be noted that no individual ensemble member shows statistically significant recovery by 2017, and only when considering data from all the ensembles together when calculating trend uncertainties can significant trends be identified.

Once the difference in projected recovery trends is accounted for, these findings are consistent with those of Weatherhead et al. (2000), who also identified the mid-latitudes as the best location to identify early signs of ozone recovery. However, there is an offset in when trends are expected to become significant between the two studies, with projected detection years modelled in this study generally occurring earlier than those of Weatherhead et al. (2000), particularly in the tropics. This discrepancy in the tropics is unsurprising, as recovery of the tropical ozone column is dependent on the competing influences of declining CFCs, decreasing stratospheric temperatures and changing BDC speeds (see e.g. Eyring et al., 2013; Meul et al., 2016; Keeble et al., 2017), with increases to the BDC offsetting ozone recovery in the lower stratosphere and resulting in smaller column ozone recovery trends. There is poor agreement in modelled projections of future BDC speeds, and as a result projections of tropical column ozone differ significantly between models (e.g. WMO, 2011). In addition, ozone depletion in the tropics resulting from increasing ESC concentrations is weak in comparison to the mid and high latitudes, and as a result identifying significant recovery trends is particularly sensitive to any interannual variability not accounted for in the MLR, which may differ between models.

6 Return to historic values

While identification of statistically significant increases in total column ozone is a real sign that ozone recovery is occurring, recovery can be said to be complete when column ozone values reach their pre-CFC values again. Traditionally these return thresholds are taken to be either 1980 or 1960 values; here we use 1980. It is likely that future total column ozone values will initially exceed the 1980s threshold and then fall below this value again due to interannual variability and the effects of the solar cycle and QBO. As a result two metrics are considered: the first time total column ozone exceeds the 1980s threshold, and the last time total column abundances are below the threshold. Between the two time periods total column ozone values rise above and fall below the threshold.

Figure 4 shows the year in which raw total column ozone abundances ~~are found to~~ return to their 1980s values for the first time (~~light red~~blue) and final time (~~dark~~ red) for each 10 degree latitude bin. In the tropics, total column ozone exceeds the 1980s threshold as early as 2000 ~~which is due to~~as the amplitude in total column ozone variations resulting from the solar cycle ~~being~~is greater than the ~~trends due to~~decrease in total column ozone resulting from ESC changes. However, despite this region seeing the first values greater than those of the 1980s, it is the only region in which total column ozone abundances are not greater than their 1980s values by the end of the simulation, consistent with other studies (e.g. Eyring et al., 2013; Meul et al., 2016). This is due to decreasing lower stratospheric ozone concentrations resulting from an acceleration of the ~~Brewer-Dobson circulation~~BDC under increased greenhouse gas concentrations offsetting increased upper stratospheric ozone concentrations due to decreased ESC and increased CO_2 (explored in detail in Keeble et al., 2017).

In the ~~northern hemisphere~~Northern Hemisphere mid-latitudes earliest recovery occurs by ~2020, while final recovery occurs by 2040. The closeness of these two dates is due to the ~~reduced interannual~~ large trend to variability ratio in the mid~~-~~-latitudes compared to both the tropics and Arctic. The results are similar in the Southern ~~hemisphere~~Hemisphere mid~~-~~-latitudes, although both dates are delayed by around 10 years, most likely due to the effects of Antarctic polar ozone depletion and transport of ozone poor ~~airmasses~~air masses into these latitudes upon the collapse of the Antarctic polar vortex.

Earliest recovery to historic values at high southern latitudes occurs by ~2040, with final recovery occurring by 2060. However, the signature of this recovery is very sensitive to calendar month, and earlier signs of recovery may be identified in certain months (e.g. September; Solomon et al., 2016). Arctic column ozone exhibits high interannual variability, with values exceeding the 1980s threshold as early as 2010. However, final recovery is not ~~expected~~projected until ~2060.

The future evolution of total column ozone is dependent on the emissions scenario considered, and the exact timings of recovery to historic values will vary with changes to CO_2, N_2O and CH_4 emissions as well as ESC reductions. As a result, the expected return dates for each latitude will evolve as we approach those dates, in line with our increased understanding of the emissions pathway or if future emission controls come into effect.

7 Discussion and Conclusions

~~We have analysed~~In this study we analyse total ozone values from an ensemble of UM-UKCA model runs to investigate different definitions of progress on the road to ozone recovery. ~~This approach allows us to take into account the internal atmospheric variability of the model in assessing the statistical significance of each definition and so provide insight into similar analyses of measurements.~~. In particular, we have investigated three definitions: (i) a slowed rate of decline and the date of minimum ozone; (ii) the identification of significant positive trends; and (iii) a return to historic values. The impacts of natural cycles on modelled internal atmospheric variability are accounted for by applying a multiple linear regression model to modelled total column ozone values. The use of multi-member CCM ensembles, in which each simulation constitutes a possible future evolution of stratospheric ozone, allows us to better account for the modelled internal atmospheric variability not captured by the explanatory variables in the MLR, and so provide greater confidence when assessing the statistical significance of each definition.

The first and most obvious conclusion is that recovery can be identified in the first two metrics before a return to past thresholds is achieved. ~~However, an important caveat~~ For the first definition of recovery, minimum total column ozone values averaged from 60°S-60°N occur between 1990 and 1995 for each ensemble member, driven in part by the solar minimum conditions during the 1990s. When natural cycles are accounted for, identification of the year of minimum ozone in the resulting ozone residuals is uncertain, with minimum values for each of the seven residual ozone time series occurring at different times between 1992 and 2000. As a result, identification of the date of minimum ozone values is problematic and a poor measure of ozone recovery.

For the second definition of recovery, positive trends are modelled at all latitudes from 2000-2017 for both the raw model data and ozone residuals. In contrast to recent analysis of total ozone measurements (e.g. Chipperfield et al., 2017; Weber et al., 2018), when the effects of natural cycles are removed from the data, statistically significant positive trends can be identified in the Southern Hemisphere between 80°S-30°S, and also in the Northern Hemisphere from 20°N-70°N. This increased significance results largely from the much larger sample size that arises in a multi-member ensemble and the resulting reduction in the uncertainty associated with the mean trend. Significant trends cannot be identified by 2017 at the highest latitudes, due to the large interannual variability in the data, nor in the tropics, due to the small trend magnitude. It was found that while ~~recovery may appear to be robust~~ accounting for the effects of natural cycles impacted trend uncertainty estimates, it did not significantly affect trend magnitudes for either the decline or recovery phases, indicating that natural cycles have played only a minor role in recent trends, consistent with previous studies.

It is important to note that, while a statistically significant, positive recovery trend could be calculated at a particular point of time, additional years of ~~observations~~data may lead to a reduced significance of ~~trends due to natural variability (e.g. solar cycle, QBO, ENSO). This points to~~ trend, due to either a decrease in the magnitude of the trend or an increase in interannual variability. This effect results in a need to distinguish between initial recovery (the time at which trends become significant for the first time) and robust recovery (the time after which trends remain significant despite adding further years). Accounting for this, we identify the mid-latitudes as the best place to find early signs of ozone column recovery. This is due to the combination of reasonably large trend magnitudes and comparably low variability (especially in the Southern Hemisphere). Despite the large trend magnitudes modelled in the high latitudes, interannual variability in these regions resulting from both the large dynamic interannual variability and the large changes in chemical ozone loss occurring from year to year is too large for a statistically significant signal to be easily detected. In contrast, the small trend magnitudes modelled in the tropics confound identification of statistically significant ozone recovery.

For the third definition of recovery, a return to historic values, it was found that, while robust recovery could be identified at all latitudes by ~2030, column ozone values which are lower than the 1980 annual mean can occur in the mid-latitudes until ~2050, and in the tropics and high latitudes deep into the second half of the 21st century. While projected column ozone values for the mid and high latitudes reach a point after which they are never lower than the 1980 annual mean, consistent with the projected super recovery of ozone, column ozone values lower than the 1980 annual mean occur in the tropics until the end of the period analysed in this study. This results in part from the large amplitude ozone response to natural cycles, particularly the solar cycle, and also the effects of increased BDC speeds offsetting column ozone recovery resulting from decreased CFCs, as discussed by Keeble et al. (2017).

This work further highlights the need to ensure that the ~~impact~~impacts of natural cycles (e.g. solar cycle, QBO, ENSO) on total ozone ~~is~~are correctly described ~~in the statistical model~~when performing MLR analysis. This is a challenge because of a number of factors. Firstly, the assumption in all MLR analysis of a linear relationship between the proxy variables used and the impact on ozone is not accurate, and there is growing evidence that these cycles are not isolated, but interact with one another (e.g. White and Liu, 2008; Calvo et al., 2009; Gray et al., 2010). Secondly, cycles with varying amplitudes (e.g. the solar cycle, which shows differing top of atmosphere solar flux during the last four solar maximums) or lengths (e.g. the QBO, the period of which may change in the future and has recently been observed to undergo rapid, non-periodic reversal) have different impacts on total column ozone which makes accurate estimates of the coefficients for these variables in the MLR harder to achieve. Finally, volcanic eruptions are particularly difficult to account for in the MLR, both because of the infrequent, non-periodic timings of eruptions, and because eruptions have very different impacts on stratospheric ozone when stratospheric ESC concentrations are high compared to when ESC is low (e.g. Tie and Brasseur, 1995).

~~The mean trends of the ensemble members for 2000-2017 are positive at most latitudes. Unlike a recent analyses of total ozone measurements (Chipperfield et al., 2017), these trends are statistically significant. This increased significance results largely from the much larger sample size that arises in a multi-member ensemble and the resulting reduction in the uncertainty associated with the mean trend. The dates at which statistically significant trends are reached indicates that recovery of total column ozone began before 2017 at many latitudes in this ensemble of model runs.~~

~~The influence of the natural cycles on trend determination is least at latitudes where the trends are sizeable, as shown by the small differences between the year of first return and the year of final return as well as the negligible difference between the analysis using MLR analysis and that using raw data. For this reason, the mid-latitudes are the best place outside of Antarctica to identify recovery in the ozone column. This is due to the combination of reasonably large trend signal and comparably low variability (especially in the southern hemisphere). The Arctic is too variable for a signal to be easily detected: this arises both from the large dynamic interannual variability and from the large changes in chemical ozone loss from year to year. The tropics have too small a trend compared to the natural variability resulting from the solar cycle and the QBO. This is most clearly seen in the large difference between the year at which recovery is first observed in the tropics and the year at which it remains above historic thresholds, though note that the latter year is influenced by the end of the model runs. While chemical ozone depletion is largest over Antarctica, the pronounced seasonal dependence of this depletion, with the majority occurring during austral spring, means detection of trends is sensitive to which month is chosen. Solomon et al. (2016) highlight that while both September and October show high levels of chemical ozone depletion, the higher interannual variability in October compared to September (resulting from the break-up of the vortex in late October-early November) means identification of ozone recovery is easier in September. As a result, Antarctica is a good location to look for early signs of recovery, but care must be taken when choosing which months to use, unlike in the mid-latitudes where the seasonal cycle is relatively smaller.~~

Our analysis has ~~been~~focused solely ~~focused~~ on interpreting the total ozone column record. Many studies have recently examined the trends in the vertical distribution of ozone since ESC maximized (e.g. Harris et al., 2015; Steinbrecht et al., 2017; Ball et al., ~~2017~~2018). In these studies, factors such as higher variability, greater uncertainties and poorer data quality add to the uncertainty in detection of significant trends compared to the total column. However similar studies to this one using ensembles of model runs could provide real insights into the issue, especially in the climatically important lower stratosphere where ozone may still be decreasing (e.g. Ball et al., ~~2017).~~2018). As a result we recommend the use of multi-member ensemble simulations, in conjunction with ongoing observational efforts, to better identify signs of ozone recovery for both the total column and ozone profiles.

Data availability

Data from the two 1960-2100 transient simulations are available as part of the CCMI initiative through BADC: https://blogs.reading.ac.uk/ccmi/badc-data-access/. All further data are available upon request.

Competing interesting

The authors declare that they have no conflict of interest.

Acknowledgements

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007 - 2013) under grant agreement n° 603557 (StratoClim) and the European Research Council through the ACCI project (project number: 267760). We thank NCAS-CMS for modelling support. Model ~~integrations~~simulations have been performed using the ARCHER UK National Supercomputing Service and MONSooN system, a collaborative facility supplied under the Joint Weather and Climate Research Programme, which is a strategic partnership between the UK Met

Office and the Natural Environment Research Council. We would like to thank Greg Bodeker of Bodeker Scientific, funded by the New Zealand Deep South National Science Challenge, for providing the combined total column ozone database.

References

Austin, J., Hood, L. L., and Soukharev, B. E.: Solar cycle variations of stratospheric ozone and temperature in simulations of a coupled chemistry-climate model, Atmos. Chem. Phys., 7, 1693–1706, doi:10.5194/acp-7-1693-2007, 2007.

Avallone, L. M. and Prather, M. J.: Photochemical evolution of ozone in the lower tropical stratosphere, J. Geophys. Res., 101, 1457–1461, doi:10.1029/95JD03010, 1996.

Baldwin, M. P., Gray, L. J., Dunkerton, T. J., Hamilton, K., Haynes, P. H., Randel, W. J., Holton, J. R., Alexander, M. J., Hirota, I., Horinouchi, T., Jones, D. B. A., Kinnersley, J. S., Marquardt, C., Sato, K., and Takahashi, M.: The quasi–biennial oscillation, Rev. Geophys., 39, 179–229, 2001.

Ball, W. T., Alsing, J., Mortlock, D. J., Staehelin, J., Haigh, J. D., Peter, T., Tummon, F., Stübi, R., Stenke, A., Anderson, J., Bourassa, A., Davis, S. M., Degenstein, D., Frith, S., Froidevaux, L., Roth, C., Sofieva, V., Wang, R., Wild, J., Yu, P., Ziemke, J. R., and Rozanov, E. V.: ~~Continuous~~Evidence for a continuous decline in lower stratospheric ozone ~~offsets~~offsetting ozone layer recovery, Atmos. Chem. Phys~~. Discuss.,~~., 18, 1379-1394, 2018.

Bednarz, E. M., Maycock, A. C., Abraham, N. L., Braesicke, P., Dessens, O., and Pyle, J. A.: Future Arctic ozone recovery: the importance of chemistry and dynamics, Atmos. Chem. Phys., 16, 12159-12176, doi:10.5194/acp-16-12159-2016, 2016.

Bodeker, G. E., Shiona, H., and Eskes, H.: Indicators of Antarctic ozone depletion, Atmos. Chem. Phys., 5, 2603–2615, doi:10.5194/acp-5-2603-2005, 2005.

Braesicke, P. and Pyle, J. A.: Sensitivity of dynamics and ozone to different representations of SSTs in the Unified Model, Q. J. Roy. Meteor. Soc., 99, 1-9, doi:10.1256/qj.03.999, 2004.

Braesicke, P., Keeble, J., Yang, X., Stiller, G., Kellmann, S., Abraham, N. L., Archibald, A., Telford, P., and Pyle, J. A.: Circulation anomalies in the Southern Hemisphere and ozone changes, Atmos. Chem. Phys., 13, 10677-10688, doi:10.5194/acp-13-10677-2013, 2013.

Brasseur, G.: The response of the middle atmosphere to long-term and short-term solar variability: A two-dimensional model, J. Geophys. Res., 98, 23079–23090, 1993.

Calisesi, Y. and Matthes, K.: The Middle Atmospheric Ozone Response to the 11-Year Solar Cycle, Springer New York, 23, 273–286, doi:10.1007/978-0-387-48341-2_22, 2007.

Calvo, N., Giorgetta, M. A., Garcia-Herrera, R., and Manzini, E.: Nonlinearity of the combined warm ENSO and QBO effects on the Northern Hemisphere polar vortex in MAECHAM5 simulations, J. Geophys. Res.-Atmos., 114, D20, doi:10.1029/2008JD011445, 2009.

Chehade, W., Weber, M., and Burrows, J. P.: Total ozone trends and variability during 1979–2012 from merged data sets of various satellites, Atmos. Chem. Phys., 14, 7059–7074, https://doi.org/10.5194/acp-14-7059-2014, 2014.

Chipperfield, M. P. and Feng, W.: Comment on: Stratospheric Ozone Depletion at northern mid-latitudes in the 21st century: The importance of future concentrations of greenhouse gases nitrous oxide and methane, Geophys. Res. Lett., 30, 1389, doi:10.1029/2002GL016353, 2003.

Chipperfield, M., Liang, Q., Strahan, S., Morgenstern, O., Dhomse, S., Abraham, N., Archibald, A., Bekki, S., Braesicke, P., Di Genova, G., Fleming, E. L., Hardiman, S. C., Iachetti, D., Jackman, C. H., Kinnison, D. E., Marchand, M., Pitari, G., Pyle, J. A., Rozanov, E., Stenke, A., and Tummon, F.: Multimodel estimates of atmospheric lifetimes of long-lived ozone-depleting substances: Present and future, J. Geophys. Res. Atmos., 119, 2555–2573, 2014.

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

Eyring, V., Waugh, D. W., Bodeker, G. E., Cordero, E., Akiyoshi, H., Austin, J., Beagley, S. R., Boville, B., Braesicke, P., Bruhl, C., Butchart, N., Chipperfield, M. P., Dameris, M., Deckert, R., Deushi, M., Frith, S. M., Garcia, R. R., Gettelman, A., Giorgetta, M., Kinnison, D. E., Mancini, E., Manzini, E., Marsh, D. R.,Matthes, S., Nagashima T., Newman, P. A., Nielsen, J. E., Pawson, S., Pitari, G., Plummer, D. A., Rozanov, E., Schraner, M., Scinocca, J. F., Semeniuk K., Shepherd, T. G., Shibata, K., Steil, B., Stolarski, R., Tian, W., and Yoshiki, M.: Multimodel projections of stratospheric ozone in the 21st century, J. Geophys. Res., 112, D16303, doi:10.1029/2006JD008332, 2007.

Eyring, V., Chipperfield, M. P., Giorgetta, M. A., Kinnison, D. E., Manzini, E., Matthes, K., Newman, P. A., Pawson, S., Shepherd, T. G., and Waugh, D. W.: Overview of the New CCMVal Reference and Sensitivity Simulations in Support of Upcoming Ozone and Climate Assessments and the Planned SPARC CCMVal Report, SPARC Newsletter No. 30, 20–26, 2008.

Eyring, V., Cionni, I., Bodeker, G. E., Charlton-Perez, A. J., Kinnison, D. E., Scinocca, J. F., Waugh, D. W., Akiyoshi, H., Bekki, S., Chipperfield, M. P., Dameris, M., Dhomse, S., Frith, S. M., Garny, H., Gettelman, A., Kubin, A., Langematz, U., Mancini, E., Marchand, M., Nakamura, T., Oman, L. D., Pawson, S., Pitari, G., Plummer, D. A., Rozanov, E., Shepherd, T. G., Shibata, K., Tian, W., Braesicke, P., Hardiman, S. C., Lamarque, J. F., Morgenstern, O., Pyle, J. A., Smale, D., and Yamashita, Y.: Multi-model assessment of stratospheric ozone return dates and ozone recovery in CCMVal-2 models, Atmos. Chem. Phys., 10, 9451-9472, https://doi.org/10.5194/acp-10-9451-2010, 2010.

Eyring, V., Arblaster, J. M., Cionni, I., Sedláček, J., Perlwitz, J., Young, P. J., Bekki, S., Bergmann, D., Cameron-Smith, P., Collins, W. J., Faluvegi, G., Gottschaldt, K.-D., Horowitz, L. W., Kinnison, D. E., Lamarque, J.-F., Marsh, D. R., Saint-Martin, D., Shindell, D. T., Sudo, K., Szopa, S., and Watanabe, S.: Long-term ozone changes and associated climate impacts in CMIP5 simulations, J. Geophys. Res.-Atmos., 118, 5029-5060, doi:10.1002/jgrd.50316, 2013.

Gray, L. J., Beer, J., Geller, M., Haigh, J. D., Lockwood, M., Matthes, K., Cubasch, U., Fleitmann, D., Harrison, G., Hood, L., Luterbacher, J., Meehl, G. A., Shindell, D., van Geel, B., and White, W.: Solar Influences on Climate, Rev. Geophys., 48, G4001, doi:10.1029/2009RG000282, 2010.

Gillett, N. P., Akiyoshi, H., Bekki, S., Braesicke, P., Eyring, V., Garcia, R., Karpechko, A. Yu., McLinden, C. A., Morgenstern, O., Plummer, D. A., Pyle, J. A., Rozanov, E., Scinocca, J., and Shibata, K.: Attribution of observed changes in stratospheric ozone and temperature, Atmos. Chem. Phys., 11, 599-609, https://doi.org/10.5194/acp-11-599-2011, 2011.

Haigh, J. D. and Pyle, J. A.: Ozone perturbation experiments in a two-dimensional circulation model, Q. J. Roy. Meteor. Soc., 108, 551–574, doi:10.1002/qj.49710845705, 1982.

Harris, N. R. P., Kyro, E., Staehelin, J., Brunner, D., Andersen, S.-B., Godin-Beekmann, S., Dhomse, S., Hadjinicolaou, P., Hansen, G., Isaksen, I., Jrrar, A., Karpetchko, A., Kivi, R., Knudsen, B., Krizan, P., Lastovicka, J., Maeder, J., Orsolini, Y., Pyle, J. A., Rex, M., Vanicek, K., Weber, M., Wohltmann, I., Zanis, P., and Zerefos, C.: Ozone trends at northern mid- and high latitudes – a European perspective, Ann. Geophys., 26, 1207–1220, doi:10.5194/angeo-26-1207-2008, 2008.

Harris, N. R. P., Hassler, B., Tummon, F., Bodeker, G. E., Hubert, D., Petropavlovskikh, I., Steinbrecht, W., Anderson, J., Bhartia, P. K., Boone, C. D., Bourassa, A., Davis, S. M., Degenstein, D., Delcloo, A., Frith, S. M., Froidevaux, L., Godin-Beekmann, S., Jones, N., Kurylo, M. J., Kyrölä, E., Laine, M., Leblanc, S. T., Lambert, J.-C., Liley, B., Mahieu, E., Maycock, A., de Mazière, M., Parrish, A., Querel, R., Rosenlof, K. H., Roth, C., Sioris, C., Staehelin, J., Stolarski, R. S., Stübi, R., Tamminen, J., Vigouroux, C., Walker, K. A., Wang, H. J., Wild, J., and Zawodny, J. M.: Past changes in the vertical distribution of ozone – Part 3: Analysis and interpretation of trends, Atmos. Chem. Phys., 15, 9965-9982, doi:10.5194/acp-15-9965-2015, 2015.

Hewitt, H. T., Copsey, D., Culverwell, I. D., Harris, C. M., Hill, R. S. R., Keen, A. B., McLaren, A. J., and Hunke, E. C.: Design and implementation of the infrastructure of HadGEM3: the next-generation Met Office climate modelling system, Geosci. Model Dev., 4, 223-253, doi:10.5194/gmd-4-223-2011, 2011.

Hollandsworth, S. M., Bowman, K. P., and McPeters, R. D.: Observational study of the quasi-biennial oscillation in ozone, J. Geophys. Res.-Atmos., 100, 7347–7361, doi:10.1029/95JD00193, 1995.

Iglesias-Suarez, F., Young, P. J., and Wild, O.: Stratospheric ozone change and related climate impacts over 1850–2100 as modelled by the ACCMIP ensemble, Atmos. Chem. Phys., 16, 343-363, doi:10.5194/acp-16-343-2016, 2016.

Jones, A., Urban, J., Murtagh, D. P., Eriksson, P., Brohede, S., Haley, C., Degenstein, D., Bourassa, A., von Savigny, C., Sonkaew, T., Rozanov, A., Bovensmann, H., and Burrows, J.: Evolution of stratospheric ozone and water vapour time series studied with satellite measurements, Atmos. Chem. Phys., 9, 6055–6075, doi:10.5194/acp-9-6055-2009, 2009.

Keeble, J., Braesicke, P., Abraham, N. L., Roscoe, H. K., and Pyle, J. A.: The impact of polar stratospheric ozone loss on Southern Hemisphere stratospheric circulation and climate, Atmos. Chem. Phys., 14, 13705-13717, doi:10.5194/acp-14-13705-2014, 2014.

Keeble, J., Bednarz, E. M., Banerjee, A., Abraham, N. L., Harris, N. R. P., Maycock, A. C., and Pyle, J. A.: Diagnosing the radiative and chemical contributions to future changes in tropical column ozone with the UM-UKCA chemistry–climate model, Atmos. Chem. Phys., 17, 13801-13818, https://doi.org/10.5194/acp-17-13801-2017, 2017.

Leblanc, T. and McDermid, I. S.: Quasi-biennial oscillation signatures in ozone and temperature observed by lidar at Mauna Loa, Hawaii (19.5◦N, 155.6◦W), J. Geophys. Res.-Atmos., 106, 14869–14874, doi:10.1029/2001JD900162, 2001.

Lee, H. and Smith, A.: Simulation of the combined effects of solar cycle, quasi-biennial oscillation, and volcanic forcing on stratospheric ozone changes in recent decades, J. Geophys. Res.-Atmos., 108, D2, doi:10.1029/2001JD001503, 2003.

Mäder, J. A., Staehelin, J., Peter, T., Brunner, D., Rieder, H. E., and Stahel, W. A.: Evidence for the effectiveness of the Montreal Protocol to protect the ozone layer, Atmos. Chem. Phys., 10, 12161-12171, doi:10.5194/acp-10-12161-2010, 2010.

Manzini, E.: Atmospheric science: ENSO and the stratosphere, Nat. Geosci., 2, 749–750, 2009.

McLandress, C., Shepherd, T. G., Scinocca, J. F., Plummer, D. A., Sigmond, M., Jonsson, A. I., and Reader, M. C.: Separating the dynamical effects of climate change and ozone depletion, Part II: Southern Hemisphere troposphere, J. Climate, 24, 1850–1868, 2011.

Meul, S., Langematz, U., Oberländer, S., Garny, H., and Jöckel, P.: Chemical contribution to future tropical ozone change in the lower stratosphere, Atmos. Chem. Phys., 14, 2959–2971, doi:10.5194/acp-14-2959-2014, 2014.

Meul, S., Dameris, M., Langematz, U., Abalichin, J., Kerschbaumer, A., Kubin, A., and Oberländer-Hayn, S.: Impact of rising greenhouse gas concentrations on future tropical ozone and UV exposure, Geophys. Res. Lett., 43, 2919–2927, https://doi.org/10.1002/2016GL067997, 2016.

Molina, M. J. and Rowland, F. S.: Stratospheric sink for chlorofluoromethanes: chlorine atomcatalysed destruction of ozone, Nature, 249, 810–812, doi: 10.1038/249810a0, 1974.

Morgenstern, O., Braesicke, P., O'Connor, F. M., Bushell, A. C., Johnson, C. E., Osprey, S. M., and Pyle, J. A.: Evaluation of the new UKCA climate-composition model – Part 1: The stratosphere, Geosci. Model Dev., 2, 43-57, doi:10.5194/gmd-2-43-2009, 2009.

Morgenstern, O., Hegglin, M. I., Rozanov, E., O'Connor, F. M., Abraham, N. L., Akiyoshi, H., Archibald, A. T., Bekki, S., Butchart, N., Chipperfield, M. P., Deushi, M., Dhomse, S. S., Garcia, R. R., Hardiman, S. C., Horowitz, L. W., Jöckel, P., Josse, B., Kinnison, D., Lin, M., Mancini, E., Manyin, M. E., Marchand, M., Marécal, V., Michou, M., Oman, L. D., Pitari, G., Plummer, D. A., Revell, L. E., Saint-Martin, D., Schofield, R., Stenke, A., Stone, K., Sudo, K., Tanaka, T. Y., Tilmes, S., Yamashita, Y., Yoshida, K., and Zeng, G.: Review of the global models used within phase 1 of the Chemistry–Climate Model Initiative (CCMI), Geosci. Model Dev., 10, 639-671, https://doi.org/10.5194/gmd-10-639-2017, 2017.

Nair, P. J., Godin-Beekmann, S., Kuttippurath, J., Ancellet, G., Goutail, F., Pazmiño, A., Froidevaux, L., Zawodny, J. M., Evans, R. D., Wang, H. J., Anderson, J., and Pastel, M.: Ozone trends derived from the total column and vertical profiles at a northern mid-latitude station, Atmos. Chem. Phys., 13, 10373–10384, https://doi.org/10.5194/acp-13-10373-2013, 2013.

Newchurch, M. J., Yang, E.-S., Cunnold, D. M., Reinsel, G. C., Zawodny, J. M., and Russell, J. M.: Evidence for slowdown in stratospheric ozone loss: first stage of ozone recovery, J. Geophys. Res., 108, 4507, https://doi.org/10.1029/2003JD003471, 2003.

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

Pawson, S., Steinbrecht, W., Charlton-Perez, A J Fujiwara, M., Karpechko, A. Y., Petropavlovskikh, I., Urban, J., and Weber, M.: Update on Global Ozone: Past, Present, and Future, in: Scientific Assessment of Ozone Depletion: 2014, World Meteorological Organization, Global Ozone Research and Monitoring Project – Report No. 55, chap. 2, World Meteorological Organization/UNEP, 2014.

Plumb, R. A.: A 'tropical pipe' model of stratospheric transport, J. Geophys. Res., 101, 3957–3972, 1996.

Randel, W. J~~. and Wu, F.: TOMS total ozone trends in potential vorticity coordinates, Geophys. Res. Lett., 22, 683–686.~~, Wu, F., Russell, J. M., Waters, J. W., and Froidevaux, L.: Ozone and temperature changes in the stratosphere following the eruption of Mount Pinatubo, J. Geophys. Res., 100, 16753–16764, doi:10.1029/95JD01001, 1995.

Randel, W. J., Garcia, R. R., Calvo, N., and Marsh, D.: ENSO influence on zonal mean temperature and ozone in the tropical lower stratosphere, Geophys. Res. Lett., 36, L15822, doi:10.1029/2009GL039343, 2009.

Ravishankara, A. R., Daniel, J. S., and Portmann, R. W.: Nitrous Oxide (N_2O): The Dominant Ozone Depleting Substance Emitted in the 21st Century, Science, 326(123), 682-125, 2009.

Reinsel, G. C., Miller, A. J., Weatherhead, E. C., Flynn, L. E., Nagatani, R. M., Tiao, G. C., and Wuebbles, D. J.: Trend analysis of total ozone data for turnaround and dynamical contributions, J. Geophys. Res., 110, D16306, https://doi.org/10.1029/2004JD004662, 2005.

Revell, L. E., Bodeker, G. E., Huck, P. E., Williamson, B. E., and Rozanov, E.: The sensitivity of stratospheric ozone changes through the 21st century to N_2O and CH_4, Atmos. Chem. Phys., 12, 11309–11317, doi:10.5194/acp-12-11309-2012, 2012.

Rowland, F. S. and Molina, M. J.: Chlorofluoromethanes in environment, Rev. Geophys., 13, 1-35, 1975.

Solomon, S., Ivy, D. J., Kinnison, D., Mills, M. J., Neely, R. R., and Schmidt, A.: Emergence of healing in the Antarctic ozone layer, Science, 353, 269–274, doi:10.1126/science.aae0061, 2016.

SPARC: SPARC assessment of stratospheric aerosol properties (ASAP), Tech. Rep. WMO-TD No. 1295, WCRP Series Report No. 124, SPARC Report No. 4, Berrieres le Buisson, Cedex, 2006.

SPARC: SPARC Report on the Lifetimes of Stratospheric Ozone-Depleting Substances, Their Replacements, and Related Species, Edited by M. K. W. Ko, P. A. Newman, S. Reimann, and S. E. Strahan, SPARC Report No. 6, WCRP-15/2013, 2013.

Staehelin, J., Harris, N. R. P., Appenzeller, C., Eberhard, J., and Piechowski, M.: Observations of ozone trends, Rev. Geophys., 39, 231–290, 2001.

Steinbrecht, W., Froidevaux, L., Fuller, R., Wang, R., Anderson, J., Roth, C., Bourassa, A., Degenstein, D., Damadeo, R., Zawodny, J., Frith, S., McPeters, R., Bhartia, P., Wild, J., Long, C., Davis, S., Rosenlof, K., Sofieva, V., Walker, K., Rahpoe, N., Rozanov, A., Weber, M., Laeng, A., von Clarmann, T., Stiller, G., Kramarova, N., Godin-Beekmann, S., Leblanc, T., Querel, R., Swart, D., Boyd, I., Hocke, K., Kämpfer, N., Maillard Barras, E., Moreira, L., Nedoluha, G., Vigouroux, C., Blumenstock, T., Schneider, M., García, O., Jones, N., Mahieu, E., Smale, D., Kotkamp, M., Robinson, J., Petropavlovskikh, I., Harris, N., Hassler, B., Hubert, D., and Tummon, F.: An update on ozone profile trends for the period 2000 to 2016, Atmos. Chem. Phys., 17, 10675-10690, https://doi.org/10.5194/acp-17-10675-2017, 2017.

Stolarski., R. S., and Cicerone, R. J.: Stratospheric Chlorine - Possible Sink for Ozone, Canadian Journal of Chemistry-Revue Canadienne De Chimie, 52, 1610-1615, Doi 10.1139/V74-233, 1974.

Telford, P. J., Braesicke, P., Morgenstern, O., and Pyle, J.: 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.

Tie, X., and Brasseur, G.: The response of stratospheric ozone to volcanic eruptions: sensitivity to atmospheric chlorine loading, Geophys. Res. Lett., 22, 3035–3038, 1995.

van Loon, H. and Labitzke, K.: The influence of the 11-year solar cycle on the stratosphere below 30 km: A review, Space Sci. Rev., 94, 259–278, 2000.

Weatherhead, E. C., Reinsel, G. C., Tiao, G. C., Meng, X.-L., Choi, D., Cheang, W.-K., Keller, T., DeLuisi, J., Wuebbles, D. J., Kerr, J. B., Miller, A. J., Oltmans, S. J., and Frederick, J. E.: Factors affecting the detection of trends: Statistical considerations and applications to environmental data, J. Geophys. Res.-Atmos., 103, 17149–17161, https://doi.org/10.1029/98JD00995, 1998.

Weatherhead, E. C., Reinsel, G. C., Tiao, G. C., Jackman, C. H., Bishop, L., Frith, S. M. H., DeLuisi, J., Keller, T., Oltmans, S. J., Fleming, E. L., Wuebbles, D. J., Kerr, J. B., Miller, A. J., Herman, J., McPeters, R., Nagatani, R. M., and Frederick, J. E.: Detecting the recovery of total column ozone, J. Geophys. Res.-Atmos., 105, 22201\u201322210, https://doi.org/10.1029/2000JD900063, 2000.

Weatherhead, E. C and Andersen, S. B.: The search for signs of recovery of the ozone layer, Nature, 441, 39–45, 2006.

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

White, W. B. and Liu, Z.: Non-linear alignment of El Niño to the 11-yr solar cycle, Geophys. Res. Lett., 35, L19607, doi:10.1029/2008GL034831, 2008

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

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

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

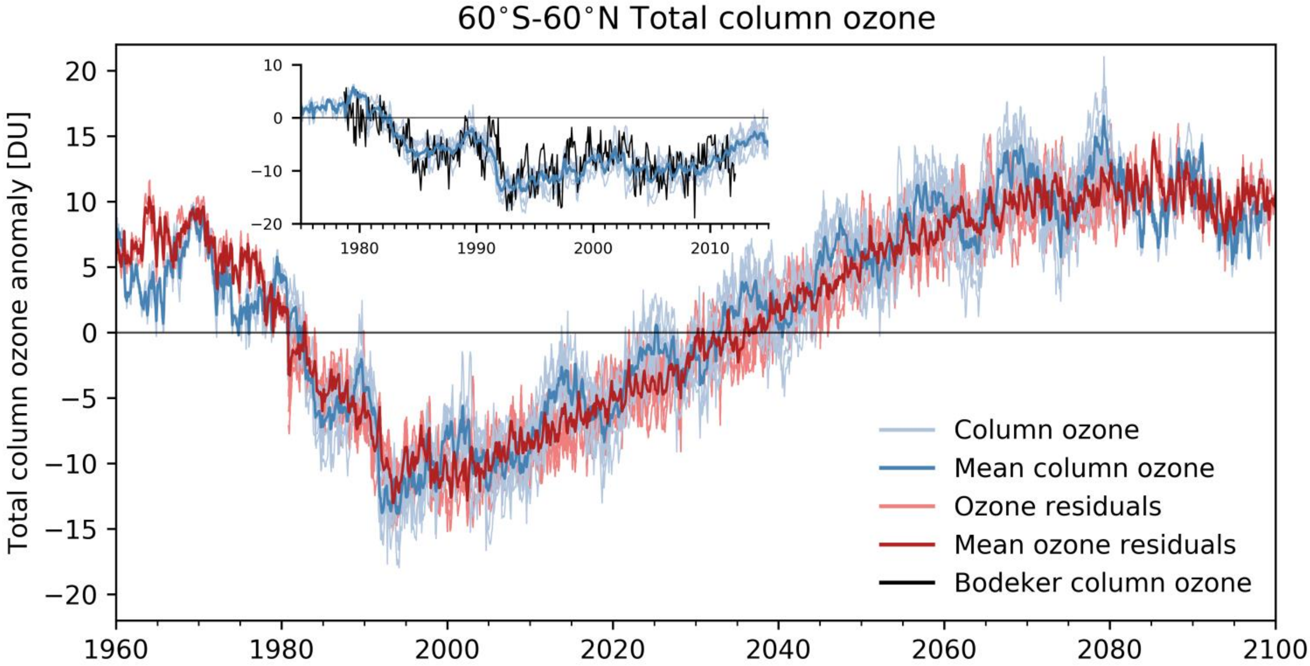


Figure 1: Deseasonalised total column ozone anomalies (in DU) relative to the 1980 mean, averaged over 60°S-60°N, for the seven UM-UKCA transient ensemble members (light blue lines) and ensemble mean (dark blue line). Also shown are the ozone residuals calculated when natural cycles are removed from each ensemble member (light red lines) and the mean of the ozone residuals (dark red line). The inset shows total column ozone anomalies for the transient UM-UKCA simulations and v2.8 of the Bodeker dataset (Bodeker et al., 2005; black line) from 1975 to 2015.

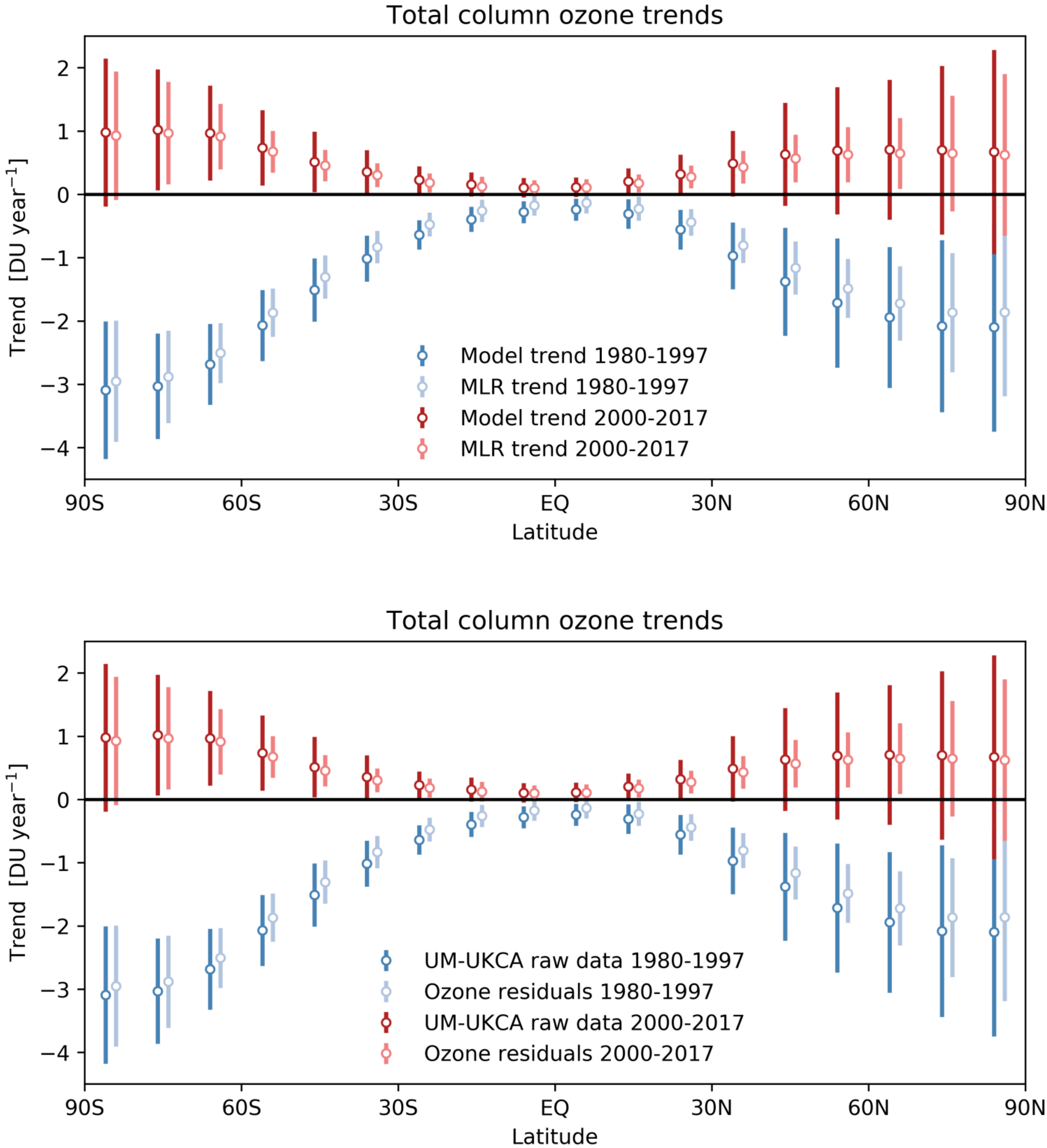
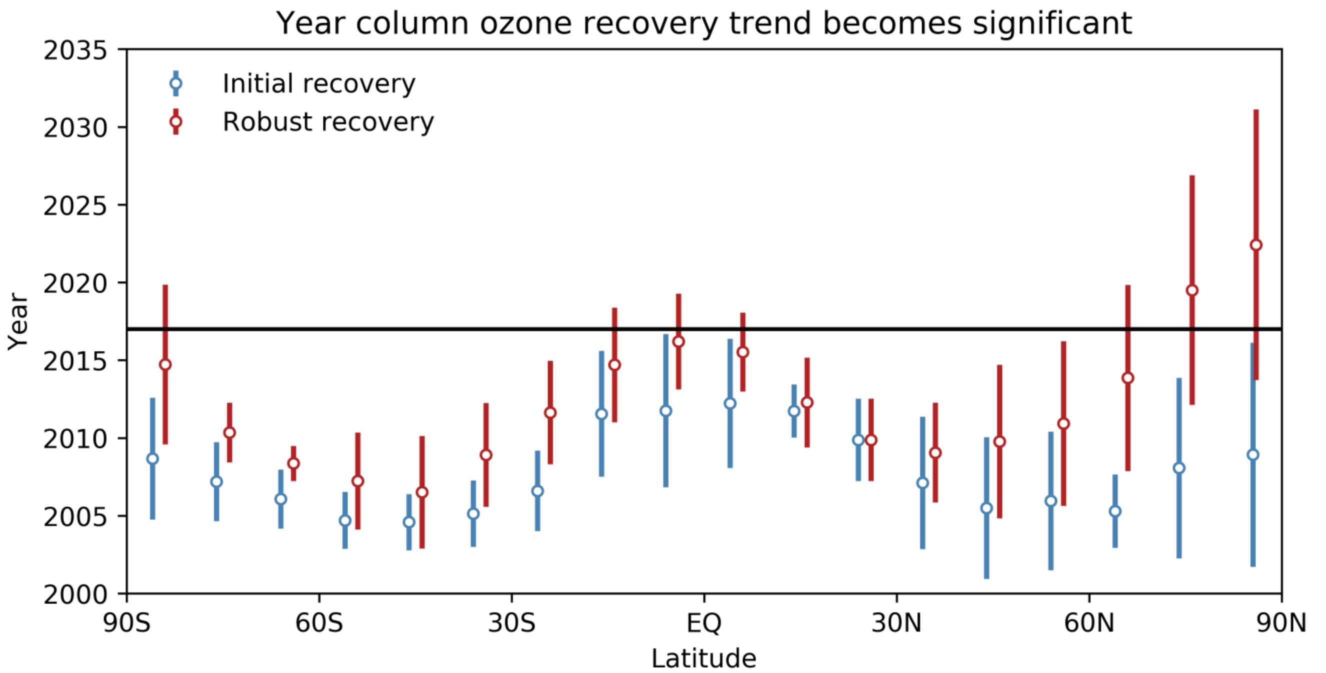


Figure 2: Ozone trends from 1980-1997 (blue points) and 2000-2017 (red points) for the raw ~~modelled~~UM-UKCA data (dark points) and ~~MLR~~ozone residuals calculated when the effects of natural cycles are removed (light points). Error bars associated with each trend represent the 95% confidence intervals, calculated as described in section 3.



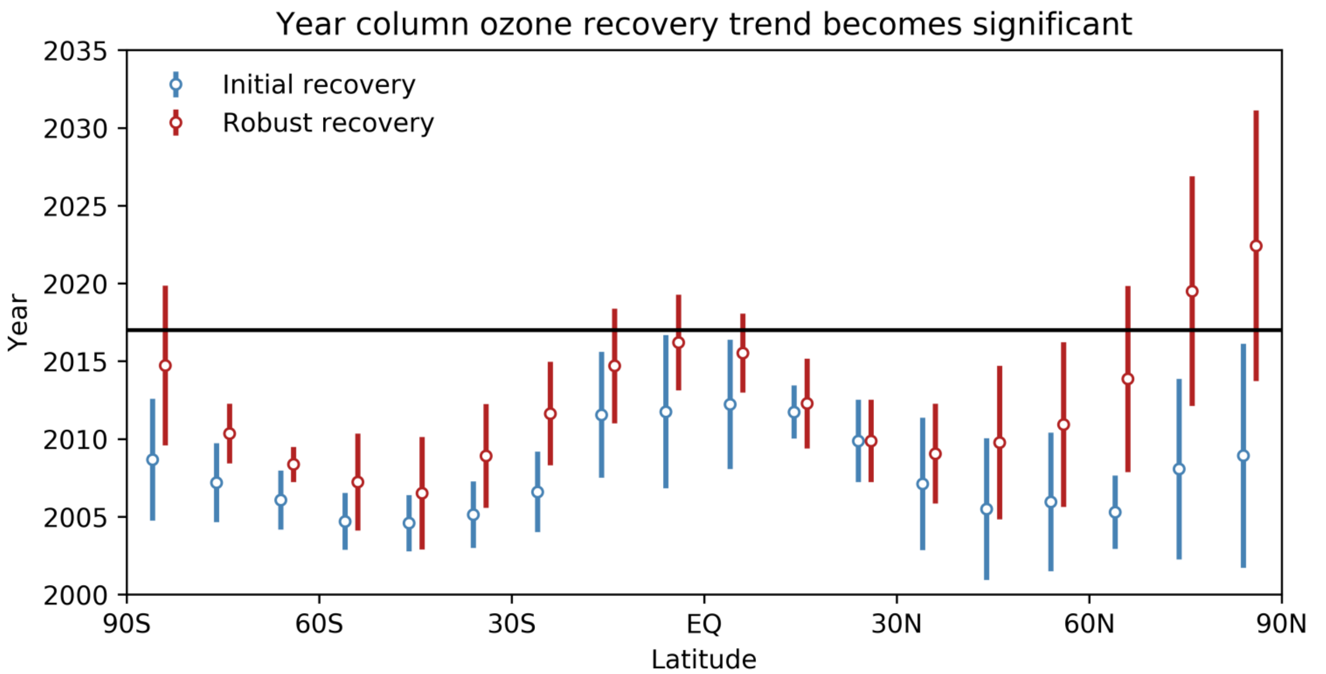


Figure 3: Year ~~when~~at which recovery trend ~~becomes significant, where trends are defined as significantly different from zero when they have a p-value < 0.05. Trend significance is determined for~~of the ozone residuals ~~calculated using the MLR coefficients, and a~~becomes significant. A distinction is made for the first time significance can be determined (blue points) and the time after which trends remain significant (red points). ~~Error bars represent the 95% confidence intervals.~~Here trend significance is calculated using the trend uncertainty obtained when all ensemble members are used, as discussed in section 3, but the trend magnitude is calculated for each ensemble member individually to reflect the impact of unaccounted for noise on the trend magnitude. Error bars represent the 95% confidence intervals, calculated as twice the standard deviation of the seven values obtained for the year trend significance is identified (one for each ensemble member)

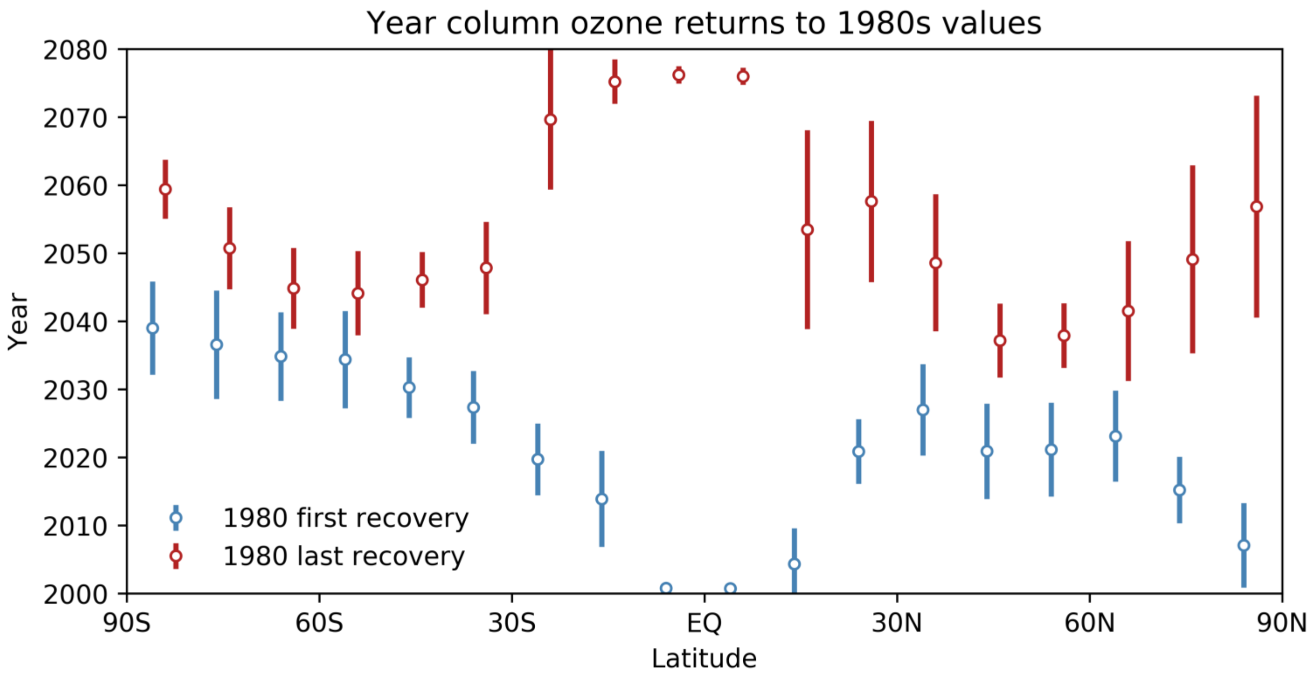


Figure 4: Year modelled total column ozone returns to 1980 annual mean values in each latitude band for the raw data from the seven UM-UKCA ensemble members. Blue points represent the first time annual mean values exceed the 1980 mean, while red points represent the final time annual mean values are lower than the 1980 mean. Error bars represent the 95% confidence intervals, calculated as twice the standard deviation of the return dates calculated for each ensemble member.