

Response to RC1 and Referee #3:

Referee #3's comments are shown with a vertical bar next to them, our comments are shown below each of them.

The Bando et al. manuscript presents a modified approach to detection and attribution that is able to account for the non-linear temporal behaviour of the forcing terms - in the case of stratospheric ozone analysed here, the rise in the atmospheric concentration of ozone depleting substances (ODSs) until the late 1990s followed by a slow decline. The authors clearly and convincingly present a comparison of their new approach with the more widely used approach of assessing signal-to-noise using linear trends and demonstrate the difficulties that arise when the magnitude of the forcing is not linearly changing with time. I have no significant concerns with the methodology or the presentation, though I will admit to having little background in detection and attribution.

My one, I believe relative inconsequential, concern is how the NAT-h timeseries was constructed. On Page 7, lines 6 – 8, the authors state:

‘In addition, we were able to isolate the response to volcanic aerosols, solar cycle, and the QBO, by differencing the sum of ensemble mean anomalies of the FIXED GHG1960 and ODS1960 simulations from the individual ensemble anomalies of ALL2.’

I can see how the arithmetic of the construction of NAT-h should work; by adding together the effects of GHGs (from the ODS1960 simulations) and ODSs (from the FIXED GHG1960 simulations) and removing these signals from the ALL2 timeseries. My concern is whether the volcanic aerosol effect will be correctly represented in the NAT-h timeseries. The response to large eruptions, such as Pinatubo, will depend critically on the concentrations of reactive chlorine in the stratosphere. Under the low chlorine loading of the ODS1960 simulation the increased volcanic aerosols will produce an increase in ozone in the mid-stratosphere, while under the higher chlorine loading of the FIXED GHG1960 the Pinatubo eruption will produce some increases in the mid-stratosphere and more significant decreases in the lower stratosphere. Given the spatial variability in the response, and regions of the atmosphere where the response to enhanced aerosols in the ODS1960 and FIXED GHG1960 simulations will be of opposite direction, I would think that it would be difficult to imagine that the actual representation of the effects of volcanic aerosols in the NAT-h timeseries would be correct.

I suggest this is probably not a significant concern because the region of the lower stratosphere from 100 to 40 hPa is below the region where volcanic aerosols have the largest impact on the reactive nitrogen chemistry, although it is the region where the opposing responses of halogen and nitrogen chemistry to aerosols is important. And I am not sure how such a relatively rare event as large volcanic eruptions would contribute to noise in long-term signals. Although I would like to stress that there are reasons to believe the NAT-h simulation is not a good representation of the effects of volcanic aerosols on ozone that one would find in a proper natural historical simulation.

Referee #3 raises an important and valid point related to the linear additivity of the stratospheric ozone response to large volcanic eruptions, and the issue of how the NAT-h simulation was constructed. In the low halogen loading stratospheric conditions of FIXED ODS1960, increases in mid-stratospheric ozone following a large volcanic eruption are expected as the loss of ozone at these altitudes is dominated by NO_x , due to enhancement of N_2O_5 hydrolysis (e.g. Tie and Brasseur, 1995). In the enhanced halogen loading conditions of FIXED GHG1960 and ALL2, the mid-stratospheric increase is limited to higher altitudes, and ozone depletion will occur throughout the lower stratosphere. As Referee #3 correctly points out, NAT-h is constructed by summing the ensemble mean ozone responses in FIXED ODS1960

and FIXED GHG1960, and then by subtracting these responses from the ALL2 realizations. We recognize that the stratospheric ozone response to volcanic aerosols may not be well represented in NAT-h in the vertical region where the reactive nitrogen chemistry is dominant. Our lower stratospheric region (from 40 to 100 hPa) may indeed encompass the region of where ozone concentrations are influenced by the competitive effects of halogen and nitrogen chemistry.

Potential errors in the ozone response to volcanic eruptions in NAT-h are now acknowledged in the revised manuscript. Below, we argue that such errors are unlikely to have substantial impact on our primary D&A results.

To identify slowly-evolving ODS and GHG signals in observational estimates of stratospheric ozone changes, it is critical to have information regarding the natural variability of ozone on timescales of 2-3 decades. It is the background multi-decadal “noise” that is of most interest here. As shown in our Figure 3, NAT-h (which was obtained by subtraction) reliably captures the observed amplitude of the decadal-timescale variability of lower stratospheric ozone arising from external solar forcing. This is the region we are most concerned with in addressing the additivity issue raised by Referee #3. NAT-h also replicates shorter-timescale (< 3 year) observed lower stratospheric ozone variability associated with the QBO and ENSO.

Errors of the form mentioned by Referee #3 (i.e., errors in the lower stratospheric ozone response to short-term, episodic volcanic forcing) are unlikely to affect our estimates of the ODSonly and GHGonly fingerprint patterns, the projections of the SWOOSH ozone data onto these fingerprints, or the estimated variability on multi-decadal timescales from NAT-h. Any errors in the simulated ozone response to volcanic forcing will have largest impact on the short-term variability shown on the x-axis in Fig. 3 – not on the longer multi-decadal variability that is most relevant for the signal-to-noise (S/N) calculations.

Figures 8 E and F help to illustrate the robustness of our results to potential errors in the response of ozone to volcanic forcing. Consider S/N results for the ODS signal in the lower stratosphere in the “tropics excluded” case. This is the case most relevant to the Referee’s concern, since the mid-latitudes are the regions where the opposing nitrogen/halogen chemistry would be important. For this particular example, S/N ratios for the full 33-year SWOOSH period are significant at the 5% level or better. To negate these significant results would require that noise trends are roughly 40% larger on the 33-year timescale. It is highly unlikely that an error of this magnitude could be caused by the ozone response to short-term volcanic forcing.

We fully agree that it would be preferable to directly estimate the ozone response to historical volcanic and solar external forcing – i.e., to have access to a simulation in which only volcanic and solar forcing are varied. Unfortunately, such a simulation was not available for our study – which is why we had to estimate the volcanic and solar responses by differencing simulations. We note that such differencing operations are not unique to our study. For example, in Gillett et al., 2011, their ODS response was estimated by differencing of combined forcing simulations and GHG single forcing simulations.

We thank Referee #3 for bringing this concern to our attention. We are confident that our attribution results would not differ significantly if we had access to a more comprehensive historical natural simulation (rather than estimating the volcanic and solar signals by subtraction). Nevertheless, the revised manuscript discusses Referee #3’s concern and provides a better description of how NAT-h was actually calculated.

Aside from this footnote to the NAT-h simulation, my other comments are minor and are given below.

- The first five paragraphs in the introductory text, over Pages 2 and 3, bounce around a bit from topic to topic, in particular the fourth paragraph (Page 3, Lines 1-9) that discusses linear trends for D&A in the middle of a general discussion of stratospheric ozone. Personally, I found it a bit difficult to follow the thread through the introduction.

We agree with Referee #3 and have reorganized the introductory text in the revised manuscript to ensure that the discussion of general attribution theory and stratospheric ozone changes are separated.

- Page 6, Lines 17-30 – The term ‘emission’ is frequently used through this section when I think the more accurate term would be concentration. For example, at Line 27 there is the statement ‘...while another keeps GHG emissions fixed at their 1960 conditions...’ which would suggest the atmospheric concentrations continued to increase after 1960.

Yes Referee #3 is absolutely correct, it should be concentration and not emission. We have changed all such occurrences in the main text.

- Page 6, Line 28 – Do the time-varying concentrations of ODSs in the FIXED GHG1960 simulations affect the radiative forcing of these coupled simulations?

Yes, the time varying concentrations of ODSs in FIXED GHG1960 are radiatively active. It is important, and has been made explicitly clear in the revised manuscript.

- Page 9, Lines Lines 11-18 – Here you argue for conducting the analysis of the lower stratosphere for both the global and extratropical regions. Given the discrepancy with SWOOSH observations for the tropics after 2005, I think this is fully warranted. My point would be about the possible explanation for the discrepancy in tropical ozone being related to the behaviour of volcanic aerosols. If the small volcanoes had significantly impacted ozone and was not properly accounted for in the database of specified aerosols used in the modelling, wouldn’t the effect be most pronounced in the extratropical lower stratosphere? Reactive chlorine levels would be much higher than in the tropical lower stratosphere and since many of these small eruptions were in the extratropics I would think the aerosols would also be more prevalent in the extratropical lower stratosphere. It must also be kept in mind that the agreement of different observational databases for the tropical lower stratosphere is not terribly good. See Section 2.2.4.3 of the WMO 2014 Ozone Science Assessment for the discussion of differences in post-2000 trends for the lower tropical stratosphere.

We have taken Referee #3’s advice, and substituted our speculation in the revised manuscript about the difference being due to volcanic eruptions (which as Referee #3 points out is not a reasoned hypothesis) with the difference possibly due to errors in the observational data in the tropical lower stratosphere. We thank Referee #3 for offering a more rational explanation.

- Page 12, Lines 13-18 – On the underestimation of variability in the upper stratosphere, part of the discrepancy may be due to observational uncertainty as different ozone datasets have some significantly different representations of the magnitude of the solar cycle – see Maycock et al., Atmos. Chem. Phys., 16, 10021-10043, 2016. Chemistry climate models also tend to have solar cycle variations in ozone that are towards the lower end of observational estimates – see Chapter 8.5 of SPARC CCMVal (2010) (SPARC Report on the Evaluation of Chemistry-Climate Models, V. Eyring, T. G. Shepherd, D. W. Waugh (Eds.), SPARC Report No. 5, WCRP-132, WMO/TD-No. 1526).

Referee #3 is correct: we failed to mention that the underestimation of variability may also be due to observational uncertainty, particularly in terms of known deficiencies in the ability of chemistry-climate models to capture ozone variations related to the solar cycle. We have included the reference provided by Referee #3 and added a brief discussion of this point.

- Page 16, Line 10 – the statement that ‘NAT-h is nudged to reanalysis temperature and wind fields.’ seems a bit misleading as it makes it sound like a ‘Specified Dynamics’ simulation where nudging is applied everywhere. The statement should be more specific to the nudging used here to produce the QBO.

We agree with Referee #3 that we did not properly describe how the QBO is implemented in WACCM. As with many comparable climate models, the QBO is imposed as an artificial forcing. This is achieved by nudging tropical stratospheric zonal-mean winds to either cyclic or fixed-phase winds, or to the observed winds. We have now revised and improved the description of how the QBO is forced in WACCM.

- Page 17, Line 8 – I am unclear what is meant by ‘...the noise data set $N(x,p,t)$, which is constructed by concatenating the NAT, NAT-h and CTL simulations.’ Does this mean a single timeseries was created by splicing all three of these simulations together, thus creating a ~1250 year timeseries? If so, how would the resulting timeseries be used with the S/N analysis that begins in 1984?

All of the natural and internal climate variability simulations, NAT (1000 years), NAT-h (3x50 years), and CTL (200 years) were concatenated together to form a 1350 year noise data set, $N(x,p,t)$. This was not that clear in the original manuscript. The revised manuscript contains a clarification of this concatenation procedure in Section 6.

- Page 21, Line 18 – there is an erroneous bracket at ‘.. with methods 1 and 2 (respectively.’

Corrected in text, thank you!

- Page 21, Line 28 – there is a word missing at ‘..method 2 yielded markedly S/N ratios...’

Corrected in text, thank you!

Response to RC2 and Referee #2:

Referee #2's comments are shown with a vertical bar next to them, our comments are shown below each of them.

This paper certainly uses a sledgehammer to crack a nut, I am afraid. It seems to construct a point of ozone trends assumed to be linear, even though the trend is nonlinear and by doing so seems to mix two slightly different types of non-linearity.

The first type of linearity (non-linearity) focusses on the temporal behaviour of the ozone time series. However, I do not know of anybody who assumes that the full behaviour of the time series is linear. This is exactly the reason why two carefully chosen periods are fitted with (independent) linear trends, or why people assume a fit to (E)ESC. In addition, non-linear (quadratic) terms have been used to consider early starting points (pre-1980) of the time series (most recently in Langematz et al., 2016, citing the earlier work as well), clearly acknowledging the complexity of the long-term trend.

The second type of linearity (non-linearity) focusses on the attribution problem. Are the attributed variabilities in ozone a sum of the different terms or not. MLR uses the implicit assumptions that the different factors are a sum, which is presumably a good approximation when different terms in the equation are largely independent of each other. However, it has been realised that this is not always the case and that we have to think carefully of how to choose our proxies (a lot of work is covering this question).

The paper does not clearly separate the two issues and in a way circumnavigates its own problems by choosing two different vertical regimes (nothing wrong with this). However, it would be interesting to know how the method would fair in a more holistic approach.

We appreciate Referee #2's comments above, and hope we can address the issue concerning linearity below. Before addressing specifically, the concerns about the temporal behavior of stratospheric ozone and attribution, we think it would aid to reiterate and clear up any misunderstanding to the detection and attribution (D&A) method used in this study. As addressed on pages 4 and 5 in the original manuscript, there are two established methodologies used in D&A studies:

1) The first one is the optimal regression methodology, which combines the spatial and temporal climate response into a single space-time vector, with the observations being modeled as a linear sum of the simulated responses to individual forcings (see e.g. Allen and Tett, 1999). Each response is scaled by a regression coefficient, expressing the strength of the space-time response pattern in observations. The underlying premise in this methodology is that the observations can be well represented as a linear combination of the input model signal response fields with an additive noise term due to internal variability. Thus, it is assumed that the response patterns to individual forcings are statistically separable and the sum of the responses is equivalent to the response obtained when all forcings are varied together. The important difference between this methodology and the one used in our study and described hereafter, is that it combines the spatial and temporal response into a single space-time vector. The optimal regression methodology was used by the only other formal stratospheric ozone attribution paper (Gillett et al., 2011). In their study, they could not separate the detectability of the ODS and GHG responses of stratospheric ozone change, from 1979-2005, but only the combined response. They hypothesized this was due to intrinsic degeneracy between ODS and GHG space-time patterns. Our study was motivated to use the second methodology for D&A, to investigate if we could confidently detect individually the ODS and GHG signals in observations, and compare their signal-to-noise ratios. The inherent drawback of the space-time optimal regression is that the method fails to separate the spatial and temporal components. It is difficult, therefore, to “deconstruct” the D&A results, and to determine

whether it is spatial and/or temporal correspondence between the model fingerprint and observations that yields positive identification of the fingerprint.

2) The second D&A methodology does not combine spatial pattern and time evolution into a single vector. Instead, it uses pattern similarity statistics to assess the time evolution of the spatial correspondence between the time-varying observations and time-invariant fingerprints (see e.g., Santer et al., 2003, 2013a,b). Fingerprint patterns are typically estimated from model simulations with individual or combined external forcings. Fingerprints are also compared with model-based estimates on natural variability. The key statistical question is whether the change over time in the spatial correspondence between an individual fingerprint and the observations is greater than the random correspondence between the fingerprint and realization of internal and natural climate variability. The underlying assumption in this approach (which we refer to as method 1) is that the fingerprint pattern does not change markedly as function of time – which we verified for stratospheric ozone by estimating the ODSonly and GHGonly fingerprints over different time periods. In most previous applications of method 1, it has been reasonable to assume that the anthropogenic signal component evolves quasi-linearly (Santer et al., 2003, 2013a,b). This assumption is not justifiable for the ODSonly signal, as Reviewer 2 correctly notes, and as we clearly discuss at multiple points in the text. Our modification of method 1 (which we refer to as method 2) directly addresses the non-linearity in the time evolution of the ODSonly signal. By comparing the signal-to-noise (S/N) ratios for method 1 and method 2, we show that for the non-linear ODSonly signal, there is a substantial enhancement of S/N in method 2. For the relatively linear GHGonly ozone signal, S/N ratios are very similar in method 1 and method 2.

In our opinion, it is scientifically valuable to compare and contrast the results obtained with purely linear and non-linear representation of signal evolution. We are not aware of other studies that have done this in a D&A context. We are comparing the efficacy of different D&A approaches – not “using a sledgehammer to crack a nut”.

Referee #2 states “*The first type of linearity (non-linearity) focusses on the temporal behaviour of the ozone time series. However, I do not know of anybody who assumes that the full behaviour of the time series is linear*”. We address the first critique related to fitting linear trends to the entire observational period. Evident in Figure 1 of the manuscript, is that linear regression of stratospheric ozone anomalies from 1984-2016 is not ideal with the SWOOSH, ALL1/2, and ODSonly anomalies. Although this is apparent, in our revised manuscript, we make clear that in order to describe the long-term changes for observations and certain simulations, a linear fit over the entire record would have large errors. However, in the model realm of GHGonly, a linear regression for global upper stratospheric ozone anomalies would indeed be appropriate to describe ozone changes over the entire time period. Indeed, in the cases where a linear fit is clearly inadequate, we could use piece-wise linear trends to describe changes in the depletion and recovery eras or use nonlinear methods such as one Referee #2 points out from Langematz et al., 2016. Many D&A studies assume that, to first order, an anthropogenically forced signal evolves linearly. That is clearly not the case with ozone. It is of interest to compare and contrast the D&A results obtained with a (sub-optimal) linear representation of the ODS signal and a representation capturing the non-linearity of the ODS signal. That is what we do here. Furthermore, unlike the piece-wise linear regression studies referred to by Referee #2, our method 2 does not require any subjective decisions to be made regarding the temporal boundary between the ozone depletion and ozone recovery periods.

In D&A studies, it is advantageous to use the full observational record. A longer observational record enables analysts to better characterize forced signals and unforced noise, and therefore improves the ability to separate an anthropogenically forced signal from internal or “total” natural climate variability. For most climate variables, the amplitude of internally generated variability decreases as the analysis period increases (see, e.g., our Figs. 7,8, and 10 in the manuscript). We therefore prefer to rely on the entire 33-year ozone record rather than on shorter, noisier segments of the record.

This is why in our Figure 4, we contrast linear and nonlinear (EESC) estimated changes 1984-2016, to show that changes associated with specified external forcings depend critically on whether the trends account for nonlinearities. Clearly fitting an EESC curve to the GHGonly simulation is not adequate, just as fitting a line to ODSonly from 1984-2016 – this is key to motivate exactly why we developed the nonlinear method in the attribution methodology. Unlike the trends shown in Figure 4, our nonlinear signal method does not depend on finding arbitrary proxies like EESC, or doing quadratic fits to proxies, as it uses information from the time evolution of the forcing in the model simulations. If the forcing is increasing linearly with time, as in the GHGonly case, there is no difference in detectability (signal-to-noise ratios) between using the traditional linear method to our nonlinear method, as shown in the manuscript. The additional purpose of the linear and EESC trends Fig 4, which Referee #2 takes issue with, is to compare the long-term change patterns with our model-derived fingerprints in Figures 5 and 6. The calculated fingerprints are not based on the regression techniques used in Fig 4., but shows that ODSonly spatial structure is qualitatively similar to the EESC regression, and the GHGonly spatial pattern is qualitatively similar to linear trends.

We agree with Referee #2, and have modified the text to further emphasize that past studies have examined long-term changes in stratospheric ozone and have used numerous methods to quantify such changes. We now cite studies that have used piece-wise linear trends to analyze changes over the ozone depletion and ozone recovery eras. As noted above, however, we seek to estimate the detectability of the ODS and GHG fingerprints over the entire observational record. Use of the entire record is beneficial for signal detection – we find that the noise amplitude is decreased by more than a factor of 2 by increasing the record length from 10 years to 30 years. This decrease in noise amplitude substantially increases our estimated S/N ratios. We have no arbitrary start/end points in our method 2, and we do not use any arbitrary proxies. For the ODSonly signal, we clearly show the value of the nonlinear signal method for D&A analysis relative to the traditional linear signal method.

With regard to the second issue of Referee #2 of the attribution problem of linearity, the referee is entirely correct that if we used a multiple linear regression (MLR) methodology to attribution (#1 above) there could be problems with the assumption that the individual response sum is not always equal to the response where all the forcings are varied together. However, we used methodology #2, so that we can compare the relative detectabilities of the GHG and ODS fingerprint patterns in observations and do not have this problem. Larger S/N ratios of one fingerprint pattern over another indicates that one pattern has a greater expression in observations, and the ratios can be compared and if they are detectable above the 95% confidence level from natural/internal variability. Related to Referee #2's concern; in our method, the spatial fingerprint patterns of ODS and GHG are not orthogonal to each other (linearly independent), as discussed in the manuscript. If two different forcings have identical fingerprints it would indeed be impossible to separate the two using the traditional linear signal method. Though the GHG and ODS do differ in their spatial distributions, even if they are not completely orthogonal, and we take into account the model-estimated time evolution in our attribution. The latter is key to distinguish between ODS and GHG in our signal-to-noise results. If Referee #2 is referencing in their second issue, not the attribution problem of linearity (which is answered above), but the linear additivity of the GHG and ODS signals, we present in Figure R1 global upper and lower stratospheric ozone anomalies to answer the question if the sum of the individual ODSonly and GHGonly ozone signals is equivalent to the ozone signal obtained when ODS and GHG concentrations are varied simultaneously? In WACCM simulations analyzed here, we see that to first order, linearity holds ($ODS_{only} + GHG_{only} \approx ALL2 - NAT-h$) for both the upper and lower stratosphere.

Referee #2 is correct that we should have recognized more coherently that many past studies have tackled the question of the nonlinear long-term changes in ozone to describe stratospheric changes, and we have added additional text to emphasize this. However, our study, shows precisely how this nonlinearity can be

used to our advantage in D&A work that can be applied to other cases in the climate system. There is a clear partitioning of the “linear additivity” issue and the “non-linear signal evolution” issue in the text of the paper. These issues are not convolved in our discussion. In regards to the comment about using a sledgehammer to crack a nut, this type of formal methodology is necessary if we want to be able to confidently ascertain whether human-caused imprints on the climate system are detectable. In an environment where there is still political debate regarding the reality of the human effect on global climate, this type of research may seem like a ‘sledgehammer’ but is needed to have sound science. We are using a published, extensively-tested D&A method. We have modified the method – in a relatively straightforward way – to deal with nonlinearity in the evolution of the ODSonly signal.

In summary, I believe the paper to be a nice little exercise in advanced statistics. It is certainly worth publishing after mayor revisions, but the paper needs to simplify its message should clearly acknowledge that the problem of linearity is well recognized (in both aspects – the temporal behaviour and the summing-up of contributing terms). Testing the limits of linear assumptions is always interesting, but it can be done in simple ways with idealised model simulations, alleviating the need for very fancy statistical models. However, I admit that this is a personal preference and that the paper will be a nice contribution to this discussion when revised.

We disagree with Referee #2, we are not using “*very fancy statistical models*”. As noted above, we are using a modification of a published, well-tested, and relatively straightforward D&A method. Space-time optimal detection is “*fancy*” – but that is not what we are doing here. The advantage of our method 1 and method 2 is that we preserve information about the spatial structure and time evolution of signal and noise (we don’t lump the spatial and temporal information into a single vector). This partitioning of spatial and temporal information is helpful in: 1) separating signal and noise; and 2) understanding the spatial and temporal differences between the ODSonly and GHGonly signals.

Some more specific comments:

- Abstract, line 26: One “the” too many . . .

Corrected, thank you

- Page 3, paragraph 1: strange discussion - non-linear versus piecewise linear (fit EESC), see comment above. This discussion and scoping of the paper needs to change most.

We have modified this section of the discussion related to how past studies have dealt with quantifying long-term changes in stratospheric ozone, including the reference Referee #2 provided.

- Page 4, line 14: I assume you talk about the absolute value, otherwise I suggest “regression coefficient significantly different from zero”.

Significantly greater than zero is correct here. A regression coefficient significantly greater than zero indicates a detectable response to the forcing, and one that is close to unity indicates that simulated and observed responses are similar in magnitude. A regression coefficient that is negative, indicates that the observed response is in the opposite direction as simulated, thus for detection in the optimal regression method, the lower limit of regression coefficient has to be greater than zero then the signal is said to be detected (see e.g. Hegerl et al., 1996).

- Having the NAT run with no QBO worries me slightly – the authors mention the fact, but I would hope for a slightly more critical assessment of this shortcoming, given that many people try to eliminate the QBO signal in their trend estimates.

It is a challenge to correctly simulate the QBO in general circulation models. In most models, the equatorial waves responsible for driving the QBO (Kelvin, Rossby gravity, inertial gravity, and mesoscale gravity waves) are not well-represented. As discussed in our paper, the QBO in WACCM is imposed as an artificial forcing. The different ways to force the QBO in WACCM are to nudge the tropical stratospheric zonal-mean winds to either fixed-phase, or observed winds. For NAT-h, the historical simulation with natural forcing only, the QBO phases are nudged to match observations from 1960-2016; after 2016, the QBO has a fixed phase through to 2099. The NAT simulation, which spanned the period from 850-1850, is forced by a historical reconstruction of solar variability, but does not have an imposed QBO. As Referee #2 correctly points out, it would be preferable to have QBO variability in the NAT simulation. We note, however, that: 1) we have no millennial-timescale information on fluctuations in the QBO; and 2) historical changes in the QBO are represented in the NAT-h simulation, which we use for estimating the noise in method 1 and method 2.

- You say: “. . . and there are post-2005 differences between the historical WACCM model simulations and SWOOSH data that are relevant to the interpretation of the D&A results.” I certainly agree. However, I would hope for a clearer discussion of what the implications are.

Because of the described differences in the post-2005 behavior of tropical lower stratospheric ozone in ALL1 and SWOOSH, we did partition our S/N analysis for two cases: a global domain and for one poleward of 30S and 30N (excluding the tropics). As discussed in the paper, the ODSonly fingerprint signal was found to be detectable with both inclusion and exclusion of the tropics, with higher S/N ratios if excluded.

Referee #2 is correct that in the conclusion/discussion we did not discuss the implications of the differences between post-2005 lower stratospheric ozone in ALL1 and SWOOSH, it was only discussed Section 3. We have now added such a discussion (see the new Section 7 of revised manuscript). A brief summary of the discussion is that because of the noticeable divergence between simulated and observed post-2005 ozone changes in the tropical lower stratosphere, inclusion of the tropics reduces S/N ratios for our method 2 (i.e., the temporal evolution of ozone differs over for the last decade in observations and in the ODSonly signal, thus reducing the regression coefficient in method 2). We showed that leaving out the tropical lower stratosphere (where the post-2005 divergence between simulations and observations is most pronounced) yields higher S/N ratios in Figure 8 of the manuscript. As pointed out by the other referee for the paper, another important motivation for performing the analysis with both the tropical lower stratosphere included and excluded, is that there is well-known disagreement of different observational databases for the tropical lower stratosphere - as the much lower abundances of ozone in that region lead to larger instrumental uncertainties. The WMO 2014 report discusses differences in post-2000 trends for the lower tropical stratosphere.

- You say: “The decadal variability is of key interest in D&A studies, since it constitutes . . .” What indications do we have that the modelled decadal variability is similar to the observed? Many models show distinct attenuations of amplitudes when free running (compared to SD runs). Is this of no concern for WACCM, or are there no sizeable differences for the free running model compared to the SD configuration?

A detailed paper on the comparison of WACCM’s climate and its variability can be found in Marsh et al., 2013. Referee #2 is correct that our D&A study relies on a single climate model, and we acknowledged the shortcomings of a “single model” analysis in the introduction and discussion. Model bias would affect

our results. To quantify the bias, and to ensure that the decadal variability in stratospheric ozone is close to observations, the original manuscript explicitly includes a section (3.2) that carried out such an analysis, so we refer back to that section.

- You say: “. . . simple linear regression line is not an adequate representation of ozone changes over the entire observational record (1984-2016).” As I mentioned above, nobody is stating this (any more, see comment above). Please clarify this.

Yes, as discussed earlier this has been clarified, as this was to motivate the development of our nonlinear signal method in the attribution section.

- You use spectral filters to construct a comparison of variability on different time scales. I would prefer simple power spectra comparing the variability. The filtering you do, makes me feel uncomfortable, give that you have one time window up-to 20 years with a time series of ~33 years.

The band-pass filter focuses on variability on the timescale of 10 years with half-power points at 5 and 20 years. Figure R2, included below, shows the response functions for the Butterworth band-pass and high-pass filters. This is the same filtering approach used by Santer et al., 2011, who examined variability of the temperature of the lower troposphere in model simulations and observations.

In our original submission, we included power spectra in the supplement Figure S2. As expected, the NAT simulation has peaks around 11 years for the solar cycle in the upper stratosphere, and NAT-h has peaks in the 2 to 3 year range related to the QBO.

- Figure 3: typo in title

This has been fixed, thank you!

Langematz, U., Schmidt, F., Kunze, M., Bodeker, G. E., and Braesicke, P.: Antarctic ozone depletion between 1960 and 1980 in observations and chemistry–climate model simulations, *Atmos. Chem. Phys.*, 16, 15619-15627, <https://doi.org/10.5194/acp-16-15619-2016>, 2016.

Response Figures:

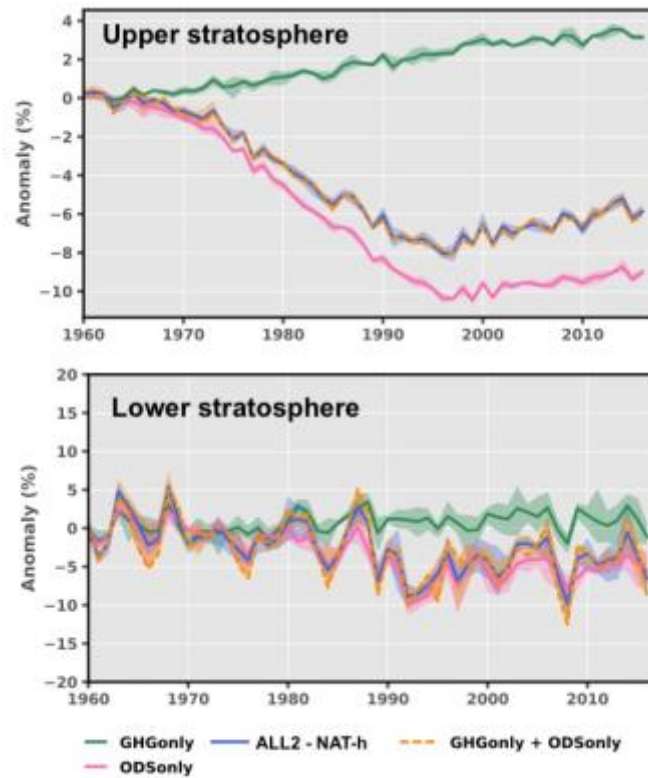


Figure R1: Global ozone anomalies in the upper and lower stratosphere, similar to Figure 1 in the manuscript, but here we subtract the NAT_h response from ALL_2 to investigate the agreement between GHGonly+ODSonly and ALL2 - NAT_h.

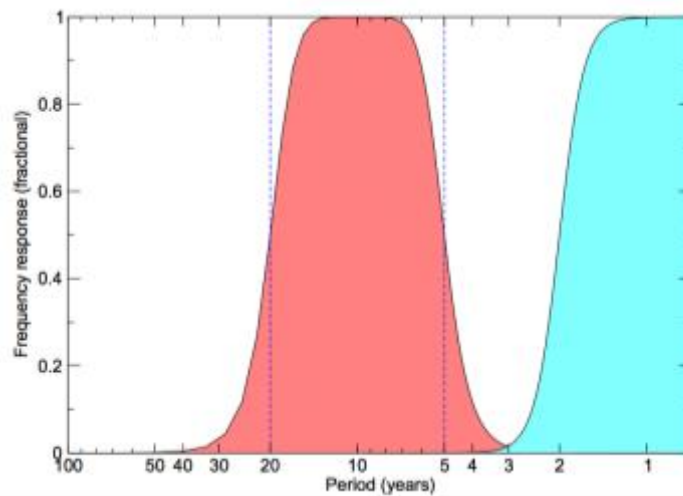


Figure R2: Response functions for Butterworth high-pass and band-pass filters used in the main text.

Response References:

Allen, M. R. and Tett, S. F. B.: Checking for model consistency in optimal fingerprinting, *Clim. Dyn.*, 15(6), 419–434, doi:10.1007/s003820050291, 1999.

Gillett, N. P., Akiyoshi, H., Bekki, S., Braesicke, P., Eyring, V., Garcia, R., Karpechko, A. Y., 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(2), 599–609, doi:10.5194/acp-11-599-2011, 2011.

Hegerl, G. C., von Storch, H., Hasselmann, K., Santer, B.D., Cubasch, U., and Jones, P.D.: Detecting anthropogenic climate change with an optimal fingerprint method. *J. Climate*, 9, 2281–2306, 1996.

Marsh, D. R., Mills, M. J., Kinnison, D. E., Lamarque, J.-F., Calvo, N. and Polvani, L. M.: Climate Change from 1850 to 2005 Simulated in CESM1(WACCM), *J. Clim.*, 26(19), 7372–7391, doi:10.1175/JCLI-D-12-00558.1, 2013.

Santer, B., Wigley, T. and Jones, P.: Correlation methods in fingerprint detection studies, *Clim. Dyn.*, 1993.

Santer, B.D., Taylor, K.E., Wigley, T.M.L., Penner, J.E., Jones, P.D., and Cubasch, U.: Towards the detection and attribution of an anthropogenic effect on climate, *Climate Dynamics* 12(77). <https://doi.org/10.1007/BF00223722>, 1995.

Santer, B. D., Mears, C., Doutriaux, C., Caldwell, P., Gleckler, P. J., Wigley, T. M. L., Solomon, S., Gillett, N. P., Ivanova, D., Karl, T. R., Lanzante, J. R., Meehl, G. A., Stott, P. A., Taylor, K. E., Thorne, P. W., Wehner, M. F. and Wentz, F. J.: Separating signal and noise in atmospheric temperature changes: The importance of timescale, *J. Geophys. Res. Atmos.*, 15 116(D22), doi:10.1029/2011JD016263, 2011.

Santer, B. B. D., Painter, J. F. J., Mears, C. a, Doutriaux, C., Caldwell, P., Arblaster, J. M., Cameron-Smith, P. J., Gillett, N. P., Gleckler, P. J., Lanzante, J., Perlwitz, J., Solomon, S., Stott, P. a, Taylor, K. E., Terray, L., Thorne, P. W., Wehner, M. F., Wentz, F. J., Wigley, T. M. L., Wilcox, L. J. and Zou, C.-Z.: Identifying human influences on atmospheric temperature., *Proc. Natl. Acad. Sci. U.S.A.*, 110(1), 26–33, doi:10.1073/pnas.1210514109, 2013a.

Santer, B. D., Painter, J. F., Bonfils, C., Mears, C. A., Solomon, S., Wigley, T. M. L., Gleckler, P. J., Schmidt, G. A., Doutriaux, C., Gillett, N. P., Taylor, K. E., Thorne, P. W. and Wentz, F. J.: Human and natural influences on the changing thermal structure of the atmosphere., *Proc. Natl. Acad. Sci. U. S. A.*, 110(43), 6–11, doi:10.1073/pnas.1305332110, 2013b.

Detectability of the Impacts of Ozone Depleting Substances and Greenhouse Gases upon Stratospheric Ozone Accounting for Nonlinearities in Historical Forcings

5 Justin Bandoro¹, Susan Solomon¹, Benjamin D. Santer², Douglas E. Kinnison³, Michael J. Mills³

¹Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139

²Program for Climate Model Diagnosis and Intercomparison (PCMDI), Lawrence Livermore National Laboratory, Livermore, CA 94550

10 ³Atmospheric Chemistry Observations and Modeling Laboratory, National Center for Atmospheric Research, Boulder, CO 80307

Correspondence to: Justin Bandoro (jbandoro@mit.edu)

Abstract. We perform a formal attribution study of upper and lower stratospheric ozone changes using observations together with simulations from the Whole Atmosphere Community Climate Model. Historical model simulations were used to estimate the zonal-mean response patterns (“fingerprints”) to combined forcing by ozone depleting substances (ODS) and well-mixed greenhouse gases (GHG), as well as to the individual forcing by each factor. Trends in the similarity between the searched-for fingerprints and homogenized observations of stratospheric ozone were compared to trends in pattern similarity between the fingerprints and the internally and naturally generated variability inferred from long control runs. This yields estimated signal-to-noise (S/N) ratios for each of the three fingerprints (ODS, GHG, and ODS+GHG). In both the upper stratosphere (defined in this paper as 1 to 10 hPa) and lower stratosphere (40 to 100 hPa), the spatial fingerprints of the ODS+GHG and ODS only patterns were consistently detectable not only during the era of maximum ozone depletion, but also throughout the observational record (1984-2016). Furthermore, we also develop a fingerprint attribution method to account for forcings whose time evolutions are markedly nonlinear over the observational record. When the nonlinearity of the time evolution of the ODS and ODS+GHG signals are accounted for used in the trend regression, we find that the S/N ratios obtained with the stratospheric ODS and ODS+GHG fingerprints are enhanced relative to standard linear trend analysis. Use of the nonlinear signal detection method also reduces the detection time - the estimate of the date at which ODS and GHG impacts on ozone can be formally identified. Furthermore, by explicitly considering nonlinear signal evolution With this method, the complete observational record can be used in the S/N analysis, without applying piece-wise linear regression and introducing arbitrary break points. The GHG-driven fingerprint of ozone changes was not statistically identifiable in the either the upper or lower stratospheric SWOOSH data, irrespective of the method used. Use of the nonlinear signal method, instead of directly operating on ozone trends, also reduces the detection time—the estimate of the date at which ODS and GHG impacts on ozone can be formally identified. The GHG-driven fingerprint of ozone changes was not statistically identifiable in either the upper or lower stratospheric SWOOSH data, irrespective of the signal detection method used. In the WACCM future simulations of future climate change, the GHG signal is statistically identifiable between 2020-2030. Our findings demonstrate the importance of

continued stratospheric ozone monitoring to improve estimates of the contributions of ODS and GHG forcing to global changes in stratospheric ozone.

1 Introduction

- 5 Climate change detection and attribution (“D&A”) studies ~~are concerned with~~ seek to identify and formally quantify an anthropogenic component of change in observed climate data. Formal identification of an anthropogenic climate change “fingerprint” has been successfully achieved with observations of atmosphere and ocean temperatures, sea level, ocean acidity, various components of the water cycle and the cryosphere, and certain climate extremes (Bindoff et al., 2013). To date, however, few formal D&A methods have been applied in studies involving stratospheric ozone (see Gillett et al. 2011 for one
10 exception to this). There is evidence that stratospheric ozone is transitioning from an era of widespread and readily detectable depletion (linked to changes in anthropogenic chlorofluorocarbons) to an era characterized by early signs of recovery or healing (Solomon et al., 2016). Our motivation for this work is to determine whether formal D&A methods can provide a more confident and quantitative attribution of ozone depletion and recovery signals.
- 15 Global changes in the physical climate system are driven by both internal variability and external influences (Hegerl et al., 2007; Karl et al., 2006). Internal variability is generated through complex interactions of the coupled atmosphere–ocean system. External influences include human-caused changes in well-mixed greenhouse gases (GHG), ozone depleting substances (ODS), and other radiative forcing agents, as well as natural fluctuations in solar irradiance and volcanic aerosols. Past D&A studies have found that each of these external influences has a unique “fingerprint” in the detailed zonal-mean
20 latitude/altitude pattern of temperature change (Hansen et al., 2005; Karoly et al., 1994; Santer et al., 1996a; Tett et al., 1996; Thorne et al., 2002; Vinnikov et al., 1996). The use of such profiles of atmospheric temperature change has proved particularly useful in separating human, solar, and volcanic influences on climate, and in discriminating between externally forced signals and internal variability. In this study, we use the latitude/altitude patterns of both upper and lower stratospheric ozone change in response to individual and combined anthropogenic forcings to understand the relative detectability of ODS and GHG
25 signals in observations.

Stratospheric ozone depletion has been a significant international concern since it was first recognized as a consequence of anthropogenic emissions of ODS (Molina and Rowland, 1974). Following implementation of the Montreal Protocol, there was a decline in emissions of ODS and a consequent decrease in halogen-containing compounds in the stratosphere. In the 2014
30 World Meteorological Organization (WMO) Scientific Assessment of Ozone Depletion (WMO, 2014), a major topic of interest was the statistical significance of ozone trends over the last decade, and the extent to which “ozone recovery” is taking place. Assessing the statistical significance of observed ozone trends does not yield definitive information on the causes of trends. In

addition to determining whether observed negative or positive trends in ozone are unusually large or small relative to model-based estimates of unforced trends in ozone, it is also important to understand and quantify the contributions of different climate forcings to the observed ozone changes, and to assess whether the agreement between an externally forced fingerprint and observations could have been obtained by natural causes alone.

5

~~The use of least squares linear trends was reasonable for quantifying stratospheric ozone loss during the depletion era. For analyses spanning both the depletion and “ozone recovery” periods, purely linear trends are inadequate for capturing the more complex nonlinear evolution of ozone change in certain parts of the stratosphere (as we will show later). Many ozone trend studies address this nonlinear behavior by performing piecewise linear regression with a break point around 1997 (Bourassa et al., 2014; Chehade et al., 2014; Jones et al., 2009; Kyrölä et al., 2013; Laine et al., 2014). The slopes of the piecewise trends are not constrained by physical and chemical considerations, and are typically arbitrarily chosen to enhance the slopes for each of the eras. Here we employ a rigorous and consistent approach for attribution over the entirety of the observational record. Our approach accounts for temporal changes in both the ODS and GHG forcing time series.~~

10

15 There is a clear scientific consensus that man-made chlorofluorocarbons were the dominant driver of global stratospheric ozone decline from the 1970s to 2000s (e.g., Solomon, 1999 and references therein). It is also well established that increasing greenhouse gases contribute to a ~~cooling~~EE of the middle to upper stratosphere (Boville, 1986; Fels, 1980). Colder temperatures result in a slowing down of gas-phase catalytic reactions that destroy ozone in the upper stratosphere, which in the absence of any other changes, leads to ozone increases. Thus, both decreasing ODS and increasing GHG concentrations
20 can act to increase upper stratospheric ozone. Understanding the extent to which GHG increases may confound the attribution of ozone changes due to ODS is essential to the identification of ozone recovery at these altitudes (see WMO, 2014).

20

Another potential confounding factor is change in the Brewer-Dobson circulation (BDC), the meridional overturning circulation in the middle atmosphere that transports trace gases from the tropics to the poles (Brewer, 1949; Dobson, 1956).

25 Previous studies have found that a strengthening of the BDC results in lower ozone concentrations in the tropical lower stratosphere and an increase in concentrations in the extratropics, and modeling studies have suggested that increases in GHG concentrations lead to a strengthening of the BDC (Eyring et al., 2010; Fleming et al., 2011; Garcia et al., 2008; Gillett et al., 2011; Oman et al., 2010). In contrast, a recent study by Polvani et al. (2017) found that trends in ODS, and not in GHG levels, have been the primary driver of trends in tropical upwelling. This complicates the problem of attributing ozone changes that
30 are primarily BDC-related, since both GHG and ODS forcing may be implicated in driving changes in the BDC.

30

Another forcing that can affect the concentration of ozone in the stratosphere is the solar cycle. The total solar irradiance changes by about 0.1% during the 11 year solar cycle; however, UV radiation can change by about 4-8% (Lean, 2000). Stratospheric ozone changes during the solar cycle have been well documented (e.g., Merkel et al., 2011) with the largest

percentage change in ozone between solar maximum and minimum occurring between 30-45 km. Stratospheric ozone is also influenced by internal variability, primarily via the quasi-biennial oscillation (QBO). The effect of the QBO on upper stratospheric ozone has been shown to be small compared to its larger signal in interannual ozone variability in the lower and middle stratosphere (Hasebe, 1994; Zawodny and McCormick, 1991). These natural and internal changes must also be considered in attributing ozone changes to specific causes.

Most D&A studies have been focused on the attribution of changes in the climate system to anthropogenic GHG forcing, which has been increasing linearly since the early 20th century. In contrast, ODS forcing has had a unique nonlinear time evolution because of the implementation of the Montreal Protocol. For analyses spanning both the depletion and “ozone recovery” periods, purely linear trends are inadequate for capturing the more complex nonlinear evolution of ozone change in certain parts of the stratosphere (as we will show later). Many ozone trend studies address this nonlinear behavior by performing piecewise linear regression with a break point around 1997 (Bourassa et al., 2014; Chehade et al., 2014; Jones et al., 2009; Kyrölä et al., 2013; Laine et al., 2014). The slopes of the piecewise trends are not constrained by physical and chemical considerations, and are typically arbitrarily chosen to enhance the slopes for each of the eras. Another approach is to quantify changes over the entire record by regression to a nonlinear proxy, such as the halogen loading in the stratosphere, described by the equivalent effective stratospheric chlorine (EESC, see e.g. Newchurch, 2003). Many studies have employed EESC-based regression to quantify trends in ozone (e.g. Langematz et al., 2016; Stolarski et al., 2006a; Wohltmann et al., 2007). This approach utilizes the entire observational record, with no arbitrary start or end points. While EESC-based regression offers a number of advantages for the detection of nonlinear ODS-driven ozone changes, it also has certain disadvantages (Kuttippurath et al., 2015). In section 4, we use EESC-based regression to address the problem of nonlinear ODS evolution over the full observational record. Our D&A approach also accounts for the much smaller temporal changes in the GHG forcing time series.

To the best of the authors knowledge, there is only one previous study that has looked at considered the formal attribution of anthropogenic emissions to observed stratospheric ozone changes: Gillett et al. (2011). Their study investigated zonal-mean Solar Backscatter Ultraviolet (SBUV) ozone measurements (McLinden et al., 2009) over the 27-year period from 1979 to 2005 (27 years) in the middle and upper stratosphere (50-1 hPa). Model results were taken from simulations performed as part of the Chemistry-Climate Model Validation activity (CCMVal, Eyring et al. 2010). Gillett et al. analyzed multi-model simulations performed with greenhouse gases only, combined anthropogenic factors only, and combined anthropogenic and natural external forcings. The ozone response to changes in ODS and the response to changes in natural external forcings were estimated through subtraction of simulations. Gillett et al. (2011) used a standard space-time optimal regression methodology for signal detection and attribution (see, e.g., Allen and Tett, 1999). In this approach, the observations are modeled as a linear sum of simulated responses (“fingerprints”) to individual forcings, with each response scaled by an estimated regression coefficient (expressing the strength of the space-time response pattern in observations). A regression coefficient significantly greater than zero indicates a detectable response to the forcing, and a coefficient close to unity signifies that simulated and

observed responses are similar in magnitude (attribution). The underlying premise here is that the observations can be well-represented by a linear combination of the input model signal response fields and an additive noise term due to internal climate variability. It is also assumed that the response patterns to different individual forcings are statistically distinct (i.e., are separable), and that the sum of the individual responses is equivalent to the response obtained when all forcings are varied simultaneously.

Gillett et al. (2011) were unable to separate the individual ODS and GHG responses in the SBUV ozone data, but found a clear combined anthropogenic signal that was consistent with observations. In their study, the authors hypothesized that the difficulty of separating the individual ODS and GHG responses was due to multiple factors: the limited ensemble size, and the degeneracy between the patterns of stratospheric ozone response to ODS and GHG forcing. Since the optimal regression methodology combines the spatial and temporal response into a single space-time vector, it was not possible to determine whether the degeneracy between the ODS and GHG response patterns was primarily due to spatial similarity or to similarity in temporal evolution.

We use a different approach here to understand and quantify the relative contributions (and detectability) of the ozone responses to ODS and GHG forcing. Rather than combining spatial pattern and time evolution information in a single vector, we use pattern correlations to assess the time evolution of the spatial similarity between time-invariant fingerprints and time sequences of: 1) observed ozone patterns; and 2) model-based estimates of the natural internal climate variability of ozone. Such methods rely on some form of spatial covariance statistic (e.g., Santer et al., 1993, 1995), and may involve rotation of the fingerprint in a low-noise direction in order to optimize signal-to-noise ratios (Hasselmann, 1993; Hegerl et al., 1996). The conventional strategy in this method is to search for a long-term, positive trend in the pattern correspondence statistic, which would indicate an increasing expression of the searched-for signal in the observations. The underlying assumption here is that the spatial pattern of the fingerprint does not change markedly as a function of time, which is a reasonable assumption for historical GHG forcing, but is probably not reasonable for ODS forcing (see Solomon et al., 2016).

Our study differs from that of Gillett et al. (2011) in a number of ways. In addition to using a different D&A method and a longer ozone dataset, we are also addressing different statistical questions. First, we seek to determine whether the observed changes in stratospheric ozone are unusual relative to estimates of both internal climate variability and total natural variability (Santer et al., 2012, 2013). Second, we consider the time evolution of signal-to-noise (S/N) ratios, both for the individual fingerprints of ODS-forced and GHG-forced ozone changes, as well as for the ozone fingerprint arising from combined ODS+GHG forcing. This allows us to determine the detection time – the time at which S/N exceeds and remains consistently above some stipulated significance level. Third, we calculate S/N ratios separately for ozone changes in the upper and lower stratosphere rather than for a single region only (the 1 to 50 hPa region considered by Gillett et al.). The reasons for this decision are explained below. Fourth, because we do not combine spatial and temporal information, the spatial aspects of S/N

behavior are easier to evaluate (e.g., in terms of the pattern similarity between the searched-for fingerprints and the observations).

There are several disadvantages of our selected approach. Unlike Gillett et al. (2011), who performed a multi-model analysis, we rely on a single model only, and are unable to evaluate the sensitivity of our results to possible errors in the model-based estimates of internal variability and the response to external forcing (North et al., 1998). Additionally, since we do not explicitly incorporate time evolution information in the searched-for fingerprints, pronounced differences in the time evolution of the ozone responses to ODS and GHG forcing cannot be used in the separation of these responses (at least not in the “linear trend” representation of these signals).

As noted above, we perform separate D&A analyses for ozone changes in the upper and lower stratosphere. Our premise is that different dominant processes control the changes in ozone concentrations in these two regions ~~have differing dominant processes controlling changes~~. In the lower stratosphere, particularly in the tropics and extratropics, ozone concentrations are affected by direct transport, both through interannual variability and via changes in the BDC (Shindell and Grewe, 2002). A further influence on lower stratospheric ozone concentrations occurs by means of heterogeneous chemistry on the surfaces of polar stratospheric clouds and volcanic aerosols; the heterogeneous reactions are extremely sensitive to temperature changes (Solomon, 1999). The ozone chemical lifetime is on the order of weeks in the lower stratosphere in the Antarctic, and on the order of months at mid-latitudes.

In the upper stratosphere, the processes controlling ozone abundance are quite different; the photochemical lifetime of ozone is on the order of hours, so that direct transport of ozone is not important. The local concentrations of upper stratospheric ozone are determined by gas-phase chemistry, which is well understood and well replicated in current chemistry-climate models (Austin et al., 2002). Transport of other greenhouse gases (such as methane) can indirectly affect ozone concentrations in the upper stratosphere (Solomon and Garcia, 1984). The anticipated response from GHG forcing radiatively cools the upper stratosphere uniformly with latitude (Aquila et al., 2016), increasing ozone concentrations, and potentially impacting atmospheric circulation through the BDC. A strengthening of the BDC would result in lower stratospheric ozone decreases in the tropics and increases in the extratropics.

Based on this understanding from atmospheric physics and chemistry, we expect that different forcings should yield distinctly different latitudinal and vertical patterns of change. Below, we examine in detail the use of these patterns for attribution of the causes of ozone changes in the upper and lower stratosphere.

2 Model Simulations and Observed Ozone Datasets

The model used for all simulations analyzed in this study is version 1 of the Community Earth System Model, with version 4 of the Whole Atmosphere Community Climate Model as the atmospheric component (CESM1(WACCM); see Marsh et al. 2013). WACCM is a coupled chemistry-climate model that extends from the Earth's surface to the lower thermosphere.

WACCM's representation of heterogeneous chemistry has been shown to be in broad agreement with observations of polar ozone and related chemical species (Solomon et al., 2015). A wide range of experiments were performed with WACCM:

i) Free-running simulations that follow the experimental design of the Chemistry-Climate Model Intercomparison (CCMI) REFC1 experiment (Morgenstern et al., 2017), with volcanic aerosols prescribed according to CCMI (Arfeuille et al., 2013).

These simulations include coupled chemistry and dynamics (Marsh et al., 2013) with sea surface temperatures prescribed according to observations for the period 1955-01 to 2014-12. Emissions-Concentrations of ozone depleting substances and greenhouse gases vary over time, as specified in the REFC1 scenario. A set of 5 ensemble members was available, each with identical forcings but starting from slightly differing initial conditions. We refer to these simulations subsequently as ALL1.

ii) Free-running simulations that follow the same REFC1 scenario as in ALL1, with the same atmospheric coupled chemistry and dynamics, but including an interactive fully-coupled ocean component. These simulations begin in 1960-01 and extend through 2099-12. One set of runs includes both historical time-varying greenhouse gases and ozone depleting substance emissionsconcentrations (referred to herein as ALL2), while another keeps GHG emissionsconcentrations fixed at their 1960 conditions but allows the evolution of time-varying ODS concentrations (referred to as FIXED GHG1960). A third set complements the second with ODS concentrationsemissions fixed at 1960 conditions but with time-varying GHG concentrationsemissions (FIXED ODS1960). Each of the aforementioned sets of experiments contains 3 ensemble members that were performed with slightly different initial conditions. FIXED GHG1960 and FIXED ODS1960 follow the convention of the CCMI REFC2 experiments (Eyring et al., 2013). It should be noted that the time-varying concentrations of ODSs in FIXED GHG1960 are radiatively active.

All sets of simulations in *i)* and *ii)* include identical historical external forcings from solar insolation changes and prescribed CCMI volcanic aerosol evolution. As in many other general circulation models (see Xue et al., 2012), the quasi-biennial oscillation (QBO) is imposed by nudging the tropical stratospheric zonal winds to observed winds. ~~The simulations are nudged to produce the observed zonal winds of the tropical quasi-biennial oscillation (QBO).~~ From the sets of fully-coupled simulations, we were able to estimate responses to greenhouse gases alone (referred to herein as GHGonly) and ozone depleting substances (referred to herein as ODSonly) by independently differencing the individual ensemble anomalies from the FIXED GHG1960 and ODS1960 simulations from the ALL2 ensemble mean. In addition, we were able to isolate the combined response to volcanic aerosols, the solar cycle, and the QBO, by differencing the sum of ensemble mean anomalies of the

FIXED GHG1960 and ODS1960 simulations from the ~~individual~~ ensemble anomalies of ALL2. We designate this as the natural historical ~~signal~~~~mulation~~ (referred to herein as NAT-h). The underlying assumption for ~~our method~~ of estimating the GHGonly, ODSonly and NAT-h responses is that the stratospheric ozone responses to individual forcings add linearly. Eyring et al. (2010) found this assumption to hold in the stratosphere, but with some departures from linearity in the tropical total column ozone. In Supplementary Figure S1 we show that, in absence of the NAT-h response, the global-mean sum of the GHG and ODS response in the upper and lower stratospheric ozone are close to the all forcing simulation.

There is reason to suspect that the NAT-h signal may not thoroughly represent the ozone response to volcanic eruptions in certain regions of the lower stratosphere. In the low halogen loading stratospheric conditions of FIXED ODS1960, increases in mid-stratospheric ozone following a large volcanic eruption are expected as the loss of ozone at these altitudes is dominated by odd nitrogen, due to enhancement of N_2O_5 hydrolysis (Tie and Brasseur, 1995). In the enhanced halogen loading conditions of FIXED GHG1960 and ALL2, the mid-stratospheric increase is limited to higher altitudes, and ozone depletion will occur throughout the lower stratosphere. The 40 to 100 hPa region in the lower stratosphere encompasses a part of the vertical altitude range where the opposing effects of reactive nitrogen and halogen chemistry occur. The validity of the assumption that these opposing effects are additive is unclear. Despite the uncertainty in the simulated ozone response to volcanic forcing, we show subsequently that the NAT-h simulation successfully captures many features of the observed short- and long-term variability in lower stratospheric ozone.

iii) Free-running fully coupled simulations of the last millennium (years 850 to 1850), in which only the estimated external forcing from historical solar variability is specified. This simulation (henceforth referred to as NAT) provides insight into the range of fluctuations in stratospheric ozone arising from the combined effects of ~~internal variability and~~ the solar cycle and internal variability (other than the QBO; ~~there is no nudging to observed stratospheric winds which does not occur~~ in this simulation).

iv) Free-running fully coupled pre-industrial control run of length 200 years. Unlike *iii*), there are no temporal changes in solar forcing and all variability is intrinsic to the climate system. This is the same experiment used by Marsh et al. (2013) to provide an estimate of internal climate variability alone, and is referred to herein as CTL.

We rely primarily on the Stratospheric Water and Ozone Homogenized (SWOOSH) database (Davis et al., 2016) for observational estimates of stratospheric ozone changes. SWOOSH includes vertically resolved ozone from a subset of the limb profiling satellite instruments operating from 1984 to present day. SWOOSH's ozone product is a gridded monthly-mean zonal-mean time-series of mixing ratios on pressure levels ranging from 1 to 316 hPa. A key aspect of this merged product is that the source records are homogenized to account for inter-satellite biases and to reduce the impact of non-climatic artifacts. In this study, we use SWOOSH version 2.6 with a latitudinal resolution of 10 degrees. An advantage of the SWOOSH dataset

is that the vertical range enables investigation of the full stratosphere from 1984-01 to 2016-12. Tummon et al. (2015) found that SWOOSH ozone concentrations from 1984-2011 were within +/-10% of other datasets throughout much of the stratosphere, with the largest differences in the lower stratosphere. To test the sensitivity of our D&A results for the upper stratosphere to the choice of observational dataset, we also employ a second ozone dataset that is independent from SWOOSH: the previously-mentioned SBUV Merged Cohesive (SBUV_CDR) ozone dataset. This is based on the version 8.6 SBUV MOD dataset, and attempts to further reduce the inter-satellite differences. The dataset is produced at the National Oceanic and Atmospheric Administration (NOAA, ftp://ftp.cpc.ncep.noaa.gov/SBUV_CDR), and was available for the years from 1979-2015 at the time of this study.

For the signal-to-noise analysis, it was necessary to transform the ozone data from both observational grids and the model grid of WACCM to a common grid. The common horizontal grid chosen was SWOOSH's 10 degree latitude zonal-mean grid from 85°S-85°N. Transformation to a relatively coarse-resolution grid reduces the spatial dimensionality of the input datasets, which is of benefit in the estimation of Empirical Orthogonal Functions (EOFs) used later in the fingerprint analysis. Regridding of the vertical coordinates was also performed: the WACCM output vertical levels were transformed to match SWOOSH's vertical levels (which contains 31 pressure layers, spanning a pressure range from 1 to 316 hPa). Model ozone concentrations were masked at latitude bands and layers where no data was present in SWOOSH. As a sensitivity test, we changed the definition of the upper and lower stratosphere regions used for the D&A analysis, using +/-1 SWOOSH levels on either side of the original vertical ranges. When calculating the fingerprints for the upper and lower stratosphere, we use anomalies weighted by the cosine of latitude, but do not employ vertical pressure weighting. The latter processing choice is reasonable, since ozone concentrations do not vary by over an order of magnitude within each region and the anomalies are expressed as percent changes.

3 Global Mean Ozone Changes

Figures 1 A and B show, separately for the upper and lower stratosphere, the time series of globally averaged annual-mean ozone anomalies for the SWOOSH observations and the five WACCM ensembles (ALL1, ALL2, GHGonly, ODSonly, and NAT-h). Anomalies are expressed in terms of percent changes relative to the first 10 years (1960-1969) of the ALL1 simulation. In the upper stratosphere, there is a steady decline in stratospheric ozone from the 1960s to late 1990s in both the ALL1 and ALL2 historical simulations, which is in agreement with the SWOOSH data. After the late 1990s, the decline slows down, and there is some indication of a recovery of stratospheric ozone. The ODSonly anomalies broadly track the temporal changes in ALL1 and ALL2 but reach larger depletion before beginning to increase, supporting the large body of literature indicating that the large decline was predominantly due to ODS. The difference between the ALL1/ALL2 simulations and ODSonly can be explained by the steady increase in upper stratospheric ozone in GHGonly, consistent with upper stratospheric cooling. This supports the conclusion that upper stratospheric ozone depletion would have been greater during the depletion

era in the absence of [increases in well-mixed](#) greenhouse gases (WMO, 2014). The NAT-h anomalies show the presence of the 11-year solar cycle in the upper stratosphere, but no overall change in ozone from 1960 to 2016.

In the lower stratosphere (Fig. 1B), ozone changes are characterized by larger interannual variability. There is a dominant 2-3 year cycle that is due to the historically [imposed nudged-QBO](#) ~~prescribed~~ in all ensemble members of ALL1, ALL2, and NAT-h. Although the ODSonly and GHGonly simulations do not contain the natural and QBO response present in the other simulations, we see a large inter-ensemble spread in the lower stratosphere when compared to the upper stratosphere. This is related to dynamical differences between ensemble members as a result of the ~~importance of~~ direct transport of ozone in the lower stratosphere. The ODSonly evolution is similar to that in the upper stratosphere, and broadly tracks the ALL1 and ALL2 runs. For global GHGonly ozone anomalies in the lower stratosphere, there is no significant trend. As noted by Li et al., 2009, in the lower stratosphere, enhanced ozone advection due to a strong BDC results in decreases in the tropical ozone and increases in the extratropical ozone, and in a global mean sense, the lower stratospheric changes cancel out.

It is noteworthy that the historical simulations differ from SWOOSH after 2005, when the global-mean lower stratospheric ozone in SWOOSH does not flatten or begin to turn around, in contrast to the recovery in ALL1 and ALL2. To further examine this divergent behavior, we calculate lower stratosphere ozone changes averaged over the tropics (30°S to 30°N; see Fig. 1C), and over a global domain that excludes the tropics (Fig. 1D). It is clear from this comparison that the post-2005 observation-model difference in global-mean lower stratospheric ozone is largely due to ozone differences in the tropics. ~~One possible explanation for this result is that the CCMI aerosols used in this study fail to reproduce the observed post-2005 increase in lower stratospheric aerosol loading following a series of small magnitude volcanic eruptions (Bourassa et al., 2012; Neely et al., 2013; Vernier et al., 2011; Ridley et al., 2014). In simulations relying on a new database of volcanic plume emissions of sulfur dioxide, Mills et al. (2016) found that the CCMI aerosols underrepresented the increase in stratospheric aerosol optical depth over the post-2005 period.~~ This may be partly due to the large observational uncertainty in ozone loss in this region. As noted in the WMO (2014) report, there is considerable disagreement between observed data sets in terms of ozone changes in the tropical lower stratosphere since 2000. The report states that changes since 2000, “computed from different data sets in the tropical lowermost stratosphere remain an open question”. Because of this observational uncertainty in tropical ozone changes (and because of the discrepancy between post-2000 tropical ozone changes in SWOOSH and the ALL1/2 simulations), our subsequent D&A analysis for the lower stratosphere is performed over two domains: a full global domain, and a domain which excludes 30°S to 30°N.

To summarize, upper stratospheric ozone in SWOOSH and the ALL1 and ALL2 simulations shows a large decline of 5-8% per decade through the middle 1990s, followed by an increase of 2.5-5%/decade over the last 10 to 15 years. This decline and recovery is in agreement with findings of the WMO (2014) report. In the lower stratosphere, larger interannual variability complicates the identification of long-term ozone changes, and there are post-2005 differences between the historical WACCM

model simulations and SWOOSH data that are relevant to the interpretation of the D&A results. SBUV_CDR ozone anomalies in the upper stratosphere are similar to those of SWOOSH (see Supplementary Figure S2+).

3.1 Variability in WACCM Natural and Control Simulations

The CTL and NAT simulations provide estimates of the amplitude and timescales of natural ozone variability in the WACCM model (see Fig. 2). In the CTL simulation (upper panels), fluctuations in upper and lower stratospheric ozone concentrations arise from internal variability alone. In the last millennium NAT simulation, ozone fluctuations result from both internal variability and from solar forcing (lower panels). Previous work has shown that the solar cycle, ENSO, and the QBO are the main contributory factors to natural variability in stratospheric ozone, affecting ozone concentrations through chemistry and transport mechanisms (Kirgis et al., 2013; Nair et al., 2013). We show below the different effects of these contributory factors in the upper and lower stratosphere.

In the upper stratosphere (left panel of Fig. 2), the ozone variation caused by the 11-year solar cycle is clearly evident in the NAT simulation (see expanded time axis). The magnitude of the internally generated ozone variability in the CTL is smaller than ozone variability arising from the 11-year solar cycle. The NAT-derived estimate of the change in upper stratospheric ozone from solar maximum to solar minimum is approximately 2-4%. This is in agreement with results from previous observational and model-based studies (Lee and Smith, 2003; Newchurch et al., 2003; Randel and Wu, 2007). NAT also shows longer-timescale variability in upper stratospheric ozone concentrations in response to solar insolation changes over multiple decades to centuries. The NAT results highlight the large contribution of solar forcing to ozone variability in the upper stratosphere, and the smaller role played by internally generated variability. The NAT-h simulation (the yellow time series in Fig. 1A) indicates that there is limited influence of the QBO 2-3 year on ozone variability in the upper stratosphere. This is consistent with a previous estimate that the QBO effect on upper stratospheric ozone is less than half that of solar variability over the 11-year solar cycle (Egorova et al., 2004).

In the lower stratosphere (right panel of Fig. 2), the interannual variability of ozone in the CTL is as large as the ozone variability in NAT. In this region of the atmosphere, internal variability in dynamics and transport is the dominant driver of natural fluctuations in ozone – not solar irradiance changes. In fact, the influence of the solar cycle on ozone is difficult to discern in the lower stratosphere (see expanded axis). This is consistent with results from previous studies, which have shown that the solar cycle effect on ozone maximizes around 35 km and weakens in the lower stratosphere (Egorova et al., 2004; Lee and Smith, 2003; Newchurch et al., 2003). The QBO-driven anomalies in NAT-h (Fig. 1B) are of much larger magnitude in the lower stratosphere; QBO-driven ozone anomalies range from 4-8%, in accord with earlier results (Hasebe, 1994; Randel et al., 1996).

Newchurch (2003) found that the large interannual variability in global-mean lower stratospheric ozone cannot be explained by the solar cycle and the QBO only: a significant portion of the variance is related to monthly meteorological variability. Changes in interannual tropical upwelling and lower stratospheric ozone concentrations have been previously linked to the El Niño-Southern Oscillation (ENSO, Brönnimann et al., 2013; Oman et al., 2013). ENSO-induced variations in lower stratospheric ozone were found to exceed 5%. ~~This large noise hampers anthropogenic signal detection.~~ ENSO-related variations in tropical upwelling are partly responsible for modulating the strength of planetary waves, leading to temperature and water vapor variations in the tropical lower stratosphere that can have impacts on the chemistry and transport of ozone (Randel et al., 2009). ENSO signals in ozone are strongest just above the tropical tropopause; at higher levels, they are weaker than the QBO-related ozone signal (Randel and Thompson, 2011).

Supplementary Figure S32 shows the normalized power spectra of global upper and lower stratospheric ozone anomalies for NAT, NAT-h, and CTL ~~with the lowest frequency shown is the lowest frequency that can be resolved given the length of the observational ozone record.~~ In the upper stratosphere, the 11-year solar cycle is the strongest signal present in the NAT and NAT-h simulations. In the lower stratosphere, the 2-3 QBO signal is strongest in NAT-h (which ~~has an imposed is nudged to the observed~~ QBO).

By comparing the stratospheric ozone results in Figs. 1 and 2, we can obtain a rough estimate of the relative sizes of anthropogenically-driven and purely natural changes. This simple comparison suggests that the upper stratosphere is the region where anthropogenic signal detection would be most feasible.

3.2 Observed Versus Model Variability

Model estimates of internal climate variability are a key component of detection and attribution studies (Allen and Tett, 1999; Santer et al., 2013a+). One common strategy is to estimate and remove externally forced climate signals from the observations, and then compare the residual variability with control run internal variability (Hegerl et al., 1996). There are a number of uncertainties in such signal removal strategies (Santer et al., 1996b). Here we use an approach similar to Santer et al., 2013b ~~Santer et al., 2011~~ and directly compare estimates of total variability (arising from both internal processes and natural external factors) in the observations, ~~CTLcontrol~~, NAT, NAT-h, ~~natural~~ and ALL1 ~~“all forcing”~~ runs. If the model systematically underestimates the amplitude of stratospheric ozone variability on multi-decadal timescales, the S/N ratios are likely to be inflated. Whether such a systematic error exists is difficult to determine because the global ozone observational records are relatively short. Observations cannot, therefore, provide an unambiguous constraint on model-based estimates of low-frequency ozone changes, but they can provide useful insights into the direction and size of model variability errors.

We investigate these issues with the modeled and observed monthly-mean time series of global-mean upper and lower stratospheric ozone anomalies. After detrending the modeled and observed time series, we isolate long-term variability by

applying a band-pass filter with half-power cutoffs at 5 and 20 years to the residuals. We also used a high-pass filter with a half-power point at 3 years to exclude all variability on timescales longer than 5 years. All filtering operations were performed after detrending the time series with a low-pass filter with a half-power point of 30 years. The key point here is that the modeled and observed ozone data are filtered in exactly the same way. We use S_{IV} (“interannual variability”) and S_{DV} (“decadal variability”) to denote the temporal standard deviations of the high-pass and band-pass filtered ozone data, respectively.

The S_{IV} and S_{DV} results in Figure 3 were calculated for SWOOSH (396 months), ALL1 (for the 372-month period from 1984 to 2014), NAT (12,000 months), CTL (2,400 months) and NAT-h (2,112 months). NAT (NAT-h) provides information on solar and internal variability in the absence (presence) of variability from the QBO. The CTL variability is solely generated by processes internal to the climate system, and has no contribution from natural external forcing. Fig. 3 shows that these simulations have systematic differences in the amplitude of variability. These differences are manifest on both the sub-3-year and the 5- to 20-year timescales. In the lower stratosphere, the QBO is a dominant component of the variability on timescales < 3 years: only the ALL1 and NAT-h simulations (both of which have QBO-induced ozone changes) are close to the observed value of S_{IV} . NAT-h also includes volcanic variability, which ~~is would be~~ an important component ~~in the lower stratosphere~~ for S_{IV} ~~in the lower stratosphere~~. CTL and NAT (which both lack QBO and volcanic-driven ozone changes) underestimate the observed interannual variability. In the upper stratosphere, the observed and modeled S_{IV} all fall within 0.60 - 0.85%.

The decadal variability is of key interest in D&A studies, since it constitutes the background noise against which analysts attempt to identify gradually evolving anthropogenic signals. In the lower stratosphere, the CTL value of S_{DV} is (as expected) lower than in the other three types of simulation. There is no evidence that the simulations most directly comparable to the observations (ALL1 and NAT-h) systematically underestimate the observed values of S_{DV} obtained from the SWOOSH data. Such an underestimate would be concerning: it would spuriously inflate the S/N ratios obtained in the D&A analysis (see below). In fact, both ALL1 and NAT-h yield values of S_{DV} that are slightly larger than in observations (see Santer et al., 2013a).

In the upper stratosphere, however, all four types of model simulation have values of S_{DV} that are smaller than the observed result. As in the lower stratosphere, CTL (which does not include solar forcing) that has the lowest decadal variability. ALL1, which has most realistic time-varying external forcing, is closest to SWOOSH, but still 11% less than the observed S_{DV} value. ~~One~~A possible explanation for this difference in upper stratospheric S_{DV} between SWOOSH and ALL1, can be seen in Fig. 1A. ~~In comparing the two around~~ the mid- to late 1980s ~~where~~ SWOOSH is prominently higher than ALL1 ~~in that short period~~, which would lead to a higher S_{DV} value. ~~There is also considerable observational uncertainty in S_{DV} as different ozone datasets have significantly different representations of the magnitude of the solar cycle (Maycock et al., 2016). Chemistry climate models are also known to have solar cycle variations in ozone that are towards the lower bounds of observational estimates (see Chapter 8.5 of SPARC CCMVAL, 2010).~~

4 Latitude/Altitude Patterns of Ozone Change

Figure 1 shows that in the case of all simulations except GHGonly, a simple least-squares linear fit regression line is not an adequate representation of ozone changes over the entire observational record (1984-2016). The nonlinear behavior of ozone —which is particularly evident in Fig. 1A— occurs because ODS emissions were curtailed through implementation of the Montreal Protocol. As mentioned in the introduction, we address this nonlinearity by representing decadal changes in ozone are modeled using an EESC proxy a standard proxy for equivalent effective stratospheric chlorine (EESC) proxy, as in Randel and Wu (2007) and Newchurch (2003) (see also WMO 2002, Fioletov and Shepherd, 2005; Stolarski et al., 2006). This proxy isolates ozone changes associated with changes in the amount of ozone depleting chlorine and bromine in the stratosphere. The EESC curve was taken from the NASA Goddard Automailer, which uses the Newman and Daniel (2007) EESC values updated according to WMO (2010). The EESC calculations assume that the mean age of stratospheric air is 5.5 years and age-of-air spectrum width = 2.75 years. Sensitivity to the choice of these two parameters (e.g. varying the MAA from 3-6 years) affected the magnitude but not the significance of the regression coefficients. The advantage of the EESC fit, is that it removes some of the nonlinearity that is manifest in the ozone changes, and indicates overall agreement with ODS changes and regression to the full record.

Figures 4 A and B show the respective least-squares linear trends in ozone and in EESC over the entire period of observational record (1984-2016). Results are for the ensemble-mean model simulations and the SWOOSH ozone data. The purpose of the comparison is to determine which trend patterns the expect change from individual and combined forcings. We also compare our fingerprints, derived in the next section, to the patterns of long-term change.

We used the method outlined by Trenberth (1984) to determine the statistical significance of the trends. Here and subsequently, the stipulated significance level is chosen here to be the 5%-level. When trends are directly calculated from the ozone data, GHGonly is the sole only simulation yielding an appreciable region of statistically significant results (Fig. 4A). This is the region where GHGonly shows a 1-2% per decade increase in ozone in the upper stratosphere (1-10 hPa, Fig. 1A) and is the expected region of latitudinally coherent cooling. In GHGonly, there is a appears to be a non-significant decrease in ozone in the tropical lower stratosphere and a non-significant increase in the extratropics, as expected from an enhanced BDC, it is not statistically significant. WMO (2011, 2014) reported negative ozone trends in the tropical lower stratosphere between 1985-2005, and CCMVal simulations indicate a long-term increase in tropical upwelling and an increased BDC strength. We note, however, that the WMO (2014) analysis of shorter satellite data sets between 2002 and 2012 does not show significant tropical lower stratospheric ozone trends, which may reflect the larger noise associated with substantial decadal variability in the lower stratosphere (see Fig. 3).

Very different results are obtained using the EESC regression (Fig. 4B). As expected, ODS forcing makes the largest contribution to the ozone-change pattern in ALL1 and ALL2. Large regions with significant effects of ODS changes on ozone are evident in ALL1, ALL2 and ODSonly. Each of these simulations shows the familiar pattern of mid-latitude lobes in the upper stratosphere. Another common aspect of ALL1, ALL2, and ODSonly is a lower stratospheric response in the southern hemisphere (SH), with an Antarctic ozone hole that is strong enough to persist in annual anomalies. The SWOOSH ozone data also display the significant two-lobe structure in the upper stratosphere and the lower stratospheric response in the SH polar region. The latter feature, however, extends further equatorward in the ALL1, ALL2, and ODSonly simulations than in the observations. In the historical NAT-h simulation, there are no statistically significant trends in either the ozone or EESC data over the 1984 to 2016 period.

The key conclusion of this section is that the statistical significance of linear trends in stratospheric ozone behavior depends critically on whether trends account for nonlinearities in the ozone forcing. Previous studies with EESC-type regressions have shown a familiar pattern of minima in the upper stratosphere (~35-45km) and the polar lower stratosphere (15-25km), similar to that obtained here in the EESC-based ALL1, ALL2, and ODSonly simulations (Randel and Wu, 2007; Wang et al., 2002). The upper stratospheric changes have a symmetric latitudinal structure, which is the fingerprint of gas-phase chlorine-induced ozone loss. For both ozone and EESC, there are no significant trends in NAT-h over 1984-2016. We conclude from our results that in the WACCM model, ozone changes over 1984 to 2016 are primarily forced by human-caused changes in ODS, and cannot be explained by natural factors alone. Patterns similar to those found for the SWOOSH data (but with slightly larger magnitudes) were obtained using the SBUV_CDR observations (compare Figure 4 and Supplementary Figure S3S4).

5 Fingerprint Estimation

In most applications, the climate change fingerprint is a geographical pattern (Hegerl et al., 1996; Santer et al., 2003), a vertical profile through the atmosphere or ocean (Barnett et al., 2005; Santer et al., 1996b; Tett et al., 1996) or a vector with information on the combined spatial and temporal properties of the signal (Gillett, 2002; Stott et al., 2000; Tett et al., 2002). Here, the fingerprint is a time-invariant latitude-altitude pattern; ozone changes are zonally averaged along latitude bands. The fingerprint provides an estimate of the multi-decadal response to external forcing by combined and individual separate-human forcings. The implicit assumption in this approach is that the spatial pattern of response does not change markedly over time (Santer et al., 2013). We examine the adequacy of this assumption for the specific problem of interest here – the identification of a human-caused fingerprint pattern in the SWOOSH observations. The assumption of timescale-invariance of the fingerprint is tested by defining the fingerprint over different time intervals. We use a standard method (Hasselmann, 1979; Santer et al., 1995) to determine whether model-predicted patterns of externally forced stratospheric ozone changes can be identified in SWOOSH. Fingerprinting is performed separately for the upper and lower stratosphere.

Let $\chi(i, x, p, t)$ represent the annual-mean percent ozone anomalies at latitude band x , pressure p , and year t for the i^{th} ensemble realization for each of the ALL1, ODSonly and GHGonly simulations:

$i = 1, \dots, N_r$ (number of ensemble members, ranging from 3 to 5)

5 $x = 1, \dots, N_x$ (total number of latitude bands, 17)

$p = 1, \dots, N_p$ (total number of pressure layers, 10 and 6 for the upper and lower stratosphere, respectively)

$t = 1, \dots, N_t$ (time in years)

Note that we are not using ALL2 for fingerprint estimation because the ODSonly and GHGonly responses were derived from ALL2. As in previous work (Santer et al. 2003, 2013a), we define the fingerprint $F(x, p)$ by first averaging ozone changes over individual ensemble members, and then calculating the leading empirical orthogonal function (EOF) of the covariance matrix of $\bar{\chi}(x, p, t)$. Many fingerprint studies seek to rotate $F(x, p)$ in a direction that maximizes the signal strength relative to the control run noise (Gillett et al., 2011; Tett et al., 2002). Optimization of $F(x, p)$ generally leads to enhanced detectability of the signal. In this study, we were able to achieve high signal-to-noise levels (see below) without any optimization of $F(x, p)$, and only non-optimized results are discussed. The searched-for fingerprints were computed using ozone anomalies from 1960 to 2016 for ODSonly and GHGonly, and from 1955 to 2014 for ALL1. Sensitivity of our results to the choice of time period for fingerprint calculation was tested by using both a longer period of time (1950 to 2050 for ODSonly and GHGonly) and a shorter time period (1984 to 2016). The fingerprint patterns were found to be relatively insensitive to the choice of time period for estimating $F(x, p)$.

20

The fingerprint patterns for the upper and lower stratosphere are shown in Figures 5 and 6 (respectively) for ODSonly (A), GHGonly (B), and ALL1 (C). Below each fingerprint is the associated principal component (PC) time series showing the temporal changes in the strength and sign of the pattern in the model simulation. For both the upper and lower stratosphere, the EOF patterns are similar to the latitude-altitude trend patterns presented previously, with GHGonly EOF 1 closely matching the GHGonly linear trends in ozone (Fig. 4A), and the leading ODSonly and ALL1 EOFs closely corresponding to trends computed from the EESC proxy.

In the upper stratosphere, the leading ODSonly and GHGonly EOFs tend to have maximum amplitude at high latitudes in both hemispheres (ODSonly) and in the tropics (GHGonly). These differences are expected based on the spatial structure and time evolution of long-term changes in ozone in ODSonly and GHGonly (see Figs. 4A and B). At the global scale, however, we note that the EOF 1 patterns in both ODSonly and GHGonly have the same sign at virtually all grid points in the upper stratosphere. The standardized PC time series indicate that the ODSonly and ALL1 results are qualitatively similar to the EESC

30

curve, while PC 1 for GHGonly is more linear. We rely on these PC time series in our subsequent signal-to-noise (S/N) analysis.

In the lower stratosphere, the familiar Antarctic ozone hole is visible in the ODSonly and ALL1 fingerprints (Figs. 6A and C).

- 5 The GHGonly fingerprint (Fig. 6B) is qualitatively similar to the long-term linear trend in ozone in the GHGonly simulation (Fig. 4A), with ozone changes of opposite sign between 100 to 50 hPa in the tropical lower stratosphere.

A common feature of both the upper and lower stratospheric results in Figs. 5 and 6 is the similarity between the ALL1 and ODSonly fingerprints. This similarity indicates that in the model simulations, changes in ODS are the primary driver of the changes in stratospheric ozone over the past 50 years. A noticeable difference between the upper and lower stratospheric results in Figs. 5 and 6 is that results for the lower stratosphere are noisier. This increased noise is manifest in two ways. First, relative to the upper stratosphere, EOF 1 of ODSonly, GHGonly, and ALL1 consistently explains less of the overall variance in lower stratospheric ozone changes. Second, the amplitude of the variability of the ODSonly, GHGonly, and ALL1 PC time series is systematically larger in the lower stratosphere. As noted above, these differences between the upper and lower stratosphere reflect the influence of different processes. In the upper stratosphere, local concentrations of ozone are primarily driven by gas-phase chemistry. In contrast, ozone concentrations in the lower stratosphere receive a substantial influence from dynamical transport, which introduces larger interannual variability.

Before presenting the results of the S/N analysis, we first examine the major modes of naturally forced and internal variability estimated from the NAT-h, NAT, and CTL simulations. This is done separately for upper and lower stratospheric ozone (see Figs. 5D-F and 6D-F, respectively). For NAT and CTL we use the full simulations for calculating EOFs and PCs, while for NAT-h we examine the ensemble mean of the individual members prior to EOF and PC estimation, and rely on the years 1960 to 2016. Below each NAT and CTL EOF pattern, we show a 50-year segment of the associated principal component for a 50-year slice, thus facilitating comparisons with the PC time series of ODSonly, GHGonly, and ALL1.

25 In the upper stratosphere, the 11-year solar cycle is present in NAT (EOF1) and NAT-h (EOF3). EOF1 from the CTL is also manifest in NAT (EOF2), and appears to be associated with interannual changes in tropical upwelling. Tropical upwelling influences temperature, and thus therefore also affects ozone, via temperature-dependent reaction rate chemistry. In the lower stratosphere, solar-induced ozone changes are not clearly visible in the three leading modes of variability computed from NAT-h, NAT, and CTL. The dominant mode is similar in these simulations, and appears to be associated with changes in upwelling and direct transport of ozone through the lower branch of the BDC. EOF2 of NAT-h displays the 2-3 year QBO influence on lower stratospheric ozone (recall that NAT-h contains an imposed QBO to match observations is nudged to reanalysis temperature and wind fields).

In both the upper and lower stratosphere, the patterns of the dominant modes of variability in NAT-h, NAT, and CTL are noticeably different from the searched-for fingerprint patterns. As noted above, the fingerprints estimated from the ODSonly, GHGonly, and ALL1 simulations show upper stratospheric ozone changes that have the same sign at all grid-points. In contrast, the leading 3 modes of variability estimated from the NAT-h, NAT, and CTL runs do not have the same spatial coherence of ozone change, and are generally characterized by ozone changes of opposite sign at smaller spatial scales. The sole exception is the coherent, same-signed EOF pattern associated the 11-year solar cycle (EOF3 in NAT-h and EOF1 in NAT; see Figs. 5D and E). To quantify these similarities and differences, we calculated pattern correlations between the searched-for fingerprints and the leading noise modes (see Table 1). We rely on these correlations later for interpreting results from the S/N analysis (see below).

10 6 Signal to Noise Estimates ~~and Detectability Enhancements~~

We now seek to identify the model-predicted fingerprints of anthropogenically forced stratospheric ozone changes in observations. We project the time-varying annual-mean latitude-height ozone anomalies from the SWOOSH data, denoted here by $O(x, p, t)$, onto $F(x, p)$ the time-invariant fingerprint from the ODSonly, GHGonly, or ALL1 simulation (see Figs. 5A-C and 6A-C). This projection step yields the signal time series $c\{F, O\}(t)$:

$$c\{F, O\}(t) = \sum_{x=1}^{N_x} \sum_{p=1}^{N_p} F(x, p) O(x, p, t) . \quad (1)$$

This projection is equivalent to a spatially uncentered covariance between the patterns $F(x, p)$ and $O(x, p, t)$ at time t . The signal time series $c\{F, O\}(t)$ provides information on both the amplitude and sign of the fingerprint in observational data. We can then analyze how $c\{F, O\}(t)$ is changing with time – i.e., whether the searched-for fingerprint pattern is becoming increasingly similar to observed latitude-altitude patterns of ozone change.

Several approaches can be used to assess the significance of the changes in $c\{F, O\}(t)$: direct comparison of actual $c\{F, O\}(t)$ values with a null distribution (e.g. Wigley et al., 1998), or comparison of the trends in $c\{F, O\}(t)$ with a null distribution of trends (e.g. Santer et al., 2003~~1995~~). We use the latter approach here. To assess trend significance, we require a case in which $O(x, p, t)$ is replaced by a record in which we know *a priori* that any spatial correspondence with the fingerprint occurs by chance alone. Here, we use the noise data set $N(x, p, t)$, which is constructed by concatenating together the NAT, NAT-h, and CTL simulations. The associated noise time series $c\{F, N\}(t)$ is the spatially uncentered covariance of $F(x, p)$ and $N(x, p, t)$:

$$c\{F, N\}(t) = \sum_{x=1}^{N_x} \sum_{p=1}^{N_p} F(x, p) N(x, p, t) . \quad (2)$$

Estimates of signal to noise (S/N) are conventionally evaluated (e.g. see Santer et al., 2003+1995) by fitting linear least squares trends of increasing length L to $c\{F, O\}(t)$, and then comparing these trends with the standard deviation of the distribution of L -length trends found in the noise time series $c\{F, N\}(t)$. Signal detection is obtained when the trend in $c\{F, O\}(t)$ exceeds and remains above a stipulated significance level. The test is one-tailed, and assumes a Gaussian distribution of trends in $c\{F, N\}(t)$. Here we use the 1% significance level, which corresponds to a S/N ratio close to 2.33. S/N ratios that consistently exceed this level are highly unlikely to be due to the combined effects of natural internal variability and natural external forcing.

Since the sign of the fingerprint EOFs in Figs. 5 and 6 is arbitrary (and since the EOF patterns are very similar to the patterns of ozone trends in Fig. 4), we stipulate that the sign of the leading EOF in ODSonly, GHGonly, or ALL1 should match the sign of the corresponding simulation's ozone trend pattern. With this stipulation, negative S/N ratios indicate that the sign of the overall change in ozone in the anthropogenically forced simulation is inconsistent with the observed ozone change.

We estimate S/N ratios by fitting least-squares trends to L -length segments of $c\{F, O\}(t)$, and then comparing these signal trends with $s(L)$, the standard deviation of the sampling distribution of m maximally overlapping L -length trends in $c\{F, N\}(t)$ (i.e., for overlap by all but one year). As noted earlier, $c\{F, N\}(t)$ is the concatenation of NAT, NAT-h, and CTL, and due to the discontinuities between the simulations, we discard trends that span two different noise simulations. As L increases, m decreases. We use maximally overlapping trends to guard against excluding the largest changes in ozone from our analysis. We use a minimum trend length of for $L = 10$ years, so the first S/N ratio (and the earliest possible detection time) is for 10-year trends starting in 1984 and ending in 1993.

One innovative feature of this work is that we derive estimates of S/N behavior using two types of approach. The regression is of the standard form:

$$y = X\beta + \epsilon, \quad (3)$$

where y can be either $c\{F, O\}(t)$ or $c\{F, N\}(t)$, X contains the regressors, β contains the reported trend/regression coefficients, and ϵ is an error term.

The first method we employ for S/N estimation is of the conventional form used in such studies (see, e.g., Santer et al., 2003), where X is the time coordinate in years:

$$X = \begin{pmatrix} 1 & t_0 \\ \vdots & \vdots \\ 1 & t_0 + L \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_l \end{pmatrix}, \quad (4)$$

t_0 is the starting year (1984), β_0 is the intercept, and $\beta_l(L)$ is the reported trend coefficient of interest, with units of percent change in ozone per year.

5

The second method relies on linear regression between $c\{F, O\}(t)$ or $c\{F, N\}(t)$ and a selected PC_1 time series. There are a total of six PC_1 time series, one for each layer (upper and lower stratosphere) and for each of the ODSonly, GHGonly and ALL1 fingerprints. In this second method, the regressor X , contains the PC_1 time series for the selected trend years:

$$10 \quad X = \begin{pmatrix} 1 & PC_1(t_0) \\ \vdots & \vdots \\ 1 & PC_1(t_0 + L) \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_{PC_1} \end{pmatrix}, \quad (5)$$

where $\beta_{PC_1}(L)$ is the reported trend coefficient, and has units in percent change per unit standard deviation of PC_1 . For example, if both $c\{F, O\}(t)$ and PC_1 show similar nonlinear behavior over a common analysis period (such as 1984 to 2016), the regression coefficient $\beta_{PC_1}(L)$ will be unusually large relative to values of the regression coefficient estimated with $c\{F, N\}(t)$. In cases where PC_1 exhibits change that is nearly linear with time, then the change in $\beta_l(L)$ (the linear trend representation of ozone change) and $\beta_{PC_1}(L)$ with increasing L will be similar.

15

The advantage of the second method is that it accommodates forcings whose time evolution is markedly nonlinear over the observational record, whereas the first method implicitly assumes that the forcing evolution is quasi-linear over the period of interest. Like the first method, however, the second method still assumes that the spatial structure of the searched-for fingerprint is essentially unchanged with time (an assumption that is justifiable; see above). In summary, our second method of estimating S/N behavior uses information on both the model-derived fingerprint and its time evolution to search for the fingerprint in SWOOSH data. Since the ODSonly and GHGonly fingerprints have different time evolution properties (particularly in the upper stratosphere; compare Figs. 5A and B), explicit consideration of this time evolution information can be useful in separating the ODS- and GHG-induced ozone change signals in observations. In the following, we refer to our first S/N method as the *linear* or β_l trends method, and refer to the second method as either the PC_1 or nonlinear signal method.

20

25

The S/N results are shown in Figures 7 and 8 for the upper and lower stratosphere respectively. The left panels show S/N estimates from the β_l trends, and the right panels show the S/N estimates based on β_{PC_1} trends. Results are for the three fingerprints: ALL1, GHGonly and ODSonly. Since the β_{PC_1} trends are regressions to the associated PC_1 of the fingerprints,

30

the ALL1 simulation (which ends in December 2014) yields S/N ratios that end 2014, while the ODSonly and GHGonly simulations span the full observational record, allowing S/N ratios to be calculated through to 2016 (inclusive).

We consider first S/N ratios for the upper stratosphere, and for results based on the first method of estimating trends (see left panels of Fig. 7). The use of β_l yields virtually identical signal trends, noise trends, and S/N ratios for the ALL1 and ODSonly fingerprints. Recall that the signal trends are calculated from the time series of the projections of the SWOOSH data onto the fingerprints. The ALL1 and ODSonly signal trends are large and positive in the first 10-20 years. These large positive trends arise during the depletion era, when upper stratospheric ozone decreased steadily until the late 1990s. The positive sign of the signal trends that end in the depletion era reflects the consistency in sign between the observed ozone loss and the loss of ozone captured in the ALL1 and ODSonly simulations (see Figs. 5A and C). As the trend length L extends into the 21st century, the signal trends for ALL1 and ODSonly decline towards zero. This is a result of the stabilization of ozone loss and the emerging recovery of stratospheric ozone. Such nonlinear behavior is not well-described by fitting a straight line through the entire $c\{F, O\}(t)$ time series.

The time evolution of the β_l signal trends obtained with the GHGonly fingerprint mirrors the results for ODSonly and ALL1, but is of opposite sign (Fig. 7A). The negative sign of the GHG signal trend arises because of the sign mismatch between upper stratospheric ozone changes in SWOOSH and the GHGonly simulation (see Fig. 1A). The standard deviations of the noise trends (Fig. 7B) are very similar for the 3 fingerprints, which is expected given the similarity between the ODSonly, GHGonly, and ALL1 fingerprint patterns (see Table 1). The decrease in the amplitude of the noise trends with increasing trend-fitting period is typical behavior for many different climate variables (see, e.g., Santer et al. 2013a,b).

Because of early 21st century ozone recovery, and the impact of recovery on β_l signal trends, S/N ratios obtained with the ODSonly and ALL1 fingerprints decline from the late 1990s through 2016 (2014 in the case of ALL1; see Fig. 7C). For both of these fingerprints, however, S/N ratios remain above the stipulated 1% significance threshold, even for signal trends sampling the recovery era. This result suggests that even with an emerging “healing” signal in the early 21st century, the overall loss in upper stratospheric ozone over the entire 1984 to 2014 analysis period is still significantly larger than can be explained by WACCM-based estimates of internally generated and solar forced changes in ozone. The close agreement between the ALL1 and ODSonly S/N ratios indicates that human-caused changes in stratospheric ozone are the dominant contributor to forced ozone changes in ALL1. In the GHGonly case, the S/N ratio becomes less negative as the trend length L increases. This reflects the observed increase in upper stratospheric ozone in the early 2000s, which projects positively onto the GHGonly fingerprint (Fig. 5B). In summary, use of a standard fingerprint identification method allows us to positively detect (in SWOOSH ozone data) the upper stratospheric ozone fingerprints in response ODS forcing alone and in response to combined ODS and GHG forcing. For all three fingerprints, however, the β_l signal trends display pronounced nonlinear behavior because

of the observed “depletion followed by recovery”. This nonlinear behavior is not accounted for in the first (β_l based) signal detection method, and decreases S/N ratios [as the analysis period lengthens](#).

To address this problem, we also calculated S/N ratios using the above-described method 2, which relies on regressing the $c\{F, O\}(t)$ time series onto each of the principal component time series for the individual ODSonly, GHGonly, and ALL1 fingerprints (see Figs. 5A-C). These “signal” regression coefficients, β_{PC_1} , are shown as a function of L , the analysis period length (Fig. 7; D-F). It is instructive to compare the S/N results on the left and right panels of Fig. 7, which highlights differences between the linear and nonlinear signal detection approaches (i.e., between methods 1 and 2, respectively). As L increases and the analysis periods sample both ozone depletion and recovery, use of method 2 markedly increases S/N ratios relative to a purely linear representation of signal trends. This enhancement of S/N occurs because method 2 incorporates information about the nonlinear behavior of the upper stratospheric ozone signal (behavior that is common to the real world and the WACCM ODSonly and ALL1 simulations). Further information on the enhancement of S/N is provided in Table 2.

There are several other noteworthy features of the method 2 results in Fig. 7. First, the method 1 and method 2 S/N ratios for the GHGonly fingerprint show qualitatively and quantitatively similar behavior. This similarity is due to the fact that PC_1 for the GHGonly fingerprint (Fig. 5B) is well-described by a linear trend, so the regression coefficients between the linear PC_1 and the nonlinear $c\{F, O\}(t)$ time series change markedly as L increases. Second, S_L , the standard deviation of the sampling distribution of noise trends, shows some differences in the method 1 and method 2 cases (Fig. 7-B& and E). In method 1, S_L decreases in amplitude with increasing L ; this holds for all three fingerprints. In contrast, the method 2 results show that for the ODSonly and ALL1 fingerprints, S_L has a local minimum for $L=18$ years, and then increases slightly for longer analysis periods. This local minimum in S_L is absent in the GHGonly results for method 2.

The differences between the method 1 and method 2 S_L results are probably related to multiple factors. These include: 1) the strong influence of solar forcing on upper stratospheric ozone; 2) the periodicity in solar forcing in the NAT simulation; 3) the fact that the dominant mode of ozone variability in NAT is more similar to the ODSonly and ALL1 fingerprint patterns than to the GHGonly fingerprint (see pattern correlation results in upper triangle of Table 1); and 4) the quasi-linearity of the PC_1 time series for the GHGonly fingerprint, and the nonlinearity of the PC_1 time series for the ODSonly and ALL1 fingerprints (see Fig. 5).¹

¹ Recall that for method 2, the “noise” regression coefficients are calculated between maximally overlapping L -year segments of $c\{F, N\}(t)$ (the projection of the combined NAT and CTL data onto the ODSonly, GHGonly, or ALL1 fingerprint) and an L -year segment of the PC_1 time series for the ODSonly, GHGonly, or ALL1 fingerprint. As the L -year analysis window is being advanced through $c\{F, N\}(t)$, there will be (for certain values of L) times of random agreement in the phasing of solar variability and the nonlinear behavior in the ODSonly and ALL1 PC_1 time series. Such agreement would be expected to yield more complex behavior in S_L as a function of L .

For the upper stratosphere, we also explored whether our S/N results were robust to the choice of ozone dataset. This involved replicating the SWOOSH-based S/N analysis with the SBUV_CDR upper stratospheric ozone dataset, which was available from 1979 to 2015. The SBUV S/N analysis was performed with the same model fingerprints shown in Figs. 5A-C. Relative to the SWOOSH results, S/N ratios were generally slightly larger when calculated with SBUV data. As in the SWOOSH case, the SBUV results show that use of method 2 yields a noticeable enhancement of S/N ratios for the ODSonly and ALL1 fingerprints (see Figure S4-S5 in supplementary material). In summary, the method 2 results for the upper stratosphere show (relative to method 1) higher S/N ratios and more confident identification of the ODSonly and ALL1 fingerprints in observations. Even with method 2, however, the GHGonly ozone signal cannot be detected by the final year of the SWOOSH record (2016).

Next, we examine the lower stratosphere. Because of the above-described differences in the post-2005 behavior of tropical lower stratospheric ozone in ALL1 and SWOOSH (see Fig. 1C), we partition our lower stratospheric S/N analysis into two cases: for a global domain, and for a domain poleward of 30°S to 30°N (i.e., excluding the tropics). Figures 8A-F show the signal trends, noise trends, and S/N ratios for the lower stratosphere, estimated with methods 1 and 2 (~~respectively~~). Consider the β_l signal trends first. As in the case of the upper stratosphere, the behavior of β_l as a function of the trend length L is similar for $c\{F, O\}(t)$ time series obtained with the ODSonly and ALL1 fingerprints. These similarities are greater when the tropics are excluded from the analysis. Inclusion or exclusion of the tropics also impacts the β_l signal trends obtained with the GHGonly fingerprint: in the global analysis, trends ending after 2000 are slightly positive, while in the “tropics excluded” case, trends ending after 2000 are slightly negative. The S/N ratios for the β_l trends (Fig. 8 – bottom left panel), show that in the lower stratosphere, both the ALL1 and ODSonly fingerprint patterns are detectable in the SWOOSH observations. This holds for the global and “tropics excluded” domains. The GHGonly fingerprint is not statistically identifiable in either domain by the end of the SWOOSH record.

Recall that in the upper stratosphere, the use of method 2 yielded ~~markedly~~ S/N ratios for the ODSonly and ALL1 fingerprints that were markedly larger than for method 1. A large enhancement of S/N ratios is not evident in the lower stratosphere (Fig. 8 C&F). A small S/N enhancement occurs for ODSonly and ALL1 fingerprints, but only when the tropics are excluded. The smaller differences between the method 1 and method 2 S/N results (relative to the upper stratosphere) are likely due to the larger variability of ozone in the lower stratosphere (see Fig. 3). As in the case of method 1, use of method 2 leads to consistent detection of the ODSonly and ALL1 fingerprints in observed lower stratospheric ozone data, but does not yield positive identification of the GHGonly fingerprint.

In summary, we find that for both the upper and lower stratosphere, the observed decline and emerging recovery of stratospheric ozone is strongly influenced by secular changes in anthropogenic chlorofluorocarbons. These secular changes reflect scientific recognition of the serious consequences of ozone depletion, and the eventual formulation and implementation

of the Montreal Protocol. We were able to identify the model-predicted latitude/altitude patterns of ozone depletion in observations of upper and lower stratospheric ozone loss. We were not able to identify the model-predicted ozone changes arising from human-caused increases in well-mixed GHGs. Our results show that in the upper stratosphere, ozone depletion and recovery introduce pronounced nonlinearity in both modeled and observed ozone changes. Accounting for this common nonlinear behavior in our signal detection method (in our method 2) substantially amplifies S/N ratios. This amplification is largest in the upper stratosphere, where the nonlinear behavior is clearest and noise levels are lowest (see Table 2).

7 Discussion

In this study, we used the SWOOSH stratospheric ozone dataset and simulations performed with the Whole Atmosphere Community Climate Model (WACCM) to evaluate the relative detectability of ozone changes arising from forcing by ozone depleting substances (ODS), greenhouse gases (GHG), and combined changes in ODS and GHG. Our focus was on the period covered by the SWOOSH data (1984 to 2016). Our detection study relied on zonal-mean latitude-height profiles (“fingerprints”) of ozone change. The lower and upper stratosphere were considered separately in this investigation. This separation was made because the forcing by ODS and GHG has different ozone-change signatures in the lower and upper stratosphere, and because the amplitude and patterns of ozone variability differ in these two atmospheric regions.

The credibility of our findings rests on the reliability of the WACCM-based estimates of natural climate variability. Before conducting our signal detection study, we first investigated the skill of the WACCM model in capturing the observed variability of global-mean lower and upper stratospheric ozone. This comparison was performed using high-pass and band-pass filtered data, which isolated variability on short (less than 3 years) and long (5-20 years) timescales, respectively. We found that in the lower stratosphere, the WACCM simulation incorporating variability associated with the QBO yielded short-term variability that was closest to the SWOOSH-based estimate. Reliable simulation of the longer-timescale variability of ozone required the inclusion of solar forcing, particularly in the upper stratosphere (where solar-forced ozone changes are largest). In the simulation of historical climate change with combined ODS and GHG forcing (which is the simulation most relevant for comparison with observations), the amplitude of the longer-timescale variability is larger than in SWOOSH in the lower stratosphere, and slightly smaller than in SWOOSH in the upper stratosphere.

In global-mean terms, the upper and lower stratospheric ozone changes from the ODS+GHG simulation best matches the observed 20th century ozone depletion and early 21st century ozone recovery (see Fig. 1). In the full pattern analysis of zonal-mean latitude-altitude changes in ozone, all three anthropogenic fingerprints (ODS, GHG, and ODS+GHG) have coherent, large-scale structure. The three fingerprints are spatially dissimilar to the smaller-scale (and opposite-signed) structure of the dominant modes of internal and solar variability. The ODS and GHG fingerprints have similar large-scale structure in the

upper stratosphere, but are of opposite sign, and have distinctly different time evolution over the observational record (see Figs. 5A-C).

We applied two different fingerprint identification methods. The first (our method 1) has been routinely used in a number of detection and attribution studies. It assumes that the time evolution of the fingerprint pattern is quasi-linear over the length of the observational record. In the case of stratospheric ozone, however, the time evolution of the ODS and ODS+GHG fingerprints (which we refer to as “signal time series”) is markedly nonlinear because of 20th century stratospheric ozone depletion and early 21st century ozone recovery. This means that method 1 – which relies on linear fits to both signal and noise time series² – is sub-optimal for assessing the S/N properties of stratospheric ozone changes.

In contrast, our method 2 explicitly accounts for the nonlinear time evolution of the ODS and ODS+GHG fingerprints. In the upper stratosphere, where this nonlinear behavior is clearest, method 2 yields S/N ratios for ODS and ODS+GHG fingerprints that are markedly larger than those obtained with method 1. In both the upper and lower stratosphere, only the ODS+GHG and ODS only fingerprint patterns were detectable before the current final year (2016) of the SWOOSH ozone dataset. The GHG-driven fingerprint of ozone changes was not statistically identifiable in either the upper or lower stratospheric SWOOSH data, irrespective of whether we employed method 1 or method 2.

Our results illustrate the importance of explicitly considering the forced, nonlinear temporal changes in the signal of interest. Here, both the 20th century ozone depletion and the emerging early 21st century ozone recovery are driven by well-understood temporal changes in ODS forcing. The time structure of the ODS ozone signal is very different from the much more linear increase in ozone caused by increases in GHG levels (see Figs. 5A-C). Our findings suggest that these differences in time evolution are key to confident separation of the ODS and GHG ozone signals. While one previous detection and attribution study involving stratospheric ozone considered time information, it was not able to clearly separate ODS and GHG signals (Gillett et al., 2011). The key advantage here is that we are dealing with an observational record that extends to 2016 compared to 2005 in Gillett et al., and thus are better to capture the emerging recovery of stratospheric ozone, and the nonlinear behavior of the ODS-driven ozone signal. Figure 9 shows a schematic comparing method 1 and our method 2 which can be applied to attribution studies where there is evidence that the climate forcing has a temporal component that is changing nonlinearly over historical records. Although our study was confined to using spatial fingerprints in the latitude/altitude domain, the method is equally suitable for other domains, such as for longitude/latitude fingerprints. Not shown in Fig. 9 is the calculation of the standard deviation of the sampling distribution for the “noise” regression coefficients, $s(L)$; Noise calculations use ~~but~~ ~~follows~~ the same approach applied for the signal: projection of the combined noise simulation data onto the fingerprint to

²As described above, “noise time series” are obtained by projecting model estimates of naturally caused changes in ozone (both internally generated, and arising from solar forcing) onto the ODS, GHG, and ODS+GHG fingerprints.

retrieve $c\{F,N\}(t)$, followed by ~~and then~~ the calculation of ~~calculating~~ regression coefficients. The regressions rely on ~~from~~ L -year segments of either $c\{F,N\}(t)$ and time (in the case of method 1) or $c\{F,N\}(t)$ and the leading PC time series of the model ALL1, ODSonly, or GHGonly simulation (in the case of method 2).

As noted in section 2, there is a discrepancy between simulated and observed SWOOSH tropical lower stratospheric ozone post-2005. This is also a region where there is considerable divergence between different observational estimates of ozone changes. (WMO,2014). Both of these factors (model-data differences and observational uncertainty) motivated the additional analysis with exclusion of ozone changes in the tropical lower stratosphere from our S/N analysis. In the “tropics excluded” case, there is high confidence in our detection of the model ODS signal in the lower stratosphere. For the two reasons outlined above, we have less confidence in the interpretation of our S/N results for the global “tropics included” case, especially for method 2 where the PC time series are used. Due to the noticeable divergence between simulated and observed post-2005 ozone changes in the tropical lower stratosphere, inclusion of the tropics reduces S/N ratios for our method 2 (i.e., the temporal evolution of ozone differs over for the last decade in observations and in the ODSonly signal, thus reducing the regression coefficient in method 2).

The above-described results pertain to the identification of the model anthropogenic fingerprints in observations. A related question is the detection time of the GHG signal – the point in the future at which the currently undetectable GHG signal might become identifiable if we continued to monitor stratospheric ozone. We can obtain a purely model-based estimate of this detection time by using the ALL2 ensemble members that extend to the end of the 21st century, and treating these ensemble members as surrogate observations.

As noted above, S/N ratios for the GHG fingerprint were approaching the 1% significance level (our stipulated signal detection threshold) near the end of SWOOSH observational record (see Figs. 7 and 8). To investigate the expected GHG fingerprint detection time in “model world”, we use the same upper and lower stratospheric GHG fingerprint patterns that we searched for in the SWOOSH ozone data (see Figs. 5B and 6B). We also estimate the “model world” detection time for the two ODS fingerprints (Figs. 5A and 6A). In each case, ozone data from individual realizations of the ALL2 simulation are projected onto the GHG and ODS fingerprints yielding the signal time series for our method 1 S/N analysis. For the method 2 S/N analysis, we additionally projected the 20th and 21st century GHG (ODS) ozone anomaly data onto the GHG (ODS) fingerprints, and then regressed the projection time series obtained with ALL2 onto the projection time series obtained with the GHG (ODS) data. The method 1 and method 2 estimates of the denominator of the S/N ratio were calculated as described in Section 6, but are now computed for analysis periods spanning the range 10 to 67 years.

Results in Figure 10 are combined for the upper and lower stratosphere. The “model only” analysis clearly illustrates the critical importance of accounting for nonlinearity in the behavior of the ODS signal. Linear trends provide a reasonable fit to

the initial ozone depletion phase. As L increases, however, linear trends are not an accurate representation of the nonlinear “ozone depletion followed by ozone recovery”. In method 1, therefore (which relies on linear trends), the “model world” S/N ratio for the ODS fingerprint is initially above the detectability threshold, but then dips and remains below this threshold. This transition from a significant to a non-significant result occurs around 2012 for the upper stratosphere and between 2010 and 2020 in the lower stratosphere (see Figs. 10A and C). Since the time evolution of the GHG-induced stratospheric ozone changes is more linear, method 1 yields “model world” detection of the GHG fingerprint by roughly 2025 in the upper stratosphere and by 2012-2014 in the lower stratosphere.

Because method 2 accounts for nonlinearity in the time evolution of the ODS signal, the “model world” S/N ratios remain above the stipulated 1% significance threshold for all values of the analysis period L , both for the upper and the lower stratosphere (see Figs. 10B and D). This is a strikingly different result from the method 1 S/N ratios for the ODS fingerprint. The GHG fingerprint is also consistently identifiable in method 2, with “model world” detection times similar to those estimated for method 1. The model results suggest, therefore, that the GHG fingerprint may not be identifiable in the SWOOSH data for at least another 5 to 10 years.

Our study relies on a single global climate model (WACCM). It would be important to determine whether the S/N results obtained here are consistent with those inferred from other climate models. Such a multi-model assessment will require use of models with a well-resolved stratosphere, and with reasonable representation of observed ozone variability. This will be the focus of subsequent work.

Acknowledgements

The authors would like to thank Sean Davis and Karen Rosenlof, from the National Oceanic & Atmospheric Administration, Chemical Sciences Division, for access to the SWOOSH ozone dataset. The National Center for Atmospheric Research is funded by the National Science Foundation.

References

- Allen, M. R. and Tett, S. F. B.: Checking for model consistency in optimal fingerprinting, *Clim. Dyn.*, 15(6), 419–434, doi:10.1007/s003820050291, 1999.
- Aquila, V., Swartz, W. H., Waugh, D. W., Colarco, P. R., Pawson, S., Polvani, L. M. and Stolarski, R. S.: Isolating the roles of different forcing agents in global stratospheric temperature changes using model integrations with incrementally added single forcings, *J. Geophys. Res. Atmos.*, 121(13), 8067–8082, doi:10.1002/2015JD023841, 2016.

- Arfeuille, F., Luo, B. P., Heckendorn, P., Weisenstein, D., Sheng, J. X., Rozanov, E., Schraner, M., Brönnimann, S., Thomason, L. W. and Peter, T.: Modeling the stratospheric warming following the Mt. Pinatubo eruption: uncertainties in aerosol extinctions, *Atmos. Chem. Phys.*, 13(22), 11221–11234, doi:10.5194/ACP-13-11221-2013, 2013.
- Austin, J., Bourqui, M., Braesicke, P., Brühl, C., Butchart, N., Considine, D. B., Cunnold, D., Fahey, D. W., Fleming, E. L., De, P. M., Forster, F., Geller, M. A., Godin-Beekmann, S., Grewe, V., Haigh, J. D., Jackman, C. H., Johnston, P., Langematz, U., Law, K. S., Liley, J. B., Megretskaya, I. A., Miller, A. J., Nagashima, T., Pitari, G., Portmann, R. W., Ramaswamy, V., Rognerud, B., Rosenfield, J. E., Ross, M. N., Schnadt, C., Schwarzkopf, M. D., Shindell, D. T., Smith, C. A., Smyshlyaev, S., Stolarski, R. S., Velders, G. J. M., Wang, R., Weisenstein, D. K., Wu, F., Zander, R. J., Chipperfield, M. P., Randel, W. J., Bodeker, G. E., Dameris, M., Fioletov, V. E., Friedl, R. R., Harris, N. R. P., Logan, J. A., Mcpeters, R. D., Muthama, N. J., Peter, T., Shepherd, T. G., Shine, K. P., Solomon, S., Thomason, L. W. and Zawodny, J. M.: Global Ozone: Past and Future, in *Scientific Assessment of Ozone Depletion: 2002*, Geneva., 2002.
- Barnett, T. P., Pierce, D. W., Achutha Rao, K. M., Gleckler, P. J., Santer, B. D., Gregory, J. M. and Washington, W. M.: Penetration of Human-Induced Warming into the World's Oceans, *Science* (80-.), 309(5732), 284–287, 2005.
- Bindoff, N. ., Stott, P. A., Achutha Rao, K. M., Allen, M. R., Gillett, N., Gutzler, K., Hansingo, K., Hegerl, G., Hu, Y., Jain, S., Mokhov, I. I., Overland, J., Perlwitz, J., Sebbani, R. and Zhang, X.: Detection and Attribution of Climate Change, in *Climate Change 2013: The Physical Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge, UK, UK., 2013.
- Bourassa, A. E., Degenstein, D. A., Randel, W. J., Zawodny, J. M., Kyrölä, E., McLinden, C. A., Sioris, C. E. and Roth, C. Z.: Trends in stratospheric ozone derived from merged SAGE II and Odin-OSIRIS satellite observations, *Atmos. Chem. Phys.*, 14(13), 6983–6994, doi:10.5194/acp-14-6983-2014, 2014.
- Boville, B.: Wave-mean flow interactions in a general circulation model of the troposphere and stratosphere, *J. Atmos. Sci.*, 1986.
- Brewer, A. W.: Evidence for a World Circulation Provided by the Measurements of Helium and Water Vapour Distribution in the Stratosphere, *Quarterly J. R. Meteorological Soc.*, 75(326), 351–363, 1949.
- Brönnimann, S., Bhand, J., Franke, J., Flückiger, S., Fischer, A. M., Bleisch, R., Bodeker, G., Hassler, B., Rozanov, E. and Schraner, M.: A global historical ozone data set and prominent features of stratospheric variability prior to 1979, *Atmos. Chem. Phys.*, 13(18), 9623–9639, 2013.
- 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(13), 7059–7074, doi:10.5194/acp-14-7059-2014, 2014.
- Davis, S. M., Rosenlof, K. H., Hassler, B., Hurst, D. F., Read, W. G., Vömel, H., Selkirk, H., Fujiwara, M. and Damadeo, R.: The Stratospheric Water and Ozone Satellite Homogenized (SWOOSH) database: a long-term database for climate studies, *Earth Syst. Sci. Data*, 8, 461–490, doi:10.5194/essd-8-461-2016, 2016.
- Dobson, G. M. B.: Origin and Distribution of the Polyatomic Molecules in the Atmosphere, *Proc. R. Soc. A Math. Phys. Eng. Sci.*, 236(1205), 187–193, doi:10.1098/rspa.1956.0127, 1956.

- Egorova, T., Rozanov, E., Manzini, E., Haberreiter, M., Schmutz, W., Zubov, V. and Peter, T.: Chemical and dynamical response to the 11-year variability of the solar irradiance simulated with a chemistry-climate model, *Geophys. Res. Lett.*, 31(6), n/a-n/a, doi:10.1029/2003GL019294, 2004.
- Eyring, V., Cionni, I., Lamarque, J. F., Akiyoshi, H., Bodeker, G. E., Charlton-Perez, A. J., Frith, S. M., Gettelman, A., Kinnison, D. E., Nakamura, T., Oman, L. D., Pawson, S. and Yamashita, Y.: Sensitivity of 21st century stratospheric ozone to greenhouse gas scenarios, *Geophys. Res. Lett.*, 37(16), n/a-n/a, doi:10.1029/2010GL044443, 2010.
- Eyring, V., Lamarque, J. F. and Hess, P.: Overview of IGAC/SPARC Chemistry-Climate Model Initiative (CCMI) Community Simulations in Support of Upcoming Ozone and Climate Assessments, in *SPARC Newsletter No. 40*, pp. 48–66., 2013.
- Eyring, V., Shepherd, T. G. and Waugh, D. W., Eds.: *SPARC CCMVal Report on the Evaluation of Chemistry-Climate Models.*, SPARC Repo., WMO/TD., n.d.
- Fels, S.: Stratospheric sensitivity to perturbations in ozone and carbon dioxide: Radiative and dynamical response, *J. ...*, 1980.
- Fioletov, V. E. and Shepherd, T. G.: Summertime total ozone variations over middle and polar latitudes, *Geophys. Res. Lett.*, 32(4), L04807, doi:10.1029/2004GL022080, 2005.
- Fleming, E. L., Jackman, C. H., Stolarski, R. S. and Douglass, A. R.: A model study of the impact of source gas changes on the stratosphere for 1850–2100, *Atmos. Chem. Phys.*, 11(16), 8515–8541, doi:10.5194/acp-11-8515-2011, 2011.
- Garcia, R. R., Randel, W. J., Garcia, R. R. and Randel, W. J.: Acceleration of the Brewer-Dobson Circulation due to Increases in Greenhouse Gases, *J. Atmos. Sci.*, 65(8), 2731–2739, doi:10.1175/2008JAS2712.1, 2008.
- Gillett, N. P.: Detecting anthropogenic influence with a multi-model ensemble, *Geophys. Res. Lett.*, 29(20), 1970, doi:10.1029/2002GL015836, 2002.
- Gillett, N. P., Akiyoshi, H., Bekki, S., Braesicke, P., Eyring, V., Garcia, R., Karpechko, A. Y., 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(2), 599–609, doi:10.5194/acp-11-599-2011, 2011.
- Hansen, J., Sato, M., Ruedy, R., Nazarenko, L., Lacis, A., Schmidt, G. A., Russell, G., Aleinov, I., Bauer, M., Bauer, S., Bell, N., Cairns, B., Canuto, V., Chandler, M., Cheng, Y., Del Genio, A., Faluvegi, G., Fleming, E., Friend, A., Hall, T., Jackman, C., Kelley, M., Kiang, N., Koch, D., Lean, J., Lerner, J., Lo, K., Menon, S., Miller, R., Minnis, P., Novakov, T., Oinas, V., Perlwitz, J. J., Perlwitz, J. J., Rind, D., Romanou, A., Shindell, D., Stone, P., Sun, S., Tausnev, N., Thresher, D., Wielicki, B., Wong, T., Yao, M. and Zhang, S.: Efficacy of climate forcings, *J. Geophys. Res.*, 110(D18), D18104, doi:10.1029/2005JD005776, 2005.
- Hasebe, F.: Quasi-Biennial Oscillations of Ozone and Diabatic Circulation in the Equatorial Stratosphere, *J. Atmos. Sci.*, 51(5), 729–745, doi:10.1175/1520-0469(1994)051<0729:QBOOOA>2.0.CO;2, 1994.
- Hasselmann, K.: On the signal-to-noise problem in atmospheric response studies, *Meteorol. Trop. Ocean.*, 251–259, 1979.
- Hasselmann, K.: Optimal Fingerprints for the Detection of Time-dependent Climate Change, *J. Clim.*, 6(10), 1957–1971, doi:10.1175/1520-0442(1993)006<1957:OFFTDO>2.0.CO;2, 1993.
- Hegerl, G., Von Storch, H., Hasselmann, K., Santer, B. D., Cubasch, U. and Jones, P. D.: Detecting greenhouse-gas-induced

- climate change with an optimal fingerprint method, *J. Clim.*, 9(10), 2281–2306, 1996.
- Hegerl, G. C., Zwiers, F., Braconnot, P., Gillet, N. P., Luo, Y., Marengo, J., Nicholls, N., Penner, J. and Stott, P. A.: Understanding and attributing climate change., in *Climate Change 2007: The Physical Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by S. Solomon, Cambridge University Press, Cambridge, UK., 2007.
- 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. Discuss.*, 9(1), 1157–1209, doi:10.5194/acpd-9-1157-2009, 2009.
- Karl, T. R., Hassol, S. J., Miller, C. D. and Murray, W. .: Temperature Trends in the Lower Atmosphere: Steps for Understanding and Reconciling Differences. A report by the U.S. Climate Change Science Program and the Subcommittee on Global Change Research, Asheville, NC, NC., 2006.
- Karoly, D., Cohen, J., Meehl, G., Mitchell, J., Oort, A. H. A. A. H., Souffer, R. J., Wetherald, R. T. R. R. T., Stouffer, R., Wetherald, R. T. R. R. T., Souffer, R. J. and Wetherald, R. T. R. R. T.: An example of fingerprint detection of greenhouse climate change, *Clim. Dyn.*, 10(1–2), 97–105, doi:10.1007/BF00210339, 1994.
- Kirgis, G., Leblanc, T., McDermid, I. S. and Walsh, T. D.: Stratospheric ozone interannual variability (1995–2011) as observed by lidar and satellite at Mauna Loa Observatory, HI and Table Mountain Facility, CA, *Atmos. Chem. Phys.*, 13(9), 5033–5047, doi:10.5194/acp-13-5033-2013, 2013.
- Kuttippurath, J., Bodeker, G. E., Roscoe, H. K. and Nair, P. J.: A cautionary note on the use of EESC-based regression analysis for ozone trend studies, *Geophys. Res. Lett.*, 42(1), 162–168, doi:10.1002/2014GL062142, 2015.
- Kyrölä, E., Laine, M., Sofieva, V., Tamminen, J., Päiväranta, S.-M., Tukiainen, S., Zawodny, J. and Thomason, L.: Combined SAGE II–GOMOS ozone profile data set for 1984–2011 and trend analysis of the vertical distribution of ozone, *Atmos. Chem. Phys.*, 13(21), 10645–10658, doi:10.5194/acp-13-10645-2013, 2013.
- Laine, M., Latva-Pukkila, N. and Kyrölä, E.: Analysing time-varying trends in stratospheric ozone time series using the state space approach, *Atmos. Chem. Phys.*, 14(18), 9707–9725, doi:10.5194/acp-14-9707-2014, 2014.
- Langematz, U., Schmidt, F., Kunze, M., Bodeker, G. E. and Braesicke, P.: Antarctic ozone depletion between 1960 and 1980 in observations and chemistry–climate model simulations, *Atmos. Chem. Phys.*, 16(24), 15619–15627, doi:10.5194/acp-16-15619-2016, 2016.
- Lean, J.: Evolution of the Sun’s Spectral Irradiance Since the Maunder Minimum, *Geophys. Res. Lett.*, 27(16), 2425–2428, doi:10.1029/2000GL000043, 2000.
- Lee, H. and Smith, A. K.: Simulation of the combined effects of solar cycle, quasi-biennial oscillation, and volcanic forcing on stratospheric ozone changes in recent decades, *J. Geophys. Res.*, 108(D2), 4049, doi:10.1029/2001JD001503, 2003.
- Li, F., Stolarski, R. S. and Newman, P. A.: Stratospheric ozone in the post-CFC era, *Atmos. Chem. Phys. Atmos. Chem. Phys.*, 9, 2207–2213, 2009.
- Marsh, D. R., Mills, M. J., Kinnison, D. E., Lamarque, J.-F., Calvo, N. and Polvani, L. M.: Climate Change from 1850 to 2005

- Simulated in CESM1(WACCM), *J. Clim.*, 26(19), 7372–7391, doi:10.1175/JCLI-D-12-00558.1, 2013.
- Maycock, A. C., Matthes, K., Tegtmeier, S., Thiéblemont, R. and Hood, L.: The representation of solar cycle signals in stratospheric ozone – Part 1: A comparison of recently updated satellite observations, *Atmos. Chem. Phys.*, 16(15), 10021–10043, doi:10.5194/acp-16-10021-2016, 2016.
- 5 McInden, C. A., Tegtmeier, S. and Fioletov, V.: Technical Note: A SAGE-corrected SBUV zonal-mean ozone data set, *Atmos. Chem. Phys.*, 9, 7963–7972, 2009.
- Merkel, A. W., Harder, J. W., Marsh, D. R., Smith, A. K., Fontenla, J. M. and Woods, T. N.: The impact of solar spectral irradiance variability on middle atmospheric ozone, *Geophys. Res. Lett.*, 38(13), n/a-n/a, doi:10.1029/2011GL047561, 2011.
- Molina, M. J. and Rowland, F. S.: Stratospheric sink for chlorofluoromethanes: chlorine atom-catalysed destruction of ozone, 10 *Nature*, 249(5460), 810–812, 1974.
- Morgenstern, O., Hegglin, M. I., Rozanov, E., O’connor, F. M., 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 15 Initiative (CCMI), *Geosci. Model Dev*, 10, 639–671, doi: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(20), 10373–10384, doi:10.5194/acp-13-10373-2013, 2013.
- Newchurch, M. J., Yang, E., Cunnold, D. M., Reinsel, G. C., Zawodny, J. M. and Russell, J. M.: Evidence for slowdown in 20 stratospheric ozone loss: First stage of ozone recovery, *J. Geophys. Res.*, 108(D16), 4507, doi:10.1029/2003JD003471, 2003.
- Newman, P. and Daniel, J.: A new formulation of equivalent effective stratospheric chlorine (EESC), *Atmos. Chem. Phys.*, 7, 4537–4552, 2007.
- North, G. R., Stevens, M. J., North, G. R. and Stevens, M. J.: Detecting Climate Signals in the Surface Temperature Record, *J. Clim.*, 11(4), 563–577, doi:10.1175/1520-0442(1998)011<0563:DCSITS>2.0.CO;2, 1998.
- 25 Oman, L. D., Plummer, D. A., Waugh, D. W., Austin, J., Scinocca, J. F., Douglass, A. R., Salawitch, R. J., Canty, T., Akiyoshi, H., Bekki, S., Braesicke, P., Butchart, N., Chipperfield, M. P., Cugnet, D., Dhomse, S., Eyring, V., Frith, S., Hardiman, S. C., Kinnison, D. E., Lamarque, J.-F., Mancini, E., Marchand, M., Michou, M., Morgenstern, O., Nakamura, T., Nielsen, J. E., Olivié, D., Pitari, G., Pyle, J., Rozanov, E., Shepherd, T. G., Shibata, K., Stolarski, R. S., Teyssèdre, H., Tian, W., Yamashita, Y. and Ziemke, J. R.: Multimodel assessment of the factors driving stratospheric ozone evolution over the 21st century, *J. 30 Geophys. Res.*, 115(D24), D24306, doi:10.1029/2010JD014362, 2010.
- Oman, L. D., Douglass, A. R., Ziemke, J. R., Rodriguez, J. M., Waugh, D. W. and Nielsen, J. E.: The ozone response to ENSO in Aura satellite measurements and a chemistry-climate simulation, *J. Geophys. Res. Atmos.*, 118(2), 965–976, doi:10.1029/2012JD018546, 2013.
- 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., 2014.

Polvani, L. M., Wang, L., Aquila, V., Waugh, D. W., Polvani, L. M., Wang, L., Aquila, V. and Waugh, D. W.: The Impact of Ozone-Depleting Substances on Tropical Upwelling, as Revealed by the Absence of Lower-Stratospheric Cooling since the Late 1990s, *J. Clim.*, 30(7), 2523–2534, doi:10.1175/JCLI-D-16-0532.1, 2017.

- 5 Randel, W. J. and Thompson, A. M.: Interannual variability and trends in tropical ozone derived from SAGE II satellite data and SHADOZ ozonesondes, *J. Geophys. Res.*, 116(D7), D07303, doi:10.1029/2010JD015195, 2011.

Randel, W. J. and Wu, F.: A stratospheric ozone profile data set for 1979–2005: Variability, trends, and comparisons with column ozone data, *J. Geophys. Res.*, 112(D6), D06313, doi:10.1029/2006JD007339, 2007.

Randel, W. J., Wu, F., Randel, W. J. and Wu, F.: Isolation of the Ozone QBO in SAGE II Data by Singular-Value Decomposition, *J. Atmos. Sci.*, 53(17), 2546–2559, doi:10.1175/1520-0469(1996)053<2546:IOTOQI>2.0.CO;2, 1996.

- 10 Randel, W. J., Shine, K. P., Austin, J., Barnett, J., Claud, C., Gillett, N. P., Keckhut, P., Langematz, U., Lin, R., Long, C., Mears, C., Miller, A., Nash, J., Seidel, D. J., Thompson, D. W. J., Wu, F. and Yoden, S.: An update of observed stratospheric temperature trends, *J. Geophys. Res.*, 114(D2), D02107, doi:10.1029/2008JD010421, 2009.

Santer, B., Wigley, T. and Jones, P.: Correlation methods in fingerprint detection studies, *Clim. Dyn.*, 1993.

- 15 Santer, B. ., Wigley, T. M. L., Meehl, G. A., Wehner, M. F., Mears, C., Schabel, M., Wentz, F. J., Ammann, C., Arblaster, J., Bettge, T., Washington, W. M., Taylor, K. E., Boyle, J. S., Bruggemann, W. and Doutriaux, C.: Influence of Satellite Data Uncertainties on the Detection of Externally Forced Climate Change, *Science* (80-.), 300(5623), 1280–1284, 2003.

Santer, B. B. D., Painter, J. F. J., Mears, C. a, Doutriaux, C., Caldwell, P., Arblaster, J. M., Cameron-Smith, P. J., Gillett, N. P., Gleckler, P. J., Lanzante, J., Perlwitz, J., Solomon, S., Stott, P. a, Taylor, K. E., Terray, L., Thorne, P. W., Wehner, M. F.,

- 20 Wentz, F. J., Wigley, T. M. L., Wilcox, L. J. and Zou, C.-Z.: Identifying human influences on atmospheric temperature., *Proc. Natl. Acad. Sci. U. S. A.*, 110(1), 26–33, doi:10.1073/pnas.1210514109, 2013a2.

Santer, B. D., Taylor, K. E., Wigley, T. M. L., Penner, J. E., Jones, P. D. and Cubasch, U.: Towards the detection and attribution of an anthropogenic effect on climate, *Clim. Dyn.*, 12(2), 77–100, doi:10.1007/BF00223722, 1995.

Santer, B. D., Taylor, K. E., Wigley, T. M. L., Johns, T. C., Jones, P. D., Karoly, D. J., Mitchell, J. F. B., Oort, A. H., Penner,

- 25 J. E., Ramaswamy, V., Schwarzkopf, M. D., Stouffer, R. J. and Tett, S.: A search for human influences on the thermal structure of the atmosphere, *Nature*, 382(6586), 39–46, doi:10.1038/382039a0, 1996a.

Santer, B. D., Wigley, T. M. L., Barnett, T. P. and Anyamba, E.: Detection of Climate Change and Attribution of Causes, edited by B. D. Santer, T. M. . Wigley, T. P. Barnett, and E. Anyamba, Cambridge University Press, New York, NY., 1996b.

Santer, B. D., Mears, C., Doutriaux, C., Caldwell, P., Gleckler, P. J., Wigley, T. M. L., Solomon, S., Gillett, N. P., Ivanova,

- 30 D., Karl, T. R., Lanzante, J. R., Meehl, G. A., Stott, P. A., Taylor, K. E., Thorne, P. W., Wehner, M. F. and Wentz, F. J.: Separating signal and noise in atmospheric temperature changes: The importance of timescale, *J. Geophys. Res. Atmos.*, 116(D22), n/a-n/a, doi:10.1029/2011JD016263, 2011.

Santer, B. D., Painter, J. F., Bonfils, C., Mears, C. A., Solomon, S., Wigley, T. M. L., Gleckler, P. J., Schmidt, G. A., Doutriaux, C., Gillett, N. P., Taylor, K. E., Thorne, P. W. and Wentz, F. J.: Human and natural influences on the changing thermal structure

- of the atmosphere., *Proc. Natl. Acad. Sci. U. S. A.*, 110(43), 6–11, doi:10.1073/pnas.1305332110, 2013b.
- Shindell, D. T. and Grewe, V.: Separating the influence of halogen and climate changes on ozone recovery in the upper stratosphere, *J. Geophys. Res.*, 107(12), 3–10, 2002.
- Solomon, S.: Stratospheric ozone depletion: A review of concepts and history, *Rev. Geophys.*, 37(3), 275–316, 1999.
- 5 Solomon, S. and Garcia, R. R.: On the distributions of long-lived tracers and chlorine species in the middle atmosphere, *J. Geophys. Res.*, 89(D7), 11633, doi:10.1029/JD089iD07p11633, 1984.
- Solomon, S., Kinnison, D., Bandoro, J. and Garcia, R.: Simulation of polar ozone depletion: An update, *J. Geophys. Res. Atmos.*, 120(15), 7958–7974, doi:10.1002/2015JD023365, 2015.
- Solomon, S., Ivy, D. J., Kinnison, D., Mills, M. J., Neely, R. R. and Schmidt, A.: Emergence of healing in the Antarctic ozone
10 layer, *Science* (80-.), 353(6296), 269–274, doi:10.1126/science.aae0061, 2016.
- Stolarski, R. S., Douglass, A. R., Steenrod, S., Pawson, S., Stolarski, R. S., Douglass, A. R., Steenrod, S. and Pawson, S.: Trends in Stratospheric Ozone: Lessons Learned from a 3D Chemical Transport Model, *J. Atmos. Sci.*, 63(3), 1028–1041, doi:10.1175/JAS3650.1, 2006a.
- Stolarski, R. S., Douglass, A. R., Steenrod, S., Pawson, S., Stolarski, R. S., Douglass, A. R., Steenrod, S. and Pawson, S.:
15 Trends in Stratospheric Ozone: Lessons Learned from a 3D Chemical Transport Model, *J. Atmos. Sci.*, 63(3), 1028–1041, doi:10.1175/JAS3650.1, 2006b.
- Stott, P. A., Tett, S. F. B., Jones, G. S., Allen, M. R., Mitchell, J. F. B. and Jenkins, G. J.: External Control of 20th Century Temperature by Natural and Anthropogenic Forcings, *Science* (80-.), 290(5499), 2133–2137, doi:10.1126/science.290.5499.2133, 2000.
- 20 Tett, S. F. B., Mitchell, J. F. B., Parker, D. E. and Allen, M. R.: Human Influence on the atmospheric vertical temperature structure: Detection and observations, *Science* (80-.), 274(5290), 1170–1173, 1996.
- Tett, S. F. B., Jones, G. S., Stott, P. A., Hill, D. C., Mitchell, J. F. B., Allen, M. R., Ingram, W. J., Johns, T. C., Johnson, C. E., Jones, A., Roberts, D. L., Sexton, D. M. H. and Woodage, M. J.: Estimation of natural and anthropogenic contributions to twentieth century temperature change, *J. Geophys. Res.*, 107(D16), 4306, doi:10.1029/2000JD000028, 2002.
- 25 Thorne, P. W., Jones, P. D., Osborn, T. J., Davies, T. D., Tett, S. F. B., Parker, D. E., Stott, P. A., Jones, G. S. and Allen, M. R.: Assessing the robustness of zonal mean climate change detection, *Geophys. Res. Lett.*, 29(19), 26-1-26-4, doi:10.1029/2002GL015717, 2002.
- Tie, X. and Brasseur, G.: The response of stratospheric ozone to volcanic eruptions: sensitivity to atmospheric chlorine loading, *Geophys. Res. Lett.*, 22(22), 3035–3038, 1995.
- 30 Trenberth, K.: Some effects of finite sample size and persistence on meteorological statistics. Part I: Autocorrelations, *Mon. Weather Rev.*, 112(12), 2459–2368, 1984.
- Tummon, F., Hassler, B., Harris, N. R. P., Staehelin, J., Steinbrecht, W., Anderson, J., Bodeker, G. E., Bourassa, A., Davis, S. M., Degenstein, D., Frith, S. M., Froidevaux, L., Kyrölä, E., Laine, M., Long, C., Penckwitt, A. A., Sioris, C. E., Rosenlof, K. H., Roth, C., Wang, H.-J. and Wild, J.: Intercomparison of vertically resolved merged satellite ozone data sets: interannual

variability and long-term trends, *Atmos. Chem. Phys.*, 15(6), 3021–3043, doi:10.5194/acp-15-3021-2015, 2015.

Vinnikov, K. Y., Robock, A., Stouffer, R. J. and Manabe, S.: Vertical patterns of free and forced climate variations, *Geophys. Res. Lett.*, 23(14), 1801–1804, doi:10.1029/96GL01736, 1996.

Wang, H. J., Cunnold, D. M., Thomason, L. W., Zawodny, J. M. and Bodeker, G. E.: Assessment of SAGE version 6.1 ozone data quality, *J. Geophys. Res. Atmos.*, 107(D23), ACH 8-1-ACH 8-18, doi:10.1029/2002JD002418, 2002.

Wigley, T. M. L., Jaumann, P. J., Santer, B. D. and Taylor, K. E.: Relative detectability of greenhouse-gas and aerosol climate change signals, *Clim. Dyn.*, 14(11), 781–790, doi:10.1007/s003820050254, 1998.

Wohltmann, I., Lehmann, R., Rex, M., Brunner, D. and Mäder, J. A.: A process-oriented regression model for column ozone, *J. Geophys. Res.*, 112(D12), D12304, doi:10.1029/2006JD007573, 2007.

10 Xue, X.-H., Liu, H.-L. and Dou, X.-K.: Parameterization of the inertial gravity waves and generation of the quasi-biennial oscillation, *J. Geophys. Res. Atmos.*, 117(D6), n/a-n/a, doi:10.1029/2011JD016778, 2012.

Zawodny, J. M. and McCormick, M. P.: Stratospheric Aerosol and Gas Experiment II measurements of the quasi-biennial oscillations in ozone and nitrogen dioxide, *J. Geophys. Res.*, 96(D5), 9371, doi:10.1029/91JD00517, 1991.

WMO (World Meteorological Organization), Scientific Assessment of Ozone Depletion: 2002, Global Ozone Research and Monitoring Project—Report No. 47, 498 pp., Geneva, 2003.

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

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

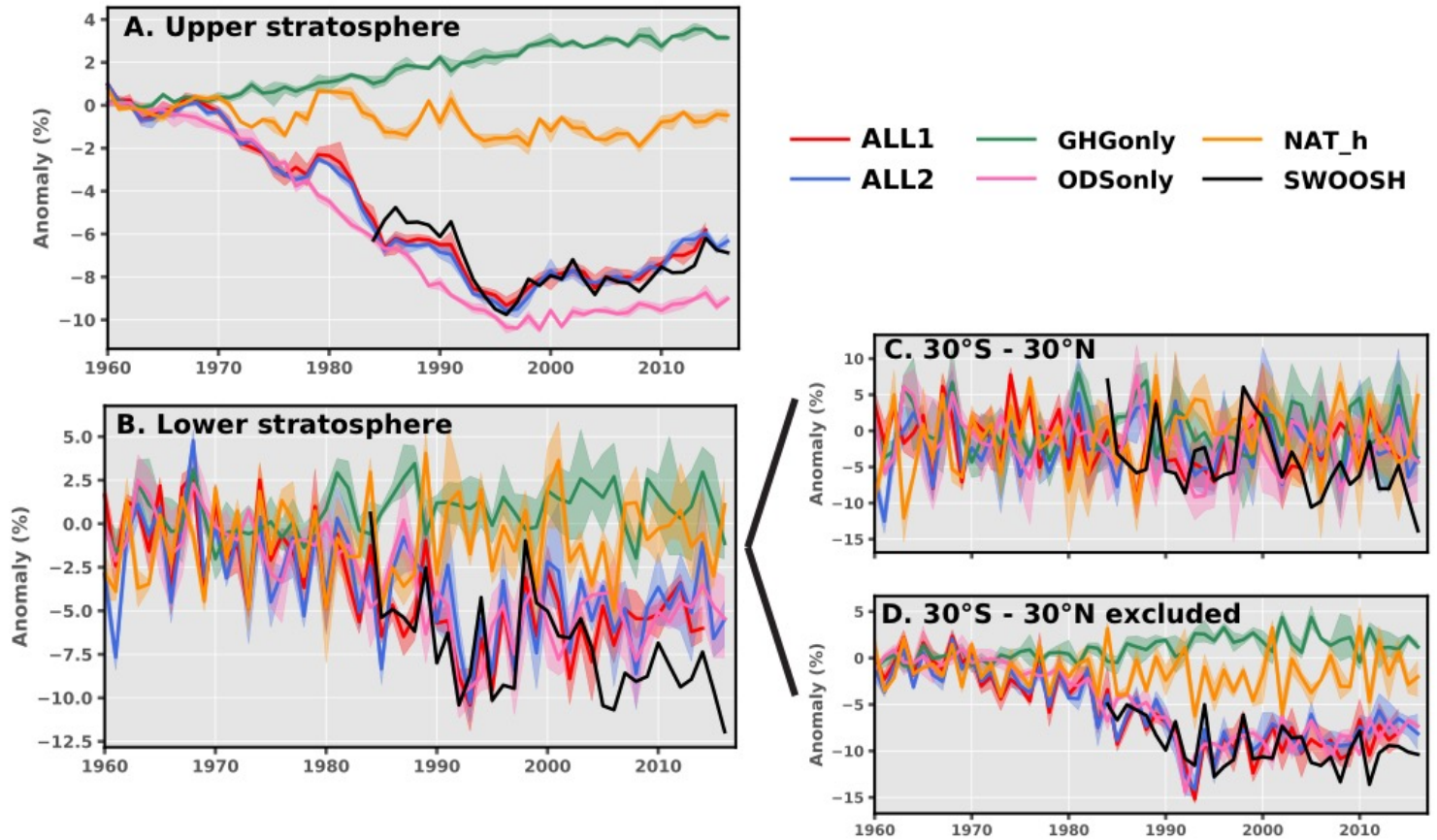
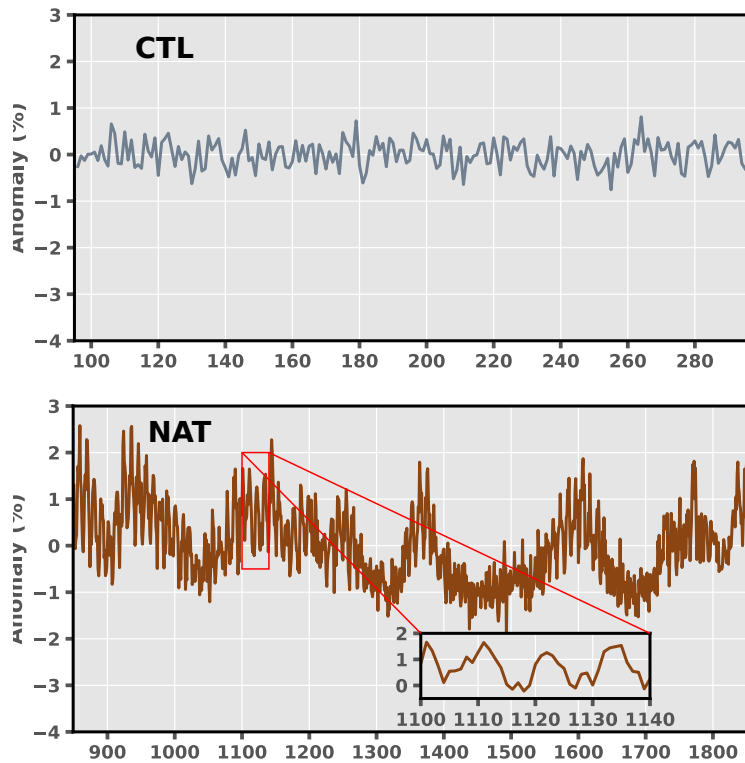


Figure 1: Time series of global mean annual ozone anomalies for A) the upper stratosphere (1-10 hPa) and B) the lower stratosphere, (40-100 hPa). The right panel shows the lower stratosphere for C) the tropics (30°S to 30°N) and D) tropics excluded. Annual anomalies were calculated from monthly-mean model output and SWOOSH ozone dataset, and both were **weighted by latitude and pressure**. All **model** anomalies are shown as percent differences from the base years of 1960-1969 in ALL1. For model simulations, the spread shows the upper and lower bounds of all ensemble **members about the mean**. The SWOOSH record spans **the period** from 1984-2016 and **the anomalies were also defined as percent differences from the base years of 1960-1969 in the ALL1 simulation**. **was forced to have the same mean as ALL1 over 1984-2016.**

A. Upper stratosphere



B. Lower Stratosphere

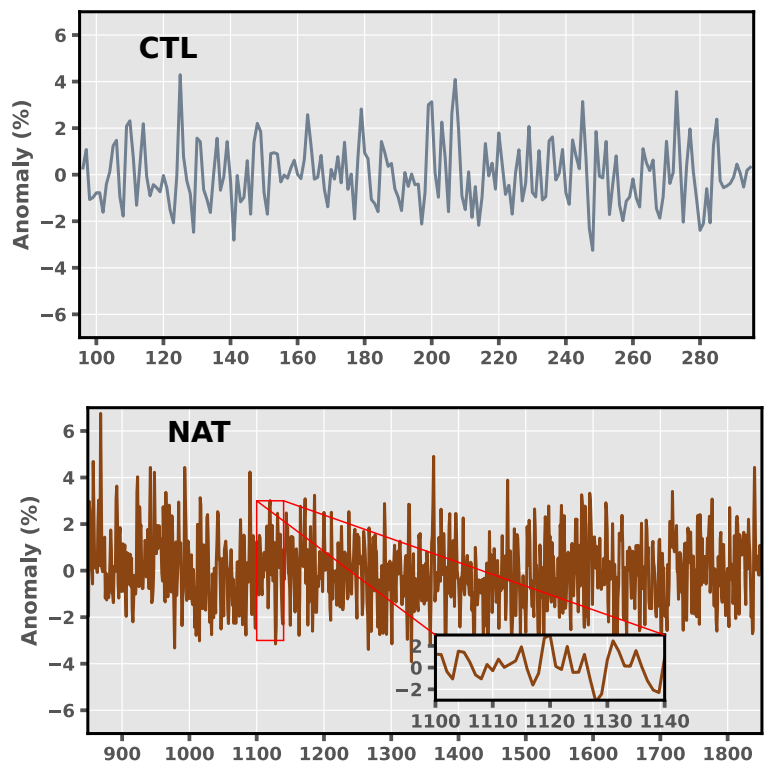


Figure 2: Same as Fig. 1 but for ozone anomalies from the pre-industrial control simulation (top), and the pre-industrial natural simulation (bottom) for the upper (A) and lower stratosphere (B). The NAT results show (as an inset) an expanded includes a 40-year segment of the anomaly data zoomed-in axis.

5

10

Simulated and observed temporal variations in stratospheric ozone

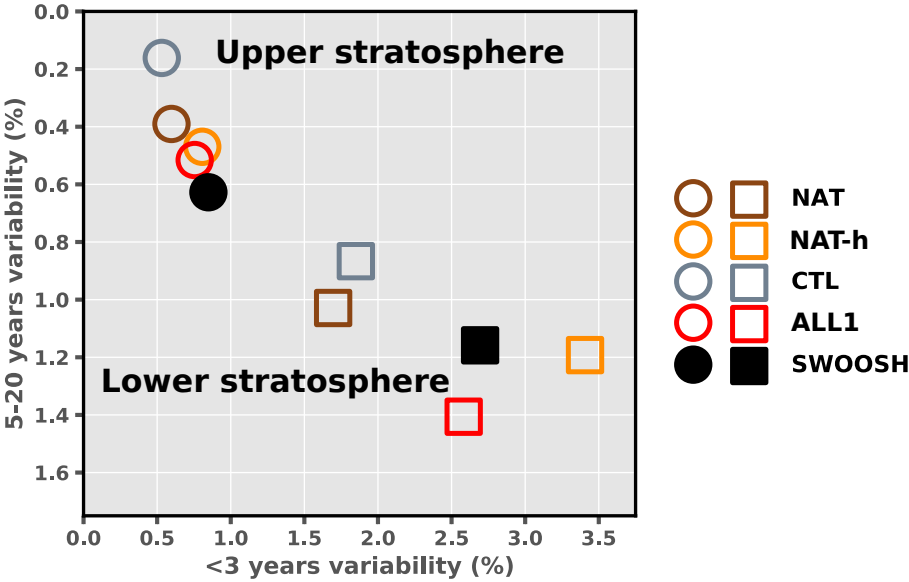


Figure 3: Comparison of simulated and observed temporal variability of global ozone anomalies in the upper and lower stratosphere. Results for the lower (upper) stratosphere are represented by square markers (circles) and the upper stratosphere is with circles. Colors correspond to the legend. Variability on monthly to interannual timescales is plotted versus variability on timescales of 5-20 years. After removal of long-term trends from raw anomalies with a 30 year low-pass filter, a Butterworth filter was used to perform band-pass and high-pass filtering. There is no overlap between the frequencies isolated by the high- and band-pass filters. Results for NAT, CTL and NAT-h are from 1000, 200 and 168 years respectively, while ALL1 and SWOOSH are from 1984-01 to 2014-12 and 2016-12 respectively. Temporal standard deviations for the band- and high-pass filtered data are given as the percent differences from the base years of 1960-1969 in ALL1. the variability, units are in %.

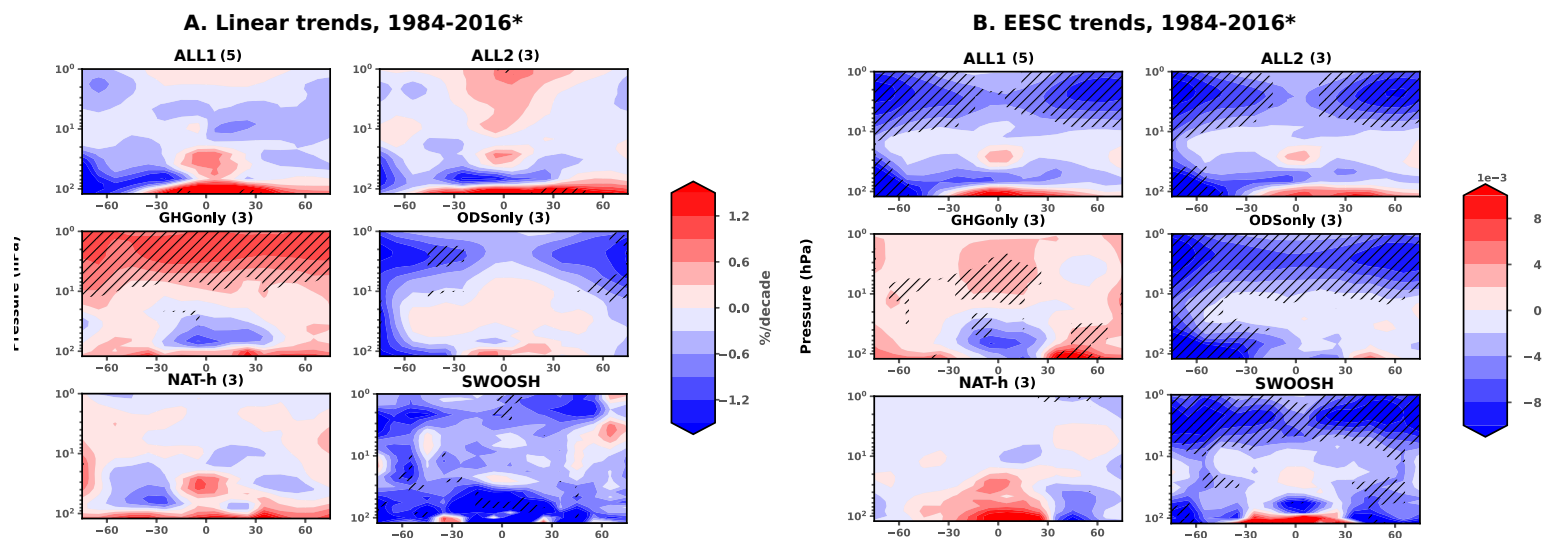


Figure 4: Zonal-mean stratospheric ozone trends estimated using A) linear regression and B) EESC regression of annual-mean anomalies from 1984-2016. For model simulations, trend estimates use the ensemble mean. Hatching indicates significance at the 95% level or better using a two-sided Student's *t*-test. FigureNumbers in brackets indicate the number of ensemble members. All of the trends shown are for 1984-2016, *except for ALL1 which is from 1984-2014.

5

10

15

20

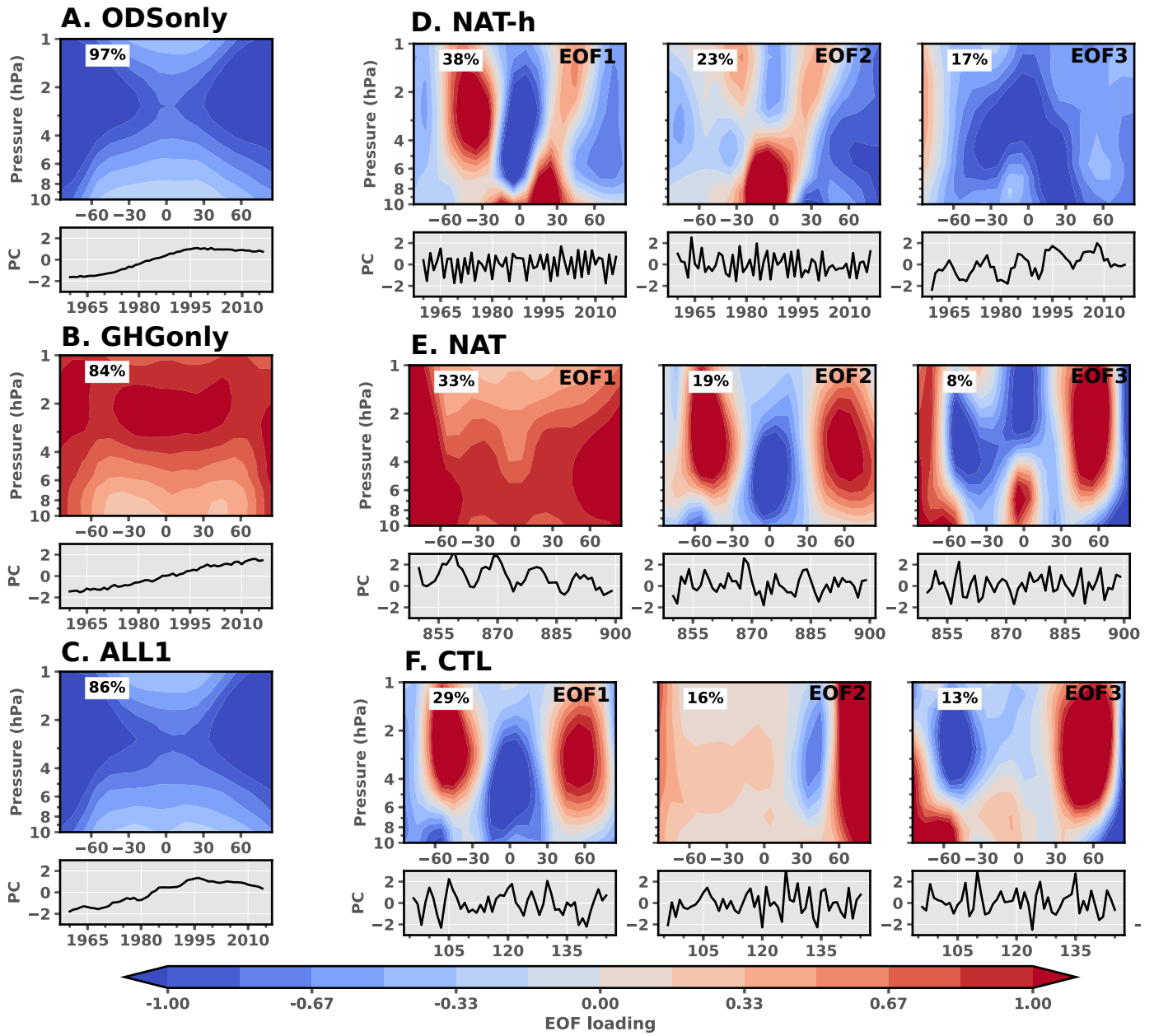


Figure 5: Leading signal and natural variability modes for upper stratospheric ozone in WACCM. The signal modes are the leading EOFs of the ensemble mean anomalies from GHGonly, ODSonly and ALL1 calculated **over** **from** 1960-2016 for GHGonly, and ODSonly and **over** 1960-2014 for ALL1. The leading natural variability modes are EOFs 1, 2 and 3 of the NAT, NAT-h, and CTL. The percentage variance explained by each mode is given in the upper left corner of the EOF. For NAT-h the EOFs were calculated from the ensemble-mean anomalies **from** **for the period** 1960-2016. **The total lengths of years for** the NAT, and CTL simulations are 1000, and 200 years respectively. **Recall that NAT-h is defined by subtraction** (see Section 2).

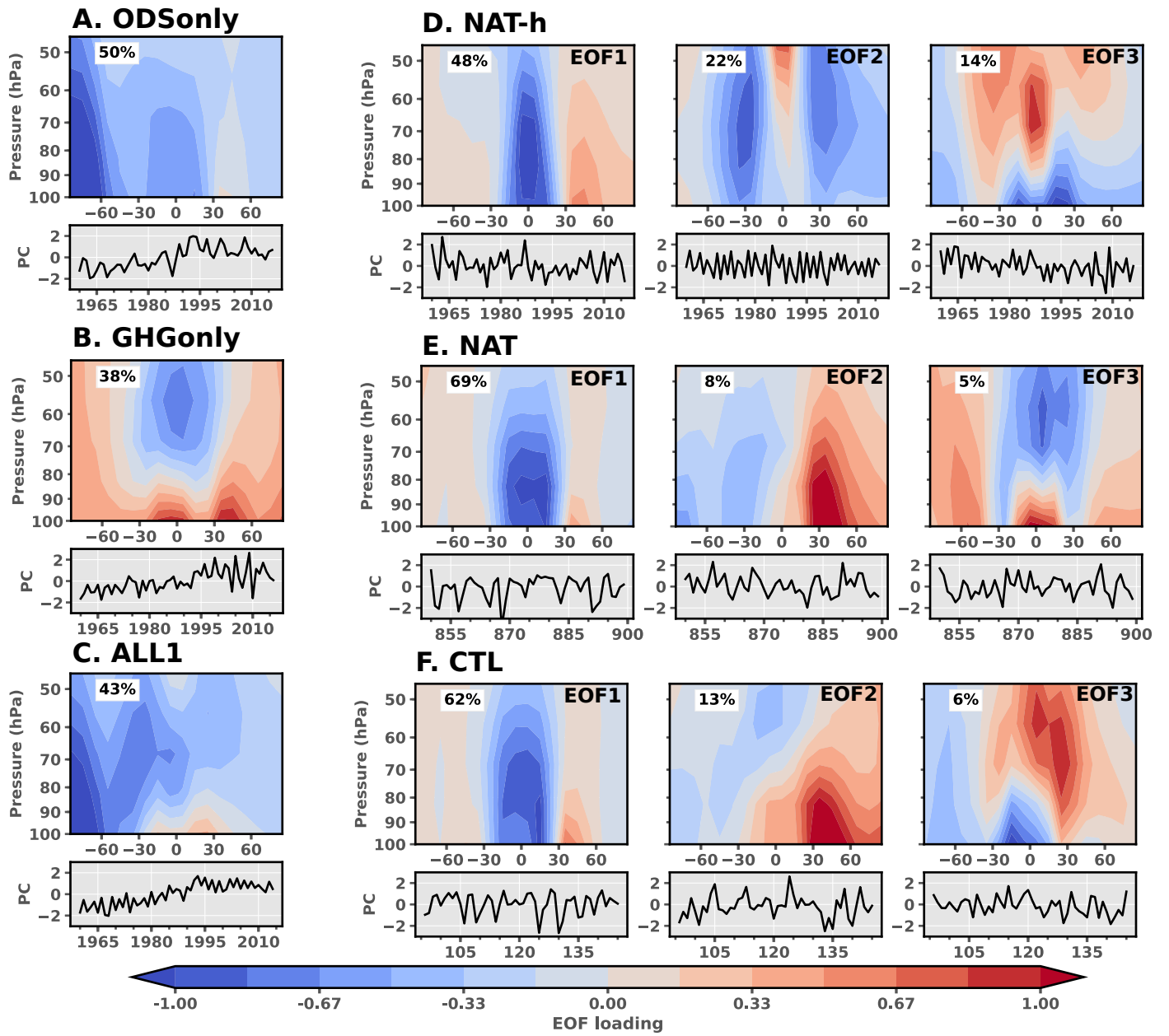


Figure 6: Same as Fig. 5 but for the lower stratosphere.

Upper Stratosphere

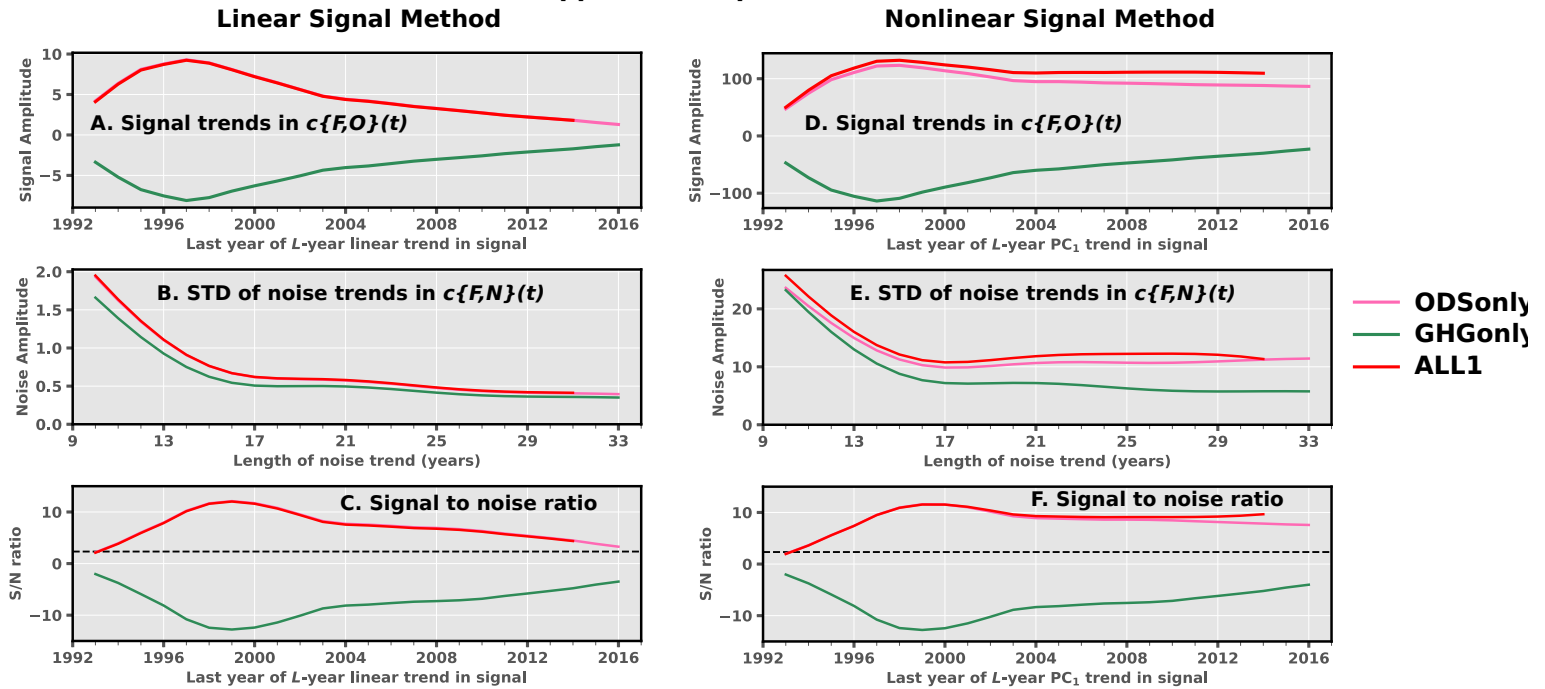
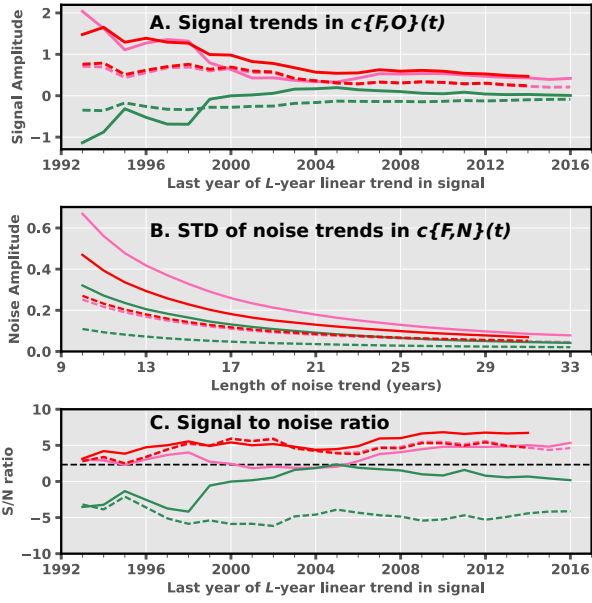


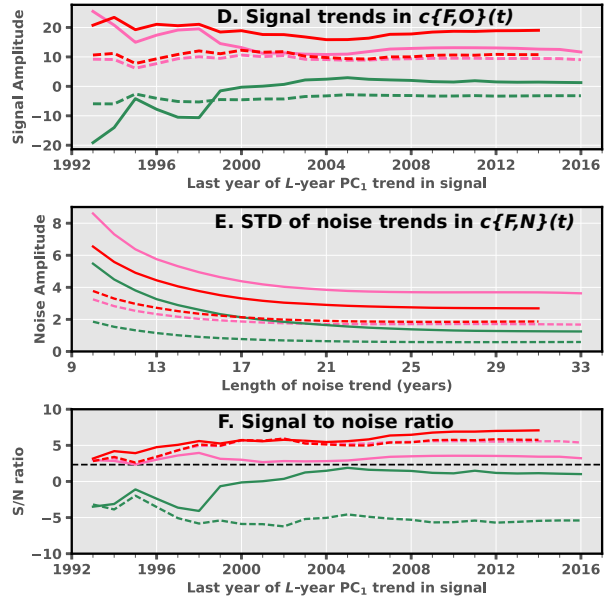
Figure 7: Results from the S/N analysis of simulated and observed changes in zonal-mean upper and lower stratospheric ozone. Signal time-series provide information on the similarity between the time-invariant ODSonly/GHGonly/ALL1 fingerprint pattern and the time-varying observed patterns of stratospheric ozone change. Signal detection relies on both simple linear regression (A-C; “method 1”) and regression between $c\{F,O\}(t)$ and the leading PC of the ODSonly, GHGonly, or ALL1 simulation the PC_1 (D-F; “method 2”). Results are a function of the analysis period L (in years). The L -year trends in these method 1 and method 2 signal time-series are plotted in the top row panel. The year on the abscissa is the end year of the L -length trend beginning in 1984. Noise time-series indicate the level of similarity between the searched-for ODSonly, GHGonly, and ALL1 fingerprints and the concatenated control CTL, NAT-h, and NAT and natural simulations for estimates of natural variability of ozone. The middle panels show the standard deviation of the distribution of maximally overlapping L -year trends in $c\{F,N\}(t)$. The S/N ratio is shown in the bottom panels. The dashed horizontal line indicates the stipulated significance level of 1% for signal detection.

Lower Stratosphere

Linear Signal Method



Nonlinear Signal Method



— ODSonly Global
- - ODSonly No Tropic
— GHGonly Global
- - GHGonly No Tropic
— ALL1 Global
- - ALL1 No Tropics

Figure 8: Same as Fig. 7 but for the lower stratosphere. The filled-solid line shows results for the global lower stratosphere, while the dashed lines exclude the tropics (defined from 30°S to 30°N).

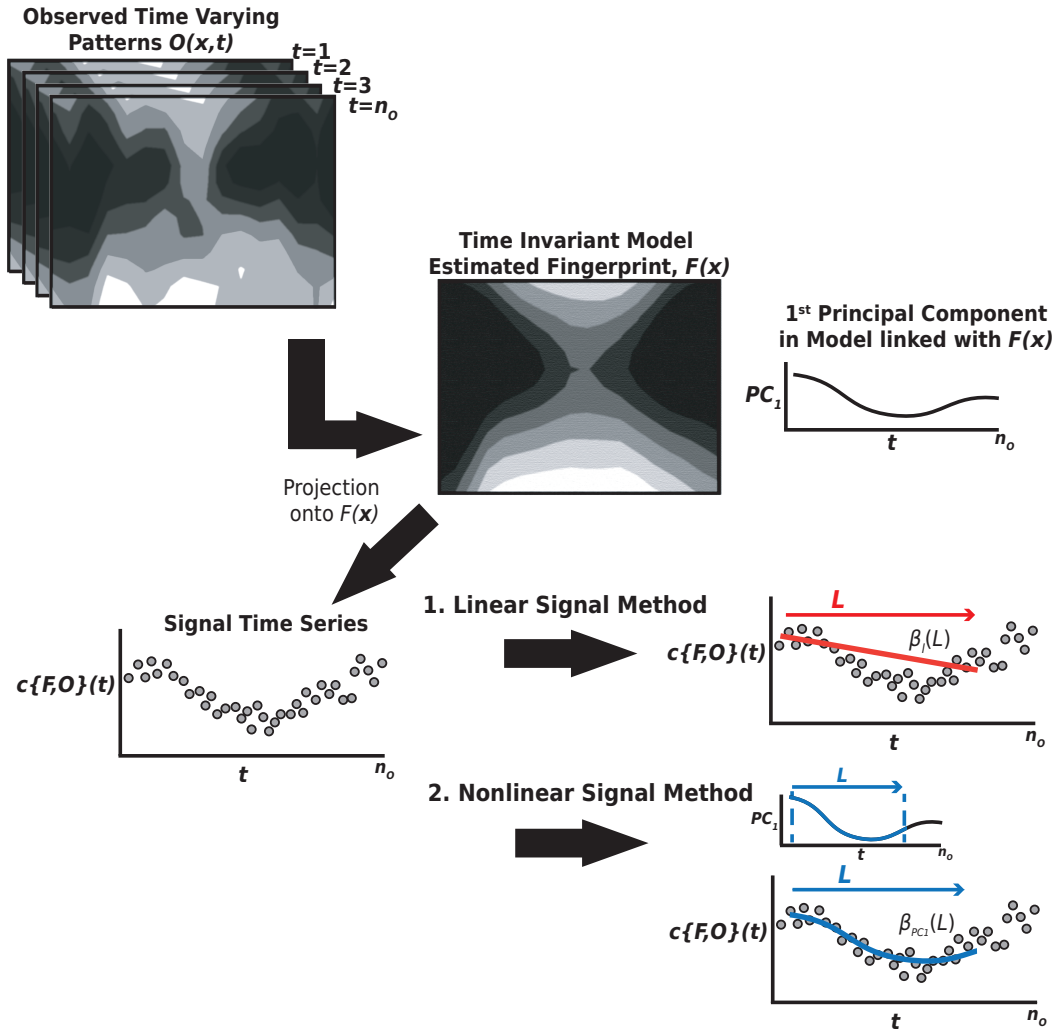


Figure 9: Schematic of the linear and nonlinear signal methods used for the fingerprint identification. $O(x,t)$ denotes the observed time-varying patterns of zonal-mean ozone anomalies; in the spatial dimension(s) of x is an index over the spatial dimensions (latitude and pressure). $F(x)$ is the searched-for spatial “fingerprint” - and is equal to the leading empirical orthogonal function of the zonal-mean ozone response in the ODSonly, GHGonly, or ALL1 simulation. from the model simulations. For each of the three fingerprints, there is an associated with $F(x)$, is the first leading principal component (PC1) time series (PC_1) which spans the same time period of the observational record. The observations are then projected onto the each fingerprint, yielding to retrieve the signal time series, $c\{F,O\}(t)$. In the linear signal method, it is assumed that the time evolution of the fingerprint pattern is quasi-linear over the length of the observational record, and L -length linear regression coefficients are calculated, between $c\{F,O\}(t)$ and the time in years; these are the coefficients, $\beta_l(L)$. In the nonlinear method, the regression is between L -year segments of $c\{F,O\}(t)$ and the model PC_1 of the corresponding fingerprint. These are the time series from PC1 corresponding the same length years is used to calculate the regression coefficients $\beta_{PC1}(L)$.

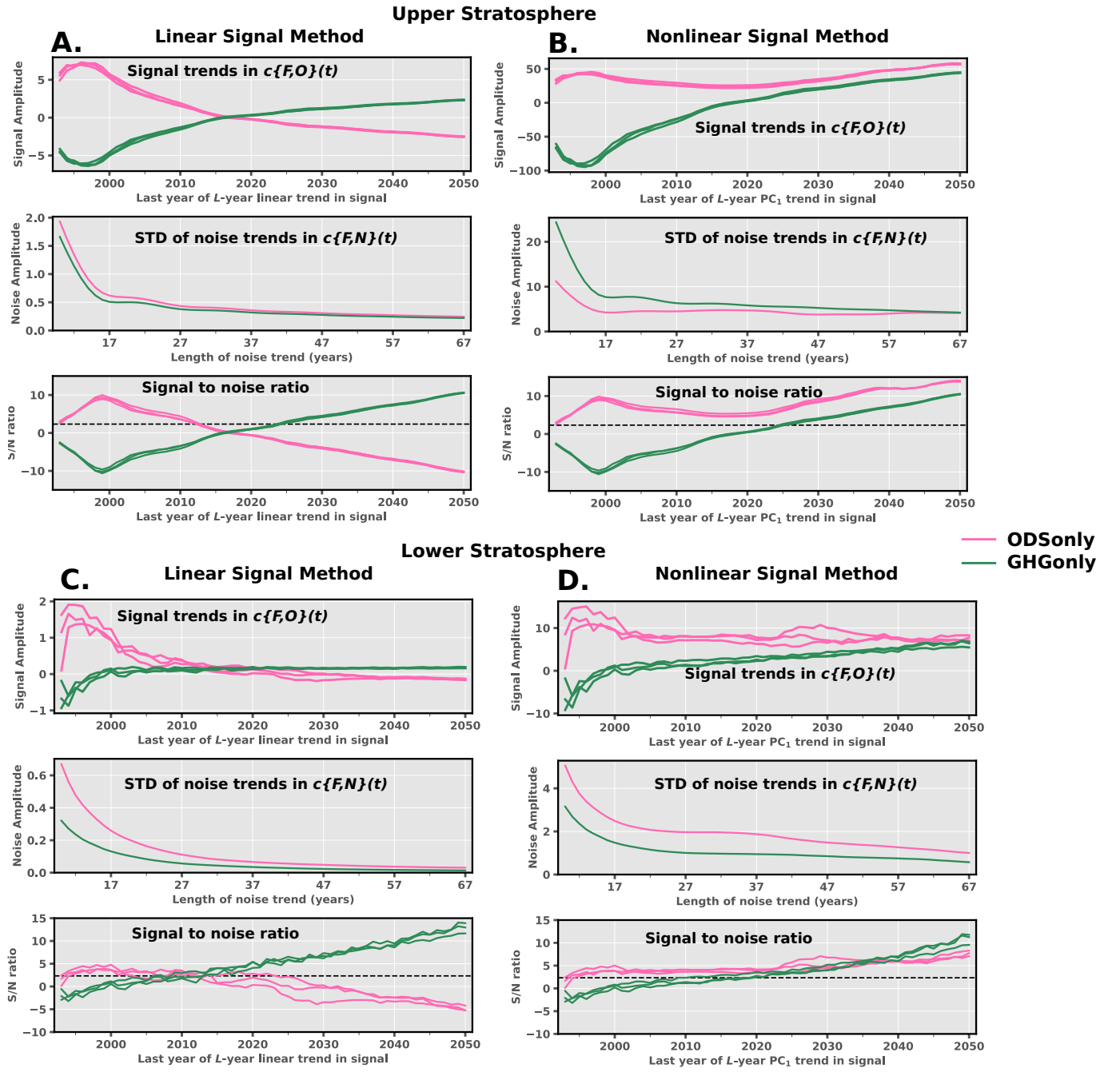


Figure 10: Same as in Figs. 7 and 8 but the ALL2 ensemble members (3 total) are used as the surrogate observations, yielding information regarding the estimated detection time for the ODSonly and GHGonly fingerprints in model data. The A) and B) are the β_I and β_{PC_1} results in A) and B) are respective trends for the upper stratosphere (respectively); results in C) and D) are the corresponding method 1 and method 2 for the same but for the lower stratosphere.

	ALL1	GHGonly	ODSonly	NAT-h	NAT	CTL
ALL1	--	-0.66	0.99	-0.18	-0.42	-0.32
GHGonly	0.19	--	-0.72	0.18	-0.10	0.19
ODSonly	0.72	-0.18	--	-0.13	-0.35	-0.38
NAT-h	0.19	-0.35	-0.13	--	0.19	0.46
NAT	-0.25	0.32	0.05	-0.93	--	-0.16
CTL	0.09	0.24	-0.03	0.90	0.97	--

Table 1: Correlation coefficients between the leading signal and noise patterns ~~in the study~~ for the upper stratosphere (top right triangle) and lower stratosphere (bottom left triangle). Signal patterns are the leading EOFs for ALL1, GHGonly and ODSonly; ~~The while the~~ noise patterns are the leading EOFs for NAT-h, NAT and CTL (~~see shown in~~ Figs. 5 and 6).

	Upper stratosphere		Lower stratosphere	
$F(x, p)$	β_l	β_{PC_1}	β_l	β_{PC_1}
ALL1	4.34	9.66	6.72 (4.68)	6.93 (5.82)
ODSonly	3.27	7.68	5.35 (4.43)	3.21 (5.42)
GHGonly	-3.49	-3.81	0.18 (-4.13)	0.32 (-4.89)

Table 2: ~~Longest L-year~~ ~~Signal-to-noise ratios~~ for the longest L -year analysis periods shown in Figs. 7 and 8. Results are for the linear (β_l) and nonlinear signal methods (β_{PC_1}). ~~F-L~~ for ODSonly and GHGonly, the longest analysis period is 33 years (1984-2016); For ALL1, the longest analysis period, and is 31 years for ALL1 (1984-2014). The figures in ~~B~~brackets for the lower stratosphere, indicate the signal-to-noise ratios when the tropics are excluded.

Supplementary Figures:

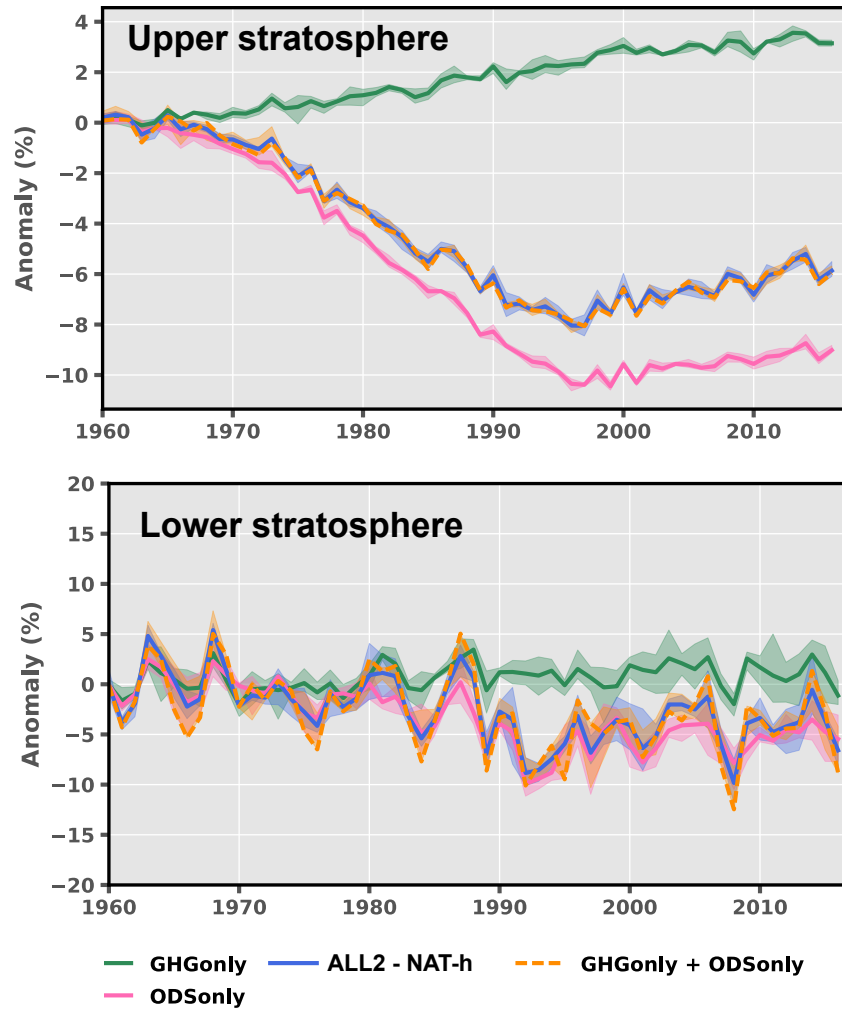


Figure S1: Same as Figs. 1A and 1B in the main text, but showing linear additivity of the GHGonly and ODSonly responses in the upper and lower stratosphere, compared to the ALL2 simulation with NAT-h subtracted.

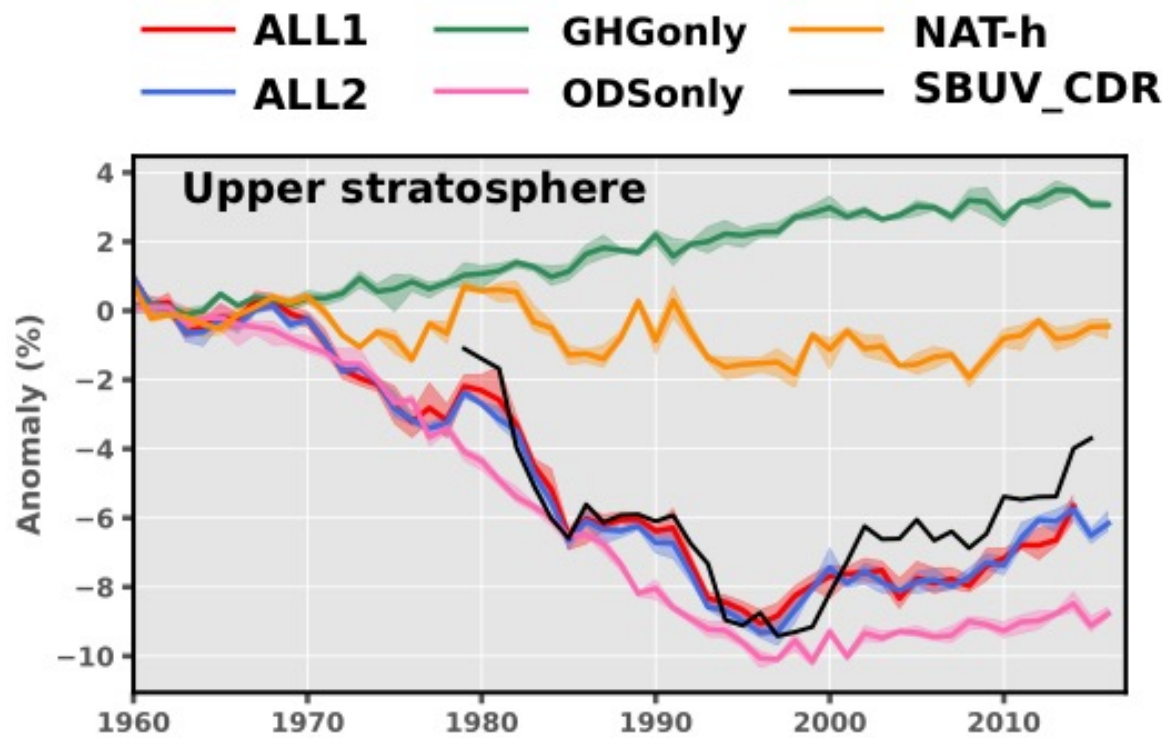


Figure S1S2: Same as Fig. 1A in the main text, but with the SBUV_CDR ozone data.

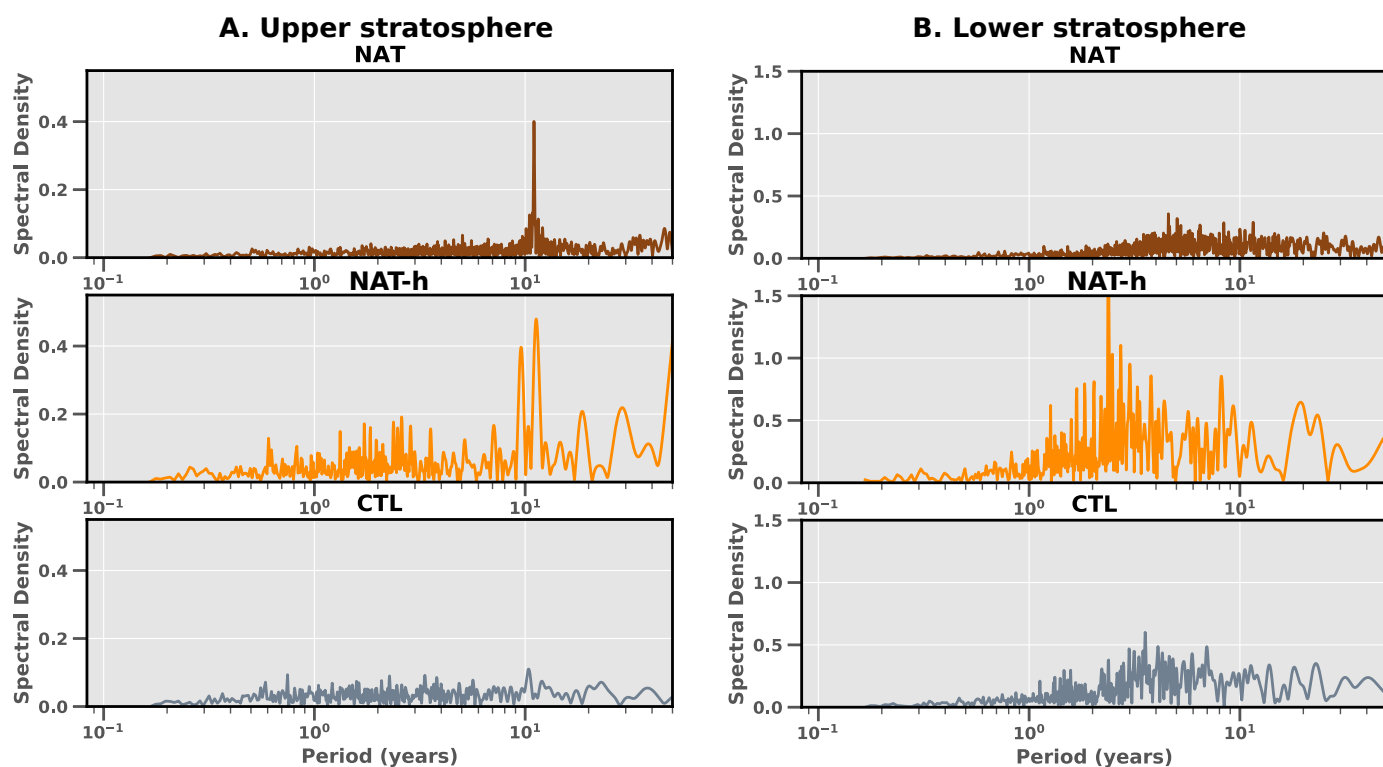


Figure S2S3: Normalized periodograms of the global stratospheric ozone anomalies shown in Fig 1 and 2 for NAT, NAT-h, and CTL. The power spectra were calculated using monthly anomalies. Note that the lowest frequency that can be resolved given the length of the observational ozone record is 16.5 years.

SBUV_CDR

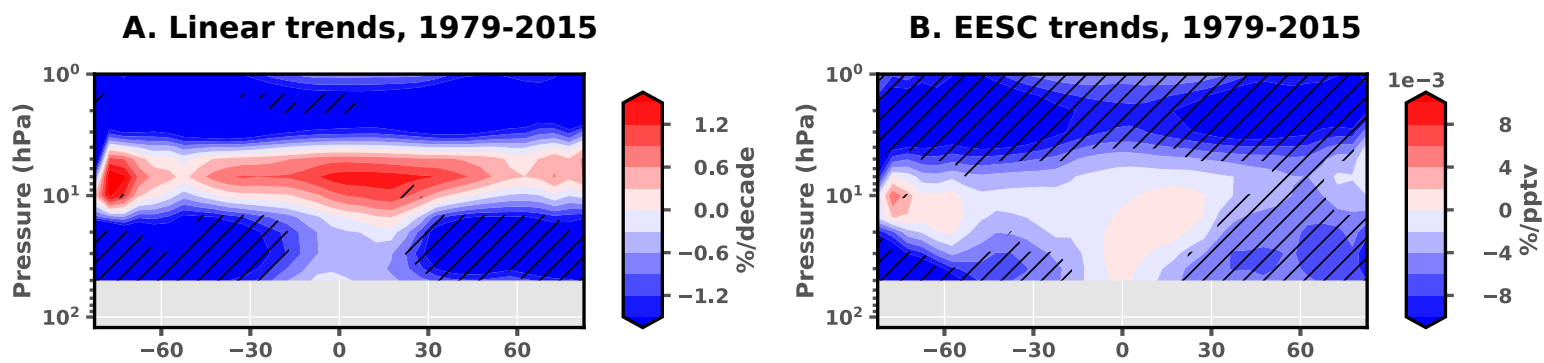


Figure S3S4: Same as Fig. 4A in the main text but with SBUV-Cohesive ozone dataset.

SBUV_CDR Upper Stratosphere

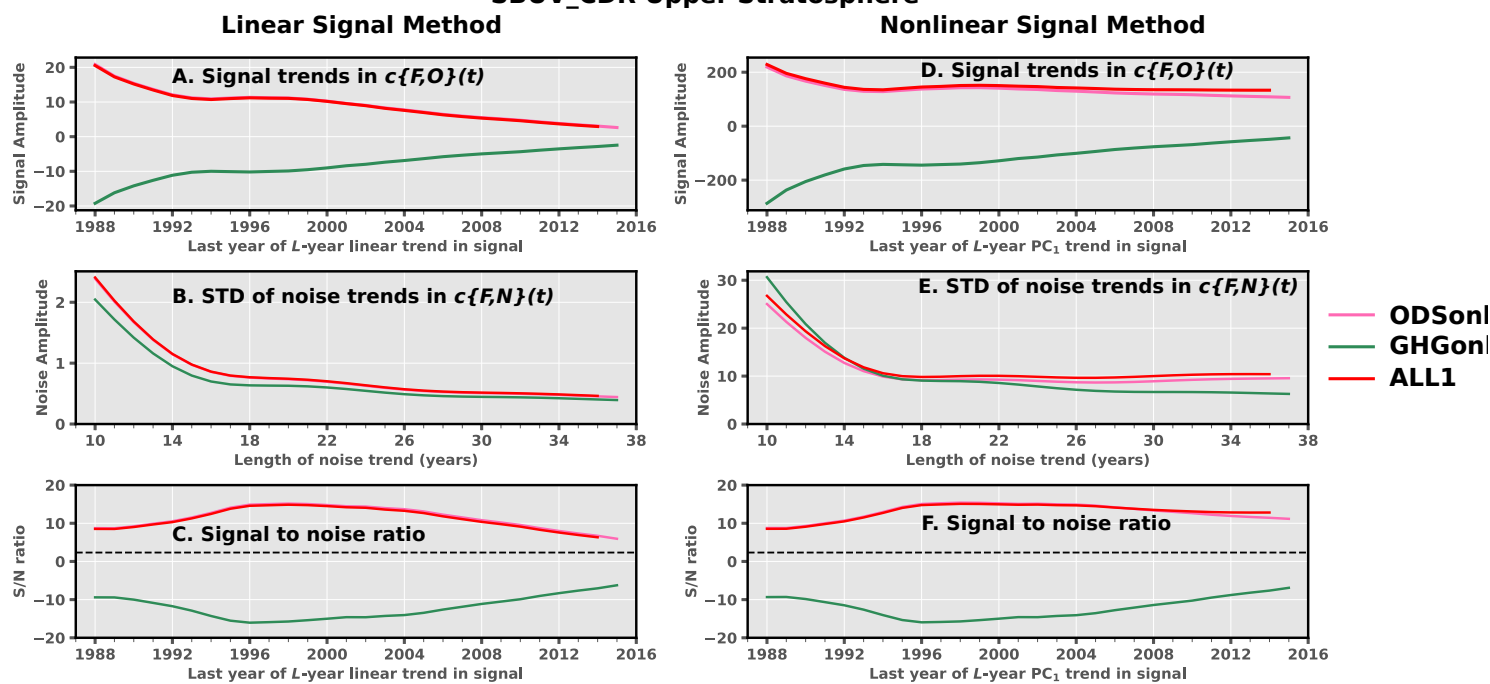


Figure S4S5: Same as Fig. 7 in the main text showing S/N estimates for the upper stratosphere, 1 to 10 hPa, but with SBUV_CDR ozone dataset instead of SWOOSH.