The authors would like to thank both referees and the editor for their time and contributions to this paper. The following pages first include the authors' responses to referee #1 followed by those to referee #2 as posted in the interactive discussion (https://www.atmos-chem-phys-discuss.net/acp-2017-575/#discussion). Thereafter we have included the revised manuscript with markup included.

The authors would like to thank the referee for taking the time to review this paper and for the many helpful comments that will be used to improve it. The referee's comments/concerns are listed below in red text, while the authors' responses to each comment are written below in black text.

The manuscript describes a new version of a regression model, but the model is nowhere described, neither briefly nor in detail. Much of it might be described in detail in Damadeo et al. (2014), however, it now is used for the first time to not just use SAGE II measurements, but simultaneously also HALOE and ACE-FTS measurements, and I think that should definitely be described mathematically.
As mentioned before, I was sometimes missing details about the methodology description and during the result discussion. I would recommend that the authors read through the manuscript again with this in mind (after considering the more specific comments below) to clarify any remaining items that are described too briefly.

An appendix has been added that summarizes the technique from Damadeo et al. (2014) and adds some additional detail regarding how multiple data sets are incorporated simultaneously. Hopefully this inclusion as well as some other edits throughout the paper make things clearer.

Section 2: I would recommend to remove the description of the data sets that are not used in the analysis that is described here in detail (POAM II, POAM III, SAGE III), as well as the description of their filtering. The results of the STS regression with all six data sources are only mentioned briefly once, and therefore the detailed description of those data sources seems unnecessary. It is always possible to refer to their description in the literature.

The same STS regression analysis was performed for all six data sets. The resulting trends are very similar and thus are not discussed here but some things like the seasonal cycle and diurnal variability of the ACE-FTS instrument were detrimentally affected slightly in their inclusion. This is why the paper discusses their exclusion. The resulting trends and the residual plots are also shown in the new supplement.

Page 4, line 11-12: Was the analysis done for more unit systems than just number density versus altitude? If yes, this should be mentioned in more detail. If no, I don't think the information about other unit systems is relevant here.

The same STS regression analysis was performed for the three main data sets in mixing ratio on pressure. The resulting trends are similar and thus are not discussed here but are shown in the new supplement.

Page 4, line 11-12: What vertical resolution is used for the different regression analyses? 1km?

The data was interpolated to 0.5 km increments. This has been added to the paper.

Page 4, line14: What is the spatial resolution of the daily means (for the STS)?

For occultation instrument sampling, the spatial resolution varies with latitude and ranges from ~3 degrees in latitude at the turnaround to up to ~10 degrees in the tropics.

Page 5, line 32 to page 6, line 16: Here is a detailed description of the correlated residuals that is not shown in a graph. It is very hard to follow the discussion without being able to look at something. I would recommend to either drop the paragraph, or add a figure that shows the correlated residuals.

This is a good point. Investigation of the total residuals do not have that much value compared to analyzing both the correlated and uncorrelated. However, in the interest of figure size, resolution, and space it is best to only show two types. The figure now shows the correlated and uncorrelated instead of the total and uncorrelated since these are what the paper discusses.

Page 7, line 33 to Page 8, line 2: The discussion about the SAGE II data filtering and the conclusion about the HALOE filtering is not clear. This section would benefit from some more details (how the conclusion was drawn) or some rephrasing.

This section has been rephrased and is hopefully clearer now.

Section 4.4: How do the results from Maycock et al. (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, 10021-10043, https://doi.org/10.5194/acp-16-10021-2016, 2016.) compare to the findings described here?

Like other recent analyses of the response of stratospheric ozone to the solar cycle using SAGE II data, Figure 4b of Maycock et al. (2016) appears to have similar results as this paper though the comparison requires a factor of 2 adjustment (peak-to-peak versus amplitude). This is now noted in Section 4.4.

Page 9, line 3-4: How much did the trends change in Millan et al. (2016) between considering the sampling bias and not considering it?

The thing about Millan et al. (2016) is that it didn't really consider the sampling bias. It chose a single "representative year" of sampling and repeated it 30 times (i.e., over 30 years) and then ran the sampling through a model. This was done so that multiple instruments that may or may not overlap in time could be evaluated on the same time scale. However, this is not the same as the actual sampling of those instruments as they change from year to year. As such, Millan et al. (2016) essentially only considers a hypothetical scenario that is not representative of how the different data sets behave. It is, however, informative in discussing the potential problems non-uniform sampling could create.

Page 9, line 7-15: It is not clear to me what that calculated bias is based on. It is given in percent, but is it percent of ozone? Or percent of something else? If it is ozone, how was the difference between the biased value and the "centered" MZM value (middle of the month and middle of the latitude band) calculated? More details would be helpful here.

The temporal and spatial offsets between the actual average of sampling and the "centered" values refer to differences in time and location. The biases are computed as ozone values, looking at the difference between the regression fits between these two times/locations, which is now clarified in the paper. In other words, after the data is regressed via the STS method and the coefficients are retrieved, a fit value to any time and place can be computed. These differences (i.e., biases) are the differences between these fit values at the two different locations and times (i.e., actual average location and time versus "centered" location and time).

These figures (6 and 7) are meant to be illustrative of the effect, but the actual correction used later is performed for each individual profile before any daily zonal means are created.

Page 10, line 17: The latitude band "20S-20N" should be "15S-15N" here? At least Figure 8 shows the results for "15S-15N".

The figure is correct and now the paper agrees.

Page 11, line 5-17: It would be good to be a bit more detailed in the description of the different methods here. It is not very obvious from the text that there are indeed 4 different regression models discussed.

The corrections that are applied are described in the paper and the regression that is applied to each of them is the same (only the input data changes with the corrections). This is now noted in the paper.

Page 11, line 18-23: It was not clear from the supplementary material how the text would change here with the updated way of calculating the uncertainties on the trends.

This paragraph has been rewritten to reflect the new methodology as detailed in the supplementary material and an appendix has been added to the paper to mathematically describe the process.

Page 11, line 33: The increase of about 1%/decade in the NH mid-latitudes is not very obvious in the updated plot. Is this a remnant description of the old plot? If not, I would recommend to adjust the description to ensure the reader knows where exactly the 1%/decade increase takes place.

It is now clarified that analyzing the impact of the diurnal correction implies comparing "MZM DCorr" and "MZM Raw" while analyzing the impact of the "seasonal" correction implies comparing "MZM DSCorr" and "MZM DCorr".

Page 13, line 6-8: It is referred here to the "recovery trend results in Fig 11", however it is not specified which results exactly. STS results? MZM? For all four results shown in Figure 11, I am not sure I see the pattern that is supposed to match the ACE-FTS drift pattern. Should this be Figure 12? If not, could you explain in more detail here where the similarities in the figures are?

The paper now mentions that it is the STS results we are comparing to. Unfortunately having some drift between the different data sets does not create a direct correlation with equal patterning into the resulting trends. The different drift patterns for the different instruments over the different time periods will alias into the different proxies in different ways. As such, the changes between Figures 11 and 14 may not be readily apparent simply from looking at Figure 13. However, Figure 14 does illustrate the aggregate effect of ignoring the possibility of drifts (but not offsets in the mean) between the different data sets.

Page 13, line 14: Where exactly are the results changed by up to 2%/decade?

The differences occur at various locations, most notably where the "recovery" trends were strongest in Figure 11.

Page 13, line 17: "limitations in these regression techniques" -> which regression techniques are referred to here? The ones used in this analysis? All four of them have the same limitations?

Page 13, line 25-26: “only a single uniform seasonal cycle should be used for these analyses” -> which analyses exactly are referred to here? Any regression model? Only the ones used here?
Page 14, line 1: “This study also highlights the limitations inherent in these techniques: : :” -> which techniques are referred to here?

It seems the word “these” implied specificity, whereas we really meant regression techniques in general, not just this one. The instances of the word “these” has been removed and the sentences corrected.

Page 14, line 8-9: “With sufficient overlap: : :” -> does MLS provide a sufficient overlap with SAGE II and HALOE to allow the suggested analysis?

This is actually both a good and very difficult question. There were/are many instruments with measurements after ~2000 that can be used to try to determine potential recovery trends. For most of these instruments, an instrument like MLS does have sufficient overlap to try to do this kind of analysis. Unfortunately, only a few of these instruments also provided data prior to 2002. The current problem is that the representation of the solar cycle can have a significant impact on derived trends and truthfully more than one solar cycle needs to be sampled by the data used in the regression to adequately constrain it. This is achievable with current measurements, but only by including data prior to 2002. Otherwise, we’ll need to wait for more measurements. This is where the difficulty of the question comes in: Is there enough overlap between any high-sampling instrument in the modern record and SAGE II or HALOE to link the time periods? While the assumption in previous works (both in regression analyses and in the creation of merged data sets such as GOZCARDS and SWOOSH) is “yes,” a definitive answer has, to our knowledge, never been investigated and is beyond the scope of this work.

Page 23: Maybe add “filtering” in the last line of the figure caption, “: : :, though results with filtering are similar”

This has been added to the paper.

The authors would like to thank the referee for taking the time to review this paper and for the many helpful comments that will be used to improve it. The referee's comments/concerns are listed below in red text, while the authors' responses to each comment are written below in black text.

It is stated that the method is described in the previous paper by Damadeo et al., 2014. It is not really the case since this paper deals with multiple instrument. The authors should add an Appendix or Supplement with the method description. In particular: Damadeo et al., 2014 analyzed data from just SAGE II. How exactly was the analysis of six satellite instruments (or just three? SAGEII, HALOE and ACE- FTS) done?

An appendix has been added that summarizes the technique from Damadeo et al. (2014) and adds some additional detail regarding how multiple data sets are incorporated simultaneously.

Was any instrument-specific weighting applied?

No. All instruments are treated equally. However, the regression is weighted and the instrument uncertainties do factor into that weighting so if one instrument is inherently less precise than another that will have an impact on the weighting.

The authors stated in many places that they try to use orthogonal functions. But the functions should be orthogonal on the dataset of available observations. Damadeo et al., 2014 mentioned that seven Legendre polynomials of latitude were used for the fit. However Legendre polynomials are orthogonal on 90S-90N, not on the 60S-60N interval where almost all measurements were taken. How was this handled?

Legendre polynomials in spherical harmonics are the logical choice to fit slowly varying data in latitude (as is done in many areas in physics) even if the data does not extend to the poles. Any polar gaps in a data set simply result in some degree of overfitting in those gaps, but ultimately it does not matter since we are not looking at those regions. Additionally, the data sets used here extend beyond 60S and 60N; we just show the results in this region because that is the primary region of focus in the community related to ozone trend studies.

Also, it seems that seven polynomials are too many. The authors should provide some justification.

Actually, this analysis uses 9 spatial terms instead of 7 (now added in the paper text) and it is very likely that more terms should be used. With a span of 180° in latitude, 9 terms yields a spatial resolution of 20°. However, some effects such as volcanic responses and the spatial extent of the peak of the QBO, for example, can be smaller than this. Choosing too many terms will create excessive overfitting through data gaps so 9 terms was chosen as a "middle ground" though no sensitivity study on this parameter was performed.

The authors compare their STS regression with the MZM method. I suggest the author reduce the part related to MZM and focus only on their STS results.

It is important to make the comparisons between the STS and the MZM as the MZM is currently the "de facto" methodology used by the community. Additionally, since we understand that it is very unlikely that the STS method will be whole-heartedly adopted, we decided that

using the STS method to create "corrected" versions for use with the MZM method would be a reasonable compromise. As such, it is necessary to detail these "corrected" versions and how they compare as well as showing the relative impact of different sampling biases.

The MZM method is used in the paper in a very peculiar way: the authors just average all data within 10-deg latitudinal belts and assigned the value to the middle of the belt at the middle of the month.

We do not understand the reviewer's comment about the MZM being implemented in a "peculiar way" as this concept of creating monthly zonal means (i.e., averaging all of the data in a single month and latitude bin) is the default methodology that has been applied for almost all regression analyses of stratospheric ozone.

Most of the ozone variability is coming from the annual cycle. The annual cycle can be estimated, for example, by the same approach as discussed in the paper: by fitting all SAGE II data by a set of spherical (for latitude and, if necessary, longitude) and sin/cos functions (for time). Then the MZM method could be applied to the deviations from the annual cycle. The annual cycle is indeed orthogonal to the other proxies, so it should not affect their estimates.

The following description of fitting the seasonal cycle sounds like the common practice of deseasonalizing the data first. One could deseasonalize first, but doing so does not include information of any collinearity in the covariance matrix during the regression process and thus this information is not represented in the resulting coefficient uncertainties.

This step would largely remove most of the sampling problems and will likely produce results similar to STS.

Deseasonalizing cannot remove the sampling problems. If the sampling problems are constant from year to year, then not deseasonalizing only results in a biased seasonal cycle but does not impact the other terms (e.g., trends). However, in the case that the sampling patterns change from year to year (as has clearly been demonstrated), neither deseasonalizing nor fitting the seasonal cycle removes the problem. This is because the sampling biases create biases in the MZM values themselves and so any correction must be implemented on an event by event basis (e.g., the "DCorr" and "DSCorr" MZM data sets). Either that, or the data needs to be handled at a resolution much closer to the native resolution (e.g., the STS method).

P.4, L. 12. What data were used for this conversion? See box 2-1 from Ozone Assessment 2014 and comment on potential conversion errors.

As stated in the paper, the data used for the conversions were the pressures, temperatures, and altitudes given in the respective data sets. Box 2-1 from the 2014 Ozone Assessment cites McLinden and Fioletov (2011) and discusses both the expected differences in trends between number density and mixing ratio but also the potential uncertainties introduced during the conversion process.

First it is important to note that there is a distinction between these two phenomena. As cited, there is an expected difference in trend values depending upon choice of unit system due to underlying trends in stratospheric temperature. This just means it is important when comparing multiple analyses done with different unit representations to not expect the values to be the same. While this difference does not imply a unit conversion error, potential unit conversion errors can have an impact on the resulting data quality used for trend analyses. McLinden and Fioletov

(2011) show the potential impact using the SAGE II data. However, it is important to note that work made use of version 6.20 of the SAGE II data while more current works use version 7.00 with the important difference of the source of (and consistency of) the meteorological data in the middle to upper stratosphere. Damadeo et al. (2013) showed that the version 7.00 ozone product was much more robust for trend analyses than version 6.20 as a result of the change in meteorological data used for the retrieval. Additionally, Hubert et al. (2016) showed that the SAGE II v7.0, HALOE v19, and ACE-FTS v3.0 data products were relatively consistent in all unit representations when compared to the ground network.

All of that having been said, the possibility for unit conversions to introduce additional uncertainties is still present and an extensive study of this impact on resulting trends was not part of this work.

P.4, L. 22. ENSO is mentioned here, but no result was shown. Is it necessary to include it?

Not really. Damadeo et al. (2014) went into much greater detail on the results of the technique, to show that the results of all of the proxies were reasonable. However, even in that work, the impact of ENSO above ~20 km is fairly negligible and thus wasn't worth going into detail either. The same is true here. Rather, in this work, we only go into detail on the proxies that are impacted by the use of multiple data sets (e.g., solar and aerosol) or the sampling (e.g., diurnal) and, of course, the trends.

P.4, L. 22. The shape of the EESC function depends of latitude and altitude. What exactly was used? The authors used 2 "orthogonal" EESC functions and show the trend results. But how does the resulting EESC signal look like? What is the "phase"/delay? If the authors want to have an additional delay for EESC, it is more logical to introduce an unknown time lag.

The EESC proxy used here is detailed in Damadeo et al. (2014). To summarize, it derives from an empirical orthogonal function (EOF) analysis of EESC data with 6 different mean ages-of-air that subsequently have different turnaround times. The leading 2 EOFs account for 99% of the variance and can recreate the original 6 functions to within ~1%. In other words, using these two proxies allows for the regression to independently adjust the turnaround time and even allows for monotonic trend results. As such, there is no manually created phase delay.

P. 8. Solar cycle. It is difficult to get the 11-year solar cycle from SAGE data. Is the estimated solar signal statistically significant? Are the differences in the solar signal at different latitudes significant?

The solar cycle is, perhaps, one of the more difficult responses to derive from regression analyses owing to the fact that its duration of 11 years is often longer than most individual data sets. SAGE data, with a 21 year record, has quite often been used for this purpose. Unfortunately, within the SAGE II mission period (1984 to 2005), the volcanic eruptions of El Chichon and Mt. Pinatubo have coincidentally coincided with solar activity. This has greatly increased the collinearity of the two effects in regression analyses and made it much more difficult to separate the two. This was discussed in Damadeo et al. (2014). As discussed in this current work, using multiple data sets that span an even longer duration including the relative volcanically quiescent time period from 1998 to 2011 seems to have sufficient length and orthogonality to better extract the solar cycle response. Additionally, we have updated Figure 5 to include statistical significance.

The Impact of Non-uniform Sampling on Stratospheric Ozone Trends Derived from Occultation Instruments

Robert P. Damadeo[1], Joseph M. Zawodny[1], Ellis E. Remsberg[1], and Kaley A. Walker[2]

[1]NASA Langley Research Center, Hampton, VA, USA

[2]University of Toronto, Department of Physics, Toronto, Ontario, Canada

Correspondence to: R. P. Damadeo (robert.damadeo@nasa.gov)

Abstract. This paper applies a recently developed technique for deriving long-term trends in ozone from sparsely sampled data sets to multiple occultation instruments simultaneously without the need for homogenization. The technique can compensate for the non-uniform temporal, spatial, and diurnal sampling of the different instruments and can also be used to account for biases and drifts between instruments. These problems have been noted in recent international assessments as being a primary source of uncertainty that clouds the significance of derived trends. Results show potential recovery trends of $\sim$2–3%/decade in the upper stratosphere at mid-latitudes, which are similar to other studies, and also how sampling biases present in these data sets can create differences in derived "recovery" trends of up to $\sim$1%/decade if not properly accounted for. Limitations inherent to all techniques (e.g., relative instrument drifts) and their impacts (e.g., trend differences up to $\sim$2%/decade) are also described and a potential path forward towards resolution is presented.

1 Introduction

Ever since the Montreal Protocol came into effect, the global scientific community has been monitoring the state of stratospheric ozone in an effort to determine at first if the loss rate was decreasing and later if ozone had begun to recover. Consequently, there has been an ongoing body of work to use single (at first) or multiple (later) sources of data, spanning the satellite record starting around 1979, for various multiple linear regression (MLR) analyses to determine the long-term trends in stratospheric ozone. A simple literature search would reveal the various techniques and studies ranging from the earlier ~~(e.g., Wang et al., 1996; Bodeker et al., 1998; Newchurch et al., 2003) works~~ works (e.g., Wang et al., 1996; Bodeker et al., 1998; Newchu revealing the loss slowdown to a recent surge in efforts to determine potential ozone recovery (e.g., Randel and Wu, 2007; Remsberg and Lingenfelser, 2010; Bodeker et al., 2013; Kyrölä et al., 2013; Bourassa et al., 2014; Gebhardt et al., 2014; Tummon et al., 2015; Harris et al., 2015; Steinbrecht et al., 2017). These works have culminated in the most recent Scientific Assessment of Ozone Depletion (WMO, 2014) that showed statistically significant recovery trends of $\sim$2%/decade in the upper stratosphere at mid-latitudes but identified three factors with a potential major impact were not readily accounted for in those analyses: diurnal variability of ozone, biases between data sets, and long-term drifts between data sets. There is an additional complication that is intricately tied to these three factors in this kind of analysis, namely the non-uniform temporal, spatial, and diurnal sampling of the different instruments used for these analyses. This non-uniform sampling can have a detrimental

impact not only on the regression techniques used to derive long-term trends in ozone but also on other analyses performed to determine diurnal variability or the magnitude of potential biases and drifts between data sets. Herein, we discuss a recently developed technique that not only accounts for the potential sampling issues, but also the perceived diurnal variability, as well as any potential bias and/or drift between instruments in a single analysis.

2 Data Sets

There have been several remote sensing instruments over the past several decades that have observed stratospheric ozone using the method of solar occultation, including but not limited to: the Atmospheric Chemistry Experiment Fourier Transform Spectrometer (ACE-FTS), the Halogen Occultation Experiment (HALOE), the Polar Ozone and Aerosol Measurement (POAM)~~II and~~ II and III, and the Stratospheric Aerosol and Gas Experiment (SAGE)~~I,~~ I, II, and III. For the purpose of this study, however, SAGE I was ignored because it does not have any overlap with the other missions.

2.1 ACE-FTS

The Atmospheric Chemistry Experiment Fourier Transform Spectrometer (ACE-FTS) was launched onboard the SCISAT–1 spacecraft in August 2003 (Bernath, 2017). The spacecraft occupies a 74°inclined orbit at an altitude of $\sim 650\,\mathrm{km}$ that allows for observations from 85°S to 85°N. The primary ACE instrument is a high spectral resolution ($0.02\,\mathrm{cm}^{-1}$) Fourier Transform Spectrometer (FTS) operating in the spectral range of ~ 2.2–$13.3\,\mu\mathrm{m}$ (750–$4400\,\mathrm{cm}^{-1}$) that measures many trace gas species and isotopologues (Bernath et al., 2005). Ozone is retrieved using the spectral features near $10\,\mu\mathrm{m}$ (Boone et al., 2005). The version of the ACE-FTS data product used here is version 3.5 (Boone et al., 2013), which produces vertical profiles of volume mixing ratio (VMR) interpolated to a $1\,\mathrm{km}$ grid with a vertical resolution of 3–4 km. The ACE-FTS instrument is still operating.

2.2 HALOE

The Halogen Occultation Experiment (HALOE) was launched onboard the Upper Atmosphere Research Satellite (UARS) in September 1991. The spacecraft occupied a 57°inclined orbit at an altitude of $\sim 585\,\mathrm{km}$ that allowed for observations from 80°S to 80°N. The HALOE instrument used a combination of broadband radiometry and gas filter correlation techniques to observe several trace gas species in the spectral range of ~ 2.4–$10.4\,\mu\mathrm{m}$ ($\sim$ 950–$4150\,\mathrm{cm}^{-1}$) and measured ozone using the spectral band near $9.6\,\mu\mathrm{m}$ (Russell et al., 1993). The version of the HALOE data product used here is version 19.0 (Thompson and Gordley, 2009), which produces vertical profiles of VMR interpolated to a $0.3\,\mathrm{km}$ grid with a vertical resolution of 2–3 km (Bhatt et al., 1999). The UARS mission was decommissioned in December 2005.

2.3 POAM II

The Polar Ozone and Aerosol Measurement II (POAM II) was launched onboard the SPOT–3 spacecraft in September 1993. The spacecraft occupied a sun-synchronous orbit, crossing the descending node at 10:30 LT, that allowed for observations in two latitude bands at 88°S to 62°S and 65°N to 71°N. The POAM II instrument used broadband radiometry to observe aerosol

and trace gases in the spectral range of $\sim$350–1070 nm and measured ozone using the spectral band near 600 nm (Glaccum et al., 1996). The version of the POAM II data product used here is version 6.0, which produces vertical profiles of number density interpolated to a 1 km grid with a vertical resolution of 1 km (Lumpe et al., 1997). The SPOT–3 spacecraft ceased functioning in November 1997.

2.4 POAM III

The Polar Ozone and Aerosol Measurement III (POAM III) was launched onboard the SPOT–4 spacecraft in March 1998. The spacecraft occupied a sun-synchronous orbit, crossing the descending node at 10:30 LT, that allowed for observations in two latitude bands at 88°S to 62°S and 65°N to 71°N. The POAM III instrument used broadband radiometry to observe aerosol and trace gases in the spectral range of $\sim$345–1030 nm and measured ozone using the spectral band near 600 nm (Lucke et al., 1999). The version of the POAM III data product used here is version 4.0 (Lumpe et al., 2002; Naval Research Laboratory, 2006), which produces vertical profiles of number density interpolated to a 1 km grid with a vertical resolution of 1 km. The POAM III instrument ceased functioning in December 2005.

2.5 SAGE II

The Stratospheric Aerosol and Gas Experiment II (SAGE II) was launched onboard the Earth Radiation Budget Satellite (ERBS) in October 1984. The spacecraft occupied a 57°inclined orbit at an altitude of $\sim$610 km that allowed for observations from 80°S to 80°N. The SAGE II instrument was a broadband spectrometer that operated in the spectral range of $\sim$375–1030 nm for aerosol and trace gas observations and measured ozone using the spectral band near 600 nm (Mauldin III et al., 1985). The version of the SAGE II data product used here is version 7.00 (Damadeo et al., 2013), which produces vertical profiles of number density interpolated to a 0.5 km grid with a vertical resolution of 1 km. The ERBS mission was decommissioned in October 2005.

2.6 SAGE III

The Stratospheric Aerosol and Gas Experiment III (SAGE III) was launched onboard the Russian Meteor–3M (M3M) spacecraft in December 2001. The spacecraft occupied a sun-synchronous orbit, crossing the ascending node at 09:00 LT, that allowed for observations in two latitude bands at 60°S to 30°S and 45°N to 80°N. The SAGE III instrument was a grating spectrometer that operated in the spectral range of $\sim$295–1025 nm for aerosol and trace gas observations and measured ozone using the spectral features near 600 nm (Mauldin et al., 1998). The version of the SAGE III data product used here is version 4.00 (Cunnold and McCormick, 2002; Wofsy et al., 2002), which produces vertical profiles of number density interpolated to a 0.5 km grid with a vertical resolution of 1 km. The M3M spacecraft ceased functioning in January 2006.

2.7 Filtering

When making use of any data set, it is important to apply the proper filtering to ensure that bad data (e.g., fill values or data contaminated by clouds) are excluded. Since this analysis is constrained to the stratosphere, all data below the tropopause are ignored. If a data set provides a tropopause height, that value is used for filtering purposes, otherwise the World Meteorological Organization (WMO) definition is used (WMO, 1992). Beyond this, the data screening procedures recommended for each data set are performed. ACE-FTS data are screened as outlined in Sheese et al. (2015). HALOE data are screened for potential problematic "constant lockdown angle" and "trip angle" events as detailed by the data producers (http://haloe.gats-inc.com/user_docs/index.php). POAM II data could be screened for interference from polar stratospheric clouds (PSCs) by looking for outliers in the 1 µm data, though this is not performed. POAM III data are screened for potential sunspot interference and heavy aerosol interference through the use of the quality flags. SAGE II data are screened for this analysis in the same way as was done in Damadeo et al. (2014). Since SAGE III data were screened prior to release, no additional screening is performed.

3 Analysis Technique

In principle, this work is a continuation of the work first performed in Damadeo et al. (2014) and so the same techniques and methodologies are used. Each data set is filtered according to the stated filtering techniques and converted to the unit system of interest (i.e., number density or mixing ratio versus altitude or pressure) using the pressures, temperatures, and altitudes provided with the respective data sets~~, though~~ . While we did apply the analysis to both combinations of unit systems, for the sake of brevity all results shown here are for regressions to data in number density on altitude ~~.~~ (some mixing ratio on pressure results are shown in the Supplement). Additionally, the data for each instrument were interpolated to 0.5 km increments. These data are then consolidated into daily zonal means for each instrument separated by both satellite and local event types. A generalized least-squares regression technique that accounts for autocorrelation, heteroscedasticity, and data gaps is then performed on all data sets simultaneously, with the autocorrelation and heteroscedasticity corrections being applied separately for each instrument. In principle, this technique is applicable to data sets with higher sampling (e.g., the Microwave Limb Sounder (MLS) on the Aura satellite) but is demonstrated here on occultation data sets only to illustrate the impact of their sparse sampling patterns on derived trends.

The same simultaneous temporal and spatial (STS) MLR model as was used in Damadeo et al. (2014) is applied, albeit with~~some~~ 9 spatial terms instead of 7 and some additional changes to account for the incorporation of multiple instruments (see Appendix A). Terms accounting for the quasi-biennial oscillation (QBO), El Niño-Southern Oscillation (ENSO), solar variability, and long-term trends (two orthogonal Equivalent Effective Stratospheric Chlorine (EESC) functions) are applied ~~globally (i.e.,~~ to all data sets ~~help constrain the coefficients)~~simultaneously. Terms accounting for volcanic eruptions (primarily the Mount Pinatubo eruption in 1991) are applied to the SAGE II and HALOE data sets only (and separately) to avoid potential overfitting of minor eruptions in data sets that do not cover the Mt. Pinatubo eruption. Diurnal variability (applied as a binary conditional term) is fit separately for each data set. While the sun-synchronous instruments (i.e., SAGE III, POAM II, and POAM III) sample both satellite event types, they do not adequately sample both local event types (Fig. 1) and so all local sun-

rises from these instruments are ignored for this analysis. The seasonal cycle is applied ~~globally~~ to all data sets simultaneously as a single seasonal cycle for all instruments. Lastly, a bias offset term and a linear drift term are applied separately for each instrument using SAGE II as the reference instrument.

Since the STS regression model uses a two-dimensional regression, it is best utilized on data that adequately covers the full range of temporal and spatial sampling to constrain the temporal and spatial variability present in the data. Occultation instruments in mid-inclination orbits tend to deliver near global coverage at somewhat reduced seasonal sampling while occultation instruments in sun-synchronous orbits tend to deliver highly localized spatial coverage at nearly full seasonal sampling. The primary focus of this work is the impact of sampling biases on long-term trends in ozone, which is typically analyzed in the stratosphere between about 60°S and 60°N. Since this work focuses on that latitude range and the sun-synchronous instruments exhibit little to no coverage within that region and thus very little influence on resulting trends there, the results presented herein derive from an STS regression using only the SAGE II, HALOE, and ACE-FTS data sets. We also applied an STS regression using all six data sets and found that the long-term trends were not significantly affected (see the Supplement) but did notice that the lack of spatial coverage in the POAM II, POAM III, and SAGE III data sets detrimentally impacted the results in the seasonal cycle and diurnal variability derived from a two-dimensional regression. In the interest of brevity and to maintain the legibility of certain figures in this paper, individualized results from the six-instrument regression are not shown here.

4 Non-trend Results

4.1 Residuals

Similarly to Damadeo et al. (2014), we investigate the residuals of the regression. The residuals from the regression can be used to ascertain the quality of the model and the data set itself, independent of any offset in the mean value. While the mean of the residuals is zero (as it should be), a clear pattern in the spread of the residuals emerges as a function of latitude at each altitude. The total residuals of the regression (i.e., the residuals from the ordinary least-squares regression) is a combination of the correlated residuals (i.e., those removed during the autocorrelation correction) and the uncorrelated residuals (i.e., the residuals from the generalized least-squares regression). The correlated residuals represent geophysical variability that is well-sampled but not well-modeled by the regression as well as any systematic instrumental variability (e.g., biased meteorological or ephemeris input data). The uncorrelated residuals represent both measurement noise and geophysical variability that is not well-sampled (e.g., geophysical variability present within each daily mean).

Figure 2 shows the spread of the ~~total~~ correlated and uncorrelated residuals for each instrument. All of the instruments exhibit increased residuals in the lower stratosphere, owing both to the increased uncertainty of measurements in that region as well as increased variability that is not adequately captured by the proxies used for this regression. Similarly, residuals are higher at higher latitudes where measurements can routinely dip into and out of the vortex both over multiple days and within a single day itself. SAGE II has greatly increased uncorrelated residuals at the highest altitudes compared to HALOE and ACE-FTS. While the influence of measurement noise and daily zonal variability in the uncorrelated residuals cannot be separated, the

fact that SAGE II and HALOE (and to a lesser extent ACE-FTS) exhibit similar sampling patterns means that the increased uncorrelated residuals in the upper stratosphere and lower mesosphere in SAGE II compared to HALOE must be a result of increased measurement noise in SAGE II. Similarly, SAGE II and ACE-FTS display slightly lower uncorrelated residuals in the lower stratosphere while HALOE and ACE-FTS display lower uncorrelated residuals in the upper stratosphere. All three instruments show comparable uncorrelated residuals in the middle stratosphere.

The correlated residuals ~~(difference between the total and uncorrelated)~~ show an increased spread in the stratosphere at high latitudes, which is expected as variability within the polar vortex is not modeled in this regression. Similarly, increases can be seen in the tropical middle stratosphere near a local peak in QBO amplitude. This is a result of a two-dimensional fit using a proxy derived only at the Equator. While modulating the QBO with the seasonal cycle better represents the QBO at higher latitudes, the inability to accurately model the QBO at higher latitudes detracts from the ability to accurately model the QBO at lower latitudes (Damadeo et al., 2014). Another interesting feature is an apparent vertical "banding" structure in the correlated residuals present in each data set. The locations of this "banding" correlate to the "turnover" latitudes in each instrument's orbit (i.e., the latitudes at which measurements go from progressively closer to the poles to progressively further away). The autocorrelation correction accounts for the degree of correlation of data from day to day. However, the locations of daily means change in latitude from day to day, with rates of motion greater at the Equator and smaller near the poles and so the degree of correlation is dependent upon both the temporal variability and the meridional variability, with the meridional variability being the primary driver. At the orbit "turnover" point, the meridional variability between each successive daily mean essentially disappears. While not explored in this study, it is possible that this additional source of correlated noise stems from the nature of how wave one action is sampled from day to day over the course of about a week until the instrument moves away from the turnover latitude. Because the measurements systematically shift in longitude over the day while the wave itself also rotates, the zonal variability is not evenly sampled and so these day to day differences will be highly correlated as the wave one action rotates and changes, thus revealing a potential additional source of sampling bias albeit more localized and on a shorter time scale.

4.2 Diurnal Variability

Occultation instruments sample one sunrise and one sunset per orbit as seen by the spacecraft, which typically correlates to one sunrise (SR) and one sunset (SS) as seen by an observer on the ground at the measurement location. This means that occultation measurements of ozone sample its diurnal variability present in the mesosphere and upper stratosphere. Diurnal variability of ozone in the mesosphere has been investigated before and is well understood to be a result of rapid photochemistry across the terminator (Chapman, 1930; Herman, 1979; Pallister and Tuck, 1983). While the full attribution of sources is still not completely understood, diurnal variability in the stratosphere is well represented in various data sets. Analysis of the diurnal variability from occultation instruments is typically performed by looking for periods where the instrument's diurnal sampling "crosses itself" (i.e., local sunrise and sunset measurements occur at roughly the same latitude at roughly the same time). Sakazaki et al. (2015) used this method to analyze the diurnal variability present in SAGE II, HALOE, and ACE-FTS and found that not all data sets agree and the differences between SR and SS values differ typically by up to $\sim$5%. The STS

regression can extract the mean diurnal variability present in each data set and the results shown in Fig. 3 compare quite well with those in Figure 5 of Sakazaki et al. (2015).

4.3 Impacts of Aerosol

Volcanic eruptions periodically inject sulfur dioxide into the stratosphere where it goes on to form sulfate aerosols that can impact ozone either via chemical effects (Rodriguez et al., 1991; Solomon, 1999) or through changes in dynamics via changes in radiative forcing (McCormick et al., 1995; Robock, 2000). In either case, it is possible for volcanic aerosols to have a significant impact on stratospheric ozone levels such that their presence can complicate these regression analyses. Since ozone trend analyses utilize data from the past $\sim 30\,\mathrm{years}$, usually only the Mt. Pinatubo eruption in mid 1991 is considered for special treatment. If the analysis goes back further, sometimes the El Chichon eruption in early 1982 is also considered. The punctuated nature of the eruptions and not completely characterized impacts on data quality often leads to many works simply excluding data from one to several years after these eruptions (e.g., Wang et al., 1996; Randel and Wu, 2007; Harris et al., 2015) while some works attempt to include a term in the regression to model the impact (e.g., Bodeker et al., 2001; Stolarski et al., 2006; Bodeker et al., 2013; Tummon et al., 2015), although the nature of these terms tends to be different between different analyses.

For this work we include an aerosol proxy that was derived in Damadeo et al. (2014). The proxy is a "volcanic" one, meaning that eruptions occur and the proxy rises, peaks, and subsequently decays back to zero. The proxy only covers the SAGE II mission time period and thus is zero throughout most of the ACE-FTS mission period. However, given that it takes a relatively large eruption (e.g., Mt. Pinatubo) to register any noticeable changes in stratospheric ozone in these regression analyses and the fact that only minor eruptions have occurred since (Vernier et al., 2011), this is assumed to be sufficient. Given that occultation instruments can (depending upon their spectral channels) have reduced measurement sensitivity in the presence of heavy aerosol loading (Wang et al., 2002; Bhatt et al., 1999), the volcanic proxy is applied separately for SAGE II and HALOE. The regression was applied under two conditions with regard to aerosol: one in which no filtering of events for the influence of aerosols was performed and another in which SAGE II was filtered under the recommendations in Wang et al. (2002) and HALOE was filtered under the recommendations in Bhatt et al. (1999).

Figure 4 shows the peak of the volcanic regression term surrounding the Mt. Pinatubo eruption for both the aerosol filtered and unfiltered cases. In the unfiltered case, both SAGE II and HALOE show similar responses of ozone to the eruption in the tropics between ~ 24 and $35\,\mathrm{km}$. Both instruments show a large region of negative correlation between ozone and aerosol in the lower stratosphere surrounding the aerosol layer itself and another large region of positive correlation in the middle stratosphere above the aerosol layer (the anomalously large responses in the lowermost stratosphere are a result of overfitting due to missing data). These results are in reasonably good agreement with Aquila et al. (2013) and Bodeker et al. (2013), which show results of the impact of the eruption on ozone levels from modeling and data respectively, and in surprisingly good agreement between the two separate instruments. The effect of the eruption on ozone derived from HALOE data is typically more difficult to quantify since HALOE did not begin to take measurements until shortly after the eruption, which has a tendency to negatively impact studies of long-term variation using only the HALOE data set (Remsberg, 2008).

When ~~aerosol filtering is applied , however, HALOE responses~~ the regression was run with the stated aerosol filtering criteria applied to the data, the results of the volcanic regression term from HALOE remain unchanged (not shown)~~while~~ . However, compared to the unfiltered case (middle plot in Fig. 4) the SAGE II responses with aerosol filtering applied (left plot in Fig. 4) remain unchanged above 28 km but significantly reduced in amplitude in the tropics below that. ~~In~~ The aerosol filtering has no effect in the middle stratosphere in the region of positive correlation ~~,~~ because the aerosol loading levels were not so high as to detrimentally affect the retrievals of ~~the instruments , which is why the aerosol filtering has no effect there. However, between there~~ these occultation instruments there. Between the middle and the lowermost stratosphere the data quality declines until measurements were no longer possible. The Wang et al. (2002) filtering criteria were meant to exclude anomalous ozone values based on aerosol extinction/ratio values ~~though it may be, given these results , that these~~ in the regions where data quality declines. However, the apparent agreement of the unfiltered results suggest that the Wang et al. (2002) filtering criteria are overly conservative and need to be revisited. Either that, or the SAGE II filtering criteria and results are reasonable and perhaps the HALOE data require a better aerosol correction in the retrieval algorithm than what is already applied (Hervig et al., 1995).

4.4 Solar Cycle Response

The impact of the $\sim$ 11-year solar cycle on stratospheric ozone has been an ongoing topic of study ~~(e.g., Wang et al., 2002; Soukharev and H~~ (e.g., Wang et al., 2002; Soukharev and Hood, 2006; Randel and Wu, 2007; Remsberg, 2008, 2014; Maycock et al., 2016; Dhomse et al., 2 . As such, it is worthwhile to show the results of the solar response in this work as well as to point out a few things about data usage and the determination of the solar cycle response to ozone when using MLR-based studies on SAGE II and HALOE data. The cited works show different solar cycles when using SAGE II data as well as different solar cycles between using SAGE II and HALOE data, with the latter exhibiting the greatest difficulty in determining (Soukharev and Hood, 2006). Figure 5 shows the latitude and altitude dependent amplitude of the solar cycle response derived from this work, which is similar to other recent works based on the usage of the SAGE II data set (e.g., Maycock et al., 2016; Dhomse et al., 2016) and naturally very similar to those from Damadeo et al. (2014). One important distinction between the previous work and this one is the impact of the use of one or two solar terms. Previously, when applied to only SAGE II data, using two solar terms shifted the solar cycle response by about 2 years in the presence of the Mt. Pinatubo eruption in agreement with Remsberg (2014), though this was believed to be the regression algorithm simply trying to attribute some of the aerosol response to the solar cycle (Solomon et al., 1996). The inclusion of HALOE and ACE-FTS in this study, however, seems to better constrain the solar cycle such that using one or two solar cycle terms no longer creates temporal shifting in the presence of the eruption (not shown) and thus only a single solar cycle term is required for the regression. While also not shown here, we attempted to apply the STS regression to only HALOE data and found that no combination of proxies exhibited realistic looking solar cycle responses, most likely due to the data having insufficient duration capable of constraining the solar cycle, aerosol, and trend terms simultaneously. This could potentially explain the often different solar cycle responses derived when using the instruments separately, while using them simultaneously creates SAGE II-like responses as well as very similar aerosol responses, though this requires further

study. Lastly, it is worth noting that the large amplitude tropical response below $\sim 23\,\mathrm{km}$ is a result of the previously discussed anomalous aerosol response in that area.

5 Sampling Biases

5.1 Seasonal Sampling

Traditionally, data sets are reduced to monthly zonal mean (MZM) values for regression analyses to determine long-term trends. Practically speaking, these MZM values are utilized as though they are representative of the center of the month and the center of the latitude bin. Though this assumption holds mostly true for highly sampled data sets (e.g., nadir and limb sounders), it generally fails when applied to occultation data sets. This fact is well known and has been studied before. Toohey et al. (2013) and Sofieva et al. (2014) both investigated non-uniform temporal sampling as an added source of noise and uncertainty that could be characterized and included in trend analyses. Using deseasonalized anomalies for trend analysis can mitigate the impacts of sampling bias if the bias is constant with time. However, owing to the observational geometry of occultation instruments and orbital parameters (i.e., altitude, inclination, and precession rates) the sampling patterns often tend to systematically drift over time as shown in the top row of Fig. 6. Millán et al. (2016) investigated the impacts of non-uniform sampling biases on resulting trends from different instruments by using a "representative year" of sampling for each data set and repeating it over 30 years to analyze the effect on trends. While illustrative, this did not account for the actual sampling bias as it changed from year to year.

The systematic drift in sampling combined with the presence of sampling biases precludes the use of the MZM method to accurately determine the seasonal cycle that is represented by an occultation data set. The STS regression, however, is less sensitive to sampling biases and can thus be used to quantitatively assess the sampling biases that would be present in the MZM method. It is relatively straight forward to compute the temporal and spatial offset between the average time and location of sampling within a given month and latitude band and the center of that month and band that is considered the representative location for the MZM method. The spatially varying seasonal ozone cycle from the STS regression can then be used to compute the difference in fitted ozone values between the actual center of sampling and the representative center of sampling to compute a seasonal sampling bias for each month, latitude band, and altitude bin. Some typical results of these biases are shown in the bottom row of Fig. 6. It is evident from a simple visual inspection of these results that drifting sampling patterns create patterned monthly biases.

While it is clear that these sampling biases will create problems attempting to use the MZM method to assess the seasonal cycle or how it changes over time, this investigation is more focused towards the effects on long-duration variability. For each year, latitude band, and altitude bin an average of the monthly sampling biases can be computed to produce a yearly-averaged bias shown in the top row of Fig. 7. While the magnitudes of yearly-averaged biases are smaller than those of monthly biases, systematic patterns are still evident. To illustrate the potential impact these sampling biases can have when incorporated into regression analyses, we can look at an individual sampling bias time-series by extracting data from the top row of Fig. 7 and plotting it along with the low frequency variability from the STS regression. The bottom row of Fig. 7 shows this data in

black with the solar cycle (red) and long-term trend (blue) overplotted to demonstrate how easily the drifting sampling patterns create patterned biases that alias into interannual and long-term geophysical variability. This will ultimately interfere with the ability of any analysis to accurately determine the "true" long-term trends. While it may appear that these results are being "cherry-picked" (and since only so many figures can be shown, they are), in actuality it is a "fruitful-tree" and results shown here are common (see the Supplement for more plots).

It should be noted that while the presence of seasonal sampling biases that alias into longer-duration terms is pervasive (in altitude and latitude for each data set), the actual degree of correlation with terms such as the solar cycle or trend is somewhat more "random" as it is dependent upon the chance combination of drifting sampling patterns, spatially-varying seasonal gradients, and frequency of interannual variability. Additionally, the seasonal sampling biases will correlate with multiple terms simultaneously, making a simple and concise quantitative evaluation of their impact on the analysis results almost impossible. Ultimately, however, it is readily apparent that the use of an MZM analysis method on data with obvious seasonal sampling biases will produce biased results in derived long-term variability.

5.2 Diurnal Sampling

With a few exceptions (e.g., Kyrölä et al., 2013; Remsberg, 2014; Damadeo et al., 2014), most analyses of ozone trends make use of MZM values where SR and SS measurements are treated equally. This has been done with the assumption that the mean value will fall between the SR and SS means but that any sampling biases are a source of random noise and do not affect the trend. As a result, should the distribution of diurnal sampling not be evenly distributed, the risk of a diurnal sampling bias becomes apparent.

In a similar way as the seasonal sampling, the nature of the orbit of the spacecraft dictates how the instrument will sample local sunrises and sunsets as a function of time of year and latitude over the mission lifetime. An example of the diurnal sampling of the SAGE II instrument over its lifetime is shown in Fig. 8. The most apparent features are the increased rate of sampling at mid-latitudes versus high and low latitudes and the presence of instrument problems during the mission that caused asymmetric diurnal sampling between mid-1993 and mid-1994 and after 2000. However, it is by taking the difference between the sunrise and sunset sampling that the true diurnal sampling differences become apparent. A close investigation of the bottom panel of Fig. 8 for any given latitude reveals a rapid oscillation of monthly biases between SR and SS dominant months. In the presence of significant (i.e., a few percent) diurnal variability such as in the upper stratosphere, this sampling bias will interfere with the derivation of the seasonal cycle for a MZM analysis.

To get a better idea of the systematic long-term nature of the diurnal sampling bias, we have looked at larger latitude bands (i.e., 35–45°N/S and ~~20~~15°~~S–20~~S–15°N) using the data in Fig. 8. By smoothing the data over a year, we can dampen some of the seasonal effects and more easily investigate the long-term changes. Also, to intercompare different latitude bands, it is preferable to look at the differences between SR and SS sampling as a percentage of the total events rather than the absolute number of events. This actually raises the question of whether to convert the differences as a number of events to a percentage of the total number of events (for each month) and then smooth or the other way around. It is interesting because this question draws a corollary with the concept of computing unweighted or weighted monthly mean values. If MZM values were computed

by first calculating a mean value for each month and then computing a mean July (for example) by simply taking the mean of all Julys, that would be unweighted (i.e., all Julys are treated equally in the overall mean regardless of how many events are in each July) and would be analogous to our converting to a percent first and then smoothing. If, however, one were to compute the overall July mean by factoring in how many events went into each month, that would be weighted and would be analogous to our smoothing in number of events and then converting to a percentage.

Figure 9 illustrates this approach for the three latitude bands for each of the three main data sets (i.e., SAGE II, HALOE, and ACE-FTS) where the "unweighted" approach is shown in the left column and the "weighted" approach is shown in the right column. As can be seen, the "unweighted" approach is more susceptible to creating a diurnal sampling bias that aliases into longer-duration variability than the "weighted" approach. However, even the "weighted" approach reveals that diurnal sampling biases cannot be avoided. The previously noted SAGE II instrument problem periods create large diurnal sampling biases with the net effect of creating large discrepancies in derived potential recovery trends (i.e., post-1997/1998) between the MZM and STS approaches. However, the diurnal sampling bias for HALOE appears to have some QBO-like periodicity that is hemispherically anticorrelated and that for ACE-FTS appears to have an overall trend in the tropics. For each case, it becomes apparent that even an attempt to account for the diurnal sampling of these instruments in an MZM analysis (i.e., the "weighted" case) will still introduce biases unless the diurnal variability is specifically modeled or corrected for beforehand.

6 Trends

The non-uniform temporal, spatial, and diurnal sampling patterns present in occultation instruments detrimentally impact trend results derived from the MZM method. To illustrate this, we also employ an MZM regression to compare with the STS regression. The MZM method employed is a one-dimensional (i.e., time only) regression that utilizes monthly means with a minimum of 5 events in 10° wide latitude bins without differentiating between sunrise and sunset events, but otherwise uses the same proxies and statistical analysis as the STS method. Since the MZM method cannot compensate for the various sampling biases but is the de facto methodology for data product usage (e.g., trend analyses or incorporation into models), we also used the results of the STS method to create corrected versions of the different data sets for incorporation into the MZM method. The first is a diurnally corrected data set, that simply applies the derived diurnal variability to bring all individual sunrise events into the sunset regime. The second applies the diurnal correction and also uses the spatially varying seasonal ozone gradient to compute a correction based on the difference between the location and time an event occurred versus the center of that month and the latitude bin it would fall within for a particular MZM averaging scheme. It is important to note that this "seasonal correction" retains variability between events within a month and bin (i.e., it does not make all values the same) and is specific to the latitude bin (i.e., width of the bin and center of the bin). The MZM regression is then applied to each of these three data versions (i.e., uncorrected or "Raw", diurnally corrected or "DCorr", and diurnally and seasonally corrected or "DSCorr").

We compute trends and uncertainties using the resulting two orthogonal EESC-proxy functions ~~. Unlike a piecewise linear trend term, the EESC-proxy terms are comprised of two separate temporal coefficients and uncertainties with functional shapes that are nonlinear, making a simple determination of the resulting overall trends and uncertainties impossible. Instead,~~

~~we take the EESC-proxy component of the fit evaluated at 10 points per year~~ and the method described in Appendix B. This uses a simple linear fit to the EESC-component of the regression results evaluated over a desired time period ~~and their uncertainties and compute a simple weighted~~ to derive the trend and makes a correlation between the EESC-fit uncertainties and the functional form of the linear fit to ~~this data. The resulting slope and uncertainty in the slope yield the trend and uncertaintyvalues~~derive the associated trend uncertainty. The derived trends from the MZM and STS methods for a typical decline period (1985–1995) are shown in Fig. 10. As expected, the difference between the MZM and STS methods during this time is small. The diurnal correction (i.e., comparing "MZM DCorr" with "MZM Raw") has some limited impact in the upper stratosphere at mid-latitudes while the seasonal correction (i.e., comparing "MZM DSCorr" with "MZM DCorr") has larger influence at higher latitudes (at all altitudes) as well as some minor influence in the tropical middle stratosphere, though trends in this area are not significant. Overall, however, the resulting trends are typical of other studies, though we would like to note that the positive trends in the tropical lower stratosphere below $\sim 23\,\mathrm{km}$ are, similar to the solar cycle response, detrimentally affected by the anomalous aerosol response.

The derived trends from the MZM and STS methods for a potential recovery period (2000–2012) are shown in Fig. 11. There are significant differences between the ~~MZM and~~ raw MZM results and the STS results most noticeably from the diurnal sampling biases. Trends in the upper stratosphere at mid-southern latitudes decrease by $\sim 1\%$/decade while trends in the upper stratosphere at mid-northern latitudes increase by~~about~~ $\sim 1\%$/decade, which is consistent with the expectations from diurnal sampling biases in the SAGE II data set. The seasonal correction, as in the decline period, influences the trends at higher latitudes as well as some minor influence in the tropical middle stratosphere. It is worth noting that generally, the fully corrected MZM data results agree much better, though expectedly not identically given the different data resolutions and techniques, with the STS results when compared to the raw MZM results. Overall the results show statistically significant trends of about 2–3%/decade in isolated parts of the upper stratosphere at mid-latitudes as well as in the tropical middle stratosphere. However, as discussed in the next section, there are other factors that affect these results that may indicate these trends are not only statistically insignificant but potentially biased as well.

7 Limitations and Orthogonality

One of the biggest issues in every regression technique is the combination of multicollinearity and orthogonality. Multicollinearity refers to the fact that the proxies used in the regression are not orthogonal to every other proxy used and that individual proxies or linear combinations of proxies are correlated with other proxies. The larger the collinearity between two or more proxies, the more difficult it is to separate their influences on the data. Sometimes proxies are sufficiently independent as to be useable, but when sampled in a particular way (e.g., to match the sampling of a particular data set) the resulting sub-sampled proxies exhibit larger collinearity. A clear example of this is seen in the diurnal and seasonal sampling patterns of the three instruments. Over their mission lifetimes, the diurnal and seasonal sampling patterns in SAGE II and HALOE are sufficiently orthogonal such that the regression can extract both the diurnal variability and seasonal cycles in each instrument separately. However, this is not the case for ACE-FTS as its diurnal and seasonal sampling patterns are highly correlated.

Figure 12 illustrates the diurnal variability for each instrument when the regression allows each instrument to have its own seasonal cycle. When compared with Fig. 3, the results for SAGE II and HALOE are the same illustrating sufficient orthogonality in their sampling patterns and the fact that their seasonal cycles are essentially the same as well. However, the results for ACE-FTS lose coherence and agreement with other studies. It is for this reason that this work made use of a single seasonal cycle as it allowed SAGE II and HALOE to constrain the seasonal cycle and thus make the extraction of the diurnal variability in the ACE-FTS data set possible. Furthermore, the fact that using a single seasonal cycle allows the independent extraction of diurnal variability that agrees well with other studies suggests that all ~~all~~ three instrument do, in fact, observe the same seasonal cycle.

The most recent Scientific Assessment of Ozone Depletion (WMO, 2014) noted that a primary problem when attempting to derive long-term trends in ozone when incorporating multiple data sets is that of instrument offsets and drifts. Given any overlap between two instruments, the offset between instruments is easily characterizable though many trend analyses are performed on anomalies and so these offsets are inherently removed. Drifts between instruments, however, are much more difficult to characterize. Hubert et al. (2016) performed an extensive analysis of ground and satellite data sets in an attempt to assess the average drifts present in each satellite data set relative to the ground network. The results showed that some instruments were more stable than others (e.g., SAGE II, HALOE, ACE-FTS, and MLS), though the degree of overlap between the satellite sampling patterns and the available ground stations did preclude the ability to determine the full spatial extent of drifts for every instrument (e.g., ACE-FTS).

This work incorporates an offset and a drift term for HALOE and ACE-FTS relative to SAGE II. The offset terms (not shown) are similar to those found in other studies comparing these instruments and are not a focus here. Figure 13 shows the linear drifts relative to SAGE II. Throughout most of the stratosphere, HALOE shows a negative drift of $\sim$ ~~2–4~~2–3%/decade relative to SAGE II, which is in good agreement with other studies (e.g., Morris et al., 2002; Nazaryan et al., 2005; Hubert et al., 2016). The drift results from ACE-FTS, however, require a different interpretation. A quick comparison of the ACE-FTS drifts in Fig. 13 and the STS "recovery" trend results in Fig. 11 shows that the patterns in the drifts somewhat match the patterns in the trends. This suggests that trends in the ACE-FTS data set are different from those in the SAGE II data set and highlights another example of the orthogonality problem. Over the course of the ACE-FTS mission period, the long-term trend terms and the drift terms are highly correlated, which is not the case for HALOE because HALOE spans the ozone turnaround time in the late 1990s, creating sufficient orthogonality between the long-term trend terms and its drift term. This means that the long-term trends are constrained by SAGE II and HALOE (and an independent HALOE drift can also be determined), but any difference in what the ACE-FTS data may suggest the trend is goes entirely into the drift term. This is further complicated by the fact that ACE-FTS data only has two years of overlap with SAGE II and HALOE. When the regression is run without any drift term (Fig. 14), the "recovery" trend results can be changed by up to $\sim$ 2%/decade, indicating a potential additional uncertainty originating from possible drift between this particular combination of data sets (similar to what was shown in Harris et al. (2015)) and that derived recovery trends are sensitive to how potential drifts are incorporated or accounted for. Overall, the issue of orthogonality highlights the limitations ~~in these~~ of regression techniques and illustrates how it is actually impossible

to simultaneously determine both potential recovery trends and relative instrument drifts using data from only after the ozone turnaround.

8 Conclusions and Future Work

A simultaneous temporal and spatial regression applied to multiple occultation data sets simultaneously without homogenization has been presented. The technique allows for a stratospheric ozone trend analysis that natively compensates for the non-uniform temporal, spatial, and diurnal sampling patterns of the data sets and results on data quality, diurnal variability, response to aerosol, and the solar cycle were shown. The STS regression shows the natural derivation of the diurnal variability captured in each instrument and highlights the impact of how the seasonal cycle is incorporated, revealing that only a single uniform seasonal cycle should be used for ~~these~~ regression analyses. Comparison of the aerosol responses in SAGE II and HALOE suggests the need to potentially revisit suggested data usage filtering criteria and the increased temporal extent of data used in the study helps to separate apparent aerosol and solar cycle responses to reveal how only a single solar cycle term should be used. Additionally, a detailed discussion of the nature of the sampling biases reveals how they impact the retrieval of long-term trends when performing regressions on MZMs causing differences in potential recovery trends up to $\sim 1\%$/decade, though we also introduce corrected versions of the data sets for use with MZM methods that apply a first-order sampling bias correction for use with trend analyses. While these corrected MZM data sets naturally do not produce identical results as the STS, they are in better agreement. This study also highlights the limitations inherent in ~~these~~ regression techniques and details how problems with multicollinearity and lack of orthogonality can impede accurate determination of long-term trends in ozone.

For future work, we would like to continue to address the topic of drifts and orthogonality as this study has shown impacts of the drifts on derived trends of up to $\sim 2\%$/decade. It is currently impossible to simultaneously determine both potential recovery trends and relative instrument drifts but it is also impossible to ascertain a global picture of drifts for every satellite instrument due to lack of necessary coverage overlaps. That being said, an analysis could be performed where a relatively stable and long-lived dense sampler (e.g., MLS) is used as the reference instrument while incorporating all other desired instruments as well. With sufficient overlap with all other data sets, an STS regression could determine the globally resolved drifts between the reference instrument and all other instruments. The derived trends, however, would come only from the reference instrument but a follow-up analysis where the reference instrument is compared to the ground network could ascertain its drift and use it as a transfer standard. In this way, all instruments could be "drift-corrected" and then fed into a final STS regression (without a drift term) so that all of the data is used to constrain the trend.

Appendix A

This work is primarily a continuation and expansion of Damadeo et al. (2014). That work discusses the application of a simultaneous temporal and spatial (STS) multiple linear regression (MLR) analysis applied to SAGE II stratospheric ozone data. This work uses the techniques described in Damadeo et al. (2014) and expands them to include multiple occultation data sets. For the sake of brevity and to assist the reader, this appendix will summarize the methodology and detail how it was expanded to incorporate multiple data sets.

Occultation instruments provide observations at two distinct latitude bands each day separated by spacecraft event type (i.e., sunrise or sunset as seen by the spacecraft). These observations are evenly distributed in longitude and span about 3 degrees in latitude at the highest latitudes to about 10 degrees in latitude in the tropics. The location of these bands gradually move from day to day according to the spacecraft's orbit and can occasionally cross each other. The data for each instrument are averaged according to these daily zonal bands and are separated by both the local and spacecraft event types so that each day can produce up to 4 data points for a single instrument. When multiple instruments are used, this process is done separately for each instrument meaning that, on a given day, it is possible to have multiple data points at the same latitude from different instruments feeding into the regression simultaneously.

The regression model applied to all of this averaged data has the following form:

$$\eta(\theta,t) = \sum_i \sum_j \beta_{i,j}\, \Theta_i(\theta)\, T_j(t)$$

where η is the concentration of O_3, $\Theta_i(\theta)$ is the functional form of the latitude dependence (Legendre polynomials in spherical harmonics), $T_j(t)$ is the functional form of the temporal dependence, and $\beta_{i,j}$ are the coefficients of the regression. The $T_j(t)$ represent all of the typical proxies used in MLR analyses (e.g., QBO, ENSO, solar, etc.) as well as several conditional proxies. Conditional proxies are simply 0 or some value (typically 1 to make it a binary conditional) depending upon whether a condition is met or not for each data point. For example, the diurnal variability proxy is 0 for every data point that is a sunset and 1 for every data point that is a sunrise. In this way, the diurnal coefficient (or rather set of coefficients because there are multiple "i" values for each "j") represents the difference between sunrise and sunset events. Additionally, some proxies are applied separately by adding another condition. Continuing the diurnal example, the diurnal variability is actually applied separately for each of the instruments so instead of a single $T_j(t)$ there are 3 (one for each instrument). The condition is a simple logical "AND" between the diurnal condition just described and a test to see if the data point of interest comes from the instrument to which the proxy applies. Similarly, there are two mean offset binary conditional terms (i.e., one for HALOE relative to SAGE II and one for ACE-FTS relative to SAGE II) and there are two drift conditional terms with forms $T_j(t) = t - t_{0,j}$ where $t_{0,j}$ is chosen at the middle of each instrument's mission period for HALOE and ACE-FTS. This process of creating conditional proxies can be repeated to apply certain temporal proxies separately to data points from different data sets (e.g., having a single seasonal cycle applied to all data sets or having each data set have its own).

Once the regression is applied, autocorrelation and heteroscedasticity corrections are applied as detailed in Damadeo et al. (2014). These corrections use the total and uncorrelated residuals from the regression to improve the uncertainties in the coefficients

that would otherwise be underestimated. When using multiple data sets, these corrections are applied separately by first subsetting the residuals to only those from a single instrument, applying the corrections, and then repeating the process for each instrument. Applying these corrections separately ignores correlations between the data sets and their impacts on the uncertainties. However, we believe this to be a second order effect as the occurrence of global perturbations is negligible and the number of coincidences between the occultation instruments is small when compared to the ensemble.

Appendix B

The goal of this work is determine ozone trends and their uncertainties from the proxies used in the regression. In the case of a piecewise linear trend (PWLT) proxy, the trend is simply the coefficient corresponding to that particular time period (or, in the case of the STS regression, an aggregate coefficient evaluated at a particular latitude). Unlike a PWLT term, the EESC-proxy terms are comprised of two separate temporal coefficients and uncertainties with functional shapes that are nonlinear, making a simple determination of the resulting overall trends and uncertainties impossible. Instead, we begin by taking the EESC-proxy component of the fit and its associated uncertainties that have the following forms:

$$y(\theta_0,t) = C_{EESC_1}(\theta_0)\, T_{j=EESC_1}(t) + C_{EESC_2}(\theta_0)\, T_{j=EESC_2}(t) \tag{B1}$$

and

$$\sigma_y^2(\theta_0,t) = \sigma^2_{EESC_1}(\theta_0)\, T^2_{EESC_1}(t) + \sigma^2_{EESC_2}(\theta_0)\, T^2_{EESC_2}(t), \tag{B2}$$

where $C_{EESC_{1,2}}(\theta_0)$ are the aggregate coefficients from the regression evaluated at a particular latitude θ_0 computed as:

$$C_{EESC_{1,2}}(\theta_0) = \sum_i \beta_{i,j=EESC_{1,2}}\, \Theta_i(\theta_0), \tag{B3}$$

with

$$\sigma^2_{EESC_{1,2}}(\theta_0) = \sum_i \sigma^2_{\beta_{i,j=EESC_{1,2}}}\, \Theta^2_i(\theta_0), \tag{B4}$$

and $T_{EESC_{1,2}}$ are the EESC proxies. The equivalent trend is then computed by performing a simple linear fit to these data over a desired time period (e.g., 2000–2012) and using the resulting slope as the trend. The resulting uncertainty in this slope, however, is more complicated because the uncertainty from the linear fit can vary with the arbitrary number of points used to create the EESC-fit. We have concluded that the best way to relate a linear fit to the EESC-fit was to draw a corollary to the uncertainties associated with a straight line fit. A linear fit to the EESC-fit data and its uncertainty have the following forms:

$$y^{'}(t) = c_0 + c_1\,(t - t_0) \tag{B5}$$

and

$$\sigma_{y'}(t) = \sqrt{\sigma_{c_0}^2 + \sigma_{c_1}^2\ (t-t_0)^2}, \tag{B6}$$

where $y^{'}(t)$ is the best fit to $y(\theta_0,t)$ and c_0 and c_1 come from the linear fit but the difficulty is determining σ_{c_0} and σ_{c_1}. It is worth noting that, for the linear fit to the EESC-fit, the choice of t_0 is arbitrary when we only care about c_1. From these equations, the correlation is made between the linear equation and its functional uncertainties (i.e., $\sigma_{y'}$ that are unknown) and the actual uncertainties from the EESC-fit (i.e., σ_y). From the above we have

$$\sigma_{c_0} = \sigma_{y'}(t_0), \tag{B7}$$

to which we draw the corollary

$$\sigma_{c_0} = MINIMUM\left\{\sigma_y(\theta_0,t)\right\} = \sigma_y(\theta_0,t_0) \tag{B8}$$

that yields σ_{c_0} and t_0. From there, it is simple to look at σ_{c_1}:

$$\sigma_{c_1} = \sqrt{\frac{\sigma_{y'}^2(t) - \sigma_{c_0}^2}{(t-t_0)^2}}, \tag{B9}$$

to which we draw the corollary

$$\sigma_{c_1} = MEAN\left\{\sqrt{\frac{\sigma_y^2(\theta_0,t) - \sigma_{c_0}^2}{(t-t_0)^2}}\right\}. \tag{B10}$$

Thus, using a direct correlation between the EESC-fit and the functional form of the linear fit, the uncertainties in the EESC-fit can be used to derive a reasonable estimate for the uncertainty in the fitted slope.

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

Acknowledgements. The ongoing development, production, assessment, and analysis of SAGE data sets at NASA Langley Research Center is supported by NASA's Earth Science Division. The Atmospheric Chemistry Experiment (ACE), also known as SCISAT, is a Canadian-led mission mainly supported by the Canadian Space Agency and the Natural Sciences and Engineering Research Council of Canada.

References

Aquila, V., Oman, L. D., Stolarski, R., Douglass, A. R., and Newman, P. A.: The Response of Ozone and Nitrogen Dioxide to the Eruption of Mt. Pinatubo at Southern and Northern Midlatitudes, J. Atmos. Sci., 70, 894–900, https://doi.org/10.1175/JAS-D-12-0143.1, http://dx.doi.org/10.1175/JAS-D-12-0143.1, 2013.

Bernath, P. F.: The Atmospheric Chemistry Experiment (ACE), J. Quant. Spectrosc. Ra., 186, 3–16, https://doi.org/10.1016/j.jqsrt.2016.04.006, http://www.sciencedirect.com/science/article/pii/S0022407316300176, 2017.

Bernath, P. F., McElroy, C. T., Abrams, M. C., Boone, C. D., Butler, M., Camy-Peyret, C., Carleer, M., Clerbaux, C., Coheur, P.-F., Colin, R., DeCola, P., DeMazière, M., Drummond, J. R., Dufour, D., Evans, W. F. J., Fast, H., Fussen, D., Gilbert, K., Jennings, D. E., Llewellyn, E. J., Lowe, R. P., Mahieu, E., McConnell, J. C., McHugh, M., McLeod, S. D., Michaud, R., Midwinter, C., Nassar, R., Nichitiu, F., Nowlan, C., Rinsland, C. P., Rochon, Y. J., Rowlands, N., Semeniuk, K., Simon, P., Skelton, R., Sloan, J. J., Soucy, M.-A., Strong, K., Tremblay, P., Turnbull, D., Walker, K. A., Walkty, I., Wardle, D. A., Wehrle, V., Zander, R., and Zou, J.: Atmospheric Chemistry Experiment (ACE): Mission overview, Geophys. Res. Lett., 32, https://doi.org/10.1029/2005GL022386, http://dx.doi.org/10.1029/2005GL022386, 2005.

Bhatt, P. P., Remsberg, E. E., Gordley, L. L., McInerney, M. J., Brackett, V. G., and Russell, J. M.: An evaluation of the quality of Halogen Occultation Experiment ozone profiles in the lower stratosphere, J. Geophys. Res.: Atmospheres, 104, 9261–9275, https://doi.org/10.1029/1999JD900058, http://dx.doi.org/10.1029/1999JD900058, 1999.

Bodeker, G. E., Boyd, I. S., and Matthews, W. A.: Trends and variability in vertical ozone and temperature profiles measured by ozonesondes at Lauder, New Zealand: 1986–1996, J. Geophys. Res.: Atmospheres, 103, 28 661–28 681, https://doi.org/10.1029/98JD02581, http://dx.doi.org/10.1029/98JD02581, 1998.

Bodeker, G. E., Scott, J. C., Kreher, K., and McKenzie, R. L.: Global ozone trends in potential vorticity coordinates using TOMS and GOME intercompared against the Dobson network: 1978–1998, J. Geophys. Res.: Atmospheres, 106, 23 029–23 042, https://doi.org/10.1029/2001JD900220, http://dx.doi.org/10.1029/2001JD900220, 2001.

Bodeker, G. E., Hassler, B., Young, P. J., and Portmann, R. W.: A vertically resolved, global, gap-free ozone database for assessing or constraining global climate model simulations, Earth. Syst. Sci. Data, 5, 31–43, https://doi.org/10.5194/essd-5-31-2013, http://www.earth-syst-sci-data.net/5/31/2013/, 2013.

Boone, C. D., Nassar, R., Walker, K. A., Rochon, Y., McLeod, S. D., Rinsland, C. P., and Bernath, P. F.: Retrievals for the atmospheric chemistry experiment Fourier-transform spectrometer, Appl. Opt., 44, 7218–7231, https://doi.org/10.1364/AO.44.007218, http://ao.osa.org/abstract.cfm?URI=ao-44-33-7218, 2005.

Boone, C. D., Walker, K. A., and Bernath, P. F.: Version 3 Retrievals for the Atmospheric Chemistry Experiment Fourier Transform Spectrometer (ACE-FTS), in: The Atmospheric Chemistry Experiment ACE at 10: A Solar Occultation Anthology, edited by Bernath, P. F., pp. 103–127, A. Deepak Publishing, Hampton, Virginia, U.S.A., 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. and Phys., 14, 6983–6994, https://doi.org/10.5194/acp-14-6983-2014, http://www.atmos-chem-phys.net/14/6983/2014/, 2014.

Chapman, S.: A theory of upper atmosphere ozone, Mem. R. Metrol. Soc., 3, https://www.rmets.org/sites/default/files/chapman-memoirs.pdf, 1930.

Cunnold, D. and McCormick, M.: SAGE III Algorithm Theoretical Basis Document (ATBD) Transmission Level 1B Products, Tech. Rep. LaRC 475-00-108, NASA, 2002.

Damadeo, R. P., Zawodny, J. M., Thomason, L. W., and Iyer, N.: SAGE version 7.0 algorithm: application to SAGE II, Atmos. Meas. Tech., 6, 3539–3561, https://doi.org/10.5194/amt-6-3539-2013, http://www.atmos-meas-tech.net/6/3539/2013/, 2013.

Damadeo, R. P., Zawodny, J. M., and Thomason, L. W.: Reevaluation of stratospheric ozone trends from SAGE II data using a simultaneous temporal and spatial analysis, Atmos. Chem. and Phys., 14, 13 455–13 470, https://doi.org/10.5194/acp-14-13455-2014, http://www.atmos-chem-phys.net/14/13455/2014/, 2014.

Dhomse, S. S., Chipperfield, M. P., Damadeo, R. P., Zawodny, J. M., Ball, W. T., Feng, W., Hossaini, R., Mann, G. W., and Haigh, J. D.: On the ambiguous nature of the 11 year solar cycle signal in upper stratospheric ozone, Geophys. Res. Lett., 43, 7241–7249, https://doi.org/10.1002/2016GL069958, http://dx.doi.org/10.1002/2016GL069958, 2016.

Gebhardt, C., Rozanov, A., Hommel, R., Weber, M., Bovensmann, H., Burrows, J. P., Degenstein, D., Froidevaux, L., and Thompson, A. M.: Stratospheric ozone trends and variability as seen by SCIAMACHY from 2002 to 2012, Atmos. Chem. and Phys., 14, 831–846, https://doi.org/10.5194/acp-14-831-2014, http://www.atmos-chem-phys.net/14/831/2014/, 2014.

Glaccum, W., Lucke, R. L., Bevilacqua, R. M., Shettle, E. P., Hornstein, J. S., Chen, D. T., Lumpe, J. D., Krigman, S. S., Debrestian, D. J., Fromm, M. D., Dalaudier, F., Chassefière, E., Deniel, C., Randall, C. E., Rusch, D. W., Olivero, J. J., Brogniez, C., Lenoble, J., and Kremer, R.: The Polar Ozone and Aerosol Measurement instrument, J. Geophys. Res.: Atmospheres, 101, 14 479–14 487, https://doi.org/10.1029/96JD00576, http://dx.doi.org/10.1029/96JD00576, 1996.

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

Herman, J. R.: The response of stratospheric constituents to a solar eclipse, sunrise, and sunset, J. Geophys. Res.: Oceans, 84, 3701–3710, https://doi.org/10.1029/JC084iC07p03701, http://dx.doi.org/10.1029/JC084iC07p03701, 1979.

Hervig, M. E., Russell, J. M., Gordley, L. L., Daniels, J., Drayson, S. R., and Park, J. H.: Aerosol effects and corrections in the Halogen Occultation Experiment, J. Geophys. Res.: Atmospheres, 100, 1067–1079, https://doi.org/10.1029/94JD02143, http://dx.doi.org/10.1029/94JD02143, 1995.

Hubert, D., Lambert, J.-C., Verhoelst, T., Granville, J., Keppens, A., Baray, J.-L., Bourassa, A. E., Cortesi, U., Degenstein, D. A., Froidevaux, L., Godin-Beekmann, S., Hoppel, K. W., Johnson, B. J., Kyrölä, E., Leblanc, T., Lichtenberg, G., Marchand, M., McElroy, C. T., Murtagh, D., Nakane, H., Portafaix, T., Querel, R., Russell III, J. M., Salvador, J., Smit, H. G. J., Stebel, K., Steinbrecht, W., Strawbridge, K. B., Stübi, R., Swart, D. P. J., Taha, G., Tarasick, D. W., Thompson, A. M., Urban, J., van Gijsel, J. A. E., Van Malderen, R., von der Gathen, P., Walker, K. A., Wolfram, E., and Zawodny, J. M.: Ground-based assessment of the bias and long-term stability of 14 limb and occultation ozone profile data records, Atmos. Meas. Tech., 9, 2497–2534, https://doi.org/10.5194/amt-9-2497-2016, http://www.atmos-meas-tech.net/9/2497/2016/, 2016.

Kyrölä, E., Laine, M., Sofieva, V., Tamminen, J., Päivärinta, 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. and Phys., 13, 10 645–10 658, https://doi.org/10.5194/acp-13-10645-2013, http://www.atmos-chem-phys.net/13/10645/2013/, 2013.

Lucke, R. L., Korwan, D. R., Bevilacqua, R. M., Hornstein, J. S., Shettle, E. P., Chen, D. T., Daehler, M., Lumpe, J. D., Fromm, M. D., Debrestian, D., Neff, B., Squire, M., König-Langlo, G., and Davies, J.: The Polar Ozone and Aerosol Measurement (POAM) III instrument and early validation results, J. Geophys. Res.: Atmospheres, 104, 18 785–18 799, https://doi.org/10.1029/1999JD900235, http://dx.doi.org/10.1029/1999JD900235, 1999.

Lumpe, J. D., Bevilacqua, R. M., Hoppel, K. W., Krigman, S. S., Kriebel, D. L., Debrestian, D. J., Randall, C. E., Rusch, D. W., Brogniez, C., Ramananahérisoa, R., Shettle, E. P., Olivero, J. J., Lenoble, J., and Pruvost, P.: POAM II retrieval algorithm and error analysis, J. Geophys. Res.: Atmospheres, 102, 23 593–23 614, https://doi.org/10.1029/97JD00906, http://dx.doi.org/10.1029/97JD00906, 1997.

Lumpe, J. D., Bevilacqua, R. M., Hoppel, K. W., and Randall, C. E.: POAM III retrieval algorithm and error analysis, J. Geophys. Res.: Atmospheres, 107, ACH 5–1–ACH 5–32, https://doi.org/10.1029/2002JD002137, http://dx.doi.org/10.1029/2002JD002137, 2002.

Mauldin, L. E., Salikhov, R., Habib, S., Vladimirov, A. G., Carraway, D., Petrenko, G., and Comella, J.: Meteor-3M(1)/Stratospheric Aerosol and Gas Experiment III (SAGE III) jointly sponsored by the National Aeronautics and Space Administration and the Russian Space Agency, vol. 3501, pp. 355–365, https://doi.org/10.1117/12.317767, http://dx.doi.org/10.1117/12.317767, 1998.

Mauldin III, L., Zaun, N., McCormick Jr., M., Guy, J., and Vaughn, W.: Stratospheric Aerosol and Gas Experiment II instrument: a functional description, Opt. Eng., 24, 307–312, https://doi.org/10.1117/12.7973473, http://dx.doi.org/10.1117/12.7973473, 1985.

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. and Phys., 16, 10 021–10 043, https://doi.org/10.5194/acp-16-10021-2016, https://www.atmos-chem-phys.net/16/10021/2016/, 2016.

McCormick, M. P., Thomason, L. W., , and Trepte, C.: Atmospheric Effects of the Mt. Pinatubo Eruption, Nature, 373, 399–404, https://doi.org/10.1038/373399a0, 1995.

Millán, L. F., Livesey, N. J., Santee, M. L., Neu, J. L., Manney, G. L., and Fuller, R. A.: Case studies of the impact of orbital sampling on stratospheric trend detection and derivation of tropical vertical velocities: solar occultation vs. limb emission sounding, Atmos. Chem. and Phys., 16, 11 521–11 534, https://doi.org/10.5194/acp-16-11521-2016, http://www.atmos-chem-phys.net/16/11521/2016/, 2016.

Morris, G. A., Gleason, J. F., Russell, J. M., Schoeberl, M. R., and McCormick, M. P.: A comparison of HALOE V19 with SAGE II V6.00 ozone observations using trajectory mapping, J. Geophys. Res.: Atmospheres, 107, ACH 10–1–ACH 10–9, https://doi.org/10.1029/2001JD000847, http://dx.doi.org/10.1029/2001JD000847, 2002.

Naval Research Laboratory: Overview of the validation of POAM III version 4 retrievals, https://eosweb.larc.nasa.gov/sites/default/files/project/poam3/readme/poam3_ver4_validation.pdf, 2006.

Nazaryan, H., McCormick, M. P., and Russell, J. M.: New studies of SAGE II and HALOE ozone profile and long-term change comparisons, J. Geophys. Res.: Atmospheres, 110, https://doi.org/10.1029/2004JD005425, http://dx.doi.org/10.1029/2004JD005425, 2005.

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

Pallister, R. C. and Tuck, A. F.: The diurnal variation of ozone in the upper stratosphere as a test of photochemical theory, Q. J. R. Meteorol. Soc., 109, 271–284, https://doi.org/10.1002/qj.49710946002, http://dx.doi.org/10.1002/qj.49710946002, 1983.

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.: Atmospheres, 112, D06 313, https://doi.org/10.1029/2006JD007339, http://dx.doi.org/10.1029/2006JD007339, 2007.

Remsberg, E. and Lingenfelser, G.: Analysis of SAGE II ozone of the middle and upper stratosphere for its response to a decadal-scale forcing, Atmos. Chem. and Phys., 10, 11 779–11 790, https://doi.org/10.5194/acp-10-11779-2010, http://www.atmos-chem-phys.net/10/11779/2010/, 2010.

Remsberg, E. E.: On the response of Halogen Occultation Experiment (HALOE) stratospheric ozone and temperature to the 11-year solar cycle forcing, J. Geophys. Res.: Atmospheres, 113, https://doi.org/10.1029/2008JD010189, http://dx.doi.org/10.1029/2008JD010189, 2008.

Remsberg, E. E.: Decadal-scale responses in middle and upper stratospheric ozone from SAGE II version 7 data, Atmos. Chem. and Phys., 14, 1039–1053, https://doi.org/10.5194/acp-14-1039-2014, http://www.atmos-chem-phys.net/14/1039/2014/, 2014.

Robock, A.: Volcanic eruptions and climate, Rev. Geophys., 38, 191–219, https://doi.org/10.1029/1998RG000054, http://dx.doi.org/10.1029/1998RG000054, 2000.

Rodriguez, J. M., Ko, M. K., and Sze, N. D.: Role of heterogeneous conversion of N2O5 on sulphate aerosols in global ozone losses, Nature, 352, 134–137, https://doi.org/10.1038/352134a0, 1991.

Russell, J. M., Gordley, L. L., Park, J. H., Drayson, S. R., Hesketh, W. D., Cicerone, R. J., Tuck, A. F., Frederick, J. E., Harries, J. E., and Crutzen, P. J.: The Halogen Occultation Experiment, J. Geophys. Res.: Atmospheres, 98, 10 777–10 797, https://doi.org/10.1029/93JD00799, http://dx.doi.org/10.1029/93JD00799, 1993.

Sakazaki, T., Shiotani, M., Suzuki, M., Kinnison, D., Zawodny, J. M., McHugh, M., and Walker, K. A.: Sunset–sunrise difference in solar occultation ozone measurements (SAGE II, HALOE, and ACE–FTS) and its relationship to tidal vertical winds, Atmos. Chem. and Phys., 15, 829–843, https://doi.org/10.5194/acp-15-829-2015, http://www.atmos-chem-phys.net/15/829/2015/, 2015.

Sheese, P. E., Boone, C. D., and Walker, K. A.: Detecting physically unrealistic outliers in ACE-FTS atmospheric measurements, Atmos. Meas. Tech., 8, 741–750, https://doi.org/10.5194/amt-8-741-2015, http://www.atmos-meas-tech.net/8/741/2015/, 2015.

Sofieva, V. F., Kalakoski, N., Päivärinta, S.-M., Tamminen, J., Laine, M., and Froidevaux, L.: On sampling uncertainty of satellite ozone profile measurements, Atmos. Meas. Tech., 7, 1891–1900, https://doi.org/10.5194/amt-7-1891-2014, http://www.atmos-meas-tech.net/7/1891/2014/, 2014.

Solomon, S.: Stratospheric ozone depletion: A review of concepts and history, Rev. Geophys., 37, 275–316, https://doi.org/10.1029/1999RG900008, http://dx.doi.org/10.1029/1999RG900008, 1999.

Solomon, S., Portmann, R. W., Garcia, R. R., Thomason, L. W., Poole, L. R., and McCormick, M. P.: The role of aerosol variations in anthropogenic ozone depletion at northern midlatitudes, J. Geophys. Res.: Atmospheres, 101, 6713–6727, https://doi.org/10.1029/95JD03353, http://dx.doi.org/10.1029/95JD03353, 1996.

Soukharev, B. E. and Hood, L. L.: Solar cycle variation of stratospheric ozone: Multiple regression analysis of long-term satellite data sets and comparisons with models, J. Geophys. Res.: Atmospheres, 111, https://doi.org/10.1029/2006JD007107, http://dx.doi.org/10.1029/2006JD007107, 2006.

Steinbrecht, W., Froidevaux, L., Fuller, R., Wang, R., Anderson, J., Roth, C., Bourassa, A., Degenstein, D., Damadeo, R., Zawodny, J., Frith, S., McPeters, R., Bhartia, P., Wild, J., Long, C., Davis, S., Rosenlof, K., Sofieva, V., Walker, K., Rahpoe, N., Rozanov, A., Weber, M., Laeng, A., von Clarmann, T., Stiller, G., Kramarova, N., Godin-Beekmann, S., Leblanc, T., Querel, R., Swart, D., Boyd, I., Hocke, K., Kämpfer, N., Maillard Barras, E., Moreira, L., Nedoluha, G., Vigouroux, C., Blumenstock, T., Schneider, M., García, O., Jones, N., Mahieu, E., Smale, D., Kotkamp, M., Robinson, J., Petropavlovskikh, I., Harris, N., Hassler, B., Hubert, D., and Tummon, F.: An update on ozone profile trends for the period 2000 to 2016, Atmos. Chem. and Phys., 17, 10 675–10 690, https://doi.org/10.5194/acp-17-10675-2017, https://www.atmos-chem-phys.net/17/10675/2017/, 2017.

Stolarski, R., Douglass, A., Steenrod, S., and Pawson, S.: Trends in Stratospheric Ozone: Lessons Learned from a 3D Chemical Transport Model, J. Atmos. Sci., 63, 1028–1041, https://doi.org/10.1175/JAS3650.1, http://dx.doi.org/10.1175/JAS3650.1, 2006.

Thompson, R. E. and Gordley, L. L.: Retrieval Algorithms for the Halogen Occultation Experiment, Tech. Rep. NASA/CR–2009-215761, National Aeronautics and Space Administration, 2009.

Toohey, M., Hegglin, M. I., Tegtmeier, S., Anderson, J., Añel, J. A., Bourassa, A., Brohede, S., Degenstein, D., Froidevaux, L., Fuller, R., Funke, B., Gille, J., Jones, A., Kasai, Y., Krüger, K., Kyrölä, E., Neu, J. L., Rozanov, A., Smith, L., Urban, J., von Clarmann, T., Walker, K. A., and Wang, R. H. J.: Characterizing sampling biases in the trace gas climatologies of the SPARC Data Initiative, J. Geophys. Res.: Atmospheres, 118, 11,847–11,862, https://doi.org/10.1002/jgrd.50874, http://dx.doi.org/10.1002/jgrd.50874, 2013.

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. and Phys., 15, 3021–3043, https://doi.org/10.5194/acp-15-3021-2015, http://www.atmos-chem-phys.net/15/3021/2015/, 2015.

Vernier, J.-P., Thomason, L. W., Pommereau, J.-P., Bourassa, A., Pelon, J., Garnier, A., Hauchecorne, A., Blanot, L., Trepte, C., Degenstein, D., and Vargas, F.: Major influence of tropical volcanic eruptions on the stratospheric aerosol layer during the last decade, Geophys. Res. Lett., 38, https://doi.org/10.1029/2011GL047563, http://dx.doi.org/10.1029/2011GL047563, 2011.

Wang, H. J., Cunnold, D. M., and Bao, X.: A critical analysis of Stratospheric Aerosol and Gas Experiment ozone trends, J. Geophys. Res.: Atmospheres, 101, 12 495–12 514, https://doi.org/10.1029/96JD00581, http://dx.doi.org/10.1029/96JD00581, 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.: Atmospheres, 107, ACH 8–1–ACH 8–18, https://doi.org/10.1029/2002JD002418, http://dx.doi.org/10.1029/2002JD002418, 2002.

WMO: International Meteorological Vocabulary, 182, World Meteorological Organization, Geneva, 2 edn., ISBN 978-92-63-02182-3, 1992.

WMO: Scientific Assessment of Ozone Depletion: 2014, Global Ozone Research and Monitoring Project-Report No. 54, World Meteorological Organization, Geneva, Switzerland, http://www.esrl.noaa.gov/csd/assessments/ozone/2014/report.html, 2014.

Wofsy, S., Michelsen, H., and McCormick, M.: SAGE III Algorithm Theoretical Basis Document (ATBD) Solar and Lunar Algorithm, Tech. Rep. LaRC 475-00-109, NASA, 2002.

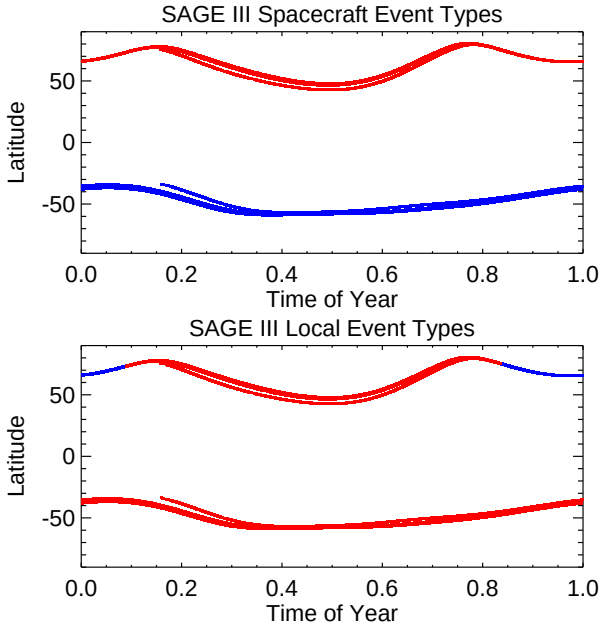

Figure 1. Location of all SAGE III occultation events for both spacecraft (top) and local (bottom) event types. In each case, sunrises are shown in blue while sunsets are shown in red. While there is a clear hemispheric distinction between spacecraft event types, nearly all local event types are sunsets with the exception of spacecraft sunset events in polar winter. Other occultation instruments in sun-synchronous orbits such as POAM II and POAM III exhibit similar behavior.

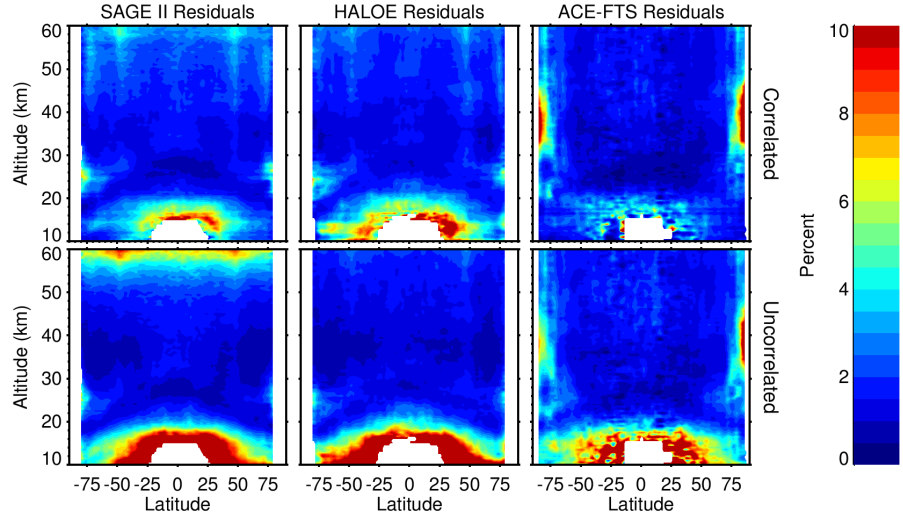

Figure 2. Spread of the ~~total~~ correlated and uncorrelated residuals as a function of latitude and altitude for each instrument from the regression. White regions show areas where insufficient data exists.

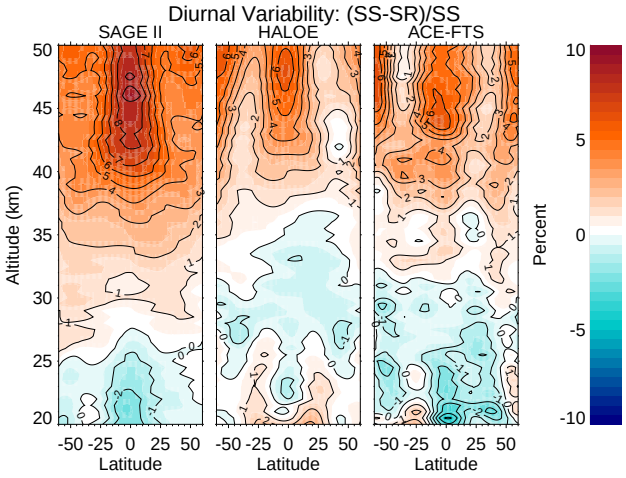

Figure 3. Results from the regression depicting the mean diurnal variability present in each data set plotted as the percent difference between sunrise and sunset events. These results compare well with those of Sakazaki et al. (2015).

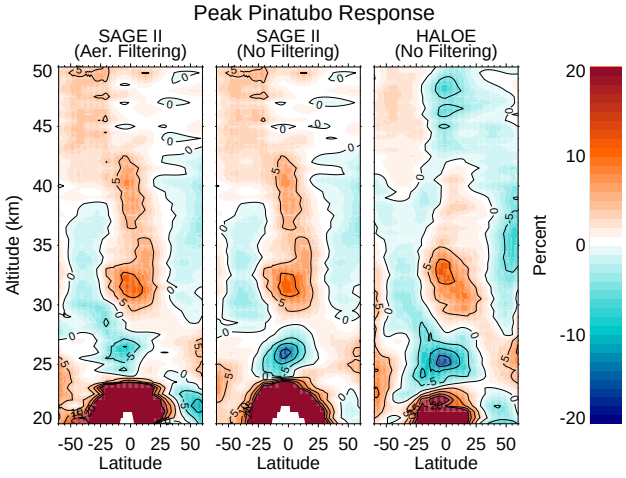

Figure 4. Peak of the volcanic term near the eruption of Mt. Pinatubo as a percentage of the local mean for both SAGE II and HALOE under different regressions. Results for SAGE II are shown both with and without the Wang et al. (2002) filtering criteria. Results for HALOE are shown without any aerosol filtering, though results with filtering are similar.

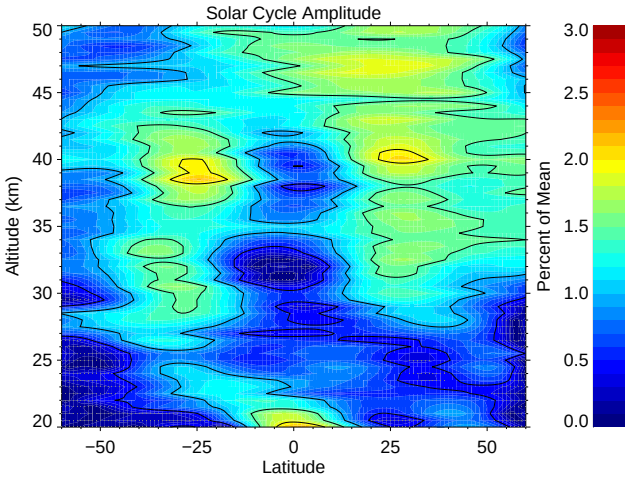

Figure 5. Amplitude of oscillation of the solar cycle response as a percentage of the local mean. Stippling denotes areas where the values are not significant at the 2σ level. Contour lines are plotted at intervals of 0.5%.

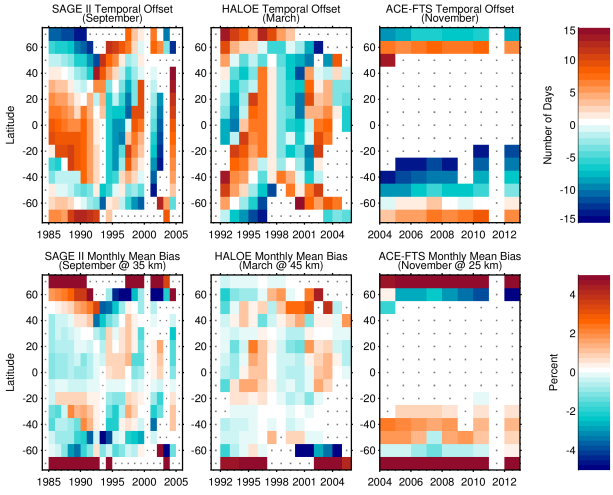

Figure 6. Top row: The MZM temporal sampling bias shown as the difference between the average time of sampling in a given month and latitude band and the center of that month. Results shown here are for different months and different data sets, though systematic biasing of results is ~~similar~~ common for most months for each data set. Bottom row: The MZM seasonal sampling bias shown as the difference in ozone between the actual center of sampling for a given month and latitude band and the center of that month and band as computed using the seasonal cycle and the local mean from the STS regression. Results are shown here for different altitudes illustrating the pervasiveness of the problem.

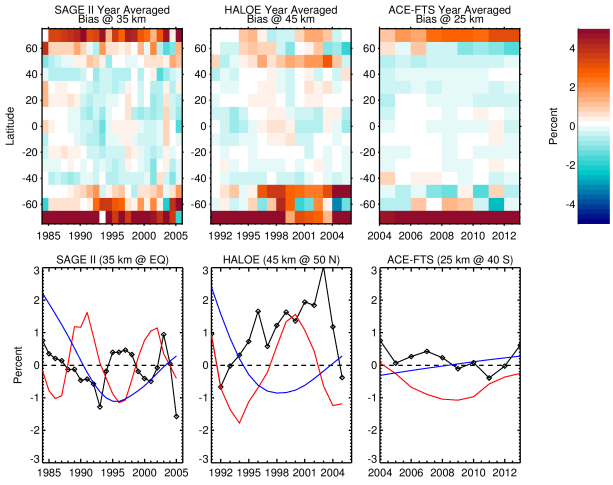

Figure 7. Top row: Yearly average of the MZM seasonal sampling biases illustrated in the bottom row of Fig. 6. While the amplitude of systematic biases decreases from the individual months, systematic biases are still apparent. Bottom row: Data extracted from the specified latitude band in the top row is plotted in black in each case. The solar cycle (red) and long-term trend (blue) from the STS regression for those altitudes and latitudes are overplotted to illustrate the potential correlation between the systematic sampling biases and long-duration variability.

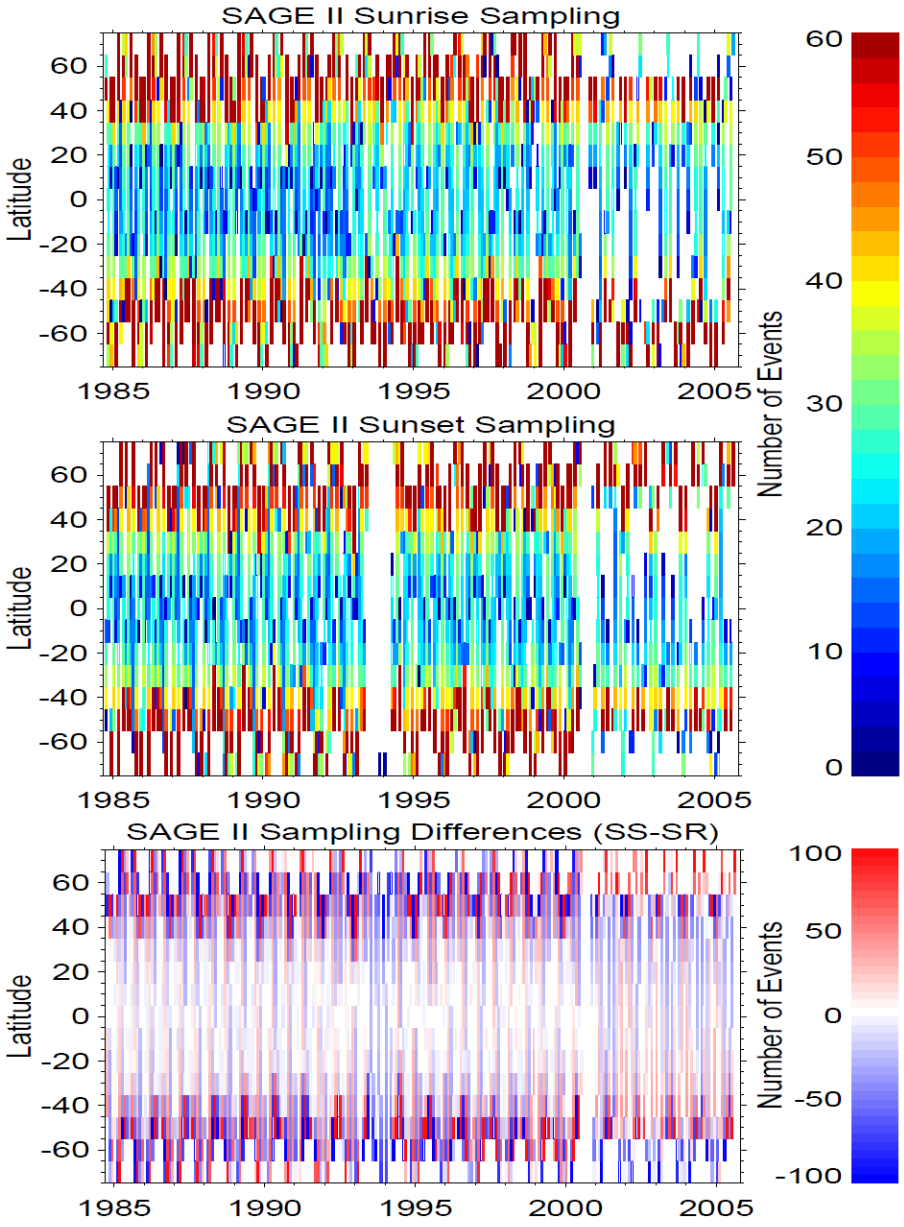

Figure 8. Monthly zonal sampling for SAGE II separated by local event type (top/middle). There was a problem with the battery that caused shortened sunset events between mid-1993 and mid-1994 and an issue with the azimuthal pointing system after late-2000 that caused a hemispheric asymmetry in sampling. The bottom panel is the difference between the top and middle panels, revealing the rapid oscillation between SR and SS dominated months as well as whole periods dominated by one event type.

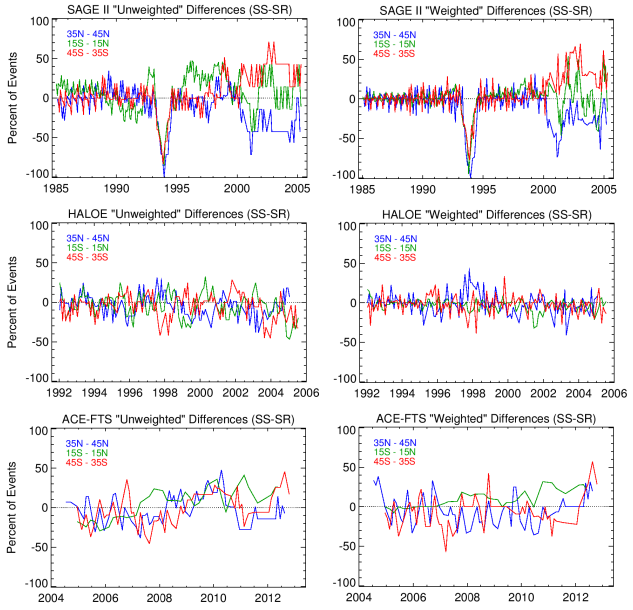

Figure 9. Long-term evolution of the diurnal sampling bias for three different data sets. The wider latitude bins are representative of data from Fig. 8. To remove the influence of the rapid monthly variability, the data is smoothed over 12 months and converted to a percent of total events. The left column first converts differences in total number of SR/SS events to percentages and then smooths while the right column first smooths in number of events and then converts to a percentage.

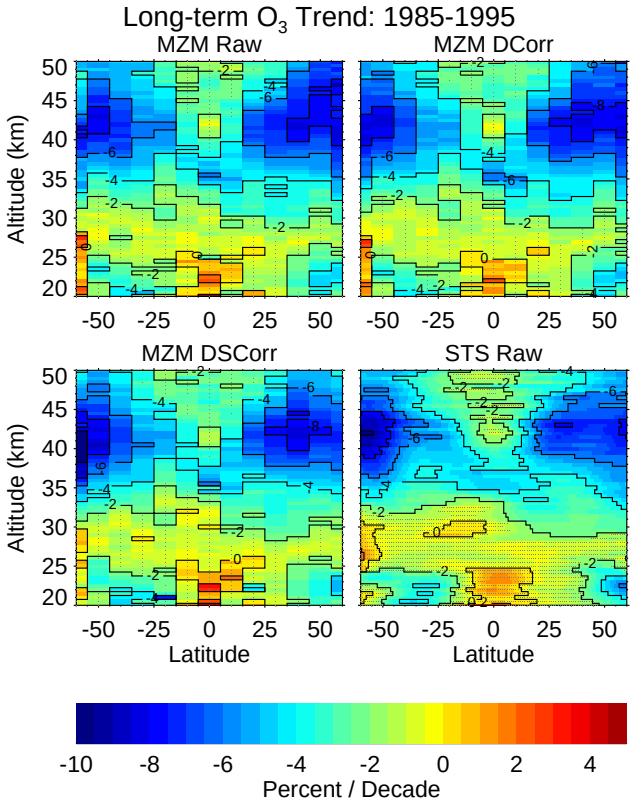

Figure 10. Long-term trends derived from both the MZM and the STS regressions during the typical decline period. Results are also shown when using the STS regression results to create a diurnally corrected (DCorr) and a diurnally & seasonally corrected (DSCorr) data set for use with the MZM regression. The diurnal correction has the greatest influence on the upper stratosphere while the seasonal correction has the greatest influence at higher latitudes. Stippling denotes areas where the trend results are not significant at the 2σ level. Contour lines are plotted at 2% intervals.

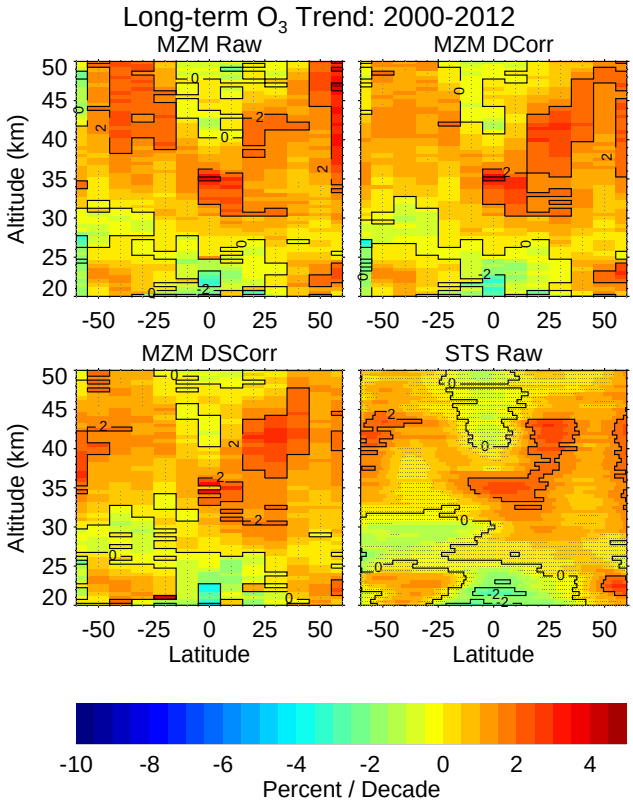

Figure 11. Long-term trends derived from both the MZM and the STS regressions during the potential recovery period. Results are also shown when using the STS regression results to create a diurnally corrected (DCorr) and a diurnally & seasonally corrected (DSCorr) data set for use with the MZM regression. The diurnal correction has the greatest influence on the upper stratosphere while the seasonal correction has the greatest influence at higher latitudes. Stippling denotes areas where the trend results are not significant at the 2σ level. Contour lines are plotted at 2% intervals.

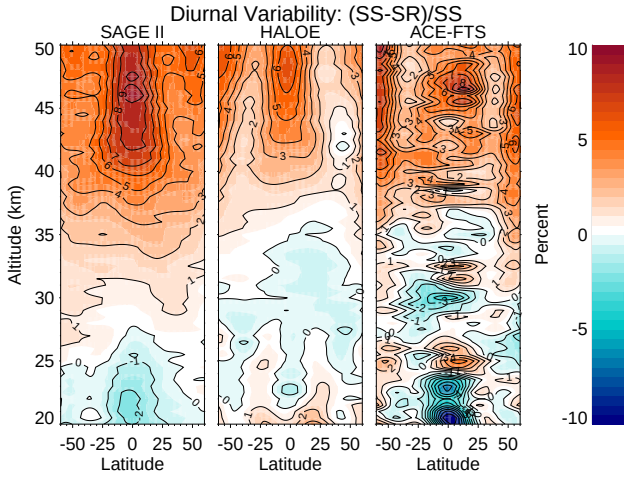

Figure 12. Same as Fig. 3 except the regression is allowed to fit different seasonal cycles for each instrument. The lack of orthogonality between the diurnal and seasonal sampling patterns in ACE-FTS makes it impossible to differentiate between the two. SAGE II and HALOE remain unaffected illustrating sufficient orthogonality and the fact that their seasonal cycles are essentially the same.

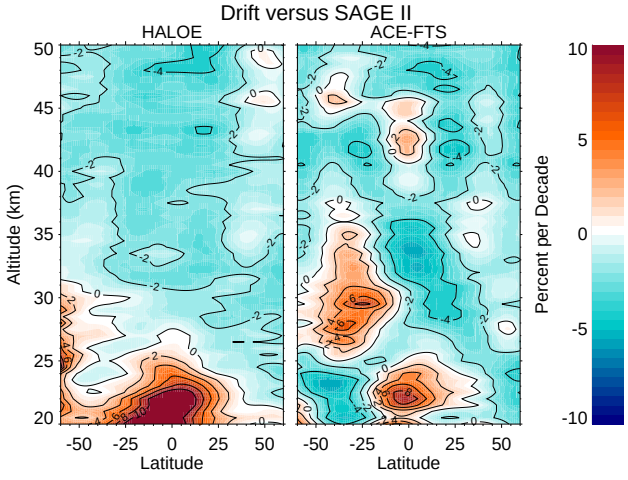

Figure 13. The result of the independent drift term used in the regression showing the relative drift from the SAGE II data for each of the other instruments. Derived drifts of $\sim$ 2–3%/decade through most of the stratosphere for HALOE agree well with earlier studies but the lack of orthogonality between trend and drift terms during the overlap between the ACE-FTS and SAGE II missions causes anomalous results.

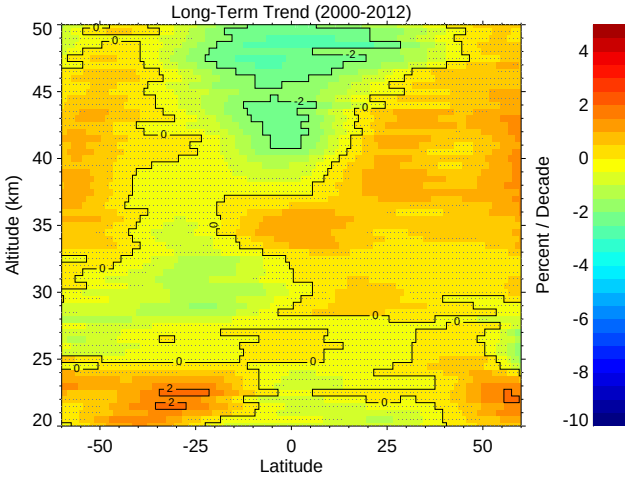

Figure 14. Same as the STS results in Fig. 11 except the regression no longer assumes any kind of drifts between the instruments. Reduction in potential recovery trends can be as high as $\sim 2\%$/decade for this particular combination of data sets.