

# **Reevaluation of stratospheric ozone trends from SAGE II data using a simultaneous temporal and spatial analysis**

**R. P. Damadeo, J. M. Zawodny, and L. W. Thomason**

NASA Langley Research Center, Hampton, VA, USA

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

## Abstract

This paper details a new method of regression for sparsely sampled data sets for use with time-series analysis, in particular the Stratospheric Aerosol and Gas Experiment (SAGE) II ozone data set. Non-uniform spatial, temporal, and diurnal sampling present in the data set result in biased values for the long-term trend if not accounted for. This new method is performed close to the native resolution of measurements and is a simultaneous temporal and spatial analysis that accounts for potential diurnal ozone variation. Results show biases, introduced by the way data is prepared for use with traditional methods, can be as high as 10%. Derived long-term changes show declines in ozone similar to other studies but very different trends in the presumed recovery period, with differences up to 2% per decade. The regression model allows for a variable turnaround time and reveals a hemispheric asymmetry in derived trends in the middle to upper stratosphere. Similar methodology is also applied to SAGE II aerosol optical depth data to create a new volcanic proxy that covers the SAGE II mission period. Ultimately this technique may be extensible towards the inclusion of multiple data sets without the need for homogenization.

## 1 Introduction

The Stratospheric Aerosol and Gas Experiment (SAGE) II flew onboard the Earth Radiation Budget Satellite (ERBS) for over 20 years from its launch in October 1984 until its retirement in August 2005. It employed the solar occultation technique to measure multi-wavelength slant-path atmospheric transmission profiles at seven channels during each sunrise and sunset encountered by the spacecraft. During its operation, SAGE II produced high precision vertical profiles of atmospheric ozone ( $\sim 1\%$  one-sigma uncertainty in the middle-stratosphere) with excellent vertical resolution ( $\sim 1$  km) (Damadeo et al., 2013). The combination of precise measurements and a long data record has seen SAGE II data consistently used for long-term ozone trend analysis (e.g., WMO, 1988, 2011; SPARC, 2010). Traditionally this is performed via multiple linear regression of monthly zonal mean ozone data to a

set of predictor variables. However, given the sparse sampling of SAGE II measurements ( $\sim 30$  observations per day), biases can be introduced if the data are not carefully treated. Herein we present a new way to perform time-series analysis on a sparsely sampled data set, in particular SAGE II, and compare the results and influence on trends with monthly zonal mean methods. The method outlined in this paper is similar to that of Bodeker et al. (2013) in that it utilizes a simultaneous temporal and spatial regression (though the terms used and how the regression is applied are different), but differs fundamentally in that it regresses to daily mean values separated by event type.

## 2 Predictor Variables

The choices of predictor variables are important and thus are chosen based on atmospheric variability that ozone has historically shown to be responsive to, namely: seasonal cycle, quasi-biennial oscillation (QBO), El Niño southern oscillation (ENSO), solar variability, volcanic eruption, and long-term trend terms. Ideally, each predictor variable is orthogonal to each other predictor variable. Since this is almost never the case, predictor variables are pretreated to create normalized orthogonal functions from multiple component data sets using empirical orthogonal function (EOF) analysis (also known as principal component analysis). When the use of EOF analysis becomes impossible due to having only one component data set, an orthogonal function is created by shifting the original function in time until the dot product between the shifted function and original function is zero over the overlap period, hereafter referred to as the temporal shift method. The use of orthogonal functions for predictors is better than the use of a single term because it allows the regression model to account for both magnitude and phase changes in the response. Ultimately, however, each set of orthogonal functions, while orthogonal internally within a set, are not necessarily orthogonal between sets.

With these tools in mind, a full set of predictor variables is created from component data sets in order to reduce the amount of multicollinearity. The seasonal cycle is simply a Fourier series with periods of 12 (annual), 6, 4, and 3 months (semiannual). To create a set of QBO

predictor variables, EOF analysis is performed on compiled monthly mean equatorial wind data (<http://www.geo.fu-berlin.de/en/met/ag/strat/produkte/qbo/qbo.dat>) at seven pressure levels (70, 50, 40, 30, 20, 15, and 10 hPa), resulting in seven orthogonal basis functions. The leading four terms account for over 99% of the variance in the QBO data so these are used as the QBO predictor variables. To create a pair of ENSO predictor variables, the temporal shift method is applied to Multivariate ENSO Index data (<http://www.esrl.noaa.gov/psd/enso/mei/>). A pair of solar predictor variables is created by applying the temporal shift method to solar 10.7 cm radio flux data ([ftp://ftp.geolab.nrcan.gc.ca/data/solar\\_flux/](ftp://ftp.geolab.nrcan.gc.ca/data/solar_flux/)). Each of these ancillary data sets is smoothed before the creation of orthogonal functions (Fig. 1) in order to minimize the effect of noise on the creation of orthogonal functions.

Two separate terms are explored for use to represent long-term changes in ozone. One is a simple piecewise linear term joined at the beginning of 1997 (i.e., both terms are linear with one being zero everywhere before 1997 and the other being zero everywhere during 1997 and after) like that in Kyrölä et al. (2013). The other is the use of terms representing equivalent effective stratospheric chlorine (EESC), which represents the total amount of chlorine and bromine loading in the stratosphere that contributes to ozone depletion. EOF analysis is performed on EESC data sets (Newman et al., 2007) for multiple mean ages of air (1, 2, 3, 4, 5, and 6 years) to retrieve two primary orthogonal functions (Fig. 1) that account for over 99% of the EESC data variance. A simple linear term in addition to EESC terms could also be included, but we found that it results in pathological behavior in the tropics in extrapolated data and is thus omitted.

The creation of a predictor variable to represent volcanic eruptions is ideally performed with atmospheric aerosol data. Often data from periods of heavy aerosol loading are omitted (e.g., Randel and Wu, 2007; Kyrölä et al., 2013; Bourassa et al., 2014) or a simple functional form for an eruption is used (e.g., Stolarski et al., 2006; Bodeker et al., 2013), but this does not take into account the varying times of injection, the change in rate of accumulation via transport, or varying decay rates, which are all functions of latitude. Other aerosol databases exist, but these are representative of total aerosol instead of just eruptive effects and seasonality cannot be trivially removed. A seasonal cycle in a predictor variable would



alias into the seasonal cycle in ozone. Since the seasonal variation of ozone is related to but not entirely dependent upon aerosol, it is preferential to have a purely eruptive term. The same is true for QBO effects or spatially varying means in the aerosol data set. SAGE II has aerosol measurements alongside ozone measurements, so in theory these data could be used as a predictor variable. However, these data have noise that is autocorrelated, making them a poor choice for use as a predictor variable. In addition, the effects of aerosol on ozone are not purely local (i.e., chemical reactions with local aerosol) but also dependent upon radiative effects from aerosol layers above and below the altitude layer in question. In the end, we chose to create our own volcanic predictor variable based on stratospheric aerosol optical depth in the 1020 nm channel. This procedure is described in Appendix A.

In addition to predictor variables that act as proxies for geophysical variability, a number of cross-terms (products of terms) can also be considered. In the end, only one is included. The data used for the QBO proxy comes from equatorial data up to an altitude of  $\sim 32$  km. However, it has been shown that the frequencies at which the QBO oscillates are different at higher latitudes than at the equator (Tung and Yang, 1994) and at higher altitudes (Remsberg and Lingenfelter, 2010). While the use of a proxy is better than simply using oscillating functions of different frequencies (since the QBO changes frequencies over time) and the use of orthogonal functions allows for the change in amplitude and phase of the response, it cannot account for the change in frequencies. In other words, regressing a QBO proxy from the equator at higher latitudes will not capture all of the variation, nor will regressing a QBO proxy from lower altitudes at higher altitudes even at the equator. It has been shown, however, that the annual cycle modulates the QBO at higher latitudes (Tung and Yang, 1994), and thus the inclusion of this cross-term would better fit the response of ozone to the QBO at higher latitudes. Ideally a multi-dimensional QBO proxy would be used that captures global variability, but to the authors' knowledge, no such proxy exists.

### 3 Pretreatment of Data

#### 3.1 SAGE II Observations

SAGE II was launched into a  $57^\circ$  inclined orbit and took two measurements per orbit. Each measurement was either a sunrise or sunset as seen by the spacecraft (spacecraft event type) and also either a sunrise or sunset as would be seen by an observer on the ground (local event type). Most of the time the spacecraft and local event types are the same, except occasionally at high latitudes. Given the nature of the orbit and observation technique, the instrument took measurements in two ground-track swaths across the surface of the Earth in a given day: one made of spacecraft sunrises and one made of spacecraft sunsets (Fig. 2). Each swath spans about  $3^\circ$  to  $10^\circ$  in latitude for high to low latitudes respectively, and  $\sim 360^\circ$  in longitude.

#### 3.2 Data Filtering

For the purpose of this work, the same basic methodology is applied to the same source data with the difference being how those data are pretreated. The process begins by extracting SAGE II version 7.0 ozone ( $O_3$ ) number density profiles for all events not flagged as "dropped" in the SAGE II inversion algorithm and for all altitudes above the reported tropopause. A modification of the Wang et al. (2002) filtering criteria is applied, which includes the following: exclusion of any profile where the  $O_3$  uncertainty exceeds 10% between 30 and 50 km, exclusion of all data below where the  $O_3$  uncertainty exceeds 200% below 30 km, and exclusion of any data where the  $O_3$  uncertainty exceeds 100% above 30 km. These filtering criteria also include aerosol filters to remove data within and below clouds (typically within the troposphere) and also to remove periods of heavy aerosol interference from the volcanic eruption of Mount Pinatubo. However, since this regression includes a volcanic term, data within the eruptive periods are not filtered out via aerosol criteria.

### 3.3 Data Binning

The binning of data is one of the primary differences between the two methods. The first method, hereafter referred to as MZM (monthly zonal mean), is simply to take all data within a latitude zone (in this case 10 degrees wide) and within a particular month and compute the mean value at each altitude. The time associated with this mean value is the center of the month and the latitude is the center of the zone. Different event types are not separated for these monthly zonal means. The second method, hereafter referred to as STS (simultaneous temporal and spatial), utilizes the data on a daily basis for each altitude. For each day, the events are separated into four subsets governed by the combination of local and spacecraft sunrises and sunsets (most of the time each day only contains two subsets of events). The mean of each subset is taken and the time associated with each mean is at the center of the day and the latitude is the mean of the latitudes of each event in the subset. The time of day is actually irrelevant here, since no diurnal model is used and each daily mean is already separated by event type, which will be treated differently as shown later.

### 3.4 Data Averaging

We can consider, for each subset of events that we are taking a mean, that we have a collection of  $N$  data points ( $Y$ ). The daily mean and standard deviation of the mean are simply the following:

$$\bar{Y} = \frac{1}{N} \sum_{i=1}^N Y_i \quad (1)$$

$$\sigma_{\bar{Y}} = \sqrt{\frac{1}{N(N-1)} \sum_{i=1}^N (Y_i - \bar{Y})^2} \quad (2)$$

Each data point also has some uncertainty value ( $\delta$ ) such that:

$$\delta_{\bar{Y}} = \frac{1}{N} \sqrt{\sum_{i=1}^N \delta_i^2} \quad (3)$$

So that, for each subset, the mean value  $\eta$  is simply  $\bar{Y}$  and the uncertainty in  $\eta$  is:

$$\delta_{\eta} = \sqrt{\sigma_{\bar{Y}}^2 + \delta_{\bar{Y}}^2} \quad (4)$$

- 5 In this way, the uncertainty in the daily or monthly mean is representative both of the uncertainty in the measurements as well as the overall variance. This is important, particularly at higher latitudes where it is possible one or two measurements of a subset dip into the vortex and produce abnormally low values that are not representative of the entire zonal band.

## 10 4 Regression Methodology

After filtering and binning the data for both the MZM and STS methods, the following functional form is regressed to the data:

$$\eta(\theta, t) = \sum_i \sum_j \beta_{i,j} \Theta_i(\theta) T_j(t) \quad (5)$$

- where  $\eta$  is the concentration of  $O_3$ ,  $\Theta(\theta)$  is the functional form of the latitude dependence,  $T(t)$  is the functional form of the temporal dependence, and  $\beta$  are the coefficients of the regression. The concept of a two-dimensional regression is similar to the work of Bodeker et al. (2013). The measure of time for the purposes of regression is year fraction (e.g., 1994.655). This regression is done separately for each altitude. For the MZM method,  $\Theta(\theta)$  is identically 1, since the data is regressed separately within each latitude zone. For the STS method,  $\Theta(\theta)$  is a series of seven Legendre polynomials in spherical harmonics (no

longitudinal dependence), which have the properties of a zero derivative at the poles and mutual orthogonality. The temporal dependence is simply the sum of the predictor variables and a constant. However, the STS method also includes a conditional temporal term based on the local event type. This accounts for any differences in the mean values between sunrise and sunset events based on diurnal variation in the data be it geophysical or algorithmic in origin.

A generalized least squares regression technique, outlined in Appendix B, is applied, which accounts for autocorrelation, heteroscedasticity, and data gaps. Due to the nature of the data, very careful consideration is given when calculating the autocorrelation coefficient  $\phi$  for the STS regression method. Only daily means with the same satellite event type can be correlated, and even then the temporal and spatial separation must be considered due to gaps in the data. We make the assumption that the level of autocorrelation does not change significantly over sufficiently small separations, because the amount of autocorrelation is related to geophysical variability that is not well modeled, which is nearly constant over sufficiently small time and spatial scales. Thus, any consecutive data points that are within 5 days and  $20^\circ$  in latitude are included in the calculation of  $\phi$ . No dependence upon the temporal or spatial separation was found within these limiting criteria and so a simple lag-1 autocorrelation was still considered. In this way, a set of pairs of points was created for each event type, which were then all fed into the calculation of  $\phi$  (Eqn. (B12)) to determine a single value for use in the autocorrelation correction. The autocorrelation correction is then iterated until  $\phi$  converges to within 0.05.

To account for the iterative correction of the heteroscedasticity, it was assumed that values of the adjustment vector  $k$  in Eq. (B13) had a simultaneous spatial and seasonal dependence (or purely seasonal for the MZM method). Regression was thus performed to a combination of the seasonal and spatial predictor variables. The fit to these data was used as the values of  $k$  and this process is iterated until  $k$  converges to values sufficiently near 1.

Residual filtering is performed using the deweighted, uncorrelated residuals ( $\epsilon$  in Eq. (B7), which have zero mean and unit variance) to remove outliers in the data. In order to create a

filtering criterion, the median absolute deviation (MAD) is computed (Muller, 2000). Values of  $\epsilon$  that deviate from the median by six MADs are omitted in future processing. Because residual filtering is performed on the deweighted, uncorrelated residuals, only data that both disagree with the model and are highly uncorrelated with any other data are omitted. This process is iterated and the number of additional data points omitted decreases rapidly after each iteration. Since the filtering is performed with the use of MADs instead of standard deviations, an iteration can converge to exclude no additional data points, though in practice only a few iterations are required.

When performing regressions to data of this type, a common problem can be the excess of predictor terms. To overcome this, one can use a priori knowledge of what terms are and are not significant, or one can use all terms and manually determine which terms are statistically insignificant and omit those terms. This can, however, prove to be a tedious process, particularly when performing the same regression repeatedly for multiple latitudes and altitudes and with potentially several hundred terms. We instead choose to automate this process. A predictor term could be considered statistically significant if it is statistically different from zero at the two-sigma ( $\sim 95\%$ ) level. In other words, if the ratio of the one-sigma uncertainty in a coefficient ( $\sigma_{\beta_{i,j}}$ ) to the coefficient itself ( $\beta_{i,j}$ ) is less than 0.5, it can be considered significant. After the residual filtering process is completed, an analysis of this ratio is done. During the first iteration, all coefficients with a ratio greater than 8 are omitted from future processing and the entire process is repeated (from the beginning). This threshold is iteratively lowered until only those coefficients that are statistically significant at the two-sigma level remain. In this way, all final coefficients are statistically significant at the two-sigma level. This does not, however, mean that all resulting terms are non-negligible. Though excluding predictor terms that are negligible (i.e., their contribution to the overall variance is minimal) is, in practice, unnecessary. Additionally, while each coefficient is statistically significant at the two-sigma level, groups of terms (e.g., piecewise linear slopes or EESC) can collectively become statistically insignificant depending upon their interactions.

## 5 Regression Quality and Results

### 5.1 Residual Analysis

A quick look at some examples of the fit (Fig. 3) shows that the algorithm works well, with the exception of potential overfitting in the regions of data gaps. However, as with any regression technique, great care must be taken in interpreting the results. The process is ultimately a numerical one, and just because the solution converges and yields a result, does not mean that result is accurate. As such, a proper investigation of the total residuals is still required to ensure that the data and model reasonably agree. If the data and model agree well, one would expect no systematic biases to emerge in the residuals. A quick look at the total and uncorrelated residuals (Fig. 4) shows this to be the case. The total and uncorrelated residuals show no systematic bias with respect to time or latitude. However, a clear pattern in the spread of the residuals can be seen as a function of latitude, though this is expected. The total residuals are a combination of the correlated and uncorrelated residuals. Correlated residuals (or those removed from the lag-1 autocorrelation correction) represent geophysical variability that is well sampled, but not well modeled. Uncorrelated residuals (or those that remain after the lag-1 autocorrelation correction) represent instrumental noise present in the data as well as geophysical variability that is not well sampled, which mathematically represent the heteroscedasticity in the data.

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. Since the mean of the residuals is nearly zero (as it should be), the spread of the residuals is used instead. To avoid the over-influence of outliers in the data, a weighted running mean of the absolute value of the residuals as a function of latitude at a particular altitude is taken. The result is shown in Fig. 5. Analysis of the uncorrelated residuals reveals the amount of uncertainty in the data that is being regressed (in this case, daily means). The uncorrelated residuals will increase in the presence of increased noise in the instrument data (e.g., at higher and lower altitudes) or in the vicinity of increased geophysical variability within a daily mean (e.g., in the tropics where each daily mean spans a greater range in latitude or at high

latitude in the local winter where measurements may dip into and out of the vortex), but the attribution of this uncertainty to each source separately cannot be determined. However, the contribution to the uncertainty from unresolved geophysical variability could be minimized by applying the regression model to the data at its native resolution. The difference between the total and uncorrelated residuals are the correlated residuals. These correlated residuals are a measure of the discrepancy between the model and the data. Namely, they are a result of the geophysical variability that is well sampled but not well modeled as well as any instrumental variability (e.g., biased meteorological or ephemeris input data). Residual analysis is useful, because applying the same model to different data sets could be used to independently assess the quality of the measurements via the uncorrelated residuals, as well as to ascertain deficiencies in the model or the data itself via analysis of the total residuals. A preliminary version of this technique was applied in Damadeo et al. (2013) to both the previous (v6.2) and current (v7.0) versions of the SAGE II data set to demonstrate and assess improvements made to the new version.

## 5.2 Predictor Coefficient Analysis

One of the problems with multiple linear regression is the issue of multicollinearity, or the possibility that two or more predictors are highly correlated. Multicollinearity, or even the presence of any correlation between predictors, does not affect the regression results as a whole (short of the possibility of poor inversion for numerical algorithms), but it does affect the interpretation of individual predictors. For example, if aerosol data were used (instead of just a volcanic proxy) that had an annual cycle in addition to the fitting of an annual cycle term, the annual cycle term would be biased because some of the annual variation in ozone would be attributable to the aerosol term. Fortunately, this effect is captured in the uncertainties in the predictor coefficients, but it still illustrates a problem when attempting to interpret single predictor coefficients. One could analyze the covariance matrix that results from the regression to determine the level of correlation between predictors, but nonetheless, care should be taken when interpreting results.



Due to these possible shortcomings, the analysis of any single predictor requires the analysis of all of the predictors in order to ensure they are reasonable. The annual cycle is fairly trivial with the exception of some overfitting in regions of missing data (Fig. 3) and ENSO lacks any substantial contribution above 20 km (not shown). Since long-term trends are discussed later, this section will take a brief look at the remaining influential terms: QBO, solar cycle, and volcanic. It has been shown that SAGE II data quality is best between 20 and 50 km (Damadeo et al., 2013) and there are fewer gaps in sampling below 60° in latitude. As such, the following analysis will focus on this region.

### 5.2.1 QBO

After the annual cycle, the QBO is the largest source of variation in ozone. Figure 6 shows the amplitude of the response of ozone to the QBO as a function of latitude and altitude as a percentage of the local mean value (i.e., the mean value is a function of latitude and altitude). The amplitude is computed as the root mean square amplitude multiplied by  $\sqrt{2}$ , and is analogous to half of the peak-to-peak amplitude for a sine wave. As can be seen, the influence of the QBO is largest in the tropics in the lower and middle stratosphere as well as in the middle stratosphere at mid-latitudes. Figure 7 shows examples of the QBO term at the equator as a function of time and altitude and at 23 km as a function of time and latitude. The altitude dependent and latitude dependent phase lags are easily noticeable in the two figures. However, it should be noted that some deficiencies still remain due to the fact that the QBO proxy term originated from data at the equator. The frequencies at altitudes above where the proxy is available do not change and the frequencies at mid-latitudes are slightly different only because of the inclusion of a cross-term with the annual cycle. It should also be noted that the amplitude of the QBO is larger around the time of the Pinatubo volcanic eruption, which may be a physical response of the QBO itself to the Pinatubo eruption (Thomas et al., 2009) or it may simply be a byproduct of correlation with the volcanic term. Additionally, the fact that the regression is both temporal and spatial means that the inability to accurately model the QBO at higher latitudes will detract from the ability to accurately model the QBO at lower latitudes.

## 5.2.2 Solar

To include a response of ozone to the solar cycle, the regression model can include either one or two solar predictor variables. The biggest differences between one or two terms are seen in the tropics between 25 and 35 km. The amplitudes of the oscillation are similar, as shown in Fig. 8, with the exception of a very weak oscillation in this region if only a single term is used. This is because, when two terms are used, the solar term, if allowed to change phase, exhibits strong correlations with the volcanic term in this region as shown in Fig. 9. Here, the solar cycle is shifted later in phase by about two years to coincide with the peak of volcanic increase surrounding the Pinatubo eruption, which is similar to results shown in Remsberg (2014).

The inclusion of a volcanic term reduces the overall residuals regardless of whether one or two solar terms are included (not shown), but there is a negligible difference in residuals and resulting trends between the use of one or two solar terms if a volcanic term is also used. It is unclear if the response of ozone to the solar cycle really does lag by about two years in the mid-stratosphere in the tropics or if the algorithm is simply trying to attribute some of the response of ozone to aerosol to the solar cycle instead (Solomon et al., 1996).

## 5.2.3 Volcanic

The results of the volcanic term need to be interpreted very carefully. On one hand it is clear from the data that ozone responds to changes in aerosol, particularly after Pinatubo. On the other hand, the SAGE II inversion algorithm can produce biases in ozone in the presence of high aerosol loading (Wang et al., 2002) and so some of the response to aerosol, particularly at lower altitudes in the tropics, can have algorithmic rather than physical origins. However, omitting data based on aerosol extinction (e.g., Wang et al., 2002) and assuming that the influence of aerosol has been removed would be incorrect.

A look at the response of ozone near the Pinatubo eruption reveals both physical and algorithmic responses as well as regressive responses. Figure 10 shows the peak of the volcanic term in the few years after the Pinatubo eruption as a percent of the mean. Ozone

shows a positive response above 28 km with a corresponding negative response just below that in the tropics, which are similar to results in Bodeker et al. (2013). These responses can be the result of local (i.e., chemical) effects of aerosol, radiative (e.g., thermal and/or photochemistry) effects of aerosol at other altitudes, or algorithmic responses of ozone to aerosol retrievals. The anomalously large responses at low altitudes are the result of overfitting to data gaps (e.g., Fig. 3 top). However, given the results from the QBO and solar terms, some correlation between these terms exists. Regardless of these correlations, however, it is clear that ozone does respond to changes in aerosol in SAGE II data and that the use of a volcanic term in these regressions is necessary.

## 6 SAGE II Sampling and Biases

As previously mentioned, SAGE II took  $\sim 30$  observations per day in two ground-track swaths that each span  $3^\circ$  to  $10^\circ$  in latitude and  $\sim 360^\circ$  in longitude (Fig. 2). This sparse sampling caused SAGE II measurements at a particular latitude to occur at roughly the same times of the year, resulting in full seasonal coverage at mid-latitudes, and restricted (or sparser) seasonal coverage at high (or low) latitudes. Figure 11 shows SAGE II sampling at both the beginning and end of the mission. Sampling is sparser at the end of the mission due to problems with the azimuth pointing system forcing the instrument to operate at 50% duty cycle starting in late 2000. The increased spread during the later period is a result of an increased rate of precession of the orbit. This demonstrates a form of potential bias due to sampling present throughout the mission, though more pronounced in the later period, where the orbit crosses a particular latitude but at progressively earlier times each successive year. If the sampling were constant over the lifetime of the instrument, it would only result in biases in the MZM seasonal cycle. However, because the sampling drifts over time, this bias also aliases into the MZM long-term trend.

Given the nature of this sampling, another potential problem could clearly arise if any difference existed between the mean values of sunrise and sunset events. Figure 12 illustrates the differences in the means of local sunrise and sunset event types. These differences can

be the result of geophysical variability and/or algorithmic biases (Damadeo et al., 2013). Regardless of the source, however, these differences are present in the data and must be accounted for in the regression.

Due to this non-uniform sampling, every monthly zonal mean value computed for ozone is biased. The primary reason is that the data sampled within a given month and zonal band has a mean sampling time and place that is not at the exact center of the month and zone. The true spatial and temporal center of each monthly zonal band can be computed and values can be extracted from the results of the STS regression method. These values can then be compared to the results of the MZM regression method to produce Fig. 13. Figure 13 shows the spatial and temporal dependence of the bias (i.e., how the MZM method is biased compared to the STS method) at two altitudes. Since the MZM method does not differentiate between event types, the bias is computed against the STS method for each type, showing how the MZM method is biased against each type, but only for where data of that type exist. As can be seen, biases in individual monthly zonal bands exist as large as 10% due to non-uniform spatial and temporal sampling, and large systematic biases exist at higher altitudes due to differences between event types. Large gaps in sampling that are asymmetric in event type (both in location and bias) are seen later in the mission, illustrating the problem with not accounting for the differences in sampling and event type in the regression.

## 7 Long-term Trends

The primary focus of time-series analysis of long-term ozone data sets is typically the long-term trend of ozone. Most often this has been done using two piecewise linear trend terms joined at some predetermined time. The regression is performed four different ways utilizing the combinations of the MZM and STS regression methods and two piecewise linear trend terms or two orthogonal EESC trend-like terms. The resulting mean trends are computed both for the traditional decrease in ozone (in this case between 1985 and 1995) and for the

traditional increase in ozone (in this case between 1998 and 2005) for all four analyses. The results for the earlier period are shown in Fig. 14 and the later period in Fig. 15.

At first glance, the four plots in Fig. 14 seem very similar. Each plot shows regions of significant decreases in ozone between 35 and 50 km at middle to high latitudes, as well as some slight positive trends in the tropics at lower altitudes. However, there are some important discrepancies to point out. The MZM method, regardless of which pair of trend terms is used, shows a slight positive trend in the tropics between 30 and 35 km, which is consistent with other studies (e.g., Randel and Wu, 2007; Kyrölä et al., 2013; Bourassa et al., 2014), though those studies show this increase to be statistically insignificant. However, the use of the STS method removes this feature, regardless of which trend terms are used. In addition, the magnitude of the trends, when using the same trend terms, is biased slightly negative for the MZM method than for the STS method. This is a result of the biases from non-uniform sampling, and is explained in more detail in the next paragraph.

The results shown in Fig. 15 are very different. Whereas the results of the MZM method are consistent with other studies (e.g., Kyrölä et al., 2013; Bourassa et al., 2014, which make use of multiple data sets extending to 2013), showing regions of large ozone recovery in the southern hemisphere and smaller recovery in the northern hemisphere, the results from the STS method show a significant increase only in the northern hemisphere, which is slightly smaller than in the MZM method. To understand this difference, one should take another look at Fig. 12 and Fig. 13. In the southern hemisphere at mid-latitudes before the pointing problems in late 2000, the concentration of sampling shows a roughly equal mix of sunrise and sunset events. Given the difference in ozone between sunrise and sunset events in this region at higher altitudes, the mean of the data is expected to be somewhere in the middle of the sunrise and sunset mean. However, later in the period, there is a significant decrease in sunrise events in this region, which results in the mean of the data skewing more towards the sunset mean. With the beginning of the potential recovery period starting at the overall mean, and the end of the calculable recovery period residing at the sunset mean, the computed trend is artificially biased high. Proper treatment of the differences

between sunrise and sunset events accounts for this effect and results in smaller recovery trends in the STS method.

It is clear that the differences caused by the SAGE II non-uniform sampling are important, and that the STS method is preferable to the MZM method regardless of which pair of long-term trend terms is used. However, there are still some differences in trends between the two pairs of trend terms as shown in Figs. 14 and 15. To understand this, one needs to look at the time evolution of the long-term trend for the use of each pair of trend terms. Figure 16 illustrates the difference between the two pairs of trend terms at 50°N. The piecewise linear trend terms force any turnaround in ozone to occur at 1997, while the use of the two orthogonal EESC terms allows this turnaround point to move in time. As shown in Fig. 16 (top right), the turnaround time is earlier at lower altitudes. This is consistent with the fact that stratospheric ozone is inversely related to stratospheric chlorine, and the EESC proxies from Newman et al. (2007) show that the EESC peaks earlier for smaller mean ages of air. Figure 16 would suggest then, that the mean age of air in the northern hemisphere decreases with decreasing altitude, which is consistent with results shown in Waugh and Hall (2002).

The study outlined in Waugh and Hall (2002) is performed only for the northern hemisphere and the assumption is made that the hemispheres are symmetric. However, the time evolution of the long-term trend at high southern latitudes (Fig. 16, bottom right) shows no clear change in turnaround time with altitude, and in some cases never turns around (i.e., is always decreasing). It is, at present, unclear if this hemispheric asymmetry is geophysical or a result of correlation between the long-term trend and other terms (e.g., solar or volcanic).

## 8 Conclusions and Future Work

A new method for performing time-series analysis of sparsely sampled data, in particular SAGE II, has been presented. The differences between the MZM method and the STS method have been discussed and the impacts on the long-term trends in ozone detailed. It

has been shown that the non-uniform sampling in SAGE II data will produce biased long-term trend values in ozone if not properly accounted for. The STS method shows declines in ozone that are similar to those from other studies in the upper stratosphere at middle to high latitudes but very different results for the presumed recovery period, namely a noticeable reduction in the magnitude of ozone increase in the southern hemisphere. The use of two orthogonal EESC predictor variables instead of a piecewise linear trend allows for a variable turnaround time in ozone due to differing mean ages of air. Results show a hemispheric asymmetry in the middle to upper stratosphere, with an earlier turnaround time with lower altitude and latitude in the northern hemisphere but no coherent pattern in the southern hemisphere. It has also been shown that the STS method can be used to assess the quality of a data set's measurements independent of other data sets. In addition to ozone, the STS method was applied to SAGE II aerosol optical depth data to create a new volcanic proxy that covers the SAGE II mission period.

For future work, we would like to extend this technique to other ozone data sets as well as to include multiple data sets to better constrain the long-term trends in the presumed recovery period. The benefit of this technique for the creation of a single time-series derived from multiple data sets is that it does not require the homogenization of the different data sets prior to regression. Instead, instrument-dependent conditional terms representing mean offsets, different diurnal variation, time-dependent drifts, or other terms could be included as necessary. Another consideration is to expand upon the creation of a volcanic proxy term to one that is altitude dependent, so that the response of ozone to both local (i.e., at the same altitude) and total aerosol can be assessed. Lastly, it could be beneficial to experiment with other coordinate systems in order to reduce uncertainties in regions of larger variance. For example, regression could be done on the data at its native resolution, using models for diurnal and longitudinal variation, as this would reduce some of the variance in the tropics where each daily mean spans  $\sim 10^\circ$  in latitude as opposed to  $\sim 3^\circ$  in latitude at high latitudes. Another coordinate transformation would be to perform this regression methodology on equivalent latitude instead of latitude, as it would remove much of the variance at high latitude where observations constantly dip into and out of the polar vortex.

## Appendix A

The creation of a volcanic proxy term is achieved via the simultaneous temporal and spatial regression to SAGE II aerosol data. This is done using the same STS regression technique as is done for ozone. The process begins by extracting 1020 nm aerosol extinction coefficient profiles. During periods of high aerosol loading, SAGE II profile retrievals stopped at altitudes well within the stratosphere. To compensate, data below retrieval termination are filled in via the process outlined in SPARC (2006, Chapter 4.3.1). Each profile is then integrated from the top (40 km) down to 3 km above the reported tropopause. These event specific stratospheric aerosol optical depth values are then compiled into daily means, except that no distinction is made between local event types as no significant diurnal cycle in aerosol is seen. The regression uses the same latitudinal dependence (albeit with 11 terms) and a temporal dependence that includes annual, QBO, and eruptive terms for each significant volcano during the SAGE II mission. The iterative regression technique outlined in this paper is applied, though the data that is regressed is the logarithm of optical depth. The logarithm is used instead of the raw data because many physical effects, such as the annual cycle or QBO, are inherently multiplicative effects (i.e., their magnitudes are related to the magnitude of the instantaneous mean). The same is also true of ozone, but ozone does not vary by several orders of magnitude over time at a particular altitude. The primary reason for this appendix, however, is the difficulty regarding the creation of the eruptive terms for each volcano used as predictor variables in the regression.

The creation of a model term for use with linear regression that accurately represents changes in aerosol as a result of a volcanic eruption is not trivial. At any given location after a major volcanic eruption, changes in aerosol are characterized by a delay after the eruption (i.e., time it takes for aerosol to reach that location from the eruption), followed by an increase in aerosol up to some peak value over time, followed lastly by a long decay back to background levels (unless another eruption occurs). It makes sense to create a piecewise function to model this rise and fall, but the choice of these functions is important. Previous attempts (e.g., SPARC, 2006, Chapter 5.4.2) use a simple polynomial from eruption to peak



values followed by an exponential decay with some characteristic decay constant. However, a simple exponential decay model would assume that the data, when plotted in log-space, is linear, which it clearly is not (see Fig. 17).

While analyzing the logarithm of the aerosol data, we choose a piecewise pair of second order polynomials in order to fit the eruptive effects on aerosol. However, the two functions are restricted to maintain continuity of both the functions and their derivatives where they join, as well as to assume the eruptive term returns to zero (i.e., background) after some amount of time has elapsed. In this way, a pair of piecewise second order polynomials can be defined by the declaration of three parameters: time of injection ( $t_I$  or time aerosol first arrives at a location after an eruption), peak time ( $t_P$  or time after injection at which aerosol values are at their peak), and return time ( $t_R$  or time after injection at which the eruptive term and its derivative return to zero). The time at which these two functions join is also constrained by these three parameters.

The downside to this methodology is that the functional form of the eruptive term is not linearly dependent upon the parameter times ( $t_I$ ,  $t_P$ , and  $t_R$ ). Additionally, the times themselves are functions of latitude and different for each eruption. Since we have no intrinsic knowledge of the value of these times, or their spatial dependence, a non-linear least squares fitting technique is applied to binned data. The process begins with data taken in a  $10^\circ$  wide bin in latitude centered at a particular latitude. Initial guesses are made for the three parameters for each of seven volcanic eruptions: El Chichon (1982), Nevado del Ruiz (1985), Kelut (1990), Pinatubo (1991), Cerro Hudson (1991), Ruang (2002), and Manam (2004). The MPFIT algorithm (Markwardt, 2009) is used, as it allows for restrictions on solvable parameters to be placed, which greatly aids in convergence. Too few data are available to constrain  $t_I$  or  $t_P$  for El Chichon, so these parameters are tied to Pinatubo. Likewise,  $t_R$  for Manam was set constant at 5 years. Some examples of the non-linear regressions can be seen in Fig. 17.

This non-linear regression was performed for each latitude between  $70^\circ\text{S}$  and  $70^\circ\text{N}$  in increments of  $1^\circ$ . The parameters ( $t_I$ ,  $t_P$ , and  $t_R$ ) were then smoothed, and any iterations that did not converge properly were ignored. To create the eruptive terms for the STS re-

gression, the parameters were interpolated (or held constant at last value for extrapolation) to all latitudes to create functional forms for each volcano for the entire record. These eruptive terms are then easily linearly regressed to, where the STS regression is allowed to determine spatial dependence but not temporal variation for each volcano as a function of latitude. The final product to be fed into the regression for ozone is a single volcanic term that represents the eruptive changes in aerosol at all times during the record for all latitudes (Fig. 18).

Ultimately the creation of a volcanic proxy is an empirical result. Theoretically, one could look at the parameters from the non-linear least squares regression individually, but most of the terms would have no real meaning. For example, there are a large number of volcanic eruptions around the time of the eruption of Nevado del Ruiz. While the parameters near the eruption make sense, the parameters at higher latitudes merely represent the algorithm's attempt to fit the overall increase in aerosol from a multitude of eruptions in that time period (e.g.,  $t_P \sim 2$  years and  $t_R \sim 8$  years). The regression fits the overall data well, but each individual term is not necessarily representative of that eruption alone.

## Appendix B

The principle of multiple linear regression is predicated upon the simple assumption that a dependent variable ( $Y$ ) is linearly dependent upon a set of predictor variables ( $X$ ) that produce a simple equation of the following form:

$$Y = \mathbf{X}\beta + R, \quad (\text{B1})$$

where  $\mathbf{Y}$  is a vector of  $N$  data points (index  $i$ ),  $\beta$  is a vector of coefficients for  $M$  predictor variables that include a constant (index  $j$ ),  $\mathbf{X}$  is an  $N$  by  $M$  matrix of each predictor variable corresponding to each data point, and  $\mathbf{R}$  is a vector of  $N$  residual differences between the data and the fit. It should be noted that generally  $X_{i,j=0}$  is identically 1 for all  $i$  from 1 to  $N$  and  $\beta_{j=0}$  is simply the overall constant of the fit. These coefficients can be solved for using a simple ordinary least squares (OLS) regression technique, which can be found in any of

a number of textbooks related to statistics. In fact, the methods outlined in this appendix derive from Kutner et al. (2005). The uncertainties in the coefficients ( $\sigma_\beta$ ) can also easily be solved for provided the following assumption holds:

$$\text{var}(\mathbf{Y}) = \sigma_0^2 \mathbf{I}, \quad (\text{B2})$$

- 5 where  $\text{var}(\mathbf{Y})$  denotes the variance-covariance matrix of  $\mathbf{Y}$ ,  $\sigma_0$  is some constant, and  $\mathbf{I}$  is the identity matrix. If this assumption holds, then the residuals have the property of being Gaussian with a constant conditional variance of  $\sigma_0^2$ .

If the assumption of Eq. (B2) does not hold, then Eq. (B2) reverts to a general form:

$$\text{var}(\mathbf{Y}) = \Sigma_0. \quad (\text{B3})$$

- 10 This produces coefficients that are still unbiased, but the estimates of their uncertainties are biased small. To overcome this problem, we turn to generalized least squares (GLS), which applies a transformation matrix  $\mathbf{G}$  to Eq. (B1) to obtain

$$\mathbf{Y}^* = \mathbf{X}^* \beta + \mathbf{R}^*. \quad (\text{B4})$$

- 15 If  $\mathbf{G} = \sigma_0 \Sigma_0^{-\frac{1}{2}}$  (where  $\Sigma_0^{-\frac{1}{2}} \Sigma_0^{-\frac{1}{2}'} = \Sigma_0^{-1}$ ), then  $\mathbf{R}^*$  has the property of being Gaussian with conditional variance  $\sigma_0^2$ . The coefficients and their respective uncertainties can be computed in the following way:

$$\beta = (\mathbf{X}' \Sigma_0^{-1} \mathbf{X})^{-1} \mathbf{X}' \Sigma_0^{-1} \mathbf{Y} = (\mathbf{X}^{*'} \mathbf{X}^*)^{-1} \mathbf{X}^{*'} \mathbf{Y}^* \quad (\text{B5})$$

$$\text{var}(\beta) = (\mathbf{X}' \Sigma_0^{-1} \mathbf{X})^{-1} = \sigma_0^2 (\mathbf{X}^{*'} \mathbf{X}^*)^{-1}. \quad (\text{B6})$$

- 20 It follows that  $\sigma_{\beta_j}$  is the square-root of the  $j^{\text{th}}$  diagonal element of Eq. (B6). It is worth noting that, when solved explicitly using  $\Sigma_0$ , the values of the coefficients and their uncertainties

are invariant to the value of  $\sigma_0$ , but when a transformation of variables is used, the equations revert to the form of solutions from OLS regression.

In time series analysis it is often the case that the assumption of Eq. (B2) does not hold. The residuals are, in fact, both heteroscedastic (have a non-constant variance) and serially correlated (temporally autocorrelated). If we assume the residuals have the following properties:

$$R_i = \phi R_{i-1} + \sigma_i \epsilon_i, \quad (\text{B7})$$

where  $\phi$  is the autocorrelation coefficient,  $R_i$  are the total residuals,  $\phi R_{i-1}$  are the correlated residuals,  $\sigma_i \epsilon_i$  are the uncorrelated residuals, and  $\epsilon$  is Gaussian with unit conditional variance, then Eq. (B2) reverts to Eq. (B3), where  $\Sigma_0$  is an  $N$  by  $N$  symmetric matrix with components of the following form:

$$\Sigma_{0i,j} = \frac{\sigma_i \sigma_j \phi^{|i-j|}}{1 - \phi^2}, \quad (\text{B8})$$

where, for this case,  $j$  goes from 1 to  $N$ . Computing  $\mathbf{G}$  leads to the following transformation of variables:

$$Y_i^* = \frac{Y_i - \phi Y_{i-1}}{\sigma_i},$$

$$X_{i,j}^* = \frac{X_{i,j} - \phi X_{i-1,j}}{\sigma_i}, \quad (\text{B9})$$

$$R_i^* = \frac{R_i - \phi R_{i-1}}{\sigma_i} = \epsilon_i.$$

This is just the Cochrane-Orcutt transformation (Cochrane and Orcutt, 1949), which ignores the first data point. The Prais-Winsten transformation (Prais and Winsten, 1954) can be used to include the first data point and an additional modification outlined in Savin and

White (1978) can be used to account for data gaps. It should be noted that, when performing OLS regression to the transformed variables, it will be necessary to force regression about the origin if using a packaged algorithm that performs regression and always assumes a constant exists in the regression.

- 5 If the heteroscedasticity and autocorrelation are known precisely and everything about the regression is perfect, then theoretically  $\sigma_0 = 1$ . In reality this is never the case. A good estimate of  $\sigma_0$ , however, can be obtained from the weighted mean-square error (also known as the reduced, weighted chi-squared error statistic):

$$\sigma_0^2 = \frac{\epsilon' \epsilon}{N - M} \quad (\text{B10})$$

- 10 It is important to compute  $\sigma_0$  so as to not underestimate the uncertainties in the coefficients in Eq. (B6) when regressing to transformed variables.

- In theory, one would want to know a priori what the values for  $\phi$  and  $\sigma$  are. Instead, these parameters are solved for iteratively towards convergence. The value for the autocorrelation coefficient is solved for by first performing OLS regression and then computing  $\phi$  in the usual manner:

$$\phi = \frac{\sum_{i=1}^{N-1} (R_i - \bar{R})(R_{i+1} - \bar{R})}{\sum_{i=1}^N (R_i - \bar{R})^2}, \quad (\text{B11})$$

which is itself a simple modification of the Pearson product-moment correlation coefficient:

$$\phi(X, Y) = \frac{\sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^N (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^N (Y_i - \bar{Y})^2}} \quad (\text{B12})$$

As can be seen,  $X$  and  $Y$  have been substituted with adjacent values of the residuals as well as a slight modification of the limits of summation to account for the number of pairs versus the number of total points.

The nature of the heteroscedasticity can be slightly more complicated. In practice, one only has an estimate for the heteroscedasticity ( $\delta$ ) such that:

$$\sigma_i = \delta_i k_i. \quad (\text{B13})$$

If the initial guess of the heteroscedasticity is correct, then  $k$  is identically 1. However, generally  $k$  is more complex, having a dependence on the predictor variables themselves. A practical way to solve for  $k$  is to first assume that  $\sigma = \delta$  and solve for  $\epsilon$ . If  $k = 1$ , then the mean value of  $\epsilon^2$  should also be 1. As such, one can regress a function  $f = \log(\epsilon^2)$  to predictor variables and obtain a fit value ( $f_i$ ) for each  $\epsilon_i$ , then  $k_i = \sqrt{e^{f_i}}$ . In this way,  $\sigma$  can be iteratively updated until  $k$  converges towards 1. However, the choice of predictor variable dependence of  $k$  may or may not be straightforward.

From a practical standpoint, this regression methodology is applied by first performing the regression (Eq. (B4)) with the assumption that there is no autocorrelation ( $\phi = 0$ ). The resulting residuals are used to compute the autocorrelation coefficient (Eq. (B12)) and the regression is repeated. The heteroscedasticity correction (Eq. (B13)) can then be applied. This process of applying the GLS regression, applying the heteroscedasticity correction, and recomputing  $\phi$  can be iterated towards convergence of  $\phi$ . Any residual filtering to be performed would require iteration of everything performed thus far. If filtering of regression coefficients is desired, it too would require an additional level of iteration of all steps performed thus far (including residual filtering).

*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.

## References

- 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, doi:10.5194/essd-5-31-2013, <http://www.earth-syst-sci-data.net/5/31/2013/>, 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, doi:10.5194/acp-14-6983-2014, <http://www.atmos-chem-phys.net/14/6983/2014/>, 2014.
- Cochrane, D. and Orcutt, G.: Application of least squares regression to relationships containing auto-correlated error terms, *J. Am. Stat. Assoc.*, 44, 32–61, doi:10.1080/01621459.1949.10483290, 1949.
- 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, doi:10.5194/amt-6-3539-2013, <http://www.atmos-meas-tech.net/6/3539/2013/>, 2013.
- Kutner, M. H., Nachtsheim, C. J., Neter, J., and Li, W.: *Applied Linear Statistical Models*, McGraw-Hill/Irwin, 5 edn., ISBN 0-07-238688-6, 2005.
- Kyrölä, E., Laine, M., Sofieva, V., Tamminen, J., Päiväranta, S.-M., Tukiainen, S., Zawodny, J., and Thomason, L.: Combined SAGE II–GOMOS ozone profile data set for 1984–2011 and trend analysis of the vertical distribution of ozone, *Atmos. Chem. and Phys.*, 13, 10 645–10 658, doi:10.5194/acp-13-10645-2013, <http://www.atmos-chem-phys.net/13/10645/2013/>, 2013.
- Markwardt, C. B.: Non-Linear Least Squares Fitting in IDL with MPFIT, in: *Astronomical Data Analysis Software and Systems XVIII*, vol. 411, pp. 251–254, <http://purl.com/net/mpfit>, 2009.
- Muller, J. W.: Possible Advantages of a Robust Evaluation of Comparisons, *J. Res. Natl. Inst. Stan.*, 105, 551–555, <https://archive.org/details/jresv105n4p551>, 2000.
- Newman, P. A., Daniel, J. S., Waugh, D. W., and Nash, E. R.: A new formulation of equivalent effective stratospheric chlorine (EESC), *Atmos. Chem. and Phys.*, 7, 4537–4552, doi:10.5194/acp-7-4537-2007, <http://www.atmos-chem-phys.net/7/4537/2007/>, 2007.
- Prais, S. J. and Winsten, C. B.: *Trend estimators and serial correlation*, Chicago: Cowles Commission, <http://cowles.econ.yale.edu/P/ccdp/st/s-0383.pdf>, 1954.

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, D06313, doi:10.1029/2006JD007339, <http://dx.doi.org/10.1029/2006JD007339>, 2007.

5 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, doi:10.5194/acp-10-11779-2010, <http://www.atmos-chem-phys.net/10/11779/2010/>, 2010.

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, doi:10.5194/acp-14-1039-2014, <http://www.atmos-chem-phys.net/14/1039/2014/>, 2014.

10 Savin, N. E. and White, K. J.: Testing for autocorrelation with missing observations, *Econometrica*, 46, 59–67, <http://www.jstor.org/stable/1913645>, 1978.

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, doi:10.1029/95JD03353, <http://dx.doi.org/10.1029/95JD03353>, 1996.

15 SPARC: Assessment of Stratospheric Aerosol Properties (ASAP), SPARC Report No. 4, WCRP-124, WMO/TD-No. 1295, <http://www.sparc-climate.org/publications/sparc-reports/sparc-report-no4/>, 2006.

20 SPARC: Chemistry-Climate Model Validation, SPARC Report No. 5, WCRP-30, WMO/TD-No. 40, <http://www.sparc-climate.org/publications/sparc-reports/sparc-report-no5/>, 2010.

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, doi:10.1175/JAS3650.1, 2006.

25 Thomas, M. A., Giorgetta, M. A., Timmreck, C., Graf, H.-F., and Stenchikov, G.: Simulation of the climate impact of Mt. Pinatubo eruption using ECHAM5 – Part 2: Sensitivity to the phase of the QBO and ENSO, *Atmos. Chem. and Phys.*, 9, 3001–3009, doi:10.5194/acp-9-3001-2009, <http://www.atmos-chem-phys.net/9/3001/2009/>, 2009.

Tung, K. K. and Yang, H.: Global QBO in Circulation and Ozone. Part I: Reexamination of Observational Evidence, *J. Atmos. Sci.*, 51, 2699–2707, 1994.

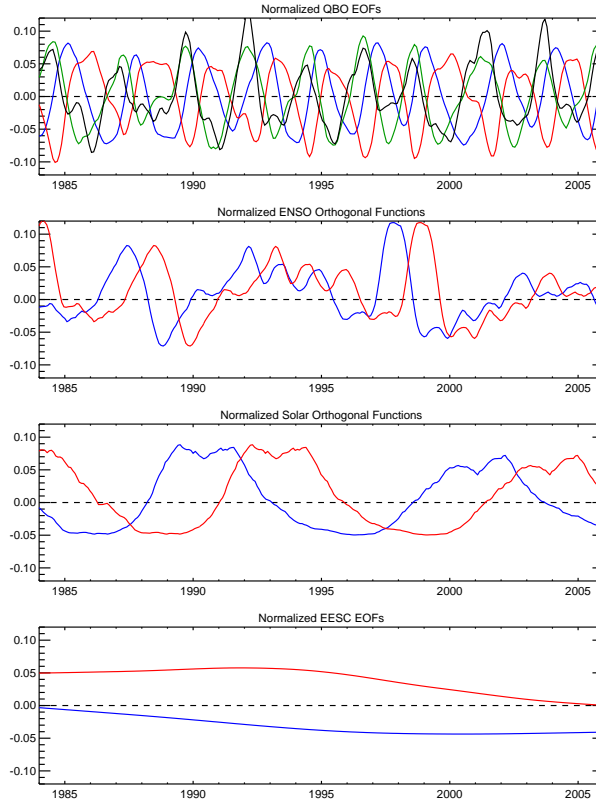
30 Wang, H. J., Cunnold, D. M., Thomason, L. W., Zawodny, J. M., and Bodeker, G. E.: Assessment of SAGE version 6.1 ozone data quality, *J. Geophys. Res.: Atmospheres*, 107, ACH 8–1–ACH 8–18, doi:10.1029/2002JD002418, <http://dx.doi.org/10.1029/2002JD002418>, 2002.



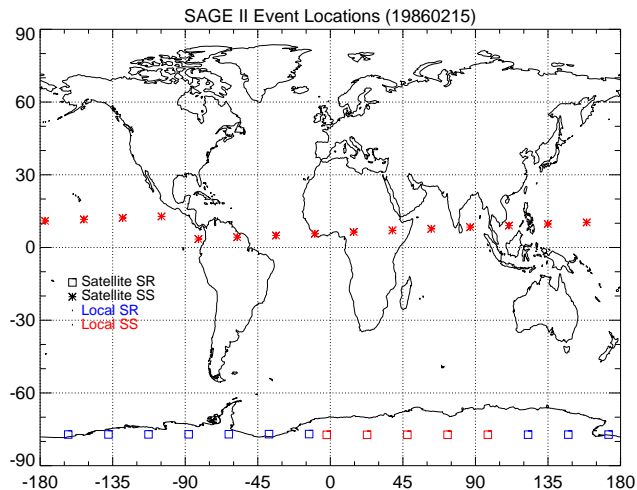
Waugh, D. and Hall, T.: Age of stratospheric air: Theory, observations, and models, Rev. Geophys., 40, 1–1–1–26, doi:10.1029/2000RG000101, <http://dx.doi.org/10.1029/2000RG000101>, 2002.

WMO: Report of the International Ozone Trends Panel 1988, Global Ozone Research and Monitoring Project-Report No. 18, World Meteorological Organization, Geneva, Switzerland, <http://www.esrl.noaa.gov/csd/assessments/ozone/1988/report.html>, 1988.

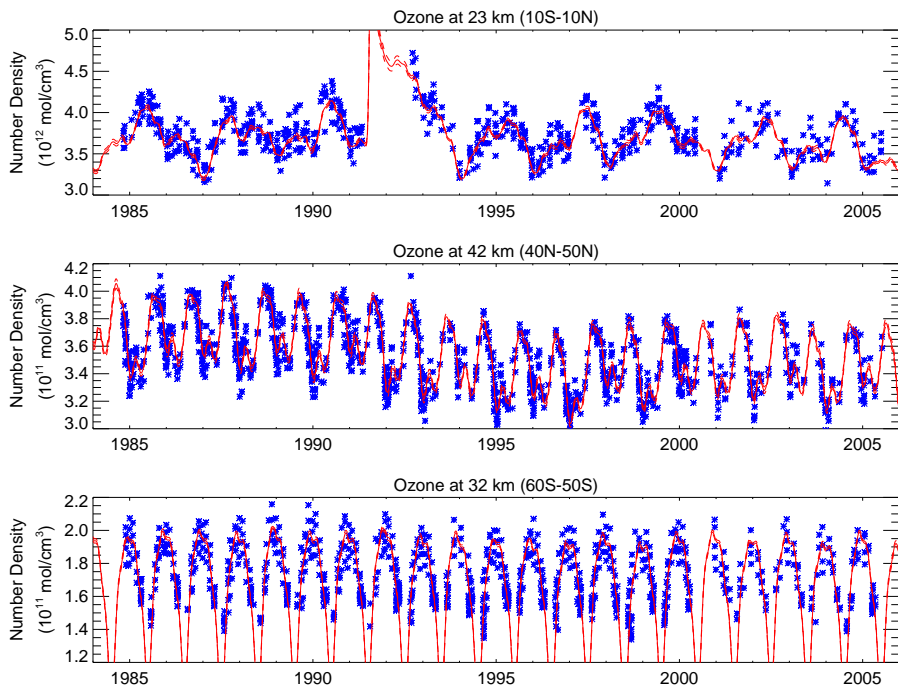
WMO: Scientific Assessment of Ozone Depletion: 2010, Global Ozone Research and Monitoring Project-Report No. 52, World Meteorological Organization, Geneva, Switzerland, <http://www.esrl.noaa.gov/csd/assessments/ozone/2010/report.html>, 2011.



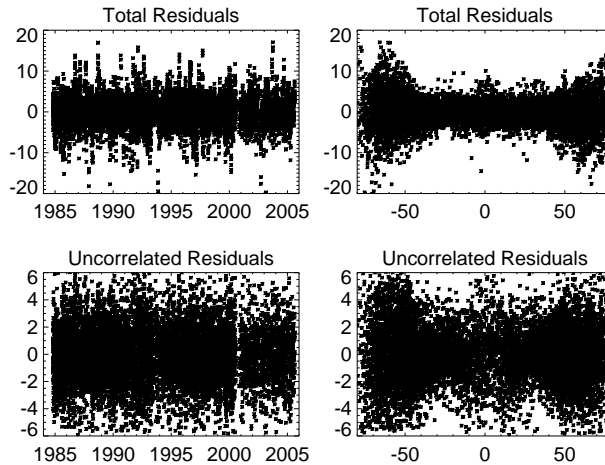
**Figure 1.** Orthogonal functions used for the regression. There are four QBO EOFs (blue [1], red [2], green [3], and black [4]), two time-shifted ENSO orthogonal functions (blue [1] and red [2]), two time-shifted solar orthogonal functions (blue [1] and red [2]), and two EESC EOFs (blue [1] and red [2]).



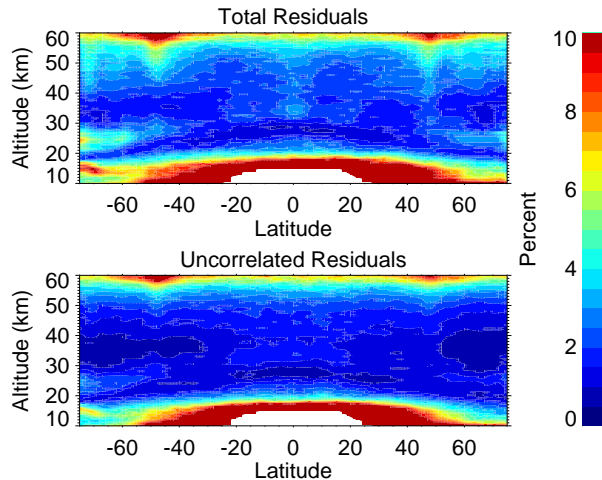
**Figure 2.** Locations of SAGE II events for a single day. Ground-track swaths are separated by satellite event type. While local event types are typically uniform within a swath, they can occasionally be different at high latitudes.



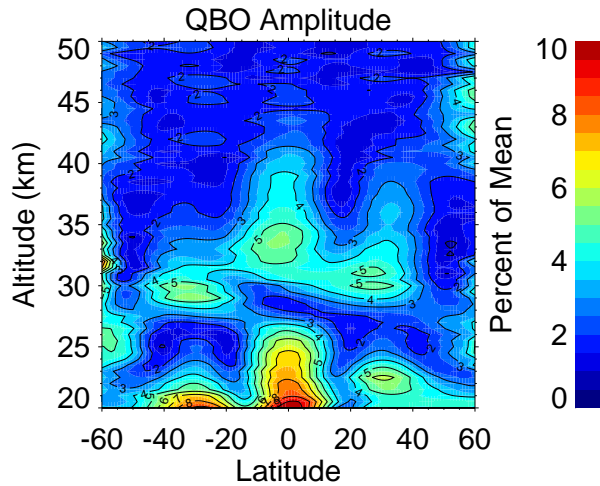
**Figure 3.** Some examples of the STS regression. Data within the stated latitude bin are shown in blue while a fit at the center of the bin is shown in red. The data shown have autocorrelation and diurnal variation removed for the purposes of plotting.



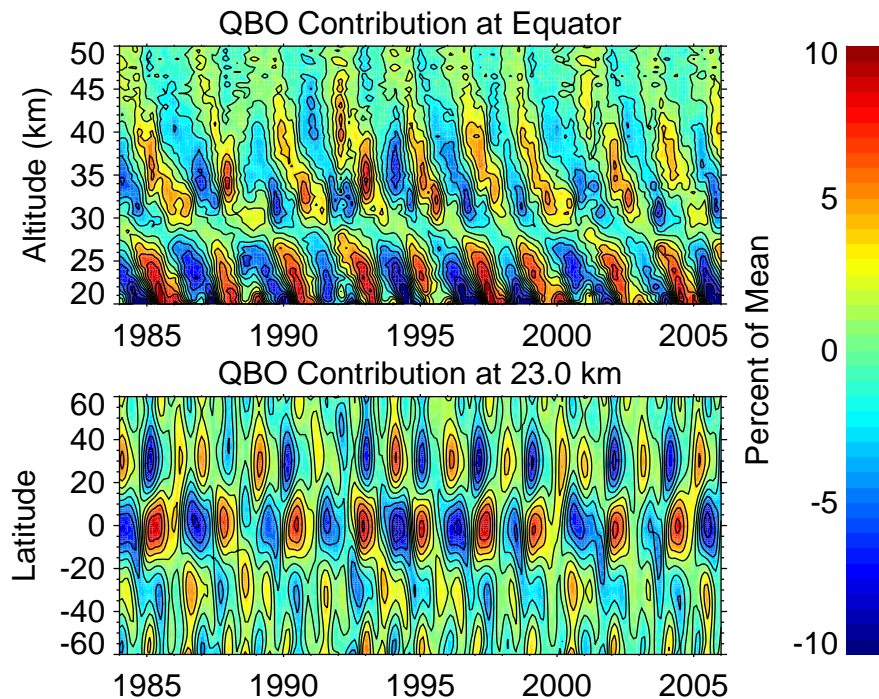
**Figure 4.** Total and uncorrelated residuals (as percents) as a function of both time and latitude at 25 km from the STS regression (see Eq. (B7)). Results are similar at other altitudes, though scales may change.



**Figure 5.** Spread of the total and uncorrelated residuals as a function of latitude and altitude from the STS regression. White regions show areas where insufficient data exist, generally due to being in the troposphere or early profile termination during retrievals.

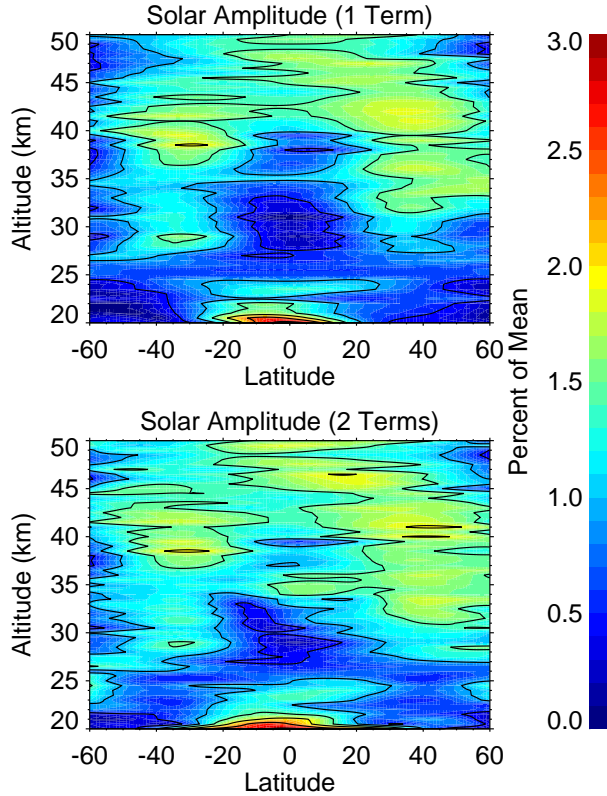


**Figure 6.** Amplitude of oscillation of the QBO term as a percentage of the local mean.

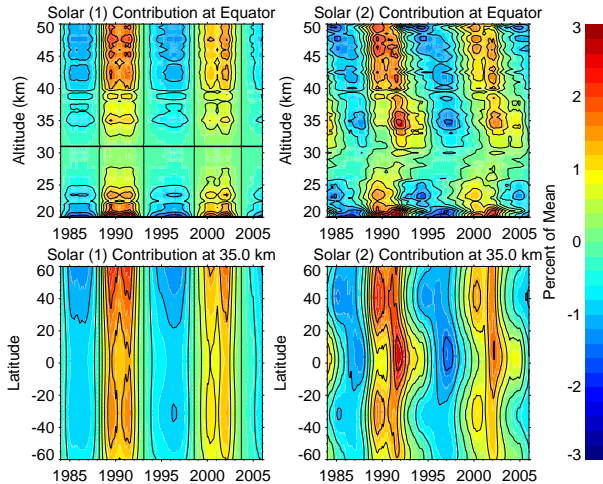


**Figure 7.** Examples of the QBO term. Contour lines are plotted at intervals of 2%.

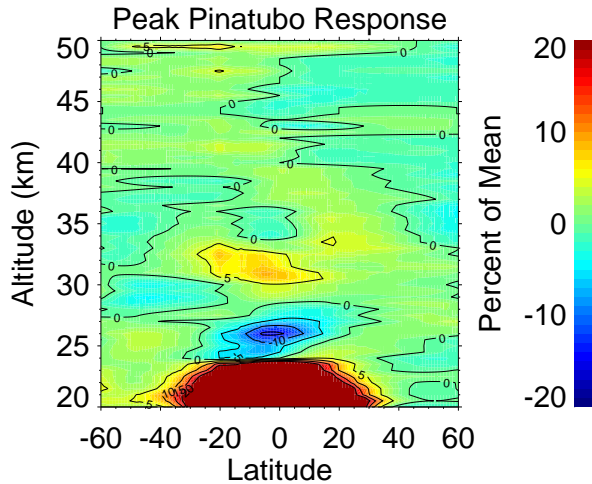




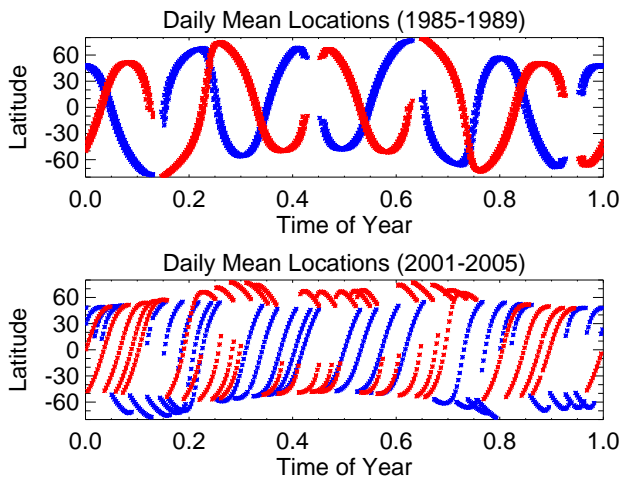
**Figure 8.** Amplitude of oscillation of the solar term as a percentage of the local mean for the use of one and two solar proxy terms while including a volcanic term. Contour lines are plotted at intervals of 0.5%.



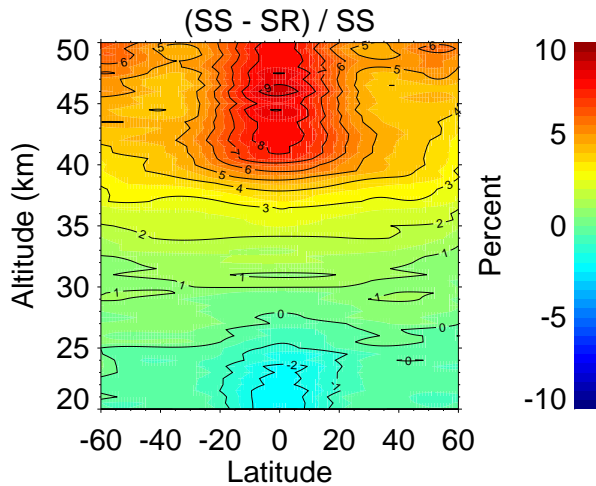
**Figure 9.** Examples of the solar term for the use of one and two solar proxy terms while including a volcanic term. Contour lines are plotted at intervals of 0.5%.



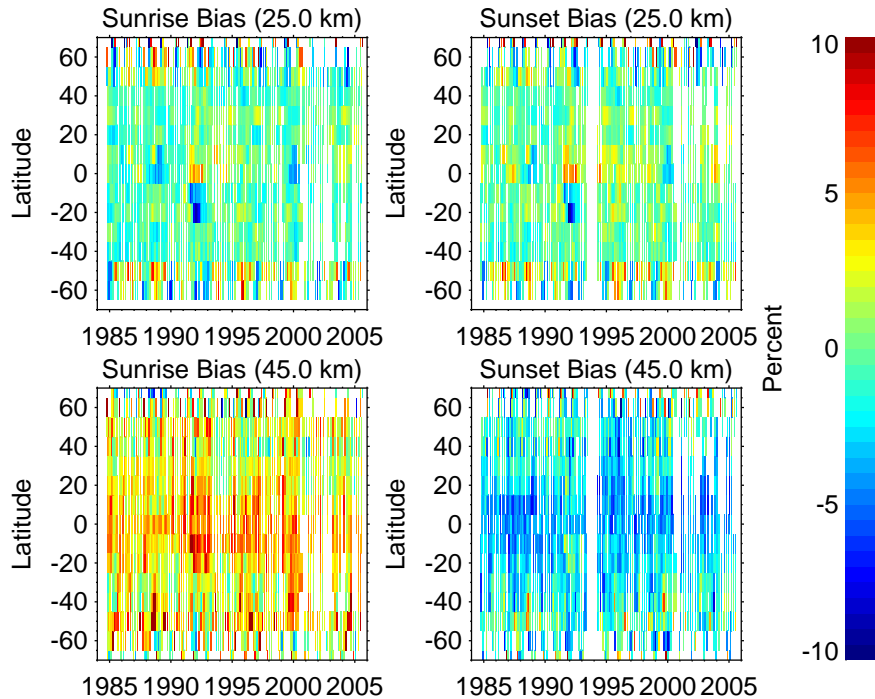
**Figure 10.** Peak of volcanic term near the time of Pinatubo as a percent of mean. Anomalous high values at lower altitudes are the result of overfitting to gaps in data as shown in Fig. 3.



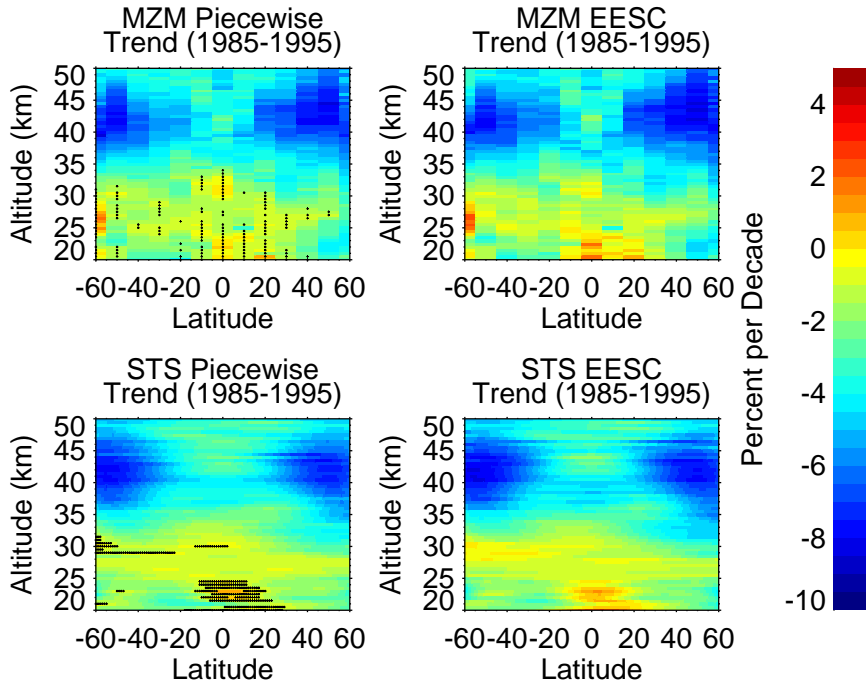
**Figure 11.** Locations of daily means for each satellite event type (sunrises are blue and sunsets are red) at the beginning and end of the mission.



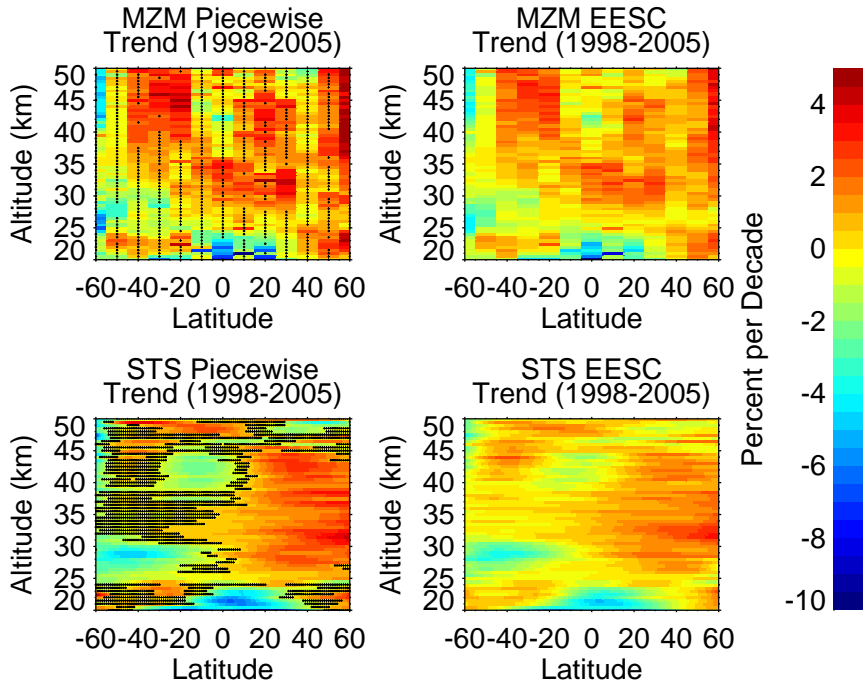
**Figure 12.** Results of the local event type piecewise term from the STS regression plotted as the percent difference between sunrise and sunset events.



**Figure 13.** Biases for each monthly zonal mean for the MZM method compared to the STS method computed as  $(\text{MZM}-\text{STS})/\text{STS} \times 100$ . The MZM method does not differentiate between event types so it does not have different values for each type. White areas show regions where data do not exist or where one or both regression methods failed to converge to a solution.

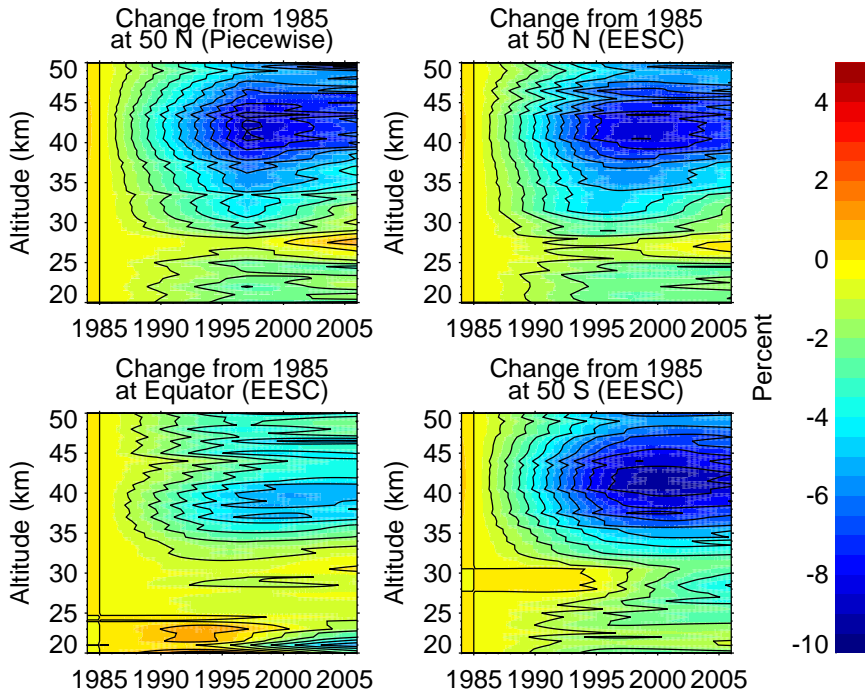


**Figure 14.** Mean trend between 1985 and 1995 for four analysis scenarios computed as  $(Tr(1995) - Tr(1985)) / Tr(1985) * 100$  where  $Tr(t)$  represents the value of the long-term trend term at time  $t$ . The analysis was run for both the MZM and STS regression methods, each using either a piecewise linear term joined at 1997.0 or two orthogonal EESC terms. Stippling shows regions where the linear slope is not significant at the two-sigma level. No similar calculation can be done for multiple EESC terms.

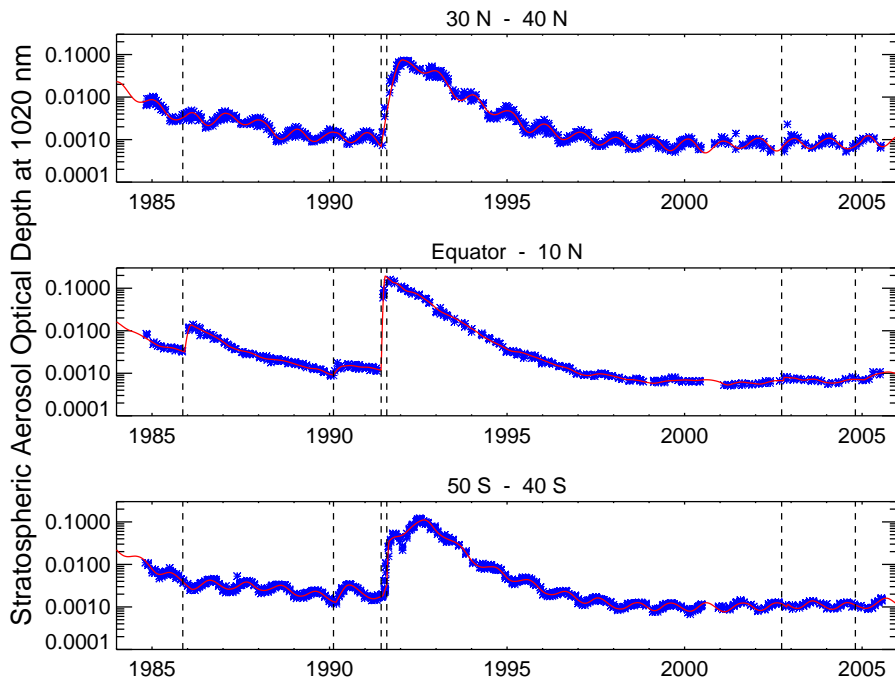


**Figure 15.** Mean trend between 1998 and 2005 for four analysis scenarios computed as  $(Tr(2005) - Tr(1998)) / Tr(1998) * 100$  where  $Tr(t)$  represents the value of the long-term trend term at time  $t$ . The analysis was run for both the MZM and STS regression methods, each using either a piecewise linear term joined at 1997.0 or two orthogonal EESC terms. Stippling shows regions where the linear slope is not significant at the two-sigma level. No similar calculation can be done for multiple EESC terms.

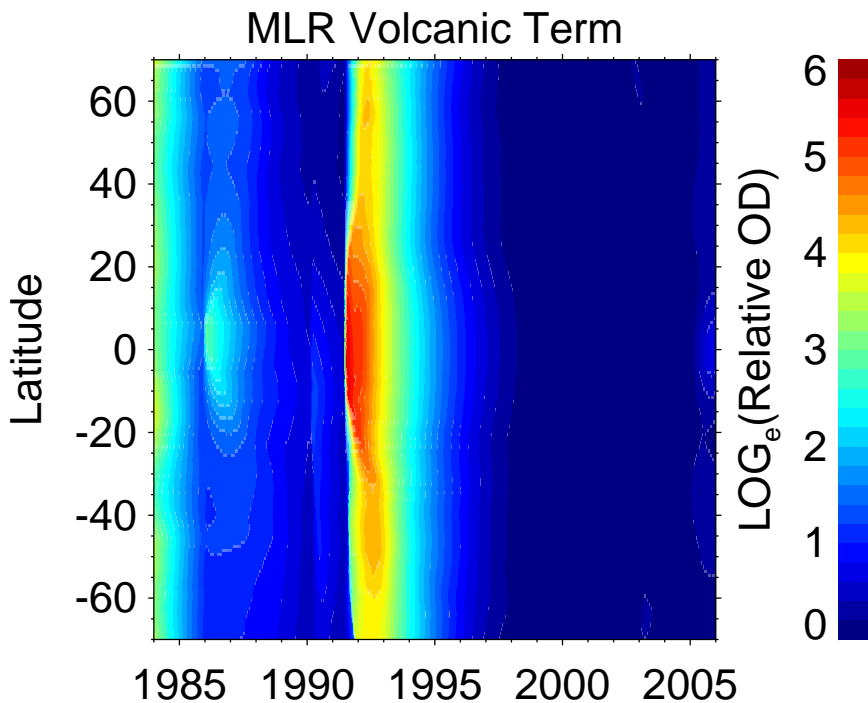




**Figure 16.** Some examples of the long-term trend term computed using the STS regression. The term in the top left comes from the use of a piecewise linear term while the other three come from the use of two orthogonal EESC terms. Contour intervals are 1%.



**Figure 17.** Some examples of the non-linear fit to aerosol at three different center latitudes ( $35^{\circ}\text{N}$ ,  $5^{\circ}\text{N}$ , and  $45^{\circ}\text{S}$ ). Data within each band are shown in blue while fits are shown in red. The vertical lines denote the times of the seven eruptions used for the fit (El Chichon is off scale). Note the difference in rates of rise and peak times at different latitudes, most easily visible for Pinatubo.



**Figure 18.** The volcanic term resulting from the STS regression to stratospheric aerosol optical depth in the 1020 nm channel. This is the term used as a volcanic proxy term for the regression to ozone data. Relative optical depth means relative to background values.