University of Reading

Chapter 4: The LOTUS regression model

Book or Report Section

Published Version

Damadeo, R., Hassler, B., Zawada, D. J., Frith, S. M., Ball, W. T., Chang, K.-L., Degenstein, D. A., Hubert, D., Misios, S., Petropavlovskikh, I., Roth, C. Z., Sofieva, V. F., Steinbrecht, W., Tourpali, K., Zerefos, C. S., Alsing, J., Balis, D., Coldewey-Egbers, M., Eleftheratos, K., Godin-Beekmann, S., Gruzdev, A., Kapsomenakis, J., Laeng, A., Laine, M., Maillard-Barras, E., Taylor, M., von Clarmann, T., Weber, M. and Wild, J. D. (2019) Chapter 4: The LOTUS regression model. In: Petropavlovskikh, I., Godin-Beekmann, S., Hubert, D., Damadeo, R., Hassler, B. and Sofieva, V. (eds.) SPARC/IO3C/GAW, 2019: SPARC/IO3C/GAW Report on Long-term Ozone Trends and Uncertainties in the Stratosphere. SPARC, pp. 37-50. doi: 10.17874/f899e57a20b (SPARC Report No. 9, GAW Report No. 241, WCRP-17/2018, available at www.sparc-climate.org/publications/sparc-reports) Available at https://centaur.reading.ac.uk/82511/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

Identification Number/DOI: 10.17874/f899e57a20b
<https://doi.org/10.17874/f899e57a20b>

Publisher: SPARC

University of Reading

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Chapter 4: The LOTUS regression model

One of the primary motivations of the LOTUS effort is to attempt to reconcile the discrepancies in ozone trend results from the wealth of literature on the subject. Doing so requires investigating the various methodologies employed to derive long-term trends in ozone as well as to examine the large array of possible variables that feed into those methodologies and analyse their impacts on potential trend results. Given the limited amount of time, the LOTUS group focused on the most common methodology of multiple linear regression and performed a number of sensitivity tests with the goal of trying to establish best practices and come to a consensus on a single regression model to use for this study. This chapter discusses the details and results of the sensitivity tests before describing the components of the final single model that was chosen and the reasons for that choice.

4.1 Regression methodology

MLR methods have been used for trend detection in ozone time series for decades. They evolved into the most commonly used approach in the community, with many smaller or more substantial variations of a baseline method having been developed. For the LOTUS project, we decided to base our sensitivity tests on a MLR with an iterative lag-1 autocorrelation correction (see *Appendix B* of *Damadeo et al.,* 2014).

In general, the regression problem can be written as,

$$y = X\beta + \varepsilon \qquad (4.1),$$

where y is the length n vector of observations, β is the length m vector of proxy coefficients, X is the $n \times m$ matrix of proxies, and ε are the fit residuals. The goal of the regression procedure is to find the values of β which minimise the quantity

$$(y - X\beta)^T \Omega^{-1} (y - X\beta) \qquad (4.2),$$

where Ω is the covariance matrix of the observations. The problem admits a direct solution,

$$\beta = (X^T \Omega^{-1} X)^{-1} \Omega^{-1} y \qquad (4.3),$$

which can also be used to obtain an error estimate for the proxy coefficients assuming the covariance matrix is correctly specified.

The regression is performed in an iterative procedure (*Cochrane and Orcutt*, 1949) with Ω set to unity for the first iteration. The first iteration is equivalent to an unweighted least squares fit. After the first iteration, the autocorrelation coefficient, ρ, is calculated through,

$$\rho = \frac{\sum_{i=2}^{N} (\varepsilon_i - \overline{\varepsilon})(\varepsilon_{i-1} - \overline{\varepsilon})}{\sum_{i=1}^{N} (\varepsilon_i - \overline{\varepsilon})^2} \qquad (4.4),$$

where $\overline{\varepsilon}$ is the mean value of the residuals. Typically, the autocorrelation coefficient is on the order of 0.2–0.3. For the next iteration, the covariance matrix is updated taking into account the observed autocorrelation (*Prais and Winsten,* 1954) with modifications by *Savin and White* (1978) if gaps are present in the data. The procedure is repeated until the autocorrelation coefficient has converged within a tolerance level of 0.01. The final error estimate is calculated by scaling the covariance matrix to match the observed variance of the residuals.

This baseline MLR was used for sensitivity tests to decide which proxies to use in the final "LOTUS regression" model, for evaluations of possible lags for proxies, and for the evaluation of weighted or unweighted regressor data. The final set-up of the "LOTUS regression" model is described in more detail in *Section 4.5*.

4.2 Proxies

Proxies are used in multiple regression analyses to represent the observed variability in the parameter being modeled, in this case ozone. There is a wealth of literature concerning the viability of various proxies to represent dynamical and chemical processes that affect ozone (*e.g.*, *WMO,* 2011; *WMO,* 2014; and references therein).

We briefly describe the most common proxies for ozone trend analyses below and provide information on where these proxies may be found. Our focus is therefore not to provide detailed studies about the effects of these proxies on ozone distribution but rather a short estimate about their influence mechanism (dynamical or chemical) and a description on how the proxy has been implemented in regression models before. The listed links for the proxies are not exhaustive and should only be seen as a subset of all available possible sources.

4.2.1 Non-trend proxies

11-year solar cycle

The 11-year solar cycle has different effects on ozone in the different regions of the atmosphere.

The solar ultraviolet spectral irradiance reaching the Earth's atmosphere changes over the course of the cycle. In the upper stratosphere this leads to changes in radiative heating and photochemistry (production rate of ozone) which then affects the ozone distribution. In the lower stratosphere the changes in ozone are thought to occur mainly through a dynamical response to solar ultraviolet variations. The exact mechanisms of this dynamical response are not yet fully understood (*WMO,* 2014), but the 11-year solar cycle proxy is important for all latitudes. Effects of the solar cycle on ozone are described, for example, in *Lee and Smith* (2003).

There are different possibilities to describe the 11-year solar cycle as a proxy. The most common ones are: 10.7 cm solar radio flux, sunspot number, Mg II core-to-wing ratio, and as a more recent alternative the 30.0 cm solar radio flux (as suggested by *Dudok de Wit et al.,* 2014). Note, all of these time series are highly intercorrelated, so only one is chosen for a solar cycle representation in a regression model. Additionally, while these proxies could theoretically be phase shifted to account for any potential lagged response (*e.g.*, in dynamical forcings in the lower stratosphere), in practice adding this additional degree of freedom can cause aliasing due to correlations with volcanic effects (*Chiodo et al.,* 2014; *Damadeo et al.,* 2014) that can negatively impact trend analyses.

Time series of these proxies can be found here:

i. Solar flux (10.7 cm):
   ftp://ftp.geolab.nrcan.gc.ca/data/solar_flux
ii. Solar flux (30.0 cm):
   ftp://ftpsedr.cls.fr/pub/previsol/solarflux/observation
iii. MG II index:
   http://www.iup.uni-bremen.de/UVSAT/Datasets/mgii
iv. Sunspot number:
   http://sidc.oma.be/silso/datafiles

QBO

The quasi-biennial oscillation (QBO) is a modulation of the zonal wind and temperature in the tropical stratosphere over time and pressure region, measured by radiosondes. These changes in wind and temperature affect ozone in this region, as well as ozone outside the tropics. The QBO time series are given as wind measurements at several different stratospheric pressure levels. Effects of QBO on ozone are described, for example, in *Baldwin et al.* (2001).

Due to the oscillating nature of the QBO, and its region of influence beyond the tropics, it is important to account for phase shifts when it is used as a proxy. There are different ways to accomplish this: (1) at least two (or more) measured QBO time series from different pressure levels that are mostly orthogonal are used simultaneously (as done, for example, in *Steinbrecht et al.*, 2017), (2) one measured QBO time series is chosen and an orthogonal time series to this QBO time series is artificially created (as done, for example, in *Harris et al.,* 2015), or (3) QBO time series at multiple pressure levels are taken and a principle component (CP) analysis is performed with them to get to orthogonal QBO time series (as done, for example, in *Damadeo et al.,* 2014).

Time series of QBO values at the different pressure levels can be found here:

http://www.geo.fu-berlin.de/met/ag/strat/produkte/qbo/qbo.dat

ENSO

The El Niño–Southern Oscillation (ENSO) is an important mode of interannual variability in wind and sea surface temperatures over the tropical Pacific Ocean. These variations cause variability in tropical upwelling and therefore changes in lower stratospheric temperature and water vapor. These then affect the ozone concentration in the tropics chemically and dynamically. Through atmospheric teleconnections, ENSO also affects ozone distributions in regions beyond the tropics (*WMO,* 2014; *Oman et al.,* 2013). ENSO time series can be given as sea level pressure difference between Darwin, Australia, and Tahiti (Southern Oscillation Index), as differences in sea surface temperatures, or as a combination of several different indices.

ENSO as a proxy is often used only as a single time series. However, it has been shown that ENSO effects outside the tropics can be delayed by some time compared to the original signal. Therefore ENSO proxies are either lagged (by a variable number of months) or an orthogonal ENSO time series is created, and the original and orthogonal ENSO proxy are used in combination to account for the time lags.

Time series of ENSO values can be found here:

https://www.esrl.noaa.gov/psd/enso/mei
(*Wolter and Timlin*, 2011)

AO

The Arctic Oscillation (AO), also known as the Northern Annular Mode (NAM), is a description of North-South

movement of the westerly winds that circle the Arctic associated with pressure anomalies of one sign in the Arctic with the opposite anomalies centred about 37°N–45°N. The effect of the AO on ozone is discussed, for example, by *Thompson and Wallace* (2000).

Time series of AO values can be found here:

https://www.cpc.ncep.noaa.gov/products/precip/CWlink/daily_ao_index/ao.shtml

AAO

The Antarctic Oscillation (AAO), also known as the Southern Annular Mode (SAM), is a description of North-South movement of the westerly winds that circle Antarctica associated with pressure anomalies of one sign centred in the Antarctic and anomalies of the opposite sign centred about 40°S–50°S. It can have a clear influence on ozone in the Southern Hemisphere polar regions as has been shown by *Thompson and Solomon* (2002).

Time series of AAO values can be found here:

https://www.cpc.ncep.noaa.gov/products/precip/CWlink/daily_ao_index/aao/aao.shtml

NAO

The North Atlantic Oscillation (NAO) index is based on the pressure difference at sea level between the Icelandic low (Subpolar Low) and the Azores high (Subtropical High). Its variations impact the strength and direction of westerly winds across the North Atlantic, which can then affect ozone distribution, mainly in the lower stratosphere. The effects of NAO on ozone is described, for example, by *Weiss et al.* (2001).

NAO can be given as a time series of air pressure differences between a location in Iceland and a location in the Azores or Portugal, or as empirical orthogonal functions (EOFs) of surface pressure defined regionally about these locations.

Time series of NAO values can be found here:

https://climatedataguide.ucar.edu/climate-data/hurrell-north-atlantic-oscillation-nao-index-station-based

EHF

The Eddy Heat Flux (EHF) is a metric used to describe the stratospheric meridional circulation, based on wave energy moving from the troposphere upward into the stratosphere. It is calculated as a product of North-South (meridional) wind and temperature departures from their respective zonal-mean values, and it is mainly important in the winter hemisphere in the polar regions. More information about the influence of EHF on ozone can be found, for example, in *Gabriel and Schmitz* (2003).

Eddy Heat Flux time series are calculated from reanalysis data, for example ERA-Interim.

Tropopause pressure

Long-term changes in tropopause pressure can be used as a proxy for tropospheric expansion and therefore as an indicator for the influence of climate change on the atmosphere. The connections between ozone and tropopause height has been discussed, for example, by *Steinbrecht et al.* (1998). The average tropopause pressure for use as a proxy in the regression has to be calculated specifically, depending on the analysed data: For station data (*e.g.*, a sonde station), normally the tropopause pressure that is recorded directly in the data files is used; for satellite data analyses, tropopause pressures are normally retrieved from reanalysis data.

Aerosol

Stratospheric sulfate aerosol concentrations can have an impact on ozone concentrations, especially in the polar regions. With increasing sulfate concentrations the surface area of atmospheric particles increases and therefore offers more opportunity for heterogeneous chemical processes and the effects on stratospheric temperature and transport changes. A major source of sulfate aerosols comes from volcanic eruptions. Two major eruptions took place in the last 40 years (*i.e.*, El Chichón in 1982 and Pinatubo in 1991) and have to be considered with ozone trend analyses. The peak of the atmospheric aerosol concentration varies regionally due to the transport times of the aerosols. Some satellite and ground-based instruments had substantial problems measuring ozone during periods of very elevated stratospheric aerosol, which is normally accounted for by removing several years of data after major volcanic eruptions from these data sets. More information about the effects of stratospheric sulfate aerosols on ozone can be found, for example, in *Solomon et al.* (1998).

There are different ways to describe aerosol as a proxy: (1) a theoretical functional form that describes the injection and loss of aerosol in the atmosphere, or (2) time series based on aerosol-related measurements. In both cases lags for the time series have to be considered to take into account the transport times.

Time series of aerosol optical depth (AOD) values can be found here:

> Mean AOD at 550 nm:
> **https://data.giss.nasa.gov/modelforce/strataer/tau.line_2012.12.txt**
> (Note, this time series is only available until mid-2012; if it was to be used as a proxy in a regression model, it would have to be extended. This is normally done by extending the average value of the last few years or the last value of the time series, assuming that during those years the mean AOD was representative of background values.)

4.2.2 Trend proxies

In addition to the periodic or punctuated influences of geophysical variability detailed above in *Section 4.2.1*, studies of long-term ozone levels also reveal an additional trend-like behavior. These long-term changes stem primarily from long-term variability in chemically reactive halogens (*e.g.*, *Molina and Rowland,* 1974*; WMO,* 2014 and references therein) and the effects of steadily increasing GHG on stratospheric temperatures and dynamical transport processes (*e.g.*, *Shepherd et al.,* 2008; *Li et al.,* 2009). Halide compounds will photodissociate, releasing halogens (*e.g.*, chlorine or bromine) that chemically interfere in the Chapman cycle resulting in catalytic ozone loss and a subsequent inverse relationship between halide loading and ozone levels that is most prominent in the upper stratosphere. The continued injection of GHGs (*e.g.*, carbon dioxide or methane) into the atmosphere results in stratospheric cooling, which slows the reaction rates of chemical ozone loss, and tropospheric heating that have a combined effect on dynamical transport mechanisms such as increasing the Brewer-Dobson circulation. This increased tropical upwelling is predicted to result in the decrease of ozone in the tropical UTLS and increase in ozone at mid-latitudes in the lower stratosphere and these dynamical influences can potentially be more influential than chemical forcing in the lower stratosphere. However, while changes in greenhouse gases exhibit a linear behavior, halogen concentration peaked in the mid-1990s and then began to decline. This combination of long-term effects complicates the ability of regression models to accurately derive long-term trends in ozone. Whereas early (*i.e.*, before the mid-1990s) works could make use of a simple linear trend to model long-term changes in ozone, studies thereafter have utilised more complicated trend proxies in regression analyses that are detailed below.

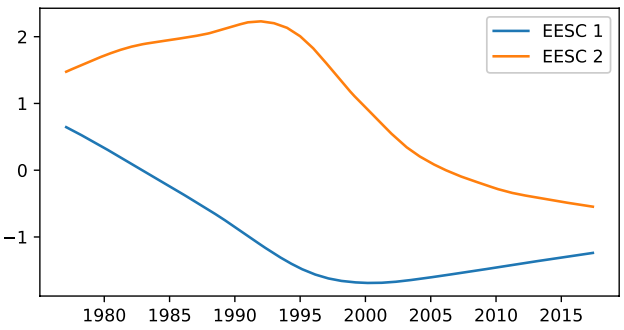

Figure 4.1: *The leading two EESC EOF terms derived from multiple mean age-of-air EESC time series proxies.*

EESC

The equivalent effective stratospheric chlorine (EESC) proxy describes the total halogen loading (chlorine and bromine) of the stratosphere that contributes to ozone depletion (*Newman et al.,* 2007). The shape and timing of the peak of the EESC time series depend on the strength of the Brewer-Dobson circulation and are therefore different for different locations in the atmosphere. To account for this in ozone regression analyses, two different methods have been used: (1) creation of individual EESC time series depending on the age of air at each chosen location (see for example *Bodeker et al.,* 2013), (2) creation of two orthogonal functions that allows the regression to determine the shape that best fits the data (see **Figure 4.1** and *Damadeo et al.,* 2014). Note that the combination of EESC-based orthogonal functions allows for an EESC-like function but with maximum halogen loading occurring at any point in the time series, or not at all (*i.e.*, monotonic trend), depending on the best fit to the data. That is, the fit is not constrained to the range of classically-defined EESC curves. Significant differences between actual and EOF-based EESC proxy fits indicate the data are diverging from a linear fit to actual EESC or responding to other forcing not explicitly represented in the regression model.

Time series of EESC values can be found here:

> **https://acd-ext.gsfc.nasa.gov/Data_services/automailer/index.html**

PWLT

A piecewise linear trend proxy (PWLT) is a combination of two linear trend terms. The first is a regular linear trend term, while the second is a linear trend term that is set to 0 until a specific time (inflection point) and is a simple linear trend afterwards. The two lines of the trend proxies are forced to meet at the inflection point. The trends in the two periods (before and after the inflection point) are therefore linked. The inflection point is chosen to coincide with the peak concentration of ODSs in the atmosphere. Since the timing of this peak changes depending on the location in the atmosphere, ideally the PWLT proxy takes this variability into account. Often, however, PWLT is applied with the same inflection point (end of 1997) at every location. Both trend terms of PWLT are fit simultaneously with the other proxies. A PWLT was used, for example by *Harris et al.* (2015).

ILT

The independent linear trend (ILT) method is also based on two different trends (like PWLT) to describe the ozone decrease in the 1980s and 1990s and the slow ozone increase since the early 2000s. Unlike the PWLT, ILT results are not linked but fit independently for the two chosen time periods. The analyses with ILT can be done in one or two steps. For the two step approach all available proxies for the ozone regression are used in a first regression fit. Then the contributions of the proxies beside the trend proxies are removed from the original data, which leaves only the contribution of the trend proxies and the residuals. In the last step the trend proxies (two different proxies for two different time periods) are fit to these already modified data to determine the trends. A regression using ILT was used, for example, by *Steinbrecht et al.* (2017). In a one step approach the two steps explained above are combined but with one additional regression term representing a constant for describing the period between the two trend terms.

4.3 Sensitivity tests

Part of the difficulty in deriving long-term trends in stratospheric ozone is the large variety of choices to be made during the process. There are a large number of data sets from different instruments that have been combined in different ways and with different merging techniques (see *Chapter 2*). There are also a large number of potential proxies to choose from to use in regression analyses, each with some apparent merit for use in different temporal and/or spatial regimes. Indeed, the SI2N effort saw a great number of different analyses applied to various different combinations of data sets performed using different combinations of proxies and different regression methodologies by different groups (*Harris et al.,* 2016). In order to attempt to disentangle the effects of these variables (*i.e.*, data sets used, proxies used, and analysis technique), a number of sensitivity tests were performed. These sensitivity tests were designed to determine what variables are influential versus non-influential and to try to establish best practices if possible.

4.3.1 Survey of existing regression models

Perhaps the main difference between recent ozone trend studies has been the varied use of regression-based models and data sets. As such, a logical first step in performing sensitivity tests would be to apply the same set of regression-based methodologies to a single data set. This serves two purposes: (1) to validate the consistency of execution against previous studies and (2) to probe the sensitivity of resulting trends to different combinations of proxies and methods. In considering this, groups involved in past ozone trends studies were asked to run their regression models on a single data set, SBUV MOD, provided by the LOTUS working group. This data set was recomputed over large latitude bins (*i.e.*, 50°S–35°S, 20°S–20°N, and 35°N–50°N) and nine different pressure levels between ~40 hPa and ~0.6 hPa (modified from *Frith et al.,* 2017) and was chosen for its ease of use for this particular sensitivity test. The data were given to each of 15 different groups to apply their regression methodology on and provide the LOTUS group with all of the relevant results (*i.e.*, coefficient values and uncertainties and proxies used). Each separate regression analysis was applied as the group has done in the past (*i.e.*, with regard to regression methodology and choice of proxies) with the exception of restricting the long-term trend term to a piecewise linear trend with a turnaround time at the end of 1997 (*Jeannet et al.,* 2007; *Zerefos et al.,* 2012; *Coldewey-Egbers et al.,* 2014; *Damadeo et al.,* 2014; *Fragkos et al.,* 2016; *Misios et al.,* 2016; *Ball et al.,* 2017; *Sofieva et al.,* 2017; *Steinbrecht et al., 2017; Weber et al.,* 2018).

Figure 4.2 shows ozone trend results for each of the 15 different regression models in two separate latitude bands (50°S–35°S and 35°N–50°N) for the time periods before/after 1997 with dashed/solid lines. Since this sensitivity test was applied to only a single data set, the emphasis here is not on the trend values themselves but rather how they compare.

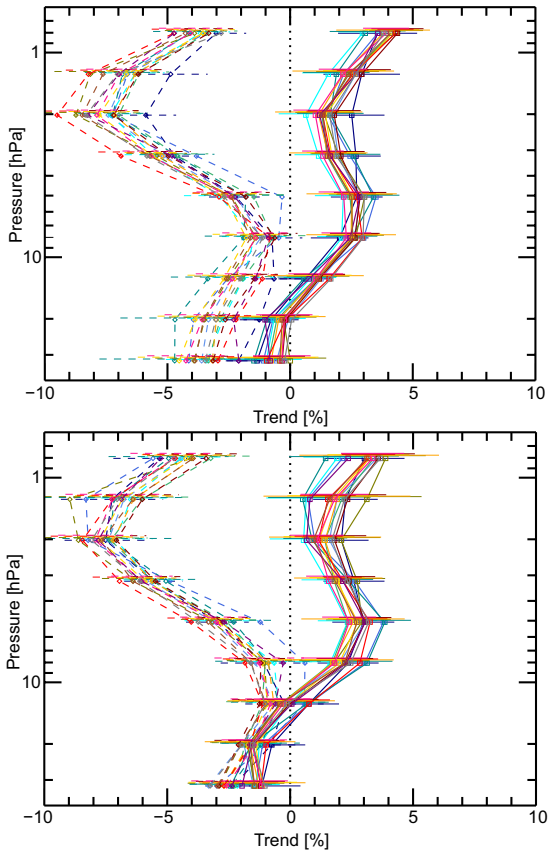

Figure 4.2: *Derived ozone trends in percent per decade from the 15 different regressions applied to the same SBUV MOD data set between 35°N–50°N (top) and 50°S–35°S (bottom). In each plot, the dashed/solid lines represent trend values before/after 1997 at each of the 9 pressure levels.*

The different regression models are mostly in agreement (*i.e.*, within their respective error estimates) at all latitudes, pressure levels, and time periods showing how the resulting trends are fairly robust with respect to the choice of regression methodology and non-trend proxies. However, there is still a spread in the trend values of about 1.5 % per decade in the lower to mid-stratosphere to 3 % per decade in the upper stratosphere, particularly in the Southern Hemisphere. With current estimates of "recovery" trends in the upper stratosphere at mid-latitudes of about 2–3 % per decade, this spread is a potentially large source of uncertainty. Results for the latitude bands 20°S–20°N and 50°S–50°N (not shown here) are very similar in their spread of results between the different regression models. As such, it was deemed prudent to perform more extensive sensitivity tests to see how trend results would change given different regression methodologies and proxies.

4.3.2 Weighted versus unweighted regression

With a few exceptions, nearly every regression analysis of long-term trends in ozone have involved unweighted regressions. The downside to this, of course, is that the uncertainties in the observations (or the resulting mean values) are never taken into consideration for the calculation of regression coefficients or the resulting uncertainties. As such, we applied both unweighted and weighted regression techniques to the data using the standard error of the mean as a weighting factor. The weighted regression and associated heteroscedasticity correction were applied as detailed in *Appendix B* of *Damadeo et al.* (2014).

Weighted regressions use the inverse of the variance in the data as the ideal weights for the regression, which is typically substituted with the inverse of the square of the uncertainties (*i.e.*, standard errors). However, an often forgotten assumption of this technique is that the variances in the data are known precisely. Since, in practice, this assumption almost never holds, a heteroscedasticity correction is necessary to attempt to modify the uncertainties using the nature of the residuals. The form of this correction, however, can be more complicated as it requires some a priori knowledge of the nature of the modification. Given that the standard errors change with both geophysical variability and the number of samples, the heteroscedasticity correction was assumed to have a seasonal form and allowed to vary for time periods containing different collections of data sets (see *Appendix B* of *Damadeo et al.*, 2014).

As an example we investigated the impacts of using weighted regressions on the merged SAGE-OSIRIS-OMPS ozone data set. **Figure 4.3** illustrates the importance of the heteroscedasticity correction. Because the standard error has a strong dependence on the sampling frequency, the weights increase significantly as the data set moves from lower samples (*i.e.*, SAGE II) to higher samples (*i.e.*, OSIRIS and OMPS). This creates mismatched weights over different time periods and results in trends that are dictated entirely by the latter data sets. Incorporating the heteroscedasticity correction helps to "even out" the weights to a more reasonable representation. Analysis of the residuals (**Figure 4.4**) reveals that the heteroscedasticity also reduces the variance in the residuals, resulting in a more robust fit. The standard deviation of the residuals is both generally reduced and more uniform over time after the heteroscedasticity correction. However, **Figure 4.4** also reveals the limitations of this correction. As previously mentioned, the heteroscedasticity correction can be complicated as it requires some a priori knowledge of the nature of the uncertainty modification. Since the standard error is strongly dependent on the sampling, it can change not only as new instruments are added but from month to month even within a single latitude bin, particularly during periods of overlap between instruments with very different sampling. As such, the true nature of the heteroscedasticity is much more complicated than can be simply modeled and it becomes extremely difficult to accurately correct for data sets that are already merged. Instead, this correction should be performed on each individual data set prior to merging. Since the analyses within the LOTUS investigation only use pre-merged data sets, an accurate heteroscedasticity correction cannot be computed and thus a weighted regression technique is not included in the final results.

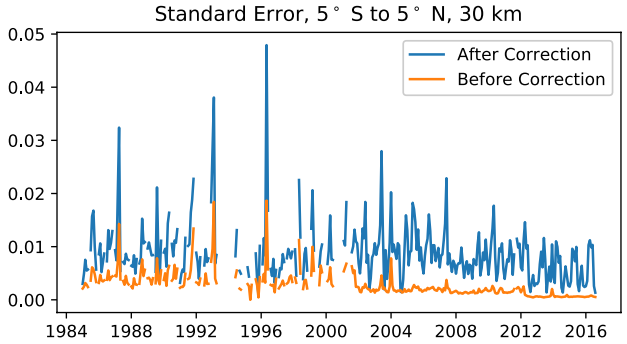

Figure 4.3: *Standard errors between 5°S–5°N at 30 km used for the weighted regression applied to the merged SAGE-OSIRIS-OMPS ozone data set both before and after the heteroscedasticity correction. Note how the errors/weights before the correction drop/rise dramatically once OSIRIS (added in 2001) and OMPS (added in 2012) are added as a result of increased sampling of the instruments.*

4.3.3 Non-trend proxy sensitivity

When performing MLR of a dependent variable (*e.g.*, ozone) to a set of proxies (*e.g.*, sources of geophysical variability), the primary motivation is to determine attribution (*e.g.*, how much variability in ozone is caused by solar flux). Precisely determining attribution requires that each proxy is orthogonal to (*i.e.*, has zero correlation with) every other proxy. Unfortunately this is almost never the case and the proxies used for the regression often have some degree of multicollinearity (*i.e.*, an individual proxy or combination of proxies is somewhat correlated to another individual or

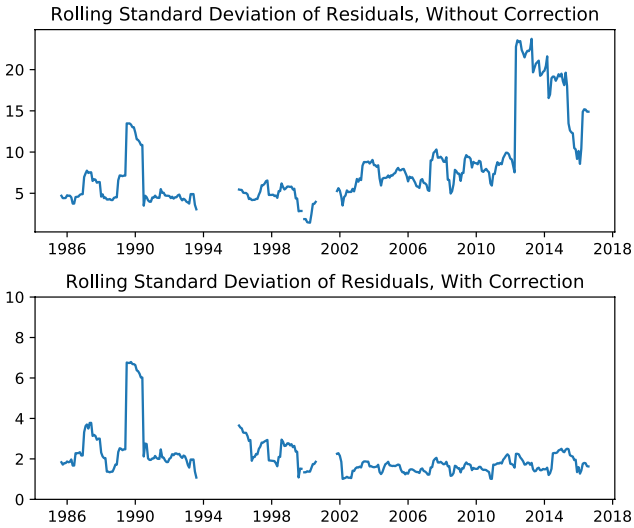

Figure 4.4: *Rolling (running average) standard deviation of the residuals of the weighted regression between 5°S–5°N at 30 km both with and without the heteroscedasticity correction. In addition to generally reducing the spread of the residuals of the fit, the heteroscedasticity correction also makes the rolling standard deviation more uniform throughout the data. Prior to the correction, the rolling standard deviation (not the standard error) increased when moving from SAGE II data to OSIRIS and later OMPS data.*

combination of proxies). While multicollinearity can affect the regression results, this information is captured in the covariance matrix and the final uncertainty estimates of the regression coefficients. However, our focus is on the impact of variability that, if not properly accounted for, may alias into the long-term trend estimate. As such, proxies that do not impact the long-term trend estimate or uncertainty only serve to complicate the statistical model and possibly alter the fits to other regression terms. To test the impact of various proxies on the trend, in the following sections we compare trend results from regressions run with and without the proxy included.

Many different MLR analyses of ozone have been performed over the years using many different combinations of proxies. To test the sensitivity of the regression to these various proxies, a simple model consisting of the leading two QBO EOFs, ENSO with zero lag, solar f10.7, and a PWLT with inflection point in 1997 was applied to each of three different data sets (*i.e.*, SAGE-OSIRIS-OMPS, GOZCARDS, and SBUV MOD) and then a single proxy was either added, subtracted, or substituted from this model. Since the aim of LOTUS is to investigate trends, the following figures compare the derived trend results (**Figures 4.5** and top row of **4.6**) and their uncertainties (**Figure 4.6,** bottom row) before and after each proxy change. Trend differences are shown as differences after the proxy change minus before the proxy change with units of percent per decade. Uncertainty / significance differences are calculated as differences after the change minus before the change with units of percent with the calculated value being the ratio of the trend to its uncertainty. Thus, for example, if initially a trend value was significant to the 2-sigma level and then afterwards it was significant to the 1-sigma level, the significance difference would be -50 %.

First it is worth noting what proxies had a very small impact on the trends differences and their uncertainties. Including any of the AO, AAO, NAO, or EHF proxies had a negligible impact on both the pre-1997 and post-1997 trends value for all three satellite data sets. Also, while some small differences are apparent in the resulting significances, none of the changes were sufficient to make trends that were not significant become so. It is also worth noting that each test that removed the QBO, ENSO, or solar proxies had a significant impact on the trends (1–2% per decade depending upon altitude and latitude) and a general decrease in overall significance levels, though there was some dependence on which data set was used.

The QBO proxy is often taken as the first two EOFs derived from the Singapore zonal winds. Even though higher order terms could be used, adding a third QBO EOF into the regression had negligible impact on the trend and uncertainty results. The ENSO proxy was applied without any lag to it, though often some lag between one to several months is used. Applying the regression with an ENSO proxy lagged anywhere between 1 and 5 months had negligible impacts on the determined trend values but significant impacts on the uncertainties. Including a lag (any lag) generally increases the overall significance, but the results are sporadic in terms of the location in the atmosphere and the degree to which the significance increases. Additionally, the results are not uniform across different data sets and so this analysis does not reveal any optimal lag. As such, the final "LOTUS regression" retains zero lag (see *Section 4.5*).

The two most commonly used solar proxies are the f10.7 and the Mg II proxies, with the f10.7 proxy being the most common. These two proxies yield nearly identical trend results but different uncertainty results.

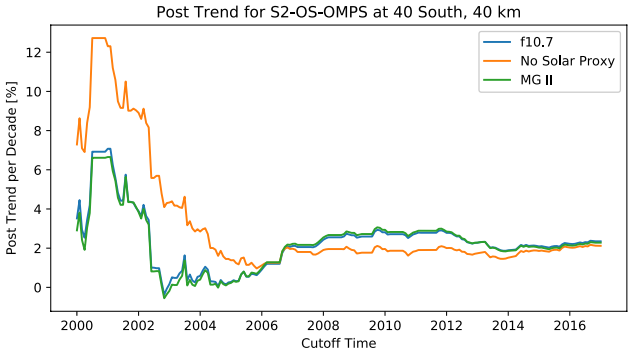

Figure 4.5: *Impact on potential recovery trends depending on the cutoff time of the regression for different solar proxies when applied to the SAGE-OSIRIS-OMPS data set centred at 40°S at 40 km. The impact on resulting trends of ignoring the solar proxy is evident for all but the longest data records while it is also apparent that the f10.7 and Mg II proxies have negligibly different impacts relative to each other.*

However, the changes in trend significance do not clearly indicate one proxy as better than the other and so the f10.7 proxy is retained as the baseline. Even though the different proxies do not yield different results, where any solar proxy shows its real influence is in the low frequency nature of the proxy. The solar cycle has a period of ~11 years, with an amplitude of influence on ozone variability of about 2 %. There is an expectation that proxies with longer periods can potentially have a greater influence on long-term trends in the regression process. This stems from the fact that resulting trends can be subject to endpoint anomalies if fitted over sufficiently short durations of data, particularly in the presence of other sources of variability with periods similar to the fitted duration. As such, the length of data used for the regression and the corresponding phase of the solar cycle at the endpoints can impact the trend results greatly. **Figure 4.5** shows the impact on potential recovery trends for different stopping dates. It is clear that not including the solar proxy is significantly different from including either the f10.7 or the Mg II proxies, though results start to converge with sufficiently long data records. It is also apparent that the length of the data record as it relates to the phase of the solar cycle is important, though the influence starts to level off to within 1 % per decade uncertainty after 2008 as the number of total solar cycles captured during the post-trend analysis increases.

The use of an aerosol proxy can be contentious. While it is well known that sulfate aerosols released from volcanic eruptions can influence stratospheric ozone through both chemical and dynamical effects, the exact relationship between volcanic aerosol and stratospheric ozone levels is not well characterised. Most regression analyses acknowledge the need to account for the El Chichón and Mount Pinatubo eruptions but do not agree on how this should be done. Some analyses simply ignore data immediately after the eruption (*e.g., Wang et al.,* 1996; *Randel and Wu,* 2007; *Harris et al.,* 2015), while others include a regression proxy of some form in an attempt to model the impact (*e.g., Bodeker et al.,* 1998; *Stolarski et al.,* 2006; *Bodeker et al.,* 2013; *Tummon et al.,* 2015). The net ozone response to aerosols depends on the ambient abundance of chlorine and dynamical conditions (*e.g., Tie and Brasseur,* 1995; *Aquila et al.,* 2013). To account for this, some studies include separate regression terms for eruptions of El Chichón and Mt. Pinatubo (*e.g., Stolarski et al.,* 2006; *Frith et al.,* 2014, *Weber et al.,* 2018). However, few aerosol proxies exist, though the most commonly used one is the NASA Goddard Institute for Space Studies (GISS) AOD proxy, which is what is tested here (**Figure 4.6**). The use of an aerosol proxy primarily influences the trend results in the lower stratosphere but only in the SAGE-OSIRIS-OMPS data set, suggesting this data may be more heavily influenced by aerosol interference. Some smaller, coherent patterns do appear in the middle to upper stratosphere, indicating a potential signal in these regions. The influence of adding an aerosol proxy on uncertainties is somewhat mixed (strong positive and negative deviations without a consistent pattern for all three analysed data sets), indicating that the use of this proxy needs further consideration to understand its impact. However, given the need to account for aerosol in some way and the desire not to simply omit data (as the period to omit is a question in itself), the GISS AOD proxy is included in the baseline (see *Section 4.2.1* for an explanation on how the GISS AOD proxy was extended beyond 2012, the last values reported for this proxy). We use a single AOD proxy as Mt. Pinatubo is the only significant eruption in the time period considered.

In general, varying the proxies applied in a regression model can affect the derived trends, though the effect can be mitigated by using data with a sufficiently long record. Additionally, the sensitivity of the trend to other proxies may vary with the resolution of the analysed data set and extent of spatial averaging. A recent study (*Zerefos et al.,* 2018) used 35 years of ozone data from the SBUV MOD data set evaluated both as zonal means and at select lidar station overpasses. As part of that study, the authors applied a similar regression model as that in LOTUS to

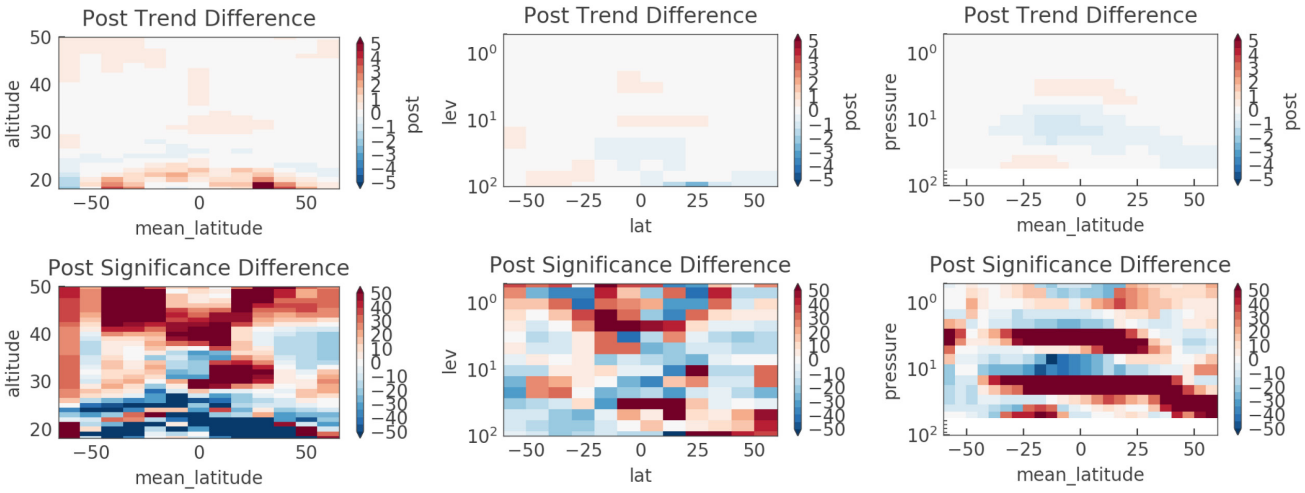

Figure 4.6: *Influence on the "Post-2000" trends (top row) and significances (bottom row) when adding the GISS aerosol proxy for the SAGE-OSIRIS-OMPS (left column), GOZCARDS (middle), and SBUV COH (right) data sets.*

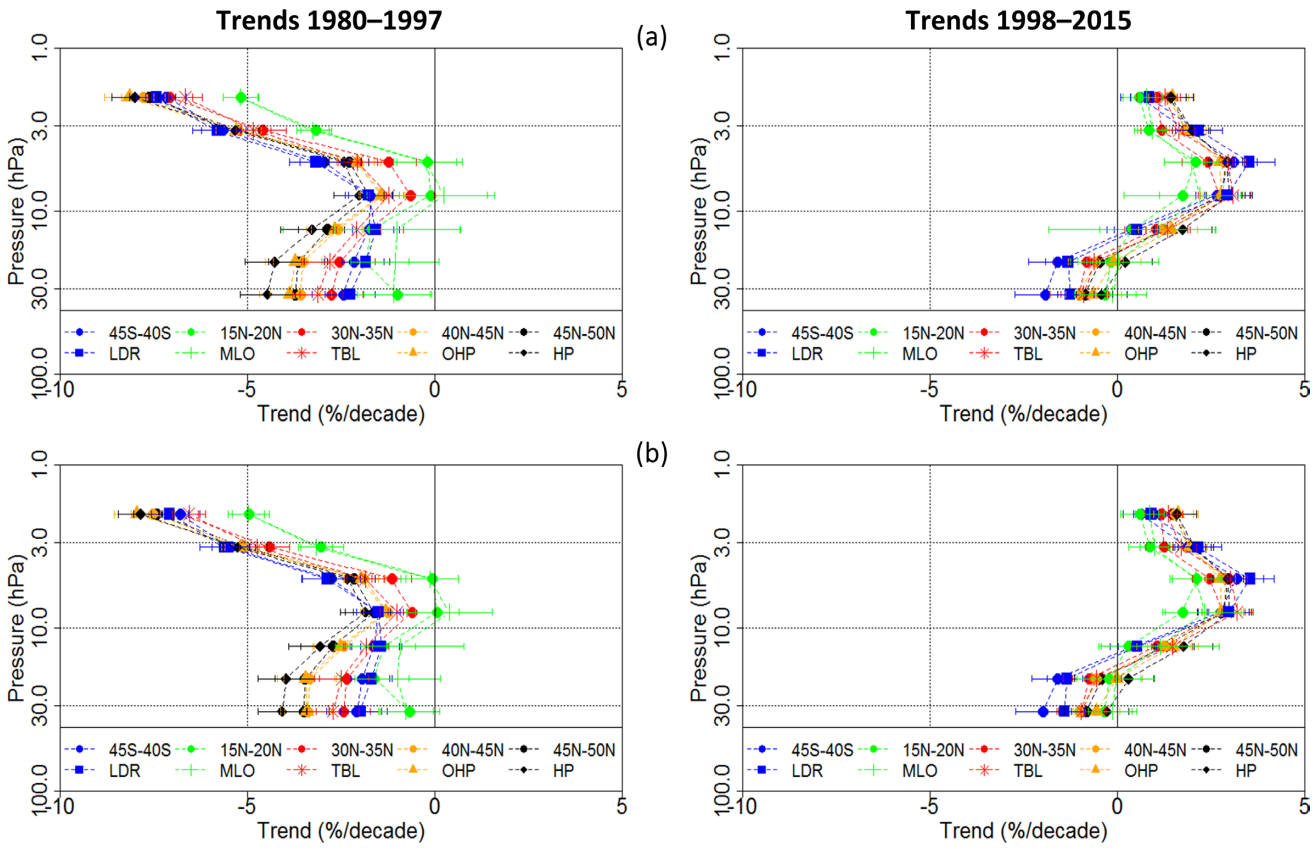

Figure 4.7: *Adapted from* ***Figure 7*** *of Zerefos et al. (2018) showing trends in the vertical distribution of ozone for the pre-1998 and post-1998 period, using (a) two linear trend terms (PWLT method) and volcanic effects and (b) the PWLT method including all proxies. The results are based on SBUV zonal means and SBUV overpasses over five lidar stations (LDR=Lauder, MLO=Mauna Loa Observatory, TBL=Table Mountain, OHP=Haute Provence, HP=Hohenpeissenberg).*

derive ozone trends and tested the effect of including and excluding almost all of the non-trend proxies at once. They found that a model using just the aerosol proxy and PWLT terms and one that also included all of the other non-trend proxies (*i.e.*, QBO, ENSO, AO/AAO, solar, and tropopause pressure) produced little difference in the resulting trend values and uncertainties when applied to SBUV MOD data with a 35 year duration (**Figure 4.7**). This is similar to the results of the sensitivity tests shown here, though we note the sensitivity of the trend is likely less in SBUV MOD due to the reduced vertical resolution of the data.

4.3.4 Trend proxy sensitivity

As detailed in *Section 4.2.2*, there have been four different proxies used to model the long-term trend in ozone for MLR analyses: A PWLT proxy, an ILT proxy, a single EESC proxy, and two EESC EOFs. The single EESC proxy represents the expected linear response of ozone to long-term variability in chemically reactive halogens. The two EESC EOFs are also meant to simulate the chemical forcing of ozone, but the extra degree of freedom allows for non-linearity in the ozone variations due to an imperfectly prescribed EESC shape (*i.e.*, incorrect age of air). The PWLT and ILT are less constrained and structured to better conform to the mean changes in the data from all long-term effects. However, the observational data are not yet sufficient to distinguish changes due to halogen chemistry from those due to other long-term variations induced by increasing GHGs. This means that in regions where chemical forcing is dominant and ozone responds directly to halogen levels (*e.g.*, the upper stratosphere), the EESC-based proxies are a better choice, but in regions where the effects of GHGs are dominant and the correlation between EESC and ozone degrades (*e.g.*, apparent monotonically decreasing ozone trends in the lower stratosphere despite decreasing EESC), the PWLT and ILT proxies are a better choice. While the focus of this Report is on the net changes in ozone from all long-term forcings, analyses of differences among the various trend proxies, in conjunction with longer data records, should allow for better attribution in future studies.

To determine the proxy that best represents observed ozone changes it is necessary to explore the strengths and weaknesses of each by focusing on their respective impacts on derived trends and uncertainties. Most notably, variability in the potential turnaround time can be problematic particularly when combined with nonlinear ozone changes. Stratospheric ozone levels decreased from the earliest satellite observations and this decrease appeared to abate over time.

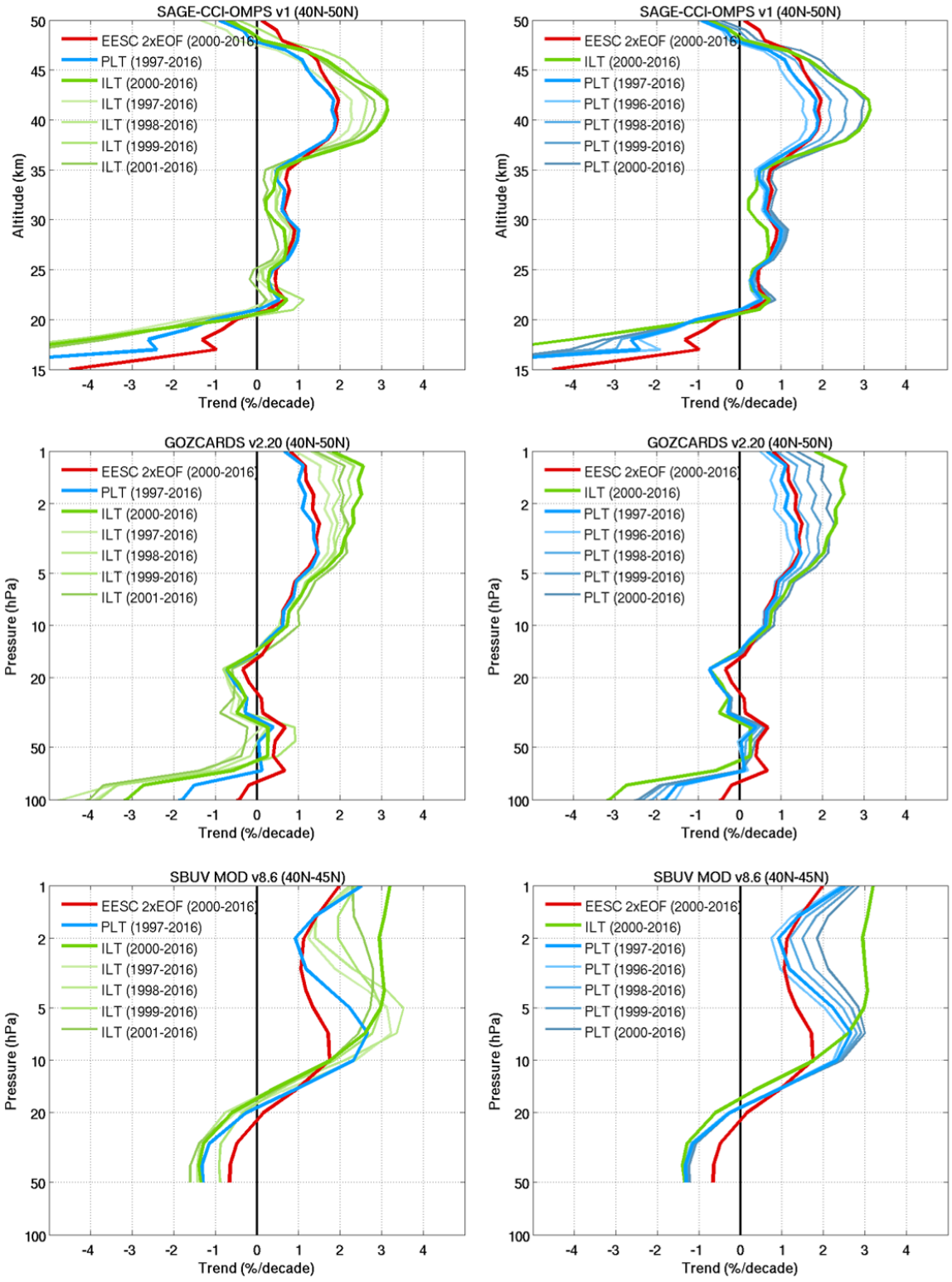

Figure 4.8: *Effect of time period start and end points on ozone trends in the past two decades obtained by regressions with PWLT and ILT.*

MLR Proxy Term	Allows for Curvature?	Allows for Variable Tornaround Time?	Allows for Monotonic Trends?
PWLT	No	No	Yes
ILT	No	Yes	Yes
Single EECS	Yes	No	No
Two EESC EOFs	Yes	Yes	Yes

Table 4.1: *Summary of the pros and cons of the different-long-term-ozone-trend proxies.*

Sensitivity tests have shown an optimal turnaround time of 1997 for the upper stratosphere (*Kyrölä et al.,* 2013; *Laine et al.,* 2014), but this likely changes with latitude and altitude and can be difficult to determine. However, the need for a predetermined turnaround time is the primary problem with using a PWLT proxy for linear regression models as choosing the wrong turnaround time will cause endpoint anomalies in the trend results, particularly in the presence of significant curvature in the actual long-term variability. **Figure 4.8** shows trend profiles derived from GOZCARDS and SBUV MOD where the end points for the PWLT and ILT proxies have been varied. The changes in trends that can be seen with changing time periods are a result of the endpoint anomalies due to the curvature of the long-term ozone variability and unexplained variability in the record. For example, the larger variations in ILT trends in SBUV MOD likely result from a known discontinuity in the data in late 2000. The ILT fits are more sensitive to this discontinuity than the PWLT fits because the endpoints of the PWLT fits are more constrained. Similarly, a single EESC proxy has the same problem as the PWLT (*i.e.*, the turnaround time varies with the mean age-of-air and must be predetermined), except it also cannot account for monotonic trends (*i.e.*, no turnaround in ozone) resulting from radiative and dynamical forcings. Instead, using a single EESC proxy in the presence of monotonic trends (*e.g.*, those seemingly present in the tropical lower stratosphere) will yield biased trend results (*Kuttippurath et al.,* 2015). Thus, a single EESC should never be used to represent the net long-term variability in ozone but only as a tool for determining chemical attribution (provided the correct age of air is known). Avoiding this particular pitfall, the ILT and two EESC EOFs allow for a variable turnaround time albeit in different ways. The ILT accomplishes this by avoiding fitting any trend term during a particular time period (*e.g.*, between 1997 and 2000), instead assuming the data to be constant during this period but allowing for shifts in the regression. This allows for two separate trends to be fit while attempting to avoid endpoint anomalies near the turnaround. The EESC EOFs can actually recreate the variable turnaround time (or lack thereof) and variations in curvature near the turnaround and can potentially allow for an independent assessment of what that turnaround time is. All told, the strengths and weaknesses of the different long-term variability proxies are summarised in **Table 4.1**.

Given the potential problems highlighted above, three different trend proxies were tested as part of a sensitivity study to examine their influence on derived trend results. For the sake of historical comparison, the PWLT proxy with a global turnaround at the beginning of 1997 was used for the baseline model. For the sake of comparison with the WMO Ozone Assessment (*WMO,* 2014), the ILT proxy with the declining period ending at the end of 1996 (hereafter called 'pre-1997' period) and the potential recovery period starting at the beginning of 2000 (hereafter called 'post-2000' period) were used. Lastly, the two EESC EOFs were also used and the resulting trends were determined as described in *Damadeo et al.* (2018) by extracting the combined EESC component of the fit and thereafter fitting a straight line to it over the given time periods to determine the mean trend during those times. Trends were evaluated pre-1997 and post-2000 for each of the three tests and the results are shown for three different data sets in **Figure 4.9**, **Figure 4.10**, and **Figure 4.11**. Not surprisingly, the general pattern of trend results is not that different between the three proxies. The pre-1997 trends are all about -8 % per decade in the upper stratosphere extra-tropics and the post-2000 trends are all about +2–3 % per decade in the same region. Trends in the middle stratosphere are generally about -2 % per decade in the pre-1997 period and about +0–1 % per decade in the post-2000 period. Additionally, all of the trends are noisy (*i.e.*, statistically insignificant) in the UTLS due to lack of high precision data in that region. There are, however, some subtle differences between the proxies. In the pre-1997 trends, upper stratospheric values are smallest (*i.e.*, least negative) in the ILT case and largest (*i.e.*, most negative) in the EESC EOFs case. Trend values are opposite in the post-2000 time frame; they are largest in the ILT case and smallest in the EESC EOFs case.

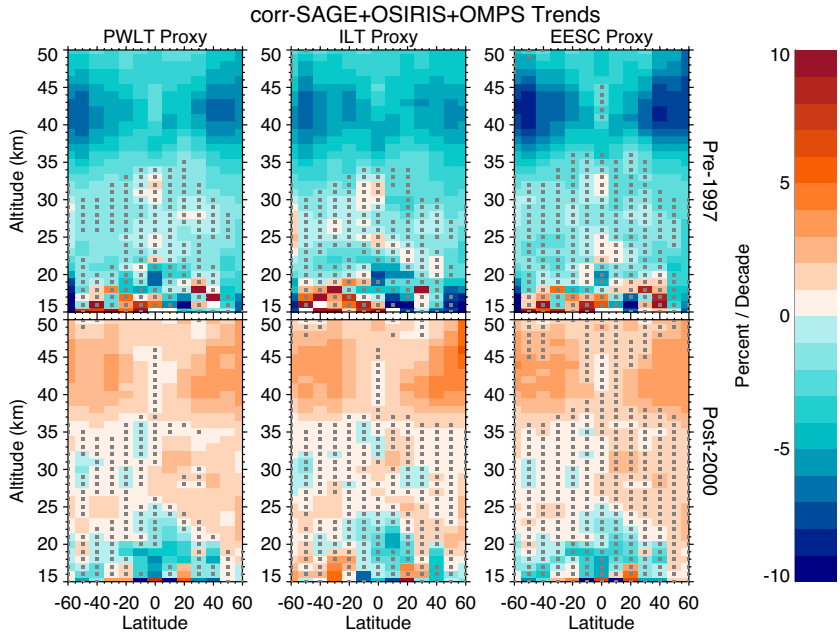

Figure 4.9: *Derived trends in ozone in percent per decade for the SAGE II-OSIRIS-OMPS data set (using the sampling bias adjusted SAGE II data from Damadeo et al., 2018) for both the pre-1997 (start of 1985 to end of 1996, top row) and post-2000 (start of 2000 to end of 2016, bottom row) time periods. Results are shown for each of the three trend proxies: The PWLT (left), ILT (middle), and EESC EOFs (right) proxies. Stippling denotes results that are not statistically significant at the 2-sigma level.*

In the later analyses, where trends are derived in broadband latitude ranges and merged, it made sense for the sake of brevity to pick a single MLR trend proxy term to use. The inability of the single EESC proxy term to capture variability beyond halogen chemistry automatically disqualifies it for use. Similarly, the sensitivity of the PWLT proxy term to the turnaround time suggests it is not ideal for use either. While both the ILT and EESC EOFs are acceptable, the desire to both investigate only the mean trends (*i.e.*, ignore the potential lack of direct correlation between the actual long-term ozone variability and ODSs) and have a more direct analog to compare with results from the last ozone assessment led us to choose the ILT proxy term for the work performed in *Chapter 5*.

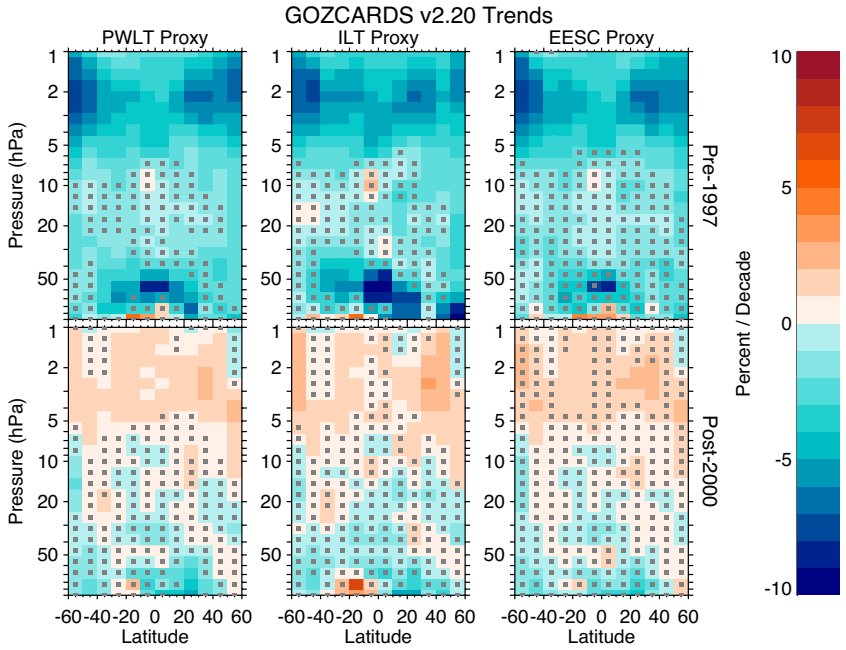

Figure 4.10: *As Figure 4.9 but for the GOZCARDS data set.*

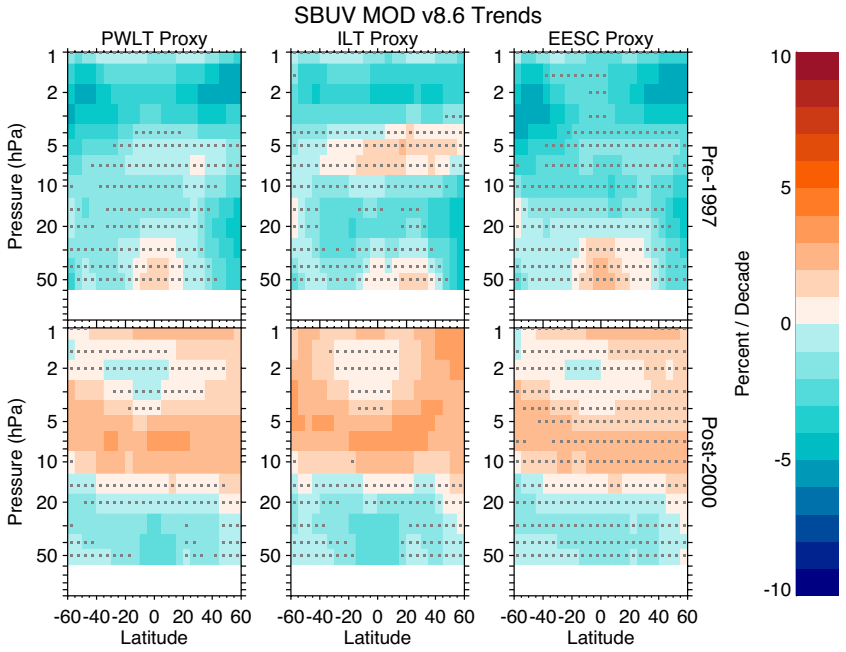

Figure 4.11: *As Figure 4.9 but for the SBUV MOD data set.*

4.4 Alternative approaches

Approaches other than MLR have been used in the community to quantify ozone changes over time. One of them is DLM (*Laine et al.,* 2014*; Ball et al.,* 2017, 2018). The regressors used for DLM are similar, though not identical, to those used in the "LOTUS regression" model but were kept identical to the analysis of *Ball et al.* (2017, 2018). The regressors include: A solar proxy (30 cm radio flux), a volcanic proxy (latitude dependent surface area density (SAD), based on *Thomason et al.*, 2018), two QBO proxies (30 hPa and 50 hPa wind fields as provided by the Freie University Berlin), and an ENSO proxy (Nino 3.4 HadSST). Seasonal cycle components, AR2 processes, and residuals are estimated together with these regressors, as well as the non-linear background trend. This non-linear background trend replaces the use of ILT, PWLT, or EESC and does not require an assumption about inflection dates, only a prior assumption about the smoothness of the non-linear background changes being estimated, which is determined from the data itself (see *Laine et al.* (2014); for further details, and *Ball et al.* (2017) for minor changes to the DLM algorithm used here). Because the background changes are non-linear, quoting a percent per decade trend is not appropriate with DLM, so nominally the overall change between two chosen dates is quoted (see for example **Figures S4.1**, **S4.2**, and **S4.3**), although the inferred non-linear background trends provide richer information about the long-term changes than a linear trend or a net change between two dates.

Another alternative to MLR is the application of a wavelet transform (WT). This method is widely used to analyse time series that contain non-stationary power at different frequencies and has been used more and more in geophysical and climatological studies (*Zitto et al.,* 2016). It allows an analysis which provides information not only on the frequencies present in the time series but also the times when the different frequency ranges are present in the sample. WT shows less sensitivity than PWLT to the choice of inflection point and therefore represents a promising alternative to MLR.

The method of empirical mode decomposition (EMD) is also well suited to deal with evolving trends over time (*Bai et al.,* 2017) and therefore as an alternative to the MLR approach. EMD decomposes any given signal into a finite number of intrinsic mode functions (IMFs), that represent simple oscillatory modes with varying frequency and amplitude along the time horizon, and a residual. This residual is then a monotonic or curved time series out of which the "trend" can be extracted. EMD is therefore not dependent on the length of data records for determining trends and is less vulnerable to outliers in the time series than MLR (*Bai et al.,* 2017).

4.5 The "LOTUS regression" model

4.5.1 General description

Based on the findings of the sensitivity tests presented in the earlier sections of this chapter, the LOTUS community agreed on one common regression model (the "LOTUS regression" model) to be used for all analyses that are presented in *Chapter 5*. The final choice of proxies and possible lags of proxies was based on finding the optimal regression for global analysis of satellite data and broad latitude band analyses. Therefore, proxies describing rather local or small-scale phenomena might not have been included in the general "LOTUS regression" model. To facilitate the comparison between satellite-based and ground-based ozone trends, the same regression model was applied to the ground-based (station) data, although proxies describing local and small-scale ozone variability might have improved the overall regression performance. Effects of this limitation in proxies for station data still need to be investigated further.

The "LOTUS regression" is applied to the ozone values without weights and so a correction for heteroscedasticity (*i.e.*, the non-constant variance in the data; see *Section 4.3.2*) is not applied. For data sets that are not already deseasonalised, Fourier components representing the seasonal cycle are also included (four sine and cosine pairs, representing the 12, 6, 4, and 3 month periodicity). No seasonal cross-terms are included in the "LOTUS regression" model as to mitigate the introduction of multicollinearity and to avoid inconsistencies between the treatment of data sets that are and are not intrinsically deseasonalised. A lag-1 autocorrelation correction was included in the regression model.

The "LOTUS regression" model uses the ILT proxy as a trend term (see *Section 4.2.2*). Additionally, it includes two orthogonal components of the QBO, the solar 10.7 cm flux, ENSO without any lag applied, and the GISS AOD. This aerosol data set was extended past 2012 by repeating the final available value from 2012 as the background AOD. To perform the ILT in a single step, the trend proxies included are:

- A linear increase until January 1997 and zero afterwards
- Zero until January 2000 and a linear increase afterwards
- Constant until January 1997 and zero afterwards
- Zero until January 2000 and a constant afterwards
- Constant between January 1997 and January 2000 and zero elsewhere

Including these proxies allows the ILT to be performed in a single step rather than the two step procedure used in *Steinbrecht et al.* (2017). The result of the regression is to obtain the coefficients A-J that correspond to the equation:

$$\begin{aligned} y(t) = {} & A \cdot QBO_1(t) + B \cdot QBO_2(t) + C \cdot ENSO(t) \\ & + D \cdot AOD(t) + E \cdot Solar(t) + F \cdot Linear_{pre}(t) \\ & + G \cdot Linear_{post}(t) + H \cdot C_1(t) + I \cdot C_2(t) + J \cdot C_3(t) \\ & + \varepsilon(t) \end{aligned} \quad (4.5),$$

where C_1 to C_3 are the three constant terms described above.

The "LOTUS regression" model has been implemented in the Python programming language. It is designed to be a flexible software package to both perform the sensitivity tests of *Section 4.3* and to run the final chosen models on the wide variety of data sets present within the LOTUS initiative. The software package and up to date documentation are available at **https://arg.usask.ca/docs/LOTUS_regression**.

4.5.2 Application to model simulations

In order to maintain comparability in the interpretation of results, we performed the analysis of trends in the vertical distribution of ozone from the CCMI-1 REF-C2 simulations (see *Chapter 2*, *Section 2.3*) using the same ILT method as for the observations. Linear trends for the pre-1997 (Jan 1985 – Dec 1996) and post-2000 (Jan 2000 – Dec 2016) periods were calculated at each grid point (*i.e.*, latitude and pressure level) of the models (and the separate ensemble simulation members). Since the models' simulated atmospheric conditions (and composition) differ from the observations, we calculated the appropriate proxies (predictors) that are included in the statistical trend analysis using model parameters. Thus, for each model/ensemble member we first calculated the QBO index, performing an EOF analysis on the simulated zonal winds at the equatorial region. Then we used the first two EOF terms as QBO1 and QBO2 indices. The ENSO index was calculated from the simulations' SSTs over the tropical Pacific, over the exact same area where the Nino3.4 index is calculated. As before, the data were deseasonalised over the period of 1998–2008.

Finally, the regression analysis was performed at the given pressure levels, using the following proxies: (1) two trend terms (identical to the method described earlier as the ILT method, *Section 4.2.2* and *Section 4.3.4)*; (2) two QBO terms (calculated as described above), in the case of models not simulating the QBO this proxy was not used; (3) one term for the ENSO effect (as described above); (4) one term for the solar forcing (we used the forcing as it was provided to the modelling groups; it should be noted that in our case the last five years of the solar forcing data are slightly different from the observations, not in terms of phase but in magnitude); and (5) one term for the volcanic effect (AOD; the same basis function that was used for the analysis of the satellite- and ground-based measurements was also used for the models). Normally the CCMI-1 REF-C2 model simulations do not include volcanic eruptions, but the effects can be present via different routes, for example SSTs or winds. All model results that are shown in *Chapter 5* (*Sections 5.2* and *5.5*) are shown as percent changes over the base period 1998–2008.

4.6 Summary

This chapter discusses several sources of uncertainties and sensitivities for trend analyses with MLR. First we tested existing regression models within the community by applying them to a common data record and comparing the resulting trend estimates. We found a general spread in derived trend values of 1–2 % per decade, with some differences as high as 3 % per decade. Next, we completed a series of sensitivity tests in an effort to identify the proxies that have the largest effect on the derived trend, leading to the observed spread in trend results across different regression models. Several sensitivity tests for proxy selection, proxy combination, and unweighted/weighted regression approaches were performed to understand their effects on the derived trend values and trend uncertainties in multiple merged satellite ozone records. We found the proxies AO, AAO, NAO, and EHF have only negligible effects on trends and significances, but excluding the QBO, solar, or ENSO proxies from the regression model had significant effects on the trend (1–2 % per decade difference) and uncertainty (around 1 % per decade) estimates. The three different trend proxies (PWLT, ILT, and EESC-based EOFs) produce generally very similar trend estimates. However, in the sensitivity tests performed here, subtle differences for the results of the trend proxies were found; for the pre-1997 trend estimates ILT produces the smallest (least negative) trend and EESC the largest trend, whereas for the post-2000 trend estimates the trend proxies behave exactly opposite. PWLT trends were shown to be affected most by end point problems caused by the chosen inflection point (and therefore the length of the analysed time series).

Based on these sensitivity tests, a "LOTUS regression" model was developed that includes two QBO proxies, a solar proxy, an ENSO proxy without any time lag applied, a stratospheric aerosol proxy, and the ILT as the trend proxy. Four Fourier components representing the seasonal cycle are also included. The "LOTUS regression" model is unweighted, and it includes a lag-1 autocorrelation correction. A detailed description of this regression model and its source code is publicly available on the LOTUS website, **https://arg.usask.ca/docs/LOTUS_regression**.