Chapter 5: Time series and trend results

Hassler, B., Damadeo, R., Chang, K.-L., Sofieva, V. F., Tourpali, K., Frith, S. M., Ball, W. T., Degenstein, D. A., Godin-Beekmann, S., Hubert, D., Maillard-Barras, E., Misios, S., Petropavlovskikh, I., Roth, C. Z., Steinbrecht, W., Vigouroux, C., von Clarmann, T., Zawada, D. J., Zerefos, C. S., Alsing, J., Balis, D., Coldewey-Egbers, M., Eleftheratos, K., Gruzdev, A., Kapsomenakis, J., Laeng, A., Laine, M., Taylor, M., Weber, M. and Wild, J. D. (2019) Chapter 5: Time series and trend results. 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. 51-70. 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/82512/

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

Publisher: SPARC

Chapter 5: Time series and trend results

The ultimate goal of LOTUS is to improve confidence in calculated ozone trend values via an improved understanding of the uncertainties. *Chapter 3* highlighted many of the challenges facing analyses of long-term ozone time series, and despite the fact that many of those challenges still need to be addressed, it is worthwhile to assess the trend results from this work in such a way as to be able to place those in the context of previous work. This chapter highlights the results of taking the "LOTUS regression" model from *Chapter 4* and applying it to the different data sets (*i.e.*, satellite, ground, and model) at different resolutions comparable to those in previous ozone assessments and comprehensive studies (*e.g.*, *WMO,* 2014; *Harris et al.,* 2015; *Steinbrecht et al.,* 2017). The individual satellite-based trend results are then combined to obtain a single mean ozone trend profile with respective uncertainty estimates. This important yet challenging final step in the assessment has been the cause of debate in the community in recent years. Different methods for combining the individual trend results are discussed and explained, and the final trend profile estimates are analysed for significance.

5.1 Satellite trends at native resolution

The regression model was applied to all satellite data sets described in *Chapter 2*, for all latitude bands and all vertical levels. In this section, only eight of the ten satellite data sets are discussed. The SAGE-OSIRIS-OMPS data set without the sampling-corrected SAGE data is excluded for reasons discussed in *Section 3.2.1.* Also, for reasons summarised in *Section 4.5.1*, *Chapter 5* only discusses results for the ILT "LOTUS regression" model. However, for the sake of completeness, the trend results for all ten data sets and with each trend proxy (*i.e.*, PWLT, ILT, and EESC EOFs) are shown in the *Supplement* (see **Figures S5.1** through **S5.6**).

5.1.1 Trend results

Figure 5.1 shows the trends derived for the pre-1997 period (this period covers January 1985 to December 1996; for simplicity it is called 'pre-1997'), for all latitude bands, all vertical levels, and all satellite data sets as well as for the CCMI-1 REF-C2 model results for comparison purposes. The general pattern of negative (4–9 % per decade) trends in the upper stratosphere (above 5 hPa / 35 km) is present in most of the satellite records (with SBUV MOD showing the smallest and SWOOSH showing the largest trends) and is consistent with previous findings (*i.e.*, *WMO,* 2014; *Harris et al.,* 2015; *Steinbrecht et al.,* 2017; and references therein). These upper stratospheric trends show a minor hemispheric asymmetry in several of the data sets, with larger (*i.e.*, more negative) trends in the Southern Hemisphere (SH) mid-latitudes than the Northern Hemisphere (NH) mid-latitudes. Pre-1997 trends in most records show positive values at 7–10 hPa (or 30–33 km) in the tropics, with slight differences between the trends derived in the SH and the NH for some data sets, though these are not statistically significant. Also, slightly positive trends, which are also not statistically significant, are found in all data sets in the lower stratosphere in NH subtropics (30–50 hPa or 20–25 km). SBUV COH is the only data set with a significant positive trend in the SH at mid-latitudes around 20 hPa (this result and the less negative SBUV MOD trend noted above are discussed in *Section 5.1.2*; see also **Figure S5.8**).

The time period of most interest is after 2000, when an ozone recovery is expected at upper stratospheric levels. **Figure 5.2** summarises the analyses for the post-2000 period (this period covers January 2000 to December 2016, or in the case of BASIC January 2000 to December 2015 (see *Chapter S4.1* in the Supplementary Material for more detail), and it is called 'post-2000' for simplicity). Overall, the results from the different data sets seem to agree on positive trends (~2–3 % per decade on average) in the mid-latitudes of both hemispheres between 5 hPa and 2 hPa (around 37 km to 45 km). As in the pre-1997 period, most of these upper stratospheric trends show a minor hemispheric asymmetry, with more positive trends in the NH than in the SH. Some asymmetry is also found in middle and lower stratospheric trends across the 20°S–20°N tropical band, including a more pronounced difference between the subtropical SH and NH. This trend asymmetry and its impact on the combined broad-band trends is noted in *Sofieva et al.* (2017), *Steinbrecht et al.* (2017), and *Zerefos et al.* (2018), and is further discussed in *Section 5.4* with respect to the representativeness of the ground-based records. However, the magnitude of these trends and their significances vary between the data sets. Out of the eight data sets, three have OMPS measurements contributing to the merged record beyond 2012 (**Table 2.2**), though the magnitude of the trend results is influenced largely by the record that is used between 2000 and 2005. The OSIRIS-based data sets produce larger recovery trends than the MIPAS-based data set, which may be related to the positive drift of OSIRIS in the upper stratosphere (see discussion of the record stability in *Section 3.1.2*). SBUV COH also appears to produce more significant recovery estimates for the upper stratosphere as compared to the SBUV MOD record (see also next section). In the middle to the lower stratosphere, SBUV-based data sets and the BASIC data set show more negative trends across all latitudes compared to GOZCARDS or SWOOSH, though the vertical resolution of SBUV is significantly reduced in this altitude range.

Additionally, SWOOSH limits negative trends to the tropical region only. OSIRIS-, CCI-, and MIPAS-based data sets show smaller negative trends in the lower stratosphere that are not clearly confined to a specific region and, in most cases, are not significant.

5.1.2 Discussion of differences

There are several possible reasons for the different trend patterns of the eight satellite data sets shown in **Figure 5.1** and **Figure 5.2**. In the following sections we discuss some of the reasons that could contribute to the trend differences based on findings from *Chapter 3* and relevant literature, such as the merging method of the data sets, application of a sampling bias correction, and conversions between different unit and grid systems to allow the construction of merged data sets.

Merging method

As detailed in *Sections 2.2.1* and *3.1.4*, although the SBUV MOD and SBUV COH are constructed from the same suite of instruments, different choices of instruments and different merging techniques allow for different, but equally valid, merged records. In the pre-1997 period, the nadir-based trends vary significantly from the limb-based trends and from each other. Data quality of SBUV instruments operating in the mid-1990s is negatively affected by on-orbit instrument problems and drifting orbits, resulting in larger intra-satellite biases and drifts during this period. SBUV MOD and SBUV COH both rely on the same instrument during the 1980s but diverge in the 1990s (see for example **Figure 5.5**, 20 hPa). The sensitivity to the mid-1990s data is enhanced when fitting over the shorter 1985–1996 time period. Furthermore, data after the eruption of Mt. Pinatubo (mid-1991–1993) are treated differently in the two merged records (*Section 2.2.1*). Estimates by *Frith et al.* (2017) suggest uncertainties of 10–15 % (2-sigma) from 1995–2000 in the merged records, suggesting trends fit over this time period for either SBUV merged record are highly uncertain. However, trends from SBUV COH and SBUV MOD computed using the "LOTUS regression" model over the longer 1979–1996 time period compare notably better than those computed over the 1985–1996 period (**Figure S5.8** in the

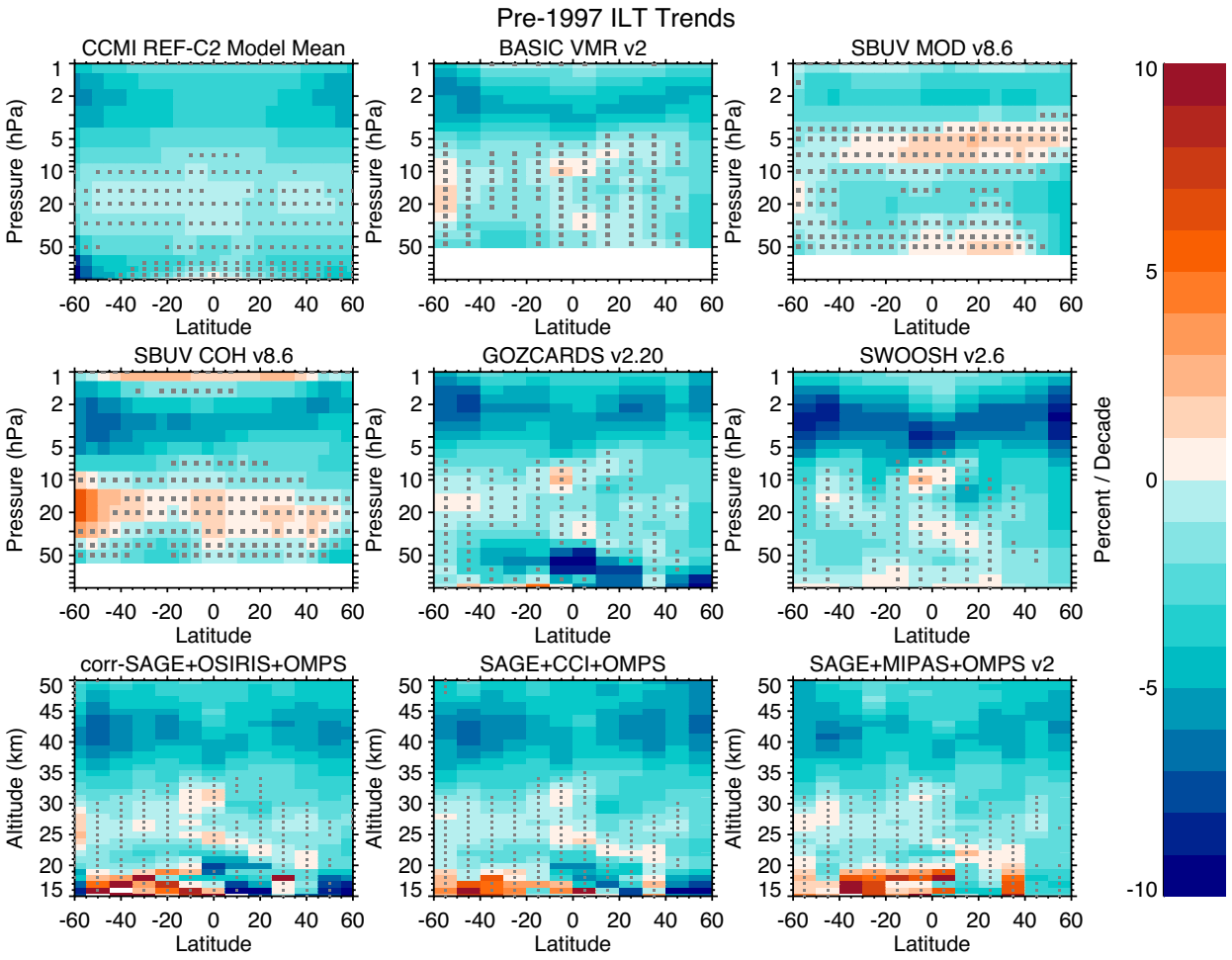

Figure 5.1: *Derived trends in satellite ozone in percent per decade for the pre-1997 period (Jan 1985 – Dec 1996) for each of the satellite data sets, using the ILT trend proxy in a regression analysis. Grey stippling denotes results that are not significant at the 2-sigma level. Data are presented on their natural latitudinal grid and vertical coordinate. For comparison, the mean of trends derived from CCMI participating models is included in the upper left panel. Results for other trend proxies can be found in the Supplement.*

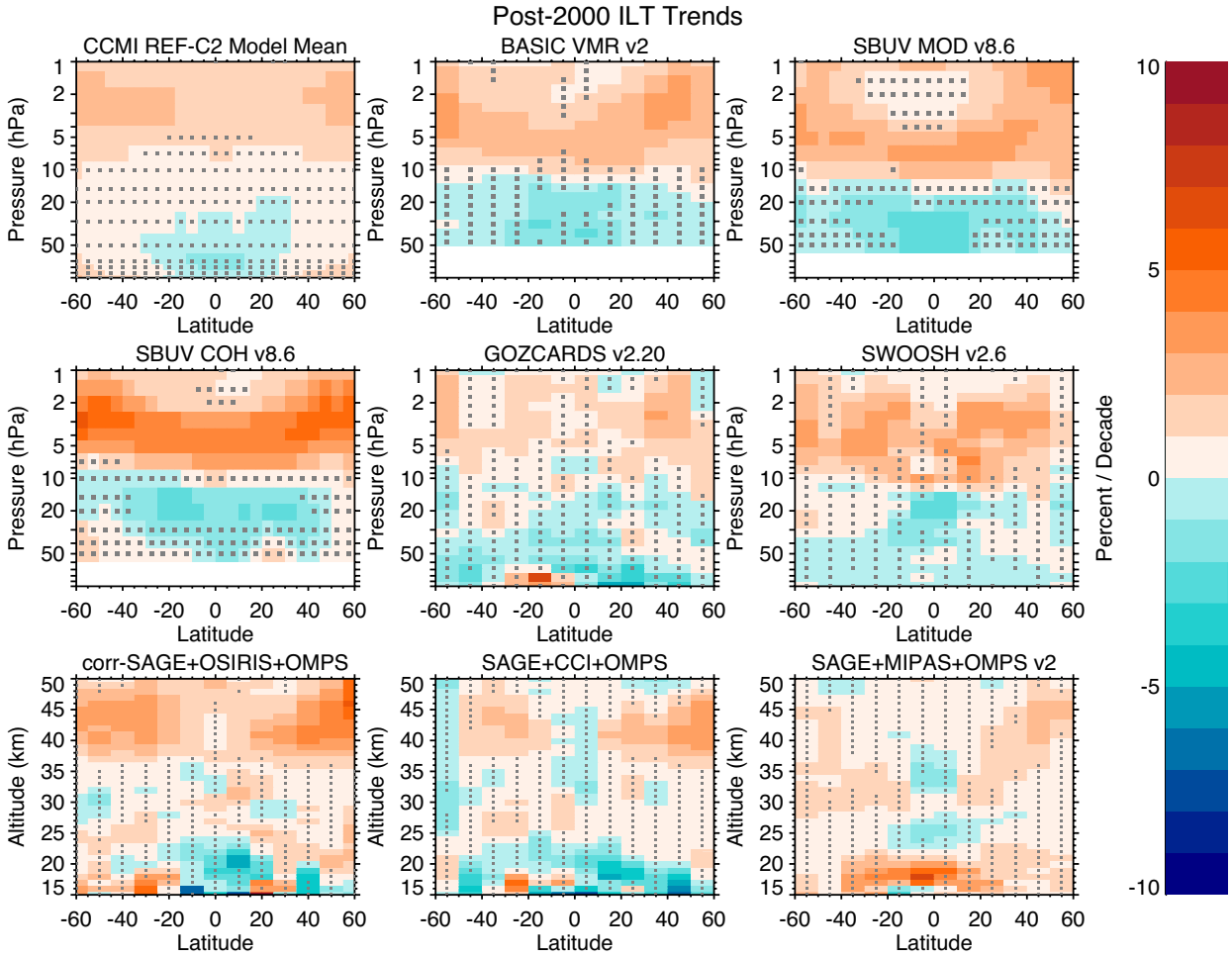

Figure 5.2: *Derived trends in satellite ozone in percent per decade for the post-2000 period (Jan 2000 – Dec 2016) for each of the satellite data sets, using the ILT trend proxy in a regression analysis. Grey stippling denotes results that are not significant at the 2-sigma level. Data are presented on their natural latitudinal grid and vertical coordinate. For comparison, the mean of trends derived from CCMI participating models is included in the upper left panel. Results for other trend proxies can be found in the Supplement.*

Supplement). Trends from both records also compare well with trends from GOZCARDS fit over the same time period (**Figure S5.8**, see also *Harris et al.*, 2015). Trends over the post-2000 period tend to be positive at NH and SH mid-latitudes between 5 hPa and 2 hPa. SBUV MOD trends are more positive at altitudes just above the 10 hPa level, particularly in the tropics, while the SBUV COH trends are more positive between 5–3 hPa at all latitudes. Because of their respective merging techniques, SBUV MOD is sensitive to successively increasing or decreasing biases in the instrument series that might alias into a trend while SBUV COH is sensitive to drifts in the reference instruments that might be propagated to other periods in the record. As discussed in *Section 3.1.4*, the higher trends in SBUV MOD are related to a positive bias in NOAA-19 data at the end of the record, while the higher trends in SBUV COH above 5 hPa are partially caused by a small drift in the NOAA-18 data used as a reference for the merged data set. *Frith et al.* (2017) used a MC approach to estimate a merging uncertainty for SBUV MOD and showed that SBUV MOD and SBUV COH trends agree within the estimated 2-sigma merging uncertainty (see **Figure 3.9**).

When combining measurements from a multitude of different instrument types (such as in GOZCARDS and SWOOSH), sources of potential differences in trends derived from these data sets increase even more. As shown by *Tummon et al.* (2015), although GOZCARDS and SWOOSH are based on almost the same data sources, differences exist for example in the annual cycle of the tropical stratosphere and in monthly mean anomalies of the SH mid-latitude stratosphere. These differences are caused by different merging approaches for the selected data sets and can have an impact on the estimated trends (for more information see *Tummon et al.*, 2015; *Harris et al.*, 2015; and *Section 2.3.3*). Both data sets, GOZCARDS and SWOOSH, have been updated since the analyses described in *Tummon et al.* (2015) and *Harris et al.* (2015), but the differences in merging technique remain and with that the possible effects on trend estimates.

Small differences in pre-1997 trends from the OSIRIS-based data set in comparison to the CCI-based data set are largely consistent with expected differences due to the use of sampling corrected SAGE data in the OSIRIS-based record (see next section).

However, small remaining differences, as well as larger differences with the MIPAS-based record (which also includes only SAGE data in the pre-1997 period), suggest differences in the latter part of the record are due to the merging technique, and data set selection can impact the pre-1997 trends through fits to other large-scale proxies such as the solar cycle. Differences up to 2 % per decade found in the post-2000 trends in the upper stratosphere are likely influenced by differences in the CCI and OSIRIS records and the respective merging methods, as the same version of OMPS-LP data is used at the end of the combined records. Upper stratospheric ozone from the OMPS-LP record compares well with Aura MLS over mid-latitudes (see **Figure 3.3** in *Chapter 3*) but shows differences over tropical regions (see **Figure 3.4** in *Chapter 3*). Also OMPS-LP and Aura MLS differences are found over the middle stratosphere in the tropics, which likely explains the difference in tropical trends derived from SAGE-CCI-OMPS and SAGE-OSIRIS-OMPS as compared to GOZCARDS and SWOOSH. Drift in the OMPS-LP record relative to lidar, microwave, and ozonesonde records (see **Figure 3.5** in *Chapter 3*) can also explain the stronger positive trends in the upper stratosphere in the combined datasets that incorporate the OMPS-LP record.

The significant differences in derived trends between SBUV and BASIC despite the inclusion of SBUV data in BASIC is likely a result of the merging technique applied to create the BASIC data set, which reduces the weight of SBUV measurements over this time period due to known instrument issues.

Sampling bias correction

The difference in evaluated trends due to the sampling bias correction of SAGE II data can be as large as 2 % per decade for the pre-1997 period and up to 1 % per decade for the post-2000 period, as illustrated in **Figure 5.3**. While the absolute differences in trends are larger for the pre-1997 period, which are derived from satellite data with coarse spatio-temporal sampling, the influence is more pronounced for the post-2000 trends as the trend differences represent a larger fraction of the trends themselves. For example, the largest impact in pre-1997 trends is found in the upper stratosphere, with typical differences of ~1 % per decade over the mid-latitudes (compared to the -6 % per decade trend), with stronger increases in trends over the NH. However, the post-2000 trends show mixed/positive differences up to 1 % per decade in the SH/NH respectively that represent as much as half of the trends there.

Conversion between different unit and grid systems

Stratospheric cooling was observed in the upper stratosphere in the 1980s and beginning of the 1990s (*Steinbrecht et al.,* 2009; *Randel et al.,* 2009), which led to a difference of 2 % per decade in trends depending on the measurement unit and vertical scale (*McLinden et al.*, 2011; also see **Box 2.1** in *Chapter 3*, *WMO,* 2014). *McLinden et al.* (2011, **Figure 3**) showed that trends derived from ozone records prior to 1997 differed, such that VMR trends computed on pressure surfaces (*i.e.*, GOZCARDS, SWOOSH, SBUV, BASIC) were found to be less negative than number density trends computed on altitude surfaces (*i.e.*, CCI or OSIRIS-based combined records; see also *Harris et al.,* 2015). These differences exist only in the presence of temperature trends due to continuous changes in conversion between pressure and altitude scales. Since the late 1990s, only small cooling trends are observed in the upper stratosphere (*e.g.*, *Thompson et al.,* 2012; *Randel et al.,* 2016), and thus the influence of unit and vertical coordinate representation on post-2000 ozone trends is expected to be small. In addition, even though all the pre-1997 limb instrument trends are based primarily on SAGE II data, the differences in trends between the different coordinate representations in the merged limb records do not match expectation (**Figures 5.1** and **5.6**). For example, the pre-1997 trend in NH for GOZCARDS (VMR on pressure) is close to that of SAGE-CCI-OMPS and SAGE-OSIRIS-OMPS (number density on altitude) but differs significantly from the SWOOSH (VMR on pressure) trend. This suggests that the influence of temperature changes on the agreement of ozone trends derived in different unit/coordinate representations is either smaller than expected or obscured by other sources of uncertainty.

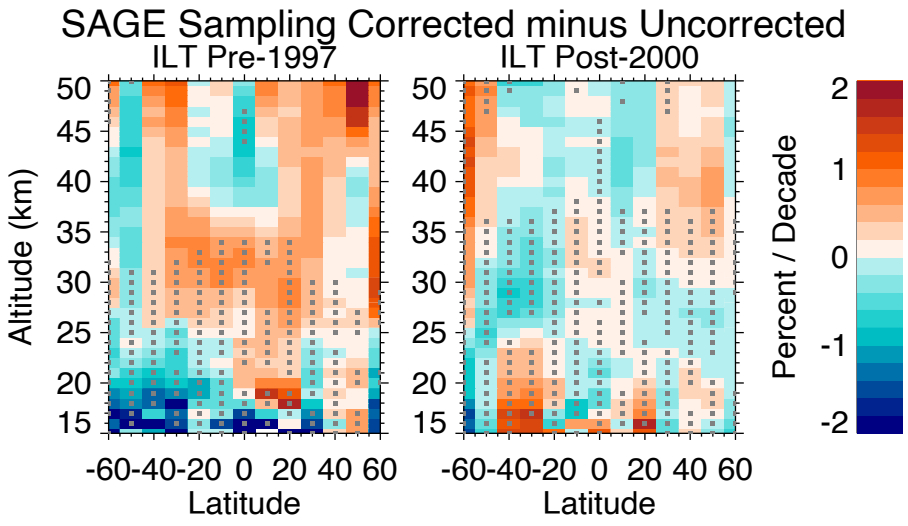

Figure 5.3: *Impact of sampling bias correction on ozone trends derived from the SAGE-ORIRIS-OMPS data set. Grey stippling denotes where the trends derived from either version of data (i.e., sampling-corrected or not) were not significant at the 2-sigmal level.*

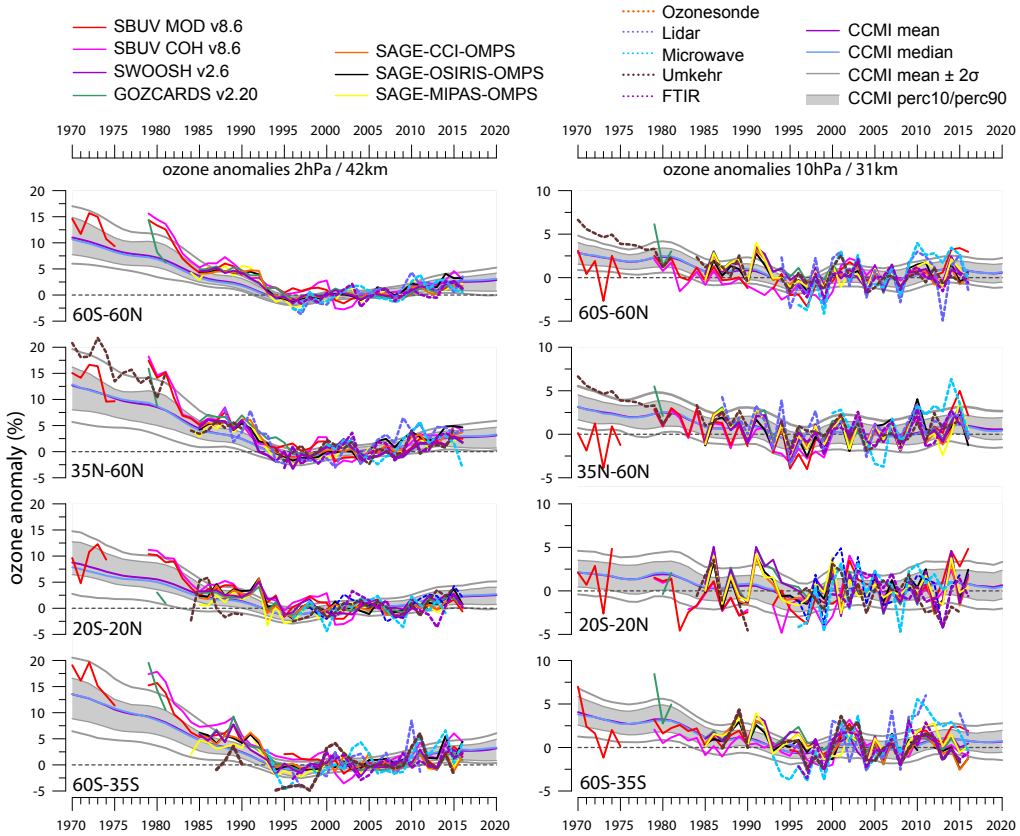

Figure 5.4: *The evolution of ozone changes as annual mean anomalies at the 2 hPa/42 km (left panel) and 10 hPa/31 km (right panel) levels. Three different latitude bands are shown. Satellite data are based on zonal means, and ground-based stations are averaged over the latitude bands. The grey "envelope" gives the CCMI-1 model results, based on the models' 10th and 90th percentile. The model mean and the median are also plotted together with the ±2 standard deviation range of the models. All anomalies are calculated over the base period 1998–2008, and the CCMI-1 models are shown as 1-2-1 year filtered averages (see text).*

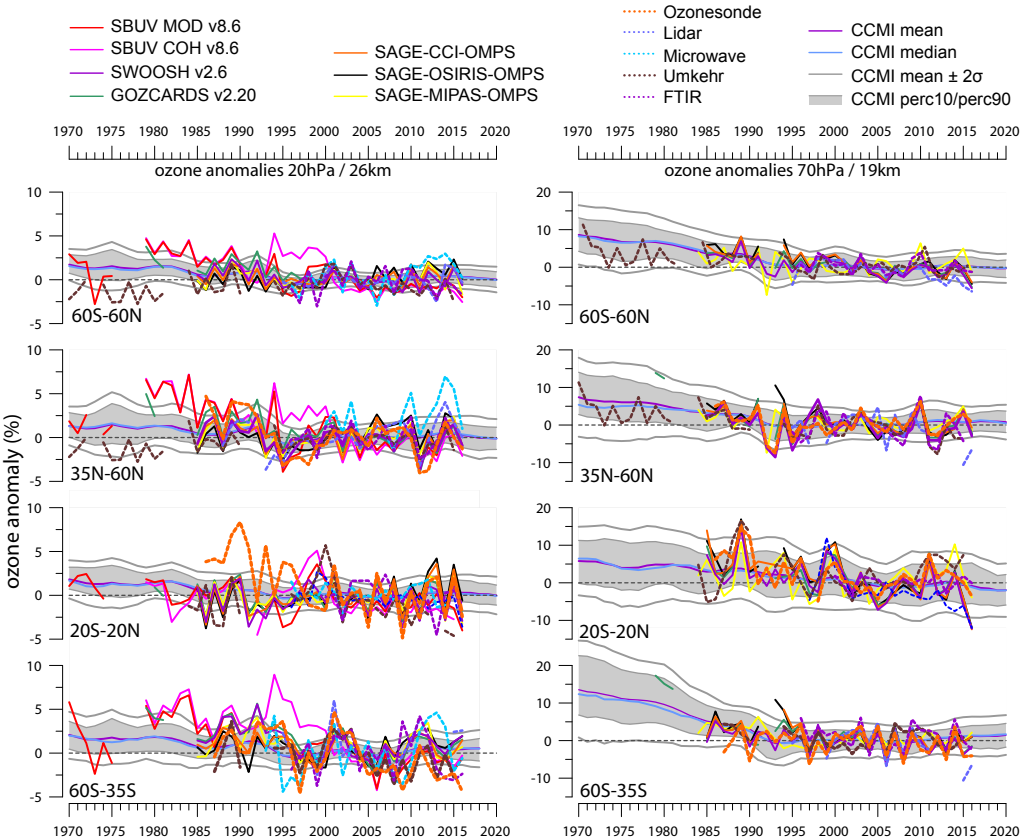

Figure 5.5: *Same as* ***Figure 5.4*** *but for the 20 hPa/26 km (left panel) and 70 hPa/19 km (right panel) levels.*

5.2 Time series in broad latitude bands

While much of the literature has analysed stratospheric ozone trends from individual data sets at their native resolution, the more expansive works (*e.g.*, *WMO,* 2014; *Harris et al.,* 2015; *Steinbrecht et al.,* 2017) have investigated trends in broad latitude bands. The three cited studies selected three broad bands: the SH at mid-latitudes (60°S–35°S), the tropics (20°S–20°N), and the NH at mid-latitudes (35°N–60°N). To place the results of LOTUS in the context of those works, the updated/modified data sets have been analysed in these broad latitude bands. An average of near global ozone profiles (60°S–60°N) have also been reported before; however, these are not the main focus here and will therefore only be shown in the *Supplement* (**Figure S5.7**).

The data sets are not natively provided for these broad latitude bands, so they must first be converted (for further details see *Sections 2.1.2*, *2.2.5*, and *2.3.2*). **Figures 5.4** and **5.5** illustrate annual mean deseasonalised anomaly time series for satellite, ground-based and model data for the three broad bands at four selected altitude or pressure levels (indicated at the top of each panel). These annual means are shown here only for illustration purposes; the regression results in *Section 5.3* are based on monthly mean anomaly data. The CCMI-1 model time series were further smoothed using a 1-2-1-year filter (described in *Section 2.3.2*). The model median (blue curve) and the model mean (dark pink) are shown with the lower 10th-percentile and the upper 90th-percentile forming the grey envelope of the CCMI-1 models' range. Grey curves surrounding the time series denote the range of CCMI-1 model mean ±2 standard deviations. Time series constructed from the satellite data and ground-based station data show interannual variability caused by natural variations such as volcanic eruptions and QBO (most pronounced in the plots for the middle and lower stratosphere), whereas in model results natural variability is smoothed out for plotting purposes (see *Section 2.3.2* for details).

Figure 5.4 shows the evolution of ozone changes in the upper stratosphere (2 hPa/42 km, left panel). The ozone decline due to the increase in ODSs up until the mid-1990s is evident, as expected. Since then, a leveling off can be seen, with the ozone values after about 2005 indicating an increase, which is most prominent in the NH mid-latitude band (35°N–60°N). Lower levels are presented in **Figure 5.4** (right panel) and **Figure 5.5**. In general, the ODS-related long-term evolution of ozone can be seen at these levels but is less pronounced than in the upper stratosphere.

The REF-C2 simulations by CCMI-1 models are capable of capturing the long-term variations in ozone, that is the solar cycle and the trends. The mean and median of the CCMI-1 REF-C2 multi-model simulations are smoother and their uncertainty ranges narrower than the anomalies calculated from observations. This is expected because the REF-C2 simulations do not include volcanic variability, and variability from forcings such as QBO and ENSO will tend to cancel in the average as their

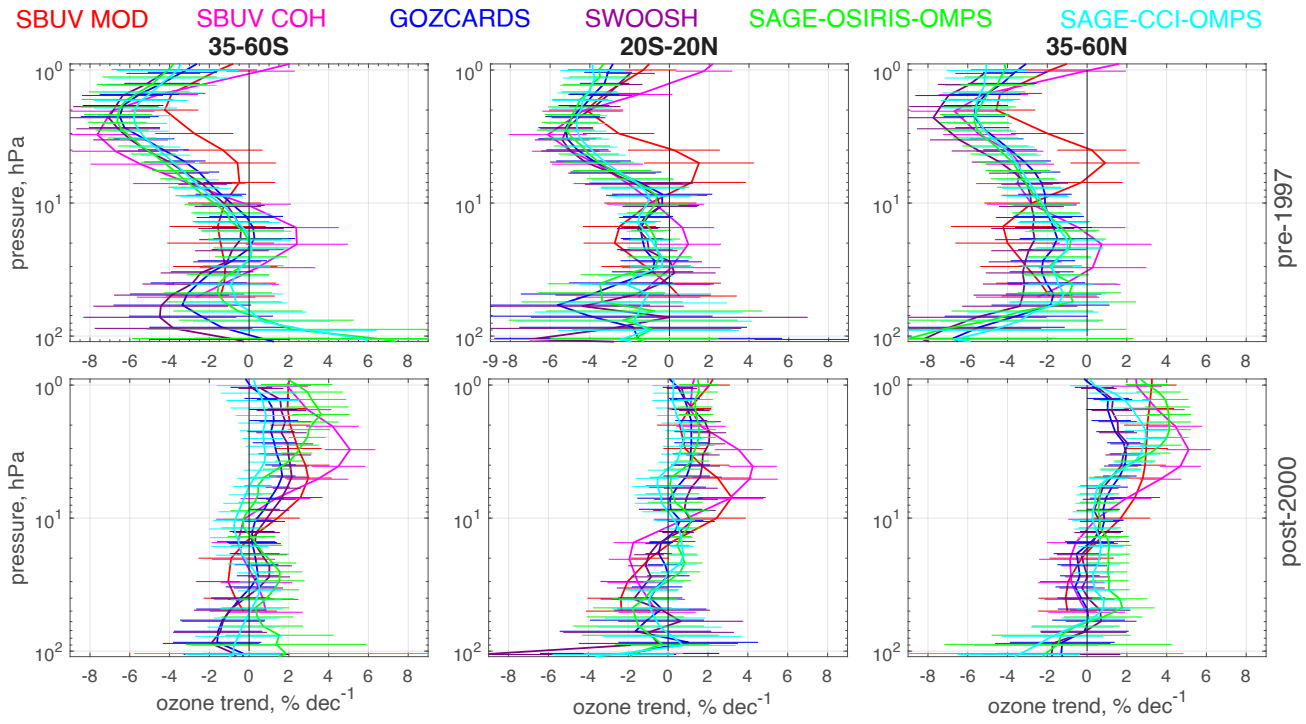

Figure 5.6: *Ozone trends with 2-sigma uncertainties for the pre-1997 (top) and post-2000 (bottom) period from the ILT regression for latitude bands 60°S–35°S (left), 20°S–20°N (centre), and 35°N–60°N. Coloured lines are the trend estimates from individual merged data sets on their original vertical grid.*

phases will vary among individual simulations. A year-to-year direct comparison between models and observational data with respect to the natural variability can therefore not be made here. Making such a comparison would require model simulations that are tied to real-world observations, such as simulations using specified dynamics or the REF-C1 runs, which were based on observed SSTs and aerosol loading. Though these runs exist, the output does not cover the entire time period analysed by LOTUS (1985–2016). The model mean only represents the range of ozone variability due to the longer term evolution of ODSs and GHGs. The lack of volcanic eruptions and other natural forcings such as solar and QBO is also the main difference between this work and the relevant figures in the 2014 WMO Ozone Assessment (*WMO,* 2014). The grey shading given by the CCMVal-2 models in the lower stratosphere shown in **Figures 2-7** and **2-8** of the 2014 WMO Ozone Assessment (*WMO*, 2014) is wider than the 10/90 percentile range presented in **Figure 5.4** and **Figure 5.5**, since the CCMVal2-based large model variability is caused mainly by volcanic eruptions.

Overall both models and observations (satellite and ground-based) follow the same evolution of ozone changes in time and at the various altitude/latitude bands. The relative differences between individual observational time series are larger during the earlier years and at the 20 hPa level. However, the ground-based measurements during the period prior to 1980 are represented only by a single Umkehr record at Arosa, Switzerland and thus may not be representative of the broad-band variability captured by the Nimbus-4 satellite. In addition, after mid-1972 the Nimbus-4 BUV data coverage was reduced due to instrument problems, adding significant noise to the record. The bi-annual variability in the observed ozone anomaly time series is associated with the QBO signal.

5.3 Combined satellite trends in broad latitude bands

5.3.1 Selection and preparation of data sets

For the evaluation of ozone trends in the three broad latitude bands (*i.e.*, 60°S–35°S, 20°S–20°N, and 35°N–60°N) we consider six of the eight merged data sets discussed in *Section 5.1*: SBUV MOD, SBUV COH, GOZCARDS v2.20, SWOOSH v2.6, corr-SAGE-OSIRIS-OMPS, and SAGE-CCI-OMPS. The SAGE-MIPAS-OMPS data set is now excluded because of concerns of larger discontinuities when switching between the instruments (see *Section 3.1.5*). This collection of data sets includes two merged data sets for each satellite data set group discussed in *Chapter 2* (*i.e.*, nadir mixing ratio profiles versus pressure, limb mixing ratio profiles versus pressure, and limb number density profiles versus altitude) and was chosen such that they are as independent as possible. For this reason, the BASIC data set, which combines four of the merged data sets in the above list (GOZCARDS, SWOOSH, SBUV MOD, and SBUV COH) is not used.

Broad latitude band trend results for individual merged data sets are derived by applying the ILT "LOTUS regression" to the relative deseasonalised ozone anomaly time series averaged over the three latitude bands as discussed in *Section 2.2.5*. However, these broad-band trends are in the native coordinate system and combining the trends requires the profiles to be expressed in the same vertical coordinate and at the same grid levels. The reference vertical scale used below is the pressure grid of GOZCARDS and SWOOSH. The SAGE-OSIRIS-OMPS and SAGE-CCI-OMPS altitude levels are first converted to pressure levels using the mean ERA-Interim altitude-pressure profile in the considered broad latitude bands and time period, and the resulting trend profiles are subsequently linearly interpolated to the reference grid. The SBUV data, on the other hand, are already on a pressure grid but one that is coarser than that of GOZCARDS, so the SBUV trend profiles are linearly interpolated to the finer reference grid.

Figure 5.6 shows the profile trends in broad latitude bands for each of the six selected data sets. Similar to **Figures 5.1** and **5.2**, the trends in **Figure 5.6** generally agree with each other with differences caused by the reasons discussed in *Section 5.1.2*. The 2-sigma error bars represent the trend uncertainty estimated by the regression model (using the fit residuals). These are very similar for limb data sets and are slightly larger for the nadir merged data sets.

5.3.2 Approach to combine trends

In order to facilitate comparison with prior studies or with model simulations, it is useful to combine the trend profiles, and their uncertainties, derived from the various observational records. Such an approach has been used in previous WMO Ozone Assessments (*e.g.*, *WMO,* 2014) and it provides an estimate of how confident we are in trend results from the global satellite observing system. Ideally, ozone time series are combined using the averaged anomaly from all available records such that a single trend can be derived (similar to, for example, the BASIC approach). However, this approach is time consuming and requires consideration of differences in the individual records pertaining to temporal and spatial sampling, stability, vertical coordinate system, different units, and vertical smoothing. Therefore, the most efficient approach for this Report was to calculate trends separately for each record and then combine those results. As mentioned in *Chapter 3*, tracking ozone recovery with multiple observations is important for redundancy, avoiding gaps in satellite operations, and avoiding impacts of drifts of individual records on the trend assessment.

While combining trends is somewhat straightforward (*i.e.*, usually computed as either an unweighted or a weighted mean), combining uncertainties is much more complicated.

One of the main challenges is to assess the independence of the trend results given that some data records (*e.g.*, SAGE II) are used multiple times and that one regression model and one set of proxies are used for all analyses. When considering the independence of the trends and their uncertainties, it is useful to consider the theoretical lower bound to the uncertainty of the regressed trend, which is defined not only by the length of the time series and the magnitude of the trend but also by the atmospheric variability not characterised by the regression model. Indeed, any variance that is not represented in the regression model will propagate into the fit residuals, which will, in turn, lead to larger uncertainties in the regression coefficients. Hence, even trends derived from an ideal data set (*i.e.*, one that is infinitely long with no sampling issues, drift, or need for merging) will have an uncertainty equal to this theoretical lower bound determined by the true, incompletely modeled atmospheric realisation. As a result, even the trend uncertainties of completely independent data sets are correlated when estimated with the same regression model.

We estimate the overall trend $\bar{t}$ as the unweighted mean of six trend estimates and evaluate its variance as the maximum of two variance terms:

$$\sigma_{mean}^2 = max\left(\frac{1}{N^2}\sum_{i,j} C_{ij}\sigma_i\sigma_j \;,\; \frac{1}{n_{eff}}\sum\frac{(x_i-\bar{x})^2}{N-1}\right) \quad (5.1),$$

where N is the number of observation records, C_{ij} are the correlation coefficients for the trend estimates x_i from data sets i and j , σ_i are the trend uncertainties estimated from the fit residuals for the individual data sets, and n_{eff} is the effective number of independent trend estimates. The first term in Equation (5.1), the quadratic form $\frac{1}{N^2}\Sigma_{i,j} C_{ij}\sigma_i\sigma_j$, is the variance of the mean of correlated values, obtained through traditional propagation of errors. The second term is the unbiased estimator of the standard error of the mean, where n_{eff} independent measurements are assumed from the N=6 different trend estimates, and can capture systematic biases in trend uncertainties between the different merged data sets that would not be captured by the first term (*e.g.*, as a result of random drifts between data sets or differing unit representations). The effective number of independent values n_{eff} in Equation (5.1) is approximated by

$$n_{eff} = \frac{N^2}{\sum_{i,j=1}^{N} C_{ij}} \quad (5.2).$$

The first term in Equation (5.1) serves as an approximation of the theoretical lower bound of trend uncertainty due to the actual realisation of the ozone time series. This special approach of using the maximum of both terms to estimate the combined trend uncertainty is done because the sample of trend estimates is small and because, in case the trend estimates within the sample coincide, the observed variance is not necessarily representative of the actual uncertainty of the combined trend. We do not use the sum of both terms in Equation (5.1) because the variance of the trend estimates in term 2 can be partly due to the uncertainties already represented in term 1, and thus using the sum would overestimate the uncertainty. However, in the case where some contribution to the standard error is independent of propagated uncertainties, the maximum of the two terms would underestimate the uncertainty. Realistically, separating out the overlap in the total uncertainty captured by each of the two terms in Equation (5.1) is impossible and, in practice, one of the two terms dominates such that the maximum is a reasonable approximation of the true uncertainty. **Figure S5.9** in the *Supplement* shows the uncertainty obtained by using both the max() and sum() approaches.

It is difficult to estimate the correlation of trend uncertainties as these correspond to the largest temporal scales in the ozone time series. In our analysis, we approximate the trend correlation by the correlation between the fit residual time series of the different regressions. These fit residuals include small-scale variations in the ozone time series from atmospheric variability that is not captured by the regression model (*e.g.*, from a sudden stratospheric warming event), which is likely to be highly correlated across all data sets, and any systematic biases in the data sets themselves (*e.g.*, from instrument behavioral anomalies), which can be highly correlated between merged data sets relying upon the same instrument. However, these correlations at short temporal scales do not necessarily translate to equivalent correlations at long temporal scales and so the correlations of the fit residuals likely overestimate the correlations of the trend uncertainties. Since the true correlations between the trend uncertainties do not have a straightforward solution, we use the correlations of the fit residuals as a conservative upper bound instead.

Detailed results of a correlation analysis of the fit residuals are presented in the *Supplement* (see *Appendix B* with **Figures S5.10** to **S5.15**). The matrices of the correlation coefficients and of the fit residuals for pre-1997 and post-2000 trend estimates are:

$$C_{pre}^{limb} = \begin{pmatrix} 1 & 0.8 & 0.4 & 0.4 & 0.4 & 0.4 \\ 0.8 & 1 & 0.4 & 0.4 & 0.4 & 0.4 \\ 0.4 & 0.4 & 1 & 0.9 & 0.6 & 0.7 \\ 0.4 & 0.4 & 0.9 & 1 & 0.6 & 0.6 \\ 0.4 & 0.4 & 0.6 & 0.6 & 1 & 0.8 \\ 0.4 & 0.4 & 0.7 & 0.6 & 0.8 & 1 \end{pmatrix}$$

$$C_{post}^{limb} = \begin{pmatrix} 1 & 0.9 & 0.6 & 0.6 & 0.6 & 0.6 \\ 0.9 & 1 & 0.6 & 0.6 & 0.6 & 0.6 \\ 0.6 & 0.6 & 1 & 0.95 & 0.6 & 0.65 \\ 0.6 & 0.6 & 0.95 & 1 & 0.65 & 0.7 \\ 0.6 & 0.6 & 0.6 & 0.65 & 1 & 0.82 \\ 0.6 & 0.6 & 0.65 & 0.7 & 0.82 & 1 \end{pmatrix} \quad (5.3).$$

The order of the data sets in these matrices is as follows: 1=SBUV MOD, 2=SBUV COH, 3=GOZCARDS, 4=SWOOSH, 5=SAGE-OSIRIS-OMPS, and 6=SAGE-CCI-OMPS. Using these correlation matrices and Equation (5.2) we obtain n_{eff}=1.6 for pre-1997 trends and n_{eff}=1.4 for post-2000 trends. Below ~50 hPa, where there are no SBUV data, n_{eff}=1.3 for both pre-1997 and post-2000 periods. The experimental correlation coefficients vary slightly with latitude and altitude (see *Supplement*) as a result of atmospheric variability, but these variations might not be

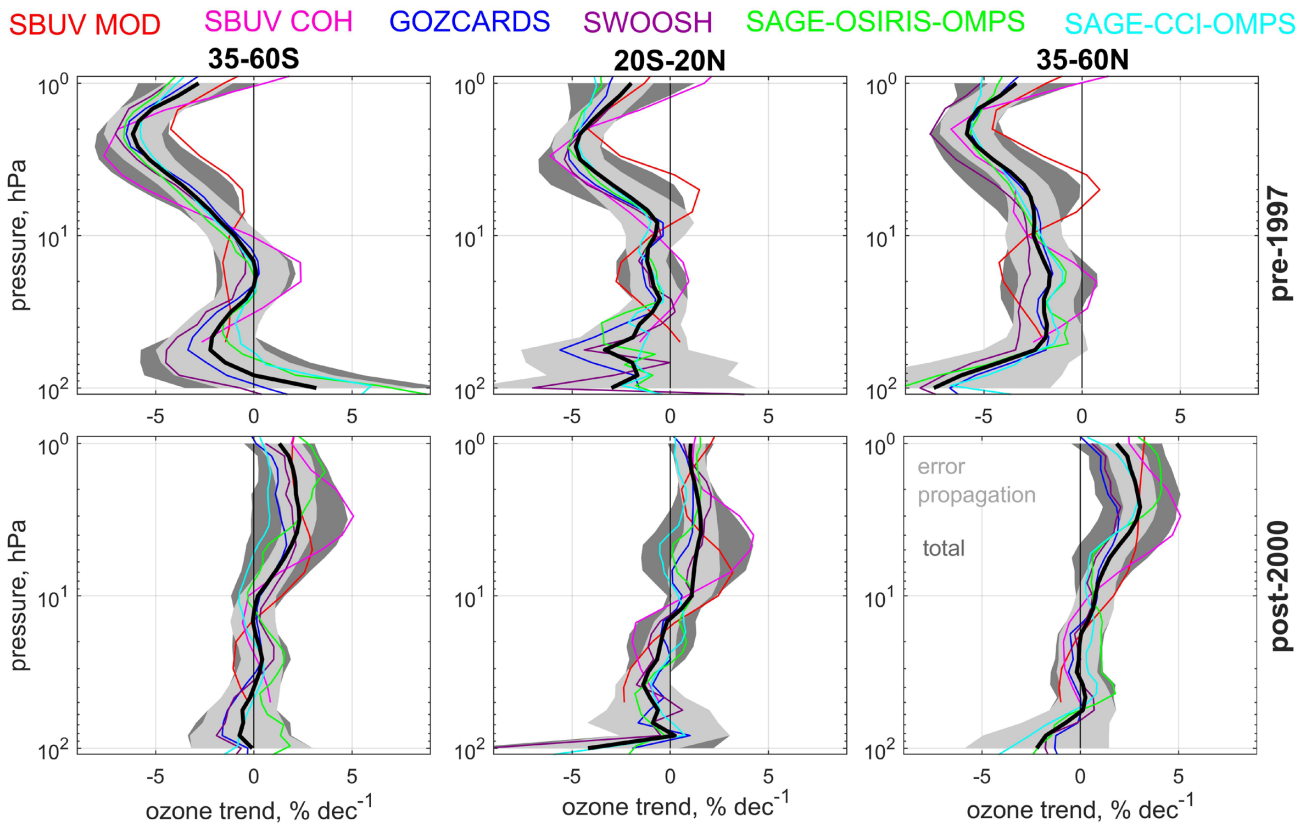

Figure 5.7: *Combining pre-1997 (top) and post-2000 (bottom) trend estimates and uncertainties (2-sigma) by Equation (5.1) from six limb profile data sets. Black solid line indicates the mean trend. The uncertainty component corresponding to error propagation (1st term in Equation (5.1)) is shown by light grey shading, while the total uncertainty is indicated by dark grey shading.*

related to the correlation of the trend estimates. Therefore, constant correlation coefficients are used for trend estimates. In any case, small variations in correlation coefficients result in insignificant changes in the effective number of independent trend values.

Figure 5.7 shows the mean overall trends and their uncertainties estimated using Equation (5.1). Light grey shading indicates the first term in Equation (5.1), which represents error propagation. In the lower stratosphere, this term generally dominates and represents the total uncertainty. In the upper stratosphere, the second term representing the variance of the sample mean dominates.

We think that Equation (5.1) overestimates the trend uncertainty for two reasons. First, the estimated trend correlations are influenced by substantial correlations at small scales, as discussed above. Second, the pre-1997 trends for SBUV would be in better agreement with limb data sets, and have smaller uncertainty estimates, if the time series from 1979 would be considered (see discussion in *Section 5.1.2* and **Figure S5.8** in the *Supplement*).

5.3.3 Alternative methods

In *WMO* (*2014*), the combined trend from several satellite records was evaluated by first adding an estimated mean drift of 2% per decade to each instrument's 2-sigma statistical uncertainty and then calculating the mean trend weighted by the total error of individual trends. The uncertainty of the combined trend was then computed by propagating these individual trend uncertainties through the weighted mean. This uncertainty estimate does not take into account the statistical dependence (*i.e.*, correlation) nor the spread between individual trend estimates. The latter is often considerable and larger than the systematic error assumed by *WMO* (2014), likely leading to an underestimate of the uncertainties.

Harris et al. (2015) used another method that combines the trend and uncertainty in a joint distribution (J-distribution; *SPARC,* 2013). The central value is computed as the arithmetic mean and its uncertainty is evaluated as

$$\sigma^2_{mean} = \frac{1}{N}\sum \sigma_i^2 + \sum \frac{(x_i - \bar{x})^2}{N-1} \qquad (5.4).$$

This J-distribution approach is appropriate to estimate the population variance based on several data sets (*e.g.*, *Sofieva et al.,* 2014). For the current application, the estimate by Equation (5.4) is quite conservative. First, it assumes that trend estimates from several merged data sets do not reduce the random uncertainty. Indeed, the first term in Equation (5.4) is the mean of individual uncertainty variances, which is equivalent to the (conservative) assumption that all individual trend estimates are fully correlated. Second, the difference between individual trend values is also due to their (random) uncertainties and thus the second term in Equation (5.4) also includes (at least partially) the first term.

Steinbrecht et al. (2017) used the arithmetic mean as the central value, like the J-distribution method. However, they suggest to estimate the variance of the combined trend as the variance of the sample mean of estimates that are not necessarily independent

$$\sigma^2_{mean} = \frac{1}{n_{eff}} \sum \frac{(x_i - \bar{x})^2}{N-1} \qquad (5.5),$$

where n_{eff} is the effective number of independent data sets. **Figure 5** of *Steinbrecht et al.* (2017) shows uncertainties for $n_{eff}=1$ (*i.e.*, all trend results assumed fully correlated), while their **Table 6** presents the uncertainty estimates for $n_{eff}=3$ at altitudes above the 50 hPa level and $n_{eff}=2$ at lower altitude levels. Only post-2000 trends are discussed in this work. This choice is motivated by the expectation that the following three groups of data sets are fairly independent in the past two decades: The SBUV records, those based on SAGE-MLS as the "backbone," and the rest relying on SAGE and another instrument. Pre-1997 trends in the limb group are highly correlated as they rely heavily on the SAGE II record (see correlations in Equation (5.3)), which would lower the effective degrees of freedom by one unit. **Figure 5.8** shows the uncertainties according to Equation (5.5) with $n_{eff}=2$ used for pre-1997 trends and $n_{eff}=3$ for post-2000 trends above the 50 hPa level (green line). Corresponding values at lower altitudes are reduced by one since no SBUV data is used in the lowermost stratosphere.

5.3.4 Discussion

Figure 5.8 compares the results from three methods to combine trend and trend uncertainty for the same six satellite trend profiles. The combined trend values, computed from the arithmetic mean, are by definition the same for each of the methods. However, differences in the combined trend uncertainties are considerable. Higher significances are found post-2000 with the method used by *Steinbrecht et al.* (2017) as a result of the assumption of 2–3 independent trend estimates (mainly affecting the upper stratosphere, see **Figure 5.7**) as well as the lack of a systematic lower bound to the uncertainties where the trend results coincide (*e.g.*, around 10 hPa and at lower altitudes). This is less of an issue in the pre-1997 period, when SBUV results diverge considerably, leading to uncertainties that are similar between the methods used by *Steinbrecht et al.* (2017) and LOTUS. The systematic inclusion of uncertainty from propagation of regression errors (J-distribution) likely overestimates the uncertainty during the ozone depletion period. As mentioned earlier, part of this error component is already present in the derived sample standard deviation.

The LOTUS and J-distribution methods have the uncertainty from error propagation as a lower bound. This limits the potential for large vertical variations in trend uncertainty, as regression errors usually have only a weak and smooth dependence with height. In most cases, however, it is the variance in the post-2000 trend sample that drives the uncertainty of the combined result, and this has a clear vertical structure as well. For all methods there is a localised large increase in uncertainty around 5 hPa in the tropics. We also note gradual increases in post-2000 trend uncertainty with decreasing pressure (higher altitudes) in the mid-latitude upper stratosphere. **Figure S5.9** in the *Supplement* shows 1-sigma uncertainty profiles of **Figure 5.8** but also adds one more curve using the sum() alternative of Equation 5.1 rather than the max(), for illustrative purposes.

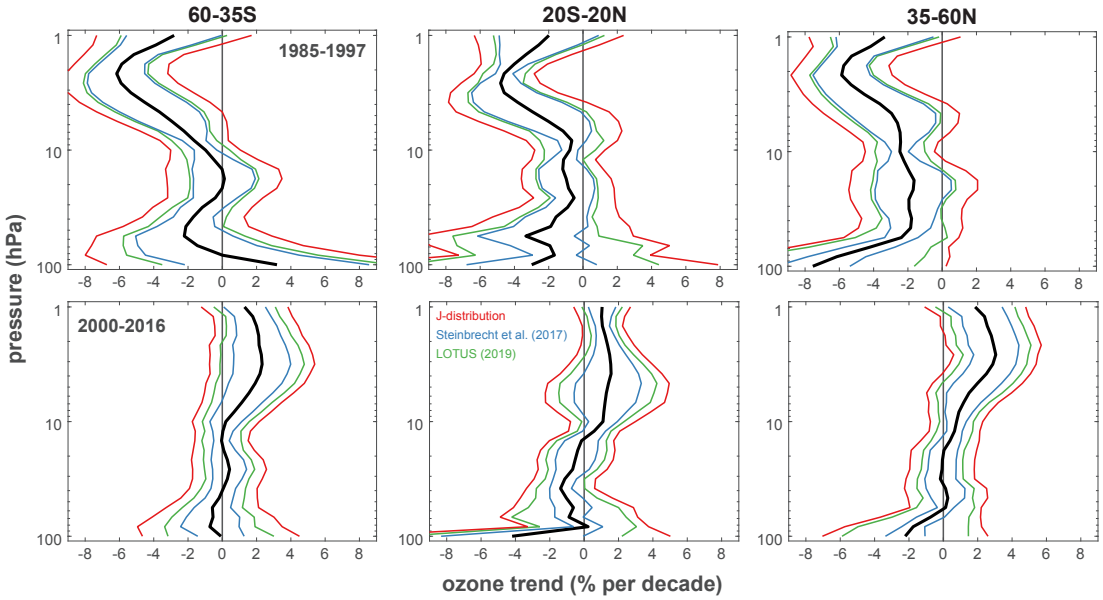

Figure 5.8: *Combined trends of six limb and nadir data records (thick black lines) with 2-sigma uncertainties (thin coloured lines) for the pre-1997 period (top) and the post-2000 period (bottom) from the ILT regression for latitude bands 60°S–35°S (left), 20°S–20°N (centre), and 35°N–60°N. Green lines denote results from the LOTUS method (Equation 5.1), blue lines denote results from the method of Steinbrecht et al. (2017; Equation 5.5) with $n_{eff}=2$ for pre-1997 and $n_{eff}=3$ for post-2000 trends, and red lines denote the results from the J-distribution method (Equation 5.4).*

The results presented here indicate that the uncertainty of the combined trend is sensitive to the method used. Each method is based on fair assumptions as the combined result is limited to the small number of trend realisations, and trend uncertainties derived with the LOTUS method lie in between those obtained according to *Steinbrecht et al.* (2017) (smaller values) and the J-distribution method (larger values), though only a limited set of methods to combine trend uncertainties were tested in detail. However, the LOTUS method not only incorporates (at least partially) the influence of systematic sources of uncertainty but also improves the calculation of uncertainties by considering the correlations between contributing trends. Ultimately this exercise in improving the trend uncertainties applies to the concept of merging trends derived from several different data sets and, while some assumptions are made about the nature of the correlations between the trend uncertainties, the most meaningful way to improve the uncertainties in future analyses would be to reconcile the discrepancies between the data sets themselves prior to the merging process.

It should be noted that the combined satellite trend uncertainties in **Figures 5.7** and **5.8** do not include contributions from the choice of regression model and proxies. Sensitivity tests in *Section 4.3.1* have shown that this error can be as large as ~1.5 % (lower/middle stratosphere) or ~3 % (upper stratosphere) per decade, although these numbers are likely a worst case scenario. Adding these in quadrature would reduce the significance of the reported trends in the upper stratosphere considerably and so it is a subject of further research to provide robust quantitative results.

5.4 Ground-based trends

This section compares pre-1997 (January 1985 – December 1996) and post-2000 trends (January 2000 – December 2016) from ground- and satellite-based ozone profiles, averaged over three broad latitude bands (**Figure 5.9**). The main purpose of these trend comparisons is to verify the robustness of the derived trends in the combined records (see discussion about the consistency of records in *Chapter 3*). The broad-band averaging smoothes the records and mitigates atmospheric noise that can be introduced in the station record by short-term meteorological variability and infrequent sampling methods. The disagreement between trends is used for the evaluation of the magnitude of uncorrected drifts in the records. Intercomparisons of multiple records and analyses of the fit residuals provide information that helps identify drifts and provide means for their correction in the future data reprocessing.

As mentioned in previous chapters, there are several ground-based measurement methods for tracking stratospheric ozone recovery, including lidars, Umkehr measurements performed with Dobson/Brewer photometers (hereafter called 'Umkehr'), Microwave radiometers

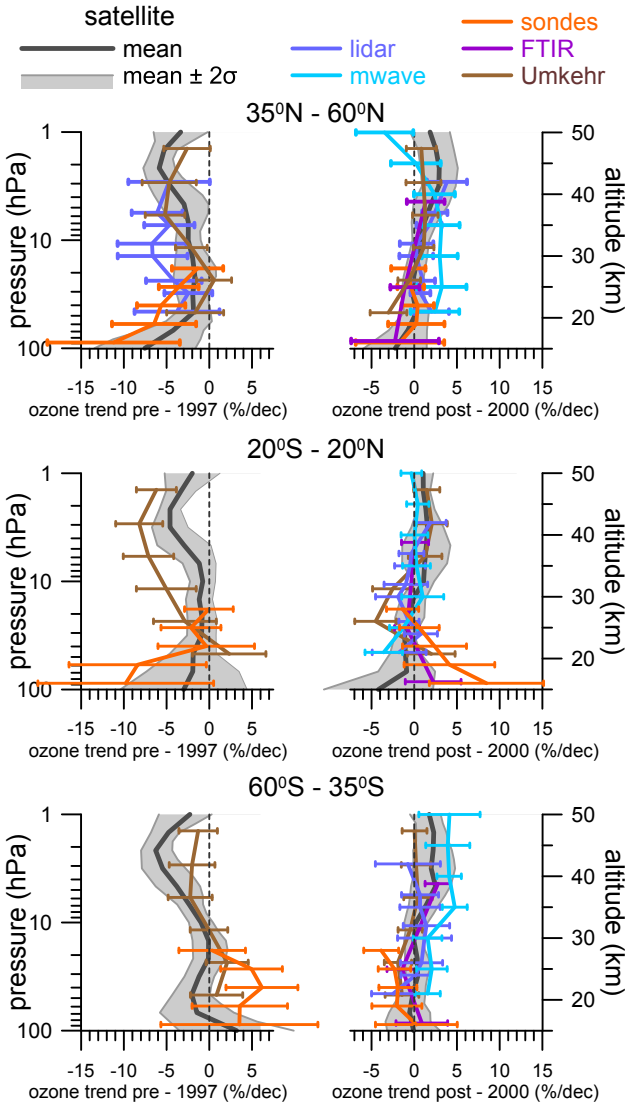

Figure 5.9: *Ozone trends for the pre-1997 and post-2000 periods from the ground-based stations, averaged over the available latitude bands. Upper panel shows 35°N–60°N, middle panel shows 20°S–20°N, and lower panel shows 60°S–35°S.*

(hereafter called 'MWR'), FTIR spectrometers (hereafter called 'FTIR'), and balloon-borne ozonesondes (hereafter called 'ozonesondes'). The length of the historical records, the temporal and vertical sampling of the different records, and the spatial distribution of the different stations variy between these methods (see *Chapter 2* for more details).

The ability of observations at several ground-based stations to capture the trends observed by satellites over the broad-band regions has been studied in this Report and is summarised in recent publications (*i.e.*, *Steinbrecht et al.*, 2009; *Zerefos et al.*, 2018; and reference therein). Results of the analyses performed by *Zerefos et al.* (2018) suggest agreement between trends derived from the subset of the SBUV MOD record selected to match the geolocation of lidar ground-based station and 5° zonally averaged SBUV-MOD records centred at the latitude of the ground-based station.

In the NH mid-latitudes, the ground-based networks are the most densely represented over Europe and North America. Moreover, according to the study of *Zerefos et al.* (2018) and the trend analyses of satellite records presented above (see **Figure 5.2**) ozone variability and trends at altitudes above 10 hPa are coherent over a wide range of latitudes (~20° or wider). Thus, several stations within that latitude range should be able to capture the trends representative of a broad-band average. However, ground-based records are generally shorter than the long SBUV records and some, such as sondes and lidars, are characterised by higher vertical resolutions. This makes trends derived from these records more sensitive to geophysical variability (*e.g.*, polar vortex influence in the winter at mid-latitudes) and incorrect evaluation of long-term atmospheric variability due to, for example, the 11-year solar cycle.

Figure 5.9 shows resulting trends for ground-based stations located within the NH mid-latitudes (top), tropics (middle), and SH mid-latitudes (bottom). Ground-based trends are compared to the satellite broad-band averages derived in *Section 5.3*. Satellite trends for the pre-1997 period (left) and post-2000 period (right) are shown with a grey envelope that represents the combined error (see *Section 5.3.2* for details). The broad-band trends for each ground-based record are calculated from the deseasonalised monthly mean anomalies averaged over the broad-band latitude ranges and at the vertical grid specific to each measurement technique and data processing method (see *Section 2.1* for further details). Deseasonalised anomaly data records are combined prior to the regression analysis in case multiple stations provide data for the broad latitude bands and for the considered measurement technique (see more details in *Section 2.1.2*). The uncertainties for trends obtained for both individual station and broad-band combined records are shown as standard errors of the ILT fit.

The trends for the pre-1997 period in the NH mid-latitudes (**Figure 5.9**, top panel) are represented by the combined Umkehr (brown) and lidar (blue) profiles (see **Table 2.1** in *Chapter 2* for the selection of the stations in the broad-band averages shown in this section). The mean trend pattern in the Umkehr trends is similar to the combined satellite and model trends (not shown). However, the error bars for combined Umkehr trends are larger than the satellite combined error bars, as would be expected. Comparisons between combined satellite and model-derived trends are discussed in the following section (see **Figure 5.11**). While not tested here, these differences are most likely based on limited temporal and spatial sampling within the 35°N–60°N latitude band that are not completely captured by three NH Umkehr stations. Lidar combined datasets tend to have stronger negative trends than found in the Umkehr, model, and satellite records between 20 hPa and 10 hPa pressure. The differences in trends can be explained by the shorter lidar records in the pre-1997 period. For lidar long-term records used in this study, a sufficient number of monthly observations was reached only by the end of the 1980s as indicated in **Figure 2.3**. Therefore, limited temporal and spatial sampling can influence the trends and their uncertainties during this early period. However, the error bars in Umkehr, lidar, and satellite trends overlap, thus indicating consistent estimates of the observed trends between different observing systems. Negative ozonesonde trends in the lower stratosphere tend to be on the far edge of the combined satellite uncertainty envelope, and they are also consistent with the lidar- and model-based trends (**Figure 5.11**).

For the post-2000 period, ground-based data are represented by FTIR, lidar, Umkehr, and MWR records. Some of these records (*i.e.*, FTIR and MWR) began past 1997. For these types of records the ILT method is essentially converted to the multiple linear regression with just a single trend term, but it still uses the same proxies. The exception is the aerosol proxy that becomes non-orthogonal to the trends itself and therefore would alias the recovery trend analyses. In these instances, the aerosol proxy was removed from the statistical MLR model. Results presented for ground-based trends in the NH for post-2000 period (right top panel of **Figure 5.9**) show general good agreement between the different instruments. A trend of -2 % per decade at 50 km (1998–2014) have been published for the Bern MWR (*Moreira et al.*, 2015); Payern MWR also shows a negative trend at 50 km (-0.5 % per decade). The possibility of an influence on the 50 km level ozone content variation of information coming from the levels above 50 km should not be neglected as the MWR averaging kernels are large at that altitude and because the measurement contribution is high for MWR at 50 km. The ground-based trends support the mean values of the recovery trends in the stratosphere detected by satellite observations and models (**Figure 5.11**). Similar to satellite records, they also suggest slightly more negative trends in the lower stratosphere (NH mid-latitudes), although error bars in both data sets are large.

Ground-based trends for tropical (**Figure 5.9,** middle panels) and SH mid-latitudes (**Figure 5.9**, bottom panels) are based on only a few station records (See **Table 2.1** in *Chapter 2*). Thus, it is expected that trends from the ground-based measurements can be biased due to their limited spatial coverage and have larger uncertainties due to limited sampling frequency. For example, the Umkehr record for the pre-1997 period at 10 hPa shows a more negative trend than the satellite (**Figure 5.9**, left middle panel) and model (**Figure 5.11**) estimates. However, there is only one Umkehr record available from the MLO in Hawaii (located at 19°N) and one ozonesonde record from Hilo station (near MLO), and thus these records are not fully representative of the trends derived from the broad 20°S–20°N latitude band. Middle and low stratospheric ozone variability is only weakly correlated between the NH and SH tropics (*i.e.*, *Zerefos et al.*, 2018; and references therein). Still, the subset of SBUV MOD, limited in space to the MLO station location, and 5° zonally averaged satellite measurements at 20°N describe

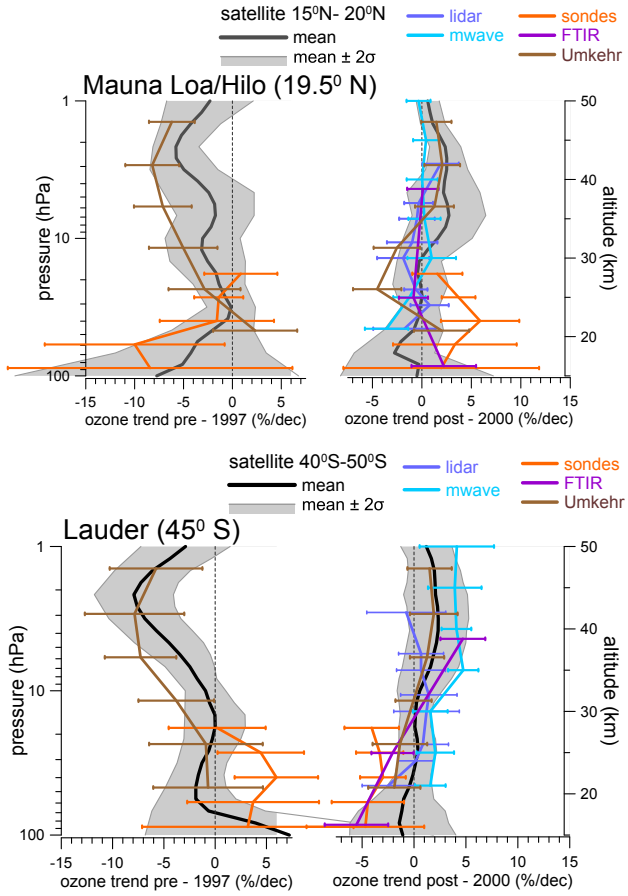

Figure 5.10: *Ozone profile trends from different ground-based data records for the pre-1997 (left) and post-2000 periods (right) at Hawaii, USA (19.5–19.7°N, 155.1–155.6°W; top row) and at Lauder, New Zealand (45.0°S, 169.7°W; bottom row). Error bars represent the 95 % confidence interval. The black line represents the multi-model mean for the latitude band 15°N–25°N and 50°S–40°S, respectively.*

very similar ozone variability (with correlations larger than 0.6 in the middle and upper stratosphere) and also comparable trends.

In order to reduce spatial sampling differences between broad-band satellite and ground-based data collected at MLO/Hilo station (and Lauder, New Zealand as discussed below) the combined satellite trends were recalculated for ~10° zonal bands. All satellite trends for this narrow zonal band are computed analogously to the reported trends in broad-band zones (*Section 5.3.2*). Since the combined datasets were provided with different zonal resolutions, the following combination of trend results were used: The SBUV MOD and SBUV COH trends from 15°N–20°N and 20°N–25°N were averaged whereas the mean of 10°N–20°N and 20°N–30°N is used for the GOZCARDS, SWOOSH, and SAGE-CCI-OMPS records. The SAGE-OSIRIS-OMPS trend is already derived for 15°N–25°N and was used as is.

When ground-based trends are compared to the satellite trends averaged over a narrower zonal band (**Figure 5.10** upper panel), uncertainties in the mean satellite trends increase and the agreement between satellite, ozonesonde, and Umkehr trends for the pre-1997 period at MLO improve. Ground-based trends prior to 1997 for SH mid-latitudes (bottom panel of **Figure 5.9**) are presented by two Umkehr records (Lauder, New Zealand and Perth, Australia) and by the combination of longer ozonesonde records from Lauder, New Zealand (begins in 1986) and shorter record from Macquarie Island (begins in 1994). The Umkehr combined trends show agreement with satellite trends averaged over the 60°S–35°S latitude band. The combined ozonesonde record detects statistically significant positive trends between 40 hPa and 30 hPa that are in agreement with Umkehr trends within their respective error bars but disagree with satellite broad-band estimates. The single station ozonesonde or Umkehr records are not expected to capture trends representative of the broad latitude range, while satellite-derived trends in 5–10° zonally averaged bands also show spatial variability in the middle stratosphere of SH mid-latitudes (see **Figure 5.2**). Variability in broad-band trends can be demonstrated by comparing combined Umkehr pre-1997 trends (bottom panel of **Figure 5.9**) against the larger negative trend derived from the Umkehr record at Lauder only (**Figure 5.10**, bottom panel). Additionally, the mean satellite trends and uncertainties used in **Figure 5.9** are changed from a broad band representation (60°S–35°S) to a 50°S–40°S degree latitude band centred at Lauder (45°S) (**Figure 5.10**). The SBUV MOD and SBUV COH trends from 50°S–45°S and 45°S–40°S were averaged. For the GOZCARDS, SWOOSH, and SAGE-CCI-OMPS records, the trends from 50°S–40°S are used, and for SAGE-OSIRIS-OMPS the average trends between 55°S–45°S and 45°S–35°S were used. The agreement between the 10-degree mean satellite, ozonesonde, and Umkehr trends at Lauder is improved as compared to the results shown in **Figure 5.9**. However, ozonesondes at Lauder at 40 hPa show trends that are different from the satellite averages and outside of the range of satellite trend uncertainties. Umkehr trends at Lauder are less negative than satellite mean trends and agree with ozonesonde trends within their respective uncertainties, thus pointing to either limited sampling at ground-based stations, shortness of the record (Umkehr and sonde records started after 1986), or inhomogeneities in the instrument record (*i.e.*, not all ozonesonde records used in LOTUS analyses were fully homogenised). The trend derived from the combined ozonesonde records is very similar to the trend derived at Lauder only. This similarity is expected because the Lauder record contributed most to the combined record before 1997, and only in the later part of time period is the data set (starting from 1994) a combination of two or three records (the third SH ozonesonde record is added in 1999).

For the post-2000 trends, in addition to Umkehr and ozonesonde MLO records, MWR and lidar records become available, but again these additional records are from MLO only. The trends from all instruments agree well and within the error bars.

There are multiple ozonesonde stations (including SHADOZ; see *Thompson et al.,* 2007) that are used to create the tropical broad-band record for post-2000 trends (see **Table 2.1** and *Section 2.1* for further details). **Figure 5.10** summarises trends derived from different instruments at MLO only (top panel); the ozonesonde trends are those at the neighbouring Hilo station. These single-station ozonesonde trends agree with the broad-band results within the uncertainties, although they indicate stronger trends between 20 km and 30 km and weaker trends at 15 km (100 hPa), which are in close agreement with FTIR measurements at the MLO station. However, ozonesonde records used in the LOTUS Report were provided before the process of homogenisation was finalised (*Sterling et al.,* 2018; *Witte et al.,* 2017; *Witte et al.,* 2018), such that these records can contain uncorrected step changes (see *Chapter 3* for discussion) that can potentially impact the derived trends. Therefore, single-station and broad-band latitude averaged trends will have to be re-evaluated after all ozonesonde homogenised records become available for trend analyses.

Ground-based trends derived from records available in the SH (bottom panel in **Figure 5.9**) indicate that ground-based trends tend to overlap with the satellite broad-band and model averaged trends (bottom panel in **Figure 5.11**) within their respective uncertainties. However, a wider range of ground-based trends in the middle and upper stratosphere is found, depending on the instrument, as compared to the combined satellite or model range of uncertainties. One reason for the instrument trend difference is the combination of two stations for FTIR and Umkehr instruments (see **Table 2.1** in *Chapter 2*), while MWR and lidar trends are derived from the Lauder record only. In order to prove the consistency in the trends derived from different instruments we plotted trends just for the Lauder station (bottom panel of **Figure 5.10**). Note that the combined satellite trend in **Figure 5.10** is based on a narrower latitude range than results shown in **Figure 5.9**. The agreement among trends derived from five instrument-specific datasets at Lauder is improved in comparison to the combined trends shown in **Figure 5.9**, thus highlighting the variability in trends at the ground-based stations away from Lauder (but still inside of the 60°S–35°S band). In addition, the subset of SBUV MOD data limited in space to the Lauder station location and 5° zonal averaged satellite measurements at 45°S captures very similar ozone variability (with correlations of 0.7-0.8 in the middle and 0.5-0.6 in the upper stratosphere; *Zerefos et al.*, 2018) and closely comparable trends (see **Figure 4.7** in *Chapter 4*). The post-2000 trends at altitudes below 20 hPa at Lauder are negative, and ozonesonde and FTIR records show statistically significant negative trends of ~5 % per decade in the lower stratosphere (below 60 hPa). The ozone trends in the lower stratosphere at Lauder are different from the near zero trends derived from the broad-band combined satellite record in the SH mid-latitudes. However, combined ozonesonde and FTIR records available within the 60°S–35°S latitude band also show near zero trends (bottom panel in **Figure 5.9**), thus indicating spatial variability of the trends detected in the SH mid-latitude lower stratosphere.

5.5 Comparison between combined satellite and CCMI model trends

Satellite ozone profile trends and uncertainties derived for the broad latitude bands at 60°S–35°S, 20°S–20°N, and 35°N–60°N are compared with CCMI-1 REF-C2 model trends in **Figure 5.11**. The satellite observed trends were combined based on the method described in *Section 5.3.2* so that only one average satellite-based trend profile and its uncertainties remain in the figure. The mean and median of trends calculated from CCMI-1 REF-C2 model simulations, averaged as described in *Section 2.3.2*, are shown in **Figure 5.11**. The model uncertainties are shown as the grey area, enveloping the model mean at ±2 standard deviations. Trends are fit to the CCMI-1 simulations over the same time periods as for the observations: Pre-1997 (Jan 1985 – Dec 1996) and Post-2000 (Jan 2000 – Dec 2016).

5.5.1 Pre-1997 period

As shown in previous sections, the left panels in **Figure 5.11** display a significant ozone decline in the pre-1997 period for all three broad latitude bands. The largest trends in satellite observations (and models) in the upper stratosphere are -5.9 % ± 1.9 % per decade (-5.4 % ± 2.9 % per decade) in the NH mid-latitudes, -4.8 % ± 1.3 % per decade (-3.8 % ± 2.6 % per decade) in the tropics, and -6.2 % ± 1.8 % per decade (-5.7 % ± 3.0 % per decade) in the SH mid-latitudes. However, smaller but nonetheless significant negative trends are present in pressure regions between 7 hPa and 1 hPa.

In the NH mid-latitudes the median trend from the CCMI-1 model simulations at 100 hPa is similar to that at 50 hPa; however, observations show an enhanced negative trend, up to -8 % per decade. At the same time, the error bars for both models and observations are large in the region just above the tropopause (~10 km/200 hPa in the extratropics and ~17 km/100 hPa in the tropics). This is due to low ozone values, large interannual ozone variability, and large vertical and horizontal ozone gradients in this region. The large error bars in the observational trends are also related to less reliable observations at these altitudes and increased variability that is not fit by the ILT regression model proxies. In the SH mid-latitudes below 20 km/50 hPa (UTLS), observations in the broad latitude bands show a small positive trend (although not statistically significant), while models show strong negative trends. The four limb profile trends exhibit coherent behaviour at altitudes below the 70 hPa level (**Figure 5.6**). The uncertainty envelope in the models tends to increase below the 50 hPa level, still the observation-based trend is found outside of the model 2-sigma uncertainty.

5.5.2 Post-2000 period

In the post-2000 period (right panels in **Figures 5.11**) satellite-based trend analyses show a statistically significant increase of 3.0 % ± 1.9 % per decade at the level of maximum response (around 3 hPa) in the NH mid-latitudes, which is identical to the 3 % per decade reported in *WMO* (2014). Even trends at 2 hPa, although not as strong as at 3 hPa, are statistically significant, indicating that observations are now covering a long enough period to detect the recovery signal on several levels of the NH mid-latitude upper stratosphere. Models suggest a slightly smaller but nonetheless statistically significant trend, 2.2 % ± 1.6 % per decade, for this latitude band in the upper stratosphere. In the lower stratosphere, models predict slightly positive trends for NH mid-latitudes, however, these have large uncertainties, indicating that the different models do not agree on the magnitude or sign of ozone evolution at this level. Satellite observations suggest small negative trends, but their uncertainties are also large, so that this trend value is not statistically significant.

In the tropical latitude band (20°S–20°N), the trends above 3 hPa or 43 km altitude are found to be positive (1.5 % ± 1.4 % per decade) in the satellite-based results. Model simulations suggest positive trends that are statistically significant for altitudes above 4 hPa with a maximum trend value of 1.8 % ± 1.2 % per decade. These satellite-based results support the trends and uncertainties reported in *WMO* (2014), while the uncertainties here are smaller than the uncertainties reported in *Harris et al.* (2015). The updated method of trend combination (see *Section 5.3.2*), quality-improved data sets (see *Chapter 2*), and four additional years of measurements help to isolate the trend signal from the background variability. In the middle stratosphere, a weak negative trend is detected in the observations that is just barely statistically significant at 30 hPa or 23 km, whereas the models suggest no trend at all at this level. In the lower stratosphere, satellite-based trends and models indicate even stronger negative trends but with larger uncertainties.

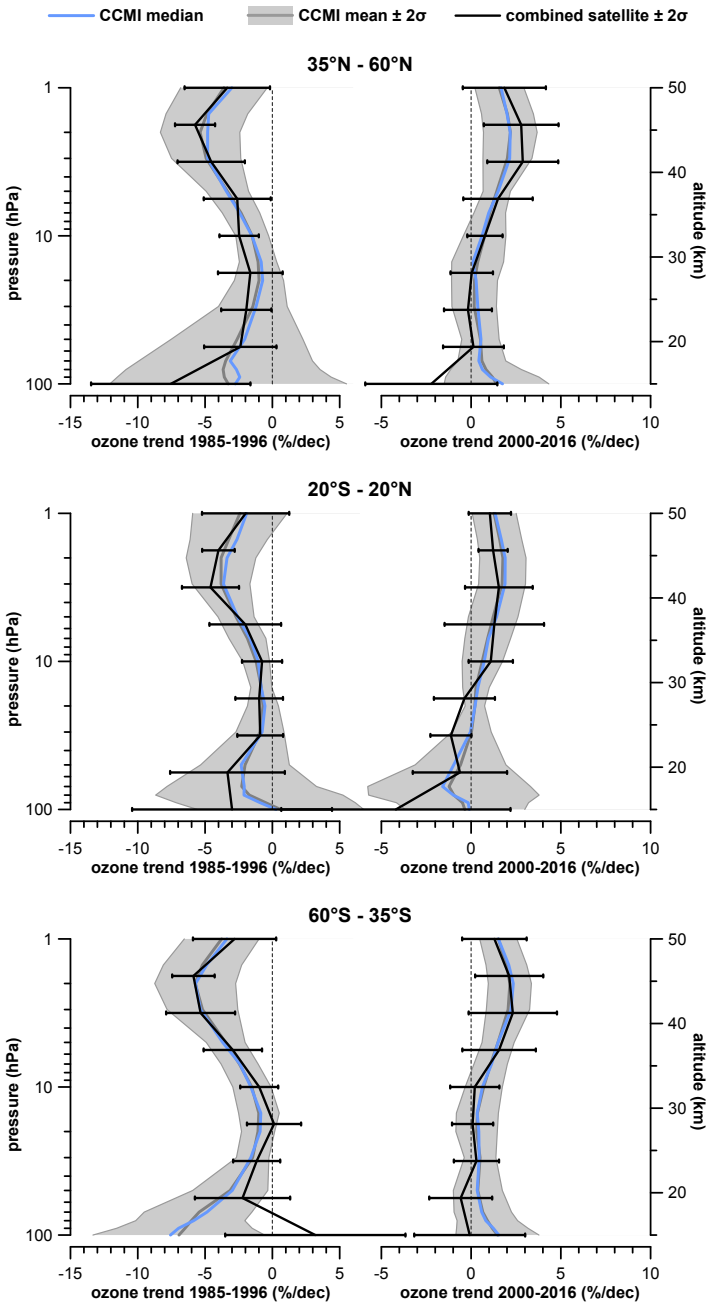

Figure 5.11: *Ozone trends for the pre-1997 and post-2000 periods from the CCMI REF-C2 models' simulation and broadband satellite data sets, averaged over the available latitude bands: 35°N–60°N (upper panel), 20°S–20°N (middle panel), and 60°S–35°S (lower panel).*

In the upper stratosphere at SH mid-latitudes, a positive trend of 2.1 % ± 1.9 % per decade that is just statistically significant is detectable at around 2 hPa/45 km in the satellite-based results given the four additional years of observations. Trends calculated in the upper stratosphere just above or below this level also indicate positive trends, but these are not statistically significant. Model results show statistically significant trends for the altitude region above 10 hPa that are of almost identical magnitude as the observational trends (2.2 % ± 1.1 % per decade at about 2 hPa/45 km). For the rest of the profile shown in **Figure 5.11**, trends calculated from satellite data are around zero, with only a small indication of negative trends at around 60 hPa or 20 km that are not statistically significant. Model trends are also close to zero between 40 hPa and 10 hPa (~23 km to ~33 km) and become slightly positive below these levels. However, the model trends are not statistically significant in the middle and lower stratosphere of the SH mid-latitudes.

5.6 Summary of observed profile trends

5.6.1 Pre-1997 period

Upper stratosphere

For the pre-1997 period we find negative trends across nearly the entire stratosphere in nearly all satellite and ground-based data records (**Table ES.1**; **Figure ES.1**; also **Figures 5.1**, **5.5**, and **5.8**). Individual and combined satellite data show highly statistically significant evidence of declining ozone concentrations since the mid 1980s and well into the 1990s in the upper stratosphere (altitudes above the 10–5 hPa level), where ozone is in photochemical equilibrium (**Figure 5.6**). The depletion reaches a maximum rate of 5.9–6.2 % per decade at mid-latitudes (near 2 hPa, ~42 km) and of 4.8 % per decade in the tropics. Ground-based measurements are much more sparsely sampled in space and time, and, as a result, the significance of the trends is not as high and trend values differ. Nonetheless, both lidar and Umkehr data corroborate the satellite findings for broad-band regions in the upper stratosphere (see **Figure 5.8** and **Figure 5.9**).

Middle stratosphere

Ozone decline rates in the middle stratosphere down to the 50 hPa level are considerably smaller, 1–2 % per decade, than in the upper layers of the stratosphere. In fact, these values are too small to be statistically significant in most regions of the globe (see **Figure 5.1**). The combined satellite trends for the broad-latitude bands do not show any statistically significant trend in the middle stratosphere. A few individual satellite data sets (*e.g.*, SBUV MOD, GOZCARDS, and SWOOSH) show statistically significant trend values in the NH mid-latitudes at their native resolution as well as in the broader latitude bands (**Figure 5.6**). Ground-based measurements generally do not show statistically significant trends in the tropics and SH mid-latitudes. Only lidar measurements in the NH mid-latitudes display significant negative trends in the middle stratosphere, but these are based on shorter records in the 1985–1997 period (**Figure 5.9**).

Lower stratosphere

In the NH lower stratosphere, both satellite and ozonesonde data point to a negative trend of 5 % per decade and more, which is seemingly significant for the satellite analysis (**Figure 5.8**). Satellite trends close to 100 hPa in the tropics and in the SH are less than 3 % per decade, but the negative trends are reversed to positive values around the Equator (**Figure 5.1**). However, our confidence in the results in the lower stratosphere is not as high as that in the upper stratosphere. Estimating trend uncertainty, and therefore statistical significance, is inherently more complicated in the lower stratosphere because of large natural variability influenced by transport and mixing processes, low ozone concentrations, waning sensitivity of satellite observations, and the lack of independent measurements[1]. Additional research will be needed to put the trend results in the lower stratosphere on more solid ground.

5.6.2 Post-2000 period

Upper stratosphere

For the post-2000 period, we find positive trends in all satellite and most ground-based records in a large part of the upper stratosphere (**Table ES.1**; **Figure ES.1**; also **Figures 5.2**, **5.3**, **5.5**, and **5.8**). The statistical significance of the estimates varies between latitude bands and between individual data sets and the combined satellite result. However, the majority of individual data sets in the upper stratosphere show significant trends over the NH mid-latitudes, while several data sets also contain statistically significant trends over (part of the) tropics and SH mid-latitudes (**Figure 5.5** and **Figure 5.8**).

In the upper stratosphere of NH mid-latitudes positive trends range between 2.0 % and 3.1 % per decade when satellite results are combined. Lidar and Umkehr data support this finding, but MWR data do not (**Figure 5.9**). However, the anomaly time series clearly indicate aberrant values in the two-station combined MWR record in recent years, which leads to a negative trend (**Figure 3.4** and **Figure 5.4**). There is currently no clear understanding of this discrepancy and additional research is needed. Combined satellite trends at NH mid-latitudes are statistically significant between 3.8–1.2 hPa (**Figure 5.13**; **Table S5.1**). Statistical significance in ground-based trends is only found for combined lidar trends at altitudes above the 5 hPa level.

In the tropics, most individual satellite and ground-based records show small positive trends that lie close to or barely beyond the 2-sigma threshold for significance. The only exceptions are SAGE-CCI-OMPS over the full range and SBUV MOD at 3-2 hPa. The combined satellite trend in the upper stratosphere ranges between 1.0 % and 1.6 % per decade and is significant between 2.6–1.0 hPa. SBUV MOD and SBUV COH results diverge largely at 8–3 hPa (**Figure 5.6**). None of the ground-based trends in the tropics are significant. Ground-based observations of tropical upper stratospheric ozone were considered at just one location (MLO, Hawaii) which complicates a direct comparison to the satellite data over the broader band.

Upper stratospheric trends derived from combined satellite data over SH mid-latitudes range between 1.8 % and 2.3 % per decade. All analyses of individual satellite and ground-based records, except for the only lidar record, show positive or zero trends in this region. The combined satellite trend

[1] Below the 50 hPa level, the SBUV profile data are not used (*Section 2.2.2.*) and all merged limb profile records rely (mostly) on SAGE II data.

data are significant over the 1.8–1.2 hPa range. However, the confidence in the positive trend significance is less than in the NH since some satellite records show trends right at the 2-sigma threshold. The MWR and FTIR trends are significant at altitudes above the 5 hPa level. Similar to the tropical belt, only one or two ground-based sites were available to assess trends in the SH mid-latitudes. In general, satellite and ground-based trends agree in the upper stratosphere within their respective uncertainty bounds.

Middle stratosphere

Trends in the middle stratosphere and over mid-latitudes are found to be smaller than 0.5 % per decade and not significant. However, in the tropical region, trends derived from the SBUV-based data sets and SWOOSH are negative and statistically significant (**Figure 5.2**). Other satellite records and ground-based records in the tropics (mostly from MLO/Hilo) show trends close to zero that are not statistically significant (**Figure 5.6**, **Figure 5.9**, and **Figure 5.11**).

Lower stratosphere

In the lower stratosphere, individual and combined satellite trends are mostly negative but statistically insignificant at both their native resolution and in broad latitude bands (**Figure 5.2**, **Figure 5.6**, and **Figure 5.11**). Most ground-based measurements are also not statistically significant for all three latitude bands in the lower stratosphere; however, in the tropics, ozonesondes show a significant positive trend at around 100 hPa of 8 % per decade, and FTIR measurements also indicate a positive trend which is not significant.

5.6.3 Comparison of LOTUS trend results with previous assessments

Figures 5.12–**5.14** display comparisons of LOTUS broadband averaged trend results with those of the most recent assessments (*i.e.*, *WMO*, 2014; *Harris et al.*, 2015; and *Steinbrecht et al.*, 2017). For both periods, pre-1997 and post-2000, combined trends with their respective uncertainties are shown in **Figure 5.12**, while **Figure 5.13** illustrates the differences in the significance of trends between the various assessments. In addition, the trend uncertainties are directly compared in **Figure 5.14**. *Steinbrecht et al.* (*2017*; or S17 hereafter) found latitude-pressure patterns in post-2000 trends similar to those reported here (or L19 hereafter); magnitudes of the trends agree within 0.5% per decade. This is not surprising since both assessments determined trends for the combined satellite records as the unweighted mean. Furthermore, trends are estimated for the same analysis period (post-2000), from similar data records, using similar non-trend proxies and similar trend proxies.

However, the uncertainty of the combined trend is computed slightly differently between *S17* and *L19* (see also *Section 5.3.4*). Both rely on the concept of variance of the sample mean, which determines the total uncertainty as a sum of random and systematic effects, where drifts are considered as relative to the zero sample mean drift. *S17* uncertainty is computed as a biased estimator of the standard deviation of the sample mean and under the assumption that all combined records contain only three independent data records in the post-2000 period, such that $n_{eff}=3$ (or *2*, at altitudes below the 50 hPa level). The *L19* approach is based on $n_{eff}=1.4$ (or *1.3*, below the 50 hPa level) independent records (derived from correlation analyses of the trend fit residuals) and includes an additional term that represents a lower bound, which is equal to the propagated uncertainties from the regression coefficients. Note that for the comparison between *S17* and *L19* results, *S17* derived trend uncertainties were corrected by a factor *N/(N-1)* in order to consider the same unbiased definition of the sample variance. In the upper stratosphere, where the sample variance is large, the difference in n_{eff} plays a decisive role in testing the null hypothesis (*i.e.*, no-trend). *S17* trends are statistically significant across the entire upper stratosphere. In the *L19* trend analysis, high significance is found only at NH mid-latitudes, and less significant trends are found in the tropics and at SH mid-latitudes. At the moment, there is not sufficient information in the trend analyses to help determine which assumption is more realistic. However, it is not unreasonable to assume that in the upper stratosphere, *S17* and *L19* uncertainties represent, respectively, a lower and upper bound. Note that for the comparison between *S17* and *L19* results in **Figure 5.12**, *S17* derived trend uncertainties were corrected by a factor *N/(N-1)* in order to consider the same unbiased definition of the sample variance.

In the lower stratosphere, there are concerns that *S17* may underestimate true uncertainty. The trend sample has only a few members and their spread is fairly, perhaps unrealistically, small at these altitudes. Also, due to the small trend sample there is the risk of underestimating the true variance of the mean when trends coincide. Even the *L19* approach, which includes a term for error propagation from the regression coefficients, is not capable of capturing all sources of uncertainty (*e.g.*, most importantly measurement drift). Therefore, *L19* assumes that these errors constitute a lower bound, which is primarily reached in the lower stratosphere and also in part of the middle stratosphere. Pre-1997 trend results (not shown in the *S17* publication but provided by authors through personal communication) are also in good agreement with *L19* (**Figure 5.12**) but differ by 1–2 % per decade, most likely as a result of the earlier start in the time series used in the *S17* analysis (1979) as compared to *L19* (1985). All LOTUS trends for satellite, ground-based, and model data were computed starting in 1985 for ease of comparison. *Steinbrecht et al.* (2017) and earlier assessments by *WMO* (2014) and *Harris et al.* (2015) analysed time series starting in the early and/or late 1970s, which has an effect on the magnitude of the derived trends.

Harris et al. (2015; or *H15* hereafter) trends are comparable to LOTUS trends, though the study used a somewhat different set of satellite records; the length of the records was shorter (by about 4 years) and some records exhibited large drifts (*i.e.*, OSIRIS).

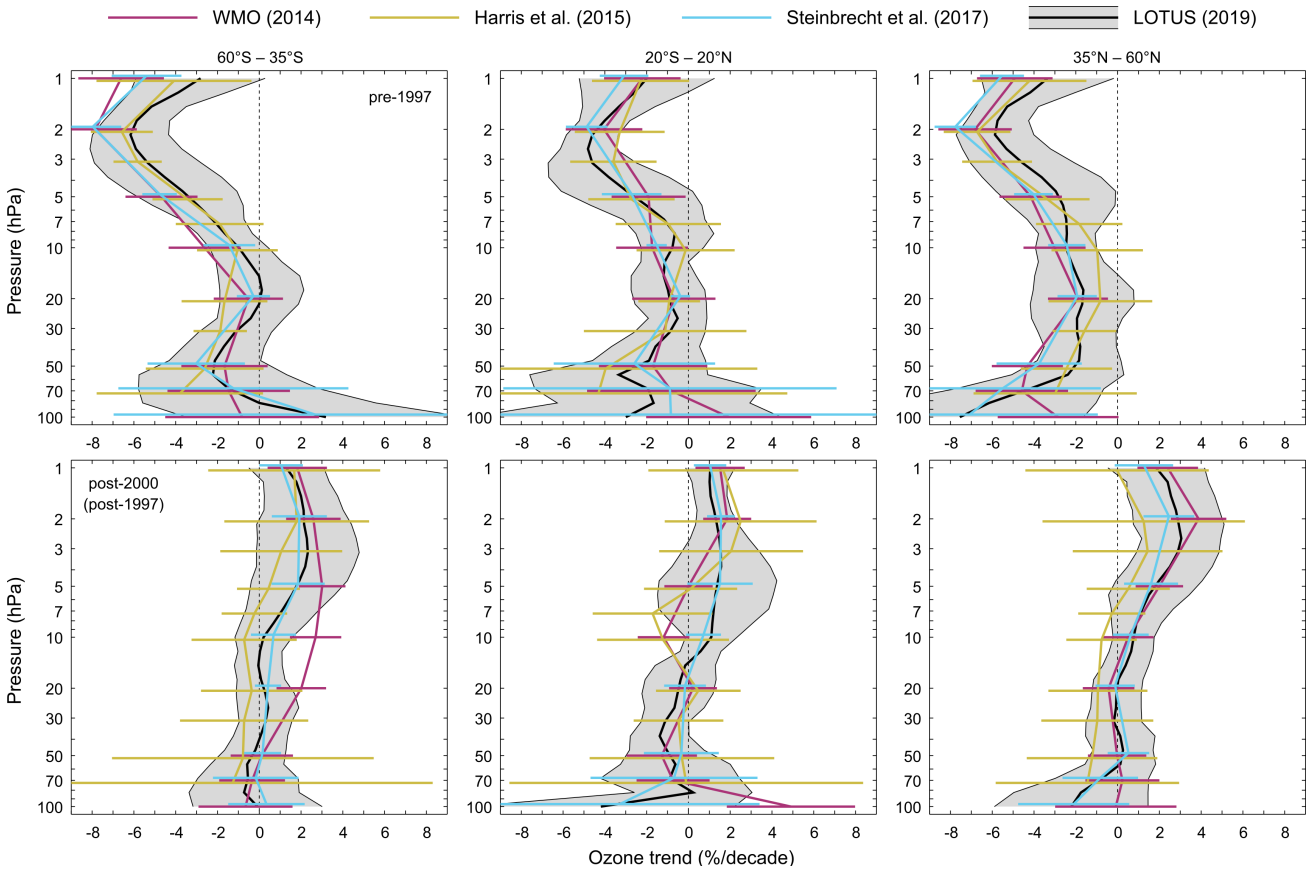

Figure 5.13: *Overview of ozone profile trends from past and recent assessments: LOTUS (this work), WMO (2014), Harris et al. (2015), and Steinbrecht et al. (2017) are shown in black, red, orange, and blue respectively. Top row shows trends before the turnaround of ODSs and bottom row since the turnaround (analysis time period differs by assessment). Shaded area and error bars represent the 95% confidence interval for the combined trend. Coloured profiles are slightly offset on the vertical axis for display purposes.*

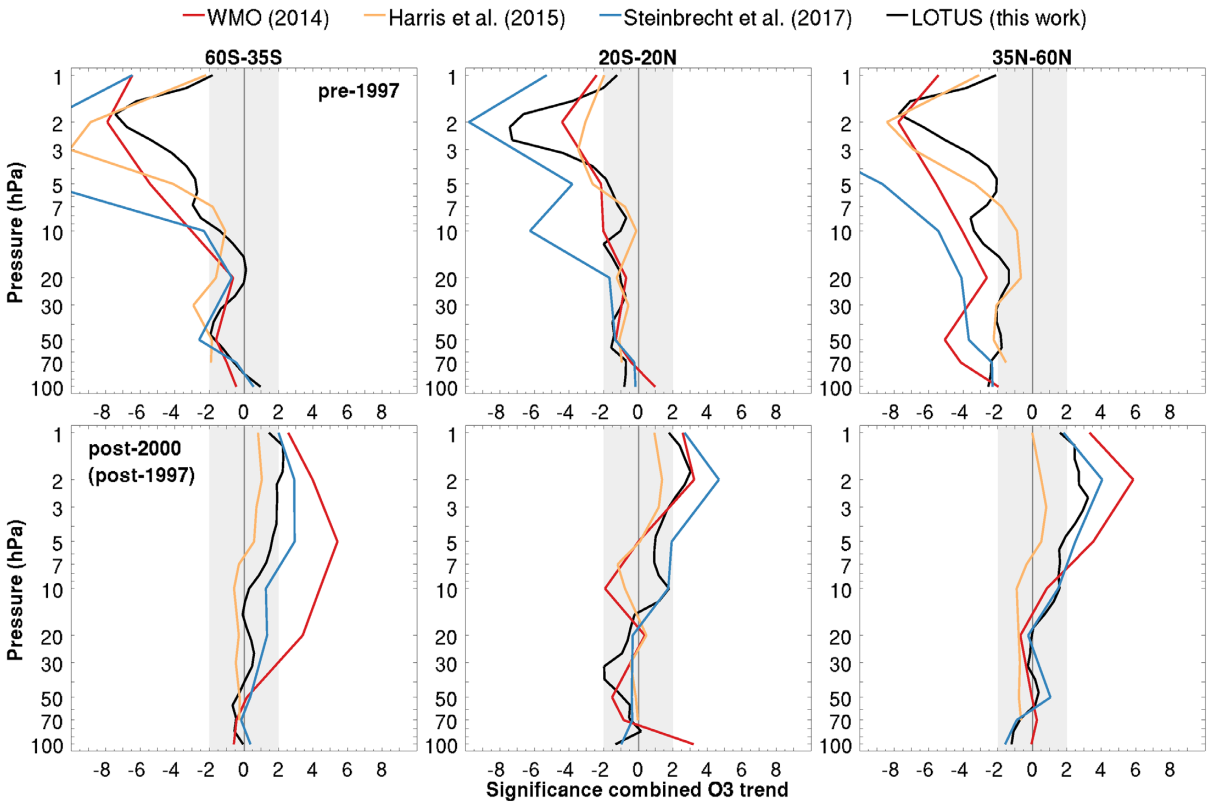

Figure 5.12: *As in* ***Figure 5.12*** *but for the significance of ozone profile trends from past and recent assessments.*

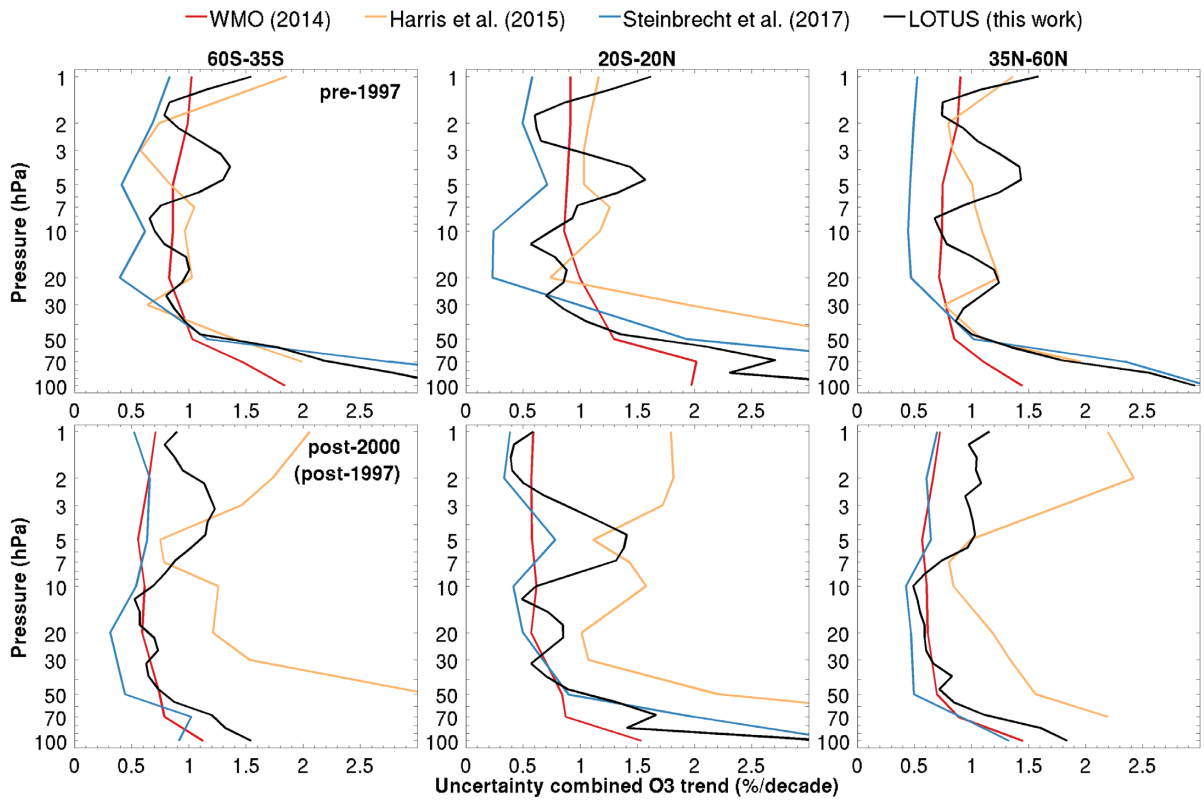

Figure 5.14: *As in* ***Figure 5.12*** *but for the uncertainty (1-sigma) of ozone profile trends from past and recent assessments.*

Nevertheless, *H15* trends in broad-band regions are consistent with LOTUS and other published analyses in the pre-1997 period, while detecting smaller trends in the upper stratosphere for NH and SH mid-latitudes, and stronger negative trends in the tropical mid-stratosphere over the post-1997 period. However, *H15* used the very conservative J-distribution approach (*Section 5.3.3*) to compute combined uncertainties and thus most of the post-2000 trend uncertainties are likely overestimated and consequently the trends are not statistically significant (*Section 5.3.4*).

Ball et al. (2017, 2018) have taken a less traditional approach for evaluation of long-term trends. The satellite datasets used in the LOTUS Report were re-combined using a Bayesian statistical approach to obtain a single BASIC time series. In this approach, common variability between multiple data sets is given greater weight than data sets that deviate from the group, thus reducing the influence of data with non-physical offsets and drifts. For trend determination, the authors used DLM (*Laine et al.*, 2014; *Ball et al.*, 2017, 2018) instead of the MLR used in the LOTUS Report. It is hard to compare the MLR and DLM results directly, because the background changes are nonlinear and the trend is represented as a change between the beginning and end of the period.

Ball et al. (2017, 2018) trends are similar but not identical to the LOTUS analyses and results are hard to compare directly given the different methodologies. The LOTUS MLR model applied to the BASIC data set shows broad agreement with trends derived from other datasets (*Section 5.1* and *Supplement*). Generally, DLM and LOTUS results agree in the upper stratosphere, while the DLM significance in the post-2000 period is lower (*i.e.*, has larger uncertainties). In addition, the NH and SH trends based on the DLM-BASIC analysis are approximately symmetric, while larger asymmetry in pre-1997 period is found in the LOTUS MLR analyses. Future versions of the BASIC dataset that use a Bayesian approach to combine data records will be more applicable to analysis of individual satellite records rather than previously merged records.

5.7 Summary

In this chapter we analysed and compared trend profiles from multiple satellite-based merged data sets and ground-based station data, as well as the estimated mean trend and uncertainties computed from the individual satellite data set trends. For comparison purposes, the multi-model mean/median of CCMI simulations were also shown.

The ozone trends derived from satellite measurements at their native resolutions show a very similar, though oppositely signed, pattern for the pre-1997 (1985–1996) and post-2000 (2000–2016) trends (*Section 5.1*). The pre-1997 results show a general pattern of negative (4–9 % per decade) trends in the upper stratosphere (above 5 hPa / 35 km), which is consistent with previous findings and post-2000 trends from the different satellite data sets show broadly positive trends (~2–3 % per decade on average) in the mid-latitudes of both hemispheres between 5 hPa and 2 hPa (around 37 km to 45 km).

Differences between results from the analysed data sets might be caused by differences in merging techniques, conversions between different native units, consideration of sampling biases, and inherent instrument measurement uncertainty.

Broad latitude band averages (60°S–35°S, 20°S–20°N, and 35°N–60°N) of data from the different satellite data sets were calculated and their trends analysed to provide information for different specific geographical regions (SH and NH mid-latitudes and tropics). The ozone trends in broad bands from individual merged datasets were inter-compared and the method for evaluation of overall combined trends and their uncertainty was proposed. The LOTUS method not only incorporates the influence of systematic sources of uncertainty but also improves the calculation of uncertainties by considering the correlations between contributing trends (*Section 5.3*).

The combined trend profile from satellite measurements (*Section 5.5*) shows significant ozone decline in the pre-1997 period for all three broad latitude bands of -5.9 % ± 1.9 % per decade (NH mid-latitudes), -4.8 % ± 1.3 % per decade (tropics), and -6.2 % ± 1.8 % per decade (SH mid-latitudes) in the upper stratosphere. Post-2000 trends are less clear, with only the upper stratosphere in the NH mid-latitudes showing a clear significant ozone increase (3.1 % ± 1.9 % per decade). This value is in line with previous studies. Trends in the upper stratosphere in the tropics and SH mid-latitudes are barely significant (1.5 % ± 1.4 % per decade and 2.1 % ± 1.9 % per decade, respectively).

Ground-based trends in broad bands support these satellite results. The agreement is improved when there are multiple stations that are used for averages and when ground-based trends are similar over the broad-band latitude and altitude regions. However, noticeable differences between Northern and Southern tropics suggest that ground-based and satellite-based trends should be compared over smaller latitude band widths. Although trends derived from satellite records limited to station overpass and averaged over 5° latitude bands agree well in the stratosphere and at different ground-based station locations, it is still advantageous to compare ground-based data with overpass satellite records to reduce spatial inhomogeneities.

Differences between instrument-based trends for co-located records need to be investigated further. Among potential causes that can impact trends and associated uncertainties are the limited or temporally inhomogeneous frequency of measurements, clear-sky sampling biases (which impact trends in lower stratosphere), differences in spectroscopic databases used to retrieve ozone profiles from initial raw measurements, differences between reanalyses-based temperature profiles used for altitude/pressure conversions and data processing, and other details of retrieval algorithms used for instrument-specific data processing. Analyses and comparisons done for this Report have brought attention to the opportunity to study instrumental artefacts at the "super stations" (*i.e.*, Lauder and Hilo) that have a wide range of co-located data sources.