Chapter 2: Observations and model data

This chapter describes the data sets used to create the ozone profile climate data records that were intercompared and analysed for trends in later chapters in this Report. *Section 2.1* describes the ozone profile data sets from ground-based and in situ instruments, *Section 2.2* is dedicated to ozone data records from satellite instruments. Finally, *Section 2.3* describes the ozone profiles produced by the chemistry-climate and chemistry-transport models of the CCMI.

2.1 Ground-based observations

We start with a brief review of the measurement techniques, data characteristics, and recent changes in the ozone profile data records collected by ground-based instrumentation. More detailed information can be found in *Hassler et al.* (2014) and references therein. The second part of this section presents the methods used to create monthly zonal mean data from these ground-based records, highlighting data set-specific limitations in spatial and temporal sampling. These broad-band, zonally averaged anomaly time series are the input to the trend analyses in *Chapter 5*.

2.1.1 Measurement techniques

2.1.1.1 Ozonesonde

Ozonesondes are a widely used method for measuring in situ ozone vertical distributions up to altitudes of 30–35 km. The balloon-borne electrochemical ozonesondes are small, lightweight, and compact instruments and ozonesonde records at several measurement stations provide the longest ozone profile time series available, with some starting in the 1960s. Ozone profiles are obtained with a height resolution of about 100–150 m. The sensing device is interfaced to a standard meteorological radiosonde for data transmission to the ground station and additional measurements of pressure, temperature, and wind speed. Three major types of ozonesondes have been in use since the 1970s (*e.g.*, *Smit,* 2012a): Electrochemical concentration cell (ECC), Brewer-Mast (BM), and carbon iodine cell (KC). Nowadays most stations have adopted the ECC ozonesonde type developed by *Komhyr* (1969).

A comprehensive review of the performance of the different ozonesondes in terms of precision and accuracy is given in *SPARC-IOC-GAW Assessment of Trends in the Vertical Distribution of Ozone* (*SPARC*, 1998). The assessment also showed inconsistencies in trends derived from data gathered from different sounding stations. A summary and update of the review have been given by *Hassler et al.* (2014) as part of the SI2N assessment. Overall, in recent decades, the random error component of sonde measurements is generally within ±5–10 % between the tropopause and altitudes less than 26 km for all types of sondes. Systematic biases between all types of ozonesondes or compared to other ozone sensing techniques are smaller than ±5–10 %. Above about 26 km altitude the results are not conclusive and the measurement behavior of the sonde types differs. The uncertainty at the top of the measured profile depends on the type of ozonesonde and sensor solution. For example, BM sondes systematically underestimate ozone with increasing altitude (*i.e.*, -15 % at 30 km altitude) (*De Backer et al.,* 1998; *Stuebi et al.,* 2008), while KC sondes tend to overestimate ozone by 10–20 % at altitudes above 30 km (*SPARC,* 1998; *Deshler et al.,* 2008). Intercomparison studies (*e.g.*, *Smit et al.,* 2007; *Smit and ASOPOS panel,* 2012b) indicate that the response of ECC sondes between 28 km and 35 km depends on the type of ECC sonde and sensing solution applied (*i.e.*, 10–20 % differences at altitudes near 35 km).

However, laboratory studies (*Johnson et al.,* 2002) and international intercomparisons like the Jülich Ozone Sonde Intercomparison Experiment (JOSIE; *Smit et al.,* 2007) and the Balloon Experiment on Standards for Ozone Sondes (BESOS; *Deshler et al.,* 2008) have also clearly demonstrated that even small differences in sensing techniques, sensor types, or sensing solutions can introduce significant inhomogeneities in the long-term sounding records between different sounding stations or within each station individually. Therefore, existing artifacts in long-term sounding records have to be resolved by homogenisation either in space (between different stations) or in time (long-term changes) through the use of generic transfer functions which have been derived from intercomparison experiments (*e.g.*, JOSIE or BESOS) and dual balloon soundings (*Deshler et al.,* 2017). A major goal of the Ozone Sonde Data Quality Assessment (O3S-DQA), which is part of this LOTUS assessment, is to reduce the uncertainties between long-term sounding records from 10–20 % down to 5–10 % through the use of generic transfer functions (*Smit and O3S-DQA panel,* 2012). Currently, a total of about 30 long-term station records have been reevaluated and homogenised through resolving known instrumental bias effects, thereby reducing the uncertainties down to 5–10 % (*Tarasick et al.,* 2016; *Van Malderen et al.,* 2016; *Deshler et al.,* 2017; *Sterling et al.,* 2018; *Witte et al.,* 2017, 2018). Some of these recently homogenised ozonesonde data sets are part of this LOTUS assessment.

There is still a potential for sudden changes in future records that can be created by abrupt radiosonde changes (often due to financial burden) or manufacturing changes, which has caused problems in the past. The ozonesonde community, including sonde manufacturers and station operators, recently performed a new JOSIE campaign where they assessed the methods and techniques used by stations in the Southern Hemisphere ADditional OZonesondes (SHADOZ) network. These exercises help to identify inconsistencies in operations and resolve changes to the stability of the record.

2.1.1.2 Lidar

Ozone lidar (Light Detection and Ranging) vertical distribution measurements are based on the Differential Absorption Lidar (DIAL) method that uses the emission of two laser wavelengths (so-called "on" and "off" wavelengths) characterised by a different ozone absorption cross-section. Range resolved measurements are provided by the use of pulsed lasers. The ozone number density is retrieved from the slope of the lidar signals originating from the atmospheric scattering of both laser wavelengths towards the optical receiving system. These signals have to be corrected for differential Rayleigh and Mie scattering as well as for differential absorption by other constituents. The laser wavelengths are chosen so that these corrections represent less than 10 % of the main term linked to ozone absorption. For stratospheric ozone measurements, the on-wavelength is usually generated by an XeCl excimer laser at 308 nm. For the non-absorbed wavelength, different techniques are used, among which the most common are the generation of a wavelength at 353 nm by stimulated Raman scattering in a cell filled with hydrogen or the use of the third harmonic of a Nd:Yag laser emission (355 nm). A more detailed description of the ozone lidar measurement technique can be found, for example, in *Mégie and Menzies* (1980), *Pelon et al.* (1986), and *Godin-Beekmann et al.* (2003). Long-term ozone lidar measurements are currently performed at several stations of the NDACC. Data records of more than 20 years are available at Haute-Provence Observatory (France), Hohenpeissenberg (Germany), Table Mountain (California, USA), Mauna Loa Observatory (MLO; Hawaii, USA) and Lauder (New Zealand). In addition, standardised definitions for the vertical resolution and uncertainty budget of the NDACC lidar ozone measurements were recently published (*Leblanc et al.,* 2016a, 2016b). The uncertainty in lidar ozone profiles ranges from a few percent below 20 km to more than 10–15 % above 45 km with vertical resolution decreasing as a function of altitude, ranging from ~0.5 km below 20 km to several kilometers above 40 km (*Godin et al.,* 1999; *Leblanc and McDermid,* 2000; *Leblanc et al.,* 2016b). Instrumental artifacts that can affect the stability of a long-term lidar record include changes in optical receiver configuration and alignment of the laser beams within the field of view of the telescope (impacting the slope of the lidar signals), changes in laser power, and changes in telescope area (impacting mainly the top of the profiles). Concerning the ozone number density retrieval, undocumented changes of ozone absorption cross-section values between data processing versions can introduce systematic biases throughout the profile.

2.1.1.3 Microwave radiometer

Microwave ozone radiometers (MWR) measure the spectra of emission lines produced by thermally excited, purely rotational ozone transitions at millimeter wavelengths. The pressure broadening effect of the line allows the retrieval of a vertical ozone profile from the measured spectrum by the use of an a priori profile, a radiative transfer simulation and the optimal estimation method based on *Rodgers* (2000). The rotational ozone transitions are measured at either 142.175 GHz or 110.836 GHz depending on the instrument. The instrument principally consists of a millimeter wave receiver and multichannel spectrometer. The measured signal is amplified and down-converted to a lower intermediate frequency which can be processed by a spectrometer. The instruments are calibrated by substituting the radiation from the sky with the thermal radiation from two black body sources at the receiver input. One source is at ambient temperature or heated and stabilised (~300 K) and the second source is cooled with liquid nitrogen at 77 K. The attenuation of the ozone signal in the troposphere is determined by measuring the tropospheric thermal emission and relating the tropospheric opacity to its emission using a radiative transfer model (*Hocke et al,* 2007; *Hassler et al,* 2014).

Ozone profiles between 20 km and 70 km altitude are given in volume mixing ratio (VMR; given in ppmv) and the pressure grid on which data are provided varies by instrument. The vertical resolution is typically 8–10 km between 20 km and 40 km, increasing to 15–20 km at 60 km (*Studer et al.,* 2013; *Nedoluha et al,* 2015, *Maillard-Barras et al.*, 2009).

The total error includes systematic error, random error, and the smoothing error term, which can be determined for each ozone profile. Based on a standard integration time of 1 h, the random and systematic errors are on the order of 3–5 % while the total error is on the order of 7–10 % in the stratosphere. The total error increases up to 20 % at 20 km and to 30–35 % at 70 km (*Studer et al.,* 2014). Lauder MWR agrees within 5–10 % with lidars and the Stratospheric Aerosol and Gas Experiment (SAGE) II between 22 km and 43 km (*McDermid et al.,* 1998). MLO MWR agrees within 10 % with lidars, Dobson Umkehr, and the Upper Air Research Satellite (UARS) Microwave Limb Sounder (MLS) at almost all altitudes, and it agrees within 5 % in the 20–45 km region (*McPeters et al.,* 1999). The 15-year climatological mean difference between the Bern (*Studer et al.,* 2013) and Payerne MWRs is within 7 % from 25 km to 65 km (*Eliane Maillard-Barras, private comm.*). Additional information on microwave radiometers can be found in *Hassler et al.* (2014).

2.1.1.4 FTIR

The ground-based FTIR (Fourier-Transform InfraRed) ozone observations are coordinated by the Infrared Working Group of NDACC. Within this network, the measurements are performed over the 600–4500 cm^{-1} spectral range, using primarily high-resolution spectrometers such as the Bruker 120M (or 125M) or Bruker 120HR (or 125HR), which can achieve a spectral resolution of 0.0026 cm^{-1}. The source of light being the sun, the spectra are recorded only during day-time and under clear sky conditions. The average number of measurements per day among all the stations is about two, with a mean of eight days of measurements per month. Despite the dependence on weather conditions, the average number of measurements per month remains very stable over the full FTIR time series.

In addition to total columns retrieved from the absorption line areas, low vertical resolution profiles can be obtained from the temperature and pressure dependence of the line shapes. The absorption line shapes also depend on the instrumental line shape, with the latter needing to be monitored regularly using gas cell measurements, and are retrieved in a harmonised way within the network (*Hase et al.,* 1999).

The profile retrievals are derived using one of the two different algorithms: PROFITT9 (*Hase,* 2000) and SFIT2 or its recent update SFIT4 (*Pougatchev et al.,*1995), both based on the optimal estimation method (*Rodgers,* 2000). The retrieval settings (*i.e.*, spectral window optimised for ozone, a priori information, *etc.*) have been harmonised within the network (see *Vigouroux et al.,* 2015 for details). There are four or five degrees of freedom for signal (DOFS) from the ground up to about 45 km. Four layers with about one DOFS can therefore be defined to provide partial columns with almost independent information (see **Figure 1** of *Vigouroux et al.,* 2015): roughly one in the troposphere and three in the stratosphere. The three partial columns used in LOTUS are located in the following altitude ranges: 12–21 km; 21–29 km; and 29–48 km, according to the FTIR vertical resolution which is between 7 km and 15–20 km depending on altitude.

The random uncertainties on these three partial columns is about 5 % (*Vigouroux et al.,* 2015; details on error budget can also be found in *García et al.,* 2012). The systematic uncertainty is dominated by the uncertainties on the spectroscopic parameters (HIgh Resolution TRANsmission (HITRAN) 2008 for the current analysis) and is about 3 % for each partial column.

Instruments and data archives	Stations (start of data record)		
	60°S –35°S	**20°S – 20°N**	**35°N – 60°N**
Ozonesonde (0–30km) http://www.ndacc.org, http://www.woudc.org/data, https://tropo.gsfc.nasa.gov/shadoz/Archive.html	Lauder (1986), Macquarie Island (1994), Broadmeadows (1999)	Hilo (1982), Ascension Island (1998), Nairobi (1998), Natal (1998), Pago Pago (1998), Kuala Lumpur (1998), Suva (1998), Hong Kong Observatory* (2000, 22.3°N)	Goose Bay (1963), Uccle (1965), Hohenpeißenberg (1966), Payerne (1968), Edmonton (1970), Wallops Island (1970), Lindenberg (1975), Legionowo (1979), Praha (1979), Boulder (1991), De Bilt (1992), Valentia (1994), Huntsville* (1999, 34.7°N)
Lidar (10–50km) http://www.ndacc.org	Lauder (1994)	Mauna Loa (1993)	OHP (1985), Hohenpeißenberg (1987), Table Mountain (1988)
Microwave (20–70km) http://www.ndacc.org	Lauder (1992)	Mauna Loa (1995)	Bern (1994), Payerne (2000)
FTIR (0–50km) http://www.ndacc.org	Lauder (2001), Wollongong (1996)	Izana* (1999, 28.3°N)	Jungfraujoch (1995)
Dobson/Brewer Umkehr (0–50km) ftp://aftp.cmdl.noaa.gov/data/ozwv/DobsonUmkehr/Stray%20light%20corrected/monthlymean	Perth (1984), Lauder (1987)	Mauna Loa (1984)	Arosa (1956), Boulder (1984), OHP (1984)

Table 2.1: *Overview of the sources of ozone profile observations by ground-based techniques used for the monthly zonal mean data considered in this Report. Stations are sorted chronologically by start year of the record; those with an asterisk are located slightly outside the attributed latitude zones.*

2.1.1.5 Umkehr

The Umkehr measurement makes use of the zenith sky UV radiation changes during the sunset or sunrise hours of the day. The earliest measurements were recorded by *Götz* (1931) at Arosa, Switzerland. Two spectral UV wavelength regions (311.4 nm and 332.5 nm) selected for the Umkehr method are subject to different levels of ozone absorption. The zenith sky ratio between two spectral channels changes with the elevation of the sun. Measurements begin at 60° solar zenith angle (SZA); the ratio gradually increases up to 85° SZA and then decreases between 85° and 90° SZA. Umkehr measurements from Dobson instruments have been collected operationally since the 1957 International Geophysical Year at a few select stations, but additional Dobson observing stations became available in the 1980s. The trend-optimised algorithm was developed by *Petropavlovskikh et al.* (2005) to derive morning and afternoon daily ozone profiles in 10 Umkehr layers based on a pressure layer system. A priori information is used to solve the optimum statistical inverse problem (*Rodgers,* 2000). The method is designed to derive ozone profiles with a vertical smoothing technique (defined by averaging kernels). This approach affects the accuracy of the retrieved ozone in a particular layer by weighting ozone variability from adjacent layers. Therefore, the method adds error in the layer-retrieved ozone amount, which is estimated to be about 5% in the stratosphere. However, this error does not impact trend analyses as it is constant in time. Time series of Umkehr ozone profiles are retrieved with the UMK04 algorithm (*Petropavlovskikh et al.,* 2005). A generic stray light correction is applied to reduce systematic biases in Umkehr retrieved profiles (*Petropavlovskikh et al.,* 2009, 2011).

2.1.2 Deseasonalised monthly mean time series

2.1.2.1 Procedure

The trend analyses in *Chapter 5* are performed on monthly averaged deseasonalised data collected at a number of stations and by five types of instruments (ozonesonde, lidar, FTIR, MWR, and Umkehr). Profiles from each instrument record are first averaged by month, separately for each station. Months with an insufficient number of profiles are discarded from further analysis. The selection criteria depend on the instrument technique (*Sections 2.1.2.2–5*). These time series created for each instrument are referred to as the Station Monthly Mean (SMM) dataset. Subsequently, the deseasonalisation process is performed in two steps.

In step 1, a site-specific seasonal cycle is computed as the average, for each calendar month (Jan, Feb, …, Dec), of all SMM data in the reference period (Jan 1998 – Dec 2008). Months with an insufficient number of years over the reference period (typically <7 but also measurement technique dependent, see below) are flagged and excluded from further analysis. This requirement ensures a more accurate determination of the observed seasonal cycle but is only satisfied for a select number of sites. This data set is referred to as the Station Seasonal Cycle (SSC) data set, one per site and per instrument.

In step 2, we compute the relative difference of each monthly mean value (SMM) to the observed climatological mean value (SSC) for that month. These deseasonalised relative anomaly time series are referred to as the Station Monthly Mean Anomaly data set (SMMa) and are defined as

$$SMMa(site,p,t) = 100 \times \frac{SMM(site,p,t) - SSC\big(site,p,m(t)\big)}{SSC\big(site,p,m(t)\big)} \quad (2.1),$$

where p stands for vertical grid level (pressure or altitude), t represents time (*i.e.*, month) and $m(t)$ the corresponding calendar month (*i.e.*, Jan, Feb, …). Hence, by construction, the (dominant part of the) seasonal cycle is removed from SMMa data and the absolute level averages to zero over the reference period. In addition, any instrument-related constant multiplicative offsets (*i.e.*, bias) are thereby removed as well.

The deseasonalisation step is motivated by the need to combine, for a wide latitude belt, the data from multiple sites, each potentially exhibiting a different bias. A Zonal Monthly Mean Anomaly data set (ZMMa) is obtained by averaging the SMMa data from each station located within the broad zonal bands. We create ZMMa's for three broad latitude bands: 60°S–35°S, 20°S–20°N, and 35°N–60°N. Only the stations listed in **Table 2.1** are used for the broad zonal bands. **Figure 2.1** demonstrates ozone anomaly time series of ZMMa ozonesonde records in the 35°N–60°N (top), 20°S–20°N (centre), and 60°S–35°S (bottom) latitude bands as a function of altitude (ground to ~30 km).

Site-dependent instrument biases can generate, in a multi-station average of SMM data sets, not only random uncertainty but also discontinuities (due to differences in time coverage). However, such sources of error are suppressed in a multi-station average of SMMa data sets (the ZMMa data sets). The intercomparisons described in *Chapter 3* (*Section 3.1.1*) identify a number of stations with clear inhomogeneities in the time series. The availability of multiple sites is expected to reduce the impact of spatial and temporal inhomogeneities in the combined ground-based data records. However, it is important to realise that it is not uncommon that a latitude belt contains just one site for the considered measurement technique.

ZMMa are created for each instrument technique separately and with equal weight given to all sites within the band (**Figures 2.2–2.6**). This effectively gives more weight to regions with more stations (*e.g.*, Europe and North America). The data from different instrument techniques are not combined in this study due to complications associated with differences in sampling frequencies, vertical smoothing and the use of different measurement units. The time series of ground-based station and zonally averaged ozone anomalies are available from the LOTUS Report data depository.

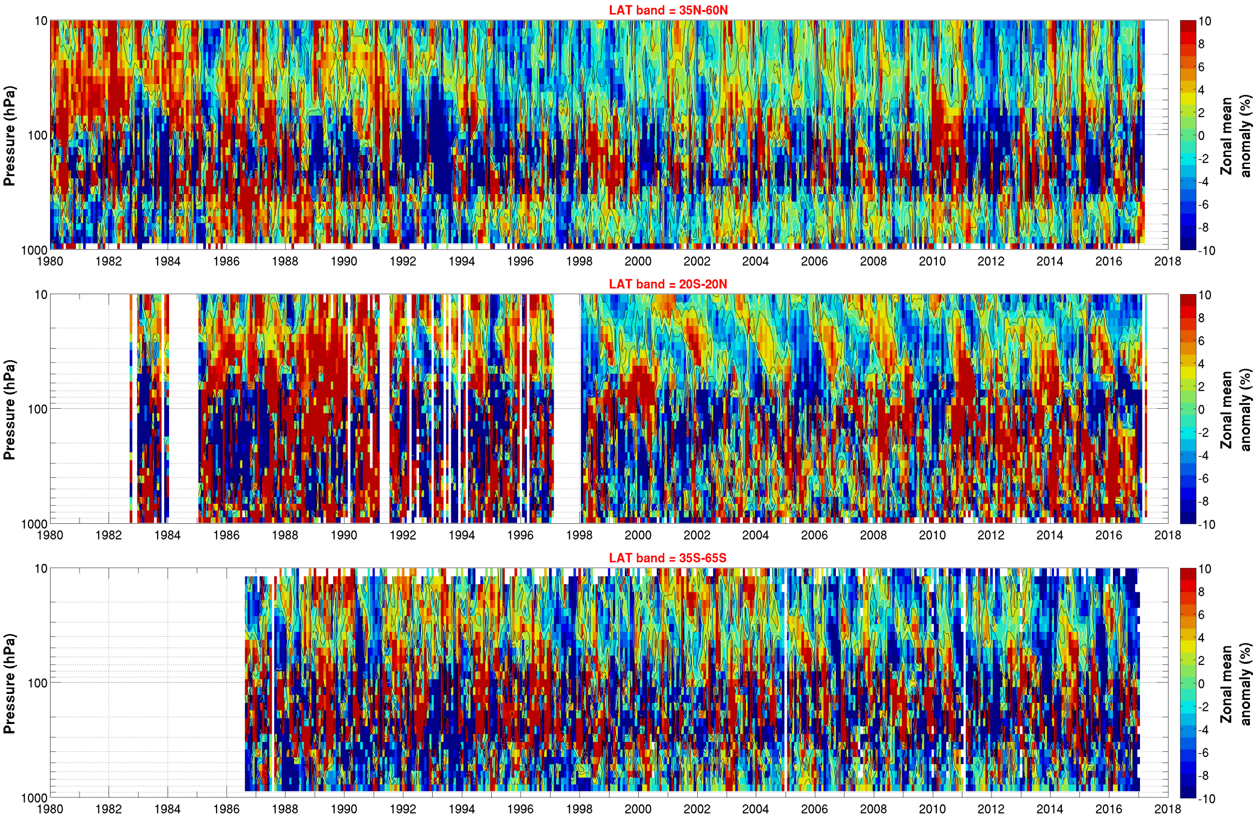

***Figure 2.1:** Example time series of monthly zonal mean relative deseasonalised anomalies computed from ozonesonde data in the 35°N–60°N (top), 20°S–20°N (centre), and 60°S–35°S (bottom) latitude bands as a function of altitude (ground to ~30 km). The sonde stations used for each band are listed in **Table 2.1**.*

Trends derived from each ZMMa time series are reported in *Chapter 5*, *Section 5.4*. The only ground-based records considered individually in this Report are the ozonesonde records from Hilo, Hawaii and Lauder, New Zealand. The trends derived from these two ozonesonde records are used for discussion of consistency in trends obtained from multiple instruments co-located at these locations (see *Chapter 5*, **Figure 5.10**).

2.1.2.2 Ozonesonde

Ozonesonde observations were retrieved from the public NDACC, World Ozone and Ultraviolet Radiation Data Centre (WOUDC), and SHADOZ data archives. The station data record differs sometimes between archives, due to different processing settings, different time periods covered, *etc*. Therefore, for a given site, the data from different archives was not mixed in order to avoid introducing inhomogeneities. Only half of the sites report total ozone normalisation correction factor (CF) values, some of these have applied the CF to the original profiles while others have not. To avoid losing a large number of sites where the CF data is missing, this information is not used to correct the reported data nor to screen the observations. Instead, the data are screened according to the criteria outlined in *Hubert et al.* (2016). German Democratic Republic sondes (GDR), mainly flown prior to the 1990s in Eastern Europe, have larger uncertainties and these profiles are hence not used (*Liu et al.*, 2009). Flights that do not reach 20 hPa are rejected as well, to avoid additional uncertainty in case the profile was normalised to a total ozone column. The VMR profile is then integrated in the pressure domain to obtain ozone partial columns of ~1 km thickness from the surface to 30 km altitude. The entire profile is discarded if at least 10 out of 30 layers are missing (quality-screened) input data.

2.1.2.3 Lidar and microwave radiometer

The monthly mean ozone profiles for lidar and microwave observations are obtained by averaging the ozone profiles available in the NDACC database (**www.ndacc.org**). For most stations, we used the profiles from the (monthly) National Aeronautics and Space Administration (NASA) Ames data files, while the recent profiles by the Bern MWR were taken from the hierarchical data format data files. Monthly mean ozone profiles for the Hohenpeißenberg lidar were obtained in a slightly different way, by retrieving the monthly mean lidar return signal (which results in some improvement above 40 km). Individual profiles are weighted by measurement length. Most stations report profiles as number density ($10^{16}\,m^{-3}$) versus altitude. For the historic microwave ozone data from Bern and Payerne stations, however, only VMR versus altitude are available.

The altitude resolution of individual microwave profiles is on the order of 10 km to 20 km. For the lidars it varies between ~0.5 km (at 15 km) to more than 5 km (above 40 km).

For monthly means, the altitude resolution is less relevant, because atmospheric changes tend to average out and tend to be coherent over many kilometers.

Three lidar and two microwave stations are available for the 35°N–60°N broad-band averages, whereas for 20°S–20°N and 60°S–35°S broad bands only single station records are available for comparisons with satellite records (see *Section 5.4*).

2.1.2.4 FTIR

As mentioned previously, FTIR solar absorption measurements are taken during the day only and only during clear-sky conditions. There are on average about three measurements per day and eight days of measurements per month. The random errors are determined by the smoothing error, which is one of the dominating error sources in FTIR profile retrievals (*Vigouroux et al.,* 2015) and is about 5 % for the three layers provided for LOTUS analyses. The systematic errors are about 3 % for the three layers. The standard deviation of the monthly means and the number of measurements used in the monthly means is also provided in the FTIR datafiles.

Two FTIR records are averaged to represent 60°S–35°S broad-band ozone variability and trends. Single station records are available for comparisons in the other two broad latitude bands (see case study in *Section 5.4*).

2.1.2.5 Umkehr

Monthly averages for Umkehr time series are calculated from all data that have passed the quality assurance (*i.e.*, iterations less than three, standard deviation of the difference between Umkehr simulated and observed values in the final retrieval less than observation uncertainty). Umkehr measurements in the years following the eruptions of El Chichón (1982–1984) and Pinatubo (1991–1993) were affected by scattering from aerosols injected into the stratosphere. These effects are not taken into account by the forward model, thus creating erroneous ozone profile retrievals. The post-processing corrections do not remove errors completely. Therefore, for trend analyses the monthly averaged Umkehr data during volcanic periods are marked as missing.

Three Umkehr records are available for the 35°N–60°N broad-band averages and two stations are used to represent the 60°S–35°S belt, whereas for 20°S–20°N only a single station record is available for comparisons with satellite records (see case study in *Section 5.4*).

2.1.2.6 Instrument and station measurement frequency

Figures 2.2 to **2.6** show the number of measurements per month for the ozonesonde, lidar, MWR, FTIR, and Umkehr techniques at all stations that are used for trend analyses in this Report. Frequency of observations varies from station to station over the records, which likely depends on the fluctuation in funding available from the supporting national programs. The minimum number of observations (two or more) required to accept a monthly mean value in the SMM dataset (see *Section 2.1.2.1*) depends on the instrument technique. In part, these rather low numbers (when compared to what is used by the satellite community) reflect limitations due to observational conditions and the sonde launch schedule.

Ozonesondes (**Figure 2.2**) are launched in all weather conditions, typically following a fixed schedule on the same day(s) of the week or month. Three European stations (Payerne, Uccle, and Hohenpeissenberg) launch sondes three times a week, while most stations do so once a week. The SHADOZ sites, located in the tropics, launch twice a month. Uncertainties in the derived monthly mean values are reduced by rejecting months and grid levels with <2 (tropics) or <3 (elsewhere) observations. Seasonal cycle entries for ozonesonde records are discarded for months and grid levels that contain <6 years (tropics) or <7 years (elsewhere) of SMM data over the reference period.

For the ground-based observations, at least two measurements are required for lidar (**Figure 2.3**), microwave (**Figure 2.4**), and Umkehr (**Figure 2.5**), while at least three measurements are required for FTIR (**Figure 2.6**). Lidars measure during clear-sky nights only and report just one profile per night. Microwave radiometers, on the other hand, measure continuously under most weather conditions and report half hourly, hourly, or six hourly profiles depending on the site. Umkehr profiles are retrieved on days of (mostly) clear sky conditions and can have two measurements per day. However, each station will have a different maximum number of days per month depending on local weather conditions (*e.g.*, overcast). The FTIR measurements also require fair weather conditions and therefore have a similar limitation on the number of profiles per month, which vary for latitude and season. Seasonal cycle entries for ground-based records are discarded for months and grid levels that contain <6 years (tropics) or <7 years (elsewhere) of SMM data over the reference period.

The non-uniform temporal sampling can have an impact on the seasonal cycle derived from each instrument record and its ability to capture the true atmospheric variability. Since composition in the lower stratosphere is strongly affected by meteorological scale variability (*Lin et al.,* 2015), the impact of the sampling frequencies on the station record seasonal cycle should be assessed for each ground-based instrument in this part of the atmosphere. Prior to trend analyses, each ground-based and ozonesonde record is deseasonalised separately prior to combining anomalies; thus, a sampling bias is expected to have small impact on the combined records and derived trends. At the time of this writing, no detailed studies were available on the impact of sampling on differences in ground-based trends. These are recommended for future analyses.

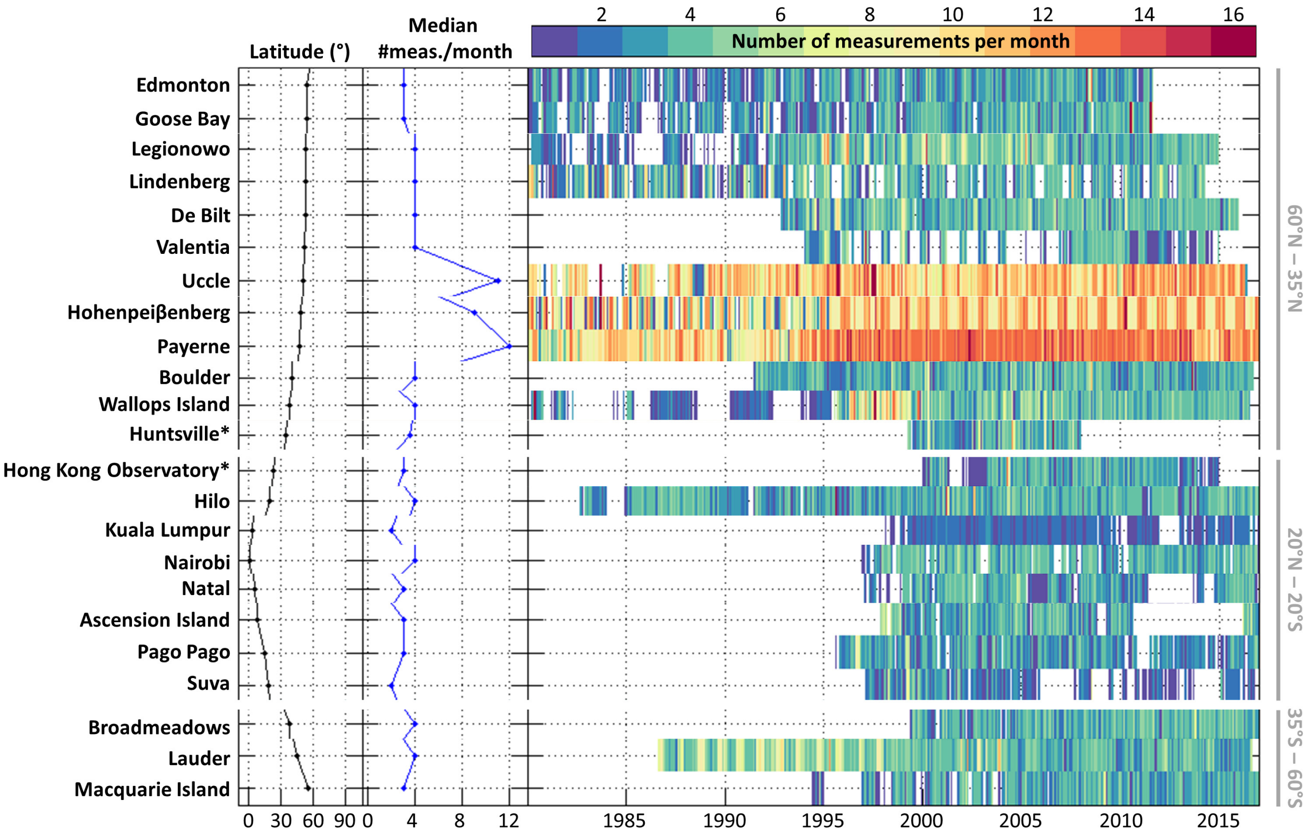

Figure 2.2: *Sampling statistics for ozonesonde station records retrieved from the NDACC, WOUDC, and SHADOZ data archives, sorted North to South. The figure shows the median number of measurements per month over the entire data record (centre) and the number of measurements for each month since 1980 (right, colour scale). Stations with an asterisk are located slightly outside the attributed latitude zones.*

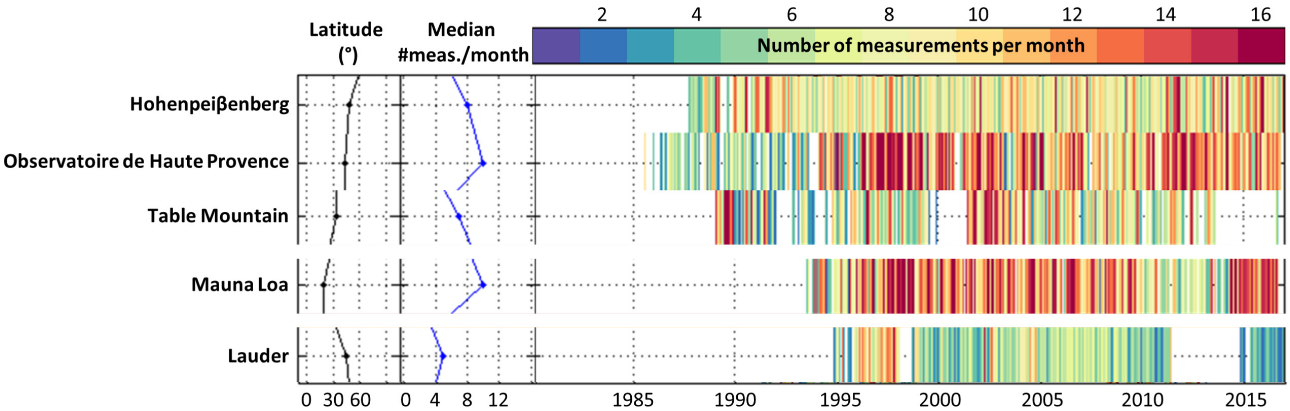

Figure 2.3: *As* ***Figure 2.2*** *but for the stratospheric ozone lidar station records retrieved from the NDACC data archive.*

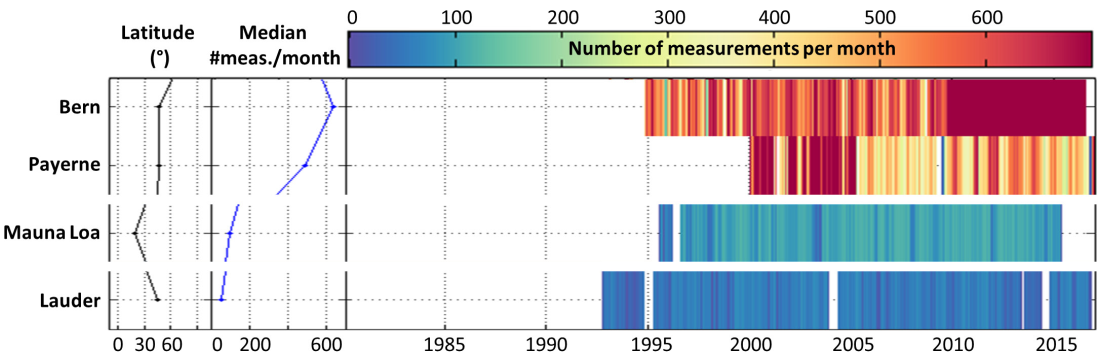

Figure 2.4: *As* ***Figure 2.2*** *but for ozone microwave radiometer station records retrieved from the NDACC data archive. Stations report half hourly, hourly, or six-hourly profiles.*

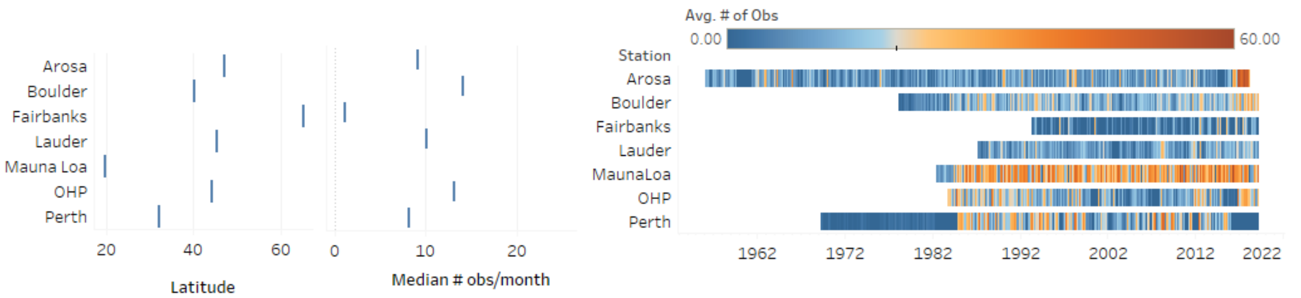

Figure 2.5: *As* ***Figure 2.2*** *but for Dobson Umkehr station records submitted by the record PIs to the LOTUS data archive. Stations report profiles once or twice a day. Note that the time axis differs from that of previous figures.*

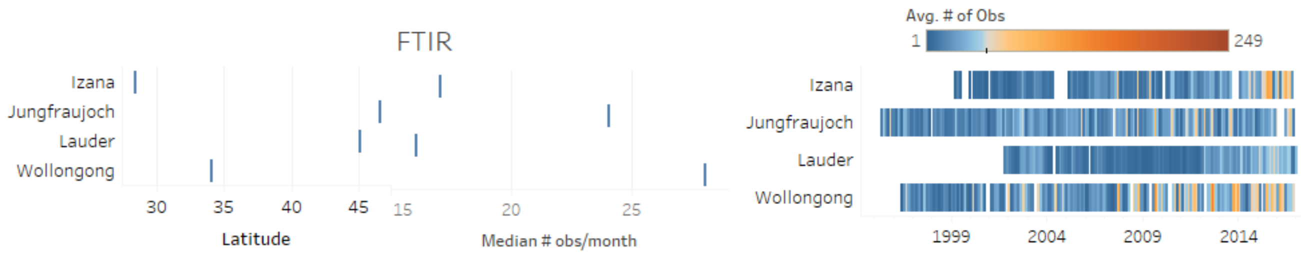

Figure 2.6: *As* ***Figure 2.2*** *but for FTIR station records submitted by the record PIs to the LOTUS data archive. Stations report profiles several times per day. Note that the time axis differs from that of previous figures.*

2.2 Satellite observations

2.2.1 General remarks

The main advantage of satellite instruments is their global coverage. For ozone trend analyses, long-term ozone data sets are needed in order to separate long-term trends from other sources of ozone variability such as solar activity. For the 2014 ozone assessment (*WMO,* 2014), several merged satellite data sets were created: SBUV Merged Ozone Data Set (SBUV MOD) and the SBUV Cohesive data set (SBUV COH), Global OZone Chemistry And Related trace gas Data records for the Stratosphere (GOZCARDS) and Stratospheric Water and OzOne Satellite Homogenized (SWOOSH), as well as SAGE-GOMOS (Global Ozone Monitoring by Occultation of Stars), and SAGE-OSIRIS (Optical Spectrograph and InfraRed Imaging System). Detailed information about these data sets and their intercomparison can be found in *Tummon et al.* (2015). An overview of satellite instruments can be found in for example *Hassler et al.* (2014).

Since the 2014 WMO Ozone Assessment, some of these merged data sets have been extended to 2016 and updated with the most recently processed versions of the ozone profile data sets from the individual satellite instruments. In addition, new merged data sets have been generated. These new merged data sets use revised data records from the individual instruments and rely on improved merging methods. The LOTUS Report evaluates these new data sets for improved accuracy and stability.

This section briefly describes the long-term merged ozone profiles data sets used in the LOTUS study. General information about the merged data sets and their main parameters is summarised in **Table 2.2**. According to measurement technique and ozone representation, the merged satellite data sets are grouped as (1) ozone profiles from nadir sensors, (2) ozone profiles from limb instruments in mixing ratio on a pressure grid, and (3) ozone profiles from limb instruments in number density on an altitude grid. In addition to measurement principles and specific features of retrieval algorithms, such a grouping is also made because ozone trends can be different in different representations due to the influence of stratospheric cooling (*McLinden and Fioletov,* 2011). The influence of the ozone representation on evaluated trends is discussed in *Chapter 5, Section 5.1.2* of the Report. For all satellite data sets, monthly zonal mean ozone profiles are used.

2.2.2 Nadir profile data records

The two nadir-based merged profile data sets in this Report are both based on the series of nine solar backscatter UV (Backscatter Ultraviolet Radiometer (BUV), SBUV and SBUV/2) nadir instruments flown over the period from 1970 to the present on NASA (*i.e.*, Nimbus 4 and Nimbus 7) and National Oceanic and Atmospheric Administration (NOAA; *i.e.*, NOAAs 9, 11, 14, 16, 17, 18, and 19) satellite platforms. The instruments are of similar design and measurements are processed using the same retrieval algorithm (Version 8.6; *McPeters et al.*, 2013; *Bhartia et al.*, 2013). Radiance measurements are calibrated using a variety of hard and soft calibration techniques, including cross-instrument calibration during periods of measurement overlap to further ensure consistency over the record (*DeLand et al.*, 2012). However, despite the instrument similarity and common retrieval algorithm, each instrument experienced unique operational conditions (*e.g.*, instrument degradation, specific on-orbit problems) and orbital characteristics, including measurement time of day, which contribute to differences among the individual records.

SBUV instruments ideally operate in late morning-early afternoon sun synchronous orbits such that measurements are made at small solar zenith angles and at the same local time each orbit. While most instruments were launched into ~2 pm local time orbits, Nimbus 4 and Nimbus 7 measured near noon local time, and NOAA 17 was launched into a ~10 am orbit. Furthermore, NOAA satellite orbits slowly drift towards the terminator, and in some cases drift through the terminator, such that the instrument evolves from making late afternoon measurements to making early morning measurements. Thus the various SBUV instruments are measuring at different local times. This can introduce differences between overlapping measurements due to both real geophysical noise (*e.g.*, diurnal variation) and instrument noise, as the data uncertainty increases when the orbit approaches the terminator (*DeLand et al.*, 2012; *Kramarova et al.*, 2013a; *McPeters et al.*, 2013). The latter is particularly true of the NOAA-9, -11, and -14 instruments, whose orbits drifted faster than other instruments in the series (*DeLand et al.*, 2012; *Kramarova et al.*, 2013a).

The primary source of error in the SBUV retrieval is the smoothing error due to the instrument's limited vertical resolution, particularly in the troposphere and lower stratosphere (*Kramarova et al.*, 2013b; *Bhartia et al.*, 2013). The SBUV instrument has a resolution of 6–7 km near 3 hPa, degrading to 15 km in the troposphere and ~10 km above 1 hPa (*Bhartia et al.*, 2013). *Kramarova et al.* (2013a) showed that SBUV ozone profiles are generally consistent to within 5 % with data from UARS and Aura MLS, SAGE II, ozonesondes, microwave spectrometers, and lidar in the region between 25 hPa and 1 hPa (also see *Frith et al.*, 2017 for updated comparisons with AURA MLS).

Data set	Satellite instruments	Ozone representation	Latitude coverage and resolution	Altitude coverage and vertical sampling	Temporal coverage
SBUV MOD v8.6 (NASA) https://acd-ext.gsfc.nasa.gov/Data_services/merged/index.html	BUV, SBUV and SBUV-2 on Nimbus 4, 7 and NOAAs 11, 14, 16, 17, 18 ,19	Mixing ratio on a pressure grid	80S–80N, 5 deg	50–0.5 hPa, 15 layers (from ~6 to ~15 km)	01/1970 – 12/2016
SBUV COH v8.6 (NOAA) ftp://ftp.cpc.ncep.noaa.gov/SBUV_CDR	SBUV and SBUV-2 on Nimbus- 7 and NOAAs 9, 11, 16, 17, 18, 19		80S–80N, 5 deg	50–0.5 hPa, 15 layers (from ~6 to ~15 km)	01/1978 – 12/2016
GOZCARDS v2.20 https://gozcards.jpl.nasa.gov	SAGE I v5.9_rev, SAGE II v7, HALOE v19, Aura MLS v4.2		90S–90N, 10 deg	215–0.2 hPa, 6 or 12 levels per pressure decade (~3 km)	01/1979 – 12/2016
SWOOSH v2.6 https://data.noaa.gov/dataset/dataset/stratospheric-water-and-ozone-satellite-homogenized-swoosh-data-set	SAGE II v7, HALOE v19, UARS MLS v5, SAGE III v4, Aura MLS v4.2		90S–90N, 10 deg (also 5 and 2.5 deg)	316–1 hPa, 6 or 12 levels per pressure decade (~3 km)	01/1984 – 12/2016
SAGE-OSIRIS-OMPS LOTUS ftp	SAGE II v7, OSIRIS v5.10, OMPS-LP USask 2D v1.0.2	Number density (anomaly) on an altitude grid	60S–60N, 10 deg	10–50 km, 1 level per km	10/1984 – 12/2016
SAGE-CCI-OMPS http://www.esa-ozone-cci.org/?q=node/167	SAGE II v7 , OSIRIS v5.10, GOMOS ALGOM2s v1, MIPAS IMK/IAAv7, SCIAMACHY UB v3.5, ACE-FTS v3.5/3.6, OMPS-LP USask2D v1.0.2		90S–90N, 10 deg	10–50 km, 1 level per km	10/1984 – 07/2016
SAGE-MIPAS-OMPS v2 https://www.imk-asf.kit.edu/english/304_2857.php	SAGE II v7, MIPAS IMK/IAA v7, OMPS-LP NASA v2.5, ACE-FTS v3.5/3.6		60S–60N, 10 deg	6–60 km, 1 level per km	10/1984 – 12/2016

Table 2.2: *General information about merged satellite data sets.*

Inter-instrument biases among the later instruments NOAAs 16–19 (since late 2000) are mostly within 3 %, while biases involving NOAA-9, NOAA-11 descending and NOAA-14 are mostly within 5 % (*Frith et al.,* 2017, see **Figure 5**; *Wild et al.,* 2019).

2.2.2.1 SBUV MOD v8.6

The SBUV MOD time series includes data from all SBUV instruments except NOAA-9, which are excluded due to increased differences with other SBUV and external data sources (*Frith et al.,* 2014; *Frith et al.,* 2017; *DeLand et al.,* 2012; *Kramarova et al.,* 2013a). The combined record provides continuous coverage of ozone profile data since late 1978. As the data have already been inter-calibrated and all known instrument problems resolved, we have no physical rational to choose one data set over another. Therefore, when constructing the merged data set no external calibration adjustments are applied, but rather the data are simply averaged during periods when more than one instrument is operational. This approach relies on the average of multiple measurements to mitigate the effects of small offsets and drifts in individual data sets rather than attempting to choose a single record as a reference calibration. To account for higher uncertainty when orbits approach the terminator, only the subset of measurements with the equator crossing time between 8 am and 4 pm are accepted into the MOD combined time series. The exception to this selection criteria is the record from NOAA-11 ascending (1989–1995) that is entirely accepted to avoid a data gap. Small remaining biases and drifts in the merged record are accounted for in the MOD uncertainty estimates (*Frith et al.,* 2017; also see *Section 3.1.4*). *Tummon et al.* (2015) showed that the MOD record agrees with the mean of other merged ozone data sets within 5 %. The MOD monthly zonal mean data are available at: **https://acd-ext.gsfc.nasa.gov/Data_services/merged/index.html.**

Monthly means are computed for each SBUV instrument separately in 5-degree wide zonal bands. Only bin averages in which the average latitude of the profiles in the bin is within 1 degree from the bin centre and the average time of the profiles is within four days from the centre of the month are included in the MOD record. Measurements are removed for a year after the El Chichón volcanic eruption and for 18 months after the eruption of Mt. Pinatubo to avoid periods when volcanic aerosols likely interfered with the algorithm (*Bhartia et al.,* 2013).

2.2.2.2 SBUV COH v8.6

The SBUV MOD approach of averaging data from all available satellites during an overlap period results in the loss of characteristics of the measurement (*e.g.*, time of measurement). Alternatively, the SBUV COH merging approach is to identify a representative satellite for each time period, thus preserving knowledge of orbital characteristics for each measurement period. Additionally, data in the overlap periods are examined to determine a correction for some satellite records. In the later period of the combined record, the overlaps between NOAA-16 to -19 ozone records are long, and each satellite can be compared and adjusted directly to NOAA-18 (*Wild et al.,* 2019). For example, NOAA-16 at 4–2.5 hPa can differ from NOAA-17 and NOAA-18 by up to 3 % at all latitudes; while the NOAA-17 record differs from NOAA-18 in the mid-latitudes especially in the upper atmosphere at 4 hPa and above where diurnal issues become significant. Recent studies (*Wild et al.,* 2019) show that NOAA-19 also differs from NOAA-18 by approximately 1–2 %. The difference is mostly found in the equatorial regions and between 10 hPa and 6.4 hPa pressure levels. Strong drifts in the early satellites (NOAA-9, -11 and -14) and poor quality of NOAA-9 and NOAA-14 data can create unphysical trends when a successive head-to-tail adjustment scheme is used in the early period (*Tummon et al.,* 2015). The current SBUV-COH data set does not adjust the Nimbus-7 or NOAA-11 data, nor does it include the NOAA-9 ascending node. Only the NOAA-9 descending data is adjusted to fit between the ascending and descending nodes of the NOAA-11 record. NOAA-14 data do not appear in the final data set, but it is used to enable a fit of NOAA-9 descending to NOAA-11 descending where no overlap exists (*Wild et al.,* 2019). The COH data is available at **ftp://ftp.cpc.ncep.noaa.gov/SBUV_CDR** as monthly or daily zonal means both as mixing ratio on pressure level, or as layer data.

The lower quality data from NOAA-9, NOAA-11 descending, and NOAA-14 lead to larger uncertainties (10–15 %) in the mid-1990s (at the time of peak halogen loading and the expected "turn-around" in ozone trends) in both merged data sets and complicate efforts to establish a long-term calibration over the full record (from 1980s to 2000s). Error propagation and trend uncertainty estimates for the SBUV merged records are discussed in more detail in *Chapter 3*, *Section 3.1.4*.

2.2.3 Limb profile data records in mixing ratio on pressure grid

2.2.3.1 GOZCARDS v2.20

The Global OZone Chemistry And Related trace gas Data records for the Stratosphere (GOZCARDS) v1.01 data set, used in the previous ozone assessment (*WMO,* 2014; *Froidevaux et al.,* 2015), has been extended to the present. Recently, a GOZCARDS merged data set v.2.20 has been created. GOZCARDS provides VMRs on a pressure grid for 10-degree latitude bins (starting at 0–10 degrees) and is a combination of various high quality space-based monthly zonal mean ozone profile data. The GOZCARDS pressure levels are regularly spaced in log-space, with 12 (6) levels for each decade change in pressure for pressures larger (smaller) than 1 hPa. The recommended data range is 215 hPa to 0.2 hPa; at tropical latitudes the recommended range is 100 hPa to 0.2 hPa to ensure only stratospheric data are considered. Caution is recommended for the upper stratospheric / lower mesospheric levels, given the existence of incompletely accounted for diurnal and seasonal effects (for both source and merged data, particularly when considering occultation data sets). The GOZCARDS monthly mean ozone record includes SAGE I (version 5.9), SAGE II (v7), the HALogen Occultation Experiment (HALOE; v19) and Aura MLS (v4.2), and covers the period from 1979-2016. SAGE II data are used as a reference for adjusting/debiasing the HALOE and Aura MLS measurements (using overlapping time periods of observation). Details of the screening criteria for each data set, the merging procedure, as well as estimated uncertainties (random and systematic) are provided by *Froidevaux et al.* (2015). This new GOZCARDS version utilises a reduced number of data sources and a finer stratospheric retrieval pressure grid, in comparison to v1.01 (*Froidevaux et al.*, 2015). UARS MLS data are not used, since they are not available on the finer vertical grid of GOZCARDS v2. While interpolation could have been used, an exact treatment of the retrieved uncertainties is not feasible. Data from the Atmospheric Chemistry Experiment - Fourier Transform Spectrometer (ACE-FTS) instrument are not used either, as the updated ACE-FTS v3.6 data version was not available in time for the data creation deadlines. The most significant change is the effect of using the SAGE II v7 data, which uses Modern-Era Retrospective analysis for Research and Applications (MERRA) temperature profiles in the retrievals, and the actual impact of the those temperatures (rather than National Centers for Environmental Prediction (NCEP) temperatures) on the conversion of SAGE II ozone (density on altitude grid) to the VMR on pressure grid used for GOZCARDS ozone. Additionally, Aura MLS v4.2 data are now included (instead of v2.2), along with HALOE v19 profiles which are interpolated to the finer pressure grid before merging. The SWOOSH record also uses SAGE II v7 ozone data and there is now closer agreement and better correlation between the SWOOSH and GOZCARDS v2.20 time series than between SWOOSH and GOZCARDS v1.01 data.

2.2.3.2 SWOOSH v2.6

The Stratospheric Water and Ozone Satellite Homogenized (SWOOSH) database was created by Chemistry Sciences division of NOAA/ESRL (NOAA Earth System Research Laboratory) in Boulder, Colorado, USA. It includes vertically resolved ozone and water vapor data from a subset of the limb profiling satellite instruments operating since the 1980s. An overview of SWOOSH is provided by *Davis et al.* (2016). The primary SWOOSH products are monthly zonal mean time series of water vapor and ozone mixing ratio on 12 pressure levels per decade from 316 hPa to 1 hPa, the same levels as from the Aura MLS instrument. SWOOSH is provided on several zonal mean grids (2.5°, 5°, and 10°), and additional products include two coarse 3D griddings (30° lon x 10° lat, 20° x 5°) as well as a zonal mean isentropic product. Here, the 10° zonal mean product is used.

SWOOSH includes data from SAGE II v7, UARS HALOE v19, UARS MLS v5, SAGE III v4, and Aura MLS v4.2. Data are compiled from both individual satellite source data as well as a merged data product. For SWOOSH, all records provided in units of number density on altitude grid (*i.e.,* SAGE II and III) are converted to mixing ratio on pressure using MERRA reanalyses, similar to the process used in GOZCARDS v2.20. A key aspect of the merged product is that the source records are homogenised to account for inter-satellite biases and to minimise artificial discontinuities in the record. The SWOOSH homogenisation process involves adjusting the satellite data records to a "reference" satellite using coincident observations during time periods of instrument overlap. The reference satellite is chosen based on the best agreement with independent balloon-based sounding measurements, with the goal of producing a long-term data record that is both homogeneous (*i.e.,* with minimal artificial discontinuities in time) and accurate (*i.e.*, unbiased). For ozone the reference instrument is SAGE II. The SWOOSH v2.6 data are publicly available at **https://data.noaa.gov/dataset/dataset/stratospheric-water-and-ozone-satellite-homogenized-swoosh-data-set**.

2.2.4 Limb profile data records in number density on altitude grid

2.2.4.1 SAGE-OSIRIS-OMPS

The merged SAGE-OSIRIS-OMPS time series has been created at the University of Saskatchewan. The basic construction technique used for this merged time series of deseasonalised anomalies is described in *Bourassa et al.* (2014). For the merged time series the data for each of the three instruments are first treated separately. They are averaged within 1 km altitude and 10° latitude bins and then individually deseasonalised. The resulting zonal mean, deseasonalised anomalies are then merged after biases are removed. The time series spans the period from 1984, when the first SAGE II measurements were made, up to the present where both OSIRIS and the Ozone Mapping and Profiler Suite - Limb Profiler (OMPS-LP) continue to produce high quality data records.

The three observation data sets merged in this data record are: SAGE II v7.0, the recently released OSIRIS v5.10 with improved pointing stability (*Bourassa et al.,* 2018), and the University of Saskatchewan OMPS-LP 2D data set (USask 2D) v1.0.2 (*Zawada et al.,* 2018). This work also further describes the merging process used to create the data record and presents results from preliminary trend analyses based on these data. These preliminary analyses indicate that the addition of the OMPS-LP data to the original SAGE II-OSIRIS merged anomaly data record only slightly changes the magnitude of the derived trends, but the additional data enhances the significance of these results. Since the OSIRIS and OMPS-LP instruments are still operational, this merged data set will be updated regularly as new measurements become available (access from the LOTUS web page).

Two versions of the merged SAGE-OSIRIS-OMPS data set are produced: One uses SAGE II data with corrected sampling effects (*Damadeo et al.,* 2018) further called corr-SAGE and another relies on the standard SAGE II v7 data (*Damadeo et al.,* 2013).

2.2.4.2 SAGE-CCI-OMPS

The merged SAGE-CCI-OMPS data set has been developed in the framework of the European Space Agency (ESA) Climate Change Initiative on Ozone (Ozone_cci). It includes data from several satellite instruments: SAGE II on the Earth Radiation Budget Satellite (ERBS), GOMOS, the SCanning Imaging Absorption spectroMeter for Atmospheric CHartographY (SCIAMACHY) and the Michelson Interferometer for Passive Atmospheric Sounding (MIPAS) on EnviSat, OSIRIS on Odin, ACE-FTS on the SCIence SATellite (SCISAT), and OMPS-LP on the Suomi National Polar-orbiting Partnership (Suomi-NPP). The data set is created specifically with the aim of analysing stratospheric ozone trends. For the merged data set, the latest versions of the original ozone data sets are used. Detailed information about the individual data sets is presented in *Sofieva et al.* (2017). Data sets from the individual sensors have been extensively validated and inter-compared (*e.g.*, *Rahpoe et al.*, 2015*; Hubert et al.,* 2016); only those data sets that are in good agreement and that do not exhibit significant drifts, with respect to collocated ground-based observations and with respect to each other, are used for merging. The intercomparison of data records from individual instruments is presented in (*Sofieva et al.,* 2017) and can also be found in *Section 3.1.3* of the current Report. The long-term data set is created by computation and merging of deseasonalised anomalies, which are estimated using monthly zonal mean profiles from individual instruments (*Sofieva et al.,* 2017).

The merged SAGE II v7, Ozone_cci, and OMPS-LP (USask 2D v1.0.2) data set consists of merged monthly deseasonalised anomalies of ozone in 10° latitude zones from 90°S to 90°N. The data are provided on an altitude grid from 10 km to 50 km, during the period from October 1984 to July 2016. The best quality of the SAGE-CCI-OMPS data set is expected in the stratosphere at latitudes between 60°S and 60°N. Ozone trends in the stratosphere have been evaluated based on the created data sets (*e.g.*, *Sofieva et al.,* 2017; *Steinbrecht et al.,* 2017). The data set is available at the LOTUS website and at the Ozone_cci website (**http://www.esa-ozone-cci.org**).

2.2.4.3 SAGE-MIPAS-OMPS v2

The SAGE-MIPAS-OMPS data set consists of deseasonalised ozone anomalies from the SAGE II v7 (1984–2005), MIPAS IMK/IAA v7 (2002–2012) and OMPS-LP v2.5 (April 2012 – March 2017) data sets which are merged using the ACE-FTS v3.6 data record as a transfer standard. Namely, time series of parent instruments are debiased by minimising the root mean square of uncertainty-weighted differences with time series of ACE-FTS (where overlapping), taking the standard error of the mean as the uncertainty. This procedure removes biases between the different data sets, including those resulting from different altitude resolutions or different prior information, sampling issues, and limited or no overlap between different data sets. The merging in overlapping periods is performed via weighted means, with weights inversely proportional to standard errors of the means of corresponding monthly means from individual data sets. Two periods of MIPAS measurements, 2002–2004 and 2005–2012, are treated as two independent data sets.

The data set is provided along with uncertainty estimates. The data are provided in 10° latitude bins, from 60°S to 60°N for the period from October 1984 to March 2017. The main differences to the SAGE-CCI-OMPS data set are:

- the OMPS data are from the NASA processor, instead of the USask 2D processor
- the MIPAS data from 2002–2004 are included in the record
- the ACE-FTS data are used as the transfer standard.

The first release of this merged data record used version 2 of the NASA OMPS-LP profile retrievals and was used in the assessment by *Steinbrecht et al.* (2017). The SAGE-MIPAS-OMPS data record used for the LOTUS assessment incorporates the newer OMPS-LP NASA v2.5 data described by *Kramarova et al.* (2018).

There exists a version of the SAGE-MIPAS-OMPS data set which uses MLS as a transfer standard, but it is not considered in this Report due to time limitations. The SAGE-MIPAS-OMPS is described in detail in *Laeng et al.* (2019) and the data set is available at **https://www.imk-asf.kit.edu/english/304_2857.php.**

2.2.5 Satellite data in broad latitude bands

In *Sections 5.2* and *5.3*, we discuss profile time series and trends in three broad latitude bands: 60°S–35°S, 20°S–20°N, and 35°N–60°N. For GOZCARDS, SWOOSH, SBUV MOD, and SBUV COH, we first computed the deseasonalised monthly anomalies (in percent) with respect to their own 1998–2008 climatology for each 5° or 10° latitude belt (**Table 2.2**), then averaged over the broader latitude zones with equal weights. The SAGE-CCI-OMPS and SAGE-OSIRIS-OMPS data records were provided as deseasonalised. However, instead of using 1998–2008 as the base period, the entire time period of the record was used for normalisation. In these two cases, we averaged the reported deseasonalised monthly anomalies (in percent) over the three belts, then offset the result to zero mean value in 1998–2008.

2.3 CCMI model data

2.3.1 Description of model data sets

We have used output from the chemistry–climate models (CCMs) and chemistry-transport models (CTMs) participating in phase 1 of the CCMI (*Eyring et al.,* 2013). CCMI is a joint activity of the International Global Atmospheric Chemistry (IGAC) and Stratosphere–troposphere Processes And their Role in Climate (SPARC) projects, with CCMI-1 being the first phase of this initiative and a continuation of previous CCM intercomparisons (CCM Validation Activity; CCMVal) such as CCMVal-1 and CCMVal-2. Model output from both CCMVal intercomparisons have been widely used in previous WMO Ozone Assessments (*WMO,* 2007, 2011, 2014).

Models participating in CCMI-1 are coupled chemistry–climate and chemistry-transport models, which are able to capture the coupling between the stratosphere and troposphere in terms of composition and physical climate processes more consistently than previous model generations. An overview of the models used in the first phase of CCMI-1, together with details particular to each model, and an overview of the available CCMI-1 simulations is given in *Morgenstern et al.* (2017).

For this Report we have used data from the REF-C2 simulation of CCMI-1. Although the most appropriate reference simulation set would have been the REF-C1, which reproduces the past, we opted to use the REF-C2 simulation, as the last year of the REF-C1 was as early as 2010 (or even earlier for some models) and would therefore not cover the entire period when observations are available. Our interest is to provide information about the long-term evolution of ozone changes until the present, and this seamless simulation from 1960–2100 was considered appropriate. REF-C2 is analogous to the REF-B2 experiment of CCMVal-2 but with a number of new and/or improved CCMs. The experiments follow the *WMO* (2011) A1 scenario for ozone depleting substances and the RCP 6.0 for other greenhouse gases, tropospheric ozone precursors, and aerosol and aerosol precursor emissions. Ocean conditions can be either modeled (from a separate climate model simulation, in 9 of the models used here), or internally generated, in the case of ocean-coupled models (7 of the models used). Details can be found in **Table S1** of the Supplement of *Morgenstern et al.* (2017). For the solar forcing, the recommendation was to use the forcing data from 1960–2011 (as in the hindcast REF-C1 simulations) and a sequence of the last four solar cycles (solar cycle numbers 20–23) until the end of the simulations. Finally, the QBO was either internally model-generated or nudged from the data set provided by Freie Universität Berlin. No volcanic forcings were used in this reference simulation. For a detailed description of the full forcings used in the reference simulations see *Eyring et al.* (2013), *Hegglin et al.* (2016), and *Morgenstern et al.* (2017).

In this work, we used a total number of 16 models submitted to the REF-C2 archive (see **Table S2.1** in the Supplemental Material that summarises the models and number of analysed runs). This final selection was based on availability of zonal averaged ozone profile data at the models' native latitude resolution between 60°S–60°N and full vertical coverage from troposphere to stratosphere. Our analysis required (a) zonal wind profiles (zonal means), (b) sea surface temperatures (SSTs, over the tropical Pacific) and (c) ozone as total column and profile (zonal means). We used all pressure levels provided by the models (at standard levels, a total of 31) and all model latitudes, using also the multiple simulations provided by many of the participating models.

2.3.2 Model data in broad latitude bands

As an initial step in this analysis, we have transferred the zonal mean ozone profile data for each model (and every ensemble member) to a common five degree latitude grid, keeping the 31 vertical levels as initially provided. All ozone profiles (at the corresponding pressure level/latitude bin) were then deseasonalised to their climatology, using 1998–2008 as the base period.

In order to create the time series analysed in *Chapter 5,* we first averaged all individual ensemble simulations for each model, so that only one time series for each model/modelling group is included in the average. This is done to avoid unequal weighting caused by the larger number of ensemble members provided by some models/groups. The deseasonalised model time series shown in *Section 5.2* and regressed in *Section 5.3* were computed as the equally weighted average over the appropriate latitude bands: 60°S–35°S, 20°S–20°N, 35°N–60°N, and 60°S–60°N. For each latitude band, the mean, standard deviation and median were calculated. The range of the model results is provided as the 10th (lower) and 90th (upper) percentiles. Moreover, the absolute minimum and maximum values at each time/level/latitude bin were calculated. The time series of model annual averages are presented in *Chapter 5* for comparisons with observations. The model data shown in **Figures 5.4** and **5.5** are smoothed with a 1-2-1-year filter to eliminate any possible shorter term natural variability, as it was included in a number of models but not in all.

2.4 Summary

Chapter 2 provides a description of long-term ozone profile data sets made available for the trend analyses discussed in *Chapter 5*. In order to be considered for trend analyses, the data have to be available from 1985 through 2016 and have no significant gaps (less than a year). In multiple regression analyses (see *Chapter 4* and *5*), longer data sets allow for a more robust fit, particularly for slowly varying proxies (*i.e.,* solar), which might otherwise alias into the trend.

Analyses published in this Report take advantage of four additional years in the long-term ozone records as compared to the results published in the 2014 WMO Ozone Assessment, *Tummon et al.* (2015), and *Harris et al.* (2015).

An exception for the length of the record was provided for several ground-based data sets in order to obtain adequate spatial distribution of trends across a wide range of latitudes. Discussion of the representativeness of individual ground-based records for broad-band trend assessment can be found in *Chapter 5.* Comparisons between satellite and ground-based trends averaged within a broad zonal band are discussed in *Chapter 5.*

The combined satellite records (*Section 2.2*) feature the addition of new satellite records (*i.e.,* two versions of the OMPS ozone data set) and recently reevaluated and stabilised historical data sets from well-established instruments (*i.e.,* removal of the drift in OSIRIS and MIPAS records). These new merged data sets are expected to be more accurate and stable due to the use of revised data records from the individual instruments. The combined data sets' stability also relies on improvements in the merging methods. Improved methods for combining satellite records became available in recent years (GOZCARDS v2.20, SAGE-CCI-OMPS, SWOOSH, and others), thus reducing unexplained features (*i.e.*, discontinuities) in the combined records and their impacts on the derived trends. Moreover, assessment of methods used to combine short satellite records led to improved understanding of the sources for propagation of errors in the combined trends and impact on the trend uncertainties (*i.e.*, see discussion about differences in the two SBUV combined records in *Section 3.1.4*). The LOTUS Report evaluates these new data sets for improved accuracy and stability.

Well-maintained long-term ozone records are also important for validation of the CCMI retrospective model runs (*Section 2.3*). Agreement between models and observational records (further discussed in *Chapter 5*) assure complete understanding of the processes that impact past ozone changes, such that we have trust in the scenarios for future ozone changes and attribution to GHG and ODS variability, though such a study of the future is not included in this Report.

Although the new and improved records are an integral part of understanding stratospheric ozone changes, there are some remaining issues that are not fully resolved in this Report. For example, comparisons of the coincident and collocated satellite and ground-based records suggest remaining intermittent drifts in the combined data sets (see *Chapter 3*). Drifts (or discontinuities) can also be found in ground-based records, which has inspired the homogenisation effort for the ozonesonde records (*Smit et al,* 2012a, 2012b). Unfortunately, only a handful of homogenised ozonesonde records were ready in time for the analyses done within the LOTUS activity. The trend analyses of the broad-band ozonesonde records will need to be repeated after all homogenised records are ready to update the broad-band averages. Other ground-based records, especially those available from the same location, need to be reevaluated to understand the causes for discrepancies and how the changes in the observational sampling or processing of the ozone measurements can potentially impact the derived trends.

Assessment of sampling biases for all satellite combined records used in this Report is not available (see discussion in *Chapter 3*). It is important to understand how sampling biases can affect the deseasonalised anomaly records. In addition, assessment of uncertainties in the combined ground-based records is needed for analyses of propagation of measurements errors in the trends analyses.

Additionally, errors in ozone satellite and ground-based combined records can be caused by their conversion to a different coordinate system and impact the resulting trends (*McLinden and Fioletov,* 2011). Error propagation is needed to evaluate the impact of the non-homogenised temperature time series (mostly prior to 2000; *Long et al.,* 2017) on the accuracy and spatial distribution of ozone records converted to new coordinates (*Douglass et. al,* 2017). Depending on the assimilation, this conversion can potentially introduce intermittent drifts and thus degrade the stability of the converted ozone record. Assessment of the impact of the conversion on trends should be addressed in the future.

Availability of satellite overpass data over all ground-based stations is needed to understand the spatial and temporal sampling limitations of the ground-based data sets. Comparisons between the overpass and broad-band derived trends is needed to understand representativeness of the ground-based and sonde ozone trends over the broad regions. Representativeness of all ground-based records for zonal averaged trends was not not fully assessed in this Report, although a limited case study of lidar records is discussed in *Chapter 3* (see *Section 3.2.2*).