

# Chapter 3: Climatology framework

This chapter discusses the datasets evaluated within the SPARC Data Initiative, including information on how the climatologies were constructed, and on the diagnostics used to evaluate them. Note that here we use the term ‘climatology’ for monthly mean zonal mean cross sections. The evaluations are based on single year cross sections, or on multi-year means compiled over particular reference periods. The resulting climatologies may be single-year or multi-year monthly or annual means.

Monthly zonal mean time series have been calculated for each trace gas species and aerosol listed in **Table ES.1** (*Executive Summary*) on the SPARC Data Initiative climatology grid, using 5° latitude bins (with mid-points at -87.5°, -82.5°, -77.5°, ..., 87.5°) and 28 pressure levels (300, 250, 200, 170, 150, 130, 115, 100, 90, 80, 70, 50, 30, 20, 15, 10, 7, 5, 3, 2, 1.5, 1, 0.7, 0.5, 0.3, 0.2, 0.15, and 0.1 hPa). Trace gas species are reported as volume mixing ratios (VMR), and aerosols as extinction coefficients. The monthly zonal mean value and the 1 $\sigma$  standard deviation, along with the number of averaged data values are given for each month, latitude bin and pressure level. The mean, minimum, and maximum local solar time (LST), average day of the month, and average latitude of the data within each bin for one selected pressure level are also provided.

For species with large diurnal variations we separate the measurements based on LST (see detailed discussion in *Section 3.1.1*). Additional climatologies are built using a photochemical box model to scale the measurements to a common LST in order to enable direct comparison between products from different instruments with different sampling patterns. All satellite-based measurements of trace gas species are imperfect estimates of the truth characterised by measurement errors. The compilation of climatologies from these measurements can introduce additional errors such as sampling biases produced by non-uniform spatial or temporal sampling, or by the use of different filtering techniques. Biases can also be introduced by applying different averaging techniques.

The climatology construction, including common methodology and information specific to each instrument, is described in *Section 3.1*. A discussion of climatology uncertainties is provided in *Section 3.2*, while the diagnostics used to evaluate the trace gas climatologies are explained in *Section 3.3*.

## 3.1 Climatology construction

### 3.1.1 Methodology

The original data products are first interpolated to the SPARC Data Initiative pressure grid using a hybrid log-linear interpolation. For instruments providing data on an altitude grid, a conversion from altitude to pressure levels is performed using retrieved temperature/pressure profiles or meteorological analyses (ECMWF, GEOS-5, or NCEP, see **Table 3.1** for detailed information). The same pressure and temperature profiles are used to convert data products retrieved as number densities to VMR.

Original data have been carefully screened according to recommendations given in relevant quality documents, in the published literature, or according to the best knowledge of the involved instrument scientists. Monthly zonal mean products are calculated as the average of all of the measurements on a given pressure level within each latitude bin and month. An exception is MIPAS, for which measurements are interpolated to the centre of the latitude bin after averaging (see *Section 3.1.3.9* for details). For some species and instruments, averaging was done in  $\log_{10}$ (VMR) space. The 1 $\sigma$  standard deviation along with the number of averaged data values are also given for each month, latitude bin and pressure level. If not otherwise mentioned, a minimum of five measurements within the bin is required to calculate a monthly zonal mean for each instrument. The mean, minimum, and maximum LST, average day of the month, and average latitude of the data within each bin are provided for one selected pressure level for each latitude bin and month. Instrument-specific information for the calculation of the monthly zonal mean values is given in *Section 3.1.3*.

For species with large diurnal variations the monthly zonal mean climatologies cannot be compared directly since the LST of the measurements can differ from instrument to instrument, and between seasons and latitudes for the same instrument. Two types of climatologies are produced for diurnally varying species; climatologies from observations binned by LST (unscaled), and climatologies from observations scaled to a common LST. Most of the instruments measure two distinct LSTs per latitude. These instruments are in polar sun-synchronous orbits, with one LST for the ascending portion of the orbit and one for the descending portion, or in the case of sun-synchronous solar occultation

Table 3.1: Instrument specifications relevant for the climatology construction.

Instrument	Latitudinal coverage	LT at equator <sup>1</sup>	LT of measurement <sup>2</sup>	Inc. <sup>3</sup>	Vert. Grid <sup>4</sup>	Alternate grid <sup>5</sup>	Meas. <sup>6</sup>	Conversion to VMR <sup>7</sup>	Data density per day
<b>LIMS</b> on Nimbus 7	64°S–84°N (daily)	a: 11:51am d: 11:51pm	a: 1pm d: 11pm	99.3°	p	N/A	VMR	N/A	3000
<b>SAGE I</b> on AEM-B	75°S–75°N (~one month)	N/A	N/A	56°	z	NCEP	ND	NCEP	30
<b>SAGE II</b> on ERBS	75°S–75°N (~one month)	N/A	N/A	57°	z	NCEP	ND	NCEP	30
<b>SAGE III</b> on Meteor-3M	60°S–30°S 40°N–80°N (~over one season)	a: 9:30am d: 9:30pm	N/A	99.6°	z	NCEP	ND	NCEP	30
<b>HALOE</b> on UARS	75°S–75°N (~over one season)	N/A	N/A	57°	p	N/A	VMR	N/A	30
<b>UARS-MLS</b> on UARS	80°S–80°N (~over two months)	N/A	N/A	57°	p	N/A	VMR	N/A	1318
<b>POAM II</b> on SPOT-3	88°S–63°S 55°N–71°N (over one year)	a: 10:30pm d: 10:30am	N/A	98.7°	z	UKMO analysis	ND	UKMO analysis	30
<b>POAM III</b> on SPOT-4	88°S–63°S 55°N–71°N (over one year)	a: 10:30pm d: 10:30am	N/A	98.7°	z	UKMO analysis	ND	UKMO analysis	30
<b>OSIRIS</b> on Odin	82°S–82°N (daily, no winter hemisphere)	a: 6:30pm d: 6:30am	a: 6:30pm d: 6:30am	97.8°	z	ECMWF operation- al analysis	ND	ECMWF operation- al analysis	300–975
<b>SMR</b> on Odin	83°S–83°N (daily)	a: 6:30pm d: 6:30am	a: 6:30pm d: 6:30am	97.8°	p	N/A <sup>8</sup>	VMR	N/A	600–975
<b>GOMOS</b> on Envisat	90°S–90°N (daily, no summer poles for night)	a: 10:00pm d: 10:00am	a: 10–12pm d: 8–10:30am	98.55°	z	ECMWF operation- al analysis	ND	ECMWF operation- al analysis	100–300 (night mea- surements)
<b>MIPAS</b> on Envisat	90°S–90°N (daily)	a: 10:00pm d: 10:00am	a: 10:00pm d: 10:00am	98.55°	z	MIPAS	VMR	N/A	1000 (1300 since 2005)
<b>SCIAMACHY</b> on Envisat	85°S–85°N (65° for winter hemisphere) <sup>9</sup>	a: 10:00pm d: 10:00am	d: 10:00am	98.55°	z	ECMWF operation- al analysis	ND	ECMWF operation- al analysis	364–1456
<b>ACE-FTS</b> on SCISAT-1	85°S–85°N (~over one season)	N/A	N/A	74°	z	ACE-FTS	VMR	ACE-FTS	30
<b>ACE-MAESTRO</b> on SCISAT-1	85°S–85°N (~over one season)	N/A	N/A	74°	z	ACE-FTS	ND	ACE-FTS	30
<b>HIRDLS</b> on Aura	65°S–82°N (daily)	a: 1:43pm d: 1:43am	a: 2:57pm d: 0:30am	98.21°	p	N/A	VMR	N/A	5600
<b>MLS</b> on Aura	82°S–82°N (daily)	a: 1:43pm d: 1:43am	a: 1:25am d: 1:25pm	98.21°	p	N/A	VMR	N/A	3500
<b>TES</b> on Aura	82°S–82°N (daily) (50°S–70°N for 2008/09; 30°S– 50°N for 2010)	a: 1:43pm d: 1:43am	a: 1:43pm d: 1:43am	98.21°	p	N/A	ln(VMR)	N/A	3145 (2126 for 2008/09; 1890 for 2010)
<b>SMILES</b> on ISS	38°S–65°N (daily)	N/A	N/A	51.6°	p	N/A <sup>8</sup>	VMR	N/A	1620

<sup>1</sup> Local time of equator crossing for satellites with sun-synchronous orbit (a=ascending, d=descending)<sup>2</sup> Local time of measurement made at equator crossing for satellites with sun-synchronous orbit (a=ascending, d= descending)<sup>3</sup> Inclination of the orbital plane<sup>4</sup> Vertical grid used for retrieval of species (altitude 'z' or pressure 'p')<sup>5</sup> Data used for conversion to alternate vertical grid<sup>6</sup> Measure of species: volume mixing ratio (VMR) or number density (ND)<sup>7</sup> Pressure/temperature data used for conversion from number density to volume mixing ratio<sup>8</sup> For SMR and SMILES the tangent-pressure is retrieved but Level 2 data are provided on altitude grids. Conversion between p and z is done using ECMWF (for SMR) or GEOS-5 (for SMILES) data.<sup>9</sup> 55° for winter hemisphere for water vapour climatologies

sounders, with measurements at sunrise and sunset as seen from the satellite. For the latter, the LSTs shift with the day of year. Climatologies of diurnally varying trace gases from instruments in a sun-synchronous orbit are generally based on measurements separated into ante meridiem (am) and post meridiem (pm) data. A representative LST can be assigned to each month and latitude bin. However, in some cases the LST variations between season and latitude bin must be considered. Instruments that observe from non-sun-synchronous orbits are characterised by drifting observation times with respect to LST. Climatologies for these instruments are generally separated into daytime and night-time measurements. Climatologies of diurnally varying trace gases from non-sun-synchronous solar occultation measurements are based on data separated into local sunrise and sunset measurements. Additional climatologies are compiled using a photochemical box model to scale the measurements to a common LST, as explained in more detail in Section 3.1.2. For chemical families ( $\text{NO}_x$ , Section 4.1.12, and  $\text{NO}_y$ , Section 4.1.17) the total family abundance is derived using all members of the family available from the instrument, supplemented with species derived from a photochemical box model if needed.

### 3.1.2 Local time scaling

For species with large diurnal variations additional climatologies are compiled by scaling the measurements with a photochemical box model to a common LST. The scaled climatologies enable a direct comparison between products from different instruments with different sampling patterns. For the diurnally varying species  $\text{NO}$ ,  $\text{NO}_2$ ,  $\text{NO}_x$  and  $\text{BrO}$  scaled climatologies are calculated for 10am and 10pm, the approximate local time of the MIPAS measurements at the equator. The  $\text{ClO}$  climatologies are scaled to 1:30am and 1:30pm, which is the approximate local time of the Aura-MLS measurements (for  $\sim 60^\circ\text{S}$ - $60^\circ\text{N}$ ).

A derivative of the University of California, Irvine photochemical box model [Prather, 1992; McLinden *et al.*, 2000; McLinden *et al.*, 2010] was applied to calculate the diurnal scaling factors used to map the VMR of a diurnally varying species from one local time ( $\text{LST}_1$ ) to another ( $\text{LST}_2$ ). This was done by scaling the measured  $\text{VMR}(\text{LST}_1)$  by the model-calculated ratio  $\text{VMR}(\text{LST}_2)/\text{VMR}(\text{LST}_1)$ , which will be referred to as scaling factor in the following text. The VMR at the new local time is then derived as:

$$\text{VMR}(\text{LST}_2) = \text{VMR}(\text{LST}_1) [\text{VMR}(\text{LST}_2)/\text{VMR}(\text{LST}_1)]_{\text{model}}$$

The scaling factors are calculated with the photochemical box model based on LST, temperature, surface albedo and concentration of various trace gases ( $\text{O}_3$ ,  $\text{N}_2\text{O}$ ,  $\text{NO}_y$ ,  $\text{CH}_4$ ,  $\text{Cl}_y$ ,  $\text{Br}_y$ ). With these parameters specified, all remaining species are calculated to be in a 24-hour steady state by integrating the model for 30 days (fixed to the prescribed Julian day and latitude). The kinetic reaction rate coefficients and photochemical data used by the box model are based on JPL-06 and JPL-09 recommendations.

The model-calculated scaling factors were provided as a function of altitude, latitude, day of year, and LST as lookup tables. The calculations were based on the photochemical box model initialised with climatological inputs. Each table consists of 25 pressure-altitudes, from 10 to 58 km in 2 km increments, with pressure-altitude  $z^* = -16 \log_{10}(p/1000)$ ,  $p$  given in hPa, and  $z^*$  given in km. The latitude grid ranges from  $77.5^\circ\text{S}$  to  $77.5^\circ\text{N}$  in  $2.5^\circ$  increments. Tables are given for the 1<sup>st</sup>, 11<sup>th</sup>, and 21<sup>st</sup> of each month for 34 local times spanning 24-hours (fewer for polar regions). The input data includes  $\text{O}_3$  and temperature from measurement-based climatologies and  $\text{N}_2\text{O}$ ,  $\text{NO}_y$ , and  $\text{CH}_4$  from three-dimensional model output. The  $\text{Cl}_y$  and  $\text{Br}_y$  families are prescribed using trace gas correlations. Surface albedo, which impacts the photodissociation rates, was set to 0.2.

OSIRIS uses a separate run of the photochemical model for each scan, initialised with OSIRIS-measured  $\text{O}_3$  abundances and ECMWF temperatures. However, this process is computationally expensive. Thus, for most instruments, the scaling is done profile-by-profile with the pre-calculated lookup tables mentioned above.

The box model can likewise be used to supply information about an unmeasured species provided it is closely coupled to one that is measured. For example, the OSIRIS  $\text{NO}_y$  climatology was obtained from the box model using OSIRIS  $\text{NO}_2$  and SMR  $\text{HNO}_3$  measurements [Brohede *et al.*, 2008].

The box model was evaluated using measurements from the JPL Mk-IV FTIR interferometer [Toon, 1991] from 10 balloon flights between 1997 and 2005. A comparison of the partitioning of stratospheric  $\text{NO}_y$  is presented in Brohede *et al.* [2008] in which good overall agreement is found except for instances near the polar day-night boundary where air mass history becomes a dominant factor. Such studies indicate that when constrained by measurements of temperature, ozone and long-lived species, the box model is able to accurately simulate the radical species. This point, combined with the fact that the diurnal scaling approach has been used successfully in numerous validation studies of diurnally-varying species [e.g., Kerzenmacher *et al.*, 2008], suggests that on average the error in the scaling factors is small. For any given profile, there may be significant errors if the assumed inputs to the model also have significant errors. However, this represents a random source of error, which is effectively minimised when averaging over a large number of profiles, as it is done in the compilation of the SPARC Data Initiative climatologies. While a rigorous error assessment has not been performed, the systematic error of these scaling factors is estimated to be less than 20% based on the above discussion.

For the scaled HIRDLS climatologies, the Specified Dynamics Whole Atmosphere Community Climate Model (SD-WACCM) is used to calculate the local time scaling factors to 10am and 10pm as a function of altitude, latitude, day of year, and LST. SD-WACCM is a global chemistry-climate model based on the Community Atmospheric Model (CAM) [Collins *et al.*, 2004] with temperature and

wind specified by the Goddard Earth Observing System (GEOS-5) reanalyses. The gravity wave drag and vertical diffusion parameterisations are described in *Garcia et al.* [2007] and the neutral chemistry modules in *Kinnison et al.* [2007].

### 3.1.3 Instrument-specific information

In the following, information relevant for the construction of the SPARC Data Initiative climatologies is described for each instrument. **Table 3.1** summarises specifications for all instruments.

#### 3.1.3.1 LIMS climatologies

The LIMS Level 3 V6 combined node (ascending and descending) daily zonal mean Fourier coefficients for O<sub>3</sub>, H<sub>2</sub>O, HNO<sub>3</sub> and ascending and descending node daily zonal mean Fourier coefficients for NO<sub>2</sub> were used to obtain the monthly zonal mean data. Note that the ascending and descending measurements were taken at approximately 1pm and 11pm local time, respectively, for the low and mid-latitudes. The LIMS Level 3 product was the more appropriate data to use for the SPARC climatology because it has no missing data, while the LIMS Level 2 product is missing data for certain orbits or even complete days. LIMS species are given in VMR, and the profiles are first interpolated to the latitudes and then to the pressure levels used within the SPARC Data Initiative. These data were then averaged per month. The LIMS V6 data retrievals near tropopause levels may contain residual effects from cloud radiances, especially at low latitudes. The LST\_MEAN, LST\_MIN, LST\_MAX, AVE\_DOM, and AVE\_LAT values provided by all other instrument climatologies are missing in the data files. Level 3 data and documentation (Level-3 README) reside at the GES DISC archive that is located at [http://disc.sci.gsfc.nasa.gov/acdisc/documentation/LIMS\\_dataset.gd.shtml](http://disc.sci.gsfc.nasa.gov/acdisc/documentation/LIMS_dataset.gd.shtml).

#### 3.1.3.2 SAGE I/II/III climatologies

The SAGE climatologies are based on retrieved Level 2 products from SAGE I (V5.9), SAGE II (V6.2) and SAGE III (V4.0). It is known that there are altitude errors in the original SAGE I (V5.9) data due to less reliable ephemeris information. An empirical altitude correction based on *Wang et al.* [1996] has therefore been applied to these data before their use in this study. All natively retrieved species from SAGE instruments are given in number density in altitude co-ordinates. In order to generate the SPARC Data Initiative climatologies, all number density profiles were first converted to VMR using NCEP temperature and pressure profiles, which are reported along with each individual number density profile in the SAGE Level 2 data files. A linear interpolation in  $\log_{10}(p)$  was then used to derive VMRs on the SPARC Data Initiative pressure levels. Additional data screenings, as described in the following, were also applied before generating the final climatologies.

Only a few studies describing how to screen SAGE I data for anomalous values exist. The main uncertainty in retrieved O<sub>3</sub>, H<sub>2</sub>O and NO<sub>2</sub> is the interference of aerosol and clouds especially in the lower stratosphere below ~15 to 20 km. For SAGE I data, all O<sub>3</sub> measurements with corresponding aerosol extinctions at  $1.0 \mu\text{m} \geq 1.0 \times 10^{-3} \text{ km}^{-1}$ , are flagged (L. W. Thomason, personal communication). For SAGE II and SAGE III O<sub>3</sub> and NO<sub>2</sub> measurements, screenings follow the approach by *Wang et al.* [2002], which removes anomalously low values and those affected by “short events” or aerosols/clouds. Due to an instrument problem, the SAGE II NO<sub>2</sub> data from satellite sunrise measurements are not included in this study. The SAGE-retrieved H<sub>2</sub>O is more sensitive to interferences from aerosol compared to O<sub>3</sub> and NO<sub>2</sub>. More stringent criteria based on *Thomason et al.* [2004] and *Taha et al.* [2004] are therefore used to screen the H<sub>2</sub>O data.

#### 3.1.3.3 HALOE climatologies

The HALOE V19 measurements starting in October 1991 and extending through November 2005 are used to create climatologies for O<sub>3</sub>, HCl, HF, H<sub>2</sub>O, CH<sub>4</sub>, NO, NO<sub>2</sub>, NO<sub>x</sub> (NO+NO<sub>2</sub>), and aerosol extinction. Each individual profile is first screened for clouds and heavy aerosols. The O<sub>3</sub>, NO<sub>2</sub>, and NO profile data are further screened for anomalous values caused by an aerosol minimum. Each individual profile is then interpolated to the SPARC Data Initiative pressure levels. These screened and interpolated data are then averaged within each SPARC Data Initiative latitude bin to produce monthly zonal means and standard deviations of the trace gases and the aerosol extinction coefficients. The diurnally varying species NO<sub>2</sub>, NO and NO<sub>x</sub> are separated into local am and local pm climatological fields. The NO<sub>x</sub> climatology is produced by first combining the screened and interpolated profiles of collocated NO and NO<sub>2</sub> measurements, and then zonally averaging them on the SPARC Data Initiative pressure-latitude grid. The aerosol extinction profiles were only screened for clouds before further processing.

#### 3.1.3.4 UARS-MLS climatologies

UARS-MLS climatologies are based on Level 3AT data (similar to Aura-MLS Level 2 along-track profiles), using V5 for O<sub>3</sub>, V6 for HNO<sub>3</sub>, and V6 for H<sub>2</sub>O. The main reference for the latest UARS data is *Livesey et al.* [2003]. The V6 HNO<sub>3</sub> files were a correction to the V5 dataset, to more properly account for emission from some of the HNO<sub>3</sub> excited vibrational states. The V6 H<sub>2</sub>O dataset (originally named V0104) is described in *Pumphrey* [1999]. These source datasets are available from the GES DISC, and the H<sub>2</sub>O dataset can also be accessed *via* the British Atmospheric Data Centre (BADC), see <http://badc.nerc.ac.uk/home/>. The above references and the UARS-MLS validation papers and data quality documentation (see individual species sections of this report) provide information about the recommended data screening for each species. The screening

methods were applied to each profile prior to the averaging and interpolation processes that were used to generate the climatological time series. This generally means that only profiles with good status values (meaning “G”, “T”, or “t” for the “MMAF\_STAT” parameter) were considered. Other screening methods are described in *Livesey et al.* [2003]; in particular, associated UARS-MLS Level 3 Parameter files contain “QUALITY” parameters that should be (and were) considered for data screening. Also, when mixing ratios are flagged negative, this indicates that the *a priori* information is playing a non-negligible role in the retrieval process, so these values are not used in this report. Vertical profiles are retrieved as VMRs *versus* a fixed pressure grid (with spacing corresponding to 6 levels per decade change in pressure). The pressure ranges used here reflect the recommended levels for UARS-MLS profiles, although some additional information often exists beyond these ranges (mostly for higher altitude regions). If average monthly values are negative, they are not used for the SPARC climatological dataset, although small negative values may be within the calculated error. Note that UARS-MLS data after 14 June 1997 are considered slightly less reliable than for the earlier dates due to a change in UARS-MLS operations after that date (in order to conserve satellite power), whereby temperature information from the MLS retrievals was lost, and meteorological temperature fields were used instead. Therefore, some small discontinuities are to be expected at this date. Furthermore, the data become increasingly sparse after 1997. Nevertheless, this report includes UARS-MLS data after mid-June, 1997, as trend analysis is not the main focus of this report.

### 3.1.3.5 POAM II/III climatologies

The POAM climatologies were constructed using Level 2 data V6.0 (POAM II) and V4 (POAM III). POAM retrieves gas number density and aerosol extinction on a uniform altitude grid (0-60 km in 1-km increments). The conversion from density to VMR for the gases is done slightly differently for the two instruments. For POAM II, the UKMO total density profile, interpolated spatially and temporally to the POAM measurement, is used for the conversion. POAM III uses a total density profile retrieved directly from the measured Rayleigh scattering above 30 km, and tightly constrained to UKMO below this altitude. Each mixing ratio/aerosol extinction profile is interpolated from the POAM altitude grid to the SPARC Data Initiative pressure grid using the co-located UKMO pressure profile. The data are then binned by month and latitude bin by calculating the median value (VMR or aerosol extinction) at each standard pressure level. A minimum of 15 valid data points are required for each month and latitude bin. Data are only used within the recommended altitude range for each species, as described in the POAM documentation. The data are screened in the binning process according to the data quality flags provided with the POAM Level 2 data (described in detail in the POAM algorithm and error analysis papers, and in documentation provided with the POAM data archives). Any suspect data were eliminated before generating the climatologies. The quality flags screen data for a number of

potential error sources. For gas species, the primary source of error is due to high aerosol loading in the presence of polar stratospheric clouds (PSCs), which can cause feedback noise in the gas retrievals. This is not an issue for O<sub>3</sub> but can be a significant source of error for NO<sub>2</sub> and H<sub>2</sub>O. Both gas and aerosol retrievals can also be flagged due to the presence of sunspots in the POAM field of view. Again, these errors are species-dependent and more significant for NO<sub>2</sub>, H<sub>2</sub>O and aerosols. Finally, optically thick PSCs can cause the POAM scan to terminate at unusually high altitudes, resulting in higher than average retrieval noise in NO<sub>2</sub> and H<sub>2</sub>O at the lowest 2-3 km of the scan. Since POAM measures at the terminator, the climatology of NO<sub>2</sub>, which has a strong diurnal variation, was generated separately for local sunrise and sunset conditions.

### 3.1.3.6 OSIRIS climatologies

Climatologies from OSIRIS are based on the following Level 2 versions: BrO V5; O<sub>3</sub> V5.07; stratospheric aerosol V5.07; and NO<sub>2</sub> V3. The derived products (NO<sub>x</sub> and NO<sub>y</sub>) are based on the NO<sub>2</sub> V3 dataset but have no specific dataset number. Note that in the case of NO<sub>y</sub>, SMR HNO<sub>3</sub> V2.0 data are also included (see *Section 2.2.7*). All quantities except aerosol are retrieved as number density on a fixed altitude grid and converted to VMR on pressure levels using temperature and pressure profiles from ECMWF operational analysis. The aerosol product is retrieved as extinction per km on a fixed altitude grid. OSIRIS can only provide daytime observations, (only profiles with solar zenith angles smaller than 92° are processed). In the Level 2 files, profiles with large pointing offsets, non-converging profiles and altitudes with clouds in the field-of-view have been filtered out. Note that due to low signal-to-noise ratios for BrO, only zonally averaged spectra (10° latitude bins) are used in the retrievals. The number of BrO profiles in each climatology bin will therefore be significantly less than for the other species and a true 5° latitude binning cannot be performed.

In the case of species retrieved using optimal estimation, *i.e.*, BrO and NO<sub>2</sub>, only levels with a measurement response above 0.67 are included in the climatologies. Note that the measurement response cut-off is not applied to individual profiles but to the average values within each climatology bin. This is done in order to reduce a bias to the *a priori* profiles in the climatological averaging. Due to NO<sub>2</sub> log(VMR) retrievals, the climatology averaging for NO<sub>2</sub> (and the NO<sub>2</sub> derivative NO<sub>x</sub>) is performed using the logarithm of the number densities. Other species are averaged in linear space.

The diurnal scaling of BrO uses lookup tables calculated from a photochemical box model initialised with climatological inputs (see *Section 3.1.2*). NO<sub>2</sub> scaling factors are obtained in a more sophisticated way from the (same) photochemical model initialised with measured OSIRIS O<sub>3</sub> abundances and temperature/pressure (from ECMWF) for each individual profile. Because of this scan-based ap-

proach, NO<sub>2</sub> (and NO<sub>x</sub>) data can be scaled to any local time without large uncertainties. For BrO, however, only am data is used to scale to am local times and pm data to pm local times. The NO<sub>x</sub> diurnal scaling factors are calculated simultaneous to the NO<sub>2</sub>/NO ratios, used to calculate NO<sub>x</sub> from NO<sub>2</sub>.

### 3.1.3.7 SMR climatologies

SMR climatologies are based on Level 2 V2.1. The sole exception is HNO<sub>3</sub>, which is based on Level 2 V2.0. In general, only 'good' quality profiles (Level 2 Quality flag = 0) have been used. Vertical profiles were retrieved as VMR or as log<sub>10</sub>(VMR) for CO, NO, and H<sub>2</sub>O from the 544.6 GHz band on an altitude grid given by the refraction-corrected tangent altitudes. Conversion to pressure was done using ECMWF profiles. Retrieved VMRs with a measurement response smaller than 0.75 were rejected (0.8 for N<sub>2</sub>O). Unphysical outliers were also filtered. The pressure range for some species was restricted: N<sub>2</sub>O: p ≥ 170 hPa; HNO<sub>3</sub>: p ≤ 1 hPa; H<sub>2</sub>O (544.6 GHz band): 150 hPa ≥ p ≥ 25 hPa. The minimum number of data values required per latitude bin and pressure level was set to a threshold of five; for H<sub>2</sub>O (both from 488.9GHz and 544.6 GHz band) and NO at least ten values were demanded. For H<sub>2</sub>O in the 544.6 GHz band, the median value was calculated instead of the mean in order to reduce the effect of unphysical outliers present in this dataset. SMR provides several Level 2 ozone data products. Ozone climatologies evaluated in this report are derived from the main stratospheric mode observations at 501.8 GHz. Climatologies have also been compiled for a second ozone product (measured in a band centred at 488.9 GHz) which has very similar characteristics compared to the 501.8 GHz SMR ozone product and is not shown in the following evaluations.

### 3.1.3.8 GOMOS climatologies

The GOMOS data used for the SPARC Data Initiative were produced by the ESA operational processor V5. GOMOS constituent data are number densities given at geographical altitudes. Data files also include ECMWF pressure and temperature data up to 1 hPa at GOMOS measurement locations. These data are used for ray tracing and estimating refractive effects. Above 1 hPa, the MSIS90 climatology is used in place of ECMWF. For the construction of the SPARC climatologies, VMRs and the altitude-to-pressure grid conversion are derived using these external data.

Here, we use GOMOS dark limb measurements only, requiring solar zenith angles greater than 107°. The solar zenith angle limit and the ability of GOMOS to follow and measure stars outside the orbital plane of Envisat leads to a variation in the LST of the measurements. This is important for measurements of diurnally varying constituents NO<sub>2</sub>, NO<sub>3</sub> and O<sub>3</sub> in the mesosphere/lower thermosphere. Envisat equator-crossing times were 10am and 10pm local time.

GOMOS tangent-point local times covered about 1.5 h near the equator and 3 h at mid-latitudes.

GOMOS occultations that used stars with magnitudes weaker than 1.9 and temperatures less than 7000 K often failed to capture the whole ozone profile from 15-100 km beginning in 2003 [Kyrölä *et al.*, 2006; 2010]. After 2003, GOMOS signal-to-noise ratios decreased due to aging of the instrument. In order to guarantee ozone data quality and consistency over the whole time period we have applied the following specific filters on ozone profiles:

- i. Estimated errors must be smaller than 50%;
- ii. VMRs must be positive in the 25-45 km range;
- iii. VMRs must be smaller than 15 ppm in the 20-45 km range;
- iv. Occultations with cool stars (cooler than 6000 K) are rejected below 45 km; the same restriction applies to star numbers 170 and 178.

For NO<sub>2</sub>, NO<sub>3</sub>, and aerosols all stars were used regardless of their magnitude and temperature. In all datasets, we rejected occultations with the obliquity angle (the angle between the occultation plane and the orbital plane of Envisat) larger than 80°. To determine the monthly zonal mean climatologies, we have used the median as a statistical average since it is more robust against outliers than the mean. The uncertainty of the median value is estimated according to Equation 1 in Kyrölä *et al.* [2010].

### 3.1.3.9 MIPAS climatologies

MIPAS trace gas profiles included in the SPARC Data Initiative climatologies were retrieved on a fixed (*i.e.*, tangent altitude independent) altitude grid. Conversion to the pressure grid relies on hydrostatics and MIPAS temperature profiles. Averaging is always performed linearly in VMR, even for species retrieved in log<sub>10</sub>(VMR) (*cf.* Funke and von Clarmann [2011] for discussion of this specific issue). For the climatologies the unweighted mean of all measurements within a month and latitude bin is used. Note that weighting the mean by the inverse squared retrieval error would bias the mean towards warmer parts of the atmosphere. The sampling pattern, particularly from 2002-2004, is such that the measurements are not representative of the full latitude range within the latitude bins. The average values within each bin are interpolated to the centre latitude of the bin, as are the standard deviations and the number of measurements (see von Clarmann *et al.* [2012] for further details). Measurements affected by clouds were discarded from the analysis, and results where the diagonal element of the averaging kernel was below a given threshold were excluded, as well as results from non-converged retrievals. Level 2 data versions distinguish between the full spectral resolution measurements (2002-2004) and reduced resolution measurements (after 2004). Species dependent version numbers are listed in **Table 3.2**. 'FR' stands for full spectral resolution, the measurement mode MIPAS operated in from 2002 to 2004, while 'RR' stands for reduced spectral resolution as applied since 2005. Data version specifiers

are composed of a prefix indicating the version of the ESA Level 1 calibrated spectra used, a gas specifier, and a suffix indicating the version of the retrieval setup. In this report the climatologies for 2002–2004, when MIPAS operated in full spectral resolution, are referred to as MIPAS(1) while climatologies for 2005–2010, when MIPAS operated in reduced spectral resolution, are referred to as MIPAS(2). Note that the version numbers in the climatology file names and in the tables in *Chapter 4* are simplified, and only consist of the retrieval version.

### 3.1.3.10 SCIAMACHY climatologies

Each data product in the scientific retrieval dataset has its own version number, which is not related to the version number of the other species. The SCIAMACHY climatology is compiled using the following versions of the Level-2 products: V2.5 for O<sub>3</sub>, V3.1 for NO<sub>2</sub>, V3.2 for BrO, V3.1 for H<sub>2</sub>O, and V1.0 for aerosol extinction coefficients.

Trace gas profiles and aerosol extinction coefficients are retrieved on an equidistant altitude grid. The retrieval is done in number density for all gases with except water vapour, which is retrieved in logarithm of the number density. The results are then converted to VMR and interpolated to the SPARC Data Initiative pressure grid using pressure and temperature information from the ECMWF operational analysis model with a spatial resolution of 1.5° x 1.5° and a temporal resolution of 6 h. The mean value of VMR for each species in each month and latitude bin is calculated through linear averaging. Aerosol extinction coefficients, retrieved in km<sup>-1</sup> are interpolated to the pressure grid and then averaged.

Because of the signal to noise ratio and radiative transfer modelling issues, only limb measurements at solar zenith angles smaller than 89° (or 85° for water vapour) are processed. Generally, these measurements are made on the dayside of the orbit (descending node, 10am equator crossing time). At high latitudes during the summer, there are also some observations on the night-side of the orbit (ascending node, 10pm equator crossing time) made at solar zenith angles smaller than 89°. However, results from these measurements are not included in the current climatology because of their substantially different local times. Furthermore, all data obtained when Envisat crosses the South Atlantic anomaly (see also *Section 2.2.10*) are excluded from the climatology. The rejected area is located between 20°S to 70°S and 0° to 90°W. For observations with clouds in the instrument field-of-view, the retrieved absorber amounts below the cloud top altitude are skipped.

### 3.1.3.11 ACE-FTS climatologies

The ACE-FTS climatology uses the Level 2 V2.2 dataset (including updates for O<sub>3</sub> and N<sub>2</sub>O<sub>5</sub>). The ACE-FTS VMR profiles are provided on an altitude grid with the pressures retrieved from the spectral measurements (as described in *Section 2.2.11*). The retrieved pressure information is used

**Table 3.2: MIPAS-IMK/IAA Level 2 data versions of different trace gases used in this report.**

	FR (2002-2004)	RR (2005-2010)
H <sub>2</sub> O	V3o_H2O_13	V4o_H2O_220
O <sub>3</sub>	V3o_O3_9	V4o_O3_220
CH <sub>4</sub>	V3o_CH4_11	V4o_CH4_220
N <sub>2</sub> O	V3o_CH4_11	V4o_N2O_220
HNO <sub>3</sub>	V3o_HNO3_9	V4o_HNO3_220
NO <sub>2</sub>	V3o_NO2_15	V4o_NO2_220
NO	V3o_NO_15	V4o_NO_220
N <sub>2</sub> O <sub>5</sub>	V3o_N2O5_10	V4o_N2O5_220
HNO <sub>4</sub>	V3o_HNO4_12	V4o_HNO4_220
ClONO <sub>2</sub>	V3o_ClONO2_12	V4o_ClONO2_220
ClO	V3o_ClO_11	V4o_ClO_220
HOCl	V3o_HOCl_4	—
CCl <sub>3</sub> F	V3o_CFC11_10	V4o_CFC11_220
CCl <sub>2</sub> F <sub>2</sub>	V3o_CFC12_10	V4o_CFC12_220
CH <sub>2</sub> O	V3o_H2CO_2	—
CO	V3o_CO_12	V4o_CO_220
SF <sub>6</sub>	—	V4o_SF6_221

for the vertical co-ordinate of this climatology. The VMR measurements for each individual profile are vertically binned using the midpoints between the pressure levels (in log-pressure), which define the bins. Since no screening flags are provided with the ACE-FTS data, we use the following filtering methods: data are excluded if the fitting uncertainty value is 100% of its corresponding VMR value and where a given uncertainty value is 0.01% of its corresponding VMR value. This is the technique used for other ACE studies [e.g., Dupuy *et al.*, 2009]. Binned data are subject to various criteria including statistical analysis (for further details, see Jones *et al.*, 2011; 2012). Observations that are larger than three median absolute deviations (MADs) from the median value in each grid cell are disregarded as they are deemed not a true representation (to a high probability, 95%) of the typical state of the atmosphere at a given time and place. Quality-controlled climatological fields are then created for each of the 17 species by considering the measurement uncertainties associated with each binned measurement. Each of the measurements in a bin is weighted by the inverse of the fitting uncertainty to calculate the mean. Furthermore, quality-controlled NO<sub>x</sub> (combination of NO and NO<sub>2</sub>) and NO<sub>y</sub> (combination of NO, NO<sub>2</sub>, HNO<sub>3</sub>, ClONO<sub>2</sub>, N<sub>2</sub>O<sub>5</sub>, and HNO<sub>4</sub>) climatologies have also been derived using a linear combination of the individual atmospheric gas climatologies that contribute to each family. Moreover, these nitrogen species have strong diurnal features and thus climatologies based on separated local sunrise or local sunset measurements have been compiled, in addition to the combined sunrise and sunset climatological fields, using the LST information for each occultation. It should be noted that only one measurement is needed per bin from each individual contributing species in order to produce an eventual NO<sub>x</sub> or NO<sub>y</sub> value for that given

bin. Scaled initial guess profiles are included as they allow for full altitude coverage to be obtained. This technique is described in detail in Jones *et al.* [2012]. A similar approach has been employed when producing the ACE-FTS climatological database [Jones *et al.*, 2011; 2012].

### 3.1.3.12 ACE-MAESTRO climatologies

The ACE-MAESTRO O<sub>3</sub> climatologies are produced using a similar methodology to that of ACE-FTS. ACE-MAESTRO VMR profiles are provided on an altitude grid, and converted to a pressure grid by linearly interpolating the ACE-FTS pressure profiles. Individual ACE-MAESTRO measurements are then binned (as described in Section 3.1.3.11) according to the SPARC Data Initiative pressures. Since no data screening flags are provided, data are only used if the uncertainty value is less than 100% of its given VMR value. Similar to the ACE-FTS climatology, we also apply a three median absolute deviation filter to the ACE-MAESTRO data so that outliers are identified and removed. Finally, a quality-controlled zonal mean average value is calculated using the measurement uncertainties associated with each individual binned measurement.

### 3.1.3.13 HIRDLS climatologies

All HIRDLS data for the SPARC Data Initiative are monthly zonal means created from the V6 Level 2 data. To minimise the impact of missing orbits or bad data points, the L3 processor is used to create a statistically best estimate for each day. These are then averaged to give the monthly mean. The L3 processor reads in all the L2 VMRs for a given product and pressure level over the entire mission and treats the data within 2° latitude bands as time series. Following a suggestion by Rodgers [1976], the data are represented as time-varying zonal means plus the amplitudes and phases of 6 zonal waves. A Kalman filter is used to make sequential estimates of all 13 values, with an estimate of their errors and the RMS difference between the estimated fit from the original measurements. This is done going forward and backward in time, and the estimates combined give the optimal values. Kohri [1981] and Remsberg *et al.* [1990] have described the method in more detail.

For quality control, parameters in each run limit the range of the data to physically reasonable values. In addition, each L2 value has an uncertainty on input, which is checked to make sure it is similar to the RMS differences from the fit. A spike detection is used so that data points that are 6σ from the estimated fit, as estimated from the covariance of the fit, have their weights reduced. This essentially means that these points have virtually no effect on the mapping or the zonal means presented here. Based on validation studies for V6 [Gille and Gray, 2011], the pressure level ranges for the resulting species have been restricted as shown in Table 3.3. It should be noted that data outside of the useful range have been eliminated from publicly released data, including the SPARC Data Initiative climatologies.

**Table 3.3: Trace gas species given with their pressure level ranges for HIRDLS.**

Species	Pressure range (hPa)
Ozone	422 – 0.1
Nitric Acid	100* – 10*
CFC 11	316 – 26.1
CFC 12	316 – 10.0
Daytime NO <sub>2</sub>	56.2 – 1.0
Night-time NO <sub>2</sub>	56.2 – 0.75

\* Best range

### 3.1.3.14 Aura-MLS climatologies

Aura-MLS climatologies are based on Level 2 V3.3. The sole exception is O<sub>3</sub>, which is based on Level 2 V2.2. This is mainly because of the more oscillatory (and poorer) UTLS tropical retrievals from the finer vertical resolution V3.3 data. The validation references and the Aura-MLS data quality documents provide information about the recommended data screening for each species (see individual species sections of this report). These screening methods have been applied for each profile prior to the averaging and interpolation processes that were used to generate the climatological time series used here. This generally means that only profiles with good “Status” and mixing ratios based on acceptable “Quality” and “Convergence” parameter values were included. An attempt to minimise cloud and outlier effects is also included per the MLS-recommended cloud screening methods, as well as other MLS data screening recommendations for each species (*e.g.*, removal of outliers). Also, when mixing ratio precision values are flagged negative, this indicates that the *a priori* information is playing a non-negligible role, and these values are typically not used for producing the averages. In general, only a small percentage of values is excluded *via* these screening methods, although this percentage can sometimes be larger than 20% for the tropical UTLS region (this applies to O<sub>3</sub>, CO, and HNO<sub>3</sub>). Vertical profiles are retrieved as VMRs *versus* a fixed pressure grid (typically with spacing corresponding to 6 levels per decade change in pressure, and double for H<sub>2</sub>O). Also, H<sub>2</sub>O is retrieved as log<sub>10</sub>(VMR). However, the Aura-MLS H<sub>2</sub>O averages are performed in the same way as the other Aura-MLS averages, using mixing ratios, so as to compare most directly with the other climatologies using this averaging method. The pressure ranges used here reflect the recommended levels for Aura-MLS profiles although some additional information often clearly exists beyond these ranges (in particular, for higher altitude regions). Retrieved negative values are sometimes obtained due to the instrument measuring close to its detection limit. Where these measurements have resulted in negative monthly averaged values in the climatologies, the results have been flagged as bad, although it may be that some of the small negative values are within the error bars, and therefore not unreasonable.

### 3.1.3.15 TES climatologies

TES climatologies are based on Level 2 V4 data. Vertical profiles are retrieved as  $\log_{10}(\text{VMR})$  on a 67-level pressure grid, and are interpolated in  $\log_{10}(p)$  to the SPARC Data Initiative pressure grid. Only good quality retrievals have been used, and there is an additional screening to eliminate “C-curve”  $\text{O}_3$  profiles. These profiles, which make up approximately 1-2% of TES V4  $\text{O}_3$  data, result from “jack-knifing” of the retrieval and convergence to an unphysical state in which the  $\text{O}_3$  profile takes on a “C” shape under particular thermal conditions.

As stated in *Section 2.2.15*, TES measures in both Global Survey and Special Observations modes; only Global Survey data are used here. TES data are normally averaged using  $\log_{10}(\text{VMR})$ , but for proper comparison to the other SPARC Data Initiative climatologies, here we use linear averaging. Simple unweighted means of the available data are calculated for each month and latitude bin. A minimum of two observations per bin is required, but in practice the minimum number of profiles is 28 and in most cases the number is  $>1000$ . While the data are provided for the full range of pressures (300 to 0.1 hPa), the sensitivity of the TES  $\text{O}_3$  retrievals drops off dramatically above 10 hPa. Data above this level should be treated with caution.

TES is a thermal instrument that measures radiances both day and night. Each global survey has measurements at two local solar times (equator crossing times of 1:43 and 13:43). The LST\_MEAN value is therefore not provided because it does not reflect an average value for the measurements within the bin. Rather, the LST\_MAX and LST\_MIN variables represent the mean of the day and night LSTs, respectively, within each latitude bin. The variability around these values is small, ranging from  $\pm 55$  minutes near the poles to  $\pm 15$  minutes near the equator.

### 3.1.3.16 SMILES climatologies

SMILES climatologies are based on the Level 2 research (L2r) product V2.0.1. There are two  $\text{O}_3$  products, Band-A  $\text{O}_3$  and Band-B  $\text{O}_3$  for the same  $\text{O}_3$  transition at 625.37 GHz with a different receiver and spectrometer to check the spectrum calibration accuracy. Level 2 data were filtered according to the quality criteria specified for this release. Measurements that were deemed of good quality based on an acceptable “measurement response” and “convergence” parameter values were included. Only clear sky data was provided for the L2r V2.0.1 data product. In this climatology, retrieved VMRs with a measurement response smaller than 0.75 have been rejected and the minimum number of data values required per latitude bin and pressure level was set to five. The pressure range has been limited to  $\geq 10$  hPa for BrO,  $\leq 1$  hPa for  $\text{HNO}_3$ , and  $\geq 25$  hPa for HOCl. Water vapour was retrieved from the continuum but is not included as a product. The quality and sensitivity of each individual

species used in this report, the recommended data screening for each species, and validation references are provided in the SMILES Mission Plan, Version 2.1, ([http://smiles.nict.go.jp/Mission\\_Plan/](http://smiles.nict.go.jp/Mission_Plan/)), and in the SMILES L2r products guide, (<http://smiles.nict.go.jp/pub/data/products.html>). L2r V2.1.5 products have been used in this report where data were made available in time for processing. In the V2.1.5 data a known issue of non-linearity in the spectrum has been improved.

The instrument is on-board the International Space Station in a  $51.6^\circ$  inclined orbit and observations drift slowly with respect to LST, so that all LSTs are sampled for each latitude over a 2-month period. Climatologies of short-lived species are separated into daytime (solar zenith angle  $\leq 87^\circ$ ) and night-time measurements ( $\geq 93^\circ$ ).

## 3.2 Climatology uncertainties

Measurements are imperfect estimates of the truth. Measurement error, defined as the difference between any measurement and the truth, can be decomposed into two parts; a random component that has, over large sample, a mean of zero, and a bias that has a non-zero mean. For satellite-based measurements of trace gas species, the magnitude of the error depends on many factors, including the measurement technique, the chemical species measured, and the time and location of the measurement.

Calculated climatological fields can be affected by the presence of errors in the measurements. Random errors, by definition, have little impact on climatological means. Measurement bias on the other hand will produce a difference between a measurement climatology and the true climatology. Measurement biases can come about due to a number of factors, including (but not limited to) retrieval errors (*e.g.*, the diurnal effect), errors in the input parameters of the retrieval that are assumed to be known but may have their own uncertainties (*e.g.*, spectroscopic data), and so-called smoothing errors related to the spatial resolution of the retrievals. Absolute bias determination for any one satellite instrument is quite difficult since the truth is rarely known, but inter-instrument biases can be deduced through validation exercises.

For limb sounders, one important aspect of the absolute measurement error is the degree to which vertical resolution can smooth the profile. This smoothing error differs between instruments, retrieval schemes and species [*cf.*, Rodgers, 2000 for details]. Therefore, the climatologies will have some instrument-specific characteristics that can be understood only by consideration of the averaging kernels (for example, instruments with better vertical resolution will see a drier hygropause). It should be noted that an instrument with poorer vertical resolution is not *per se* bad; its results are still useful, but the data user must take the instrument and retrieval characteristics properly into account when interpreting the data.

Wherever possible, differences in climatologies within the SPARC Data Initiative will be explained based on the results of prior validation work. However, in addition to the error in the raw measurements, the monthly mean climatologies contain errors introduced by the climatology production. This section will focus on highlighting important sources of climatology error, including added uncertainty due to instrument sampling (Section 3.2.1), and due to differences in averaging techniques (Section 3.2.2). Section 3.2.3 concludes with a description of the climatology error bars used in this report.

### 3.2.1 Uncertainties due to sampling

The monthly zonal mean SPARC Data Initiative climatologies are produced by binning measurements from each instrument in month and latitude bins. Each instrument obtains a finite sample of profile measurements in each bin, based on the space-time pattern of measurement locations for that instrument. The space-time sampling pattern may be dense and uniform, or sparse and highly non-uniform, or somewhere in between. The degree of non-uniformity of the sampling pattern, together with the space-time gradients in the measured field may lead to a difference between the sample mean and the true mean.

This sub-section briefly describes an exercise that aims to produce pseudo-quantitative estimates of sampling bias for a number of instruments participating in the SPARC Data Initiative. These sampling biases can be seen as example cases, and can be used to highlight regions and seasons of significant sampling bias, and its approximate magnitude. This information should help in the comparisons of instrument climatologies in other chapters.

Sampling patterns have been collected from each instrument team, and defined by day, latitude and longitude of measurement locations. For many instruments, a typical year of actual sampling locations has been used in the analysis, rather than, for instance, a time series of all possible measurements, which may differ because of *e.g.*, data download limitations. The time periods used to define each instruments' sampling pattern are the same as those used to produce the sampling density figures in the instrument descriptions of Chapter 2.2.

We have used output from the WACCM3, a fully coupled chemistry-climate model, spanning the range of altitude from the Earth's surface to the thermosphere [Garcia *et al.*, 2007]. The particular version of the model used here (3.4.58) is the same as that used for the last Chemistry-Climate Model Validation Activity [SPARC CCMVal, 2010], except that the number of vertical levels has increased to 102, and the number of chemical species included has increased to 125. The horizontal resolution is 1.9° by 2.5° (latitude by longitude). Here, we use model output with daily resolution at 0 UTC from one year of a transient simulation under current climate conditions.

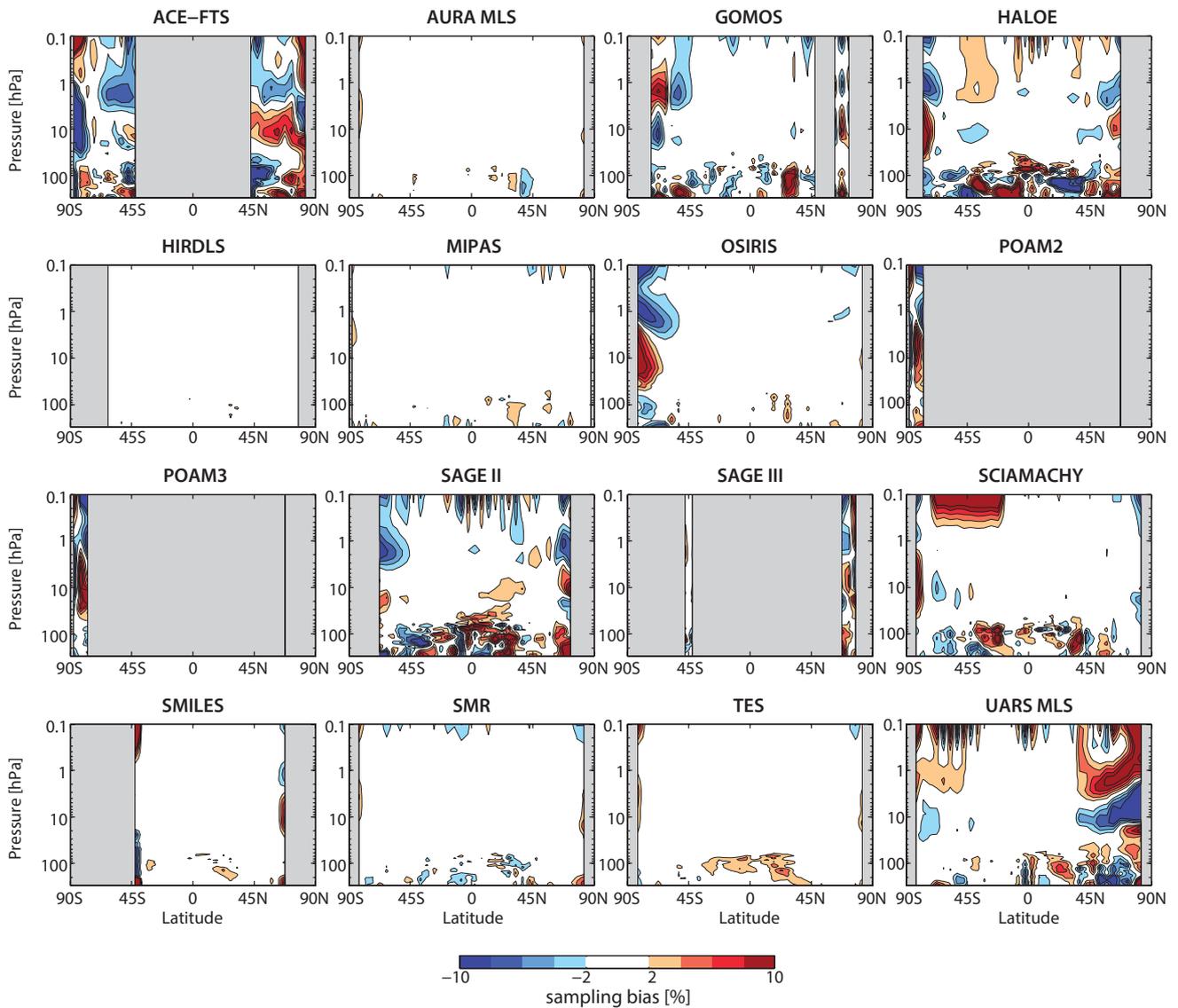
Instrument sampling patterns for each month of the year are used to subsample the model data. For each sample, model fields from the corresponding Julian day are linearly interpolated in space to the latitude and longitude of the sample location. (Interpolation is not performed to the time-of-day of the measurements, since the effect of diurnal variability on SPARC Data Initiative climatologies is explicitly dealt with for short-lived species, for which the diurnal cycle is important.) Once model data have been interpolated to each sample location, the subsampled fields are binned according to the SPARC Data Initiative latitude grid, and the mean is calculated. The "true" model climatology, or population mean, is produced by first calculating the mean of all model fields on each latitude circle of the model's latitude grid, then linearly interpolating these mean values to the midpoint of each SPARC Data Initiative latitude bin. The difference between the instrument-sampling-pattern-based field mean and the full-model-resolution field mean gives the sampling bias. For each month and for each instrument, this bias is calculated for every latitude bin in which an instrument has measurements, and at all pressure levels of the model fields.

As an example result, the monthly zonal mean sampling bias for O<sub>3</sub> in March is shown for each instrument as a function of latitude and height in Figure 3.1. Monthly zonal mean climatology sampling bias estimates from the sampling exercise for O<sub>3</sub> for all months and for all instruments are available in Appendix A. The results of the sampling bias exercise can be very briefly summarised by categorizing instruments according to the severity of their sampling bias. We see:

- i. A weak sampling bias (always <5%) for dense samplers Aura-MLS, HIRDLS, MIPAS, SMR and TES.
- ii. Strong sampling bias (>5%) for occultation instruments ACE-FTS, HALOE, POAM II, POAM III, SAGE II, SAGE III, and GOMOS which is strongest at, but not limited to high latitudes.
- iii. Occasionally (in time or space) strong (> 5%) sampling bias for OSIRIS, SCIAMACHY, SMILES and UARS-MLS.

The largest sampling biases can be understood to be a product of non-uniform sampling throughout the days of a month, as can be seen when one examines variations in ozone over a month and the correlation of these variations with instrument sampling patterns. Figure 3.2 shows the time evolution of zonal mean O<sub>3</sub> in March from the model, at pressure levels 100, 10 and 1 hPa, as anomalies from the monthly zonal mean. Superimposed on the chemical fields are latitude *versus* time sampling patterns of ACE-FTS and MIPAS, as examples of the two extremes in types of sampling patterns.

The MIPAS sampling pattern contains measurements in all latitude bins for all days, *i.e.*, there is no variation in the sampling locations with time, and as a result the sampling bias is small. ACE-FTS, on the other hand, as a solar occultation instrument, samples each latitude band over only a few days of the month. For example, in the month of March, SH



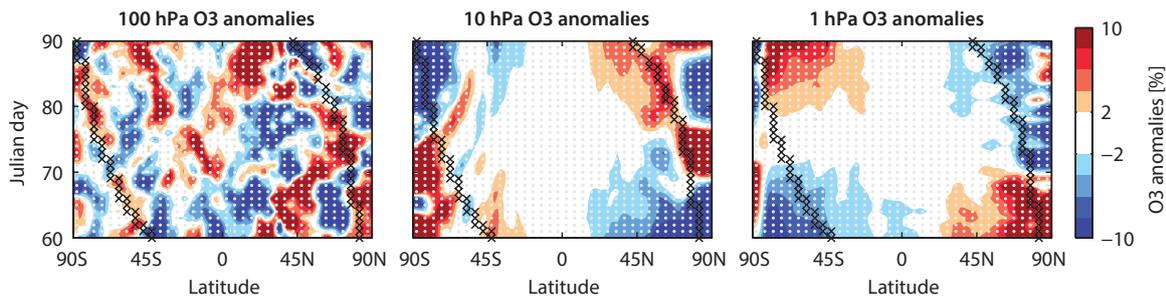
**Figure 3.1:** Latitude-height sections of calculated sampling error for  $O_3$  in March, based on sampling patterns of instruments as labelled in each panel. Grey regions denote regions of no measurements.

mid-latitudes ( $45^\circ\text{S}$ ) are sampled only at the very beginning of the month, while SH high latitudes ( $80^\circ\text{S}$ ) are sampled only at the very end of the month. At 1 hPa, ozone mixing ratios are increasing through the month over this latitude range, therefore, the ACE-FTS sampling pattern leads to negative sampling bias around  $45^\circ\text{S}$ , and slightly positive sampling bias at the highest latitudes. The seasonal cycle of ozone is comparatively reversed at 10 hPa, leading to slightly positive bias in the SH mid-latitudes, and negative bias in the SH high latitudes. In this way, it can be seen that the sampling biases of ACE-FTS can be well explained by the instrument's sampling pattern and the intra-monthly variations in ozone, which depend strongly on height and latitude. At 100 hPa, intra-monthly  $O_3$  variations are relatively noisy, and as a result the sampling bias is dependent on the sampling of the intra-monthly variability. We therefore can expect that in regions where the sampling bias is due to the non-uniform sampling of the slow seasonal variability through a month, that the sign and approximate magnitude of the sampling bias calculated through our model exercise to be a reasonably accurate estimate of the real sampling

bias for each instrument. However, in regions where variability is dominated by short-term (intra-monthly) variations, limited sampling of such a chemical field will lead to a random sampling error. In this case the sign and magnitude of the sample error calculated through our model exercise serves as an example, and should be used only to identify regions where sampling error may be important.

The sampling biases for solar occultation instruments are similar to that of ACE-FTS, and are primarily a result of the non-uniform day-of-month sampling. The sampling biases of OSIRIS and UARS-MLS come from a similar source: while these instruments have dense sampling patterns, the latitudinal coverage of their measurements changes periodically, and as a result, certain latitudes (or in fact a whole hemisphere) are often sampled for less than the full month. Such is the case for OSIRIS in the SH and UARS-MLS in the NH in the sampling error exercise results shown in **Figure 3.1**.

In general, the sampling bias for all instruments is weak in the tropics where variability is weak on both intra-seasonal



**Figure 3.2: March  $O_3$  anomalies from the March monthly mean shown as a function of latitude and Julian day for the 100, 10 and 1 hPa surfaces. The locations of the latitude bins that contain measurements according to the MIPAS (grey dots) and ACE-FTS (black crosses) sampling patterns.**

and seasonal time scales. In the extra-tropics and polar regions, where variability is more pronounced, the sampling bias becomes much larger. Between  $60^\circ$ - $65^\circ$  in both hemispheres, sampling bias has a double-peak structure, with maximum values around 20 and 2 hPa. It is interesting to note that the solar occultation instruments ACE-FTS, HALOE and SAGE II, as well as OSIRIS, show similar sampling biases for March at around 1 hPa between  $45^\circ$ - $65^\circ$  in both hemispheres due to similarities in the seasonal progression of their sampling patterns. This is one example where close agreement between data climatologies from different instruments may not imply good agreement with the true climatological mean.

In order to assess how the sampling bias can affect annual mean climatologies, we calculate the annual mean sampling bias for each instrument by averaging the sampling biases for the 12 calendar months. These annual mean sampling biases are shown for each instrument in **Figure 3.3**.

The instruments with the highest sampling density (Aura-MLS, HIRDLS, MIPAS, and TES) show small annual mean sampling biases of only a few percent, as would be expected due to the small sampling biases in their monthly means. Due to the seasonal variability of the OSIRIS and UARS-MLS sampling patterns, their sampling bias somewhat cancels out in the annual average, with maximum values of a few percent. Finally, for the occultation instruments (ACE-FTS, GOMOS, HALOE, POAM II, POAM III, SAGE II, and SAGE III), the annual mean sampling biases are on the order of 5% at latitudes  $>50^\circ$  in both hemispheres. The details of the sampling bias – its sign and magnitude – are generally different for the different instruments, however, some features are common to multiple instruments (*e.g.*, negative sampling bias at 1 hPa and  $\sim 60^\circ$  in both hemispheres) and are related to similarities in the sampling patterns.

In summary, when constructing climatologies by averaging binned atmospheric measurements, sampling bias can arise due to non-uniform sampling in time or space. We have examined sampling biases produced by the sampling patterns of a number of instruments participating in the SPARC Data Initiative using ozone from WACCM. We find that:

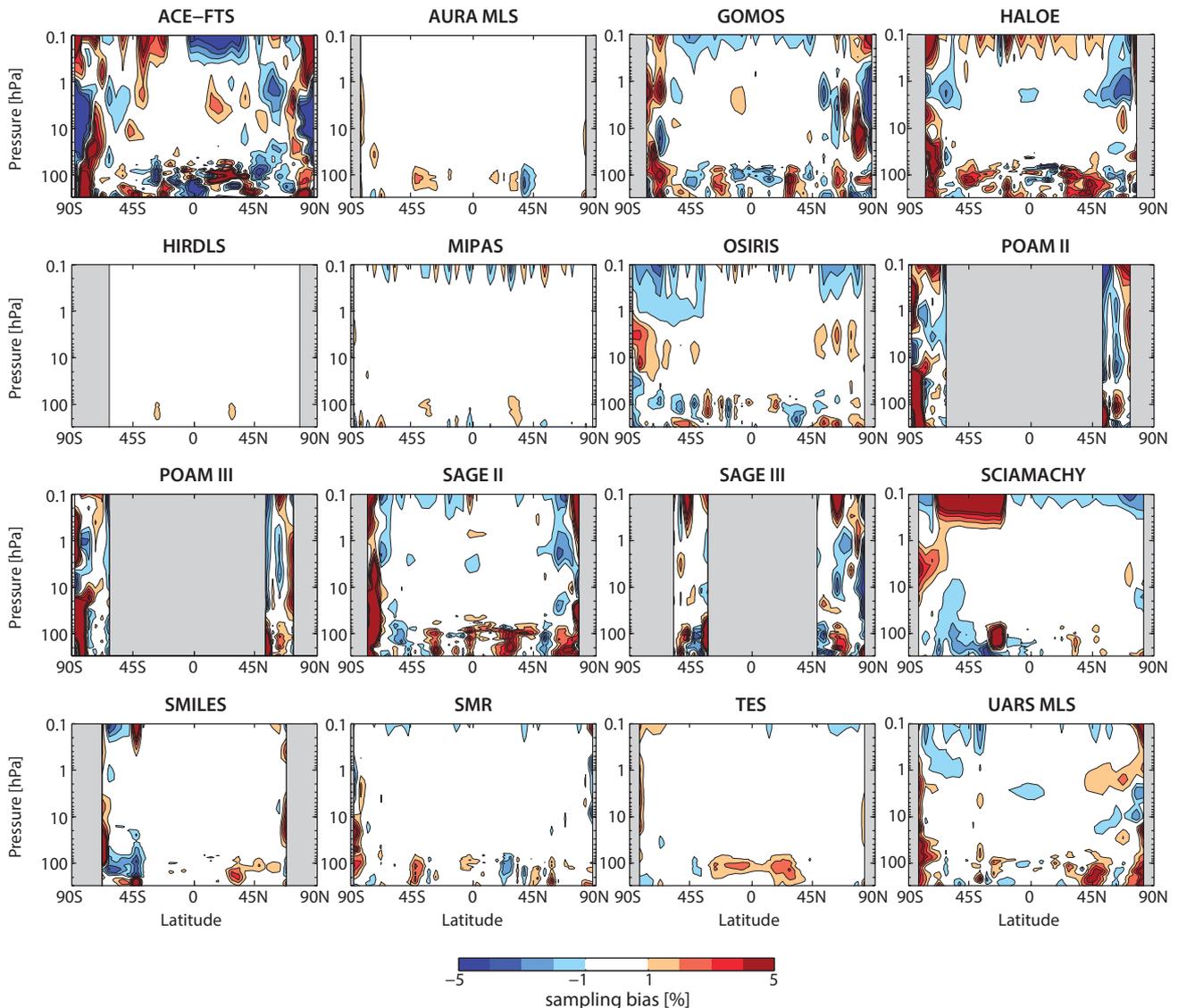
- Climatologies based on measurements from instruments with high sample density generally have small

sampling biases due to their highly uniform sampling of each latitude bin.

- Climatologies based on measurements from instruments whose latitudinal coverage varies with time can have strong sampling biases for certain months and locations. Sampling biases for  $O_3$  were found in some instances to be above 10%. This is primarily due to non-uniformity in day-of-month sampling, and occurs whenever an instrument provides measurements in one month over only a portion of that month. Whenever the atmospheric variability is dominated by the seasonal cycle, this type of sampling error could in theory be reasonably well quantified or even corrected, however, when variability is dominated by intra-seasonal (short-term) variations, only the absolute magnitude of the sampling bias can be estimated from model studies. This type of sampling bias is most relevant for solar occultation instruments, but also for instruments with high sample density when the latitudinal coverage changes with time, such as OSIRIS, SMILES and UARS-MLS.
- Annual mean sampling bias can be on the order of 5% or larger for solar occultation instruments at high latitudes, and a few percent for instruments with varying latitudinal coverage such as OSIRIS, SMILES and UARS-MLS.
- In the UTLS region, intra-monthly variations and gradients in many trace gas species are large, therefore the sampling bias is more important. The sampling bias for  $O_3$  in monthly mean climatologies is found to be often on the order of 10% (higher for  $H_2O$ ; not shown), and still significant in annual mean climatologies. For precise monthly-mean or annual-mean climatologies in the UTLS, one requires a high sample density.

### 3.2.2 Uncertainties due to averaging technique

Averaging of data may lead to biases between climatologies in cases when different averaging procedures are used to generate the climatologies. Averages are typically defined as monthly zonal mean VMRs, but averages of  $\log_{10}$ (VMR) or of median values of the spatio-temporal distributions are also used. Under particular atmospheric conditions,



**Figure 3.3:** Latitude-height sections of calculated annual mean sampling error for  $O_3$ , based on sampling patterns of instruments as labelled in each panel. Grey regions denote regions of no measurements.

these averaging methods can lead to significantly different results for many trace gas species. As an example, we show in **Figure 3.4** monthly zonal mean distributions of  $H_2O$ ,  $CO$  and  $O_3$  and their standard deviations calculated from WACCM model simulations described in *Jackman et al.* [2008] for November 2003. The monthly zonal means are calculated from 10,000 modelled mixing ratios per species for each latitude-pressure grid point, and are compared to averages calculated in  $\log_{10}(\text{VMR})$  space, as well as to their respective median values. The following conclusions can be drawn from this comparison:

- The bias between differently averaged zonal mean fields (*i.e.*, linear or logarithmic averages or median values) correlates spatially with the standard deviation of the distributions.
- Standard deviations and hence biases are most pronounced where spatial gradients are strongest, *e.g.*, in regions of transport barriers or strong vertical transport. In our example, this occurs for  $CO$  in the polar regions in the mid-stratosphere and is related to vertical transport by the meridional circulation.

$H_2O$  variability is highest in the UTLS. Additionally, averaging biases related to diurnal variations are found for  $O_3$  in the mesosphere.

- Logarithmic averaging always yields smaller values than linear averaging.
- Median values can be higher or lower than linearly averaged zonal means.

The sign of the bias depends on the asymmetry of the distribution. This is particularly evident in the case of  $O_3$  in the mesosphere where the  $O_3$  distribution is bi-modal due to diurnal effects. In the summer hemisphere, where daytime population is dominant, the median yields values closer to the daytime VMR and hence is smaller than the linear average, while the opposite occurs in the winter hemisphere.

Most of the climatologies within the SPARC Data Initiative were built on the basis of linear monthly zonal means, though exceptions exist; *e.g.*, GOMOS  $O_3$  and  $NO_2$  climatologies and SMR  $H_2O$  from the 544.6 GHz band (SMR2) are

based on median values, while OSIRIS NO<sub>2</sub> and NO<sub>x</sub> climatologies are based on log<sub>10</sub>(VMR). The comparison of these climatologies with those of other instruments (see *Chapter 4*) might therefore suffer from statistical averaging biases.

In the case of GOMOS O<sub>3</sub>, however, such pronounced mesospheric biases resulting from the use of the median as seen in **Figure 3.4** are not expected since GOMOS measures only during the night-time and issues related to different diurnal populations do not play a role in the averaging technique. Remaining biases, most likely located in the UTLS region, are expected to be within 15%, which is considerably smaller than the inter-instrumental spread observed in this altitude range. GOMOS NO<sub>2</sub> median values are likely to be smaller than linear averages at the edge of NO<sub>x</sub>-rich air masses descending in polar winter, as observed for CO. On the other hand, a slightly positive bias might occur in the core of these air masses. As in the case of O<sub>3</sub>, averaging biases related to diurnal variations are not expected to occur. Regarding the SMR H<sub>2</sub>O climatology obtained from the 544.6 GHz band (SMR2), biases related to the use of median values might be an issue. **Figure 3.4** indicates deviations on the order of ±20% in the altitude range 16–20 km (~100–60 hPa) where this data product is provided.

No important averaging biases are expected for the OSIRIS NO<sub>2</sub> and NO<sub>x</sub> climatologies since they are restricted to sunlit conditions (*i.e.*, no diurnal issues) and do not cover the polar winter regions where averaging differences related to the mixing of NO<sub>x</sub>-rich mesospheric and stratospheric air masses might occur.

Apart from these biases, which arise from the comparison of differently averaged climatologies, there exists an intrinsic source of statistical averaging errors for climatologies built from trace gas abundance data retrieved in the log<sub>10</sub>(VMR) space (*i.e.*, CO, NO, NO<sub>2</sub>, and H<sub>2</sub>O from MIPAS, SMR, OSIRIS, SCIAMACHY). A detailed discussion of this error source on basis of idealised retrieval simulations is given in *Funke and von Clarmann [2011]*. A quantitative evaluation of related errors in the context of this study is not feasible due to the complex dependence of their magnitude and sign on natural variability, measurement sensitivity, and retrieval constraints. However, efforts have been undertaken in the definition and optimisation of the instrument-specific retrieval algorithms operating in the log<sub>10</sub>(VMR) space in order to reduce these errors whenever possible.

### 3.2.3 Climatology error bars

The statistical uncertainty in a mean value, calculated from  $n$  measurements with a standard deviation  $\sigma$ , is commonly estimated through the standard error of the mean (SEM):

$$\text{SEM} = \sigma / n^{1/2}. \quad (3.1)$$

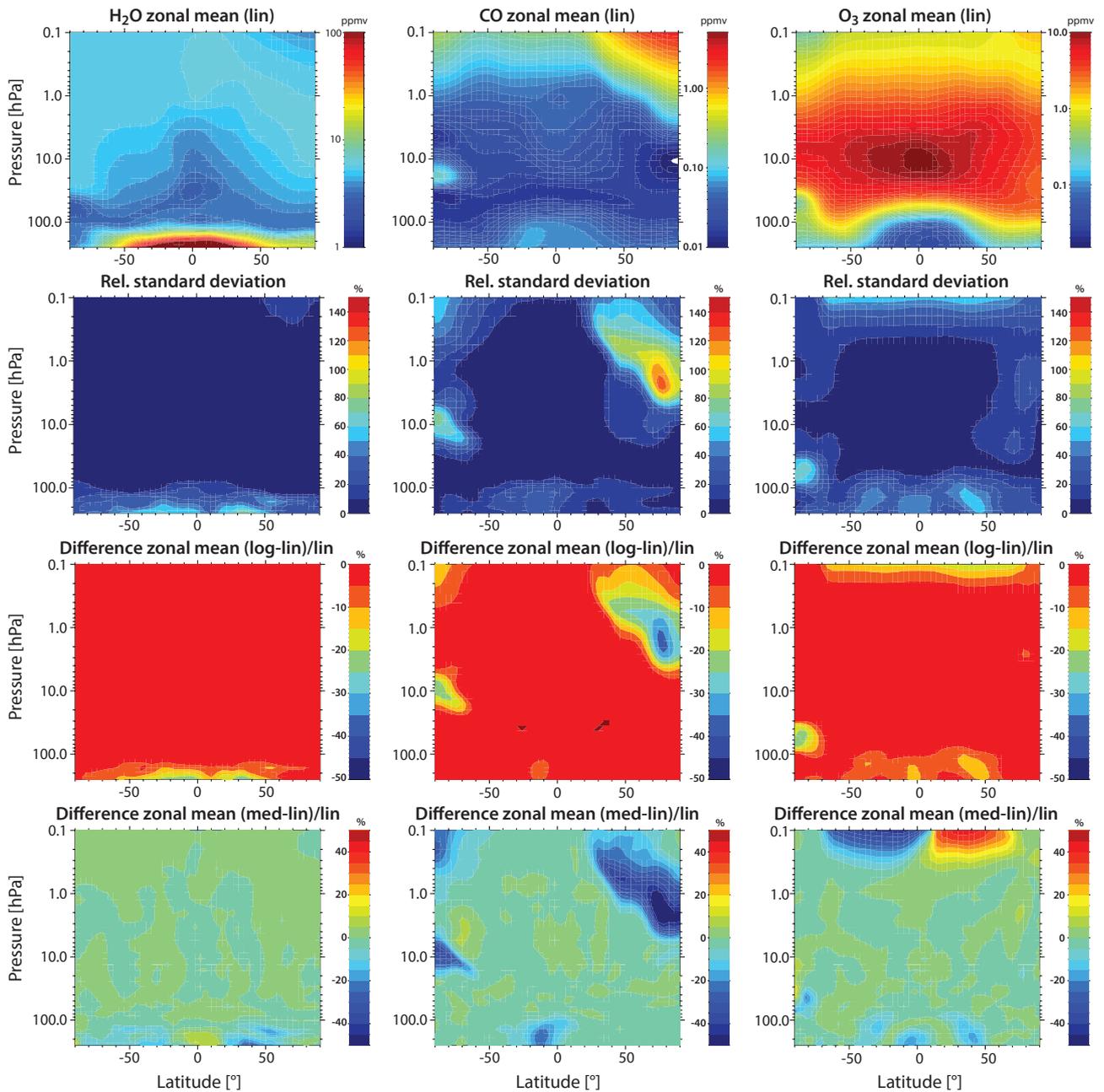
The SEM is an estimate of the standard deviation of all the possible mean values one would produce if one was able to

re-sample the original population from which the sample is drawn. The formalism of the SEM assumes that individual samples are independent. This may not be the case within the SPARC Data Initiative, since, for example, the sampling patterns of some instruments may be dense enough that closely spaced measurements are autocorrelated. In fact, satellite data sorted into latitude bands may exhibit positive or negative autocorrelations, depending on the details of the sampling pattern and latitude grid [*Toohey and von Clarmann, 2013*]. It is therefore not possible to know whether the “classical” SEM, as calculated by Equation 3.1, is in general an over- or underestimate of the true uncertainty in the mean climatologies.

Standard deviations are also affected by the climatology production. The standard deviations are themselves a function of both the random measurement error and the natural variability sampled at the spatial and temporal resolution/pattern of the instrument. Thus, the magnitude of the natural variability present in the climatological standard deviation fields is also subject to sampling error compared to the true variability within a latitude bin. In some cases, it may be preferable or necessary to interpolate the standard deviation to latitude grid midpoints (see *von Clarmann et al., 2012*). It should be noted that linear interpolation, as used to produce climatologies on a standard vertical grid, will decrease the variability of a field when the correlation between adjacent points is low (*i.e.*, when random measurement errors are large compared to natural variability). Due to this effect, the standard deviation of the climatologies will in some cases be less than the standard deviation calculated on an instrument’s native retrieval grid. This reduction in standard deviation is artificial in that any interpolation between two data points on the original grid acts to reduce the uncertainty associated with the random measurement error, as when calculating the mean of multiple data points.

Despite its shortcomings, due to its ease of computation and its frequent use in past studies, the SEM as calculated *via* Equation 3.1 using the standard deviations provided in the climatology will be used in this report to indicate an approximate measure of uncertainty in each climatological mean. In particular, uncertainties in the mean will be graphically illustrated by 2×SEM error bars, which can be loosely interpreted as a 95% confidence interval of the mean.

It should be stressed that the statistical error in the mean is in many cases much smaller than the overall error of the climatology, which contains the systematic errors of both the measurements and the climatology construction. We have briefly explored the potential importance of two types of climatology error in this subsection, but this discussion is not exhaustive. For example, potential biases introduced through filtering of retrievals used in the climatology construction (*e.g.*, including only cloud-free measurements) are not addressed here. A complete characterisation of the systematic errors of each climatology is beyond the scope of this report and would require



**Figure 3.4:** Monthly zonal mean distributions of  $\text{H}_2\text{O}$ ,  $\text{CO}$  and  $\text{O}_3$  and their standard deviations calculated from WACCM for November 2003 in the two upper rows. Differences of linear mean values to averages calculated in  $\log_{10}(\text{VMR})$  space, as well as to their respective median values are shown in the two lower rows.

a precise knowledge of the absolute measurement uncertainties for all instruments. Since such knowledge is not available in a consistent way for all instruments, it is recommended that future efforts that focus on deriving absolute measurement uncertainties. The uncertainties would need to include a range of error sources such as uncertainty in the spectroscopic data, calibration, pointing accuracy, and others. The uncertainties would need to be derived consistently between the instruments according to a common standard so to allow for apple-to-apple comparisons. In the absence of such bottom up measurement uncertainties, we will use the inter-instrument spread of the climatologies to provide a measure of the overall uncertainty in the underlying chemical fields.

### 3.3 Climatology diagnostics

A set of standard diagnostics is used to investigate and test the differences between the trace gas time series obtained from each instrument. The diagnostics include annual and monthly zonal mean climatologies, vertical and meridional mean profiles, seasonal cycles, and interannual variability. In addition, trace gas-specific evaluations such as the tape recorder for  $\text{H}_2\text{O}$  and the quasi-biennial oscillation (QBO) for  $\text{O}_3$ , which test the physical consistency of a dataset, are carried out. Such diagnostics include the latitude-time or altitude-time evolution of trace gases that are sensitive to specific transport processes, such as descent within the polar vortex or the seasonal variation in the strength of the

Brewer-Dobson circulation. The evaluation methods are described in more detail in the following.

### 3.3.1 The multi-instrument mean (MIM)

We introduce the concept of the multi-instrument mean (MIM), which we use throughout the report as a common point of reference. The MIM is calculated by taking the mean of all available instrument climatologies within a given time period of interest. Note, that the MIM is not a data product and will not be provided with the instrument climatologies. By no means should the MIM be regarded as the best estimate of the atmospheric state, since all instruments are included in its calculation regardless of their quality and without any applied weighting applied. Where instruments offer more than one data product of a given trace gas species, only one data product is included in the MIM, so not to bias the MIM towards this instrument.

Throughout the report we calculate relative differences between the trace gas mixing ratios of an instrument ( $\chi_{\text{instrument}}$ ) and the MIM ( $\chi_{\text{MIM}}$ ) using

$$100 \times (\chi_{\text{instrument}} - \chi_{\text{MIM}}) / \chi_{\text{MIM}} \quad (3.2)$$

It should be emphasised that when interpreting relative differences with respect to the MIM, one must keep in mind that the set of instruments from which the MIM was calculated may have changed in between time periods. Also, if there is an unphysical behaviour in one instrument, the MIM and thus the differences with respect to the MIM of the other instruments will most certainly reflect this unphysical behaviour. Finally, if one instrument does not have global coverage for every month a non-physical structure may be introduced into the MIM that reflects this sampling

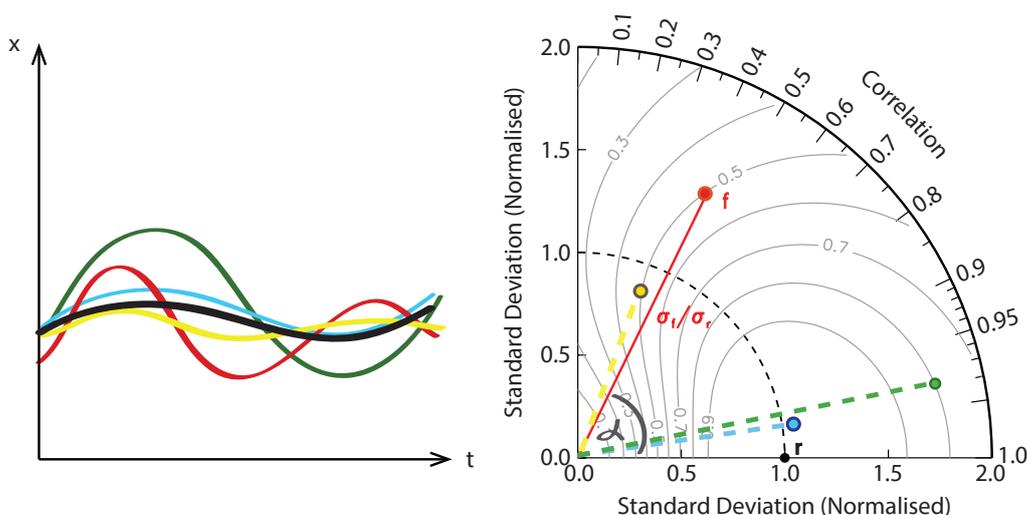
issue. Despite its shortcomings, we have chosen to use the MIM throughout the report as a common point of reference for comparison between instruments, in order to avoid singling out any particular instrument as a benchmark.

### 3.3.2 Annual and monthly mean cross sections and profiles

For the annual and monthly mean cross sections, as well as the altitude and meridional profile evaluations, multi-annual means were produced in order to reduce potential sampling errors, and to limit the influence of interannual variability, e.g., through the QBO. However, we also intended to compare a maximum number of available instruments for the same time period, so often a trade-off between number of instruments and length of the climatology had to be made. The monthly or annual zonal mean cross sections are analysed to investigate mean biases in the datasets. The vertical and meridional profiles help focus on particular height/latitude regions and months. This evaluation (along with other evaluations that follow) will also help to determine if biases between datasets are persistent over the entire year. The comparison of cross sections (or profiles) from individual instruments is based on the relative differences of each instrument to the MIM (see Section 3.3.1).

### 3.3.3 Seasonal cycles

For the seasonal cycles, the multi-year approach has been chosen. The seasonal cycle results include the MIM (see explanation above) together with its  $1\sigma$  standard deviation, which is a measure of the range of mean values obtained from the different instruments. A combined annual and semi-annual fit has been applied to all the available monthly mean values of a single instrument, in order to yield a



**Figure 3.5:** Left panel: Exemplary seasonal cycles corresponding in colour to the dots in the Taylor diagram. Right panel: Taylor diagram describing the agreement between the reference field ( $r$ ) and a test field ( $f$ ). The angle  $\alpha$  represents the correlation between the fields. The radial distance shows the amplitude in the seasonal cycle of the test field normalised by the standard deviation of the reference field ( $\sigma_f / \sigma_r$ ). The grey thin lines indicate the skill score of the test field, which is an overall metric of the agreement (see text for explanation).

seasonal cycle that is comparable even for instruments that do not measure for all months of the year. Finally, Taylor diagrams [Taylor, 2001] are used in order to compare the different instruments in a more quantitative way. Taylor diagrams offer a visual summary of the pattern statistics of how well a certain instrument's seasonal cycle reproduces the seasonal cycle of a reference field or a 'true' state. Three measures can be deduced from the Taylor plots as illustrated in **Figure 3.5**: the correlation on the azimuthal axis, which represents how well the phase of the true seasonal cycle is reproduced by the instrument; the normalised amplitude on the x- and y-axis; and the skill factor, indicated by the light grey lines, which summarises the overall performance of an instrument's field. The closer the instrument lies to the '1' on the x-axis, the better it agrees with the reference field. The Taylor diagram shown in **Figure 3.5** demonstrates that the blue seasonal cycle is closest to the reference field ( $r$ , black), with a skill score of about 0.97, green shows a similar phase, but too large an amplitude (resulting in a skill score of about 0.8), yellow shows the wrong phase but the right amplitude (skill score 0.5), and red shows the wrong phase with too large an amplitude (skill score 0.5).

Note, that the Taylor diagrams do not include information on the performance of how well the instruments reproduce the mean values of the seasonal cycles, so this measure needs to be examined in addition. Please see *Hegglin et al.* [2010] for an additional example of how to interpret Taylor diagrams.

### 3.3.4 Time series of latitude and altitude profiles

Time series of both the absolute values and deseasonalised anomalies are used to analyse intra-annual and interannual variability in the trace gas datasets. Examples of time series based on absolute values are the H<sub>2</sub>O tape recorder or polar dehydration evaluations, which show the time-pressure evolution of absolute mean values over several years. In some instances, the latitude or altitude time series are averaged over several years so to yield a more robust estimate of the mean annual evolution of monthly zonal mean values.

Deseasonalised time series are shown for selected latitude bands and pressure levels or as an altitude-time evolution of the trace gas, *e.g.*, to analyse the QBO. For each month the anomalies are calculated by subtracting the multi-year mean value of the month of the respective instrument (averaged over all years taken into account for this diagnostic) from the monthly mean values.

### 3.3.5 Summary plots

We use two different types of summary plots in order to present an overview of the findings within each trace gas chapter: one highlighting the uncertainty in our knowledge of the atmospheric mean state; and the other highlighting specific inter-instrument differences.

For each trace gas species the first type of summary plot shows the inter-instrument spread of climatologies to give some measure of the overall uncertainty in the underlying chemical fields. Annual zonal MIM, multi-instrument minimum (MIN) and multi-instrument maximum (MAX) fields are provided, with the latter two based on the minimum and maximum over all instruments estimated separately for each grid point. The difference between MAX and MIN, as well as the standard deviation over all instruments, is presented in absolute and relative values to demonstrate the maximum spread and the variations from the MIM over all instruments. Again, the two quantities are estimated separately for each grid point.

In the second type of summary plot, average deviations of each instrument from the MIM are presented for different regions showing which datasets are consistent with each other and which not. The regions are divided into different altitude ranges (300-100 hPa; 100-30 hPa; 30-5 hPa; 5-1 hPa; 1-0.1 hPa) and into the extra-tropics (40°-80°S/N) and the tropics (20°S-20°N). The tropics show somewhat smaller variability than the extra-tropics, hence trace gas evaluations are generally less sensitive to sampling issues and give a cleaner estimate of the overall measurement error. In the extra-tropics, inter-instrument differences are expected to be larger due to larger dynamical variability and hence greater sensitivity to sampling issues. The average deviation of each instrument for a particular region is calculated as the median (MED) over all values the instrument exhibits in this region. The median is regarded to be more robust against outliers. Additionally, the median absolute deviation (MAD) is provided for each instrument and region. The MAD over the sample  $x = (x_1, \dots, x_n)$  is defined as:

$$\text{MAD} = \text{MED} ( |x - \text{MED}(x)| ) \quad (3.3)$$

and represents the interval around the median that contains 50% of the data [Rousseeuw and Croux, 1993]. For comparison, the range indicating the mean  $\pm 1\sigma$  is also indicated.

