Evaluation of the inter-annual variability of stratospheric chemical composition in chemistry-climate models using ground-based multi species time series

Virginie Poulain, Slimane Bekki, Marion Marchand, Martyn P. Chipperfield, Myriam Khodri, Franck Lefèvre, Sandip S. Dhomse, Greg E. Bodeker, Ralf Toumi, Martine de Mazière, et al.

Evaluation of the inter-annual variability of stratospheric chemical composition in chemistry-climate models using ground-based multi species time series

V. Poulain[1,2], S. Bekki[1], M. Marchand[1], M. P. Chipperfield[3], M. Khodri[2], F. Lefèvre[1], S. Dhomse[3], G. E. Bodeker[4], R. Toumi[5], M. De Maziere[6], J.-P. Pommereau[1], A. Pazmino[1], F. Goutail[1], D. Plummer[7], E. Rozanov[8,9], E. Mancini[10], H. Akiyoshi[11], J.-F. Lamarque[12], and, J. Austin[13],

[1] Laboratoire Atmosphères, Milieux, Observations Spatiales ; Université Versailles Saint-Quentin-en-Yvelines ; Université Pierre et Marie Curie - Paris 6 ; INSU ; CNRS : UM8190, France

[2] Laboratoire d'Océanographie et du Climat: Expérimentations et approches numériques, Sorbonne Universités, UPMC Université Paris 06, IPSL, UMR CNRS/IRD/MNHN, F-75005 Paris, France.

[3] School of Earth and Environment, University of Leeds, Leeds, UK

[4] Bodeker Scientific, New Zealand

[5] Imperial College, London, UK

[6] Belgian Institute for Space Aeronomy (BIRA-IASB), Brussels, Belgium

[7] Canadian Ctr Climate Modelling & Anal, Victoria, BC, Canada

[8] {Physikalisch-Meteorologisches Observatorium Davos and World Radiation Center (PMOD/WRC), Davos, Switzerland }

[9] {Institute for Atmospheric and Climate Science, Swiss Federal Institute of Technology Zurich, Universitaetstrasse 16, CHN, CH-8092 Zurich, Switzerland }

[10] Univ Aquila, Dept Phys & Chem Sci, I-67100 Laquila, Italy

[11] Natl Inst Environm Studies (NIES), Tsukuba, Ibaraki, Japan

[12] National Center for Atmospheric Research, Boulder, Colorado, USA

[13] Enigma Scientific Publications, Winnersh, Berkshire, UK

*Correspondence to: LATMOS , Tour 45-46, 4e étage, Boîte 102, 4, Place Jussieu 75252 Paris Cedex 05. Phone : +33 (0)1 44 27 84 43. virginie.poulain@latmos.ipsl.fr

Abstract

The variability of stratospheric chemical composition occurs on a broad spectrum of timescales, ranging from day to decades. A large part of the variability appears to be driven by external forcings such as volcanic aerosols, solar activity, halogen loading, levels of greenhouse gases (GHG), and modes of climate variability (quasi-biennial oscillation (QBO), El Niño-Southern Oscillation (ENSO)). We estimate the contributions of different external forcings to the interannual variability of stratospheric chemical composition and evaluate how well 3-D chemistry-climate models (CCMs) can reproduce the observed response-forcing relationships. We carry out multivariate regression analyses on long time series of observed and simulated time series of several traces gases in order to estimate the contributions of individual forcings and unforced variability to their internannual variability. The observations are typically decadal time series of ground-based data from the international Network for the Detection of Atmospheric Composition Change (NDACC) and the CCM simulations are taken from the CCMVal-2 REF-B1 simulations database. The chemical species considered are column O_3, HCl, NO_2, and N_2O. We check the consistency between observations and model simulations in terms of the forced and internal components of the total interannual variability (externally forced variability and internal variability) and identify the driving factors in the interannual variations of stratospheric chemical composition over NDACC measurement sites. Overall, there is a reasonably good agreement between regression results from models and observations regarding the externally forced interannual variability. A much larger fraction of the observed and modelled interannual variability is explained by external forcings in the tropics than in the extratropics, notably in polar regions. CCMs are able to reproduce the amplitudes of responses in chemical composition to specific external forcings. However, CCMs tend to underestimate very substantially the internal variability and hence the total interannual variability for almost all species considered. This lack of internal variability in CCMs might partly originate from the surface forcing of these CCMs by analysed SSTs. The results illustrate the potential of NDACC ground-based observations for evaluating CCMs.

1 Introduction

The variability of stratospheric chemical composition occurs on a broad spectrum of time scales, ranging from hours to decades. Some of this variability involves couplings between chemistry and dynamics and, more generally, chemistry-climate interactions and is driven by external forcings (e.g. volcanic aerosols, solar activity, halogen loading (e.g. CFCs and halons), levels of greenhouse gases (GHG)) and modes of climate variability (e.g. QBO, ENSO). These forcings are thought to be responsible for most of the interannual variability in stratospheric chemical composition. Many observational studies have attempted to link variations in stratospheric chemical composition, in particular ozone, to these forcings. Although several studies have considered vertically resolved ozone datasets (e.g. Zawodny et al., 1991; Brunner et al., 2006; Randel et al., 2011), the most commonly analysed datasets are satellite column ozone data. Studies of global satellite data records such as TOMS (Total Ozone Mapping Spectrometer) or SAGE (Stratospheric Aerosol and Gas Experiment) have documented the impact of the QBO on column ozone (Bowman et al., 1989; Zawodny et al., 1991; Randel et al., 1994; Yang et al., 1994; Baldwin et al., 2001; Randel et al., 2007). Tropical column ozone variations were found to be approximately in phase with equatorial zonal wind near 30 hPa, an indicator of the QBO, whereas the extra-tropical column ozone anomalies tended to be out of phase with the tropical signal. Linear regressions showed that the mean amplitude of the QBO signal is about 2-4% of the mean column ozone (WMO, 1999). The ENSO is also known to influence column ozone (Shiotani et al., 1992; Randel et al., 1994; Steinbrecht et al., 2006). However, unlike the QBO, the effects of ENSO on column ozone are mostly visible in the zonal distribution. Another significant source of variability in ozone is the variations in stratospheric aerosol loading which are predominantly controlled by volcanic eruptions (Thomason et al., 1997; McCormack et al., 1997; Vernier et al., 2011). The strongest global ozone anomaly was observed just after the volcanic eruption of Mount Pinatubo in 1991 with a decrease of about 3-4% over a 2-year period following the eruption (Randel et al., 1995); larger decreases in column ozone, of order 5-10%, were observed locally in northern hemisphere middle and high latitudes (Bojkov et al., 1993; Randel et al., 1995; Coffey et al., 1996; Zerefos et al., 1997; Robock et al., 2000). Studies of ground-based ozone records extending over three decades have indicated the existence of a decadal variation in column ozone that is approximately in phase with the solar cycle (Angell et al., 1988; Zerefos et al., 1997; Austin et al., 2008). This is supported by analyses of global

satellite ozone records since 1979 showing evidence for a decadal oscillation of column ozone with maximum amplitude (~2-4%) at low latitudes (Chandra et al., 1994; McCormack et al., 1997; Hood et al., 1997; WMO, 2007; Randel et al., 2007). As noted by Solomon et al. (1996), the occurrence of two major volcanic eruptions nine years apart during each of the last two declining phases in solar activity could lead to some confusion in separating volcanic and solar effects on ozone. This should not strongly influence trend estimates for the long time records, but may have implications for isolation of the solar cycle in short observational records. Chandra et al. (1991) also showed the possible importance of the phase of the QBO in the ozone response to solar variability.

The effects of forcings on the stratospheric variability of other chemical species have also been studied. Although the main focus was on the trends, other components of the variability were also analysed in some of these observational studies. For example, variations in NO_2 column have already been decomposed and attributed to a range of forcings (QBO, ENSO, aerosols and solar activity) (Zawodny et al., 1991; Liley et al., 2000; Struthers et al., 2004; Gruzdev et al., 2008; Cook and Roscoe, 2009; Dirksen et al., 2011).

The links between stratospheric chemical composition and forcings have also been investigated using coupled chemistry-climate models (CCMs) (e.g. WMO, 2003). Obviously, the ability of a model to reproduce theses links depends on its ability to simulate correctly the processes involved in the relationship between forcing and response. For example, the effects of solar variability on stratospheric composition involve photochemical processes such as the photolysis of molecular oxygen, radiative heating and the associated dynamical response (Wohltmann et al., 2007; Gray et al., 2010). Also, the effects of varying stratospheric aerosol loading involve heterogeneous chemical processes and radiative heating (Robock et al., 2000; Wohltmann et al., 2007).

Within the framework of the international CCMVal/SPARC (Chemistry-Climate Model Validation activity/Stratospheric Processes And their Role in Climate) programme, state-of-the-art CCMs forced by natural and anthropogenic external forcings have been used to simulate past changes in stratospheric chemical composition. Ozone CCM simulations have been evaluated against a range of satellite observations with respect to climatologies (i.e. zonal mean distribution) and long-term trends (Eyring et al., 2006; CCMVal, 2010). Model-calculated

column ozone responses to external forcing were also evaluated using multi linear regression analysis (Austin et al., 2008; CCMVal, 2010;). The evaluation was mostly carried out over large latitudinal bands representative of the tropics, mid-latitudes and polar regions. Regarding the QBO, models with forced or internally generated QBO were generally able to reproduce the latitudinal variations of the QBO signal in column ozone. Nonetheless, CCMs forced with observed QBO tended to overestimate the amplitude of the column ozone response. Some of the models with internally generated QBO had problems reproducing the periodicity of the QBO and hence of the QBO-induced ozone signal (CCMVal, 2010). For the ENSO, the column ozone signal was comparable in most CCMs but could not be assessed against observations because of the large interannual variability and the weakness of the ozone column signal in observations (CCMVal, 2010). Regarding the aerosol loading, CCMs showed a considerable spread in their response to volcanic eruptions although most models were forced by a common dataset for the time-varying stratospheric aerosol surface area density (SAD) distribution. The models displayed differing degrees of sensitivity to aerosol levels, leading to different ozone losses. None of the models reproduced the observed hemispheric asymmetry in post-Pinatubo ozone losses, for either full hemispheric means or for mid-latitudes (CCMVal, 2010). Finally, regarding solar variability, the column ozone signal was reasonably well represented in models. Most CCMs reproduced 70-80% of the observed solar signal in global column ozone (averaged between 60°S and 60°N) (CCMVal, 2010). However, the vertical structure of the tropical solar signal in ozone did not seem to be correctly reproduced in models (Austin et al., 2008).

The purpose of the present study is to assess how well CCMs are able to reproduce the effects of specific forcings on stratospheric chemical composition at NDAAC (Network for the Detection of Atmospheric Composition Change, http://www.ndsc.ncep.noaa.gov/) measurement sites. The model performance in simulating the links between forcings and stratospheric composition is critical for the level of confidence in future model projections. Although the CCMVal programme (Eyring et al., 2006, CCMVal, 2010) did assess the performances of the models with respect to a range of chemical species, the evaluation was carried out with zonal mean data and the focus was on long-term trends and on the effects of external forcing on stratospheric ozone. Here, instead of using composite satellite data and a zonal mean representation, the evaluation uses long time series of quality-controlled ground-based observations from the NDACC. An evaluation of zonal means only is too limited because the concentration fluctuations along

latitude circles (i.e. deviations from the zonal mean) may play a very significant role in the variability of gas concentrations above certain sites. In addition, our evaluation of stratospheric composition response to forcings does not focus solely on ozone, as done previously, but also covers the species relevant to the ozone chemistry (e.g. nitrogen oxides, chlorine reservoirs and stratospheric source gases). The major limitation in NDACC data is the spatial sampling which is extremely sparse compared to satellite data. On the other hand, the NDACC database offers long time series of homogeneous data, a relatively global distribution of sites (with tens of stations covering over the globe) and simultaneous measurements of whole range of chemical species. To quantify the links between forcings (e.g. aerosol loading, solar irradiance, stratospheric halogen loading or a linear trend term, QBO, ENSO) and variations in chemical composition, we perform a multi linear regression (MLR) analysis. Both CCM and observational data records are processed in exactly the same way. The response of a chemical species to an external forcing can involve complex and multi-step mechanisms and might depend on responses of other variables, for instance temperature. However, the aim here is not to explore how well models perform with respect to intermediate steps in the forcing-response links but simply to assess the overall links in model simulations and their consistency with the links derived from observations. Therefore, the only explanatory variables in the MLR are external forcings and not model variables such as atmospheric temperature, strength of the general circulation or heat flux. Another consequence is that most of the internally generated variability in stratospheric chemical composition is not expected to be accounted for in our MLR. Finally, it is also worth pointing out that, to our best knowledge, this study is the first attempt to analyse simultaneously large NDACC multi-species datasets in the evaluation of CCMs.

The paper is divided into 5 sections. Section 2 describes the NDACC observations and the CCMs. The methodology and forcing proxies are presented in Section 3. The results are described and discussed in Section 4. First, seasonal cycles are studied, followed by an analysis of the variance and of the MLR results. Section 5 is devoted to summary and concluding remarks.

2 Observations and model simulations

2.1 Observations

We evaluate CCM responses to external forcings using different ground-based observational datasets without zonal averaging. The NDACC network is composed of monitoring and research stations providing high-quality observations of atmospheric chemical composition which are well distributed over the globe. NDACC instruments are designed to provide consistent, standardized, long-term measurements of atmospheric trace gases, particles, UV radiation reaching the Earth's surface, and physical parameters. NDACC data are also used as reference data in the calibration of satellite data. Here we use O_3 and NO_2 columns measured with Systeme d'Analyse par Observations Zenithales (SAOZ) instruments (Pommereau et al., 1988) and HCl, $ClONO_2$, CH_4, O_3, N_2O and HNO_3 columns from Fourier Transform Infra-Red (FT-IR) spectrometers (Zander et al., 2008). The quality of SAOZ data has been evaluated during several instrument intercomparisons (Hofmann et al., 1994; Roscoe et al., 1999; Hendrick et al., 2011). The FT-IR measurements have also been validated by many instrument intercomparisons and by comparison with satellite observations (Lambert et al., 2003; Griffith et al., 2003; Hauchecorne et al., 2005; Dils et al., 2006; Griesfeller et al., 2006; Vigouroux et al., 2008). The periods of observation vary for different species and stations, and are summarized in Table 1. For example, the longest time series used is SAOZ measurement of NO_2 and O_3 columns during 17 years at Dumont d'Urville, whereas the shortest time series used is FT-IR measurement of column CH_4 during 2 years at Wollongong. On average, the length of the time series is 10 years. As the periods considered and the length of the time series can have an impact on our results, the model-calculated and observational data are analysed in the same way over the same periods as the observations (including data gaps) in order to make the results from models and observations as comparable as possible. The comparisons are carried out on monthly means. In contrast to other species, due to a strong diurnal cycle, the SAOZ sunset and sunrise NO_2 data are not averaged together, but treated as two distinct datasets; like the other species, their data are averaged monthly. As no sunset or sunrise NO_2 data over specific sites are provided in the CCMVal model database, adjusted sunset/sunrise CCM values are calculated using the time-varying global daily meanNO_2 fields from CCMs and time series of monthly mean ratios of sunset and sunrise NO_2

over daily mean NO_2 for each site provided by the SLIMCAT chemical transport model SLIMCAT (Chipperfield, 2006).

2.2 Model simulations

The primary data used in this study are CCM output from the SPARC CCMVal-2 programme (Eyring et al., 2008; Morgenstern et al., 2010). The CCMVal-2 models are state-of-art chemistry-climate models designed to describe all the required stratospheric photochemistry, radiative and dynamical processes and their interactions. Since CCMVal-2 models and their forcings are already described elsewhere (Eyring et al., 2008; Morgenstern et al., 2010), only some aspects are briefly summarised here. Based on the availability of model output required for the analysis, seven models are considered here. The list of models and their references are given in Table 2, along with their horizontal and vertical resolutions, the pressure of top level, and some main parameterizations. Each individual model simulation is analysed but the discussion is rather centred on the multi-model mean (MMM) because we are interested here in systematic model biases. For the needs of specific analyses, models are sometimes divided into 3 subgroups according to the QBO description: models that internally generate the QBO, models with an externally forced QBO by nudging of equatorial stratospheric zonal winds towards observed wind profiles, and models that do not simulate a QBO. The type of simulation analysed here is the so-called REF-B1 reference simulation that covers the 1960-2006 period (Eyring et al., 2008). The CCMVal-2 REF-B1 simulations generally include all known anthropogenic and natural forcings. The temporal evolution of the surface mixing ratios of greenhouse gases (CO_2, CH_4 and N_2O) in the models follows the scenario A1B of the Intergovernmental Panel on Climate Change (McCarthy et al., 2001). The evolution of halogen Ozone-Depleting Substances (ODS) (i.e. CFCs, HCFCs and halons) surface mixing ratios follows the 2007 WMO standard scenario (WMO, 2007). The sea surface temperature (SSTs) are prescribed as monthly means following the global sea ice and sea surface temperature (HADISST1) dataset provided by the UK Met Office Hadley Center (Rayner et al., 2006) except for the CCM LMDz-reprobus that uses AMIP II sea surface data (Taylor et al., 2000). The influence of the solar variability on photolysis and heating rates is parameterized according to the intensity of the 10.7 cm solar radiations (Lean et al., 2005). All models except ULAQ have implemented a spectrally resolved solar forcing in the simulation, both for radiation and for photolysis calculations. The three major volcanic eruptions

(Agung in 1963, El Chichon in 1982 and Mount Pinatubo in 1991) are taken into account by prescribing observed global distributions of aerosol surface area densities (SADs). Chapter 2 of WMO (2011) summarizes the various ways in which the set-up of some models may deviate from the REF-B1 scenario (Eyring et al., 2008). The CCM outputs uploaded from the CCMVal database are monthly mean fields. The analysis is focused on T3M model outputs that correspond to the 3-D monthly mean fields. The modelled values at the different observation sites are obtained by simple linear interpolation of T3M fields. Few results for T2M outputs (zonally averaged monthly mean fields) are also presented for illustrating the importance of zonal deviations. Some modelling groups did not provide T3M outputs for all the species considered here (see Table 2). As a result, instead of 7 models, only 4 models are analysed for HCl and 3 models for NO_2. In the figures, model results are shown for the CCM and the corresponding individual model simulations. The model dispersion (i.e. model-to-model-variations) is simply calculated as standard deviations on the MMM calculation, and MMM inter-annual variations are calculated from the mean of the individual model variances (i.e. the mean of individual model inter-annual variability). It is worth keeping in mind that the model simulations are sampled in exactly the way as the observations in order to consider a model time series of the same length with the same gaps as in the observational time series.

3 Multiple regression model

Multiple regression models have long been standard tools in atmospheric science. For the stratosphere, they have been used to quantify the amount of ozone losses of anthropogenic origin (e.g. WMO, 1999; 2007; 2011). There have been several studies exploring the use of multiple regression models in analysing column ozone time series (Hood and McCormack, 1992; Bodeker et al., 2001; Reinsel et al., 2002; Randel et al. 2007). A “standard” MLR model for ozone monthly-mean time series would include explanatory variables such as QBO, solar cycle and halogen loading. In the last years, the number of explanatory variable (i.e. forcing proxies) has tended to increase to include model prognostic variables such as temperature in order for the regression model to account for as much as possible of temporal variations in column ozone, called the explained or dependent variable. The idea is to have a picture of ozone variability as complete as possible in order to identify the physical mechanisms driving this variability and

estimate trends of anthropogenic origin accurately, albeit within the limitations of linear regression approaches. However, as the focus of our study is the externally driven variability, we limit the external variables of the MLR to well-known external forcings such as solar cycle, aerosol and halogen loadings, ENSO, QBO. As a result, a large part of the interannually generated variability, at least the internal component, cannot in any way be accounted for by our MLR analysis. Note that, in some studies of polar ozone variability, the MLR also includes atmospheric parameters such as temperature or heat fluxes into the stratosphere as explanatory variables (Svendby et al., 2004; Brunner et al., 2006; Dhomse et al, 2006; Wohltmann et al., 2007; Harris et al., 2008). In that way, some internal variability can be accounted for in the MLR. However, the use of both external forcings and of atmospheric variables that are partly driven by both external forcings and internal variability as explanatory variables makes it very difficult to disentangle forced and internal variability. For this reason, we do not use model prognostic variables as forcing proxies in the MLR.

In the MLR approach, the column variations are assumed to be described as a sum of independent contributions following a linear model of the form:

$$Y(t) = \mu_i + \alpha_0 + \alpha_1 T(t) + \alpha_2 A(t) + \alpha_3 S(t) + \alpha_4 QBO(t) + \alpha_5 QBO'(t) + \alpha_6 ENSO(t) + \epsilon(t) \quad (1)$$

Where Y(t) is the stratospheric column monthly mean at time t (in months), μ_i is the monthly average of the i^{th} month of the year (i = 1, 2, ..., 12) calculated over all the years of the time series. α_0 to α_6 are the model coefficients calculated using a standard linear least squares regression. The coefficient α_0 is constant term. $\epsilon(t)$ are the residuals. The regression equation includes a so-called trend term T(t) that depends on the species considered. T(t) is constructed using EESC (Equivalent Effective Stratospheric Chlorine) time series (Chapter 5 of WMO (2011)) for ozone, inorganic chlorine loading time series for HCl and $ClONO_2$ or simply time for N_2O, CH_4, HNO_3 and NO_2 as trend explanatory variables. The formulation of EESC, a proxy for the combined total inorganic chlorine (Cly) and bromine (Bry) loading, is taken from WMO (2011) so that it accounts for the age-of-air spectrum and age-of-air dependent fractional release values (Newman et al., 2007). Note that EESC peaked in 1997 at mid-latitudes and 2002 at high latitudes (WMO, 2011). A(t) is the global mean surface area density (SAD) column calculated using the 1979-2004 monthly time series of zonal mean aerosol SAD which has been constructed within the CCMVal project from SAGE (Stratospheric Aerosol and Gas Experiment) I, SAGE II,

SAM (Stratospheric Aerosol Measurement) II, and SME (Solar Mesophere Explorer) instruments (Thomason and Peter, 2006). S(t) is the solar flux at 10.7 cm and is a common proxy for UV irradiance changes

(ftp://ftp.ngdc.noaa.gov/STP/SOLAR_DATA/SOLAR_RADIO/FLUX/DAILYPLT.OBS).
QBO(t) is the equatorial zonal wind at 30 hPa and the QBO'(t) term has been constructed normal to the QBO proxy QBO(t) by duplicating and shifting it by one month increments until the time integral of QBO'(t)×QBO(t) is close to zero. The use of two orthogonal QBO proxy terms allows a phase lag between the evolutions of the dependent variable and of the QBO (Austin et al., 2008). Note that the temporal shift used to construct the two orthogonal QBO proxies and estimate QBO'(t) is such in some models (e.g. LMDz-REPROBUS) that the last 2 years have to be discarded. Since all the modelled time series have to be of same length for calculating the MMM, the analysis of the time series stops at 2004 instead of 2006. In the case of model simulations, QBO indexes, QBO(t) and QBO'(t), are calculated from the modelled winds. ENSO(t) is the southern oscillation index for the El Niño-Southern Oscillation (ENSO) mode of climate variability taken as the global SST ENSO index. It is defined as the difference between the mean sea surface temperatures anomaly equatorward of 20° latitude (north and south) and the mean sea surface temperature poleward of 20° and it captures the low-frequency part of the ENSO phenomenon (http://jisao.washington.edu/data/globalsstenso/).

Uncertainties (σ_n) in regression coefficients, α_n (with n being the index of a specific forcing in Eq. (1), are calculated with a bootstrapping method (Efron and Tibshirani, 1994; Bodeker and Kremser, 2015). In our approach, we start by calculating a first set of regression coefficients by least squares fitting of the observational or model time series with the function of equation (1). The residuals (i.e. differences between the series and the fit) are then randomly added to the fit in order to create new artificial time series. In order to make sure that the results do not depend on the number of the artificial time series, 10,000 new time series are created. Finally, mean values and standard deviations can then be calculated from these 10,000 sets of regression coefficients.

In order to check that residuals are normally distributed with zero mean and constant standard deviation, adequacy chi-squared tests are performed with the null hypothesis being that there is no significant difference between the residuals' distribution and a normal distribution with zero mean and constant standard deviation. The significance level of the test is set to 5%. The

residuals from the regression model calculated by bootstrap are normally distributed with zero mean and constant standard deviation for all the cases at the 5% significance level. In the exploratory phase of this work, the time series were also analysed without bootstrapping. A first-order autogressive model AR1 was applied in order to take into account autocorrelation. The data were regressed against the model after a Cochrane-Orcutt transformation; the method is described in appendix A of Tiao et al. (1990) (see also Bodeker and Kremser, 2015). The Cochrane-Orcutt transformation does not change significantly the values of the regression coefficients but, as expected, the associated errors increase. Overall, this approach was providing satisfactory results in most cases. However, it was found not to be robust for short time series when errors could be vastly underestimated (Tiao et al., 1990). As a result, we tested the bootstrapping approach; we found that it provided results (regression coefficients, associated errors) results that were close to the results obtained with the Cochrane-Orcutt approach for most cases. However, the results were much more robust and reasonable for short time series. In order to test further the effect of autocorrelation in the time series, we tried to conserve some of the autocorrelation characteristics of the initial time series in the artificial time series generated by bootstrap by picking up and adding groups of successive residuals to the fit instead of picking up single residuals randomly. The length (i.e. number of successive residuals) of the groups was related to the autocorrelation length of the series. We tested the effect of varying lengths of the groups and found no significant difference in the regression results when the number of successive residuals per group was varied from 1 (no autocorrelation) to 10. The bootstrapping method appears to be able to generate residuals that are normally distributed with no significant autocorrelation (at the 95% confidence level). We have compared errors (i.e. standard deviations) in regression coefficients derived using this bootstrap method with standard MLR errors calculated using an autocorrelation model of type AR1 (Tiao et al., 1990). For our long time series, typically more than a decade, standard deviations derived with the bootstrap method are found to be in excellent agreement with MLR standard errors. However, for short time series, MLR standard errors are found to be unrealistically small in some cases because of the too-limited size of the sample whereas errors calculated with the bootstrap method remain reasonable, possibly because this Monte Carlo approach is better at estimating coefficient uncertainties using the residual distribution. This suggests that the bootstrap method can provide robust and reasonable estimations of errors on regression coefficients for the column time series.

The MLR approach relies on a number of assumptions that may not be valid in all cases. The main limitation is the correlation between forcings, implying that the contributions from forcings in equation Eq. (1) are not independent. For example, the aliasing between the solar and aerosol terms is related to the fact that the last 2 major eruptions occurred at the end of 11-year solar cycles. In addition, it is also possible that the relationship between the column abundance (explained variable) and a forcing is not linear. As a result, the relationship would not be properly represented in the MLR model or, at least, it would be approximated by a linear relationship. These possible deficiencies should be kept in mind when analysing the results. There is a final check to do once MLR analyses are performed. One needs to verify that a relationship exists to some degree between the explained variable (stratospheric chemical composition parameters) and the explanatory variables (external forcings). Here we use an F-test to test whether the likelihood of such a relationship at a 99% confidence level; the F test can be designed to test whether the hypothesis of an association between two variables can be due to chance (including here the null hypothesis of no relationship between explained and explanatory variables). The MLR analysis is carried out on each individual model result. Only means of all the individual model regression coefficients are presented in the figures along with error bars calculated as means of the individual model bootstrap errors.

4 Results

We assess the relative importance of different sources of variability in stratospheric chemical composition here. We first analyse the seasonal cycle, then the total versus deseasonalized variability and finally the forced variability estimated with the MLR method.

4.1 Annual cycle in stratospheric chemical composition

4.1.1 Annual cycle in ozone

Annual variations in stratospheric ozone are caused by seasonal variations in transport and photochemistry. As total ozone is largely determined by the lower stratospheric partial column, its annual cycle is mainly affected by the ozone evolution in this altitude range. Variations in transport, driven by dynamical processes, can affect ozone either directly or also indirectly through changes in the transport of ozone-depleting substances. Photochemical production of

ozone depends on annual variations in the solar irradiance. The resulting annual cycle in total column ozone is characterized by low amounts in tropics year-round, maxima in the spring of Northern Hemisphere (NH) high latitudes and Southern Hemisphere (SH) middle latitudes. Globally, the hemispheric mean total ozone is larger in the NH than in the SH. Figure 1 shows the mean seasonal cycle of total ozone measured at 6 stations grouped into three latitude bands (Figure 1a and 1b at 67°N, Figure 1c and 1d at 44°N and Figure 1e and 1f at 21-22°S). Grey areas around the NDACC mean seasonal cycle (black line) represent the inter-annual variability. Dashed red bold lines around the MMM (red thick line) represent the model dispersion (i.e. model-to-model variations) estimated as standard deviation in the MMM calculation. Naturally, NDACC inter-annual variations should not be compared to model dispersion on this Figure. These 6 stations have been selected because they correspond to pairs of stations located within very narrow latitude bands, allowing to illustrate the importance of zonal deviations. As the results are also representative of results obtained at other stations, there is no need to overload readers with many figures. The MMM reproduces the broad features of the observed annual cycle in column ozone in the northern hemisphere and in the tropics. Models also appear to be able to reproduce correctly the strong deviations from the zonal mean as revealed by NDACC observations. For example, consider stations at the same latitude band (OHP at 44°N-6°E and Moshiri 44°N-142°E or Sodankyla at 67°N-27°E and Zhigansk at 67°N-123°E) in Figure 1. The maximum observed at OHP is about 350 DU whereas it reaches 430 DU at Moshiri and the timings of the maxima and minima of the annual cycle are shifted at OHP compared to Moshiri. These differences between stations in the same latitude band also exist between Sodankyla and Zhigansk stations at 67°N and again models do reproduce them though they tend to underestimate the amplitude of these deviations from the zonal mean when compared to observations. As shown in Figures 1a and 1b, the interannual variability (characterised by the standard deviation) at NH high latitudes is larger during winter. This is due to the strong dynamical variability such as the rapidly changing position of the polar vortex brought about by the effect of wave activity that tends to deform the zonal flow. In contrast to high and middle latitudes, the seasonal cycles are very similar at the two tropical stations confirming that deviations from the zonal mean are small in the tropics (see Figure 1e and 1f). The large deviations from the zonal mean at middle and high latitudes impose the use of 3-D fields in

evaluating CCMs against site observations. Overall, the results confirm that models perform well for the ozone column.

4.1.2 Annual cycle in long-lived source gases N_2O

Information on the global transport in the stratosphere can be obtained from analyses of stratospheric chemical tracers, namely long-lived chemical constituents, whose spatial distributions and annual cycle are largely controlled by transport. Two such constituents are methane (CH_4) and nitrous oxide (N_2O). Generally, the amplitudes of the CH_4 and N_2O column annual cycles are rather small (see Figure 2) due to the large contribution from the troposphere. Nonetheless, minima are observed in fall and winter, especially at NH high latitude stations (for example Kiruna) followed by an increase in spring due to the cumulated effect of the downward transport inside polar vortex during the winter, and its breakup in spring with the mixing-in of mid-latitudes air. Models are generally better at simulting the seasonal cycle in N_2O column than the seasonal cycle in CH_4 column whose amplitude is underestimated in model simulations. There are no other salient model deficiencies in terms of seasonal variability apart from this too flat seasonal cycle in CH_4 column. Note that CCMs have a systematic bias in terms of absolute column abundance, they overestimate both columns by about 5%. It seems to be partly due to a small overestimation of tropospheric columns in models as pointed in the last SPARC CCMVal report (CCMVal, 2010), Indeed, in the REF-B1 scenario used by all CCMs, the same surface mixing ratios of CH_4 and N_2O are prescribed all over the globe, without taking into account the heterogeneities of the surface emissions distribution and hence of surface concentrations which are particularly significant for CH_4.

4.1.3 Annual cycle in NO_2

Nitrogen oxides are of primary importance in controlling stratospheric O_3 levels. We analyse NO_2 at sunset and sunrise, separately instead of daily means, due to the strong diurnal cycle of NO_2. Figure 3 shows examples of mean seasonal cycles of sunrise column NO_2 for 3 CCMs (CCSRNIES, CMAM, LMDZrepro) at six NDACC stations, grouped in 3 different latitudes bands (the same stations as in Figure 1). Results at sunset (not shown) are very similar. CCMs reproduce the general features of the seasonal cycles. As for ozone, observations indicate longitudinal differences between summer maxima at two stations on the same latitude band.

CCMs are able to reproduce these deviations from zonal mean. CCMs, or more precisely the MMM, appear to reproduce the amplitude of the NO_2 column maxima observed at Sodankyla, Moshiri and Bauru but overestimate it by about 10-20% at Zhigansk, OHP and Reunion. When individual model simulations are considered in Figure 3, one can see that the MMM overestimation at OHP and Reunion is mostly due to a very large systematic bias in one of the models (CCSRNIES).

4.1.4 Annual cycle in chlorine reservoir species

The accumulation of halogenated compounds in the stratosphere over the last half-century has been the primary driver of stratospheric ozone depletion. HCl is the two major stratospheric chlorine species. As a result, the seasonal cycle is expected to exhibit its largest amplitude at polar sites. Figure 4 shows the seasonal cycle of HCl column for 4 models (CCSRNIES, CMAM, LMDZrepro, SOCOL) at 7 stations located in NH and SH mid-latitudes. HCl is the dominant species in the chlorine family except in polar regions during the winter when chlorine activation on PSCs modify radically the chlorine partitioning. As shown in Figure 4, the general shape of the seasonal cycle of HCl column is well reproduced by CCMs. However, CCMs, or more precisely the MMM, tend to overestimate HCl columns. A careful analysis of Figure 4 shows that the CCM overestimation is mostly due to a large overestimation of about 15 to 50% (depending on the station) by one of the models (CCSRNIES). An analysis of T2M model outputs (zonally averaged fields) indicates that the sum of HCl and $ClONO_2$ columns, a good proxy for total chlorine, appears to be similar in all the models. It is not surprising because total chlorine levels are largely determined by ODS boundary conditions that are the same for all the models. However, according to the T2M outputs, this model vastly overestimates the HCl/$ClONO_2$ ratio compared to other models, pointing towards a major deficiency in its chlorine chemical scheme.

4.2 Total and interannual variability of the stratospheric chemical composition

While the total variance of most columns is dominated by the seasonal cycle, there is also a very substantial interannual variability in column time series. Our main aim here is to assess the ability of CCMs to reproduce the interannual variability, notably the externally forced component, observed at NDACC stations. We first start by comparing the variability in the observed and

MMM raw time series. We plot the standard deviations for the observed column time series as a function of the corresponding standard deviations from CCM simulations at different stations in Figures 5 to 8. Each figure corresponds to a chemical species and is composed of 3 model-versus-observation correlation plots where linear regression lines, 1-1 lines and correlation coefficient are also indicated. On each figure, the left-hand plot refers to the raw data time series (total variance) and the middle plot refers to the deseasonalized time series. In order to deseasonalize the time series, the mean seasonal cycle, defined as monthly means averaged over the corresponding months of the time series, is subtracted from the raw time series (Ziemke et al., 1997, Staehelin et al. 2001, Dhomse et al., 2006). The robustness of this deseasonalisation method has been tested by comparing the results with another widely applied deseasonalisation method using multiple harmonics as a functional for the seasonal cycle (Stolarski et al., 1991, Brunner et al., 2006, CCMVal, 2010). These two methods give very similar results. We favour the "monthly mean" method because χ^2 tests for distributional adequacy indicate that the distributions of the raw and deseasonalized time series are found to be distinct with slightly higher confidence levels (exceeding of 99.9%) in the case of the "monthly mean" deseasonalization than in the case of the "multiple harmonics" deseasonalization. The results for the deseasonalized time series provide a good estimate of the interannual variability component because, on intra-annual timescales, the overwhelmingly significant component is the seasonal one. Finally, the right-hand correlation plot in Figures 5 to 8 refers to the deseasonalized times series extracted from modelled zonal mean fields instead of modelled 3-D fields as in the middle plot. The idea is to see whether the interannual variability is fully resolved in zonal mean fields or whether the longitudinal dependency is significant. Note that the variability in zonal mean fields from CCMs has been extensively evaluated against zonally averaged satellite observations (CCMVal, 2010).

4.2.1 Ozone

Figure 5a shows a very good agreement between column O_3 total (i.e. raw data) standard deviations from observations and CCMs with values from the different stations lying around the 1-1 line (regression slope=1.07 ± 0.07) and a very high correlation coefficient between the modelled and observed time series (r=0.96). We can distinguish three groups of stations depending on latitude and amplitude of the standard deviation. The stations at high latitudes

exhibit a stronger variability than tropical stations except for the DDU (Antarctic) station where standard deviations are more typical of stations at mid-latitudes and, inversely, for the MOS (mid-latitude) station where standard deviations are more typical of high latitudes. After deseasonalisation (see Figure 5b), the standard deviations are strongly reduced at all stations for both observations and CCMs, reflecting the importance of the seasonal cycle contribution to the total ozone variability. Standard deviations after deseasonalisation remain grouped by latitudinal regions (tropics, mid-latitudes, high latitudes) indicating that the column ozone interannual variability generally increases with latitude. One can also note that, in contrast to the raw data, the standard deviation of the deseasonalized time series at MOS is of the same order of magnitude as at the other mid-latitude stations. Overall, after deseasonalization, column ozone variances in observations and models remain similar and relatively well correlated (r=0.88), even if the models tend to underestimate slightly the interannual variability (slope of 1.26 +/- 0.15), particularly at high latitudes. Modelled standard deviations for deseasonalized zonally averaged data (Figure 5c) are smaller than for the 3-D data (Figure 5b) and hence have stronger negative biases (regression slope=1.72 +/- 0.26) and a lower correlation (r=0.80) with the corresponding observational data. A very large fraction of the interannual variability in column ozone originates from interannual variations around zonal means, especially at mid- and high latitudes.

4.2.2 N_2O

The comparison between observed and modelled standard deviations for raw time series is worse for long-lived tracers such as CH_4 (not shown) and N_2O (Figure 6a) than for ozone. CCMs underestimate the total variance of the long-lived tracer columns by about 70% (regression slope=1.69). It is not possible to group the data points according to values of standard deviation and latitude. The effect of deseasonalisation on the N_2O column variance is small due to the very weak amplitude of the N_2O column seasonal cycle, especially in the CCMs (see section 4.1.2). After seasonal adjustment (see Figure 6b), the results show that CCMs generally underestimate the interannual variability in N_2O column. There is little difference between the results for zonal means (Figure 6c) and 3-D fields (Figure 6b). Therefore, the analysis of the long-lived tracers interannual variability can be restricted to the zonal component. Results for CH_4 are very similar to those for N_2O (not shown).

4.2.3 NO_2

There is a rather good agreement between modelled and observed standard deviations for the raw sunrise (sr) NO_2 column (Figure 7a) with a correlation coefficient of 0.97. As for ozone column, the amplitude of the total variance increases with latitude and allows us to group together stations from the same regions. In contrast to most other species analysed here, CCMs tend to overestimate the total variance of the NO_2 column, notably at high latitudes. This bias is much more pronounced in the case of the deseasonalized time series for both 3-D fields (Figure 7b) and zonal mean fields (Figure 7c). This particular behaviour at high latitudes is certainly related to the use of the SLIMCAT CTM NO_2sr/NO_2 ratio in estimating sunrise values from daily mean for the CCMs. The introduction of NO_2sr/NO_2 ratios calculated offline by a model forced by meteorological analyses into other models calculating their own dynamics and chemistry is bound to generate some spurious variability, particularly in polar regions where, due to the specific polar winter chemistry, NO_2 columns are low. For instance, the interannual variability in dynamics is strong at NH high latitudes with very variable winter conditions that partly drive the reactive nitrogen partitioning including the NO_2sr/NO_2 ratio. Even if CCMs are able to reproduce the magnitude of this interannual variability, their year-to-year variations are not expected to be in phase with the CTM interannual variations. As a result of this inconsistency, the derivation of NO_2sr in CCMs from the CTM NO_2sr/NO_2 ratio generates some unrealistic interannual variations in NO_2sr. The results for sunset NO_2 column are almost identical to those for the sunrise NO_2 column and hence are not shown.

4.2.4 HCl

The correlation coefficient between observed and modelled standard deviations for HCl column raw data is high (r=0.9) but the regression slope (1.56 +/- 0.30) shows that, on average, CCMs substantially underestimate the total variability (see Figure 8a). As found for the other species, the variability generally increases from tropics to polar regions with a peak at NH stations. The seasonal adjustment tends to reduce the variances but with a more pronounced impact on the modelled variability compared to the observational time series; the correlation coefficient decreases to reach 0.62 with a regression slope of 2.21 indicating a strong underestimation of the interannual variability by CCMs, especially at NH stations (see Figure 8b). The use of zonally

averaged CCM fields tends to degrade further the correlation with the observational time series (r=0.44) (see Figure 8c).

4.3 Interannual variability explained by external forcings

Results presented in Section 4.2 show that CCMs tend to underestimate the interannual variability seen in ground-based observations. The magnitude of the model bias varies with the species and latitude considered. It is usually most pronounced for long-lived species including total chlorine and at mid-latitude and polar stations. The only exception is NO_2, probably because of the spurious variability generated by the use of NO_2sr/NO_2 ratios from a CTM. The interannual variability can be decomposed in two main parts: externally forced variability and internal variability. The aim of this section is to assess the ability of CCMs to reproduce the different components of the interannual variability and, in particular, the contribution of individual external forcings to this variability. In order to estimate the forced variability and the different contributions of external forcing, we apply the MLR method described in detail in Section 3 with five explanatory variables (trend, aerosol, solar, QBO and ENSO) to the deseasonalized time series for both observations and CCMs at all stations and species. Figures 9 to 12 present the results. Panel (a) of each figure is a model-versus-observation correlation plot of the R^2 parameter of the MLR; R^2 represents the ratio of the variance accounted for by the regression model to the total variance (here, variance associated with interannual variability). The higher the value of R^2 is, the higher the fraction of variance explained by the MLR model is, with R^2=1 meaning that the regression model can explain all the variance. The other part of the variance, the non-explained fraction, is thought to represent mostly the internal variability. It is unlikely that some of this unexplained variance could be some forced variability driven by external factors that are not taken into account in the MLR because the keys external forcing for the stratosphere are already well identified. The fact that some hypotheses of the linear regression model (i.e. linear relationship between forcings and atmospheric composition responses, uncorrelated individual forcings,…) may not be entirely valid can also affect the value of the R^2 ratio. Panels (b) and (c) in Figures 9 to 12 are bar charts of the fraction of the variance explained by individual forcings in observations and in CCMs at different stations for different species; the contributions of the different explanatory variables to this fraction are indicated with different

colours (trend in yellow, solar in red, QBO in light blue, aerosol in dark blue, ENSO in green). To facilitate the reading, the contributions of individual external forcings to column variance are represented as absolute values but negative contributions are hatched in the bar charts.

4.3.1 Ozone

In the case of O_3 column (see Figure 9a), the highest fractions of explained interannual variability (i.e. highest R^2) for both observations (R^2 values ranging 50 to 80%) and CCMs (R^2 values ranging 35 to 68%) are found at the tropical stations. At the other stations, the fractions are fairly similar in observations and CCM time series with R^2 values around 20 to 40%. The R^2 correlation plot indicates a relatively high correlation coefficient (0.78) but with a regression slope of 1.4 suggesting that CCMs tend to underestimate slightly the fraction explained by external forcings. As shown by Figures 9b and 9c, the fraction of variability explained by the EESC forcing is dominant at DDU (SH polar regions) for both observations and CCMs due to the high dependency of the polar ozone loss to halogen loading. At the other stations, the EESC forcing represents only 10% of the interannual variability in CCMs which is in good agreement with observations. The solar forcing explains a substantial part of the variability at tropical stations, around 15%, except at REU where it reaches 30%. In the tropics, UV radiation is intense all year around and its fluctuations can strongly influence O3 photochemistry (Haigh et al., 1994). In addition, the solar forcing appears to be very significant at NH polar stations (about 15-20% at ZHI, SCO and NYA) where dynamical feedbacks are expected to play a role in driving the stratospheric response to solar forcing (Marchand et al., 2011). As expected, the QBO forcing explains a larger part (10-50%) of the interannual variability in tropics than in other regions (5-15%) in both observations and CCMs. The proportion of the interannual variability explained by the aerosol and ENSO forcings are relatively similar whatever the stations, except at JUN where the fraction explained by the aerosol forcing exceeds 40%. The time series of this station has the issue of starting in 1991 during the Mount Pinatubo eruption. For both observations and CCMs, the fraction of variance explained by the aerosol forcing strongly depends on the period covered by the time series, whether it is a background (non-volcanic) or volcanically active periods. The injection of sulphur from this volcanic eruption in June 1991 affected O_3 columns for a period of several years. The aerosol forcing also plays a very significant role at OHP because the time series starts just two years after the eruption, while, at stations like ZUG and MOS, aerosol

fractions are much smaller with the series beginning 4 years after the eruption (WMO, 1999). Models generally tend to underestimate the contribution of aerosol forcing compared to observations.

4.3.2 N_2O

In the case of tracer species like N_2O, the R^2 results suggest that, wherever the station, the interannual variability in CCMs is almost entirely explained by external forcings with R^2 values ranging from 75% to 96% (see Figure 10a). In contrast, R^2 for most observational time series ranges from 20 to 45% except in one station where it reaches 67% of the variance (see Figure 10a). The correlation between modelled and observed R^2 is very poor. Recall that CCMs generally underestimate the interannual variability in long-lived tracers (see Figure 6b). The results here show it is mostly due to the fact that they strongly underestimate the internal variability (i.e. externally unforced variability) for long-lived tracers As expected for N_2O column, the fraction of interannual variance explained by the trend forcing is the dominant contribution in CCMs whatever the stations. It is driven by the steady increase in N_2O emissions at the surface since early 20th century (see Figure 10b). The most salient differences in the trend contribution between CCMs and observations are observed at NH mid and high latitudes where almost all the variability is forced in CCMs whereas the forced variability is not the dominant in the observations. The solar contribution is almost negligible everywhere both in observations and CCMs except at mid-latitudes stations though it is still fairly low (see Figure 10b). The QBO contribution is generally higher in observations than in CCMs except at mid-latitudes stations. The aerosol and ENSO contributions are broadly similar in observations and CCMs whatever the stations. As the results are the same for methane, they are not shown here.

4.3.3 NO_2

In the case of sunrise NO_2 column, the R^2 parameter ranges from 20% to 80% in CCMs and from 20% to 70% in the observations (see Figure 11a); the correlation coefficient between modelled and observed R^2 is relatively low (0.61) with a regression slope of 0.78. CCMs tend to overestimate the contribution of external forcing to the total interannual variability compared to the observations, especially at NH stations, and it is generally due to a model overestimation of the NO_2 column response to the aerosol forcing (see Figure 11b and 11c). For most stations, the

dominant contributor to the forced interannual variability is the aerosol forcing, except at TAR station where the dominant contributor is the trend. Indeed, the sulphate aerosol influences NO_2 column due to well-known denoxification processes via heterogeneous chemistry. The large eruption of Mount Pinatubo in 1991 led to an increase of the aerosol loading by 2 orders of magnitude resulting in highly enhanced aerosol surface areas and hence in heterogeneous chemical processing in the stratosphere until at least the mid-1990s (Fish et al., 2000). The aerosol contribution varies with the period covered by the time series, notably covering the Mt Pinatubo period. The aerosol contribution is smaller at HAR, MAU, REU, BAU and KER station where time series start after 1991. It may be noted that the same results and behaviour have been found in the analysis of NO_2 column at sunset (not shown). The trend contribution represents 5 to 60% of the variance of the NO_2 column at SH stations with generally higher proportions in observations than in CCMs whereas the trend contribution reaches only 10% only at NH stations. The solar contribution is extremely low in both observations and CCMs. The QBO contribution is small and similar at most stations in both observations and CCMs and, as expected, is maximum at tropical stations.

4.3.4 HCl

For both HCl column, as found for most species, the proportion of interannual variability explained by external forcing (i.e. forced variability) is generally overestimated in CCMs compared to observations (see Figure 12); on average, modelled R^2 fractions are 0.12 higher for HCl, the most abundant chlorine species. Generally, the R^2 values are lower at high latitude stations compared to middle and low latitude stations both in observations and CCMs. The trend contribution originating from the chlorine loading change clearly dominates the variance, except at polar stations where the QBO contribution becomes dominant. The solar contribution is relatively small (around 5%) for all stations both in observations and CCMs. The solar contribution to HCl column variance is about 2 to 5% for CCMs and lower than 2% for observations.

4.4 Absolute contributions of external forcings to interannual variability

In this section, instead of looking at the relative contributions of individual forcing to the forced variability, we now analyse them in absolute terms by converting the fraction of variance explained by a specific forcing into a percentage change in a chemical species column per unit change of a forcing indicator. This allows us to compare directly all the forcing contributions whatever the amount of total (i.e. externally forced and unforced) variance in the time series or the amount of variance explained by the MLR model (i.e. forced variability). We attempt here to assess the ability of CCMs to reproduce the magnitude of the response of stratospheric chemical composition to given individual external forcings. Figures 13, 15 and 16 show the magnitudes of responses to the main external forcings at the different stations for O_3, N_2O, NO_2sr and HCl total columns. The sensitivity (i.e. response) to the external forcings is derived from the results of the MLR as described in Section 3 over the period covered by the observational time series. In these figures, the response is calculated by multiplying the regression coefficient corresponding to a specific external forcing by the typical amplitude of the forcing over the considered period. These values are then normalised by the average column of the chemical species over the period in order to express the so-called sensitivity (or response) to a specific forcing as a column percentage change per a typical variation in a specific external forcing (expressed in the forcing units).

4.4.1 Total ozone sensitivity

In the Figures 13a and 13b, the total ozone sensitivities are expressed as a percentage of total ozone change for 0.5 ppbv EESC (i.e. combined chlorine and bromine loading) change. Red bar charts indicate statistically significant results according to a Student test whereas grey bar charts indicate results that are not statistically significant. Errors are indicated with the vertical bars and are derived from the errors on regression coefficients that are estimated with a bootstrapping method (see section 3). CCMs reproduce generally well the latitude dependency of the total ozone column sensitivity derived from the observed ozone time series. The sensitivity is negative at middle (-5 to -10% per 0.5 ppbv EESC) and high latitudes (-10 to -15% per 0.5 ppbv EESC), peaking at the Antarctic station DDU. Trends in Arctic ozone are much more difficult to estimate due to the strong year-to-year variability in meteorological conditions. At Arctic stations, when

total ozone sensitivities from CCMs differ markedly from observed sensitivities, total ozone sensitivities are usually found to be not statistically significant (Student test) either in the observations or in the CCMs or even in both. Although CCMs reproduce the broad features seen in the observations at mid- and high latitudes, column ozone appears to be more sensitive to EESC in the observations than in the CCMs. In the tropics, both CCMs and observations show clear positive sensitivities at some stations, possibly related to solar activity. Indeed, tropical ozone column is quite sensitive to solar variability, notably the 11-solar cycle. For the period of interest (solar cycle 23), solar activity was weak around 2006 and high in 2001-2002 corresponding to relatively low tropical column ozone abundances in the mid-1990s and relatively high values at the beginning of the 2000s. Interestingly, as solar activity, EESC peaked in 2002 at high latitudes (WMO, 2011) making likely aliasing between EESC and solar forcing proxies during this period. While CCMs and observed ozone sensitivities agree at the tropical TAR station, the ozone column appears to be nearly twice as sensitive to EESC in observations than in CCMs at the MAU station. These results for tropical ozone appear to differ from the results obtained analysing much longer ozone time series (WMO, 2011) with negative tropical total ozone trends in CCMs (while EESC was increasing) and not statistically significant trend in observations. This confirms that the positive sensitivities found in the tropics in our analysis are likely to be artefacts of the MLR via aliasing between the EESC and solar forcing.

The 11-year solar cycle has a clear and direct impact on tropical ozone through radiative and chemical processes in the upper stratosphere and through less direct and more complex mechanisms involving dynamics, transport and/or chemistry throughout the stratosphere (e.g. see review by Gray et al. (2010)). The ozone response to solar variability in a CCM depends on the model representations of the relevant solar-driven processes (radiative transfer and photochemistry parametrisation) but also of dynamics representation in the tropical lower stratosphere and extra-tropical stratosphere. In Figures 13c and 13d, the sensitivity of total ozone column to a variation of 100 units solar flux show a clear positive statistically significant signature in tropics in both CCMs and observations. The tropical total ozone response to solar variations is around +1.5-4.5% per 100 solar flux unit in the observations and +1-4% in CCMs. The sensitivity to solar variations is not statistically significant at almost all the middle latitudes stations in both CCMs and observations. In polar regions, ozone sensitivities to solar variations

are of opposite sign in CCMs (negative sensitivity) and observations (positive sensitivity). Our results may be compared with other MLR results also obtained with ground based data and CCM simulations (Austin et al., 2008), although considering different time periods leads to differences in the retrieved solar signal; this point will be discussed later. Analysing both observations and CCMs, Austin et al. (2008) found an ozone sensitivity to solar forcing of about +1-2% of the annual mean column ozone per 100 F10.7 flux units in the tropics where the errors were the smallest and the response was statistically significant. These findings are consistent with our results. Away from the tropics, the errors were larger and differences between CCMs were also more important. In order to investigate the effect of the length of the time series and of the specific time period considered on the MLR results regarding the solar signal, we perform a sensitivity study with the long CCMs time series (about 30 years of simulations, from 1974 to 2006). The results are presented in Figure 14. The modelled column ozone sensitivity to solar forcing is plotted as a function of latitude for different time intervals over the 1976-2004 period. As expected, when the length of the time series is reduced, solar coefficients become less statistically significant and with larger errors, especially in the extra-tropical regions. When different periods and lengths are compared, we find large differences in the solar signals. For instance, the sensitivity to solar variations is relatively high for the 1990-1999 period and all the values are positive and increasing toward the poles. In contrast, the solar signal for the 1999-2004 period is found to be much weaker between 50°N-50°S than for the other periods and is negative and not statistically significant at high latitudes. The main reason for these variations in ozone column sensitivity is likely to be aliasing, the fact that forcing indicators that are supposed to be independent in MLR may be somewhat correlated over certain periods. For example, aliasing between the stratospheric aerosol loading and solar forcing over the recent decades has already been clearly established (Solomon et al., 1996). In addition, there can also be aliasing between the trend and solar forcing. For instance, when the periods considered covered only one or two 11-year solar cycles, as it is the case here, the fact of starting and ending the time series at different phases of the 11-year solar cycle (minimum versus maximum) may result into some of the ozone response to solar forcing appearing as a trend in the MLR. It is likely that only MLR analysis on time series of at least several decades can provide robust and reliable estimates of ozone column sensitivities to solar variations for monthly mean data, especially at high latitudes (see Figure 14). If daily data have been considered, a solar signal could have been extracted from

much shorter time series because of an important periodicity in the solar activity, the 27-day solar rotational cycle (Rozanov et al., 2006; Fioletov, 2009; Bossay et al., 2015).

Volcanic aerosols impact stratospheric chemical composition through heterogeneous chemistry and though interactions with incoming solar radiation that result in changes in actinic fluxes (photolysis rates of species), temperature and transport. Total ozone sensitivity to aerosol loading changes is plotted in Figure 13e and 13f. The results are not statistically significant at several stations. We focus the analysis on the limited number of stations where the results are statistically significant in both observations and CCMs (tropical REU, NH mid-latitude OHP and JUN, NH high-latitude SOD). At these stations, as expected, column ozone responses to an increase in surface aerosol density column of 50 $\mu m^2.cm^{-2}$ are negative in CCMs and range from -1 to -2%. These modelled values are in relatively good agreement with results for REU observations but are smaller than the values derived from observations at mid-latitudes stations (OHP, JUN).

Column ozone sensitivities to the QBO forcing for observations and CCMs are presented in Figure 15. Model results are shown on different plots according to the QBO characteristics of the CCMs. Figures 15a and 15b present the results for the multi-model mean (MMM) calculated from the mean of the results of all the CCMs. Figures 15c and 15d provide results for CCMs with internally generated QBO, Figures 15e and 15f for CCMs with a forced QBO and Figures 15g and 15h for CCMs without a QBO. Indeed, most CCMs have problems in simulating spontaneously a QBO, partly because of the difficulties in representing a realistic spectrum of upward propagating waves in the tropics. As expected, the ozone column response to the QBO for all CCMs shows a statistically significant maximum in the tropics and minimum in the subtropics. It is somewhat consistent with previous analyses of observations and CCM simulations within the framework of the WMO report (WMO, 2011) and CCMVal programme (CCMVal, 2010). Outside the tropics, in both hemispheres, the agreement between CCMs and observations is generally degraded and is accompanied with an increase in MLR errors and in the spread in CCM results (WMO, 2011). The differences between the results of the “all CCMs” case and observations rare the most pronounced at SH mid-latitude and at NH high latitude stations (Figures 15a and 15b). However, the agreement and the quality of the result are improved when only CCMs with an internally generated or forced QBO are considered (Figures 15c to 15f). The modelled maxima and minima seem also to be in better agreement with results from observations. The negative ozone response found at SH mid-latitudes for the observations is well reproduced

by CCMs with internally generated QBO (Figures 15c and 15d). As expected, column ozone responses in CCMs without QBO do not correlate with ozone responses in observations (Figures 15g and 15h).

4.4.2 Total N_2O, HCl, and NO_2 sensitivity

The trend forcing being the dominant contributor to the forced interannual variability for N_2O and HCl columns, their sensitivities to a trend forcing are shown in Figures 16a and 16b at different stations for CCMs and observations. Keep in mind that the trend forcing proxy T(t) is not the same for N_2O and HCl (see Section 3). T(t) is constructed using linearly increasing time for N_2O and, therefore, N_2O sensitivities are equivalent to time trends. As stratospheric HCl levels are expected to follow the tropospheric chlorine loading with a time lag due to stratospheric transport time scales, T(t) is simply inorganic chlorine loading time series.

Modelled and observational sensitivities for N_2O (about +3%) are in agreement at NH high latitude stations. In contrast, at middle latitudes, N2O sensitivities (statistically significant for both observations and CCMs) are lower in observations than CCMs. CCMs underestimate N_2O column trends at SH and tropical stations with values around 2%/decade while observed values are about 5-6\%/decade in the observations. HCl column sensitivities in observations and CCMs (Figure 16b) are somewhat consistent except at 2 stations (MOS and KIR). While the discrepancy at KIR station is probably due to the non-statistically significant results found in the observations, the negative trend in the observations at MOS which is in complete contradiction with the CCMs results is explained by a sharp rupture and shift found in the observational time series due to a change of instrument.

Regarding the NO_2sr column, Figure 16c and 16d show only its response to aerosol forcing for the different stations due to its dominant role in forced NO_2sr variability. There is a good agreement between statistically significant results from models and observations at almost all stations at middle and high latitudes in both hemispheres with a sensitivity of about -3 or -4% for an aerosol column increase of 50 $m^2.cm^{-2}$ (except at HAR and NYA where the results are found to be not statistically significant in the observations). Otherwise, at the other stations, in

particular, at the tropical stations, there is a lot of scatter in the results. The differences between results from CCMs and observations are large and often results are not statistically significant.

5 Summary and conclusions

The primary aim of the paper is to assess how well CCMs are able to reproduce the effects of key external forcings (QBO, ENSO, aerosol loading, solar irradiance and stratospheric halogen loading/trend) on stratospheric chemical composition (O_3, HCl, $ClONO_2$, N_2O, CH_4, HNO_3 and NO_2 columns) above specific measurement sites. The measurements sites are part of the global NDACC network and cover most of the main latitudinal regions (tropics, midlatitudes, polar regions). Because of strong deviations from the zonal mean at mid and high latitude sites, modelled chemical composition time series above measurements sites are reconstructed from 3-D CCM fields instead of the usual zonal mean fields. Observational and modelled times series are processed in exactly the same way in order to make the CCMs and observational results as comparable as possible. The relative importance of different sources of variability in stratospheric chemical composition is estimated.

First, we analysed seasonal cycles and variances of the raw and deseasonalized time series. Then, by means of multiple linear regressions on deseasonalized time series, the forced and interannual components of the variability were estimated. As seasonal variations vastly dominate the intra-annual variability, we consider that the variability left in the deseasonalized time series is the interannual variability. The contributions of individual external forcings to the forced interannual variability in modelled and observational time series are also compared.

CCMs are able to reproduce the broad features of the observed annual cycle in column ozone, including deviations from the zonal mean. Because of strong deviations from the zonal mean at mid and high latitude sites, modelled chemical composition time series above measurements sites are reconstructed from CCMs 3-D fields instead of the usual zonal mean fields. Total variances in observational and CCMs ozone column time series are very well correlated and exhibit a clear latitudinal dependency with increasing variability with latitude. The interannual variability (after seasonal adjustment) in total ozone follows the same behaviour but with reduced amplitudes reflecting the major contribution of the seasonal cycle to the total variability. Overall, CCMs tend to underestimate slightly the ozone column interannual variability. This model bias is much more

pronounced when the modelled fields are zonally averaged. The most important external forcings for the tropical interannual variability are the QBO and solar variability that are responsible for around 20-40% of the interannual variability at the extratropical measurement stations. CCMs generally reproduce correctly the latitude dependency of the total column ozone trend derived from NDACC observations. As expected, trends are found to be negative at mid and high latitudes with a peak at SH high latitudes. In the tropics, O_3 trends are found to be positive in both observations and CCMs with a varying degree of agreement depending on the station. A statistically significant solar signature is found in tropics in both CCMs and observations but the solar signature is generally underestimated in CCMs. Note that the solar component results are found to be strongly dependent on the length of the data series (Bossay et al., 2015). For both observations and CCMs, the amount of variance explained by the aerosol forcing is very variable and strongly depends on the period covered by the time series, whether it is background (non-volcanic) or volcanically active periods. The QBO forcing explains a larger part (10-50%) of the interannual variability in tropics than in other regions (5-15%) in both observations and CCMs.

Regarding long-lived species (chemical tracers), the seasonal cycle in N_2O column is better reproduced by CCMs than the seasonal cycle of CH_4 column which is found to be too flat in CCMs compared to observations. CCMs underestimate the total variability as well as the interannual variability of long-lived species. The interannual variability in CCMs is almost entirely explained by external forcings (mainly trend terms) while the forced variability represents only around 20-45% of the interannual variability in observations. This very significant underestimation of the internal variability of long-lived species suggests that the amount of internal variability in transport and hence stratospheric general circulation is not entirely reproduced in CCMs.

Generally, CCMs reproduce reasonably well the features of the seasonal cycle in total NO_2 column even if they tend to overestimate the summer maximum, especially in polar regions. In contrast to other species, the total variability of the NO2 column tends to be overestimated in CCMs, especially at high latitude, even after seasonal adjustment. It is certainly related to the use of NO_2sr/NO2 ratio from a CTM when reconstructing sunrise NO_2 (NO_2sr) column from CCMs NO_2 daily mean fields which generates some spurious variability. CCMs tend to overestimate the contributions of external forcing to the NO_2sr interannual variability, especially at high latitudes. It is mostly due to an overestimation of the NO_2 column response to the aerosol forcing. CCMs

underestimate the interannual variability of HCl column, especially in the NH. Again, the fraction of interannual variability explained by external forcings is generally overestimated in CCMs. It is largely due to an underestimation of the internal variability in CCMs. The results show a clear dominance of the trend term in the HCl column interannual variability.

Overall, the results of this study demonstrate that long time series of ground-based measurements are very useful in the evaluation of CCMs, notably the different components of the interannual variability of stratospheric composition. Although CCMs reproduce reasonably well seasonal cycles, they tend to underestimate very substantially the total interannual variability. They are able to simulate most of the externally forced variability and even specific responses to individual external forcings. However, the analysis of observational time series shows that the internally generated variability represents a very large fraction of the interannual variability and CCMs vastly underestimate this internal variability. As a result, the forcings appear to be responsible of much larger fractions of the total interannual variability in CCMs than in observations. This lack of internal variability in CCMs might partly originate from the surface forcing of CCMs by analysed SSTs which may be dampening the variability. It would be interesting to carry out similar simulations with CCMs coupled to ocean models and use the same methodology to analyse the different components of the interannual variability. The results show that some of the model evaluation within the framework of on-going joint IGAC/SPARC Chemistry-Climate Model Initiative (CCMI) programme (http://www.met.reading.ac.uk/ccmi/) can be carried out successfully using long time series of ground based observations. However, in order to take advantage of all the potential of NDACC data, CCM modelling groups should provide 3-D outputs (T3M) of more NDACC chemical species, notably $ClONO_2$ and HNO_3 in order to test the model chlorine and nitrogen partitioning.

Acknowledgements

This project was supported by the European project StratoClim (7^{th} framework programme, grant agreement 603557) and the grant 'SOLSPEC' from the Centre d'Etude Spatiale (CNES). The SLIMCAT modelling work was supported by the UK Natural Environment Research Council (MAPLE project NE/J008621/1).

References

Akiyoshi, H., L. B. Zhou, Y. Yamashita, K. Sakamoto, M. Yoshiki, T. Nagashima, M. Takahashi, J. Kurokawa, M. Takigawa, and T. Imamura, 2009. A CCM simulation of the breakup of the Antarctic polar vortex in the years 1980–2004 under the CCMVal scenarios. *J. Geophys. Res*., 114(D3).

Angell, J. K., 1988. Relation of Antarctic 100 mb temperature and total ozone to equatorial QBO, equatorial SST, and sunspot number, 1958-87. *Geophys. Res. Lett*, 15(8), 915-918

Austin, John, and R. John Wilson, 2006. Ensemble simulations of the decline and recovery of stratospheric ozone. *J. Geophys. Res*., 111,(D16).

Austin, J., K. Tourpali, E. Rozanov, H. Akiyoshi, Slimane Bekki, G. Bodeker, C. Brühl et al., 2008. Coupled chemistry climate model simulations of the solar cycle in ozone and temperature. *J. Geophys. Res*., 113,(D11).

Baldwin, M. P., L. J. Gray, T. J. Dunkerton, K. Hamilton, P. H. Haynes, W. J. Randel, J. R. Holton et al., 2001. The quasi-biennial oscillation. *Rev. Geophys*., 39(2), 179-229.

Bodeker, G. E., J. C. Scott, K. Kreher, and R. L. McKenzie, 2001. Global ozone trends in potential vorticity coordinates using TOMS and GOME intercompared against the Dobson network: 1978–1998. *J. Geophys. Res*., 106(D19), 23029-23042.

Bodeker, G.E., and Kremser S., Techniques for analyses of trends in GRUAN data, Atmos. Meas. Tech., 8, 1673–1684, 2015

Bojkov, Rumen D., Christos S. Zerefos, Dimitrios S. Balis, Ioannis C. Ziomas, and Alkiviadis F. Bais, 1993. Record low total ozone during northern winters of 1992 and 1993. *Geophys. Res. Lett*, 20(13), 1351-1354.

Bossay, Sebastien, Slimane Bekki, Marion Marchand, Virginie Poulain, and Ralf Toumi, 2015. Sensitivity of tropical stratospheric ozone to rotational UV variations estimated from UARS and Aura MLS observations during the declining phases of solar cycles 22 and 23. *Journal of Atmospheric and Solar-Terrestrial Physics*, 130, 96-111..

Bowman, Kenneth P., 1989. Global patterns of the quasi-biennial oscillation in total ozone. *J. Atmos. Sci.*, 46(21), 3328-3343.

Brasseur, Guy, and Susan Solomon, 2006. Aeronomy of the middle atmosphere: chemistry and physics of the stratosphere and mesosphere. Vol. 32. Springer.

Brunner, Dennis, J. Staehelin, J. A. Maeder, Ingo Wohltmann, and G. E. Bodeker, 2006. Variability and trends in total and vertically resolved stratospheric ozone based on the CATO ozone data set. *Atmos. Chem. Phys*., 6(12), 4985-5008.

CCMVal, S. P. A. R. C., V. Eyring, T. G. Shepherd, and D. W. Waugh, 2010. SPARC report on the evaluation of chemistry-climate models. SPARC report 5.

Chandra, S., 1991. The solar UV related changes in total ozone from a solar rotation to a solar cycle. *Geophys. Res. Lett*., 18(5), 837-840.

Chandra, S., and R. D. McPeters, 1994. The solar cycle variation of ozone in the stratosphere inferred from Nimbus 7 and NOAA 11 satellites. *J. Geophys. Res*., 99(D10), 20665-20671.

Chipperfield, M. P., 2006. New version of the TOMCAT/SLIMCAT off-line chemical transport model: Intercomparison of stratospheric tracer experiments. *Q. J. Roy. Meteor. Soc.*, 132(617), 1179-1203.

Coffey, M. T., 1996. Observations of the impact of volcanic activity on stratospheric chemistry. *J. Geophys. Res*., 101(D3), 6767-6780.

Cook, P. A., and Howard K. Roscoe, 2009. Variability and trends in stratospheric NO 2 in Antarctic summer, and implications for stratospheric NOy. *Atmos. Chem. Phys*., 9(11), 3601-3612.

Dhomse, S., M. Weber, Ingo Wohltmann, Markus Rex, and J. P. Burrows, 2006. On the possible causes of recent increases in northern hemispheric total ozone from a statistical analysis of satellite data from 1979 to 2003. *Atmos. Chem. Phys*., 6(5), 1165-1180.

Dhomse, Sandip, Mark Weber, and John Burrows, 2008. The relationship between tropospheric wave forcing and tropical lower stratospheric water vapour. *Atmos. Chem. Phys*., 8(3), 471-480.

Dils, Bart, M. De Mazière, J. F. Müller, Thomas Blumenstock, M. Buchwitz, R. de Beek, Philippe Demoulin et al., 2006. Comparisons between SCIAMACHY and ground-based FTIR data for total columns of CO, CH4, CO2 and N2O. *Atmos. Chem. Phys*., 6(7), 1953-1976.

Dirksen, Ruud J., K. Folkert Boersma, Henk J. Eskes, Dmitry V. Ionov, Eric J. Bucsela, Pieternel F. Levelt, and Hennie M. Kelder, 2011. Evaluation of stratospheric NO2 retrieved from the Ozone Monitoring Instrument: Intercomparison, diurnal cycle, and trending. *J. Geophys. Res*., 116(D8).

Efron, B., and Tibshirani, R. J., 1994. An introduction to the bootstrap. *CRC press.*

Eyring, Veronika, N. Butchart, D. W. Waugh, H. Akiyoshi, J. Austin, Slimane Bekki, G. E. Bodeker et al., 2006. Assessment of temperature, trace species, and ozone in chemistry-climate model simulations of the recent past. *J. Geophys. Res*., 111(D22).

Eyring, Veronika, Martyn P. Chipperfield, Marco A. Giorgetta, Douglas E. Kinnison, Elisa Manzini, Katja Matthes, Paul A. Newman, Steven Pawson, Theodore G. Shepherd, and Darryn W. Waugh, 2008. Overview of the new CCMVal reference and sensitivity simulations in support of upcoming ozone and climate assessments and the planned SPARC CCMVal report. *SPARC Newsl.* 30, 20-26.

Fioletov, V. E., 2009. Estimating the 27-day and 11-year solar cycle variations in tropical upper stratospheric ozone. *J. Geophys. Res*., 114(D2).

Fish, D. J., H. K. Roscoe, and P. V. Johnston, 2000. Possible causes of stratospheric NO2 trends observed at Lauder, New Zealand. *Geophys. Res. Lett.*, 27(20), 3313-3316.

Gray, Lesley J., J. Beer, M. Geller, J. D. Haigh, M. Lockwood, Katja Matthes, U. Cubasch et al., 2010. Solar influences on climate. *Rev. Geophys.*, 48(4).

Griesfeller, A., J. Griesfeller, F. Hase, I. Kramer, P. Loes, S. Mikuteit, U. Raffalski, T. Blumenstock, and H. Nakajima, 2006. Comparison of ILAS-II and ground-based FTIR measurements of O3, HNO3, N2O, and CH4 over Kiruna, Sweden. *J. Geophys. Res*., 111(D11).

Griffith, D. W. T., N. B. Jones, B. McNamara, C. Paton Walsh, W. Bell, and C. Bernardo, 2003. Intercomparison of NDSC ground-based solar FTIR measurements of atmospheric gases at Lauder, New Zealand. *J. Atmos. Oceanic Tech.*, 20(8), 1138-1153.

Gruzdev, A. N., 2008. Latitudinal dependence of variations in stratospheric NO2 content. *Izv. Atmos. Oceanic Phys.*, 44(3), 319-333.

Harris, N. RP, E. Kyrö, J. Staehelin, Dennis Brunner, S-B. Andersen, Sophie Godin-Beekmann, S. Dhomse et al., 2008. Ozone trends at northern mid-and high latitudes–a European perspective. *In Annales geophysicae: atmospheres, hydrospheres and space sciences*, Vol.26, No.5, p.1207.

Hauchecorne, A., J.-L. Bertaux, F. Dalaudier, C. Cot, J.-C. Lebrun, S. Bekki, et al., 2005. First simultaneous global climatologies of night-time stratospheric NO2 and NO3 observed by GOMOS/ENVISAT in 2003, *J. Geophys. Res.*, 110, No. D18, D18301, doi: 10.1029/2004JD005711.

Hendrick, F., J-P. Pommereau, Florence Goutail, R. D. Evans, Dmitry Ionov, Andrea Pazmino, E. Kyrö et al., 2011. NDACC/SAOZ UV-visible total ozone measurements: improved retrieval and comparison with correlative ground-based and satellite observations. *Atmos. Chem. Phys,*. 11(12), 5975-5995.

Hofmann, D. J., S. J. Oltmans, W. D. Komhyr, J. M. Harris, J. A. Lathrop, A. O. Langford, T. Deshler, B. J. Johnson, A. Torres, and W. A. Matthews, 1994. Ozone loss in the lower stratosphere over the United States in 1992–1993: Evidence for heterogeneous chemistry on the Pinatubo aerosol. *Geophys. Res. Lett.*, 21(1), 65-68.

Hood, Lon L., and John P. McCormack, 1992. Components of interannual ozone change based on Nimbus 7 TOMS data. *Geophys. Res. Lett.*, 19(23), 2309-2312.

Hood, L. L.,1997. The solar cycle variation of total ozone: Dynamical forcing in the lower stratosphere. *J. Geophys. Res*., 102(D1), 1355-1370.

Jourdain, Line, Slimane Bekki, François Lott, and Franck Lefèvre, 2008. The coupled chemistry-climate model LMDz-REPROBUS: description and evaluation of a transient simulation of the period 1980–1999. *In Annales Geophysicae*, vol. 26, no. 6, pp. 1391-1413. Copernicus GmbH, 2008.

Lamarque, Jean-François, D. E. Kinnison, P. G. Hess, and F. M. Vitt, 2208. Simulated lower stratospheric trends between 1970 and 2005: Identifying the role of climate and composition changes. *J. Geophys. Res*., 113(D12).

Lambert, J. C., J. Granville, M. Allaart, T. Blumenstock, T. Coosemans, M. De Mazière, U. Friess et al., 2002. Ground-based comparisons of early SCIAMACHY O3 and NO2 columns. In Proc. First ENVISAT Validation Workshop, ESA/ESRIN, Italy, 9-13 Dec.

Lean, Judith, Gary Rottman, Jerald Harder, and Greg Kopp, 2005. SORCE contributions to new understanding of global change and solar variability, In The Solar Radiation and Climate Experiment (SORCE). pp. 27-53. *Springer New York.*

Liley, J. B., P. V. Johnston, R. L. McKenzie, A. J. Thomas, and I. S. Boyd; 2000. Stratospheric NO2 variations from a long time series at Lauder, New Zealand. *J. Geophys. Res.*, 105(D9), 11633-11640.

Marchand, M., Keckhut, P., Lefebvre, S., Claud, C., Cugnet, D., Hauchecorne, A., Lefèvre, F., Lefebvre, M.P., Jumelet, J., Lott, F., Hourdin, F., Thuillier, G., Poulain, V., Bossay, S., Lemennais, P., David, C., Bekki, S., 2012. Dynamical amplification of the stratospheric solar response simulated with the Chemistry-Climate model LMDz-Reprobus. *J. Atmos. Sol.-Terr. Phys.*, 75-76, 147-160.

McCarthy, James J., ed. Climate change 2001: impacts, adaptation, and vulnerability: contribution of Working Group II to the third assessment report of the Intergovernmental Panel on Climate Change. Cambridge University Press, 2001.

McCormack, J. P., L. L. Hood, R. Nagatani, A. J. Miller, W. G. Planet, and R. D. McPeters, 1997. Approximate separation of volcanic and 11-year signals in the SBUV-SBUV/2 total ozone record over the 1979-1995 Period. *Geophys. Res. Lett.*, 24(22), 2729-2732.

Morgenstern, O., M. A. Giorgetta, K. Shibata, V. Eyring, D. W. Waugh, T. G. Shepherd, H. Akiyoshi et al., 2010. Review of the formulation of present-generation stratospheric chemistry-climate models and associated external forcings. *J. Geophys. Res.*, 115(D3).

Newman, P. A., J. S. Daniel, D. W. Waugh, and E. R. Nash, 2007. A new formulation of equivalent effective stratospheric chlorine (EESC), *Atmos. Chem. Phys*., 7(17), 4537-4552.

Pitari, G., E. Mancini, V. Rizi, and D. T. Shindell, 2002. Impact of future climate and emission changes on stratospheric aerosols and ozone. *J. Atmos. Sci.*, 59(3), 414-440.

Pommereau, Jean Pierre, and Florence Goutail, 1988. O3 and NO2 ground-based measurements by visible spectrometry during Arctic winter and spring 1988. *Geophys. Res. Lett.*, 15(8), 891-894.

Randel, William J., and Janel B. Cobb, 1994. Coherent variations of monthly mean total ozone and lower stratospheric temperature. *J. Geophys. Res*., 99(D3), 5433-5447.

Randel, William J., Fei Wu, J. M. Russell, J. W. Waters, and L. Froidevaux, 1995. Ozone and temperature changes in the stratosphere following the eruption of Mount Pinatubo. *J. Geophys. Res*., 100(D8), 16753-16764.

Randel, William J., and Fei Wu, 2007. A stratospheric ozone profile data set for 1979–2005: Variability, trends, and comparisons with column ozone data. *J. Geophys. Res*., 112(D6).

Randel, William J., and Anne M. Thompson, 2011. Interannual variability and trends in tropical ozone derived from SAGE II satellite data and SHADOZ ozonesondes. *J. Geophys. Res*., 116(D7).

Rayner, N. A., P. Brohan, D. E. Parker, C. K. Folland, J. J. Kennedy, M. Vanicek, T. J. Ansell, and S. F. B. Tett, 2006. Improved analyses of changes and uncertainties in sea surface temperature measured in situ since the mid-nineteenth century: The HadSST2 dataset. *J. Clim.*, 19(3), 446-469.

Reinsel, Gregory C., Elizabeth Weatherhead, George C. Tiao, Alvin J. Miller, Ronald M. Nagatani, Donald J. Wuebbles, and Lawrence E. Flynn, 2002. On detection of turnaround and recovery in trend for ozone. *J. Geophys. Res*., 107(D10), ACH-1.

Robock, Alan, 2000. Volcanic eruptions and climate. *Rev. Geophys.*, 38(2), 191-219.

Roscoe, H. K., P. V. Johnston, M. Van Roozendael, A. Richter, A. Sarkissian, J. Roscoe, K. E. Preston et al., 1999. Slant column measurements of O3 and NO2 during the NDSC intercomparison of zenith-sky UV-visible spectrometers in June 1996. *J. Atmos. Chem.*, 32(2), 281-314.

Rozanov, Eugene, T. Egorova,, W. Schmutz, & T. Peter, 2006. Simulation of the stratospheric ozone and temperature response to the solar irradiance variability during sun rotation cycle. *J. Atmos. Sol.-Terr. Phys.*, 68(18), 2203-2213.

Schraner, M., E. Rozanov, C. Schnadt Poberaj, P. Kenzelmann, A. M. Fischer, V. Zubov, B. P. Luo et al., 2008. Technical Note: Chemistry-climate model SOCOL: version 2.0 with improved transport and chemistry/microphysics schemes. *Atmos. Chem. Phys*., 8(19), 5957-5974.

Scinocca, J. F., N. A. McFarlane, M. Lazare, J. Li, and D. Plummer, 2008. Technical Note: The CCCma third generation AGCM and its extension into the middle atmosphere. *Atmos. Chem. Phys*., 8(23), 7055-7074.

Shiotani, M., 1992. Annual, quasi-biennial, and El Niño-Southern Oscillation (ENSO) time-scale variations in equatorial total ozone. *J. Geophys. Res*., 97(D7), 7625-7633.

Solomon, S., R. W. Portmann, R. R. Garcia, L. W. Thomason, L. R. Poole, and M. P. McCormick, 1996. The role of aerosol variations in anthropogenic ozone depletion at northern midlatitudes. *J. Geophys. Res*., 101(D3), 6713-6727.

Solomon, S., 1999. Stratospheric ozone depletion: A review of concepts and history. *Rev. Geophys.*, 37(3), 275-316.

Staehelin, J., N. R. P. Harris, C. Appenzeller, and J. Eberhard, 2001. Ozone trends: A review. *Rev. Geophys.*, 39(2), 231-290.

Steinbrecht, W., B. Haßler, C. Brühl, M. Dameris, M. A. Giorgetta, V. Grewe, E. Manzini et al., 2006. Interannual variation patterns of total ozone and lower stratospheric temperature in observations and model simulations. *Atmos. Chem. Phys*., 6(2), 349-374.

Stolarski, Richard S., Peter Bloomfield, Richard D. McPeters, and Jay R. Herman, 1991. Total ozone trends deduced from Nimbus 7 TOMS data. *Geophys. Res. Lett.*, 18(6), 1015-1018.

Struthers, H., K. Kreher, J. Austin, Robyn Schofield, G. Bodeker, P. Johnston, H. Shiona, and A. Thomas, 2004. Past and future simulations of NO2 from a coupled chemistry-climate model in comparison with observations. *Atmos. Chem. Phys*., 4(8), 2227-2239.

Svendby, T. M., and A. Dahlback, 2004. Statistical analysis of total ozone measurements in Oslo, Norway, 1978–1998. *J. Geophys. Res*., 109(D16).

Taylor KE, Williamson D, Zwiers F. 2000. 'The sea surface temperature and sea-ice concentration boundary conditions of AMIP II simulations'. Report 60, Program for Climate Model Diagnosis and Intercomparison. Lawrence Livermore National Laboratory: Livermore, California.

Thomason, L. W., L. R. Poole, and T. Deshler, 1997. A global climatology of stratospheric aerosol surface area density deduced from Stratospheric Aerosol and Gas Experiment II measurements: 1984–1994. *J. Geophys. Res*., 102(D7), 8967-8976.

Thomason, L., and Th Peter, 2006. SPARC assessment of stratospheric aerosol properties. WCRP-124, WMO/TD-No. 1295. No. 4. *SPARC Report.*

Tiao, G. C., G. C. Reinsel, Daming Xu, J. H. Pedrick, Xiaodong Zhu, A. J. Miller, J. J. DeLuisi, C. L. Mateer, and D. J. Wuebbles, 1990. Effects of autocorrelation and temporal sampling schemes on estimates of trend and spatial correlation. *J. Geophys. Res*., 95(D12), 20507-20517.

Vernier, J-P., Larry W. Thomason, J-P. Pommereau, Adam Bourassa, Jacques Pelon, Anne Garnier, Alain Hauchecorne et al., 2011. Major influence of tropical volcanic eruptions on the stratospheric aerosol layer during the last decade. *Geophys. Res. Lett.*, 38(12).

Vigouroux, C., M. De Mazière, Philippe Demoulin, Christian Servais, F. Hase, T. Blumenstock, I. Kramer et al., 2008. Evaluation of tropospheric and stratospheric ozone trends over Western Europe from ground-based FTIR network observations. *Atmos. Chem. Phys*., 8(23), 6865-6886.

WMO, 1999. Scientific assessment of ozone depletion: 1998, World Meteorological Organisation, Global Ozone Research and Monitoring Project–Report 44.

WMO, 2003. Scientific assessment of ozone depletion: 2002, World Meteorological Organisation, Global Ozone Research and Monitoring Project–Report 47.

WMO, 2007. Scientific assessment of ozone depletion: 2006, World Meteorological Organisation, Global Ozone Research and Monitoring Project–Report 50.

WMO, 2011. Scientific assessment of ozone depletion: 2010, World Meteorological Organisation, Global Ozone Research and Monitoring Project–Report 52.

Wohltmann, Ingo, Ralph Lehmann, Markus Rex, Dennis Brunner, and J. A. Mäder, 2007. A process-oriented regression model for column ozone. *J. Geophys. Res*., 112(D12).

Yang, Hu, and Ka Kit Tung, 1994. Statistical significance and pattern of extratropical QBO in column ozone. *Geophys. Res. Lett.*, 21(20), 2235-2238.

Zander, Rodolphe, Emmanuel Mahieu, Philippe Demoulin, Pierre Duchatelet, Ginette Roland, Christian Servais, M. De Mazière, Stefan Reimann, and Curtis P. Rinsland, 2008. Our changing atmosphere: Evidence based on long-term infrared solar observations at the Jungfraujoch since 1950. *Science of the Total Environment*, 391(2) 184-195.

Zawodny, Joseph M., and M. Patrick McCormick, 1991. Stratospheric Aerosol and Gas Experiment II measurements of the quasi-biennial oscillations in ozone and nitrogen dioxide. *J. Geophys. Res*., 96(D5), 9371-9377.

Zerefos, C. S., K. Tourpali, B. R. Bojkov, D. S. Balis, B. Rognerund, and I. S. A. Isaksen, 1997. Solar activity-total column ozone relationships: Observations and model studies with heterogeneous chemistry. *J. Geophys. Res*., 102(D1), 1561-1569.

Ziemke, J. R., S. Chandra, R. D. McPeters, and P. A. Newman, 1997. Dynamical proxies of column ozone with applications to global trend models. *J. Geophys. Res*., 102(D5), 6117-6129.

Figure 1: Seasonal cycles of ozone column (DU) at 6 stations. The black thick line corresponds to the NDACC mean seasonal cycle and its 1 standard deviations, indicated as grey areas, represent the inter-annual variability in the NDACC time series. The red thick line corresponds to the MMM seasonal cycle for 7 models (AMTRAC3, CAM3.5, CCSRNIES, CMAM, LMDZrepro, SOCOL, ULAQ); each red thin lines corresponding to the mean seasonal cycle of a single individual model simulation and the two dashed red bold lines represent the model dispersion (i.e. model-to-model variations) estimated as standard deviation in the MMM calculation. Naturally, NDACC standard deviations (i.e. inter-annual variations) cannot be compared to model dispersion shown here.

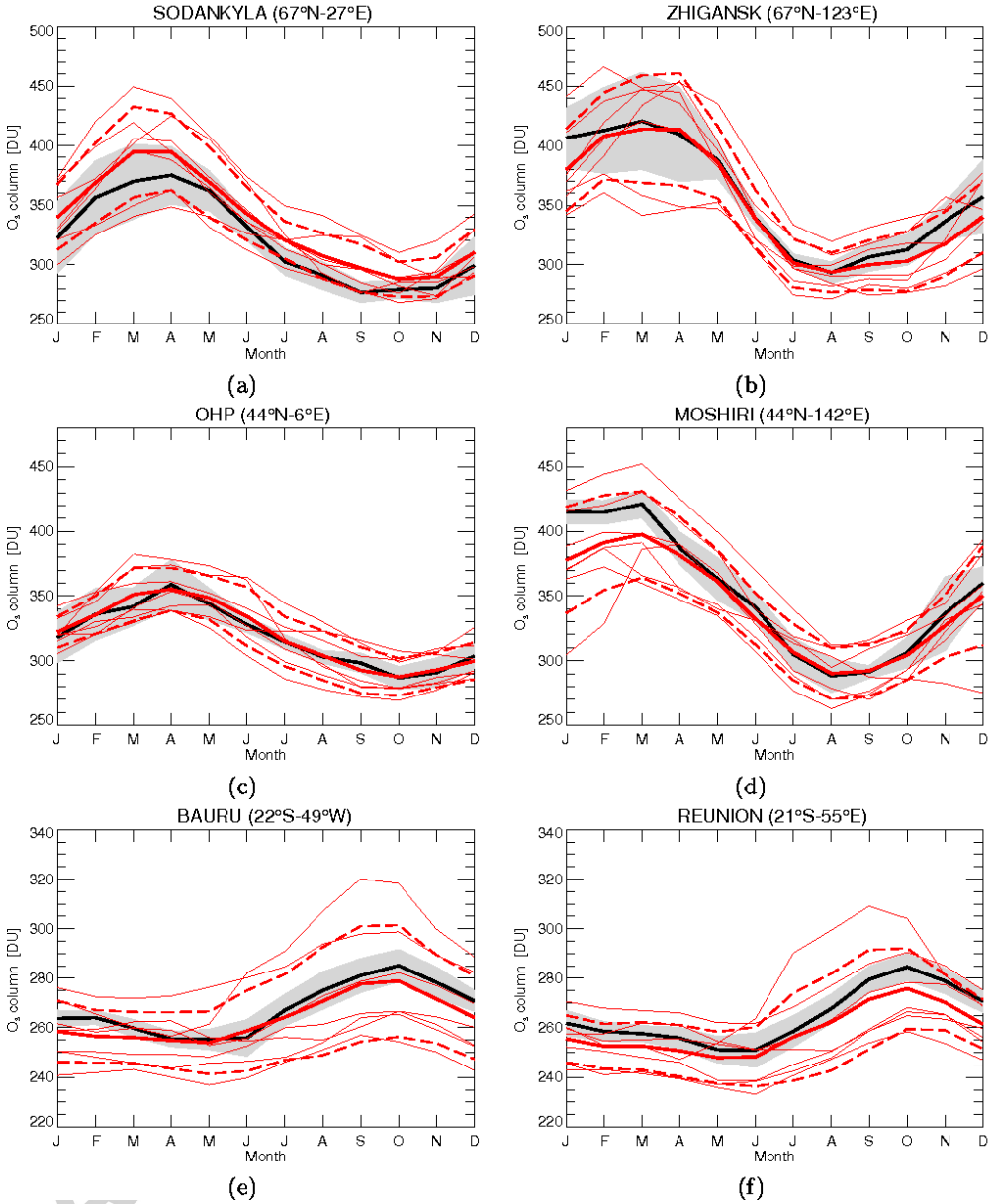


Figure 2: Same as figure 1 but for (left) CH_4 (10^{19} molec.cm^{-2}) and (right) N_2O column (10^{18} molec.cm^{-2}) at 4 stations for 5 models (CCSRNIES, CMAM, LMDZrepro, SOCOL, ULAQ).

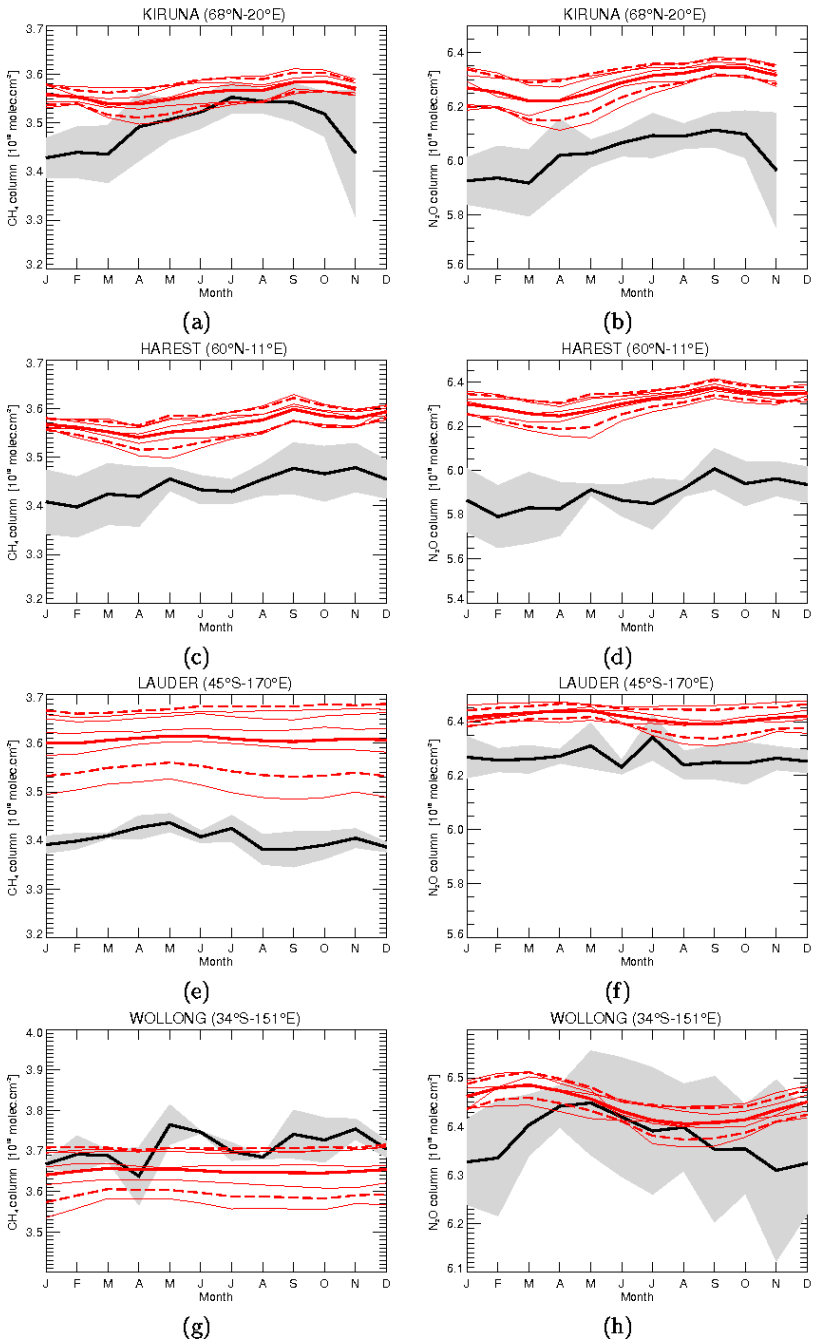


Figure 3: Same as figure 1 but for sunrise NO_2 (NO_2sr) column (10^{15} molec.cm^{-2}) at 6 stations for 3 models (CCSRNIES, CMAM, LMDZrepro).

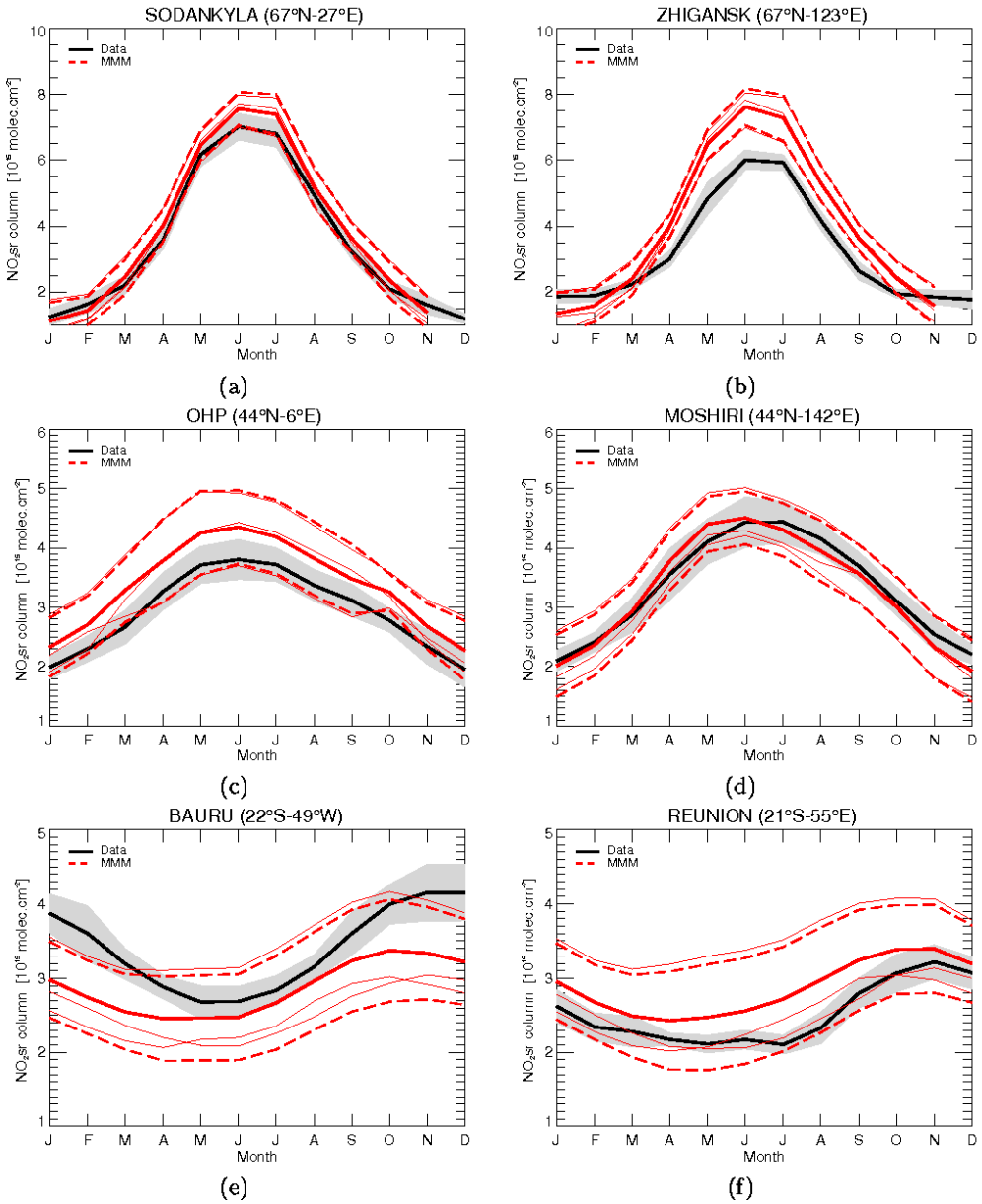
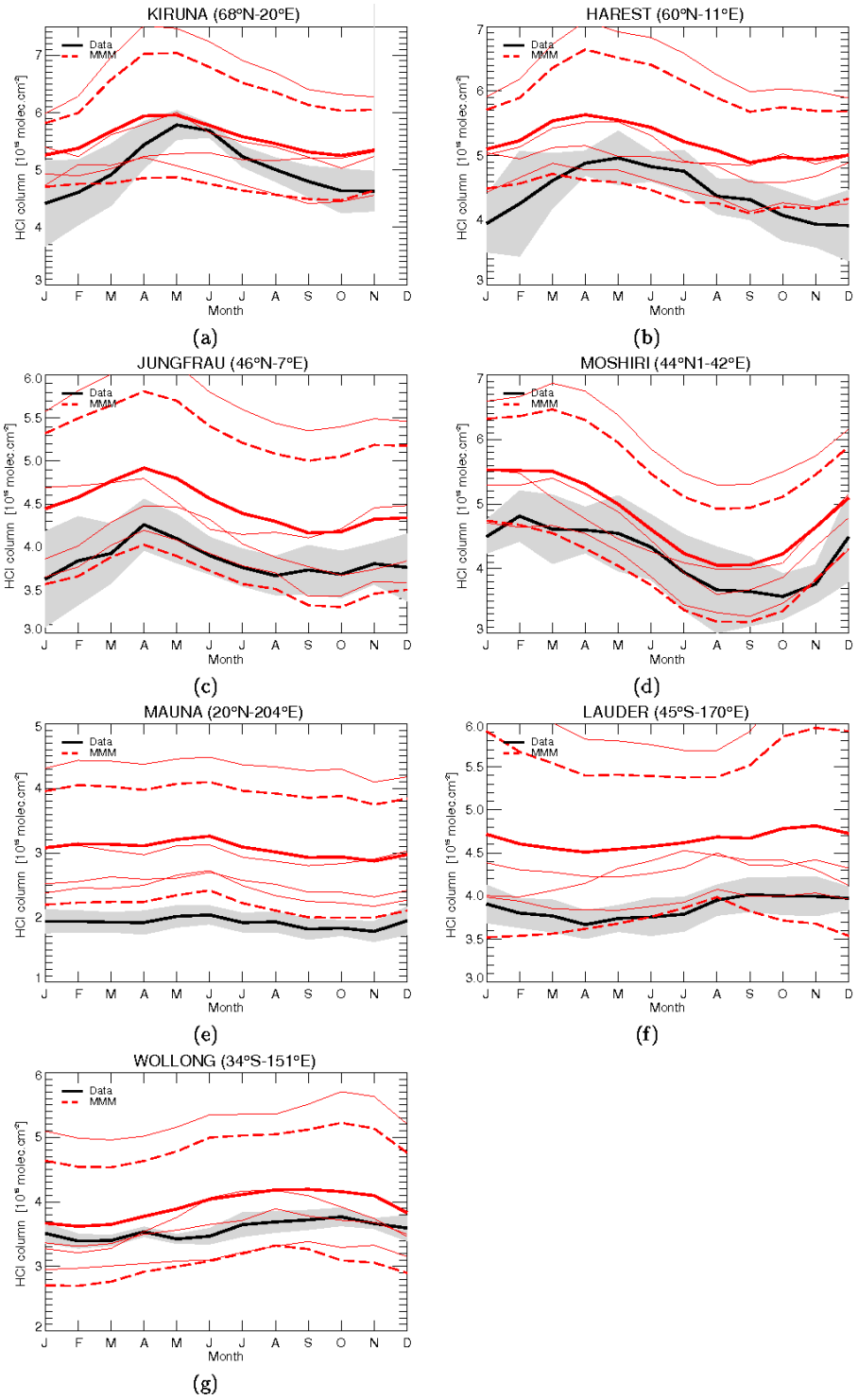
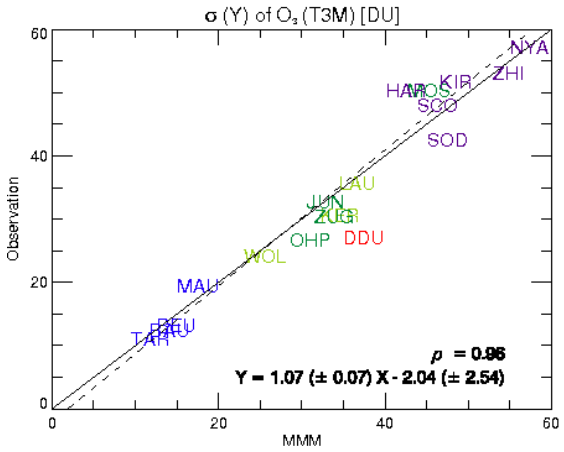
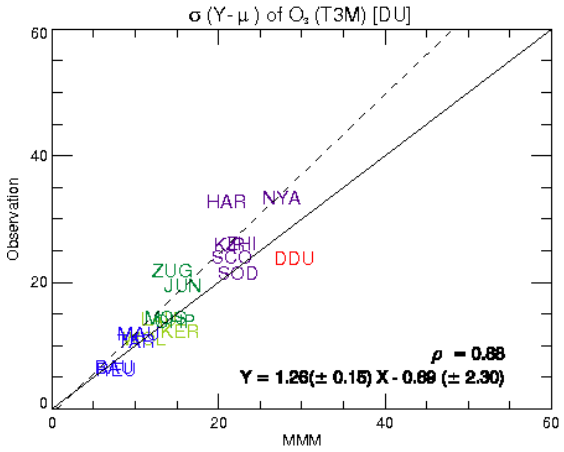


Figure 4: Same as figure 1 but for HCl column (10^{15} molec.cm^{-2}) at 7 stations for 4 models (CCSRNIES, CMAM, LMDZrepro, SOCOL).

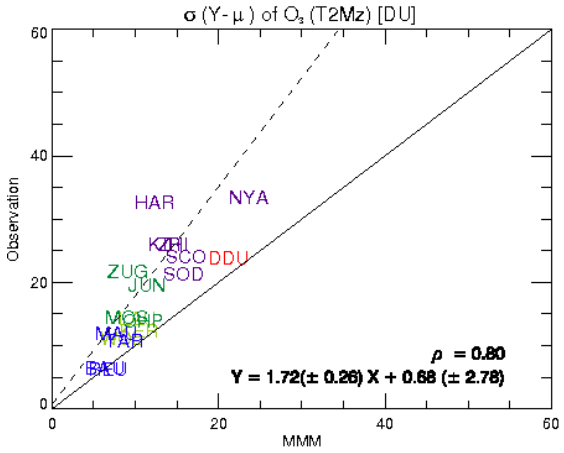




(a)



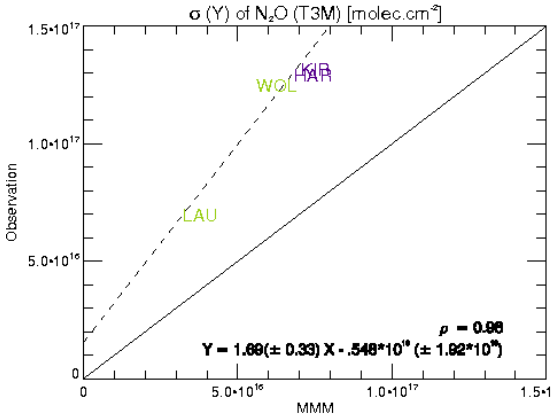
(b)



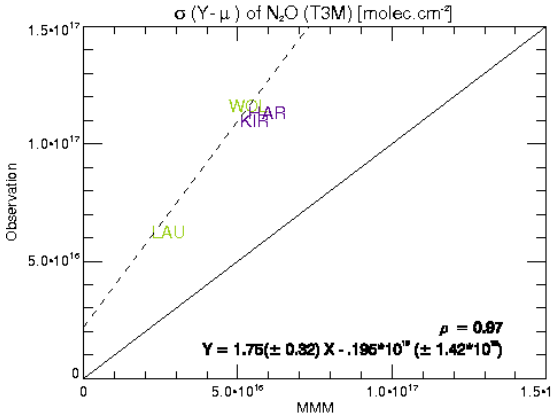
(c)

Figure 5: NDACC standard deviation as a function of MMM standard deviation for ozone column time series at 18 stations (see Table 1). NDACC and MMM standard deviations correspond to inter-annual variations. MMM standard deviations are derived from the average of the variances from the 7 individual model simulations and, as such, represent an estimation of the model inter-annual variability. (a) MMM standard deviations calculated from column time series taken from 3-D CCMs outputs, (b) MMM standard deviations calculated from deseasonalized column time series taken from 3-D CCMs outputs and (c) standard deviation from deseasonalized column time series taken from zonally averaged CCMs outputs. Colour coding is red for southern hemisphere (SH) high latitude stations, light green for SH middle latitude stations, deep green for northern hemisphere (NH) middle latitude stations, purple for NH high latitude stations and blue for tropical stations. In contrast to figures 1 to 4, both NDACC and MMM standard deviations represent inter-annual variations and hence can be compared directly.

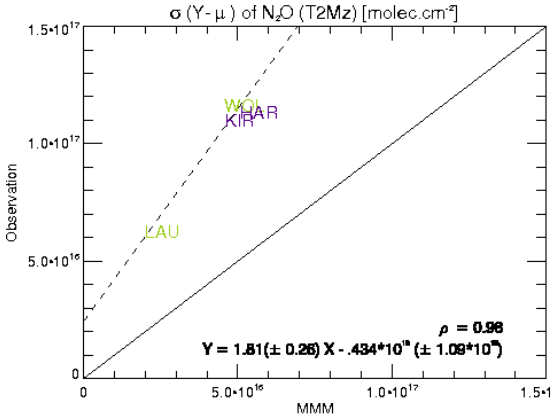
Figure 6: Same as figure 5 but for N_2O column at 4 stations.



(a)

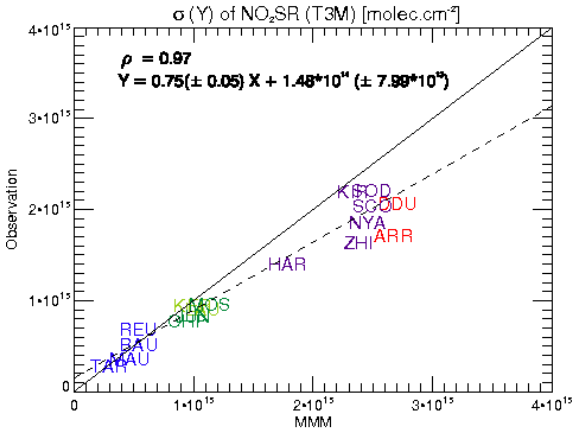


(b)

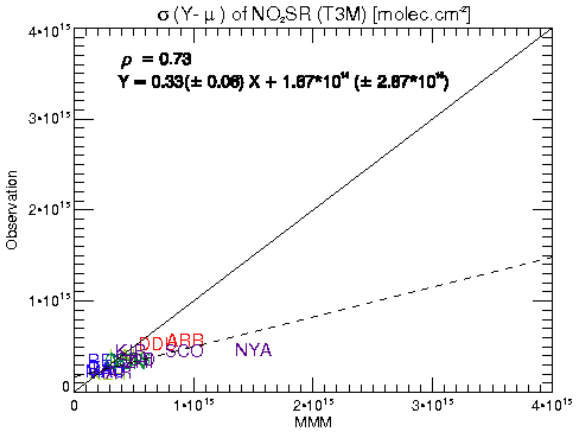


(c)

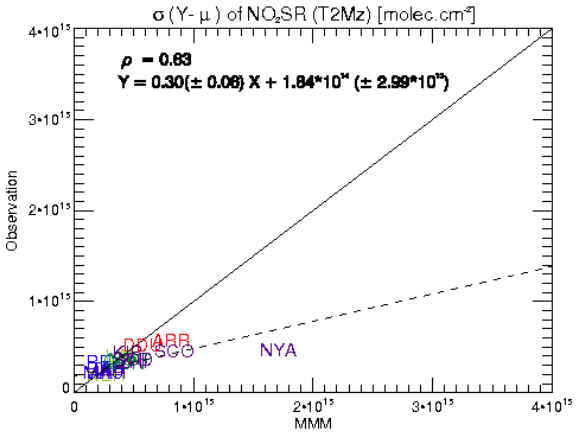
Figure 7: Same as figure 5 but for NO_2sr column at 17 stations.



(a)

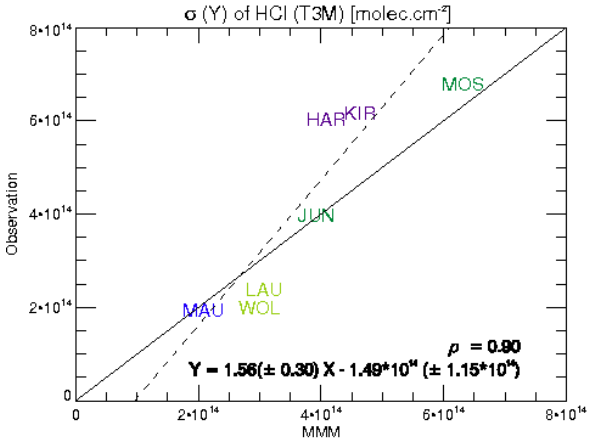


(b)

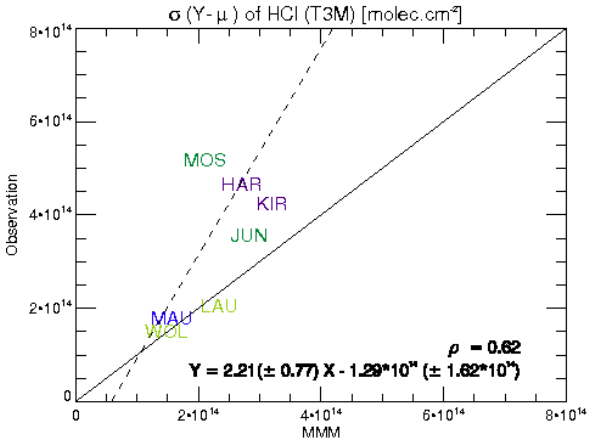


(c)

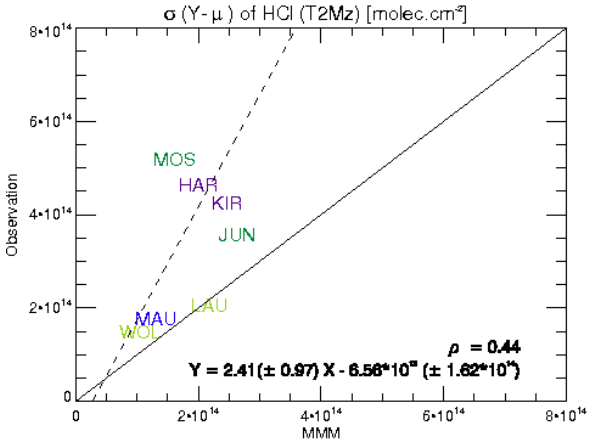
Figure 8: Same as figure 5 but for HCl column at 7 stations.



(a)

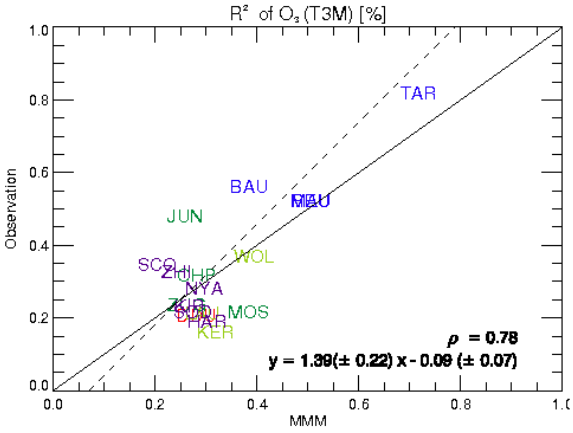


(b)

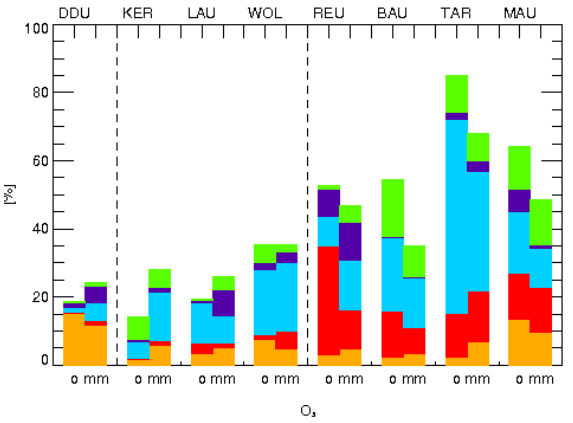


(c)

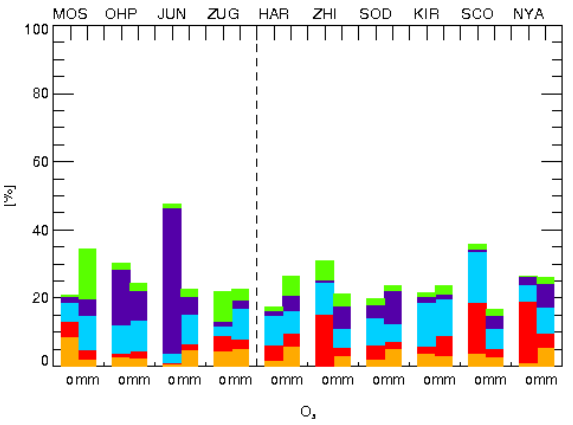
Figure 9: (top) R^2 from NDACC data as a function of R^2 from model calculations for ozone column time series at 18 stations; R^2 represents the ratio of the variance accounted for by the MLR model to the total variance. (middle, bottom) Contributions of individual external forcings to the ozone column total variance explained by the MLR model for different stations. Colour coding is yellow for trend forcing, red for solar forcing, light blue for QBO forcing, dark blue for aerosol loading forcing and green for ENSO. Filled areas correspond positive fractions of variance explained by explanatory variable and hatched areas correspond to negative fraction of variance explained by explanatory variable.



(a)

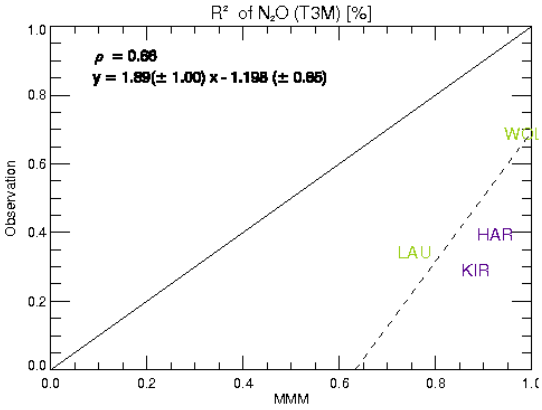


(b)

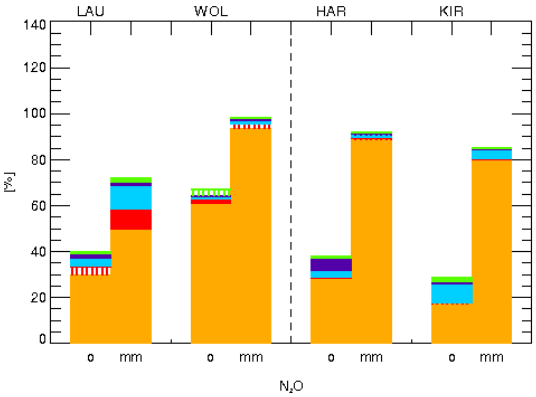


(c)

Figure 10: Same as figure 9 but for N_2O column at 4 stations.

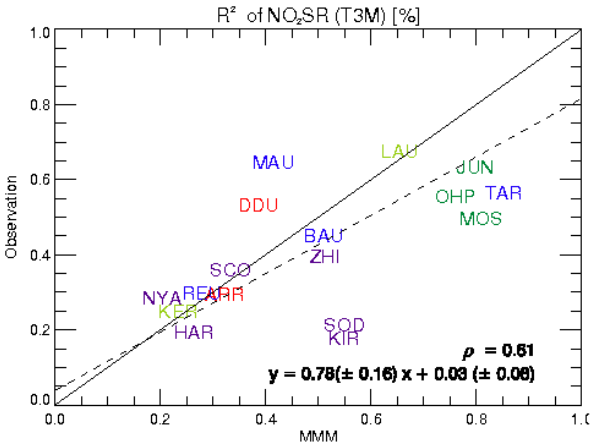


(a)

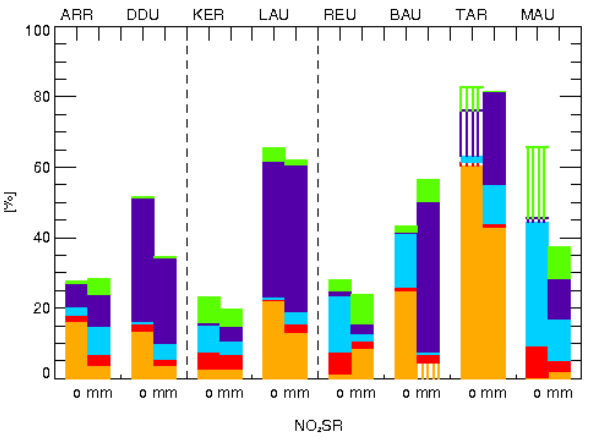


(b)

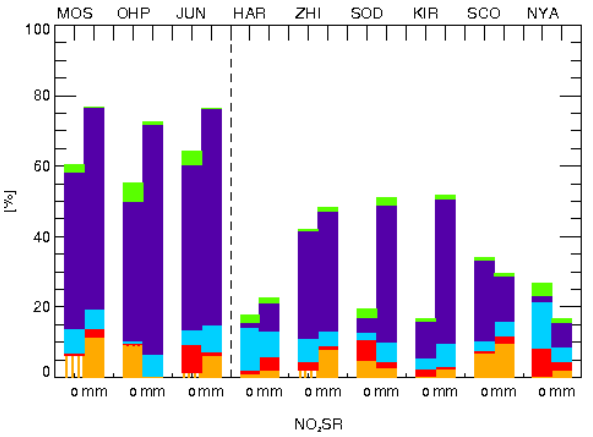
Figure 11: Same as figure 9 but for NO_2sr column at 17 stations.



(a)

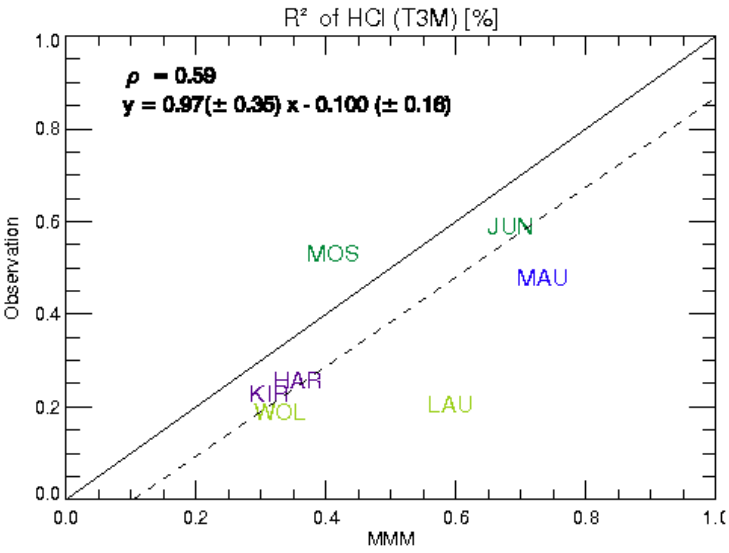


(b)

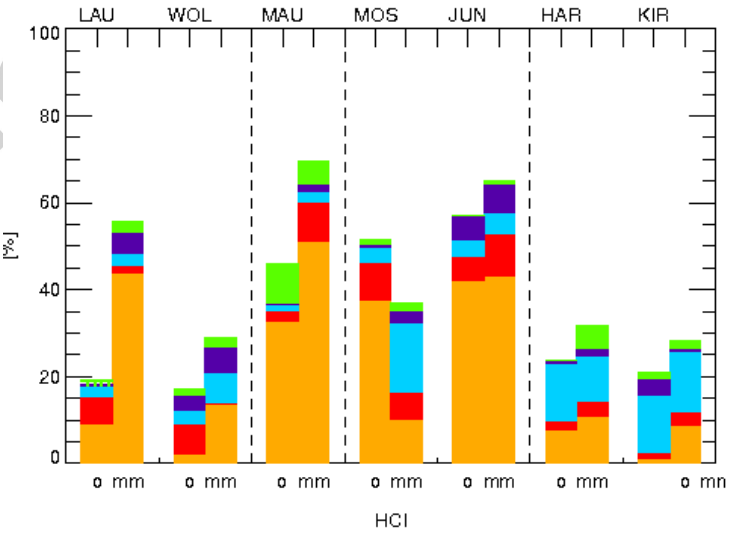


(c)

Figure 12: Same as figure 9 but for HCl column at 7 stations.

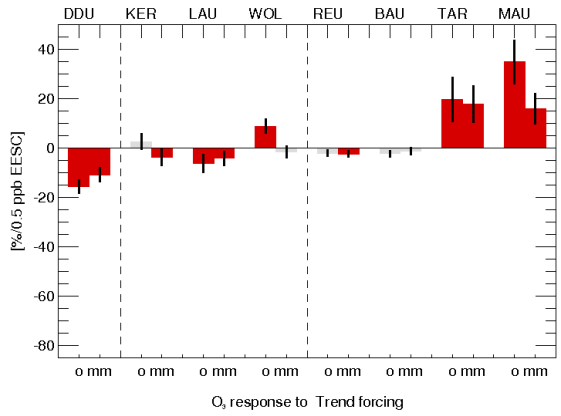


(a)

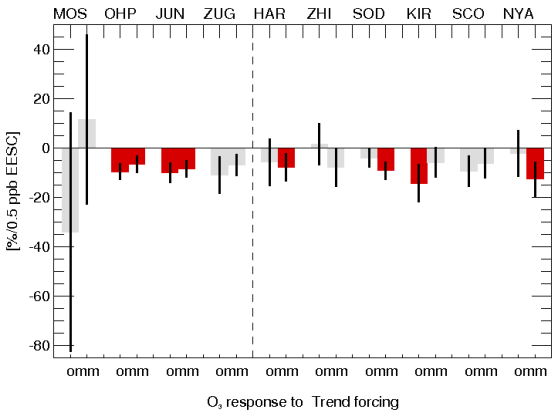


(b)

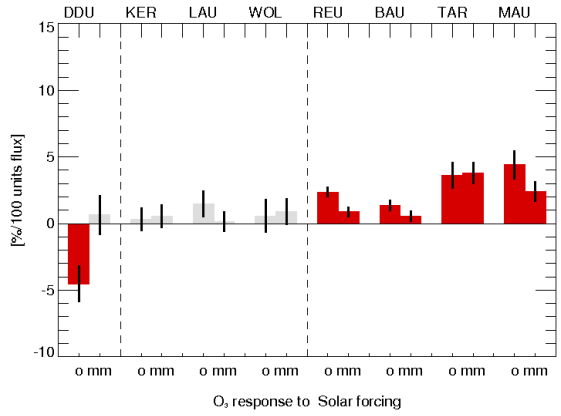
Figure 13: (a) and (b) Ozone column sensitivity to EESC (expressed as % change in ozone column per 0.5 ppbv change in EESC) at 18 stations for NDACC and MMM; (c) and (d) idem but for ozone column sensitivity to the solar forcing (expressed as % change in ozone column per change of 100 units of F10.7 flux); (e) and (f) idem but for ozone column sensitivity to the aerosol forcing (expressed as % change in ozone column per change of 50 $\mu m^2.cm^{-2}$ in surface aerosol density column). Observed (O) and MMM results are next to each other for each station. Red indicates statistically significant results whereas grey indicates results that are not statistically significant. Regression errors are indicated with thick black vertical bars. MMM sensitivities and errors are calculated as an average of the 7 individual model regression coefficients and of associated regression errors.



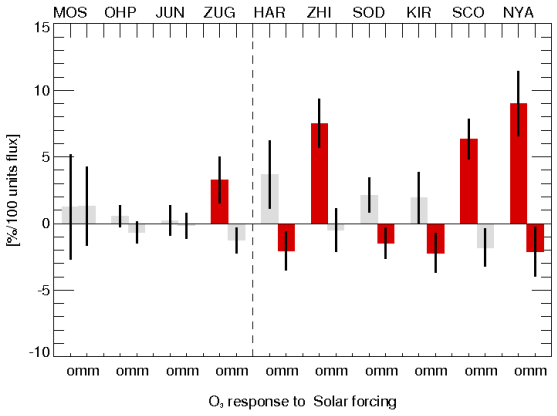
(a)



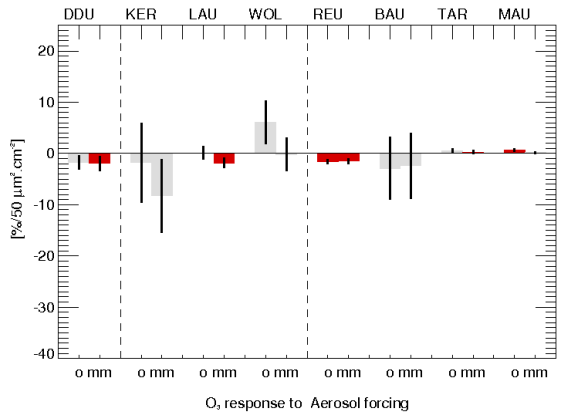
(b)



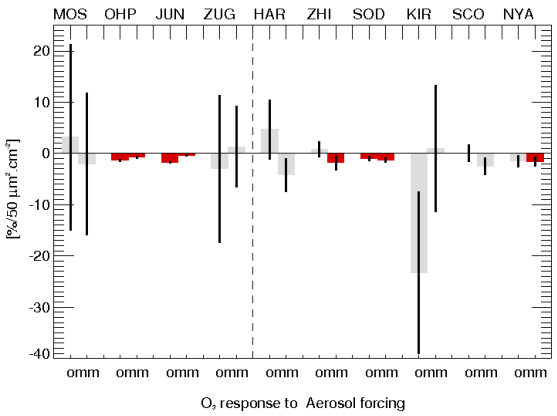
(c)



(d)



(e)



(f)

Figure 14: O_3 column sensitivity to solar forcing (expressed as O_3 column % change per 100 unit flux change) as a function of latitude for different periods. The periods of CCM simulations considered for MLR analysis are: 1976-2004 (black), 1990-2004 (blue), 1990-1999 (yellow), 1995-2004 (green), and 1999-2004 (red). Vertical bars denote the errors on regression coefficients. Circles indicate the statistically significant coefficients and the crosses are for coefficients that are not statistically significant.

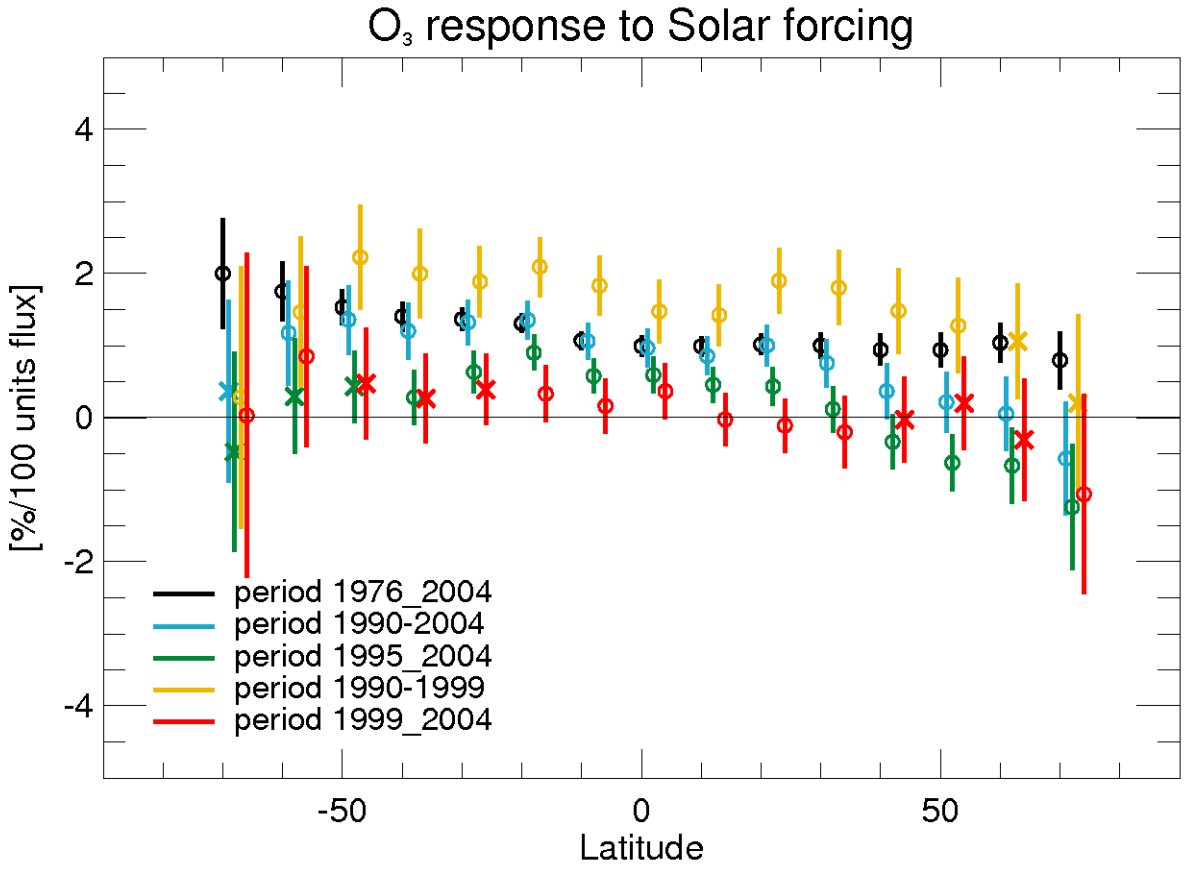


Figure 15: O_3 column sensitivity to the QBO forcing (expressed as O_3 column % change per $m.s^{-1}$ change in tropical zonal winds) for: (a) and (b) all the CCMs, (c) and (d) CCMs with internally generated QBO, (e) and (f) CCMs with a forced QBO and (g) and (h) CCMs without a QBO. Observed (O) and modelled (MM) results are next to each other for each station. Red indicates statistically significant results whereas grey indicates results that are not statistically significant. Errors are indicated with vertical bars.

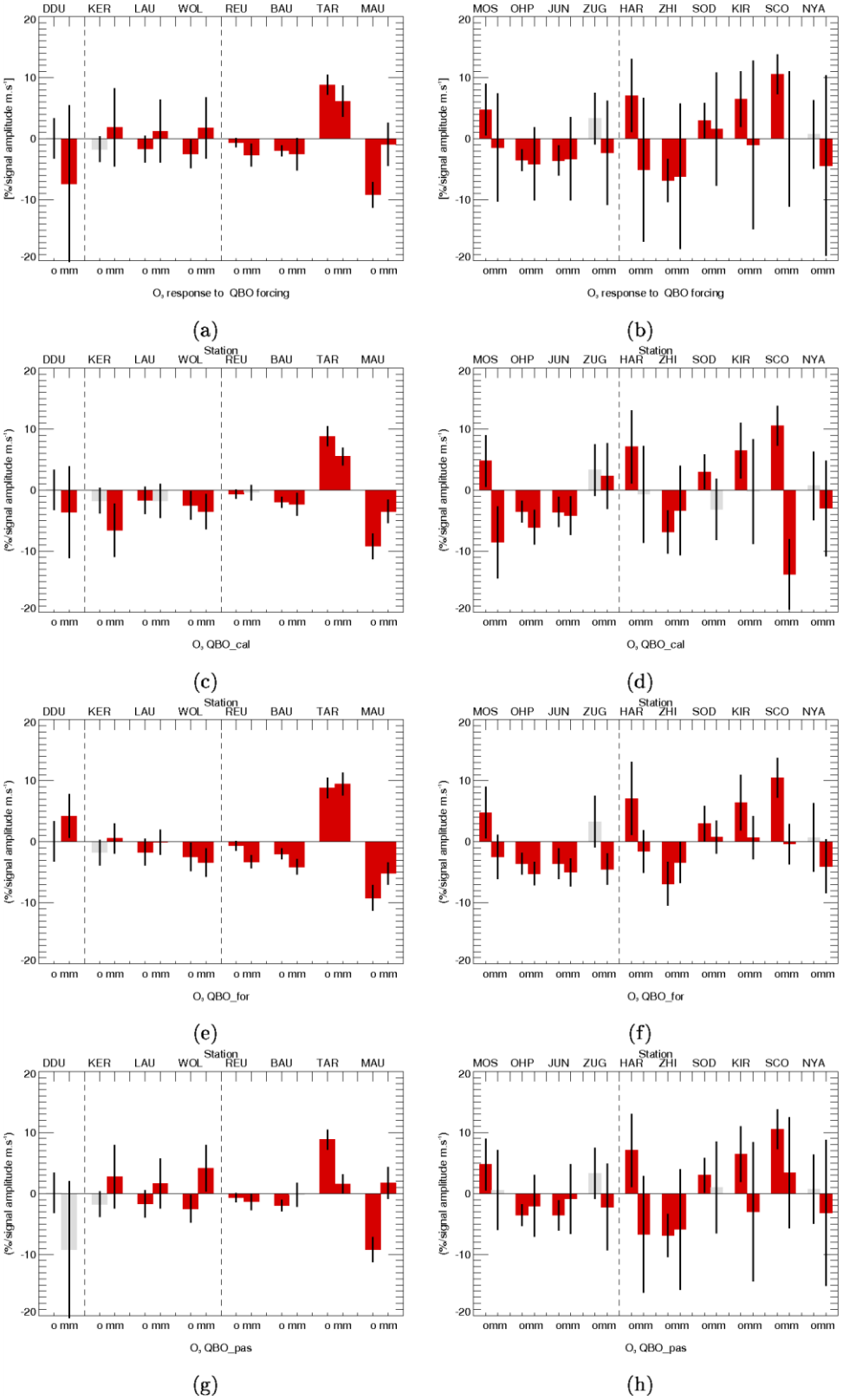


Figure 16: (a) N_2O column trend (expressed as % change per decade); (b) HCl column sensitivity to total chlorine loading (expressed as % change in HCl column per 65 pptv change in total chlorine loading); (c) and (d) sunrise NO_2 column sensitivity to aerosol forcing (expressed as % change in NO_2 column per change of 50 $\mu m^2.cm^{-2}$ in surface aerosol density column). Observed (O) and multi-model modelled mean (MM) results are next to each other for each station. Red indicates statistically significant results whereas grey indicates results that are not statistically significant. Errors are indicated with the vertical bars.

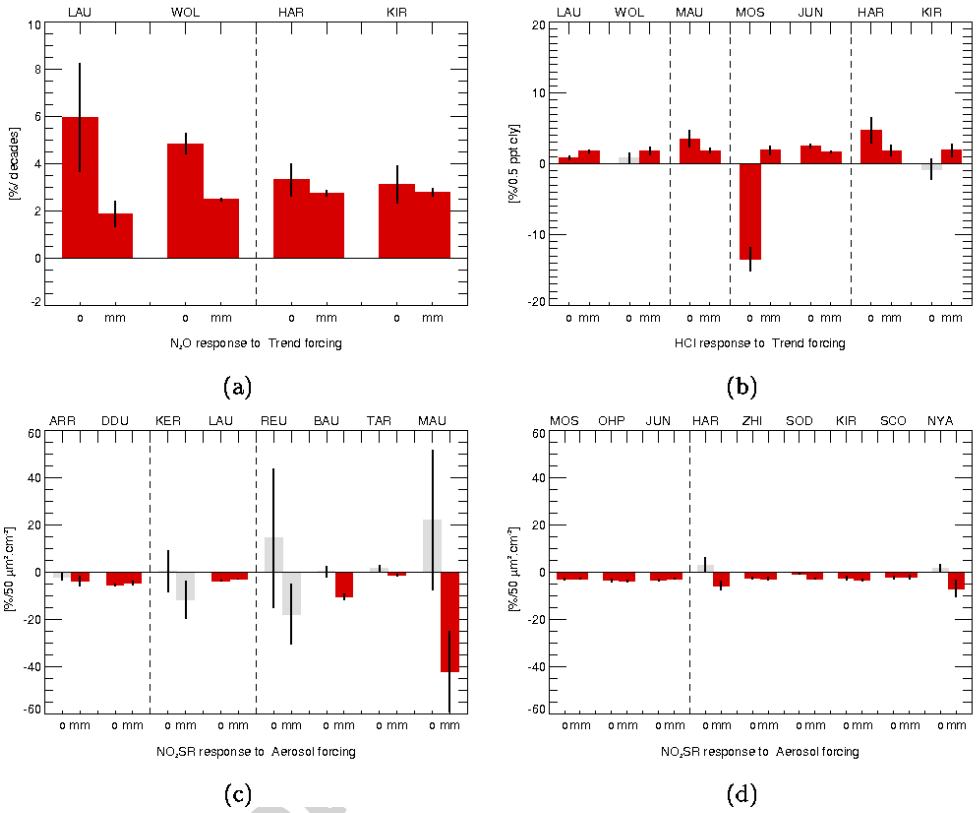


Table 1: Station used here, with location, abbreviation, instrument and period.

Stations	Location	Abbreviation	Instrument	Species	Period
Nyalesund	79°N-12°E	NYA	SAOZ	O_3, NO_2	1991-2004
Scoresbysund	70°N-22°W	SCO	SAOZ	O_3	1994-2004
			SAOZ	NO_2	1992-2004
Kiruna	68°N-20°E	KIR	SAOZ	NO_2	1991-2004
			FT-IR	O_3, HCl, CH_4, N_2O	1996-2004
Sodankyla	67°N-27°E	SOD	SAOZ	O_3, NO_2	1990-2004
Zhigansk	67°N-123°E	ZHI	SAOZ	O_3	1994-2004
			SAOZ	NO_2	1992-2004
Harestua	60°N-11°E	HAR	SAOZ	NO_2	1995-2004
			FT-IR	O_3, HCl, CH_4, N_2O	1995-2004
Zugspit	47°N-11°E	ZUG	FT-IR	O_3	1996-2004
Jungfrajoch	46°N-8°E	JUN	SAOZ	O_3	1991-2004
			SAOZ	NO_2	1996-2004
			FT-IR	HCl	1990-2004
Observatoire de Haute Provence	44°N-6°E	OHP	SAOZ	O_3	1993-2004
			SAOZ	NO_2	1992-2004
Moshiri	44°N-142°E	MOS	FT-IR	O_3, HCl,	1996-2001
			SAOZ	NO_2	1991-2004
Mauna Loa	20°N-156°W	MAU	FT-IR	O_3, HCl	1992-2001
			SAOZ	NO_2	1996-2004
Tarawa	1°N-173°E	TAR	SAOZ	O_3	1992-1999
			SAOZ	NO_2	1993-1999
La Réunion	21°S-55°E	REU	SAOZ	O_3, NO_2	1994-2004
Bauru	22°S-49°W	BAU	SAOZ	O_3, NO_2	1996-2004
Wollongong	34°S-151°E	WOL	FT-IR	O_3, HCl, N_2O	1995-2004
			FT-IR	CH_4	2003-2004

Lauder	45°S-170°E	LAU	FT-IR	O_3	1994-2004
			FT-IR	HCl	1991-2004
			FT-IR	CH_4 and N_2O	2001-2004
			SAOZ	NO_2	1982-2004
Kerguelen	49°S-70°E	KER	SAOZ	O_3, NO_2	1996-2004
Dumont d'Urville	67°S-140°E	DDU	SAOZ	O_3, NO_2	1988-2004
Arrival Heights	78S-167E	ARR	SAOZ	NO_2	1991-2004

Table 2: Models used here, with institution, horizontal and vertical resolutions, top level pressure, some main parameterizations, chemical species with T3M outputs, and references.

Chemistry-Climate Models	Institution	Horizontal Resolution, number of levels, top level	QBO	Solar variability	Aerosol heating rate	Species output in T3M	References
AMATRA C3	NOAA, US	~200 km, L48, 0.017 hPa	Yes	Yes	GISS data	O_3	Austin and Wilson (2006)
CAM3.5	NCAR, US	1.9°x2.5°, L26, 3.5 hPa	Forced	Yes	No	O_3	Lamarque et al. (2008)
CCSRNIES	NIES, Japan	T42, L34, 0.0012 hPa	Forced	Yes	GISS data	O_3, N_2O, CH_4, NO_2 and HCl	Akiyoshi et al. (2009)
CMAM	Environment Canada	T31, L71, 0.00081 hPa	No	Yes	SAD data	O_3, N_2O, CH_4, NO_2 and HCl	Scinocca et al. (2008)
LMDz-REPOBUS	IPSL, France	2.5°x2.5°, L50, 0.07 hPa	No	Yes	No	O_3, N_2O, CH_4, NO_2 and HCl	Jourdain et al. (2008)
SOCOL	PMOD/WRC, Switzerland	T30, L39, 0.01 hPa	Forced	Yes	Mix sources	O_3, N_2O, CH_4 and HCl	Schraner et al. (2008)
ULAQ	University of Aquila, Italy	R6, L26,	Forced	No	Volcanic injections of SO_2 and	O_3, N_2O and CH_4	Pitari et al. (2002)

	0.04 hPa	aerosol calculation

Highlights

- Variability in stratospheric composition evaluated in models and NDACC observations.
- Models reproduce the externally forced inter-annual variability seen in NDACC data.
- But models vastly underestimate the internal and hence total inter-annual variability.