

A vertically resolved, monthly mean, ozone database from 1979 to 2100 for constraining global climate model simulations

B. HASSLER*†, G. E. BODEKER†, I. CIONNI† and M. DAMERIS‡

†National Institute of Water and Atmospheric Research, Lauder, P. B. 50061, Omakau, Central Otago 9182, New Zealand

‡Deutsches Zentrum für Luft- und Raumfahrt, Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany

A new global three-dimensional (latitude, pressure, time) zonal mean monthly mean ozone database, from 1979 to 2100, has been created. The database extends from 87.5°S to 87.5°N at a 5° latitude resolution and from 878 hPa to 0.0460 hPa in 70 pressure levels ~1 km apart. The full period is divided into two sub-periods, viz: (1) observations from 1979 to 2006 obtained from several sources of high vertical resolution ozone profile measurements, and (2) extensions from 2007 to 2100 based on chemistry climate model (CCM) projections. At each latitude and pressure level the CCM ozone time series are shifted and scaled to better match the measurements from 1997 to 2006 to ensure a smooth transition from observations to model output. This new database is expected to provide ozone boundary conditions suitable for long-term global climate model simulations.

1. Introduction

The global climate model (GCM) simulations for the IPCC fourth assessment report differed in their treatment of ozone radiative forcing and in particular in future ozone radiative forcing (Miller *et al.* 2006). For the twentieth century, some simulations used the ozone data set of Randel and Wu (1999) to constrain ozone radiative forcing, others used the data set of Kiehl *et al.* (1999), while others excluded ozone radiative forcing altogether. The Randel and Wu database is the output from a linear least squares regression model incorporating decadal trends, solar cycle and quasi-biennial oscillation (QBO) terms. The decadal trends are modelled using equivalent effective stratospheric chlorine (EESC; Daniel *et al.* 1995). However, because only ozonesonde profiles from Syowa (69°S) are used to characterize the dependence of ozone on EESC poleward of 60°S, and because Syowa is close to the vortex edge and will occasionally sample extra-vortex air, ozone trends over Antarctica in this database are likely to be underestimated.

There was a similar range of treatment of twenty-first century stratospheric ozone changes; some models assumed that stratospheric ozone remained constant at present day values while others assumed a slow recovery to pre-industrial values. Others, again, excluded ozone radiative forcing completely. Those models that did not include future increases in ozone over southern high latitudes are unlikely to reliably simulate

*Corresponding author. Email: birgit.hassler@gmail.com

future changes in the southern annular mode (Perlwitz *et al.* 2008) and hence changes in Antarctic surface climate (Thompson and Solomon 2002).

While models such as HadCM3 generated their own ozone fields interactively (Johns *et al.* 2003), for those models that do not have this functionality, for future GCM simulations there is an urgent need for a homogeneous global and vertically resolved ozone data set spanning the full model simulation period. This paper presents one such possible data set.

The GCMs which are likely to use such an ozone boundary condition database will differ in their horizontal and vertical resolution. Therefore, to avoid spatial sampling biases, we have created the ozone database at high latitudinal and vertical resolution. Zonal means are calculated within 5° latitude zones centred at 87.5°S, 82.5°S ... 82.5°N, 87.5°N. The database has 70 pressure levels approximately 1 km apart extending from 878 hPa to 0.0460 hPa. This results in a total of 2520 monthly mean ozone number density time series from 1979 to 2100.

CCM simulations are available for the full period (1979–2100). However, if these time series alone were to be used to prescribe the ozone boundary conditions for GCM simulations, it is unlikely that the true ozone radiative forcing would be reproduced in all parts of the atmosphere. The approach taken in this study is to combine the advantages derived from an observational time series (accurate reproduction of ozone radiative forcing) with the advantages derived from CCM simulations of ozone (temporal coverage to 2100). To this end, monthly mean zonal mean ozone number density time series, from 1979 to 2006, were calculated from ozone observations (section 2). To extend these time series to 2100 an ensemble mean of ozone simulations from 10 different CCMs was calculated. Because the absolute values of the CCM time series may be biased, at each latitude and altitude the CCM data are shifted and scaled to match the observations during an overlap period (1997 to 2006), as described further in section 3. Some GCM simulations are planned to also cover the period before 1979, extending possibly as far back as 1850. An ozone database extending back this far is therefore also required. A brief overview of one method by which this may be achieved is provided in section 4.

2. Observational core (1979–2006)

2.1 Data sources

The ozone measurements for the observational core of the database are provided by the Binary DataBase of Profiles (BDBP; Hassler *et al.* 2008). The BDBP is constructed from high vertical resolution ozone profile measurements from the SAGE I, SAGE II, HALOE, POAM II and POAM III satellite instruments and from ozonesonde flights made at 135 stations globally.

2.2 Calculation of unfilled monthly means

Monthly mean zonal mean ozone number density time series, on the latitude/pressure grid described above, were calculated from all available data sources in the BDBP. A minimum of six data values per month was necessary to calculate a reliable monthly mean. Because version 1.0 of the BDBP does not remove any offsets and drifts between the different data sources, sub-sampling of the available measurements in each month was used to avoid anomalously high and low values (the highest and lowest quartile of the available data were discarded). Since for each individual data

value an error value was available in the BDBP, monthly means for each latitude band and pressure level were calculated using error weighting (Paradine and Rivett 1964):

$$\bar{x} = \frac{\sum w_i x_i}{\sum w_i} \quad (1)$$

with

$$w_i = \frac{1}{\sigma_i^2} \quad (2)$$

where x_i is the i -th data value, and σ_i is the error on the i -th data value.

Although multiple data sources were used to calculate the monthly mean ozone values, there were periods and locations where there were insufficient data to calculate a valid monthly mean. However, since GCM simulations require a database with no missing values, regression constrained interpolation, as described in section 2.3, was used to fill gaps in the observational time series.

2.3 Calculation of filled monthly means

The approach used to fill gaps in the observational time series was to first capture the effects of known sources of geophysical variability in monthly mean ozone through the application of a linear least squares regression model. Regression model estimates of ozone during the periods when ozone is missing, together with interpolated regression anomalies, can then be used to fill the missing values. To this end, at each latitude and pressure, a linear least squares regression model was fitted to the available monthly mean ozone time series. The regression model includes basis functions for a mean annual cycle, EESC, that is sensitive to the mean age of air (Waugh and Hall 2002, Newman *et al.* 2007), QBO, solar cycle, El Niño Southern Oscillation (ENSO), and for stratospheric aerosol loading resulting from the eruptions of the El Chichón and Mt. Pinatubo volcanos.

If there are insufficient monthly means to reliably calculate the regression model fit, the data gaps cannot be filled. To this end, a method was developed that permits values to be ‘borrowed’ from neighbouring pressure levels or latitude bands to ensure that sufficient data are available for the regression model fits. The regression model is fitted first at those latitudes and altitudes where the fewest data are missing. For time series where more than 70 values are missing (~20% of the data) values are borrowed from surrounding time series (that may have already been filled) until only 70 values are missing. This ensures that there are enough data to constrain the regression fit. Specifically, values up to 10° in latitude and 2 km in altitude away are considered for borrowing. Where 10 or more such values are available for borrowing, they are first corrected for latitudinal and vertical gradients using a prescribed monthly mean climatology calculated using unfilled values from 1979 to 2006. The median of these corrected values then provides the borrowed value. Note that this ‘borrowing’ procedure is used only to obtain sufficient data to reliably constrain the regression model, not to actually fill the data gaps. The next step is to calculate the residuals (measurements minus regression model) and linearly interpolate between these residuals. If data are missing at the beginning or end of the time series, interpolation is not possible and in this case the residuals are simply exponentially decayed to zero with an e-folding time of 12 months. Linearly interpolated residuals would result

in filled monthly means not showing the same month-to-month variability as the observations. Therefore 'noise', based on quantified variability during periods when observations are available, and including the auto-correlation in that variability, was added to the interpolated residuals. The first order auto-correlation for each ozone time series was determined from consecutive residual pairs, and the remaining 'white noise' determined from the residuals of the auto-correlation model. The auto-correlation and 'white noise' statistics are used to generate a synthetic noise time series which is added to the linearly interpolated residuals. Where there were insufficient residual pairs to calculate these noise statistics, means of these statistics were obtained from the same surrounding time series as used for the borrowing of values for the regression fit. A graphical example of this procedure for filling the data gaps is provided in figure 1 for the zonal mean ozone between 30°N and 35°N and 58.2 hPa.

This method has the advantage of exactly reproducing the measurements where they are available, and providing realistic estimates of ozone through the data gaps based on known sources of variability in ozone. The addition of realistic noise, matching both the magnitude and auto-correlation characteristics of the observed month-to-month variability, to the interpolated residuals results in filled values that are statistically almost indistinguishable from the observed values; long-range correlation in the observations would not be captured in the first order auto-correlation model used to fill the missing monthly means (Vyushin *et al.* 2007). However, this is a second order effect and has been neglected in our procedure for filling missing values.

An example of the effects of data filling on the monthly mean ozone times series between 30°N and 35°N is shown in figure 2.

3. Extension into the future (2007–2100)

3.1 Data sources

CCMVal (Eyring *et al.* 2005a) reference simulations have exact specifications for anthropogenic and natural forcings to ensure comparability between output from different models. For the reference simulations from CCMVal where future ozone development is considered (REF2) only anthropogenic forcings are included. Natural forcings, such as the influence of the 11-year solar cycle, were excluded. A QBO was only included in the simulations if the models internally generate a QBO (Eyring *et al.* 2005b).

REF2 simulations from 10 different CCMs were chosen to extend the database beyond the observations and out to 2100 (Eyring *et al.* 2007). The 10 models are listed in table 1 with the start and end dates of their respective REF2 runs. For three models more than one REF2 experiment was available. For each of those three models the mean of their time series was calculated to ensure only one entry per model for the overall ensemble mean. As detailed in table 1, the start and end dates of the REF2 experiments differ from model to model.

3.2 Monthly means

In CCMVal all ozone data from the CCMs are given in mass mixing ratio (mole/mole). The conversion to ozone number density required for this database was done by using the zonal mean temperatures of each model. Beside the conversion of the

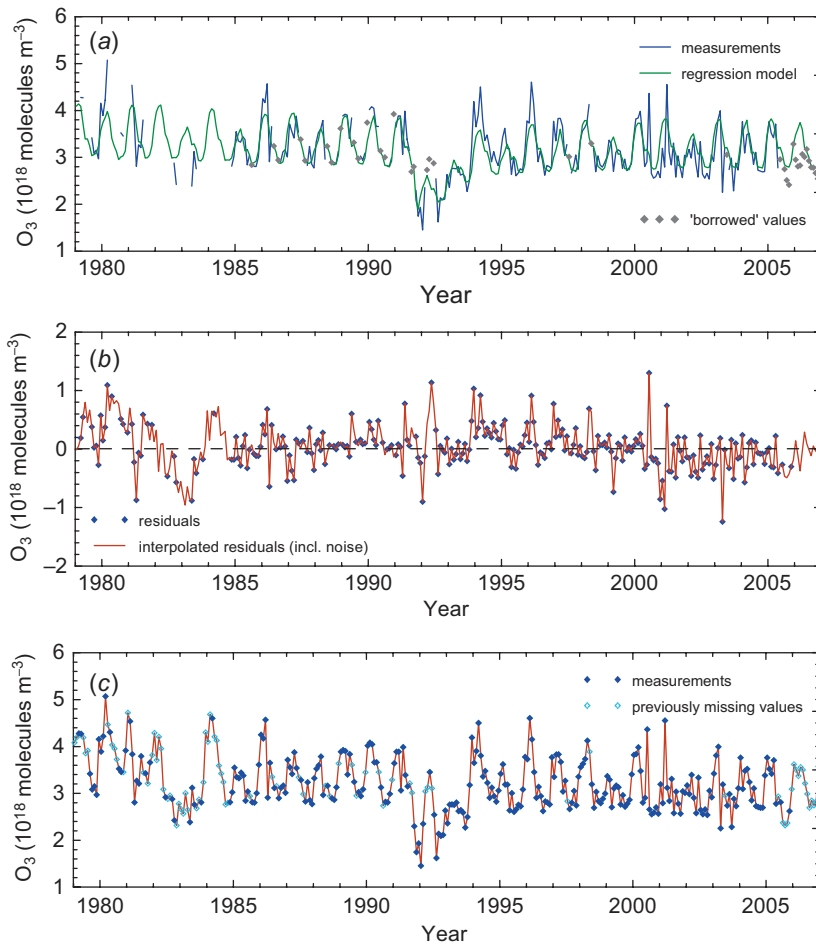


Figure 1. Ozone time series for the latitude band 30°N – 35°N and at the pressure level of 58.2 hPa. (a) Time series with data gaps (blue line), 'borrowed' values (grey diamonds) and regression model output (green line). (b) Residuals (blue diamonds) and linearly interpolated residuals with added noise (red line). (c) Complete time series with measurements (blue diamonds) and filled values (turquoise diamonds).

ozone units, interpolation between the model native latitude and vertical grids was required to match that of the observations. Interpolation between latitudes was done linearly after ensuring that the ozone distribution was not changed by the interpolation, whereas interpolation between pressure levels was logarithmic. No extrapolation above the upper boundary of the model was permitted.

Finally, the ensemble mean was calculated using an error weighted averaging approach similar to that described in section 2. In this case, however, the 'uncertainty' (1σ value) for any model monthly mean is prescribed as the standard deviation of the differences between the model monthly means and the unweighted ensemble means of all 10 CCMs. In this way, when the weighted ensemble mean is calculated, those models that originally showed the greatest discrepancy from the unweighted ensemble mean, are given reduced emphasis. No provision was made for the fact that the start

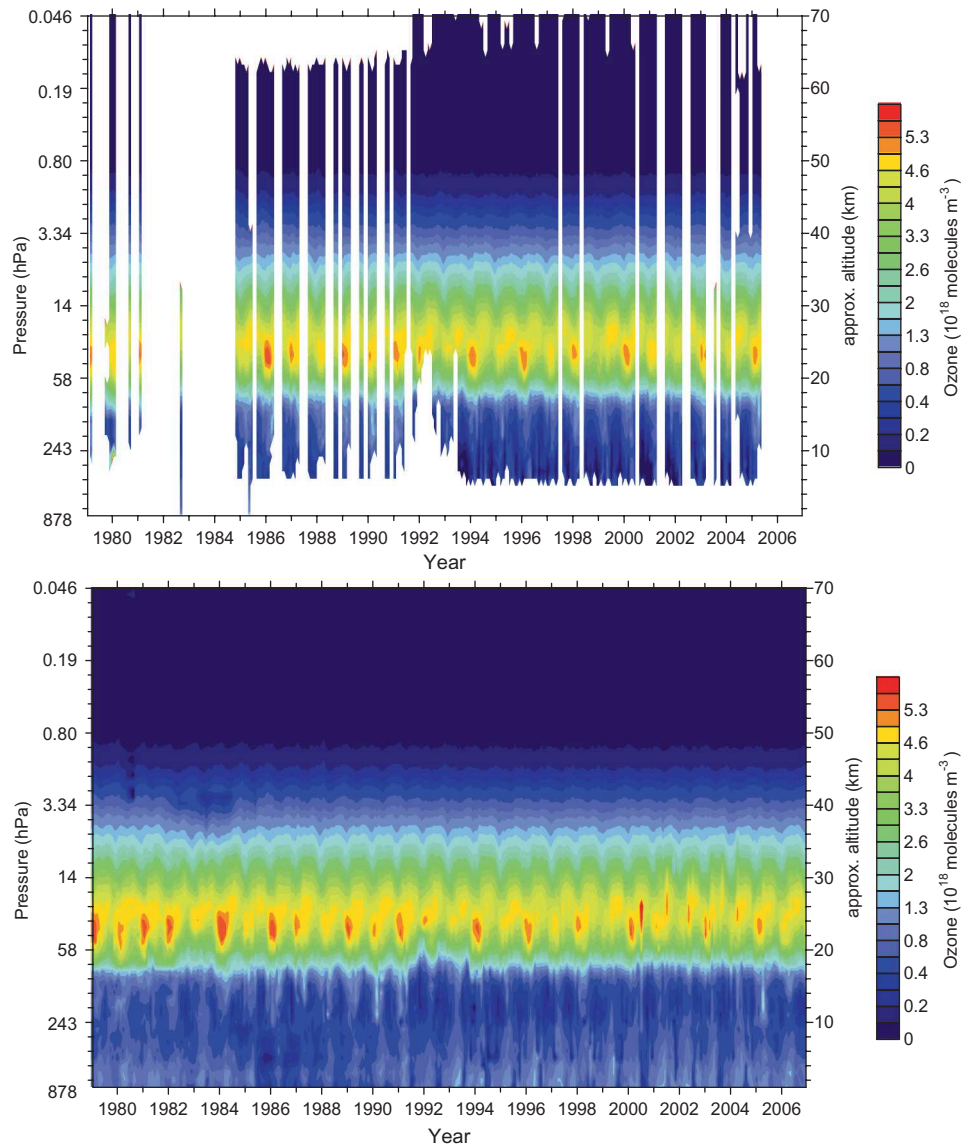


Figure 2. Monthly mean ozone values for the latitude band 30°N–35°N for all 70 pressure levels. The upper panel shows the unfilled data field (white areas indicate missing data values) and the lower panel shows the data field after the regression constrained interpolation has been applied.

and end dates of the several simulations differ. The result is that while the ensemble mean was calculated from 10 different models from 2000 to 2020, up to 10 hPa, only three different models provided values for the period 2050–2100 (see table 1).

3.3 Combining the observational core with CCM output

The data from the observational core and the CCM ensemble mean were combined separately for each latitude band and pressure level. During a 10 year

Table 1. CCMs contributing to the calculated ensemble mean with the start and end dates of their REF2 simulations, and details on the spatial resolution of the model output.

Model	Start date	End date	Upper boundary (hPa)	Number of vertical levels	Domain/Resolution
AMTRAC	1990	2099	0.0017	48	$2^\circ \times 2.5^\circ$
CCSRNIES	1980	2050	0.01	34	T42
CMAM ^a	1960	2099	0.0006	71	T32
E39C ^a	2000	2020	10	39	T30
GEOSCCM	2000	2099	0.01	55	$2^\circ \times 2.5^\circ$
MAECHAM4CHEM	2000	2020	0.01	39	T30
SOCOL	1980	2050	0.01	39	T30
ULAQ	1960	2050	0.04	26	$10^\circ \times 22.5^\circ$
UMSLIMCAT	1980	2019	0.01	64	$2.5^\circ \times 3.75^\circ$
WACCM ^a	1980	2050	4.5×10^{-6}	66	$4^\circ \times 5^\circ$

^aThe mean of three independent experiments was calculated to generate the data set of these models.

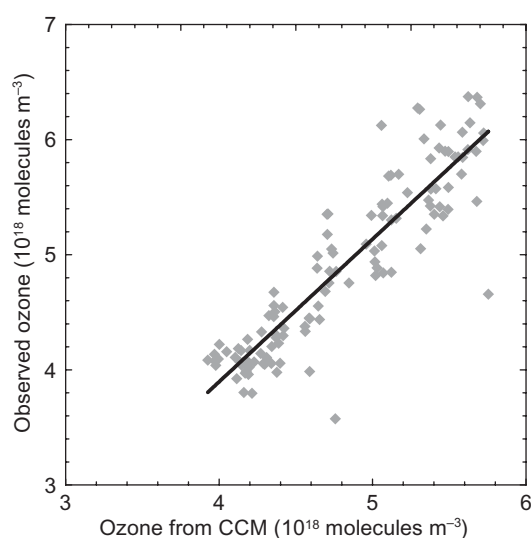


Figure 3. Monthly mean ozone values from the observational core and the CCM ensemble mean time series for the latitude band 60°N – 65°N and pressure level 58.2 hPa. Only monthly means from 10 overlapping years (1997–2006) are plotted. The coefficients from the straight line fit to the scatter plot (best fit equation: $Y = 1.2393 X - 1.0605$; $R^2 = 0.82$) are then used to scale the CCM ensemble mean time series to better match the observations.

overlap period (1997–2006), measured and modelled ozone values were plotted against each other (see figure 3) to determine a scaling (slope of the straight line fit to the scatter plot) and offset (Y -axis intercept of the straight line fit) to be applied to the CCM ensemble mean values to bring them into better agreement with the observations. An example of the effect of the shift and scaling applied to the CCM ensemble mean is shown in figure 4.

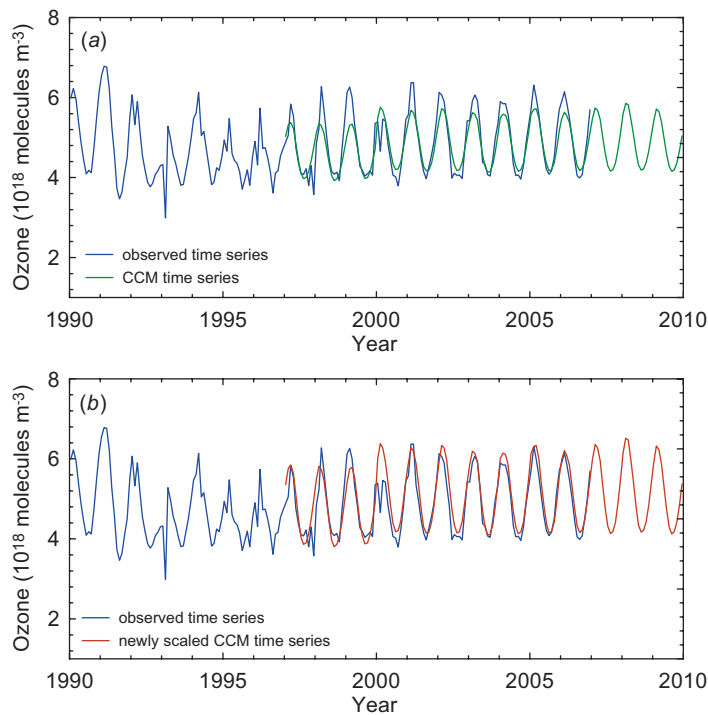


Figure 4. Ozone time series for the latitude band 60°N – 65°N and pressure level 58.2 hPa. (a) Time series from the observational core (measurements and filled data gaps; blue line) and CCM ensemble mean time series (green line). (b) Time series from the observational core (blue line) and adjusted CCM ensemble mean time series (red line).

4. Future plans

The next step in this project will be to extend the database back to 1850. While good global coverage of total column ozone measurements is available from 1964 onwards (Fioletov *et al.* 2002), global vertical ozone profile data are available mainly with the advent of satellite-based measurements; operational ozonesonde stations before this period are not sufficiently global. Therefore, from 1964 to 1978, a monthly climatology of ozone profiles, calculated from the monthly means over the period 1979–1983, will be scaled to match total column ozone measurements available for this period (Fioletov *et al.* 2002). Before 1964, the ground-based total column ozone measurements are not sufficiently spatially dense to permit such an approach. Therefore, for the period 1850–1963, regression model fits to the monthly means from 1964 to 2006 will be used to extend the database back to 1850 assuming that data for all basis functions can be obtained back to that date.

5. Conclusions

There are a number of metrics which can be used to gauge the success of the Montreal Protocol. One such metric is changes in ozone in response to changes in ozone depleting substances in different regions of the stratosphere. To quantify these changes in ozone, a long-term global vertically resolved database of ozone

concentrations is required. The creation of such a database, extending from close to the surface to the lower mesosphere, and covering the period 1979–2100, has been described in detail above. Future work will focus on the attribution of changes in ozone derived from this new database to changes in emissions of ozone depleting substances brought about by the Montreal Protocol and will thereby support the growing body of scientific evidence that demonstrates the success of the Protocol. This database will also be suitable for providing boundary conditions suitable for prescribing ozone radiative forcing in global climate model simulations that do not generate their own ozone fields interactively.

Acknowledgements

We thank the CCM groups of AMTRAC (GFDL, USA), CCSRNIES (NIES, Tsukuba, Japan), CMAM (MSC, University of Toronto and York University, Canada), E39C (DLR, Oberpfaffenhofen, Germany), GEOSCCM (NASA/GSFC, USA), MAECHAM4CHEM (MPI Mainz, Hamburg, Germany), SOCOL (PMOB/WRC and ETHZ, Switzerland), ULAQ (University of L'Aquila, Italy), UMETRAC (UK Met Office, UK, NIWA, NZ), and WACCM (NCAR, USA) for providing their model output for this analysis. We also thank the Chemistry-Climate Model Validation Activity (CCMVal) of WCRP-SPARC (World Climate Research Programme - Stratospheric Processes and their Role in Climate) for organizing the model data analysis activity, and the British Atmospheric Data Center (BADC) for collecting and archiving the model output. We further gratefully acknowledge D. Smale's input concerning the calculation of the error weighted monthly means.

References

- DANIEL, J.S., SOLOMON, S. and ALBRITTON, D.L., 1995, On the evaluation of halocarbon radiative forcing and global warming potentials. *Journal of Geophysical Research*, **100**, D1, pp. 1271–1285.
- EYRING, V., HARRIS, N.R.P., REX, M., SHEPHERD, T.G., FAHEY, D.W., AMANATIDIS, G.T., AUSTIN, J., CHIPPERFIELD, M.P., DAMERIS, M., FORSTER, P.M., GETTELMAN, A., GRAF, H-F., NAGASHIMA, T., NEWMAN, P.A., PAWSON, S., PRATHER, M.J., PYLE, J.A., SALAWITCH, R.J., SANTER, B.D. and WAUGH, D.W., 2005a, A strategy for process-oriented validation of coupled chemistry-climate models. *Bulletin of the American Meteorological Society*, **86**, pp. 1117–1133.
- EYRING, V., KINNISON, D.E. and SHEPHERD, T.G., 2005b, Overview of planned coupled chemistry-climate simulations to support upcoming ozone and climate assessments. *SPARC Newsletter*, **25**, pp. 11–17.
- EYRING, V., WAUGH, D.W., BODEKER, G.E., CORDERO, E., AKIYOSHI, H., AUSTIN, J., BEAGLEY, S.R., BOVILLE, B.A., BRAESICKE, P., BRÜHL, C., BUTCHART, N., CHIPPERFIELD, M.P., DAMERIS, M., DECKERT, R., DEUSHI, M., FRITH, S.M., GARCIA, R.R., GETTELMAN, A., GIORGETTA, M.A., KINNISON, D.E., MANCINI, E., MANZINI, E., MARSH, D.R., MATTHES, S., NAGASHIMA, T., NEWMAN, P.A., NIELSEN, J.E., PAWSON, S., PITARI, G., PLUMMER, D.A., ROZANOV, E., SCHRANER, M., SCINOCCA, J.F., SEMENIUK, K., SHEPHERD, T.G., SHIBATA, K., STEIL, B., STOLARSKI, R.S., TIAN, W. and YOSHIKI, M., 2007, Multimodel projections of stratospheric ozone in the 21st century. *Journal of Geophysical Research*, **112**, D16303, doi:10.1029/2006JD008332.
- FIOLETOV, V.E., BODEKER, G.E., MILLER, A.J., MCPETERS, R.D. and STOLARSKI, R., 2002, Global and zonal total ozone variations estimated from ground-based and satellite measurements: 1964–2000. *Journal of Geophysical Research*, **107**, 22, 4647, doi:10.1029/2001JD001350.

- HASSLER, B., BODEKER, G.E. and DAMERIS, M., 2008, Technical Note: A new global database of trace gases and aerosols from multiple sources of high vertical resolution measurements. *Atmospheric Chemistry and Physics*, **8**, pp. 5403–5421.
- JOHNS, T.C., GREGORY, J.M., INGRAM, W.J., JOHNSON, C.E., JONES, A., LOWE, J.A., MITCHELL, J.F.B., ROBERTS, D.L., SEXTON, D.M.H., STEVENSON, D.S., TETT, S.F.B. and WOODAGE, M.J., 2003, Anthropogenic climate change for 1860 to 2100 simulated with the HadCM3 model under updated emissions scenarios. *Climate Dynamics*, **20**, pp. 583–612.
- KIEHL, J.T., SCHNEIDER, T.L., PORTMANN, R.W. and SOLOMON, S., 1999, Climate forcing due to tropospheric and stratospheric ozone. *Journal of Geophysical Research*, **104**, D24, pp. 31239–31254.
- MILLER, R.L., SCHMIDT, G.A. and SHINDELL, D.T., 2006, Forced annular variations in the 20th century Intergovernmental Panel on Climate Change Fourth Assessment Report models. *Journal of Geophysical Research*, **111**, D18101, doi:10.1029/2005JD006323.
- NEWMAN, P.A., DANIEL, J.S., WAUGH, D.W. and NASH, E.R., 2007, A new formulation of equivalent effective stratospheric chlorine (EESC). *Atmospheric Chemistry and Physics*, **7**, pp. 4537–4552.
- PARADINE, C.G. and RIVETT, B.H., 1964, Theory of errors. In *Statistical Methods for Technologists* (3rd edn), D.G. Christopherson (Ed), pp. 157–170 (London: The English Universities Press Ltd.).
- PERLWITZ, J., PAWSON, S., FOGT, R.L., NIELSEN, J.E. and NEFF, W.D., 2008, Impact of stratospheric ozone hole recovery on Antarctic climate. *Geophysical Research Letters*, **35**, L08714, doi:10.1029/2008GL033317.
- RANDEL, W.J. and WU, F., 1999, A stratospheric ozone trends data set for global modeling studies. *Geophysical Research Letters*, **26**, 20, pp. 3089–3092.
- THOMPSON, D.W.J. and SOLOMON, S., 2002, Interpretation of recent southern hemisphere climate change. *Science*, **296**, pp. 895–899.
- WAUGH, D.W. and HALL, T.M., 2002, Age of stratospheric air: theory, observations and models. *Reviews of Geophysics*, **40**, 4, 1010, doi: 10.1029/2000RG000101.
- VYUSHIN, D.I., FIOLETOV, V.E. and SHEPHERD, T.G., 2007, Impact of long-range correlations on trend detection in total ozone. *Journal of Geophysical Research*, **112**, D14307, doi:10.1029/2006JD008168.