REanalysis of the TROpospheric chemical composition over the past 40 years
A long-term modelling study of tropospheric chemistry

Funded under the 5th EU framework programme
EU-contract No. EVK2_CT-2002-00170

Work Package 3
Deliverable D3-2

Analysis of model performance for seasonal variations

N.H. Savage[1], S. Dalsøren[2], T. van Noije[3], S. Rast[4], M. Schultz[4,5], and S. Szopa[6]

1. Department of chemistry, University of Cambridge, UK
2. Institute of Geophysics, University of Oslo, Norway
3. Koninklijk Nederlands Meteorologisch Instituut KNMI, The Netherlands
4. Max-Planck-Institute for Meteorology, Hamburg, Germany
5. now at ICG-II, Forschungszentrum Jülich, Jülich, Germany
6. Laboratoire des Sciences du Climat et de l'Environnement, Gif-sur-Yvette, France

1. Introduction

Over the last half century, anthropogenic emissions of ozone precursors and greenhouse gases have continuously increased on a global scale. The changes in tropospheric composition due to these increases in emissions are not well quantified as a result of the sparsity of observations, especially before the 1990s. Global modelling studies provide an essential tool in developing our understanding of budgets of anthropogenic pollution and have been widely used in examining the production, transport and destruction of tropospheric ozone. However, before using any of these results it is necessary to test the validity of the results against observation. Previous intercomparisons include: the GIM/IGAC intercomparison exercise (Kanakidou et. al., 1999); IPCC Ox-Comp (IPCC, 2001); the WCRP intercomparison (Rasch et. al., 2000) and the ACCENT/IPCC study (Dentener et al., 2006).

The RETRO project aims to use the newly available ERA-40 meteorological data set to carry out multi-decadal CTM and GCM integrations to achieve reliable estimates of trends and variability for tropospheric trace species. Using the period 1997-2000 from these runs an exercise has been performed to validate the models when forced with ERA-40 data and the RETRO emissions database. This paper describes the overall model results for the year 1997 with a focus on the global budgets of tropospheric ozone and OH, and it summarizes the validation of the models with observations from longer-term surface measurements of ozone and CO and from ozone sondes. In order to arrive at some objective criteria for assessing the performance of individual models, skill scores are defined as the fraction of modelled values which are within a predetermined threshold of the observations. This methodology highlights some common strengths and weaknesses for all models (which may also point to limitations in the observational data), and it identifies deficiencies in individual models for specific tasks. Further development of such methods and common quality objectives is urgently needed as a larger community effort.

2. Model and input data descriptions

2.1 Model Descriptions

Five global chemistry models were used in this study: three Chemistry Transport models (CTMs), i.e. p-TOMCAT, TM4 and Oslo CTM2, and two nudged General Circulation Models (GCMs), i.e. MOZECH and LMDZ-INCA. Details of the resolution, physical parameterisations and chemistry scheme of these models is presented in Table 1. Further model details can be found in the RETRO document from workpackage 4 describing the long runs, D4-4.

	MOZECH	LMDz-INCA	TM4	p-TOMCAT	CTM2
Participants	MPG-IMET S. Rast M. Schultz	LSCE S. Szopa D. Hauglustaine	KNMI T. van Noije P. van Velthoven	UCamb N. Savage J. Pyle	UiO S. Dalsoren J. Sundet M. Gauss
Resolution	2.8x2.8 L31	3.75x2.5 L19	3.0x2.0 L25	2.8x2.8 L31	2.8x2.8 L60
Total number of species (tracers)	65 (63)	75 (73) or 89 (87)	37 (22)	60 (42)	61 (51)4
Chem. scheme (reactions)	NMHC (135)	NMHC (332)	NMHC (95) Sulphur aerosols	NMHC (131)	NMHC (140) Aerosols
Advection	Lin+Rood (or SL)	Van-Leer	Slopes	Prather	Prather
Convection	Tiedtke	Tiedtke	Tiedtke	Tiedtke	CTM2
Boundary Layer	Monin Obukhov	LMDz	Holtslag+Boville	Holtslag+Boville	Holtslag
CPU (Wallclock) 1year	300 h (24 h)	35-48h	22h	24h	200 h
Biogenic emissions including NOx	Interactive MEGAN model by Guenther et al. [1998], soil CO from POET	ORCHIDEE climatology [Lathiere et al., 2005]	ORCHIDEE climatology [Lathiere et al., 2005]	ORCHIDEE data for 1997 [Lathiere et al., 2005]	POET
Lightning emissions	Interactive Grewe et al., 2001	Interactive	Interactive [Meijer et al., 2001]; scaled to 5 Tg N/yr for 1997 (ERA-40)	Interactive	Interactive
Height distribution of biomass burning emissions	Altitude profiles without segregation of vegetation types	no	Altitude profiles with segregation into 4 vegetation types	No	Altitude profiles with segregation into 4 vegetation types
Ocean CO and other VOC emissions	POET (CO, C2H6, C3H8, higher alkanes, C2H4, C3H6)	Erickson and, Taylor (GRL 1992) rescaled: CO: 20.0, Isoprene: 0.88, C_2H_4: 0.688, C_3H_6: 0., C_2H_2: 0.20, alkenes: 0.27, CH_3COCH_3: 12.41Tg.yr-1	POET (CO and CBM-4 categories from C2H6, C3H8, C2H4, C3H6)	POET (CO, C2H6, C3H8)	POET (CO, C2H6, C3H8, higher alkanes, C2H4, C3H6)
Aircraft NOx emissions	Grewe [2003] ?	Annual total from Grewe [2003] distributed according to ANCAT	Grewe [2003]	Grewe [2003] ?	Grewe [2003] ?
2 key references	Horowitz et al., 2003 Aghedo et al., 2006	*Hauglustaine et al.* [2003] *Brunner et al.* [2003]	Dentener et al. [2003] van Noije et al. [2004, 2006]	*Law et al.* [1998] *Stockwell et al.* [1999]	

Table 1. Summary of model properties

2.2 Model Inputs

2.2.1 Emissions

Emissions used in this study were taken from the RETRO emissions inventory for 1997 to 2000. These emissions include monthly variations of all types of emissions including industrial emissions (which in many previous studies had only annual average values). Unfortunately, the version of the RETRO emissions database used for the model runs presented here contained an error in the seasonality of the anthropogenic emissions. The seasonal cycle for the Southern Hemisphere should be the opposite to that of the Northern Hemisphere and a factor to account for this was to be applied to the data. Unfortunately, a mistake was made in applying this factor and it was applied not to the Southern Hemisphere point but to those longitudes west of the Greenwich meridian. This will cause a minor error in the seasonal cycle of these emissions. Fortunately, the most significant seasonal cycles in anthropogenic emissions of interest to RETRO are small in areas affected by this error. For more details see RETRO report D1-6. After this analysis was completed, an error in the MOZECH simulations between 1992 and 1999 was found. Due to a data processing problem, a wrong inventory for fire emissions was used with much lower emissions than there should be and located in wrong locations. Therefore, the MOZECH results presented in this report must be treated with caution.

2.2.2 Meteorology

All models runs used in this study were forced or nudged with meteorological data based on the ERA-40 meteorological reanalysis. The forty years of gridded meteorological data with 6-hour time resolution needed to perform the comprehensive atmospheric GCM/CTM simulations of this project constitute a huge amount of data. These data have been retrieved from the archive of the European Centre for Medium Range Weather Forecast (ECMWF), processed or reformatted in order to be used as inputs to the various models. Most models used the six-hour forecast meteorology from ECMWF, as these fields are in better balance than the analyses; but other models (TM and Oslo) used longer forecast periods. The two GCMs used different fields (LMDz INCA: 6-h forecasts of u, v, Wind speed 10m and surface pressure; MOZECH: analyses of surface pressure, T, vorticity and divergence), and they applied a nudging technique in order to constrain the meteorology to ERA-40.

3 Observations used for validation

3.1 Carbon Monoxide Observations

Flask samples are regularly analysed for carbon monoxide samples from a global network by CMDL. For more information see Novelli et al. 2003 and references therein.

Here we analyse observations from 42 stations which all have a good data coverage for the period of interest. Further details of the stations are given in the table below.

Station name	Code	Lat	Lon	Alt (m)
Alert, Nunavut, Canada	ALT	82.45	-62.52	210
Ascension Island, United Kingdom	ASC	-7.92	-14.42	54
Assekrem, Algeria	ASK	23.18	5.42	2728
Terceira Island, Azores, Portugal	AZR	38.77	-27.38	40
Baltic Sea, Poland	BAL	55.35	17.22	28
St. Davids Head, Bermuda, United Kingdom	BME	32.37	-64.65	30
Tudor Hill, Bermuda, United Kingdom	BMW	32.27	-64.88	30
Barrow, Alaska, United States	BRW	71.32	-156.6	11
Black Sea, Constanta, Romania	BSC	44.17	28.68	3
Cold Bay, Alaska, United States	CBA	55.2	-162.72	25
Cape Grim, Tasmania, Australia	CGO	-40.68	144.68	94
Christmas Island, Republic of Kiribati	CHR	1.7	-157.17	3
Crozet Island, France	CRZ	-46.45	51.85	120
Easter Island, Chile	EIC	-27.15	-109.45	50
Mariana Islands, Guam	GMI	13.43	144.78	6
Halley Station, Antarctica, United Kingdom	HBA	-75.58	-26.5	33
Hegyhatsal, Hungary	HUN	46.95	16.65	344
Storhofdi, Vestmannaeyjar, Iceland	ICE	63.34	-20.29	127
Tenerife, Canary Islands, Spain	IZO	28.3	-16.48	2360
Key Biscayne, Florida, United States	KEY	25.67	-80.2	3
Cape Kumukahi, Hawaii, United States	KUM	19.52	-154.82	3
Sary Taukum, Kazakhstan	KZD	44.45	75.57	412
Plateau Assy, Kazakhstan	KZM	43.25	77.88	2519
Park Falls, Wisconsin, United States	LEF	45.93	-90.27	868
Mace Head, County Galway, Ireland	MHD	53.33	-9.9	25
Sand Island, Midway, United States	MID	28.21	-177.38	7.7
Mauna Loa, Hawaii, United States	MLO	19.54	-155.58	3397
Niwot Ridge, Colorado, United States	NWR	40.05	-105.58	3526
Palmer Station, Antarctica, United States	PSA	-64.92	-64	10
Ragged Point, Barbados	RPB	13.17	-59.43	45
Mahe Island, Seychelles	SEY	-4.67	55.17	7
Shemya Island, Alaska, United States	SHM	52.72	174.1	40
Tutuila, American Samoa	SMO	-14.24	-170.57	42
South Pole, Antarctica, United States	SPO	-89.98	-24.8	2810
Ocean Station M, Norway	STM	66	2	5
Syowa Station, Antarctica, Japan	SYO	-69	39.58	14
Tierra Del Fuego, La Redonda Isla, Argentina	TDF	-54.87	-68.48	20
Wendover, Utah, United States	UTA	39.9	-113.72	1320
Ulaan Uul, Mongolia	UUM	44.45	111.1	914
Sede Boker, Negev Desert, Israel	WIS	31.13	34.88	400
Mt. Waliguan, Peoples Republic of China	WLG	36.29	100.9	3810
Ny-Alesund, Svalbard, Norway and Sweden	ZEP	78.9	11.88	475

Table 2. CMDL Stations with CO data used in this study.

3.2 Surface Ozone Observations - CMDL

Ozone data from the CMDL network from Bermuda, Barrow, Mauna Loa, Niwot Ridge, Samoa and the South Pole have been used. The locations are the same as in the table above. These are hourly measurements, but only the 00, 06, 12 and 18 UT observations have been used in the analysis to coincide with the frequency at which the models were sampled. Note that the model output for comparison with surface data took account of the elevation of the stations. This can be very important for mountain stations as the height of these stations can be several model levels above the surface level due to the very uneven topography in mountain areas. For more information on the data see Oltmans and Levy, 1994.

3.3 Surface Ozone Observations - EMEP

Parties to the Convention on Long-Range Transboundary Air Pollution perform ozone monitoring by UV-absorption at regional ground-based sites across Europe. These stations, about 151 (depending of the considered year), are widespread over Europe as shown by the Figure 1 (the details regarding each station location and altitude are described in annex of the D4-4 report). These sites are complementary to the CMDL network as there are far more stations but they are restricted in spatial coverage to Europe.

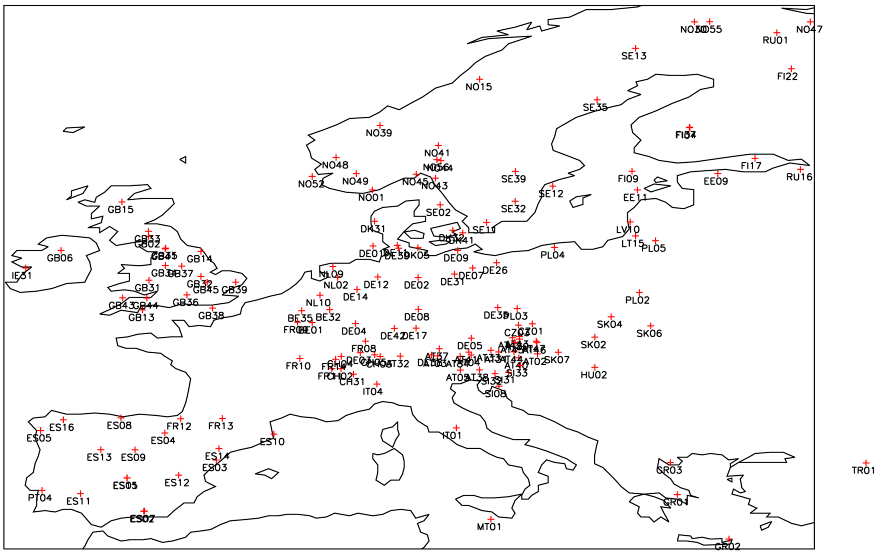

Figure 1. Distribution of the EMEP stations for ozone monitoring

3.4 Ozone sondes

Vertically resolved ozone data comes from the WOUDC ozone sonde network (http://www.woudc.org) which began to operate in 1960. It is one of the five WMO Global Atmosphere Watch (GAW) data centres. The WOUDC is operated by the Experimental Studies Division of the Meteorological Service of Canada (MSC), Environment Canada. Data is provided by a large number of agencies. From 2005-04-04 to 2006-04-04, 71 agencies contributed WOUDC data. For more details see the WOUDC website.

Data from the following stations have been used in this study. The stations were chosen on the basis of whether they provided a good data record in all the years of interest. This unfortunately leads to a fairly restricted dataset with only one station based in the Southern Hemisphere. Coverage in tropical latitudes and in the Southern Hemisphere increased only after the year 2000 when the Southern Hemisphere Additional OZone sondes (SHADOZ) network began to operate (Thompson et al., 2001/2).

Station Name	Station Abbreviation	Country	Longitude	Latitude
Resolute	RESO	Canada	-94.98	74.72
Churchill	CHUR	Canada	-94.07	58.75
Edmonton	EDMO	Canada	-114.1	53.55
Goose Bay	GOOB	Canada	-60.3	53.32
Uccle	UCCL	Belgium	4.35	50.8
Hohenpeissenberg	MOHP	Germany	11.02	47.8
Payerne	PAYE	Switzerland	6.57	46.49
Wallops Island	WALL	USA	-75.483	37.933
Tateno	TATE	Japan	140.1	36.05
Kagoshima	KAGO	Japan	130.55	31.55
Lauder	LAUD	New Zealand	169.68	-45.044

Table 3. WOUDC sites used for this study

4 Overview of model results

4.1 Global Ozone fields

Figures 2 to 5 below show the ozone concentrations for all five models, averaged over the months of January to March for the year 1997.

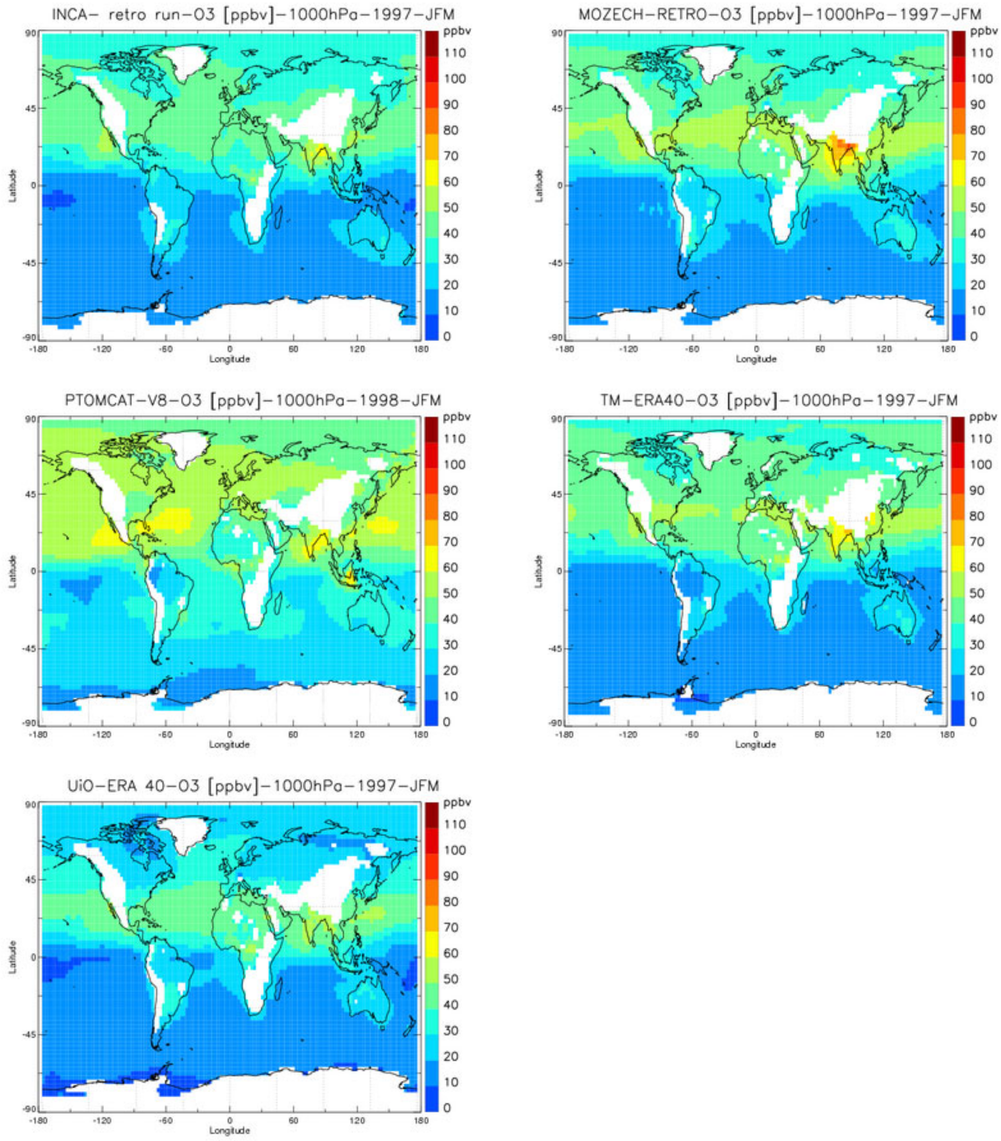

Figure 2. Model 1000 hPa Ozone fields for January-March 1997.

All models show higher surface ozone concentrations in the Northern Hemisphere than in the Southern Hemisphere and have elevated concentrations over or near India and China. These results are expected and are consistent with previous studies. The higher ozone concentrations in the Northern Hemisphere are a result of the greater emissions spread across a greater land surface area in the Northern Hemisphere. The elevated concentrations over India and China result from a combination of high solar irradiance and large emissions of ozone precursors. In central Africa all models also show a region close to the equator of high ozone concentrations which are associated with areas of biomass burning. Ozone concentrations over Australia are also higher than the background in all the models.

Although the general patterns of the models agree well, the details are significantly different between models. The gradient between the Northern and Southern Hemispheres is smallest for Oslo CTM2. This model also shows smaller ozone concentrations in the regions around China, India and the biomass burning area in Africa. In contrast, MOZECH shows the greatest concentrations in these regions which may indicate that the models have different ozone production rates for a given emission rate of ozone precursors. Oslo CTM2 has the lowest concentrations in the Arctic region, while p-TOMCAT has very little difference between middle and high latitude concentrations of surface ozone in the Northern Hemisphere.

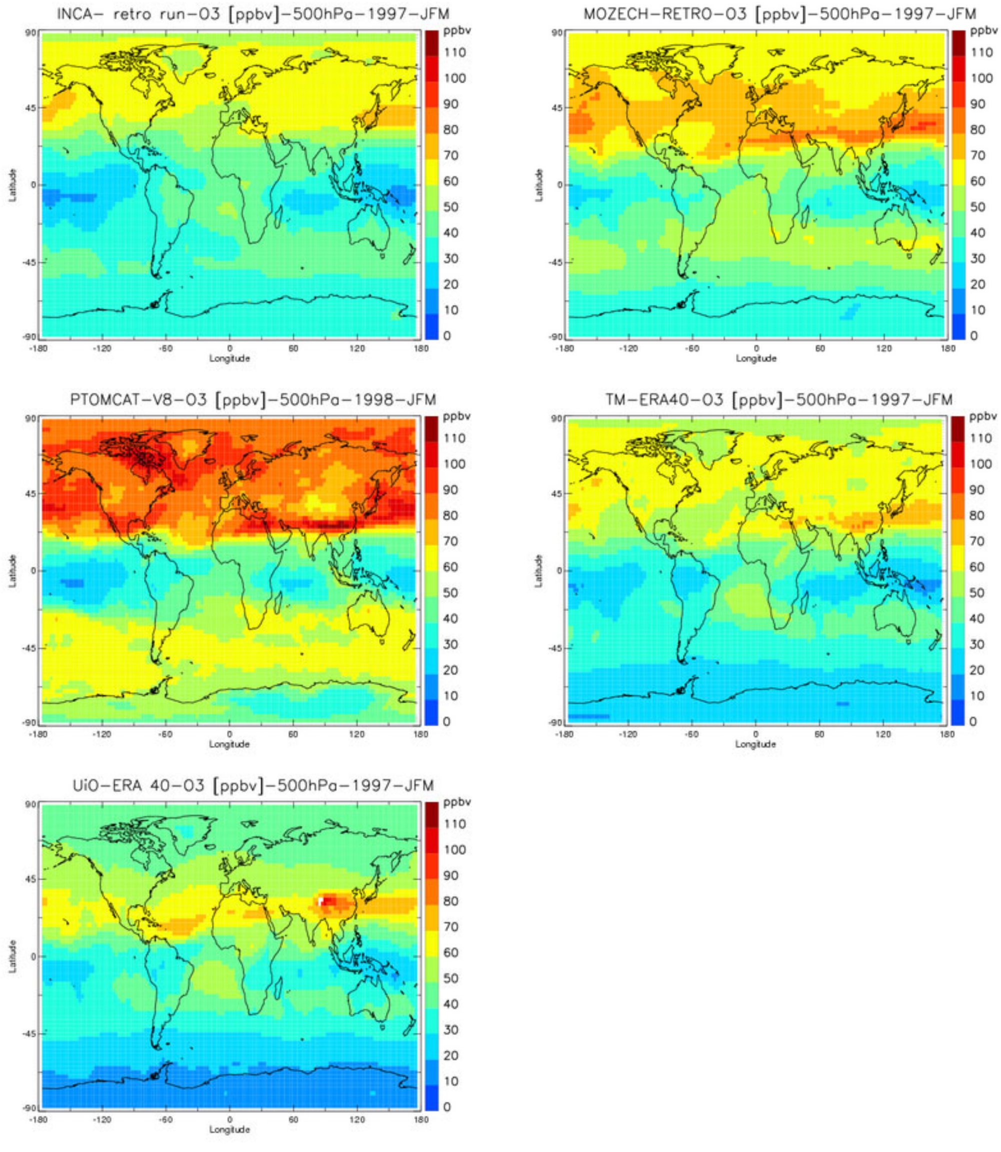

Figure 3. Model 500 hPa ozone fields for January-March 1997.

As with the surface ozone fields there are many similarities between the model results at the 500 hPa level. All models have a band of enhanced ozone concentrations at between 25° and 45° N. There is a similar but less pronounced feature in the Southern Hemisphere. They all have low concentrations in the equatorial Pacific and a region of higher concentrations off the west coast of Africa. This region of higher ozone concentrations off Africa may be the result of air with higher ozone concentrations being transported from the regions of biomass burning (cf. the analysis of the SAFARI and TRACE-A campaigns in a special issue of J. Geophys. Res., Vol. 101, NO. D19, October 30, 1996).

Again, the details of the distributions are very different. The p-TOMCAT model has much higher concentrations in the most northerly latitudes. The band of higher ozone concentrations in the Northern Hemisphere is least pronounced in MOZECH, while in p-TOMCAT the concentrations are both very high and have strong horizontal gradients. This is also seen for p-TOMCAT in the Southern Hemisphere whereas, while Oslo CTM 2 shows this feature strongly in the Northern Hemisphere, it is almost absent in the Southern Hemisphere.

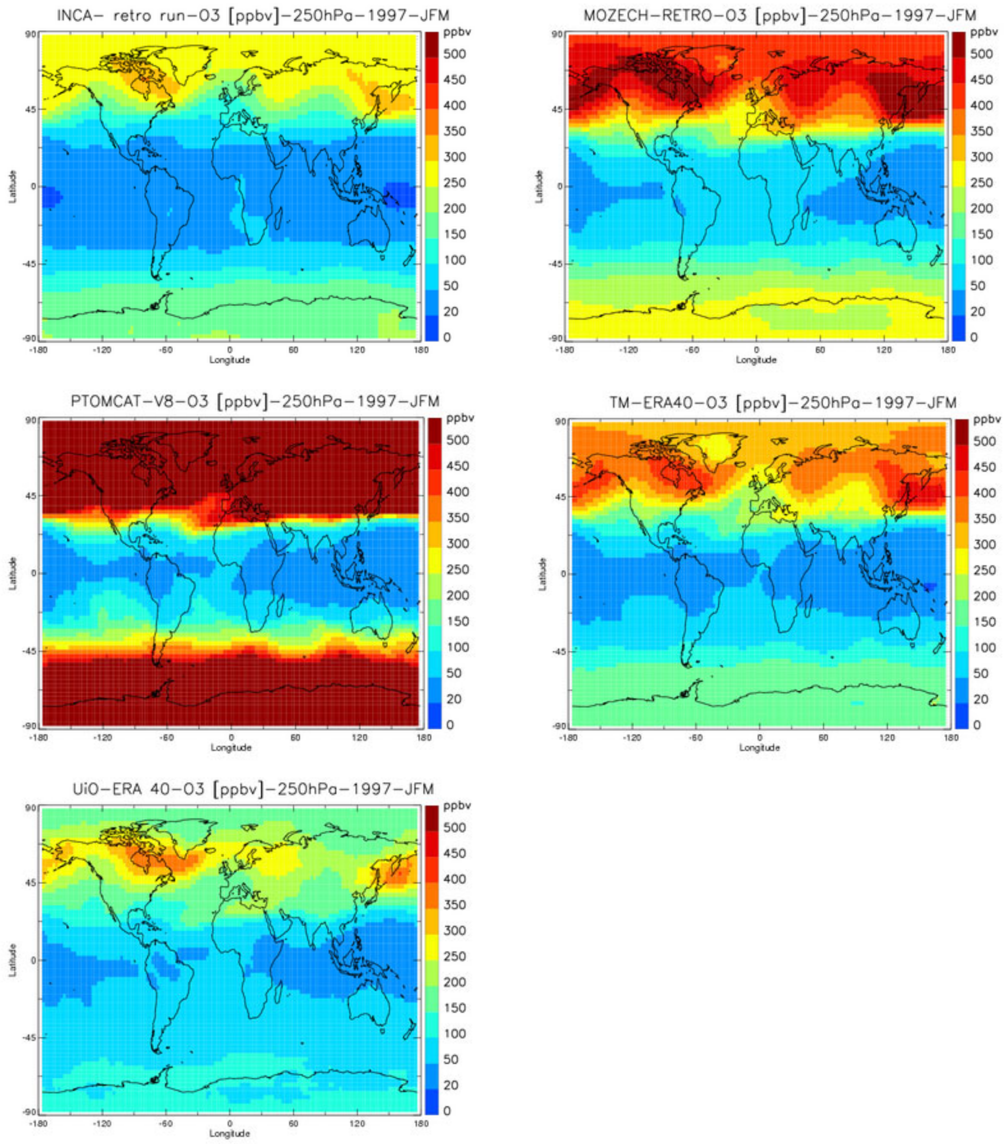

Figure 4. Model 250 hPa ozone fields for January-March 1997.

At the 250 hPa level, all models show a strong contrast between the tropics where there are low ozone concentrations (due to very high tropopause height in this region) and mid-latitudes where the ozone is higher due to the influence of stratospheric air. The models all show very different concentrations at the mid-latitudes with the highest extra-tropical ozone concentrations in MOZECH and the lowest in the Oslo CTM 2 model. This is related to transport schemes, top boundary conditions and differences in the number of model levels at this height. The main model-model difference is between p-TOMCAT which has much higher ozone concentrations at mid latitudes than the other models. The very high concentrations in p-TOMCAT may be due to stronger vertical transport, a longer lifetime for ozone in this region of the atmosphere or differences in the upper boundary condition. The other models show a wave structure in the concentrations at northern mid-latitudes with higher concentrations over Northern America, to the east of Scandinavia and just to the east of China. This may also be present in p-TOMCAT but the contour levels prevent it from being seen.

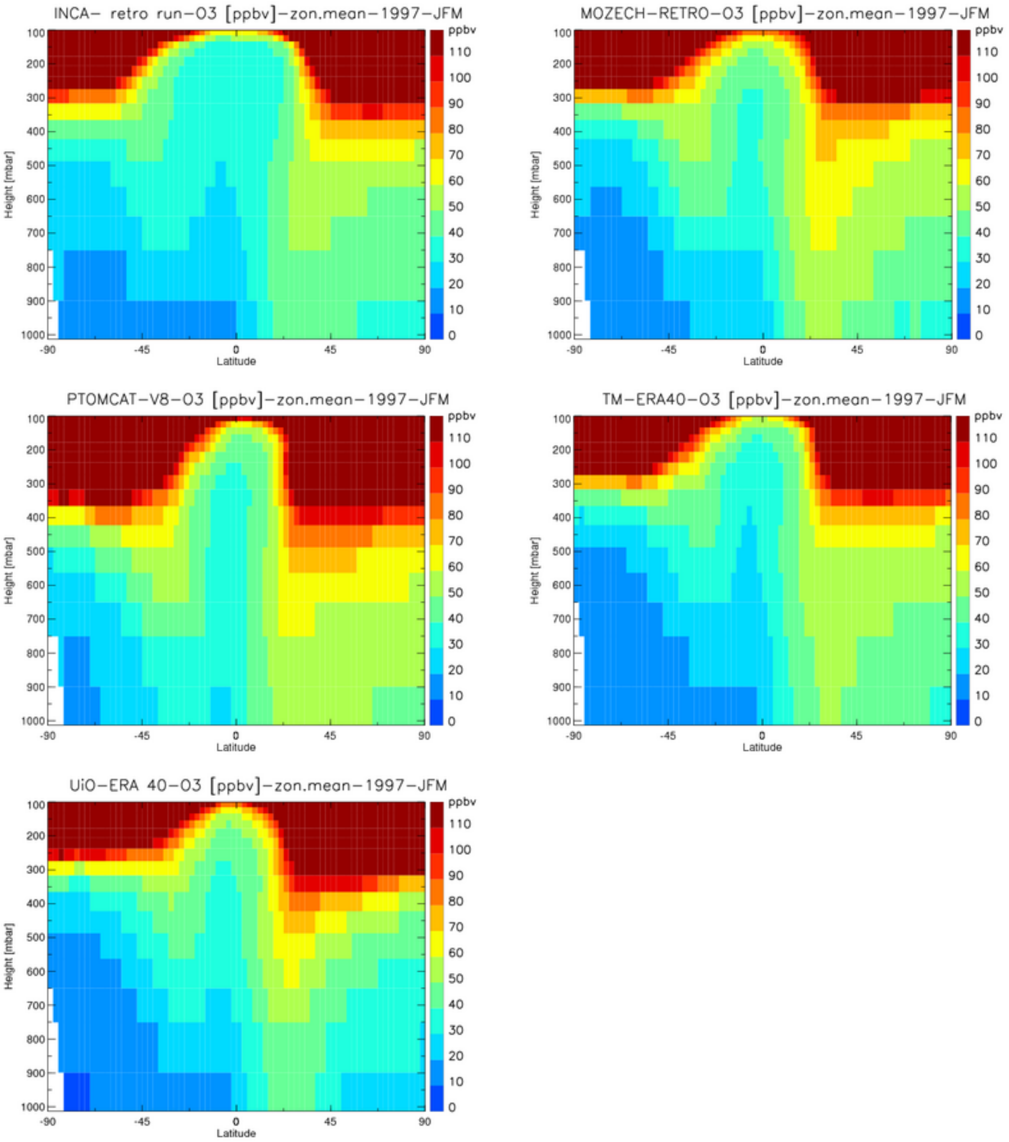

Figure 5. Model zonal mean ozone fields for January-March 1997.

All models show the expected distribution of zonal mean ozone, with the chemical tropopause (defined here as >100ppb of ozone) being higher in the tropics and lower at mid latitudes. The models also show strong downward transport of ozone at about 30° N and 30° S. The strength of this transport appears to be different between the models however. The chemical tropopause at mid latitudes in p-TOMCAT is also lower at about 400 hPa than in the other models, which is consistent with the previous field shown for 250 hPa, where there were higher ozone concentrations in mid latitudes for p-TOMCAT.

The same ozone plots as above are shown below in Figures 6-9 but averaged over the months of July-September 1997.

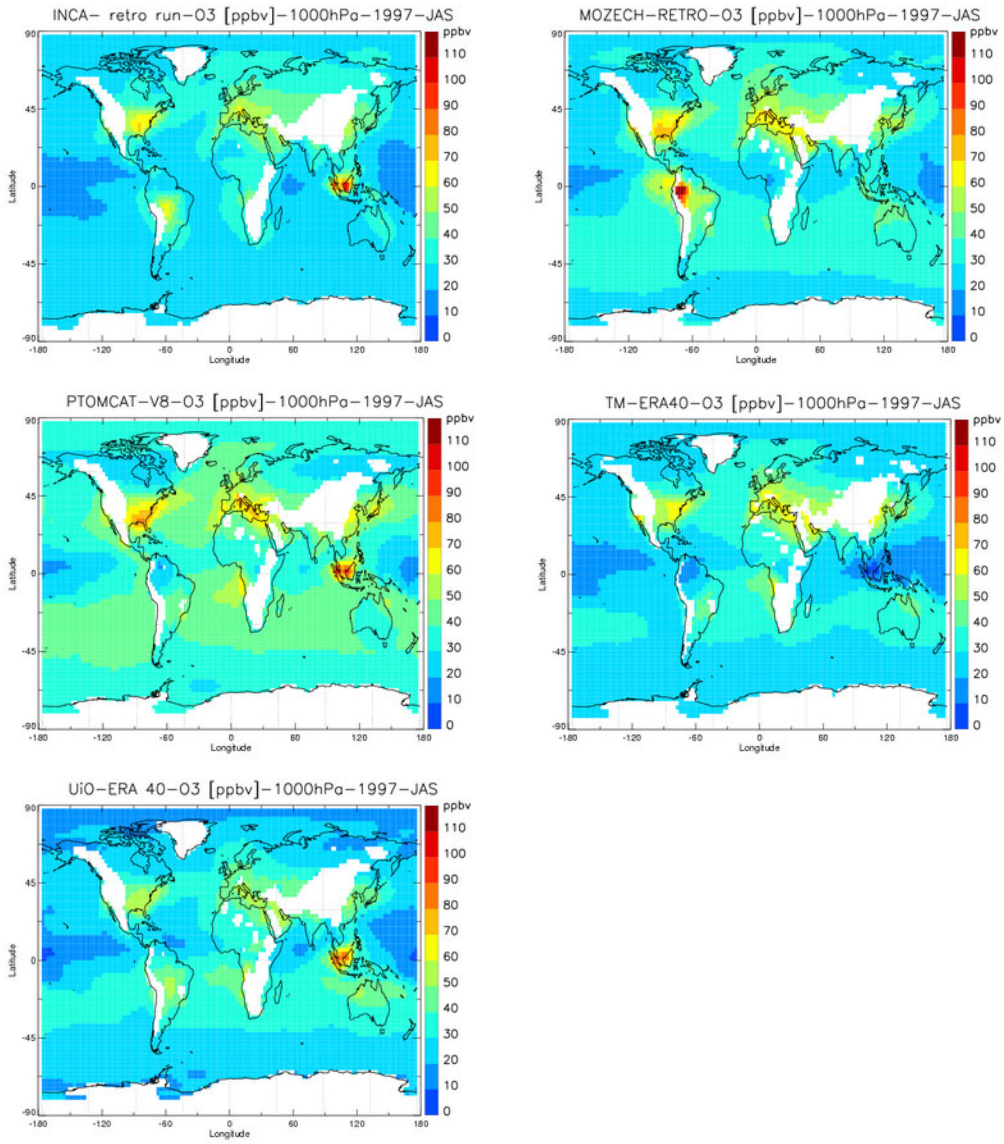

Figure 6. Model 1000 hPa Ozone fields for July-September 1997.

For the JAS season, all models show higher surface ozone concentration in polluted regions of the Northern Hemisphere than for the winter seasons, as would be expected. The ozone production in these regions for Oslo CTM2 again seems to be lower than with the other models, with TM having the highest average ozone concentrations over Europe. The effect on ozone of biomass burning in South America seems to be greatest in MOZECH, while there seem to be lower than average ozone concentrations in this region for LMDZ-INCA, p-TOMCAT and TM.

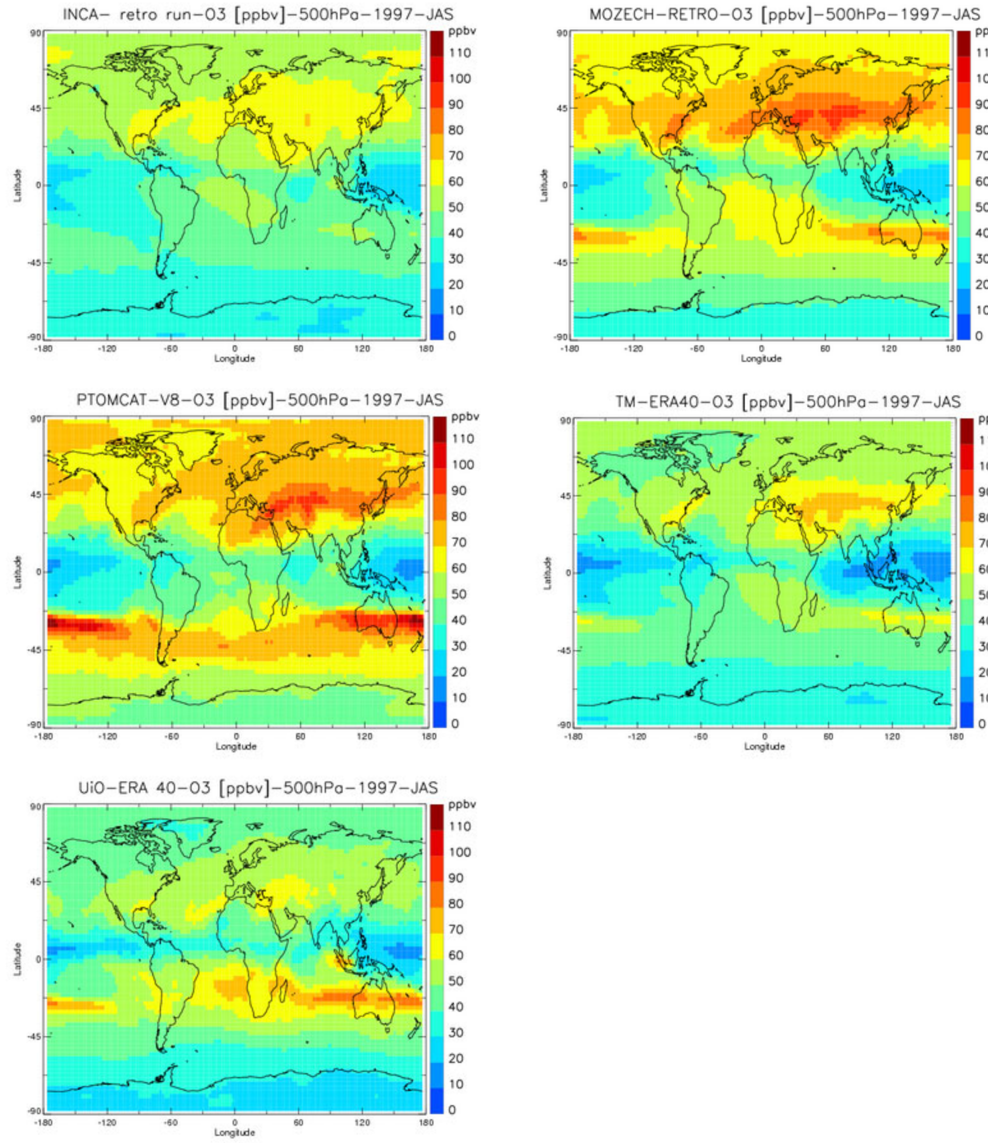

Figure 7. Model 500 hPa Ozone fields for July-September 1997.

The model results at 500 hPa are more similar to each other for the Northern Hemisphere summer than for JFM. The basic features which the models have in common are two bands of elevated ozone at around 40° N and 40° S. In the Northern Hemisphere, the concentrations in this structure are similar in the two seasons for TM, LMDZ-INCA and Oslo CTM2, whereas the concentrations are lower in p-TOMCAT and much higher in MOZECH for the JAS season. For the Southern Hemisphere, LMDZ-INCA has similar concentrations in the two seasons; TM is slightly higher in JAS than JFM; MOZECH is higher and p-TOMCAT and Oslo CTM2 are much higher.

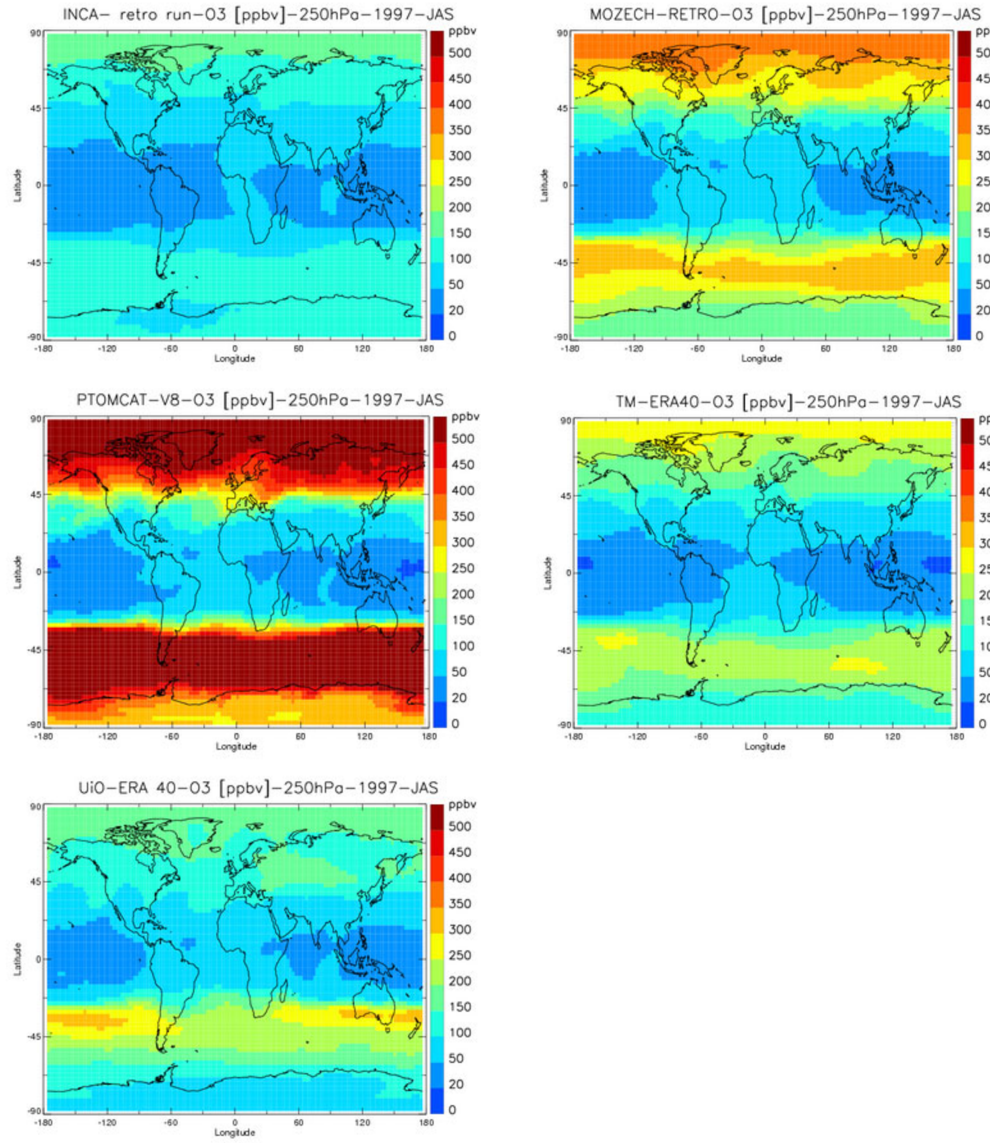

Figure 8. Model 250 hPa Ozone fields for July-September 1997.

The general ozone distribution at the 250 hPa level for JAS is similar to JFM but there are changes from one season to the other, which as for 500 hPa differ between models. In the Northern Hemisphere TM, LMDZ-INCA and Oslo CTM2 are lower in JAS. In the tropics most models have similar concentrations in the two seasons but MOZECH has higher concentrations in the JAS season over the Gulf of Guinea. In the Southern Hemisphere, TM and Oslo have increased concentrations whereas LMDZ-INCA ozone concentrations decrease. All models show evidence of the stratospheric ozone hole with lower concentrations over the South Pole than in JFM.

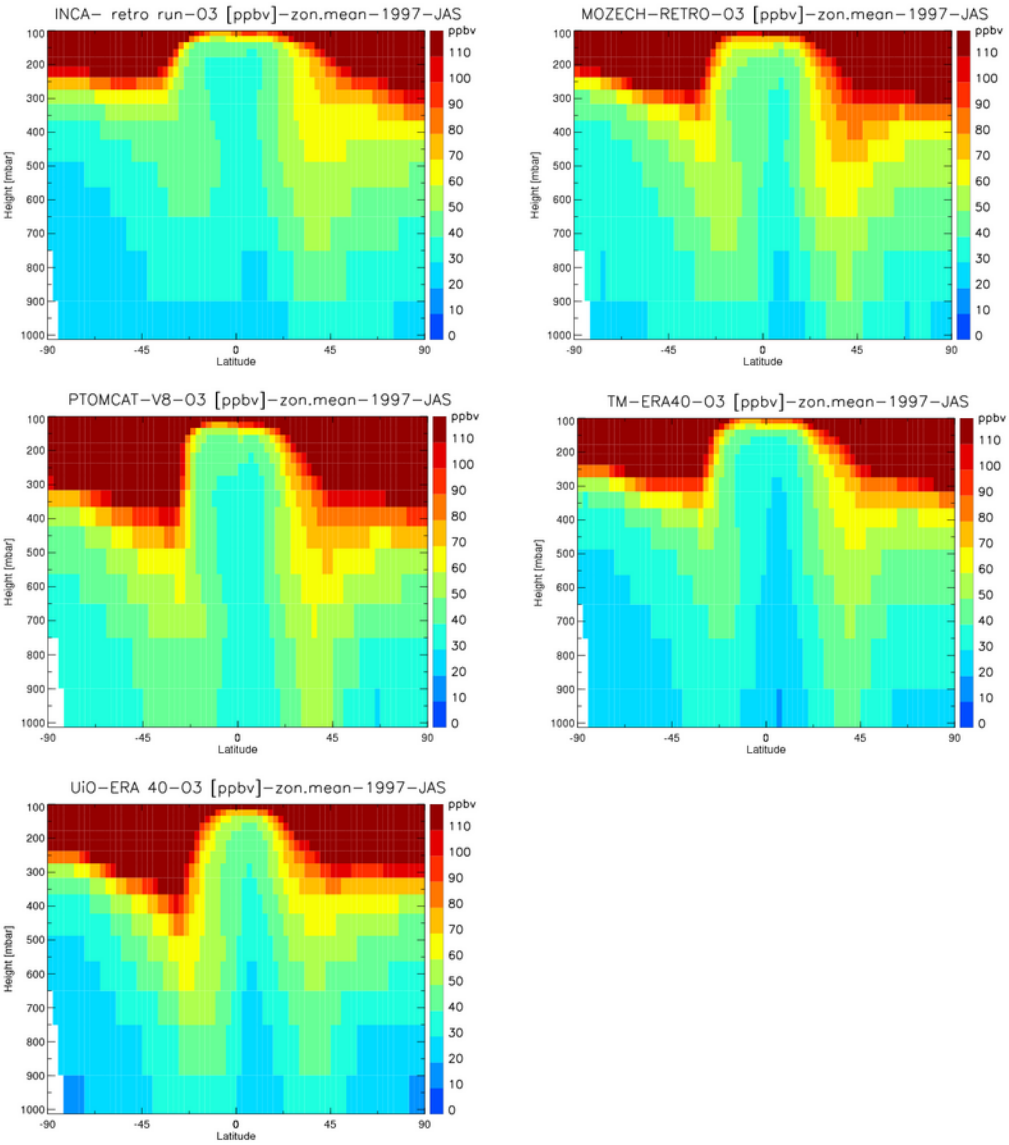

Figure 9. Model zonal mean Ozone fields for July-September 1997.

The zonal mean ozone concentrations in JAS are similar to those seen for JFM, but all models seem to have a more symmetrical pattern of downward transport at about 30° N and 30° S. As for JFM, the chemical tropopause is at much lower pressures in p-TOMCAT than the other models. The UiO model predicts a stronger stratospheric influence in the midlatitudes of the Southern Hemisphere, whereas all other models show a larger influx in the Northern Hemisphere. In all models, the pattern is more symmetric in JAS than in JFM (compare with Figure 5).

4.2 CO, NO_x and OH fields

Carbon monoxide and NOx are important chemical compounds in the troposphere both due to their own effects as air pollutants harmful to health and for their central role in the tropospheric chemistry of ozone. The OH radical is often known as the 'the detergent of the atmosphere' due to the fact that reaction with OH is the major sink of CO and other air pollutants. Therefore Figures 10, 11 and 12 show plots of these from all models for the JFM season.

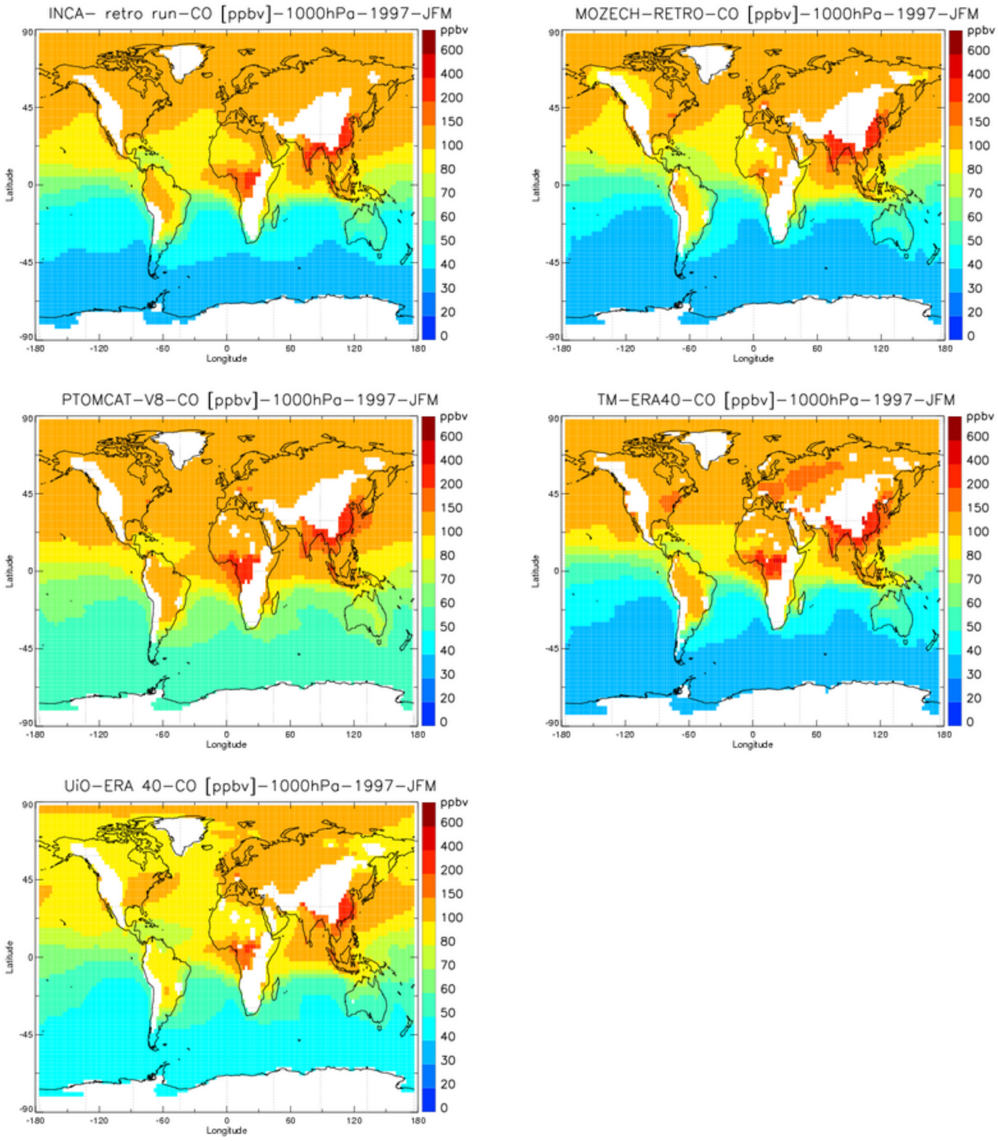

Figure 10. Modelled 1000 hPa CO concentrations for January- March. Note that fire emissions in MOZECH were erroneous.

The overall pattern of surface CO concentrations is similar in all four models. The most notable differences between the models are the high CO concentrations in the Southern Hemisphere in p-TOMCAT while the Northern Hemisphere is in better agreement with the other models. The Oslo CTM2 model also has higher concentrations than other models in the Southern Hemisphere but lower concentrations in the Northern Hemisphere, particularly over North America.

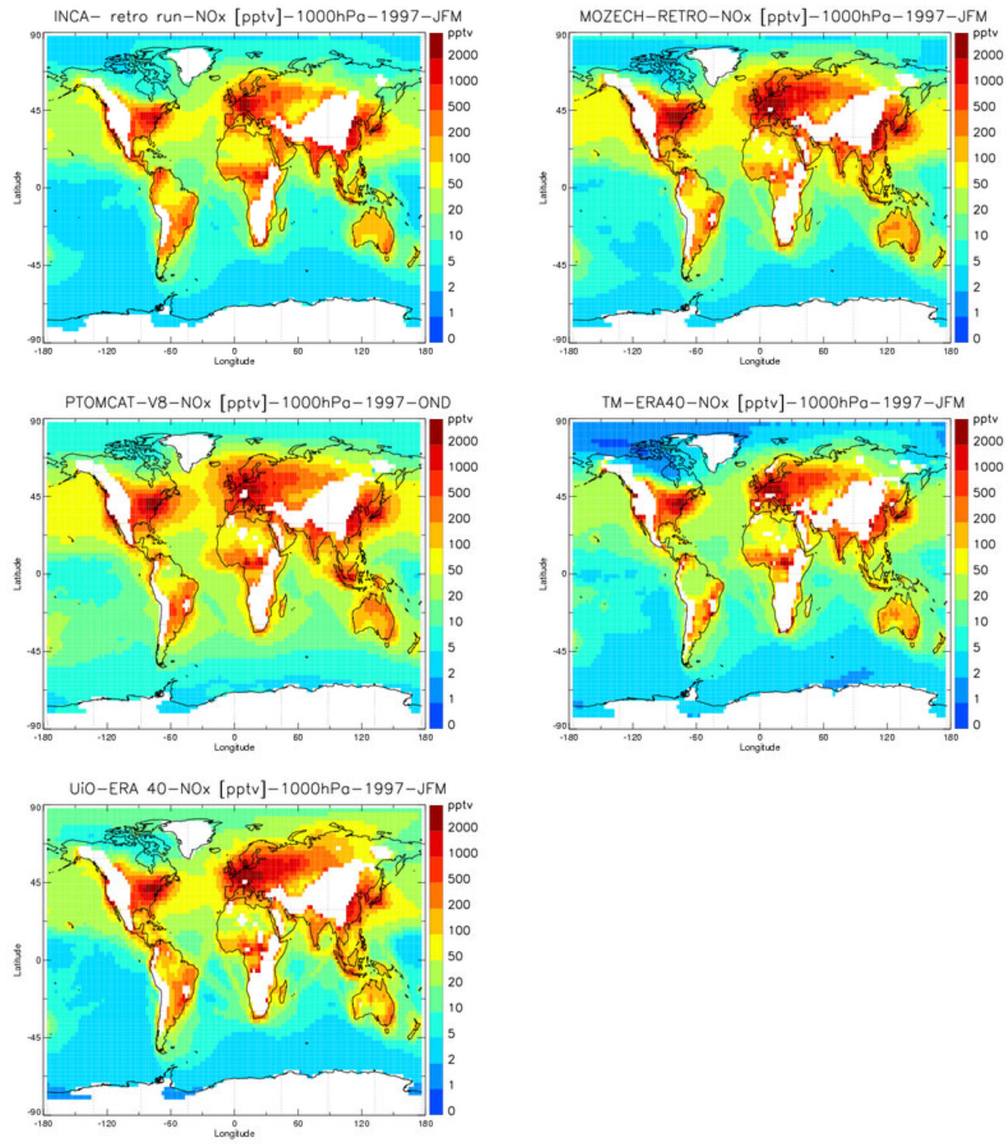

Figure 11. Modelled 1000 hPa NO_x concentrations for January- March. Note that fire emissions in MOZECH were erroneous.

For NO_x, which has a much shorter lifetime than CO, once again the overall patterns of the model fields agree well with peak values over source regions. As for CO, p-TOMCAT has the highest concentrations in the Southern Hemisphere, with Oslo CTM 2 having the highest concentrations in the Arctic region.

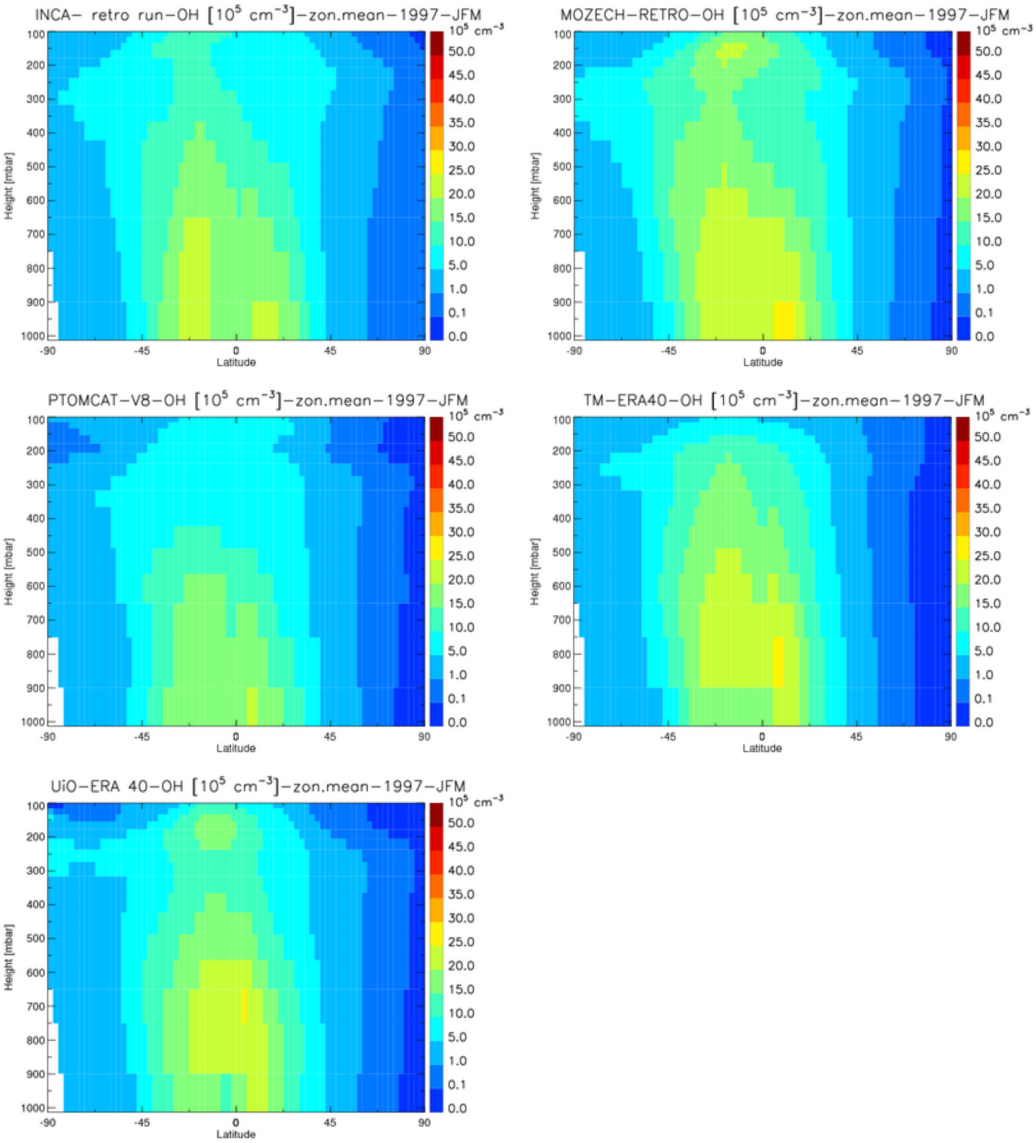

Figure 12. Modelled zonal mean OH concentrations for January-March.

Regarding the zonally averaged OH concentration fields in the Northern Hemisphere winter, the models seem to agree well with one another in mid latitudes but major differences are in the tropics with significantly lower concentrations in the LMDz-INCA and p-TOMCAT models. The lower tropical OH concentrations in p-TOMCAT help explain the higher CO concentrations in the Southern Hemisphere. Although the range of simulated OH concentrations is generally similar between all models, there are substantial differences in the chemical lifetimes for methane or methyl chloroform derived from these fields (see below).

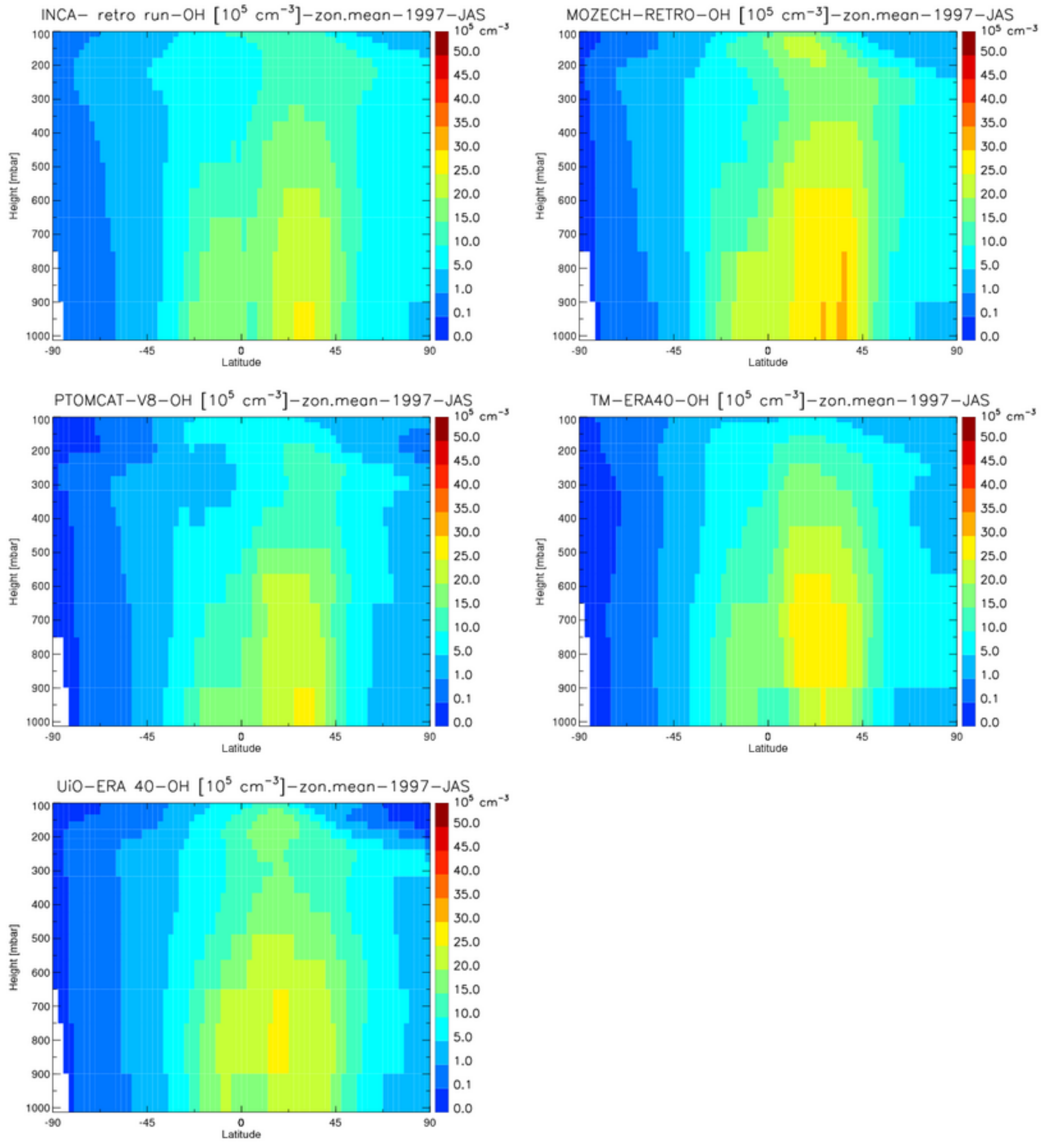

Figure 13. Modelled zonal mean OH concentrations for July-September.

For the July-to-September season, the models show a similar level of agreement for the extra tropics but again show large differences in the tropics. MOZECH now has higher tropical concentrations than the other models, and it is also notable that the differences in vertical structure in the tropics are more pronounced with a clear maximum at the surface in p-TOMCAT and LMDZ-INCA, whereas the other models have their OH maximum above the surface.

4.3 Model budget terms

In the Figures below we present for the different models the annual mean tropospheric burden and lifetime of O_3 for the years 1997–2000. For calculating tropospheric budgets, the tropopause has been defined as the 150-pbbv level of ozone following the definition in Stevenson et al., 2006. The corresponding stratosphere-troposphere exchange (STE) flux, diagnosed as residual of the other global budget terms (see Table 4), is also included in the Figure. More detailed information on the global ozone budget from all models can be found in Table 4.

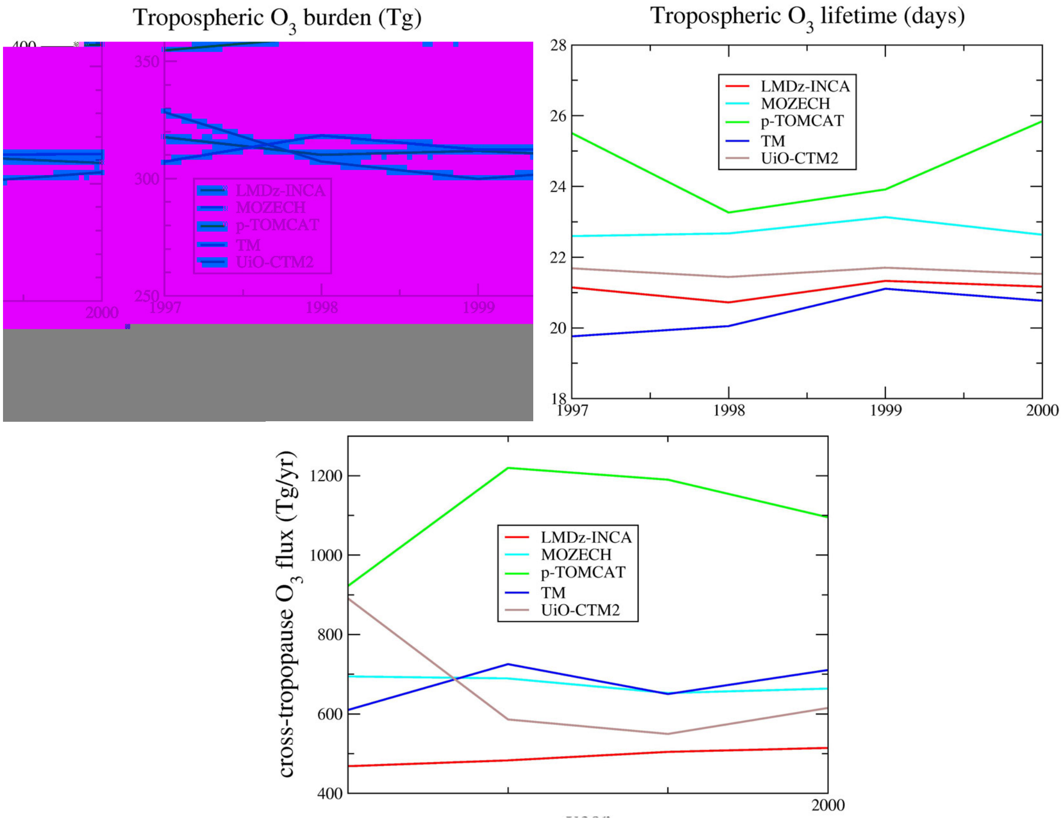

Figure 14. Tropospheric O_3 burden, lifetime and STE flux for the years 1997–2000

Variable	LMDz-INCA	MOZECH	p-TOMCAT	TM	UiO-CTM2
O_3 production (Tg/yr)	4912	5177	4209	4922	4585
O_3 destruction (Tg/yr)	4209	5058	3761	4953	3757
NET O_3 (P-L) (Tg/yr)	703	119	493	-31	828
O_3 deposition (Tg/yr)	1196	794	1564	642	1488
O_3 STE (Tg/yr)	493	675	1107	674	660
O_3 burden (Tg)	312	365	358	313	310
O_3 lifetime (days)	21.1	22.8	24.6	20.4	21.6
CO burden (Tg)	-	-	323	366	286
Mass averaged OH concentration (10^5 molec/cm^{-3})	-	-	9.77	11.47	11.40
CH_4 chemical lifetime (yr)	-	-	9.96	8.87	8.98

Table 4. Average tropospheric budgets for the years 1997–2000. The CH_4 chemical lifetime is calculated as the ratio between the atmospheric burden and the loss due to reaction with OH in the troposphere.

4.4 Conclusions

All models show distributions of ozone and its precursors which are broadly as would be expected from previous studies. However the models differ significantly in the details and to further evaluate these simulations it is necessary to compare the models to observations.

5 Validation using observational data

All comparisons of the model results with observations have been calculated using monthly mean observations. The model results have been filtered to calculate monthly mean values from only those days on which observations were made. For the surface ozone where continuous analysers were used, the data at 00, 06, 12 and 18 UT were taken to correspond with the lowest frequency of model output (from p-TOMCAT). Where there was missing data in the observation record, the models were not sampled at these times.

Several different statistical measures of model performance have been used to evaluate the models. The model skill score is a measure of deviation of model results from observations (Figure 15). It is defined as the percentage of model monthly means deviating from the observational value less than a given relative deviation. This threshold criterion is the model quality objective (MQO) and needs to be defined according to a specific analysis task.

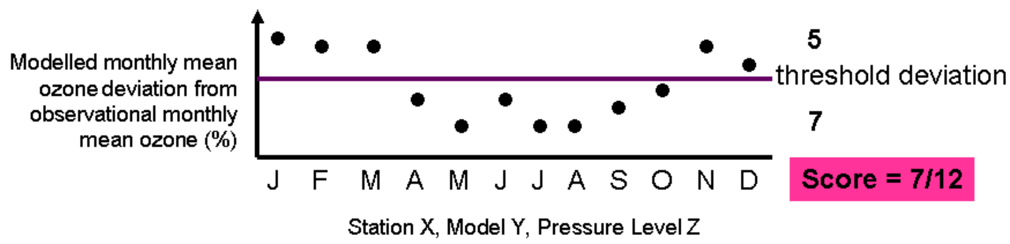

Figure 15. Definition of the model skill score.

The relative bias is a measure of deviation of model results from observations. The relative bias is calculated as the annual mean of the monthly relative deviations of model ozone from observations.

5.1 Surface Ozone - CMDL

5.1.1 Line plots and Taylor plots

As would be expected from the global fields discussed in the previous section, there is a wide range in the model predictions at all sites. Since all models are using the same emissions database and are using meteorological data based on the ERA-40 reanalysis, these differences must be due to resolution, differences in parameterisations, different chemical reaction schemes or some other model component. However to investigate in detail the reasons for the model differences would require a study in itself. We concentrate here on attempting to evaluate the model results with the data.

There is strong evidence for the impact of halogen chemistry on the Arctic ozone concentrations. For example, Barrow observations show springtime concentrations being well below the range of all model calculations. The raw hourly data contain many periods when zero concentrations of ozone are observed and this is consistent with previous observations (e.g. Bottenheim et al, 2002) which have explained this as a result of a rapid increase in Bromine concentrations. As none of the models contain these chemical reactions, they are unable to reproduce these very low ozone events and so the monthly means are too high.

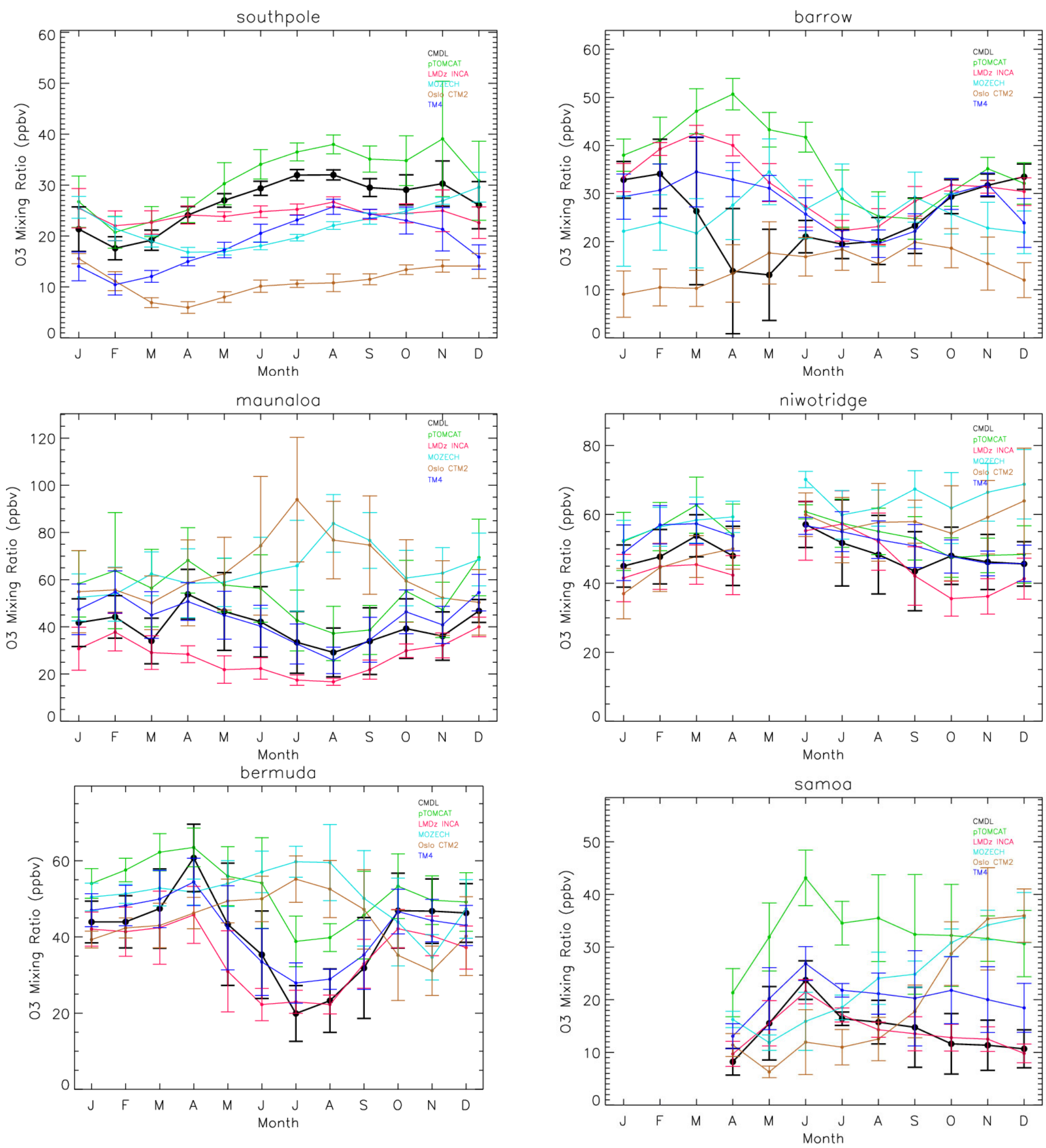

Figure 16. Model comparison to surface ozone data for 1997.

It is not possible to say that any one model shows clearly better results than all others. The models generally capture the seasonal cycle of the observations although at almost all sites the concentrations in p-TOMCAT are much too high. MOZECH generally also exhibits a high bias, whereas LMDz-INCA shows a tendency to underestimate ozone. The ozone peak in July/August at Mauna Loa seen in MOZECH and Oslo CTM is not observed in either the other models or the measurements. Figure 17 shows Taylor plots for the same data. The Taylor plot (Taylor, 2001, Brunner et al, 2003) is a means of displaying two pieces of information about model performance on the same figure: correlation coefficient and standard deviation relative to the observations. A model with a correlation of 1 with the observations would lie on the x-axis and a model with the same standard deviation as the observations would lie on the dotted quarter circle. A 'perfect' model would lie a the point indicated by the small circle on the x-axis. For some of these stations some of the models have either a strong negative correlation or a very much larger standard deviation than the observations and so they do not appear at all on the plot. Due to the influence of bromine chemistry all models show poor performance at Barrow. There does not appear to be a consistent picture of model performance with a model that does well at one station (e.g. p-TOMCAT at Samoa) doing very badly at another (e.g. at Barrow). However in general LMDZ, TM4 and p-TOMCAT seem to give the best results at these stations for ozone.

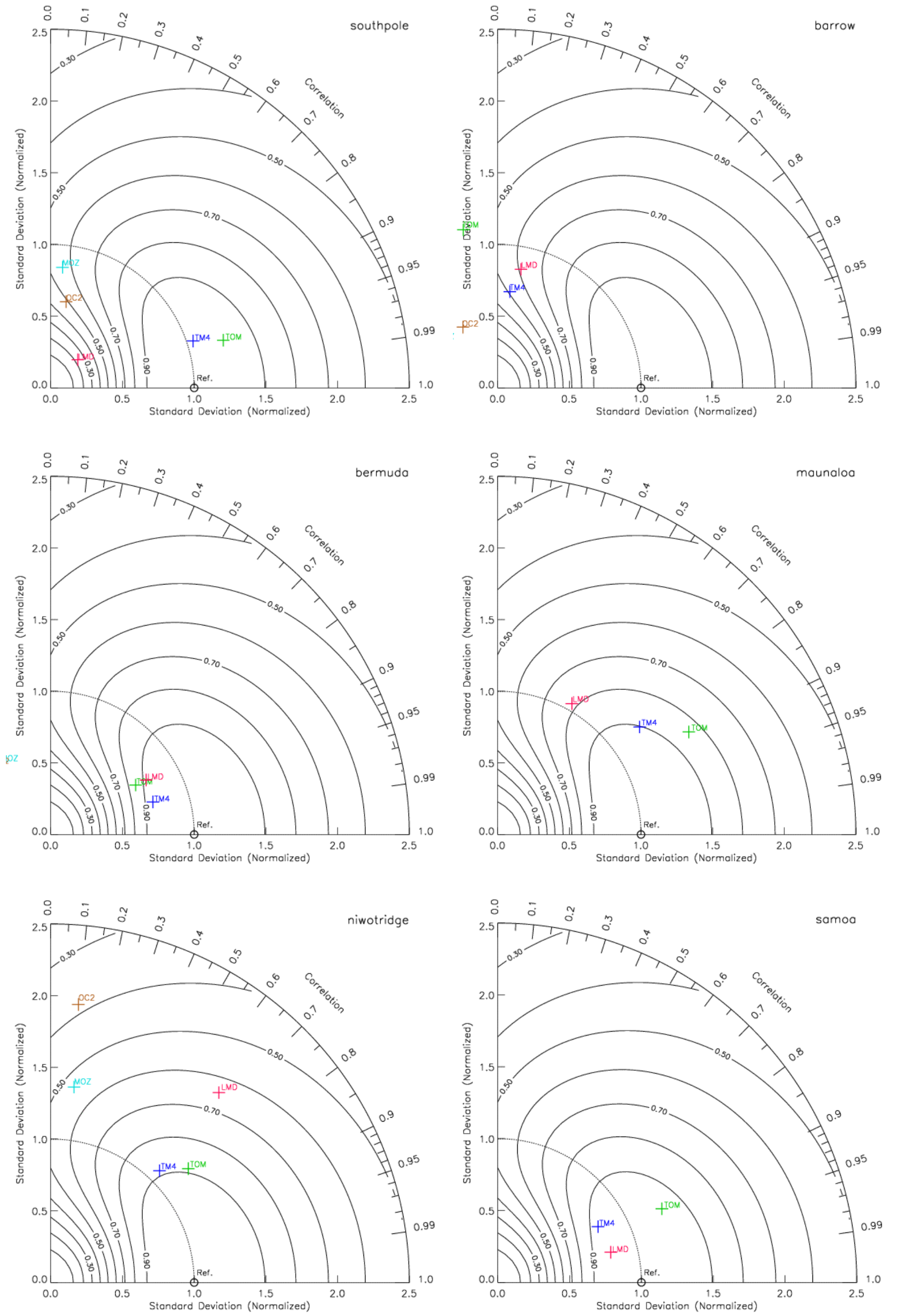

Figure 17. Taylor plots for surface ozone data for 1997.

5.1.2 Statistics

The first statistic we make use of in this section is a MQO of 20%. This means that the simulated ozone concentrations are expected to be within 20% of the observed values, and an individual monthly mean value is counted as success only if this is the case. Table 5 shows the model skill scores at all the surface ozone stations for 1997, while Table 6 shows the average skill scores at all stations for all years.

Station	num	pTOMCAT	LMDz-INCA	MOZECH	UiO	TM	model mean
Bermuda	12	0.33	0.75	0.5	0.42	0.83	0.57
Barrow	12	0.42	0.58	0.17	0.33	0.58	0.42
Mauna Loa	12	0.08	0.33	0.08	0.17	0.83	0.3
Niwot Ridge	11	0.91	0.82	0.36	0.73	1	0.76
Samoa	9	0	1	0.11	0	0.11	0.24
S. Pole	12	0.83	0.83	0.33	0	0.17	0.43
Mean		0.43	0.72	0.26	0.27	0.59	0.45

Table 5. Fraction of months where the model is within 20% of observations:1997

	pTOMCAT	LMDz-INCA	MOZECH	UiO	TM	model mean
1997	0.43	0.72	0.26	0.27	0.59	0.45
1998	0.52	0.65	0.46	0.43	0.61	0.53
1999	0.31	0.52	0.25	0.25	0.4	0.35
2000	0.17	0.49	0.17	0.24	0.43	0.3
Mean	0.36	0.60	0.29	0.30	0.51	0.41

Table 6. Average skill scores for a model quality objective of 20% agreement over all surface ozone stations

Colour code:

(red)	0.9 - 1.0
(orange)	0.8 - 0.89
(yellow)	0.7 - 0.79

On average, the models can only predict observed monthly ozone concentrations to within 20% less than half of the time. The performance of LMDz seems better than the other models', but all models have a strong variation from year to year with the mean of the models going from 0.3 in 2000 to 0.53 in 1998. For example, all models seem to do relatively well at Niwot Ridge, but all except TM have poor performance at Mauna Loa.

station	pTOMCAT	LMDz-INCA	MOZECH	UiO	TM	model mean
Bermuda	0.86	0.87	-0.56	-0.63	0.95	0.3
Barrow	-0.21	0.19	-0.78	-0.5	0.12	-0.23
Mauna Loa	0.88	0.49	-0.62	-0.45	0.8	0.22
Niwot Ridge	0.77	0.66	0.12	0.1	0.7	0.47
Samoa	0.91	0.97	-0.47	-0.48	0.87	0.36
S. Pole	0.96	0.69	0.1	0.18	0.95	0.58
Mean	0.7	0.65	-0.37	-0.3	0.73	0.28

Table 7. Correlation coefficients for 1997.

station	pTOMCAT	LMDz-INCA	MOZECH	UiO	TM	model mean
1997	0.7	0.65	-0.37	-0.3	0.73	0.28
1998	0.65	0.55	0.12	0.19	0.67	0.44
1999	0.55	0.42	0.15	0.17	0.6	0.38
2000	0.51	0.34	0.19	0.46	0.52	0.4
Mean	0.60	0.49	0.02	0.13	0.63	0.38

Table 8. Average correlations for other years.

	$r \geq 0.9$
	$0.8 \leq r < 0.9$
	$0.7 \leq r < 0.8$
	$0.6 \leq r < 0.7$
	$r < 0.6$

With the correlations in 1997 there seem to be two groups of models - p-TOMCAT, LMDZ and TM all have reasonably good correlations at all these sites while the correlation of MOZECH and Oslo CTM2 are poor at all sites. 1997 seems to have on average better correlations for the models which have good correlations, but is the worst year for the other two models.

Station	pTOMCAT	LMDz-INCA	MOZECH	UiO	TM	model mean
Bermuda	33.98	-11.67	41.48	24.01	5.55	18.67
Barrow	66.59	41.03	20.01	-32.2	23.32	23.75
Mauna Loa	35.24	-31.51	66.75	66.08	7.69	28.85
Niwot Ridge	11.46	-7.78	28.31	10.48	7	9.89
Samoa	138.37	0.9	86.15	54.71	49.32	65.89
S. Pole	17.78	-5.58	-13.94	-57.16	-30.82	-17.94
Mean	50.57	-2.43	38.13	10.98	10.34	21.52

Table 9. Annual mean relative biases at all stations for 1997 (percent)

station	pTOMCAT	LMDz-INCA	MOZECH	UiO	TM	model mean
1997	50.57	-2.43	38.13	10.98	10.34	21.52
1998	46.83	-5.29	22.82	-5.9	11.46	13.98
1999	66.64	4.39	29.05	-15.35	12.39	19.42
2000	82.7	13.3	48.08	-2.5	25.2	33.36
Mean	61.69	2.49	34.52	-3.19	14.85	22.07

Table 10. Average biases for all years.

Colour code:

	$< \pm 5$ %
	$\pm 5 < \text{dev} < \pm 10$ %
	$\pm 10 < \text{dev} < \pm 20$ %
	$\pm 20 < \text{dev} < \pm 30$ %

p-TOMCAT clearly is the model which has the largest bias with an average positive bias of over 50 pbb in 1997, a positive bias of at least 10 ppb at all sites and a huge 138% positive bias at Samoa. This is consistent with the evidence of strong stratosphere to troposphere transport in the p-TOMCAT global model plots. MOZECH also has a clear positive bias but this can be seen to be smaller than in p-TOMCAT. All models except TM have large positive or negative biases at Mauna Loa and all models except LMDz have large positive biases at Samoa. The average bias is smallest at Niwot Ridge.

5.2 EMEP surface O_3 observations over the 1990-2000 period

The measurements as well as the model results are averaged with a three day shifting mean in order to remove the high frequency variations. Figure 18 summarizes, with Taylor diagrams, the correlation and normalized standard deviations (i.e. σ_mod/σ_obs) obtained, at each EMEP stations, for each of the five models for the year 1997. For this year, the LMDz-INCA model has the lowest dispersion of the points indicating in particular its ability to reproduce the standard deviation of the ozone observations. The results of the UiO, MOZECH and TM4 models are also fairly good whereas the pTOMCAT results are quite scattered with some poor correlations at several stations.

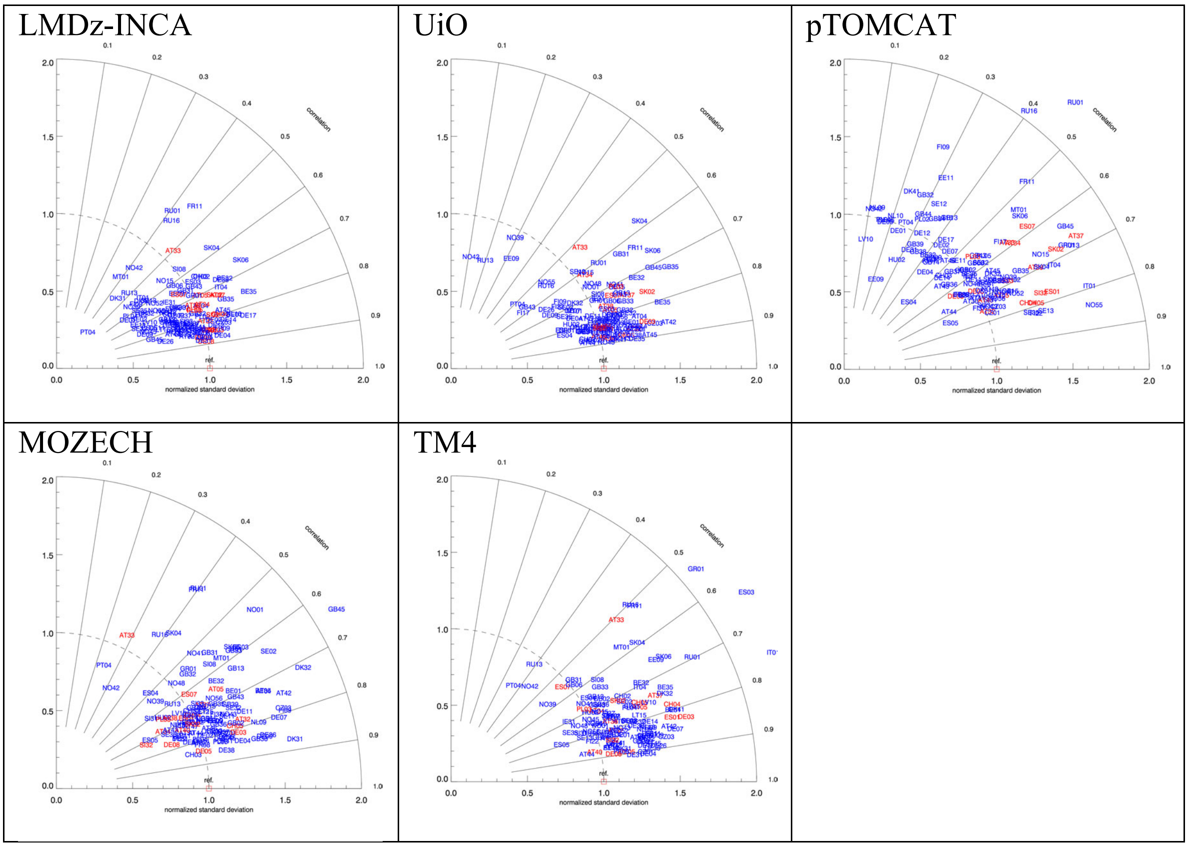

Figure 18: Taylor diagrams for 1997 with 3-day time filtered ozone mixing ratios

For a more quantitative analysis, the statistical results averaged over the EMEP stations for each year are displayed in Table 11.

Criteria for good results highlighting		
Abs. bias <10ppb	σ_mod/σ_obs in [0,9 ; 1,1]	Correlation >0,7

UiO	Abs bias (ppb)	σ_mod/σ_obs	Correlation	Skill score for threshold 20%	Skill score for threshold 30%
1996	7,39	0,94	0,69	0,49	0,67
1997	6,74	1,06	0,77	0,53	0,70
1999	7,30	0,99	0,67	0,52	0,71
2000	6,87	0,97	0,69	0,53	0,71
MOZECH	bias_abs	sdv_mod/std_obs	Correlation	Threshold 20%	Threshold 30%
1990	12,68	1,19	0,72	0,24	0,36
1991	11,84	1,20	0,68	0,28	0,42
1992	13,16	1,06	0,73	0,25	0,37
1993	12,45	1,16	0,73	0,28	0,41
1994	12,23	1,10	0,72	0,29	0,43
1995	11,93	1,17	0,67	0,31	0,45
1996	11,11	1,11	0,68	0,34	0,48
1997	11,08	1,13	0,70	0,34	0,48
1998	10,62	1,14	0,69	0,36	0,52
1999	10,59	1,11	0,66	0,38	0,54
2000	11,85	1,14	0,68	0,32	0,46
TOMCAT	bias_abs	sdv_mod/std_obs	Correlation	Threshold 20%	Threshold 30%
1997	15,72	1,14	0,56	0,23	0,34
1998	15,63	1,31	0,52	0,24	0,35
1999	16,03	1,35	0,54	0,23	0,35
2000	15,77	1,18	0,52	0,22	0,33
INCA	bias_abs	sdv_mod/std_obs	Correlation	Threshold 20%	Threshold 30%
1990	11,25	0,96	0,74	0,30	0,43
1991	11,71	0,99	0,70	0,30	0,43
1992	11,72	0,98	0,75	0,30	0,44
1993	10,78	0,91	0,74	0,34	0,47
1994	10,78	1,06	0,76	0,36	0,50
1995	10,05	1,01	0,74	0,39	0,52
1996	10,63	0,99	0,72	0,37	0,51
1997	9,59	0,92	0,73	0,41	0,54
1998	9,62	1,02	0,69	0,42	0,56
1999	8,98	1,02	0,72	0,46	0,60
2000	9,25	1,03	0,72	0,44	0,58
TM4	bias_abs	sdv_mod/std_obs	Correlation	Threshold 20%	Threshold 30%
1990	12,08	1,13	0,78	0,23	0,36
1991	12,13	1,13	0,73	0,25	0,37
1992	13,09	1,12	0,82	0,21	0,34
1993	12,16	1,11	0,78	0,26	0,40
1994	12,77	1,20	0,80	0,24	0,38
1995	12,78	1,22	0,76	0,25	0,38
1996	12,07	1,17	0,75	0,29	0,42
1997	11,46	1,23	0,77	0,30	0,45
1998	10,83	1,34	0,73	0,34	0,49
1999	10,36	1,23	0,74	0,38	0,53
2000	10,91	1,19	0,73	0,35	0,49

Table 11. Summary statistics for whole years of the 90s with 3-day time filtered ozone mixing ratios.

These results show that the models reproduce the European 3 daytime filtered ozone with an absolute bias (one year average) around 10 ppbv (ranging between 6.7 and 16 ppbv with the lowest bias obtained with UiO). Except for the UiO model, all models exhibit a positive bias and thus overestimate the ozone mixing ratios in general. Regarding the standard deviation ratio (σmod/σobs), showing the ability of the model to reproduce the amplitude of ozone variations, p-TOMCAT, TM4 and MOZECH overestimate this ratio whereas UiO and LMDz-INCA show an averaged ratio closed to one but resulting to some extent from error compensations. The correlations lie between 0.53 for TOMCAT and 0.76 for TM4 (0.73=LMDz-INCA=0.73; UiO=0.71; MOZECH=0.70).

5.3 Surface Carbon Monoxide - CMDL

5.3.1 Line plots and Taylor diagrams

Given the large number of stations used for this analysis and the four years of data the models are compared to, it is not practical to show the data for all sites and years. Instead a sample of stations is chosen to illustrate the models performance for 1997, then statistics at all stations are presented for this year and finally the mean values of the statistics are presented for all years as for surface ozone.

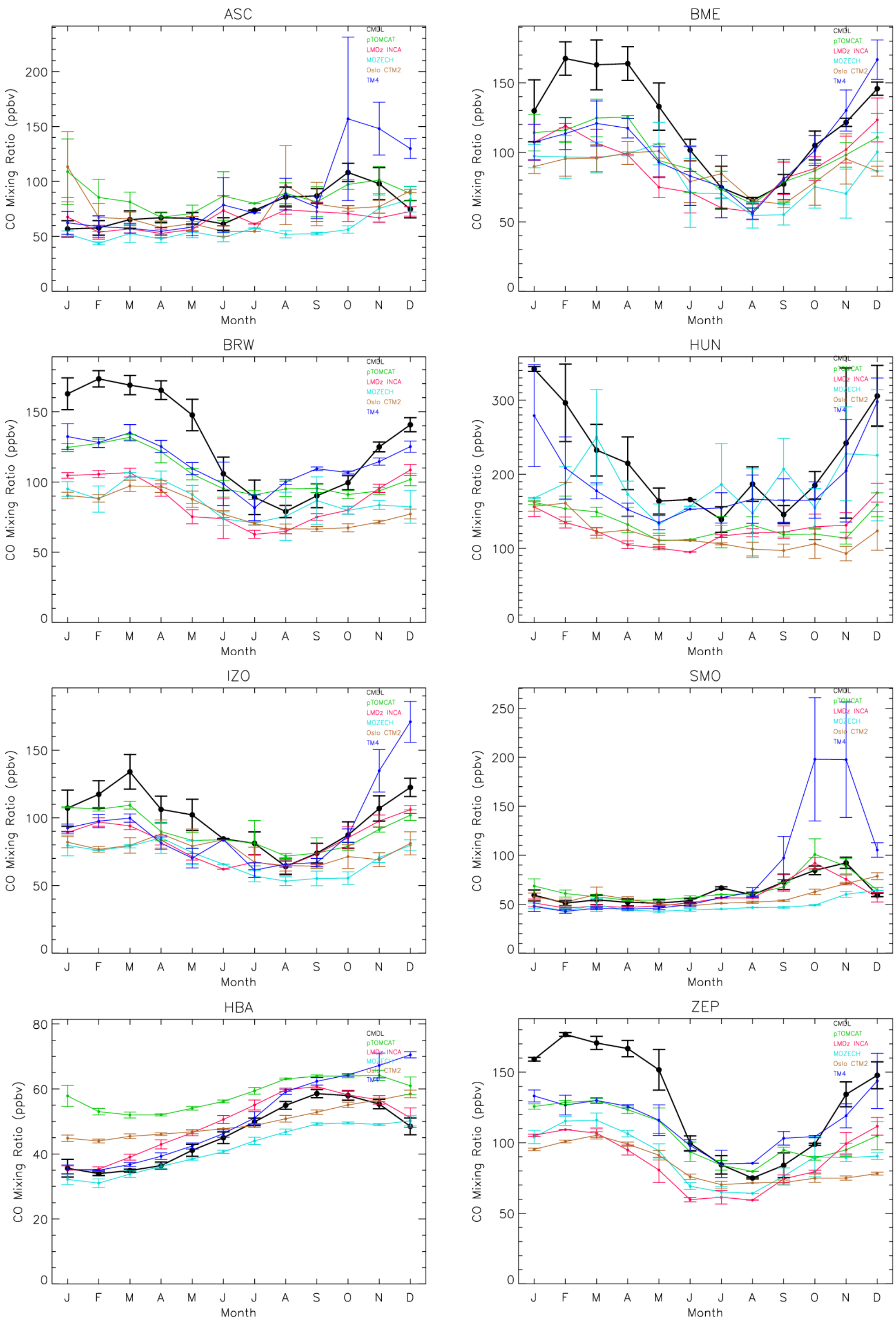

Figure 19. CO concentrations at selected sites for 1997.

It can been seen from these plots that although the models generally reproduce well the general timing of the seasonal cycles, there is a large range in the modelled concentrations and at a large number of sites the model concentrations are all consistently smaller than the observations. The most dramatic signal is the response to large biomass burning in Indonesia seen in the TM model at the Samoa (SMO) site. This is discussed in more detail in the report on process studies (D3-4). CO concentrations in TM are in general higher than in the other models, but at many sites still do not have concentrations as large as in the observations. The exception to this seems to be in the latter months of the year when at some Southern Hemisphere sites TM does have much larger CO concentrations probably due to the influence of the Indonesian fires in this year. There is different picture at the Antarctic site (Halley Bay, HBA) where p-TOMCAT has much larger concentrations than the other models and the observations. The other models are in better agreement with the data for most months but there is an increase in CO in the last months of the year which is not seen in the models (or in p-TOMCAT).

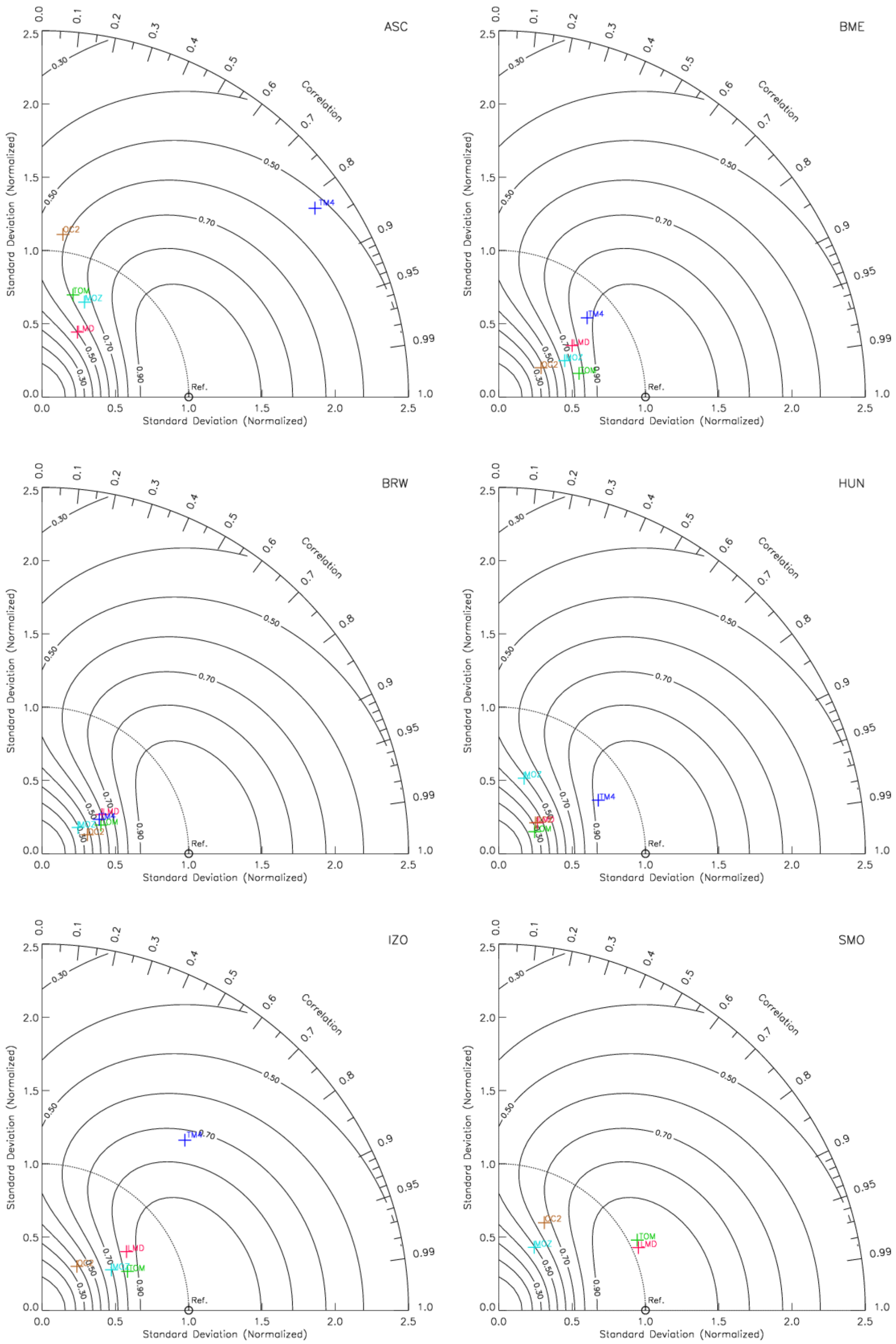

Figure 20a. Taylor plots for surface CO data for 1997 at selected sites.

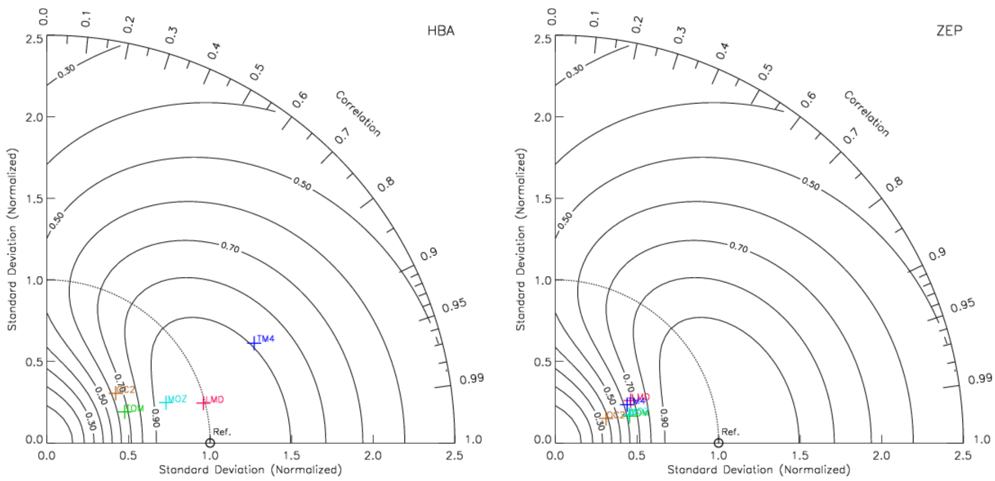

Figure 20b. Taylor plots for surface CO data for 1997 at selected sites.

In general, the models lie much closer together on this Taylor plot than on the one for surface ozone, which is probably a result of the fact that the tropospheric chemistry of ozone is more complex than that of carbon monoxide. In particular, at Barrow (BRW), it can be seen that the models lie very close together with high correlation coefficients but a standard deviation (e.g. amplitude of the seasonal variation) only about half of the observations. In contrast, at Ascension Island (ASC) the models are much more widely spread and have lower correlations with the observations. The TM model is something of an outlier at this site with a far larger standard deviation than the observations but still a fairly good correlation coefficient. At Samoa, the TM standard deviation is so large it is off the edge of this plot.

5.3.2 Statistics

We examine here the same statistics for CO as for ozone in the previous section. First the model skill scores for a quality objective of a 20% agreement is examined.

station	num	pTOMCAT	LMDz INCA	MOZECH	UiO	TM	model mean
ALT	11	0.27	0.18	0.18	0.27	0.36	0.25
ASC	12	0.67	0.75	0.42	0.58	0.67	0.62
ASK	12	0.58	0.58	0.08	0.17	0.33	0.35
AZR	12	0.58	0.17	0.08	0.33	0.67	0.37
BAL	12	0.58	0.33	0.17	0.25	0.33	0.33
BME	12	0.58	0.58	0.17	0.25	0.67	0.45
BMW	8	0.25	0.38	0.25	0.12	0.5	0.3
BRW	12	0.33	0.25	0.25	0.08	0.5	0.28
BSC	11	0	0	0.09	0	0.36	0.09
CBA	12	0.33	0.08	0.08	0.08	0.5	0.22
CGO	11	0.45	1	0.82	0.73	0.82	0.76
CHR	2	0.5	1	0.5	0.5	0.5	0.6
CRZ	7	0.71	1	0.57	0.86	1	0.83
EIC	11	0.82	1	0.27	0.82	0.82	0.75
GMI	12	0.5	0.42	0.08	0.33	0.25	0.32
HBA	12	0.42	1	1	0.58	0.83	0.77

HUN	12	0.17	0.17	0.5	0	0.75	0.32
station	num	pTOMCAT	LMDz INCA	MOZECH	UiO	TM	model mean
ICE	11	0.55	0.45	0.18	0.27	0.45	0.38
IZO	12	1	0.67	0.17	0.5	0.5	0.57
KEY	11	0.91	0.18	0.36	0.36	0.55	0.47
KUM	12	0.67	0.5	0.08	0.33	0.5	0.42
KZD	3	0	0	0	0	0.67	0.13
KZM	3	0.67	1	0	0	0.67	0.47
LEF	12	0.33	0.25	0.33	0.25	0.58	0.35
MHD	12	0.5	0.42	0.5	0.33	0.5	0.45
MID	12	0.25	0	0.08	0.25	0.58	0.23
MLO	12	0.67	0.33	0	0.17	0.25	0.28
NWR	12	0.58	0.67	0.25	0.08	0.67	0.45
PSA	12	0.42	1	1	0.67	0.83	0.78
RPB	12	0.83	0.5	0.08	0.42	0.33	0.43
SEY	8	0.88	0.75	0.25	0.62	0.62	0.62
SHM	11	0.36	0	0.09	0.09	0.55	0.22
SMO	12	1	1	0.58	0.58	0.67	0.77
SPO	10	0.5	1	1	0.6	0.8	0.78
STM	12	0.42	0.17	0.25	0.17	0.67	0.33
SYO	12	0.42	1	1	0.33	0.83	0.72
TAP	12	0.25	0.08	0.33	0.25	0.83	0.35
TDF	6	0.67	1	0.67	1	1	0.87
UTA	11	0.82	0.45	0.64	0.36	0.64	0.58
UUM	12	0.25	0.17	0.17	0	0.58	0.23
WIS	12	0.5	0.17	0.58	0.08	0.58	0.38
WLG	12	0.17	0.5	0.25	0.17	0.42	0.3
ZEP	12	0.42	0.17	0.25	0.25	0.58	0.33
Mean		0.51	0.5	0.34	0.33	0.6	0.45

Table 12. Percentage of months in 1997 where model results differ within 20% from observations.

Year	pTOMCAT	LMDz INCA	MOZECH	UiO	TM	model mean
1997	0.51	0.5	0.34	0.33	0.6	0.45
1998	0.35	0.35	0.3	0.29	0.57	0.37
1999	0.41	0.36	0.33	0.29	0.53	0.38
2000	0.48	0.41	0.42	0.34	0.61	0.45
Mean	0.44	0.41	0.35	0.31	0.58	0.41

Table 13. Skill scores for a MQO of 20%; average for all stations in every year.

Overall TM clearly shows the best comparison to surface CO with, on average, nearly 60% of the monthly means within 20% of the observed values. However when the stations are examined individually the picture is much more complex. The general agreement ranges from the worst at the Black Sea site (BSC) with an average agreement in 1997 MQO of 0.09 to the best at Tierra Del Fuego (TDF) with an average of 0.87 where 3 models (LMDZ, Oslo CTM2 and TM) are within 20% for all months. However, at this site the other two models only achieve this for two thirds of the months. Although all models achieve a perfect score of 1 for at least one site there is no site where all models achieve this. At Sary Taukum, Kazakhstan (KZD) all models except TM score zero.

Colour code:

red	0.9 - 1.0
orange	0.8 - 0.89
yellow	0.7 - 0.79

station	pTOMCAT	LMDz-INCA	MOZECH	UiO	TM	model mean
ALT	0.93	0.87	0.98	0.94	0.9	0.93
ASC	0.29	0.48	0.41	0.13	0.82	0.42
ASK	0.86	0.7	0.88	0.54	0.58	0.71
AZR	0.87	0.72	0.82	0.75	0.75	0.78
BAL	0.87	0.92	-0.08	0.85	0.88	0.69
BME	0.96	0.82	0.88	0.82	0.74	0.84
BMW	0.95	0.69	0.78	0.85	0.69	0.79
BRW	0.9	0.84	0.81	0.92	0.86	0.86
BSC	0.62	0.68	0.52	0.45	0.5	0.55
CBA	0.91	0.85	0.97	0.88	0.79	0.88
CGO	0.81	0.93	-0.18	-0.11	0.81	0.45
CHR	1	-1	-1	1	1	0.2
CRZ	0.96	0.99	0.93	0.84	0.99	0.94
EIC	0.81	0.97	0.74	0.64	0.81	0.8
GMI	0.12	0.31	0.75	0.27	0.08	0.31
HBA	0.93	0.97	0.95	0.81	0.9	0.91
HUN	0.85	0.78	0.32	0.76	0.88	0.72
ICE	0.9	0.76	0.53	0.82	0.75	0.75
IZO	0.91	0.82	0.86	0.62	0.64	0.77
KEY	0.91	0.78	0.08	-0.47	0.6	0.38
KUM	0.87	0.75	0.72	0.73	0.61	0.74
KZD	0.93	0.97	0.63	0.46	0.88	0.77
KZM	0.97	0.95	0.53	1	0.98	0.89
LEF	0.86	0.72	0.22	0.59	0.72	0.62
MHD	0.92	0.73	0.73	0.8	0.63	0.76
MID	0.9	0.89	0.85	0.86	0.78	0.86
MLO	0.81	0.76	0.76	0.81	0.63	0.75
NWR	0.86	0.16	0.74	0.95	0.56	0.65
PSA	0.93	0.94	0.94	0.76	0.88	0.89
RPB	0.85	0.72	0.87	0.5	0.4	0.67
SEY	0.96	0.97	0.46	0.61	0.98	0.79
SHM	0.89	0.81	0.97	0.9	0.77	0.87
SMO	0.89	0.91	0.49	0.46	0.93	0.74
SPO	0.87	0.97	0.96	0.85	0.9	0.91
STM	0.92	0.73	0.67	-0.17	0.72	0.57
SYO	0.84	0.97	0.95	0.03	0.88	0.73
TAP	0.24	0.25	-0.03	-0.13	0.74	0.21
TDF	0.95	0.88	0.8	0.92	0.99	0.91
UTA	0.82	0.22	0.18	0.41	0.39	0.4
UUM	0.61	0.37	0.72	0.65	0.7	0.61
WIS	0.76	0.71	0.49	0.83	0.42	0.64
WLG	0.59	0.38	0.56	-0.1	0.47	0.38
ZEP	0.94	0.87	0.93	0.9	0.88	0.91
Mean	0.83	0.71	0.61	0.6	0.74	0.7

Table 14. Correlations coefficients at all stations for 1997.

Year	pTOMCAT	LMDz-INCA	MOZECH	UiO	TM	model mean
1997	0.83	0.71	0.61	0.6	0.74	0.7
1998	0.5	0.66	0.6	0.43	0.45	0.53
1999	0.77	0.74	0.69	0.65	0.82	0.73
2000	0.63	0.75	0.71	0.66	0.74	0.7
Mean	0.68	0.72	0.65	0.59	0.69	0.67

Table 15. Average correlation coefficients for all years

red	$r \geq 0.9$
orange	$0.8 \leq r < 0.9$
yellow	$0.7 \leq r < 0.8$
grey	$0.6 \leq r < 0.7$
white	$r < 0.6$

The models all have high average correlation coefficients with model data on average. This is due to the fact that this correlation was performed with monthly mean data and the models are all able to capture the seasonal cycle of CO well. However at Guam (GMI) none of the models except MOZECH do well whereas at Alert (ALT) and Ny Alesund (ZEP) the models all do exceptionally well. On average, p-TOMCAT has the highest correlations in 1997, LMDZ the highest in 1998 and 2000 and TM in 1999 while the Oslo CTM2 had the lowest average correlations in all years.

	pTOMCAT	LMDz-INCA	MOZECH	UiO	TM	model mean
ALT	-14.59	-29.7	-29.95	-32.81	-7.52	-22.92
ASC	19.32	-11.12	-23.24	2.85	12.92	0.14
ASK	-17.75	-18.66	-34.15	-27.32	-14.44	-22.46
AZR	-17.1	-28.68	-30.27	-27.43	-12.66	-23.23
BAL	-18.82	-29.41	1.58	-29.24	-1.02	-15.38
BME	-16.37	-22.35	-29.23	-24.48	-11.94	-20.87
BMW	-21.75	-27.73	-30.32	-31.62	-16.02	-25.49
BRW	-14.15	-30.37	-29.93	-35.68	-7.84	-23.59
BSC	-42.89	-47.19	-32.58	-49.67	-25.24	-39.51
CBA	-18.12	-32.01	-32.38	-33.44	-8.21	-24.83
CGO	21.95	2.44	14.45	14.62	14.01	13.49
CHR	-12.06	-4.31	-27.04	-17.94	18.86	-8.5
CRZ	13.65	-3.37	-15.58	-2.16	-4.54	-2.4
EIC	10.37	-8.36	-20.4	-5.22	3.97	-3.93
GMI	-6.58	-16.68	-33.07	-9.67	14.27	-10.34
HBA	30.17	7.4	-8.59	10.89	9.65	9.9
HUN	-36.45	-39.58	-9.78	-43.58	-11.8	-28.24
ICE	-14.62	-28.21	-21.35	-25.27	-5.26	-18.94
IZO	-7.55	-15.39	-29.05	-21.31	-6.45	-15.95
KEY	-0.92	-29.33	30.26	26.78	12.62	7.88
KUM	-17.5	-25.52	-32.05	-24.22	-7.72	-21.4
KZD	-48.96	-44.43	-41.17	-56.25	-19.42	-42.05
KZM	-21.39	-7.32	-29.62	-37.02	4.61	-18.15
LEF	-22.17	-26.76	-24.47	-33.84	-11.21	-23.69
MHD	-15.89	-26.94	-21.22	-27.16	-4.85	-19.21
MID	-22.15	-32.41	-36.24	-30.66	-16.6	-27.61

	pTOMCAT	LMDz-INCA	MOZECH	UiO	TM	model mean
MLO	-15.67	-25.35	-36.95	-31.53	-6.96	-23.29
NWR	-15.8	-1.96	-28.43	-30.9	-8.96	-17.21
PSA	27.07	5.11	-9.06	8.08	7.76	7.79
RPB	-13.01	-22.1	-31.6	-19.03	-9.31	-19.01
SEY	8.64	-12.75	-30.75	-2.88	97.58	11.97
SHM	-19.34	-32.83	-33.57	-35.05	-9.98	-26.15
SMO	5.53	-7.34	-20.97	-6.17	22.29	-1.33
SPO	32.43	5.88	-9.12	9.53	10.28	9.8
STM	-16.71	-29.01	-23.16	-10.17	-7.2	-17.25
SYO	30.68	6.82	-8.71	81.12	9.9	23.96
TAP	-29.28	-36.26	0.1	-22.64	-8.02	-19.22
TDF	13.75	4.55	-15.44	-4.64	4.52	0.55
UTA	-8.65	-18.11	0.58	-18.96	3.64	-8.3
UUM	-24.25	-28.52	-31.42	-42.26	-13.39	-27.97
WIS	-20.17	-28.4	9.97	-29.51	13.41	-10.94
WLG	-25.56	-19	-23.01	-36.9	-13.07	-23.51
ZEP	-14.82	-30.77	-27.8	-31.43	-7.44	-22.45
Mean	-9.24	-19.67	-20.81	-18.47	-0.39	-13.72

Table 16. Annual mean Bias (%) 1997

	pTOMCAT	LMDz-INCA	MOZECH	UiO	TM	model mean
1997	-9.24	-19.67	-20.81	-18.47	-0.39	-13.72
1998	-18.29	-24.34	-26.07	-24.67	9.58	-16.76
1999	-20.18	-24.06	-22.17	-20.69	-13.76	-20.17
2000	-15.72	-20.4	-17.15	-17.48	-11.68	-16.49
	-15.86	-22.12	-21.55	-20.33	-4.06	-16.79

Table 17. Annual mean of relative biases (%)

Colour code:

red	< ± 5 %
orange	±5 < dev < ±10 %
yellow	±10 < dev < ± 20 %
grey	± 20 < dev < ± 30 %

The TM model has the smallest average bias in all years, with p-TOMCAT next smallest and the other three models having similar biases. The models (except TM in 1998) have an average negative bias and in 1997 the average bias at all sites except 9 is negative (ASC, TDF, PSA, KEY, SPO, HBA, SEY, CGO, SYO) which indicates either that emissions of CO are too small or the hydroxyl radical concentrations are too large causing the lifetime of carbon monoxide to be too short. The average bias at the sites ranges from 0.14 at Ascension to -42.05 at Sary Taukum, Kazakhstan (KZD).

These results are consistent with those presented in Shindell et al. (2006). This study examined the CO simulated by model participating in the ACCENT-IPCC model intercomparison and evaluated the results for 2000 against MOPPITT satellite and data from the CMDL network. All the RETRO models participated in this experiment. They also found that the models had large negative biases in the northern hemisphere while reproducing the observations much closer in the tropics, in particular at Samoa as found here (except for TM).

5.4 Ozone sondes - WOUDC

5.4.1 Line Plots and Taylor plots for 1997

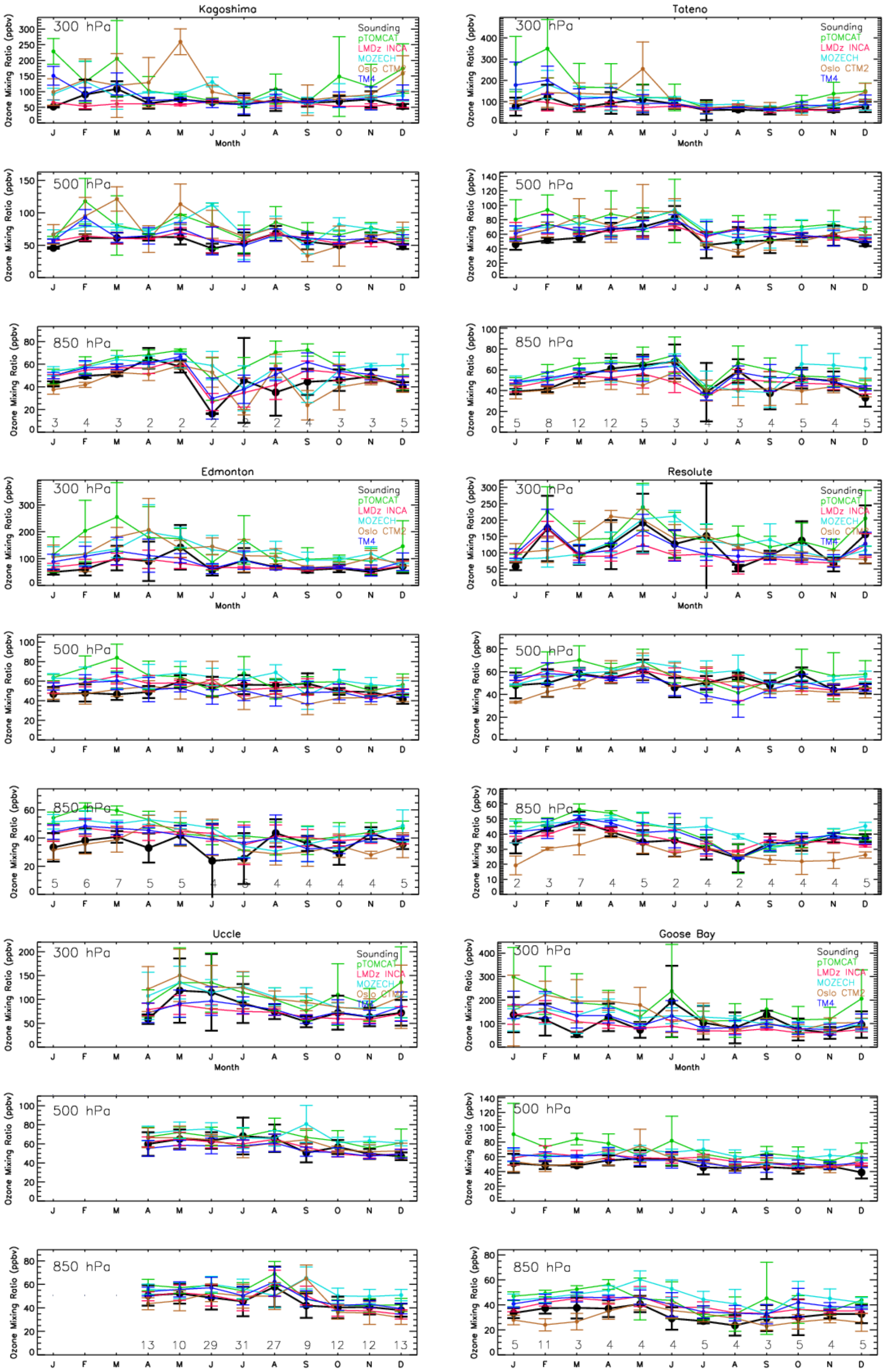

Figure 21a 1997 Monthly mean ozone concentrations at 300, 500 and 850 hPa.

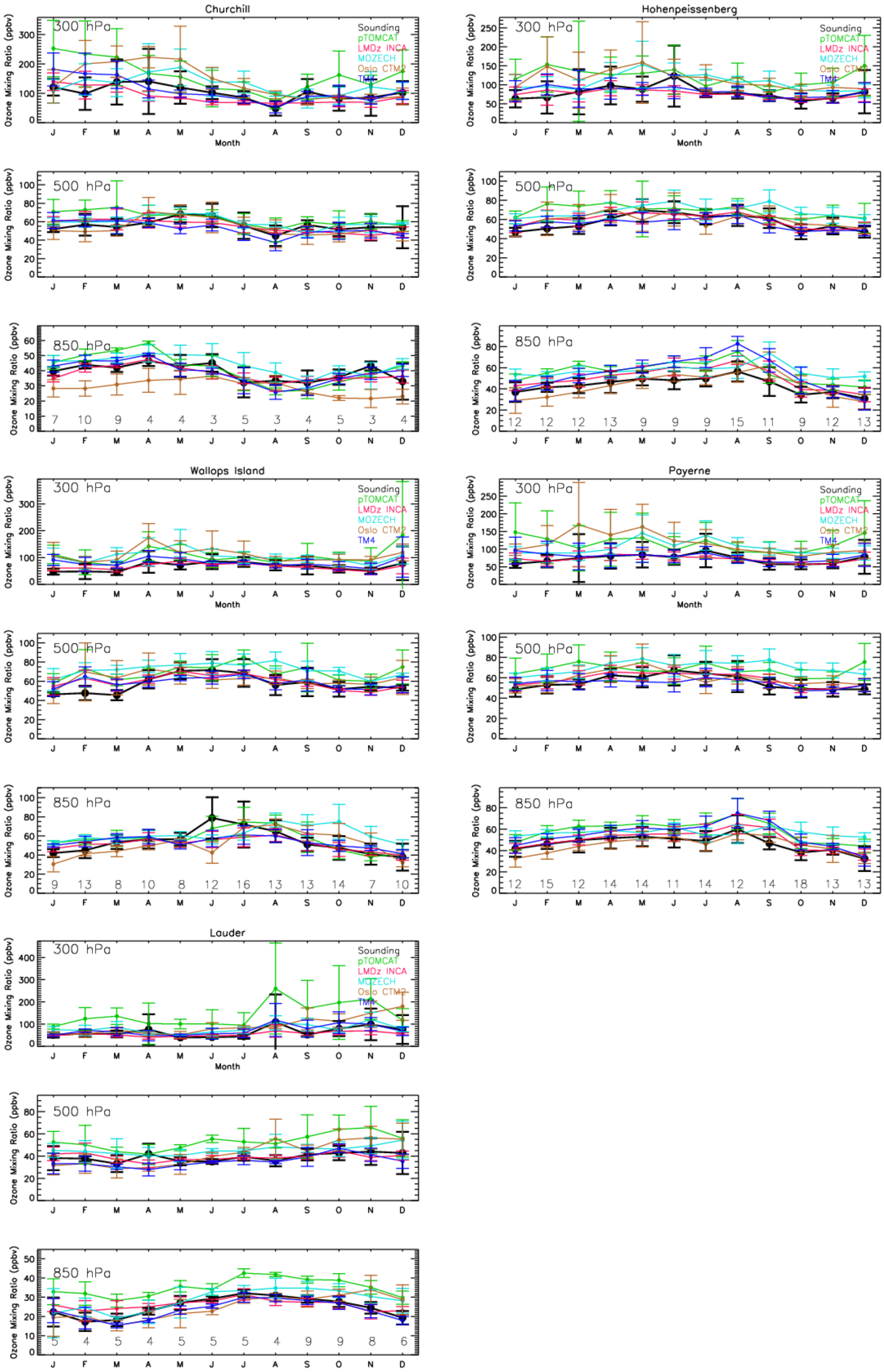

Figure 21b 1997 Monthly mean ozone concentrations at 300, 500 and 850 hPa.

All models in general seem to be able to reproduce the seasonal cycles of ozone at most sites and pressure levels. Good examples of this are the summer maximum seen at 850hPa at Hohenpeissenberg and Payerne, which is captured by all models, and the month to month variability at 300 hPa at Resolute, which all models do a reasonable job of predicting. Consistent with the surface ozone results and the global plots, p-TOMCAT can be seen to have too high ozone concentrations at most levels and locations. In contrast, near the surface the Oslo CTM2 model often has ozone concentrations which are too low while the concentrations at 300 hPa are too high (in particular at Churchill).

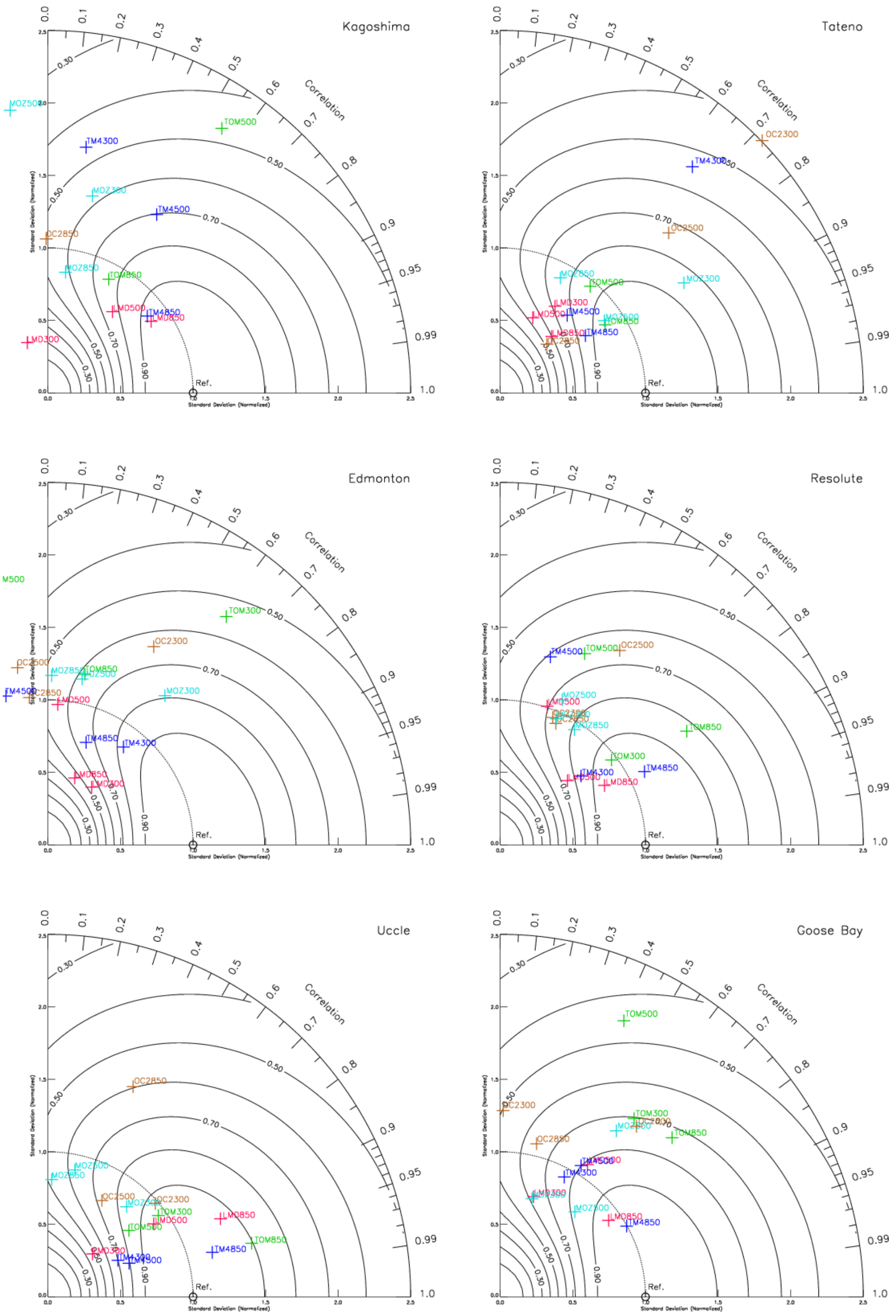

Figure 22a 1997 Taylor plots for ozone concentrations.

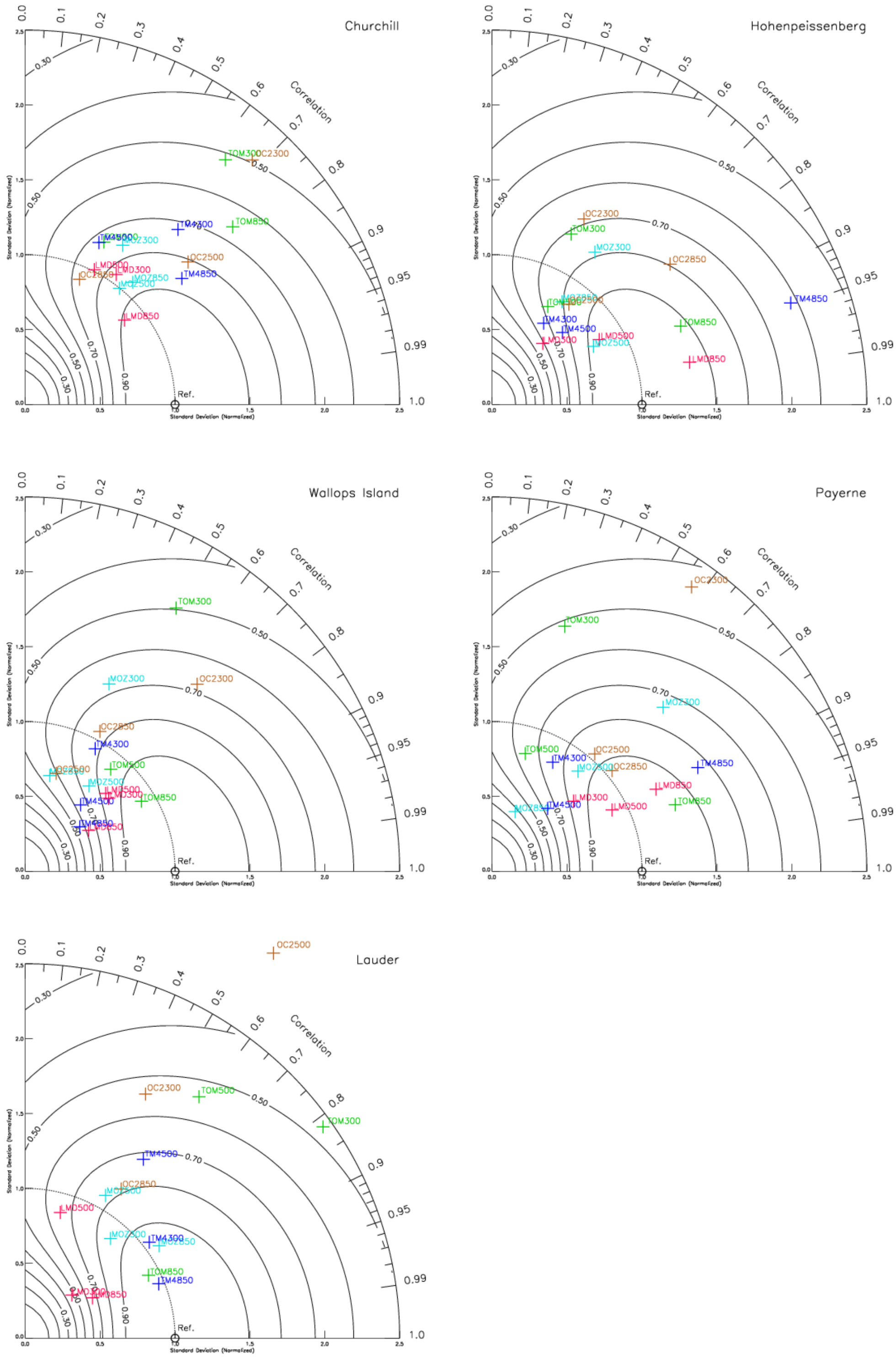

Figure 22b 1997 Taylor plots for ozone concentrations.

The models seem to frequently have the lowest performance at 500 hPa and it is highly variable from one location to another which model performs the best. At Hohenpeissenberg, the model which is closest to the observations is the LMDz model, at Uccle, it is the TM model and at Walllops Island, the model which performs best is p-TOMCAT, all the 850 hPa level. However, p-TOMCAT is also one of two models which on occasions has such poor performance that the point on the Taylor diagram falls outside the plotting area (at Kagoshima and Tatento for 300 hPa and at Edmonton for 500 hPa). The other model for which this is true is Oslo CTM2 at Kagoshima for both the 300 and 500 hPa levels. For Edmonton, all the models seem to do badly with all models lying well away from the 'ref' point. In contrast, at Uccle the model points generally lie much closer to the 'perfect model'. This may be related to either or both of two factors: there was no data in the JFM season this year and the minimum number of ascents was 9 and the maximum 29 in any given month. The larger data coverage may simply mean that the better agreement is purely a result of better statistics. On the other hand the January to March period also appears to be the time during which many of the models have the greatest problems in reproducing Northern Hemisphere ozone concentrations especially at upper levels.

5.4.2 Statistics

Skill scores.

station	num	pTOMCAT	LMDz-INCA	MOZECH	UiO	TM	model mean
RESO	12	0.58	1	0.5	0.42	0.92	0.68
CHUR	12	0.75	1	0.75	0.25	0.92	0.73
EDMO	12	0.33	0.5	0.17	0.58	0.5	0.42
GOOB	12	0.25	0.5	0	0.42	0.25	0.28
UCCL	9	0.89	1	0.56	0.89	1	0.87
MOHP	12	0.17	0.83	0.17	0.67	0.42	0.45
PAYE	12	0.17	0.92	0.5	0.83	0.75	0.63
WALL	12	1	0.92	0.42	0.92	0.83	0.82
TATE	12	0.5	0.67	0.5	0.75	0.67	0.62
KAGO	12	0.58	0.75	0.5	0.92	0.83	0.72
LAUD	12	0	0.83	0.67	0.75	1	0.65
Mean Lev	.	0.47	0.81	0.43	0.67	0.73	0.62

Table 18. Pressure level 850hPa, MQO 20%, 1997

	pTOMCAT	LMDz-INCA	MOZECH	UiO	TM	model mean
1997	0.47	0.81	0.43	0.67	0.73	0.62
1998	0.41	0.79	0.45	0.64	0.69	0.6
1999	0.39	0.8	0.59	0.51	0.76	0.61
2000	0.28	0.67	0.48	0.55	0.72	0.54
	0.39	0.77	0.49	0.59	0.73	0.59

Table 19. Pressure level 850hPa, MQO 20%, all years.

station	num	pTOMCAT	LMDz-INCA	MOZECH	UiO	TM	model mean
RESO	12	0.83	0.92	0.92	1	1	0.93
CHUR	12	0.83	1	1	1	1	0.97
EDMO	12	0.58	0.92	0.75	0.92	1	0.83
GOOB	12	0.17	0.92	0.67	1	1	0.75
UCCL	9	1	1	0.78	0.89	1	0.93
MOHP	12	0.67	1	0.75	1	1	0.88
PAYE	12	0.75	1	0.67	1	1	0.88
WALL	12	0.75	1	0.67	0.92	1	0.87
TATE	12	0.5	0.83	0.58	0.67	0.83	0.68
KAGO	12	0.33	0.83	0.67	0.42	0.67	0.58
LAUD	12	0.17	1	0.92	1	1	0.82
Mean Lev	.	0.6	0.95	0.76	0.89	0.95	0.83

Table 20 Pressure level 500hPa, MQO 30%, 1997.

	pTOMCAT	LMDz-INCA	MOZECH	UiO	TM	model mean
1997	0.6	0.95	0.76	0.89	0.95	0.83
1998	0.61	0.91	0.81	0.89	0.92	0.83
1999	0.55	0.92	0.76	0.9	0.93	0.81
2000	0.58	0.99	0.73	0.94	0.97	0.84
	0.59	0.94	0.77	0.91	0.94	0.83

Table 21. Pressure level 500hPa, MQO 30%, all years.

station	num	pTOMCAT	LMDz INCA	MOZECH	UiO	TM	model mean
RESO	12	0.42	0.75	0.67	0.58	0.83	0.65
CHUR	12	0.17	1	0.75	0.5	0.83	0.65
EDMO	12	0.08	0.75	0.33	0.42	0.58	0.43
GOOB	12	0.25	0.75	0.67	0.5	0.58	0.55
UCCL	9	0.56	0.89	0.67	0.78	1	0.78
MOHP	12	0.25	1	0.58	0.5	0.92	0.65
PAYE	12	0.17	1	0.5	0.25	0.92	0.57
WALL	12	0.25	0.92	0.42	0.25	0.67	0.5
TATE	12	0.42	1	0.83	0.42	0.75	0.68
KAGO	12	0.42	0.83	0.5	0.58	0.75	0.62
LAUD	12	0.08	0.92	0.58	0.42	0.67	0.53
Mean Lev	.	0.28	0.89	0.59	0.47	0.77	0.6

Table 22. Pressure level 300hPa, MQO 40%, 1997

	pTOMCAT	LMDz-INCA	MOZECH	UiO	TM	model mean
1997	0.28	0.89	0.59	0.47	0.77	0.6
1998	0.22	0.87	0.63	0.74	0.75	0.64
1999	0.25	0.89	0.63	0.85	0.73	0.67
2000	0.28	0.83	0.55	0.79	0.81	0.65
	0.26	0.87	0.60	0.71	0.77	0.64

Table 23. Pressure level 300hPa, MQO 40%, all years

Colour code:

red	0.9 - 1.0
orange	0.8 - 0.89
yellow	0.7 - 0.79

At 850hPa, all models except MOZECH have high scores at Uccle and Wallops Island. Overall the TM and LMDZ models are clearly best performing. At 500 hPa the models in general have higher scores than at 850 hPa (although with a less strict criterion of 30% not 20%). TM and LMDZ are again the best models with a score of 0.9 at most stations in 1997. The Oslo CTM2 model also shows quite good performance at this level and p-TOMCAT is clearly performing least well of the models, never achieving a score of >0.61 at any station in 1997 and with an overall average of just 0.59. At 300 hPa the scores are similar to those at 850 hPa with a criterion of 40%, twice as large as for 850hPa. LMDZ performing best with an average score for all years greater than 0.83. In contrast the performance of p-TOMCAT falls even lower here to an average of only 0.26. In other words almost three quarters of the time this model's monthly means have an error of more than 40%.

Correlation coefficients.

	pTOMCAT	LMDz-INCA	MOZECH	UiO	TM	model mean
RESO	0.83	0.89	0.46	0.35	0.89	0.68
CHUR	0.73	0.75	0.72	0.46	0.78	0.69
EDMO	0.26	0.35	0.09	0.02	0.19	0.18
GOOB	0.82	0.78	0.62	0.16	0.91	0.66
UCCL	0.02	-0.24	-0.42	-0.7	-0.08	-0.28
MOHP	0.91	0.97	0.59	0.73	0.95	0.83
PAYE	0.92	0.88	0.38	0.71	0.88	0.76
WALL	0.93	0.8	0.13	0.57	0.78	0.64
TATE	0.83	0.47	0.57	0.59	0.8	0.65
KAGO	0.26	0.52	0.2	0.1	0.43	0.3
LAUD	0.92	0.84	0.82	0.54	0.96	0.82
Mean	0.68	0.64	0.38	0.32	0.68	0.54

Table 24. Correlation coefficient for 1997 at 850hPa

Year	pTOMCAT	LMDz INCA	MOZECH	UiO	TM	model mean
1997	0.68	0.64	0.38	0.32	0.68	0.54
1998	0.56	0.56	0.33	0.52	0.58	0.51
1999	0.78	0.73	0.55	0.56	0.83	0.69
2000	0.46	0.67	0.59	0.51	0.74	0.59
Mean	0.62	0.65	0.46	0.48	0.71	0.58

Table 25. Average correlation coefficient for all years at 850hPa

	pTOMCAT	LMDz-INCA	MOZECH	UiO	TM	model mean
RESO	0.48	0.29	0.28	0.4	0.32	0.36
CHUR	0.5	0.55	0.74	0.7	0.49	0.59
EDMO	-0.24	0.02	0.31	0.05	-0.27	-0.03
GOOB	0.68	0.69	0.71	0.69	0.68	0.69
UCCL	0.41	0.19	-0.6	0.04	0.36	0.08
MOHP	0.63	0.87	0.81	0.8	0.74	0.77
PAYE	0.71	0.91	0.77	0.64	0.79	0.76
WALL	0.6	0.75	0.8	0.6	0.61	0.68
TATE	0.34	0.24	0.76	0.58	0.42	0.47
KAGO	0.12	-0.07	-0.18	0.13	0.04	0.01
LAUD	0.52	0.46	0.58	0.63	0.45	0.53
Mean	0.43	0.45	0.45	0.48	0.42	0.45

Table 26. Correlation coefficients at 500hPa, 1997.

Year	pTOMCAT	LMDz-INCA	MOZECH	UiO	TM	model mean
1997	0.43	0.45	0.45	0.48	0.42	0.45
1998	0.27	0.58	0.55	0.25	0.42	0.41
1999	0.5	0.68	0.76	0.69	0.7	0.67
2000	0.39	0.59	0.58	0.44	0.5	0.5
Mean	0.40	0.58	0.59	0.47	0.51	0.51

Table 27. Mean correlation coefficients at 500hPa

	pTOMCAT	LMDz-INCA	MOZECH	UiO	TM	model mean
RESO	0.52	0.39	0.36	0.44	0.46	0.43
CHUR	0.76	0.68	0.58	0.67	0.75	0.69
EDMO	0.42	0.37	0.63	0.56	0.34	0.46
GOOB	0.1	0.09	0.05	-0.06	0.12	0.06
UCCL	0.29	0.09	-0.25	0.43	0.15	0.14
MOHP	0.26	0.64	0.61	0.51	0.29	0.46
PAYE	0.31	0.66	0.78	0.63	0.4	0.56
WALL	0.16	0.63	0.62	0.55	0	0.39
TATE	0.66	0.65	0.8	0.76	0.64	0.7
KAGO	0.13	-0.41	0.3	0.19	0.03	0.05
LAUD	0.71	0.62	0.6	0.64	0.66	0.65
Mean	0.39	0.4	0.46	0.48	0.35	0.42

Table 28. Correlation coefficients at 300hPa, 1997.

Year	pTOMCAT	LMDz-INCA	MOZECH	UiO	TM	model mean
1997	0.39	0.4	0.46	0.48	0.35	0.42
1998	0.28	0.48	0.39	0.04	0.33	0.3
1999	0.63	0.67	0.63	0.6	0.68	0.64
2000	0.6	0.63	0.6	0.53	0.63	0.6
Mean	0.48	0.55	0.52	0.41	0.50	0.49

Table 29. Mean correlation coefficients at 300hPa, all years.

(red)	r ≥ 0.9
(orange)	0.8 ≤ r < 0.9
(yellow)	0.7 ≤ r < 0.8
(grey)	0.6 ≤ r < 0.7
(white)	r < 0.6

It is interesting to note that those stations at which the models have high skill scores are not necessarily associated with high correlation coefficients. For example, Uccle at which the models had some of the best skill scores for 850hPa, has an average correlation which is negative. The correlations at Hohenpeissenberg are the highest in the two lowest levels. As for the skill scores, TM and LMDz are among the best models overall, and MOZECH is ranking second. Compared to the skill score analysis for this level, p-TOMCAT performs much better relative to the other models. On the other hand, the correlations for the Oslo CTM are worse than the skill scores compared to the other models, indicating that the UiO model reproduced the annual ozone concentrations better than the seasonal cycle. At the two higher levels the model correlations are much closer to each other and are relatively small.

Relative bias.

	Mean obs	pTOMCAT	LMDz-INCA	MOZECH	UiO	TM	model mean
RESO	36.42	14.75	0.07	20.63	-21.5	7.89	4.37
CHUR	39.15	4.25	-2.5	13.27	-25.31	2.61	-1.54
EDMO	35.34	33.64	19.6	32.56	-0.94	16.17	20.2
GOOB	32.56	33.72	22.72	43.91	-5.25	26.7	24.36
UCCL	46.22	13.38	-0.17	19.55	-1.08	7.15	7.76
MOHP	43.48	28.27	14.52	33.36	-1.89	21.9	19.23
PAYE	46.66	25.23	6.71	24.06	-3.72	16.7	13.8
WALL	53.86	7.54	-4.57	19.9	-4.64	1.15	3.88
TATE	49.68	20.65	-2.34	15.63	-5.2	7.95	7.34
KAGO	45.32	39.26	9.76	33.24	13.74	19.35	23.07
LAUD	25.03	44.31	6.49	18.05	0.03	-4.77	12.82
Mean		24.09	6.39	24.92	-5.07	11.16	12.3

Table 30. Model bias, 850hPa (%), 1997

	pTOMCAT	LMDz-INCA	MOZECH	UiO	TM	model mean
1997	24.09	6.39	24.92	-5.07	11.16	12.3
1998	23.48	5.65	24.15	-10.04	8.71	10.39
1999	23.91	4.17	19.63	-13.87	6.94	8.15
2000	24.68	3.85	20.76	-13.28	6.82	8.57
	24.04	5.02	22.37	-10.57	8.41	9.85

Table 31. Model bias, 850hPa (%), all years

	Mean obs	pTOMCAT	LMDz-INCA	MOZECH	UiO	TM	model mean
RESO	51.64	21	5.03	11.43	-6.26	0.74	6.38
CHUR	56.09	14.19	-1.73	6.25	-7.82	-5.4	1.1
EDMO	51.22	29.26	9.21	20.91	-2.17	0.45	11.53
GOOB	48.61	40.47	16.84	30.72	11.17	13.95	22.63
UCCL	58.45	12.39	-1.37	17.51	3.47	-6.42	5.11
MOHP	57.01	21.76	3.62	22.11	4.98	-1.8	10.13
PAYE	55.72	22.66	5.53	25.6	8.81	-0.2	12.48
WALL	57.25	25.51	3.97	26.23	12.74	3.68	14.43
TATE	56.53	30.95	12.31	27.07	19.35	10.12	19.96
KAGO	56.32	35.34	5.63	28.22	25.36	10.98	21.11
LAUD	39.02	40.9	0.62	16.86	6.73	-5.58	11.9
Mean		26.76	5.42	21.17	6.94	1.87	12.43

Table 32. Model bias, 500hPa (%), 1997

	pTOMCAT	LMDz INCA	MOZECH	UiO	TM	model mean
1997	26.76	5.42	21.17	6.94	1.87	12.43
1998	23.38	3.18	15.95	-9	-1.73	6.36
1999	29.2	3.19	16.84	-9.95	-1.17	7.62
2000	26.94	4.32	18.35	-8.43	1.11	8.46
	26.57	4.03	18.08	-5.11	0.02	8.72

Table 33. Model bias, 500hPa (%), all years

	mean obs	pTOMCAT	LMDz INCA	MOZECH	UiO	TM	model mean
RESO	118.66	67.32	0.99	23.94	19.26	16.07	25.51
CHUR	102.64	62.53	-11.77	19.75	35.48	9.75	23.15
EDMO	75.36	115.21	12.01	55.6	62.11	34.45	55.88
GOOB	104.85	86.16	-3.4	31.44	50.6	20.66	37.09
UCCL	79.88	47.19	-4.16	35.89	42.5	4.92	25.27
MOHP	79.24	59.57	-2.63	34.74	45.56	6.02	28.65
PAYE	71.9	69.82	4.84	43.37	56.95	15.22	38.04
WALL	66.37	80.43	5.71	52.36	77.52	36.87	50.58
TATE	78.6	83	-3.62	32.89	53.85	30.88	39.4
KAGO	70.78	78.74	-10.07	35.4	69.7	21.52	39.06
LAUD	65.03	121.12	-7.58	29.12	50.31	17.36	42.07
Mean		79.19	-1.79	35.86	51.26	19.43	36.79

Table 34. Model bias, 300hPa (%), 1997

	pTOMCAT	LMDz INCA	MOZECH	UiO	TM	model mean
1997	79.19	-1.79	35.86	51.26	19.43	36.79
1998	90.42	-1.07	30.83	8.53	23.07	30.35
1999	88.76	-9.44	23.68	3.2	13.35	23.91
2000	72.75	-6.56	20.38	0.1	13.55	20.04
	82.78	-4.72	27.69	15.77	17.35	27.77

Table 35. Model bias, 300hPa (%), all years

Colour code:

red	< ± 5 %
orange	±5 < dev < ±10 %
yellow	±10 < dev < ± 20 %
grey	± 20 < dev < ± 30 %

The model biases offer a final complimentary measure of the model performance. At 850 hPa, only the Oslo CTM model has an overall negative bias which is consistent with what was found from the surface ozone data. The other models have overall positive biases, although these are small for LMDz and TM. This could indicate either too much photochemical ozone production in these models or excess stratosphere to troposphere exchange. The largest biases are in p-TOMCAT and MOZECH although these are much smaller in p-TOMCAT than for the surface ozone.

At 500 hPa, we again see that all models except Oslo CTM have a positive bias, but in most models this is smaller. This decrease in bias with altitude would seem to be more consistent with excessive photochemical ozone production in the models than with excessive STE (with the likely exception of p-TOMCAT, where STE may be more important). The bias at this level is very small in TM and LMDz, somewhat smaller in MOZECH, but unchanged in the p-TOMCAT model.

At 300 hPa, the bias in the p-TOMCAT model is very large for all years and stations, which would seem to indicate much too great a stratosphere to troposphere flux of ozone in this model. TM and MOZECH have relatively small positive biases which seems to indicate that, although there may be somewhat too much STE in these models, it is not as excessive as in p-TOMCAT. Apart from 1997, the average bias in Oslo CTM is relatively small and LMDz has a small negative bias giving no evidence of excessive STE in this model.

6 Overall model scores

To evaluate the models a final set of scores is now produced for each of the datasets above. These are averaged across all stations and years for each model.

6.1 Surface O3

	pTOMCAT	LMDz INCA	MOZECH	UiO	TM	model mean
Skill score (MQO 20%)	0.36	0.60	0.29	0.30	0.51	0.41
Correl. Coeff	0.60	0.49	0.02	0.13	0.63	0.38
Bias (%)	61.69	2.49	34.52	-3.19	14.85	22.07

Table 36.

For surface ozone the LMDz and TM models show the best overall performance with a score better than the average model for all 3 metrics. MOZECH and UiO show a very poor correlation coefficient indicating problems with the representation of seasonal cycles, while p-TOMCAT has a very large positive bias, possibly as a result of errors in the representation of STE.

6.2 Surface CO

	pTOMCAT	LMDz INCA	MOZECH	UiO	TM	model mean
Skill score (MQO 20%)	0.44	0.41	0.35	0.31	0.58	0.41
Correl. coeff	0.68	0.72	0.65	0.59	0.69	0.67
Bias (%)	-15.86	-22.12	-21.55	-20.33	-4.06	-16.79

Table 37.

All models have a negative bias which indicates either too large OH concentrations in the models or emissions of CO which are too low. Given that the methane lifetime of the models are similar to those found in previous studies (or in the case of p-TOCMAT towards the upper end of the range), it seems likely that this indicates that the emissions of CO in the inventory are too low. TM has the highest MQO and LMDZ the highest correlation coefficient.

6.3 Ozone on pressure levels

	pTOMCAT	LMDz INCA	MOZECH	UiO	TM	model mean
850 hPa						
Skill score (MQO 20%)	0.39	0.77	0.49	0.59	0.73	0.59
Correl. Coeff	0.62	0.65	0.46	0.48	0.71	0.58
Bias (%)	24.04	5.02	22.37	-10.57	8.41	9.85
500 hPa						
Skill score (MQO 30%)	0.59	0.94	0.77	0.91	0.94	0.83
Correl. coeff	0.40	0.58	0.59	0.47	0.51	0.51
Bias (%)	26.57	4.03	18.08	-5.11	0.02	8.72
300 hPa						
Skill score (MQO 40%)	0.26	0.87	0.60	0.71	0.77	0.64
Correl. coeff	0.48	0.55	0.52	0.41	0.50	0.49
Bias (%)	82.78	-4.72	27.69	15.77	17.35	27.77

Table 38.

At all levels, the TM and LMDz models are once again the best performing models with generally low biases, reasonable correlation coefficients and high MQOs. The p-TOMCAT model has large positive biases probably resulting from unresolved problems with excessive STE.

7 Conclusions

Although there are some problems related to stratosphere-troposphere exchange when using the ERA-40 data set (cf. van Noije et al., 2004) the model performance when validated against data indicates that with careful treatment of the top-boundary condition it is possible to use this data for chemistry-transport modelling with a performance comparable to that in previous studies. Progress continues to be made in global tropospheric chemistry modelling with the models able to capture the major features of the data, but it is clear that there is much room to improve the models. Model results vary widely and these models cannot achieve an objective of being within 20% of observations much of the time. However, a standard set of data and methods for comparison such as those presented here would allow progress in global tropospheric chemistry modelling to be monitored and give an objective way of evaluating future model developments.

8 Acknowledgements

This work was done as part of the RETRO project. RETRO is EU-funded project EVK2-CT-2002-00170, NH Savage acknowledges funding and supercomputer support from NERC via the NERC Centres for Atmospheric Science. The authors also would like to acknowledge NOAA CMDL and Sam Oltmans for the Surface Ozone and NOAA CMDL and Paul Novelli for the CO datasets.

References

Beljaars, A.C.M., and P. Viterbo, The role of the boundary layer in a numerical weather prediction model, in Clear and cloudy boundary layers, edited by A.M.M. Holtslag and P.G. Duynkerke, pp. 287-304, Royal Netherlands Academy of Arts and Sciences, Amsterdam, 1998.

Bottenheim JW, Fuentes JD, Tarasick DW, Anlauf KG, Ozone in the Arctic lower troposphere during winter and spring 2000 (ALERT2000), Atm. Env., 36, 15-16, 2535-2544, 2002.

Brunner, D. , J. Staehelin, H. L. Rogers, M. O. Köhler, J. A. Pyle, D. Hauglustaine, L. Jourdain, T. K. Berntsen, M. Gauss, I. S. A. Isaksen, E. Meijer, P. van Velthoven, G. Pitari, E. Mancini, V. Grewe, and R. Sausen, An evaluation of the performance of chemistry transport models by comparison with research aircraft observations. Part 1: Concepts and overall model performance, Atmos. Chem. Phys., 3, 1609-1631, 2003.

Brocheton, F., (1999), Représentation des émissions anthropiques dans les modèles globaux de chimie-transport: Sensibilité à la représentation spatiale des émissions et au degré de raffinement du schéma chimique. Ph. D. Thesis – Université Paris XII – Val de Marne.

Carver G.D., Brown P.D. and Wild O., The ASAD atmospheric chemistry integration package and chemical reaction database. Comput. Phys. Commun., 105 (2-3): 197-215, A, 1997.

Dentener, F., Peters, W., Krol, M., van Weele, M., Bergamaschi, P., and Lelieveld, J., Interannual variability and trend of CH_4 lifetime as a measure for OH changes in the 1979–1993 time period, J. Geophys. Res., 108(D15), 4442, doi:10.1029/2002JD002916, 2003.

Dentener, F.J., et. al.. The Global Atmospheric Environment for the Next Generation, Environ. Sci. and Technol., in press, 2006.

Hauglustaine, D.A., F. Hourdin, L. Jourdain, M.-A. Filiberti, S. Walters, J.-F. Lamarque and E.A. Holland (2004), Interactive chemistry in the Laboratoire de Météorologie Dynamique general circulation model: Description and background tropospheric chemistry evaluation, *J. Geophys. Res*., 109, D04314.

Holtslag, A.A.M., and Boville, B.A., Local versus nonlocal boundary layer diffusion in a global climate model. J. Clim., 6 (10), 1825-1842, 1993.

Hourdin, F., and A. Armengaud (1999), The use of finite-volume methods for atmospheric advection of trace species: 1. Test of various formulations in a general circulation model, *Mon. Weather Rev*., 127, 822-837.

IPCC (Intergovernmental Panel on Climate Change), Contribution of Working Group I to Third Assement Report on Climate Change, Cambridge University Press, 2001.

Kanakidou, M., Dentener, F. J., Brasseur, G. P., Berntsen, T. K., Collins, W. J., Hauglustaine, D. A., Houweling, S., Isaksen, I. S. A., Krol, M., Lawrence, M. G., Mueller, J. -F., Poisson, N., Roelofs, G. J., Wang, Y., Wauben, W. M. F. 3-D global simulations of tropospheric CO distributions - results of the GIM/IGAC intercomparison 1997 exercise. Chemosphere: Global Change Science, 1, 263-282, 1999.

Li, D., and K.P. Shine (1995), A 4-D ozone climatology for UGAMP models, UGAMP internal report, U.K. Univ Global Atmos. Model. Programme, Swindon.

Law, K.S., and Pyle, J. A.: Modeling Trace Gas Budgets in the Troposphere: 1. Ozone and Odd Nitrogen, J. Geophys. Res., 98, 18 377–18 400, 1993.

Novelli PC, Masarie KA, Lang PM, Hall BD, Myers RC, Elkins JW, Reanalysis of tropospheric CO trends: Effects of the 1997-1998 wildfires, J. Geophys. Res., 108 D15: art. no. 4464, 2003.

O'Connor, F.M., Law, K.S., Pyle, J.A., Barjat, H., Brough, N., Dewey, K., Green, T., Kent, J., and Phillips, G., Tropospheric Ozone Budget: Regional and Global Calculations, Atmos. Chem. Phys. Discuss., 4, 991-1036, 2004.

Olivier, J., Peters, J., Granier, C., Petron, G., Muller, J.F. and Wallens, S. Present and future surface emissions of atmospheric compounds, POET Report #2, EU project EVK2-1999-00011, 2003.
Oltmans, SJ, and Levy II, H, Surface ozone measurements from a global network, Atmos. Environ., 28, 9-24, 1994.
Penner, J. E., Lister, D., Griggs, D., Docken, D., and MacFarland, Taylor, K. E., Summarizing multiple aspects of model performance in a single diagram, Journal of Geophysical Research, 106, 7183-7192, 2001.
Prather, M.J., Numerical advection by conservation of second-order moments. J. Geophys. Res., 91, 6671-6681, 1986.
Rasch, P. J., Feichter, J., Law, K., Mahowald, N., Penner, J., Benkovitz, C., Genthon, C., Giannakopoulos, C., Kasibhatla, P., Koch, D., Levy, H., Maki, T., Prather, M., Roberts, D. L., Roelofs, G. J., Stevenson, D, Stockwell, Z, Taguchi, S, Kritz, M, Chipperfield, M, Baldocchi, D., McMurry, P., Barrie, L., Balkansi, Y., Chatfield, R., Kjellstrom, E., Lawrence, M., Lee, H. N., Lelieveld, J., Noone, K. J., Seinfeld, J., Stenchikov, G., Schwartz, S., Walcek, C., Williamson, D. A comparison of scavenging and deposition processes in global models: Results from the WCRP Cambridge workshop of 1995. Tellus B, 52, 4, 1025-1056, 2000.
Russell, G.L., and J.A. Lerner, A new finite-differencing scheme for the tracer transport equation, J. Appl. Meteorol., 20, 1483-1498, 1981.
Sadourny, R., and Laval, K., January and July performance of the LMD general circulation model, in *New Perspectives in Climate Modelling*, edited by A. Berger and C. Nicolis, pp. 173-198, Elsevier Sci., New York, 1984).
Shindell, D.T., G. Faluvegi, D.S. Stevenson, L.K. Emmons, J.-F. Lamarque, G. Pétron, F.J. Dentener, K. Ellingsen, M. Amann, C.S. Atherton, N. Bell, D.J. Bergmann, I. Bey, T. Butler, J. Cofala, W.J. Collins, R.G. Derwent, R.M. Doherty, J. Drevet, H.J. Eskes, A.M. Fiore, M. Gauss, D.A. Hauglustaine, L.W. Horowitz, I.S.A. Isaksen, M.C. Krol, M.G. Lawrence, V. Montanaro, J.-F. Müller, G. Pitari, M.J. Prather, J.A. Pyle, S. Rast, J.M. Rodriguez, M.G. Sanderson, N.H. Savage, M.G. Schultz, S.E. Strahan, K. Sudo, S. Szopa, T.P.C. van Noije, O. Wild, and G. Zeng. Multi-model simulations of carbon monoxide: Comparison with observations and projected near-future changes. Submitted to J. Geophys. Res., 2006.
Tiedtke, M. A., comprehensive mass flux scheme for cumulus parameterisation in large-scale models. Mon. Weather Rev., 117, 1779-1800,1989.
van Leer, B., (1977), Towards the ultimate conservative difference scheme, Part IV: A new approach to numerical convection, *J. Comput. Phys*., 23, 276-299.
van Noije, T. P. C., H. J. Eskes, M. van Weele, and P. F. J. van Velthoven (2004), Implications of the enhanced Brewer-Dobson circulation in European Centre for Medium-Range Weather Forecasts reanalysis ERA-40 for the stratosphere-troposphere exchange of ozone in global chemistry transport models, J. Geophys. Res., 109, D19308, doi:10.1029/2004JD004586.
van Noije, T.P.C., Segers, A. J., and van Velthoven, P.F.J.: Time series of the stratosphere-troposphere exchange of ozone simulated with reanalyzed and operational forecast data, J. Geophys. Res., 111, D03301, doi:10.1029/2005JD006081, 2006.
Vogelezang, D.H.P., and A.A.M. Holtslag, Evaluation and model impacts of alternative boundary-layer height formulations, Boundary-Layer Meteorol., 81, 245-269, 1996.