**Atmospheric
Chemistry
and Physics**

An evaluation of the performance of chemistry transport models – Part 2: Detailed comparison with two selected campaigns

D. Brunner[1], J. Staehelin[1], H. L. Rogers[2], M. O. Köhler[2], J. A. Pyle[2], D. A. Hauglustaine[3], L. Jourdain[4], T. K. Berntsen[5], M. Gauss[5], I. S. A. Isaksen[5], E. Meijer[6], P. van Velthoven[6], G. Pitari[7], E. Mancini[7], V. Grewe[8], and R. Sausen[8]

[1]Institute for Atmospheric and Climate Science, ETH, Zürich, Switzerland
[2]Centre for Atmospheric Science, Cambridge University, Cambridge, UK
[3]Laboratoire des Sciences du Climat et de L'Environnement, Gif-sur-Yvette, France
[4]Service d'Aéronomie, Paris, France
[5]Department of Geosciences, University of Oslo, Oslo, Norway
[6]Section of Atmospheric Composition, Royal Netherlands Meteorological Institute, De Bilt, The Netherlands
[7]Dipartimento di Fisica, Università L'Aquila, L'Aquila, Italy
[8]Institut für Physik der Atmosphäre, DLR, Wessling, Germany

Abstract. This is the second part of a rigorous model evaluation study involving five global Chemistry-Transport and two Chemistry-Climate Models operated by different groups in Europe. Simulated trace gas fields were interpolated to the exact times and positions of the observations to account for the actual weather conditions and hence for the specific histories of the sampled air masses. In this part of the study we focus on a detailed comparison with two selected campaigns, PEM-Tropics A and SONEX, contrasting the clean environment of the tropical Pacific with the more polluted North Atlantic region. The study highlights the different strengths and weaknesses of the models in accurately simulating key processes in the UT/LS region including stratosphere-troposphere-exchange, rapid convective transport, lightning emissions, radical chemistry and ozone production. Model simulated Radon, which was used as an idealized tracer for continental influence, was occasionally much better correlated with measured CO than simulated CO pointing towards deficiencies in the used biomass burning emission fields. The abundance and variability of HO_x radicals is in general well represented in the models as inferred directly from the comparison with measured OH and HO_2 and indirectly from the comparison with hydrogen peroxide concentrations. Components of the NO_y family such as PAN, HNO_3 and NO were found to compare less favorably. Interestingly, models showing good agreement with observations in the case of PEM-Tropics A often failed in the case of SONEX and vice versa. A better description of NO_x and NO_y emissions, chemistry and sinks is thought to be key to future model improvements with respect to the representation of chemistry in the UT/LS region.

Correspondence to: D. Brunner
(dominik.brunner@env.ethz.ch)

1 Introduction

Global chemistry transport models have become standard tools to study tropospheric and stratospheric photochemistry and the impact of different emission sources including scenarios for future emission changes. Radiative transfer calculations applied to the fields of radiatively active species such as ozone can then be used to infer a climate impact. Studies based on such models formed a central element in scientific assessments of the impact of present and future air traffic emissions (Penner et al., 1999; Brasseur et al., 1998; NASA, 1999). Other typical global model applications include studies of changes in tropospheric ozone levels since preindustrial times due to anthropogenic activity (Wang and Logan, 1998; Berntsen et al., 2000; Hauglustaine and Brasseur, 2001) and estimates of the contributions from different emission sources to the global tropospheric ozone budget (e.g. Lamarque et al., 1996; Lelieveld and Dentener, 2000).

A large number of such models have been developed over the last decades and the demand for evaluating how well they are able to reproduce available observations has increased accordingly. An important step in this direction has been the systematic compilation of aircraft observations of a large

number of measurement campaigns by Emmons et al. (1997) and Emmons et al. (2000) from which gridded composites as well as statistics for vertical profiles at a number of different sites have been generated. These composites have subsequently been used in a number of model evaluation studies (Bey et al., 2001; Wang et al., 1998; Levy II et al., 1999; Horowitz et al., 2003; Kuhlmann et al., 2003; Hauglustaine et al., 2004). Ozone measurements from commercial aircraft programs such as MOZAIC (Marenco et al., 1998), NOXAR (Brunner et al., 1998, 2001), and CARIBIC (Brenninkmeijer et al., 1999) now clearly form the most reliable and representative source of information on tropospheric ozone distributions against which models can be tested (Law et al., 2000; Bregman et al., 2001; Crowther et al., 2002; Emmons et al., 2000). For other compounds the situation is not quite as satisfactory. However, one year of continuous measurements of NO_x during NOXAR have already been used for model validation (Emmons et al., 2000; Grewe et al., 2001) and the more recent CO and NO_y measurements from MOZAIC as well as the less frequent but more complete measurements from CARIBIC will certainly be used extensively in future studies.

In the first part of this study we presented a new observation database that was compiled in the framework of the EU project TRADEOFF (Aircraft emissions: Contributions of various climate compounds to changes in composition and radiative forcing - tradeoff to reduce atmospheric impact) and a first evaluation of the models involved in TRADE-OFF using research aircraft observations of compounds relevant for tropospheric photochemistry (Brunner et al., 2003). The TRADEOFF database has been designed specifically to allow for a "point-by-point" comparison between interpolated model fields and corresponding observation points rather than for comparing averaged model fields with gridded composites. The advantage of this approach is that it can account for the actual meteorological conditions during each measurement campaign and hence for the specific origin and history of the sampled air masses. In-situ measurements from research aircraft campaigns can only cover a tiny volume of the whole space-time domain and can therefore not be regarded as being representative in a climatological sense. Observations from such campaigns are therefore often grouped into different air mass classes based on tracer-tracer correlations or on air parcel trajectory calculations which are then analysed and characterized separately. Observations within each class may then be regarded as being representative to some extent for a specific air mass type. However, the frequency at which different classes are sampled may be strongly biased by the specific goals of a project and the actual meteorological conditions prevailing during the campaign. Comparisons of climatological or monthly mean model fields with such limited observations therefore need to be treated with care and may even lead to wrong conclusions with respect to model performance.

In this second part of the study we analyse the performance of the TRADEOFF models for two selected measurement campaigns, the Pacific Exploratory Mission - Tropics A (PEM-Tropics A) and the Subsonic Assessment Ozone and Nitrogen Oxide Experiment (SONEX). This detailed analysis extends the first part which presented a description of the database, the models involved, the methods applied to judge model performance, and a first overall model evaluation. Different from the first part we here focus on the performance of the models under a range of different situations dominated by different atmospheric processes in order to learn more about the ability of the models to simulate these processes accurately. We will investigate model performance with respect to the representation of long-range transport, stratosphere-troposphere-exchange, rapid deep convective transport and mixing, lightning NO production, and radical chemistry and ozone production.

Section 2 briefly summarizes the models, data sets, and analysis methods. Results of the comparison between models and observations of the two campaigns are then shown in Sect. 3. In the first part of this section time-series of all flights and vertical profiles at two selected sites are analysed in detail, and in the second part a more quantitative analysis of model performance based on overall biases, correlations and root mean square differences represented in the form of Taylor diagrams (Taylor, 2001; Brunner et al., 2003) is shown.

2   Models, data and methods

A detailed description of the chemistry-transport (CTM) and chemistry general circulation models (C-GCM) and the methods used for evaluating the models was presented in the first part of this study (Brunner et al., 2003). Here we only repeat the main points.

The TRADEOFF project featured five different CTMs (TM3 (Meijer et al., 2000), CTM-2 (Kraabol et al., 2002), TOMCAT (Law et al., 2000), ULAQ CTM (Pitari et al., 2002), SLIMCAT (Chipperfield, 1999)) and two C-GCMs (LMDz-INCA (Hauglustaine et al., 2004), DLR E39/C (Hein et al., 2001)). A table of the main model characteristics including spatial resolution, the number of gas phase and heterogeneous reactions considered, the methods and parameterizations used for advection, convection, and diffusion, and others was given in Brunner et al. (2003). Whenever possible the models performed transient multi-year runs simulating the meteorological conditions of the years 1995 to 1998. The models were run as much as possible using the same emission fields which were generated specifically for the TRADEOFF study and which closely follow the recommendations of the recent IPCC OxComp intercomparison exercise (Houghton et al., 2001). For more details see Brunner et al. (2003).

Two different types of output were generated for model evaluation purposes and delivered to ETH Zurich where the output was processed into a uniform format and uploaded

to an online database accessible to all project partners. The first type of output was gridded monthly mean volume mixing ratio fields of the components O_3, CO, OH, HO_2, H_2O_2, H_2O, NO, NO_2, HNO_3, PAN, and $Rn222$. Additional diagnostics included gridded fields of net ozone production rates $P(O_3)$, lightning NO emissions, and wet HNO_3 deposition. The second type of output, which will be used extensively in this study, consisted of the concentrations of the species listed above interpolated to the exact times and positions of measurements obtained from a large number of aircraft measurement campaigns and ozone soundings during the period 1995 to 1998. The production of this second type of output puts relatively little additional computational load onto the model simulations. The advantages of this "point-by-point" approach have already been discussed in Sect. 1.

A listing of the campaigns included in the comparison was given in Brunner et al. (2003). The temporal (or in case of soundings the vertical) resolution of the measurements has been reduced to better match the scales resolved by the models. For both the SONEX and PEM-Tropics A campaigns 5-minute averages were used corresponding to a cruise distance of the NASA DC-8 aircraft of about 70 km, which is still significantly smaller than the grid-spacing of the models ranging from approximately 200 to 500 km.

No output for point-by-point comparisons could be produced by the University of l'Aquila CTM (ULAQ) driven by meteorological fields from a GCM, and by the E39/C C-GCM (model based on ECHAM-4 physics and operated by DLR) because these models simulated a climatological meteorology representative of the year 2000 rather than the actual conditions of the years 1995 to 1998. The LMDz-INCA GCM was nudged with wind fields provided by the European Center for Medium Range Weather Forecasts (ECMWF) analyses allowing to simulate the dynamical conditions of the selected four years very successfully as shown in Brunner et al. (2003) and Hauglustaine et al. (2004). All remaining models were driven by ECMWF analysed temperature, pressure, wind and humidity fields. Results from the SLIM-CAT model are not presented here because only very few measurements from the SONEX and PEM-Tropics A campaigns were available for comparison with this stratospheric model. Two different versions of the CTM-2 model, both operated by the University of Oslo, were used in this work. Output from a version including only tropospheric chemistry was only available for the year 1996 (and hence for PEM-Tropics A) whereas output from a second version, which will be called CTM2-Gauss in the following, including both tropospheric and stratospheric chemistry was only available for the year 1997 (and hence for SONEX).

The skill of the models has been tested by analysing average concentration biases and by using the illustrative and quantitative method introduced by Taylor (2001). A detailed description of the use of "Taylor diagrams" in the context of CTM model evaluation was provided by Brunner et al. (2003). On these diagrams the correlation coefficient R and

pattern root-mean-square (RMS) error E' between a test field f (model) and a reference field r (observations), along with the ratio of the standard deviations (σ_f and σ_r) of the two patterns are all indicated by a single point in a two-dimensional plot. The pattern RMS E' is the RMS error after subtracting the means from both the test and the reference fields. The correlation coefficient and the pattern RMS provide complementary aspects of model performance (Taylor, 2001). A skill score S weighting these two aspects was defined as

$$S = \frac{4(1 + R)^2}{\left(\hat{\sigma}_f + 1/\hat{\sigma}_f\right)^2 (1 + R_0)^2} \tag{1}$$

where R_0 is the maximum attainable correlation which is limited by the overall uncertainty of the measurements and also by the fact that the model fields can not fully resolve all features of the 5-min averaged observation data. $\hat{\sigma}_f$ is the normalized standard deviation, that is the ratio of the standard deviation of the test field to the standard deviation of the reference field. This is a measure of the amplitude of variations simulated by the model relative to that seen in the observations. Deviations from the optimal value 1 imply an enhanced pattern RMS error E'.

3 Results

In this section we compare the models in detail with measurements from two distinct campaigns, contrasting the results for a remote region of the atmosphere with results for a more polluted environment. We have selected the PEM-Tropics A and SONEX campaigns (see campaign descriptions below) for this purpose due to the comprehensiveness of the set of measured species and because the same platform was used in both campaigns (the NASA DC-8 airplane) and many species were therefore analysed by the same groups often using nearly identical or at least similar instruments. Details on the instruments and their overall performance and accuracy are given in the overview papers by Hoell et al. (1999) and Singh et al. (1999) for PEM-Tropics A and SONEX, respectively.

The Pacific Exploratory Missions (PEM) were initiated by NASA to investigate the impact of human and natural processes on the chemistry of the troposphere over the Pacific basin. As part of this program the PEM-Tropics A mission was conducted in September 1996 during the southern hemisphere dry season (season with maximum biomass burning) and focused on the tropical and subtropical regions of the South Pacific Ocean (Hoell et al., 1999). The broad objectives were to improve our understanding of the oxidizing power of the tropical atmosphere as well as investigating oceanic sulfur compounds and their conversion to aerosols. The Pacific basin is mostly remote from continental influence, and hence provides a particularly sensitive indicator of the global-scale impact of human activity on the chemistry

a) b)
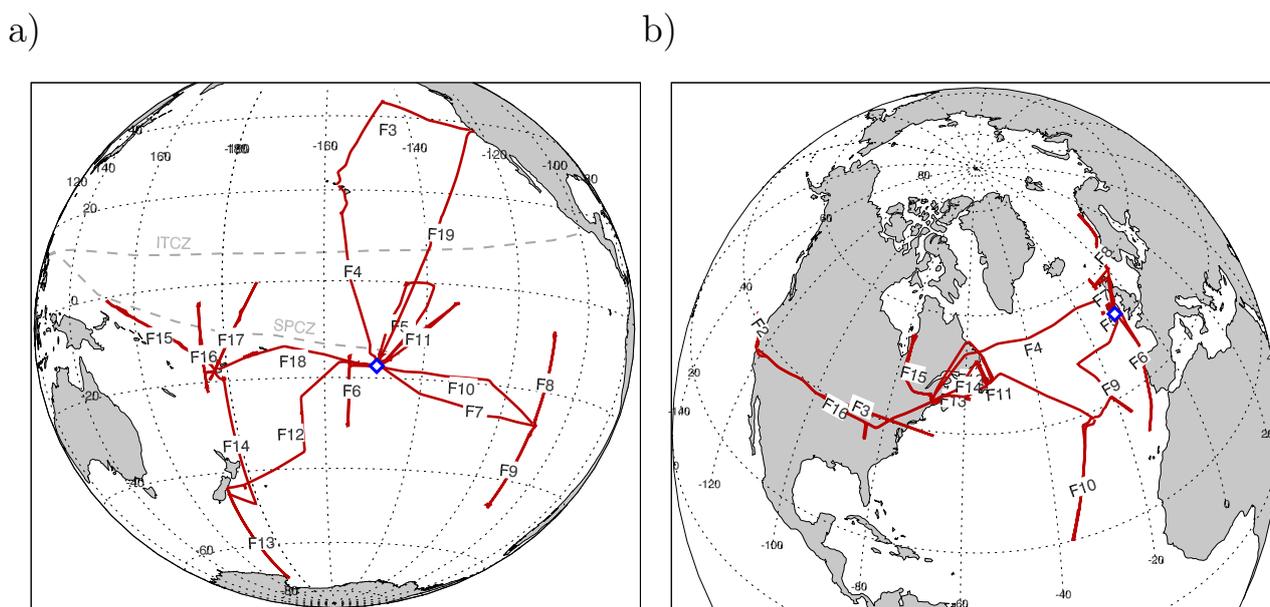
Fig. 1. Maps of **a)** PEM-Tropics A, **b)** SONEX measurement flights. ITCZ=Intertropical Convergence Zone, SPCZ=South Pacific Convergence Zone. Blue diamonds indicate the positions of the vertical profiles shown in Fig. 3 and Fig. 5.

of the troposphere. Figure 1a presents a map of the investigation area and the individual flight paths.

The budget of ozone and the influences of biomass burning and lightning to the South Pacific during the PEM-Tropics A campaign have been investigated by means of global CTMs by Schultz et al. (1999), Staudt et al. (2002), and Wei et al. (2003). Staudt et al. (2002) calculated that large amounts of the CO above the South Pacific originated from biomass burning over Africa, South America, and smaller amounts from Indonesia and Australia. Biomass burning was close to the long-term mean in September 1996 (Olson et al., 1999) but the meteorological conditions were favorable for rapid eastward transport from Africa and South America which probably resulted in above average CO and O_3 concentrations over the South Pacific (Staudt et al., 2002). The South Pacific is a region of net ozone destruction in the tropospheric column (Schultz et al., 1999; Staudt et al., 2002). Net ozone destruction is found in the lower troposphere up to about 6 km altitude due to very low NO_x concentrations. This loss is partly compensated by net ozone production found above 6 km altitude where NO_x concentrations are higher. According to Staudt et al. (2002) lightning was the dominant source of NO_x in the upper troposphere over the South Pacific whereas biomass burning dominates the budget of PAN. Decomposition of PAN is the dominant source of NO_x below 4 km altitude (Schultz et al., 1999).

Staudt et al. (2002) performed a point-by-point comparison between their model and the observations similar to ours. Other model evaluations using PEM-Tropics A measurements, however, were based on monthly mean values,

on output from climatological model runs, or on output from simulations for other years (Schultz et al., 1999; Bey et al., 2001; Horowitz et al., 2003; Wei et al., 2003; Kuhlmann et al., 2003).

Measurements from The Subsonic Assessment: Ozone and Nitrogen Oxide Experiment (SONEX) were obtained in the more polluted mid-latitudes of the Northern Hemisphere during October and November 1997. The flights were conducted over the North Atlantic ocean and near the east coast of North America and hence closer to continental sources and more strongly influenced by aircraft emissions than the PEM-Tropics A measurements (see Fig. 1b). The primary goal of SONEX was to study the impact of aircraft emissions on atmospheric concentrations of NO_x and ozone and to better understand the relative role of the different sources of NO_x.

A model analysis of the sources of upper tropospheric NO_x during SONEX and a similar point-by-point comparison with observations as performed here has been carried-out for four selected flights by Meijer et al. (2000) using the TM3 model. Air traffic emissions were shown to be the dominant source of NO_x in the North Atlantic flight corridor (about 50%) and surface sources and lightning supply another 15%–25% each. With respect to upper tropospheric O_3, however, lightning (20%) and surface sources (30%) made up much larger contributions than air traffic (10%).

3.1 PEM-Tropics A

Figure 2 shows composites of PEM-Tropics A measurements in which the time-series of all individual flights of the DC-8
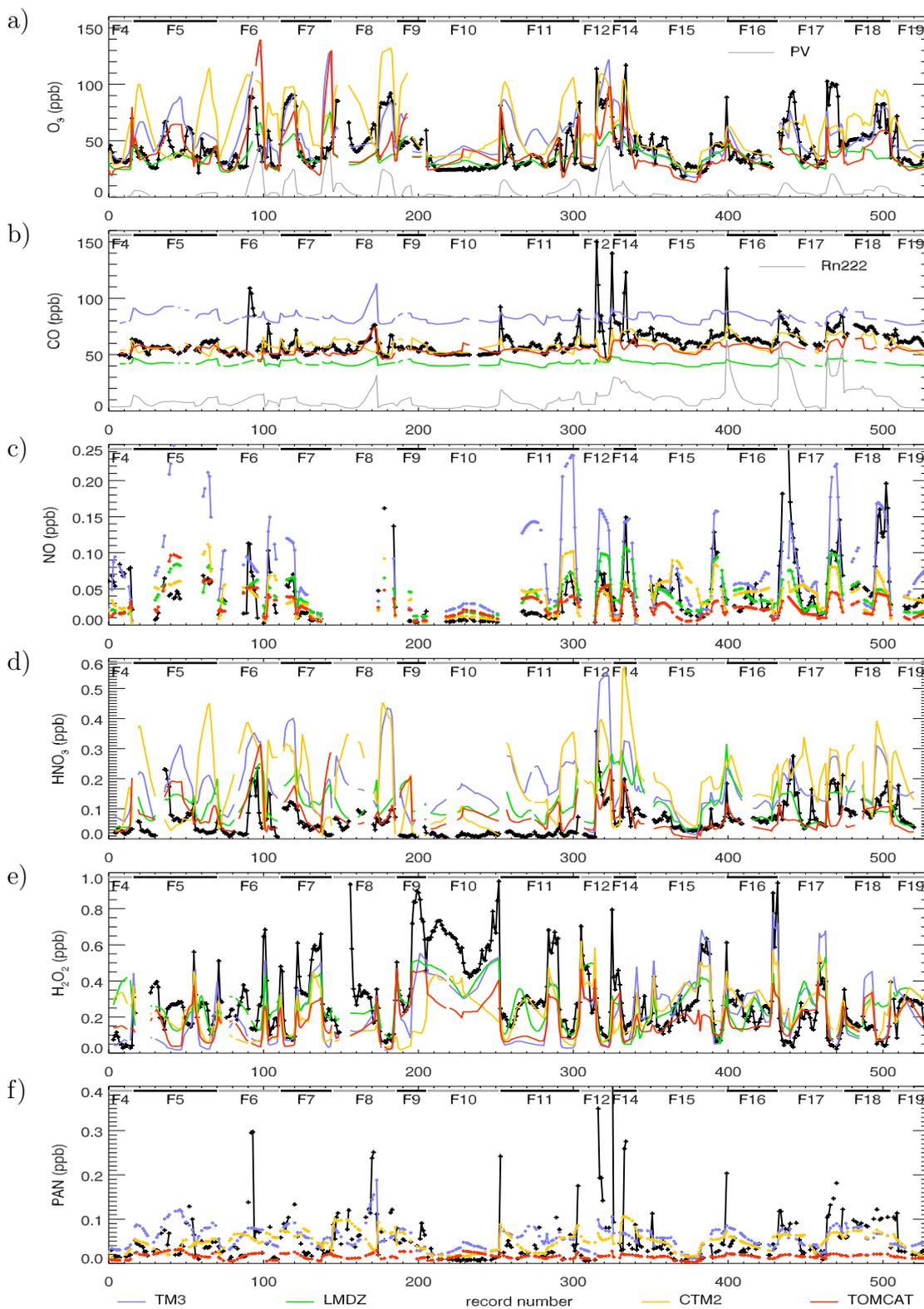
Fig. 2. Composites of PEM-Tropics A time series. Measurements at 0–35° S and 400–200 hPa (about 7.2–12 km) only. Flight numbers are indicated at the top of each panel. Black: measurements. Colors: simulations of the different models (see legend for color code). The grey line in the top panel **a)** is potential vorticity (PVU) from the LMDz-INCA model, scaled by a factor of 10 to fit on the O$_3$ axis scale. Similarly, Rn222 (mol/1e22mol) is shown in grey in panel **b)** and is displayed on the same y-axis as CO.

aircraft are merged to a single figure. Flight numbers are indicated at the top of each panel. Each data point represents a 5-min average along the flight track. Measurements are shown in black and the corresponding interpolated model values in different colors. The analysis is restricted to the upper troposphere ($p < 400$ hPa, that is about 7.2 to 12 km) and to the tropical/subtropical South Pacific ocean (0–35° S). The same domain was selected in the study of Board et al. (1999) where the chemical composition of air masses from differing source regions observed during PEM-Tropics A was characterized based on trajectory calculations. Potential vorticity (PV) and Radon (Rn222) are displayed as additional model diagnostics in the ozone and CO panels, respectively (only values from the LMDz-INCA model are shown for clarity). Potential vorticity helps identifying any stratospheric influence whereas Radon is emitted from continents and as such is an ideal species for tracing continental influence (Jacob et al., 1997). In this section we concentrate on the comparison of observed and simulated patterns rather than on overall biases. Averaged measured concentrations and model biases for the entire PEM-Tropics A domain and separately for flights 5–10 over the eastern South Pacific and flights 15–18 over the western South Pacific are summarized in Table 1 and will be discussed in more detail later in Sect. 3.3 along with a more objective analysis of model performance during PEM-Tropics A.

3.1.1 Continental and biomass burning signatures

Carbon monoxide is emitted by industrial sources and biomass burning and is formed through oxidation of methane and other hydrocarbons such as isoprene. Compared to other tracers CO shows a rather low variability over the South Pacific with several narrow spikes of strongly elevated concentrations. The models differ significantly from each other in terms of absolute concentrations but the simulated structures are quite similar. Measured background levels were around 50 to 60 ppb during the first part of the campaign (flights 5 to 10) investigating the eastern South Pacific (see Fig. 1a). During the later flights 15 to 18 focusing on the western South Pacific, background concentrations were somewhat enhanced by about 10 ppb (Table 1). This change in background levels, although less pronounced, is captured by CTM-2 and to some degree by TOMCAT, but not by LMDz-INCA and TM3. The different meteorological regimes of the eastern and western South Pacific were described by Fuelberg et al. (1999). Large-scale subsidence in the quasi-stationary subtropical anticyclone centered near Easter Island was dominating the meteorology over the eastern South Pacific. Westerly winds prevailed at 300 hPa south of about 10° S and therefore flights over the eastern South Pacific were mostly remote from continental sources as confirmed by the relatively low Radon concentrations (Fig. 2b). The Radon peak at the northernmost tip (7° S) of flight 8 is most probably due to outflow from Central or South America

by the easterly tropical winds as suggested by the trajectory analysis of Fuelberg et al. (1999). The same air mass exhibits a marked increase in CO which is very nicely captured by the models, especially by TOMCAT and TM3. In contrast, the high CO peak during flight 6 of up to 110 ppb, associated with a strong increase in ozone, is completely missed by the models. Similar peaks with strong enhancements in both ozone and CO are also seen on flights 12 and 14. All these flights were directed southwards towards the subtropical jet stream where air was advected rapidly from the west. Such air masses were identified by Board et al. (1999) as being transported over long distances by strong westerly winds during the previous 10 days, first crossing southern Africa and then Australia, and some even originating from South America. Tracer measurements belonging to this "long-range air" category showed strong signatures of biomass burning but no clear indications of urban or industrial activity (Board et al., 1999). Although biomass burning emissions were not substantially different from the long-term mean in September 1996, they were somewhat above normal along the South African coast (south of about 22° S) and below normal further north (Olson et al., 1999). The TRADEOFF models used climatological emission fields which do not represent the specific conditions during September 1996. This may partly explain the poor correlation between modelled and observed CO in these plumes.

Better correspondence between measured and modelled CO and other tracers is found on flights 15 to 18 over the western South Pacific based from Fiji. These flights were influenced by the South Pacific Convergence Zone (SPCZ, see Fig. 1a), a region associated with widespread convective activity separating the inner tropics to the north from the subtropics to the south (Fuelberg et al., 1999). The SPCZ was crossed on flights 16 and 17 and was just touched at the return point on flight 15. Relatively clean air was sampled in the inner tropics north of the SPCZ (Gregory et al., 1999; Schultz et al., 1999) where easterly winds prevailed. In contrast, air south of the SPCZ was advected from the west passing over Australia and typically originating from South Africa or Southeast Asia (Fuelberg et al., 1999), and strong signatures of biomass burning were frequently observed here (Gregory et al., 1999). This likely explains the generally elevated CO values observed on these flights. A recent continental origin of some air parcels is indeed suggested by the strongly enhanced Radon concentrations which correlate surprisingly well with measured CO (Fig. 2b). In some cases, in particular on flights 15 to 18, Radon is even better correlated with observed CO than the simulated CO values, providing some evidence for a missing or falsely distributed biomass burning source in the models.

PAN had a dominant source from biomass burning during PEM-Tropics A (Schultz et al., 1999). Thus similar problems with respect to the detection of individual plumes as seen in the CO results also apply to PAN. Individual plumes in both PAN and CO are best captured by CTM-2 and TM3,

Table 1. Mean observed concentrations and model biases (model/meas*100%) for PEM-Tropics A and SONEX. Measured mean and standard deviation values are shown in italics.
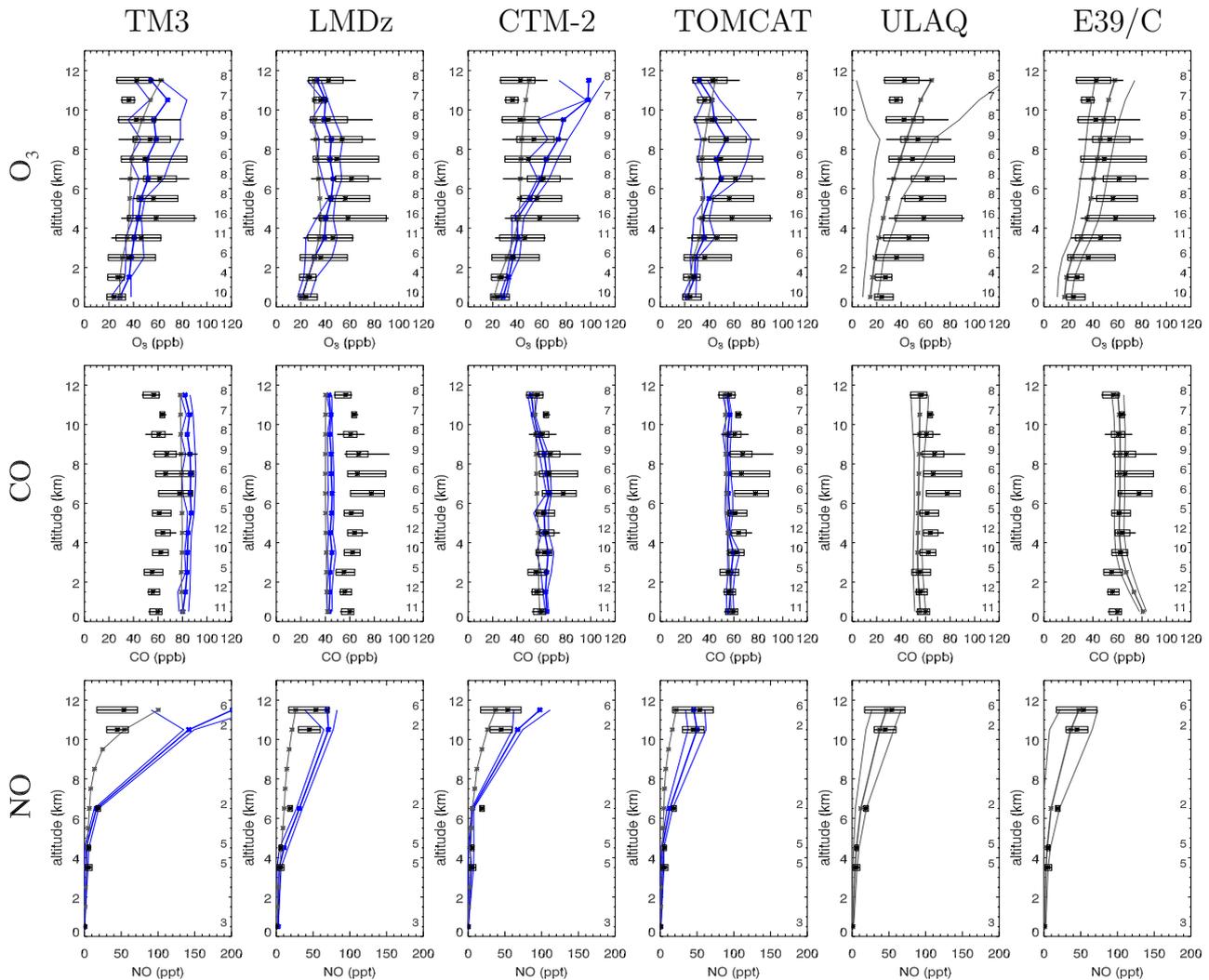
		PEM-Tropics A			SONEX		
		All flights	Flights 5–10	Flights 15–18	All flights	Flights 3–7	Flights 9–16
O_3	*(ppb)*	*43.9±19.3*	*42.7±18.4*	*46.3±19.5*	*76.7±63.4*	*55.0±45.5*	*77.0±60.1*
	TM3	+18.3%	+28.2%	−4.6%	+38.0%	+44.1%	+35.7%
	LMDz-INCA	−19.0%	−16.8%	−23.7%	−11.2%	+3.0%	−16.3%
	CTM-2	+37.9%	+54.4%	+12.9%	+26.7%	+52.6%	+25.3%
	TOMCAT	−10.4%	+1.4%	−27.9%	+20.7%	+29.9%	+12.3%
CO	*(ppb)*	*61.1±11.6*	*55.9±8.1*	*66.6±8.2*	*81.7±20.3*	*72.7±11.3*	*89.2±19.9*
	TM3	+35.7%	+49.2%	+24.7%	+35.0%	+44.8%	+27.8%
	LMDz-INCA	−30.5%	−24.3%	−36.3%	−23.9%	−19.0%	−27.9%
	CTM-2	−4.8%	+0.6%	−7.6%	+18.2%	+25.0%	+12.2%
	TOMCAT	−10.4%	−4.5%	−14.4%	−21.9%	−14.7%	−26.5%
NO	*(ppt)*	*44±39*	*30±32*	*57±45*	*94±167*	*65±57*	*126±217*
	TM3	+77.1%	+153.6%	+20.7%	−30.0%	+19.1%	−50.9%
	LMDz-INCA	−4.8%	+25.6%	−25.4%	−25.2%	+14.6%	−42.5%
	CTM-2	−8.4%	+7.9%	−16.7%	−50.9%	−21.8%	−63.4%
	TOMCAT	−39.3%	+3.9%	−60.6%	−50.5%	−20.4%	−65.1%
HNO_3	*(ppt)*	*60±53*	*48±47*	*82±49*	*204±228*	*136±204*	*207±205*
	TM3	+192.8%	+282.8%	+79.5%	−13.1%	−30.8%	−16.2%
	LMDz-INCA	+90.7%	+143.9%	+34.9%	+82.1%	+139.0%	+73.6%
	CTM-2	+213.3%	+270.8%	+134.4%	−27.0%	−42.2%	−37.0%
	TOMCAT	+25.5%	+100.5%	−37.3%	+50.9%	+70.8%	+43.3%
H_2O_2	*(ppt)*	*293±193*	*367±212*	*224±156*	*113±82*	*132±67*	*99±84*
	TM3	−22.6%	−39.6%	+27.3%	−56.0%	−71.2%	−35.6%
	LMDz-INCA	−10.4%	−22.9%	+7.8%	+127.8%	+115.3%	+156.4%
	CTM-2	−22.6%	−50.5%	+24.2%	−47.2%	−53.9%	−37.5%
	TOMCAT	−40.3%	−49.9%	−22.5%	+22.0%	+8.9%	+44.9%
OH	$(10^{-3}ppt)$	.	.	.	*90±68*	*90±39*	*90±78*
	TM3	.	.	.	−35.0%	+15.7%	−47.6%
	LMDz-INCA	.	.	.	+91.3%	+165.6%	+84.2%
	CTM-2	.	.	.	+8.6%	+56.0%	+2.8%
	TOMCAT	.	.	.	+9.7%	+61.3%	+3.3%
PAN	*(ppt)*	*54±64*	*50±51*	*52±40*	*67±43*	*56±36*	*75±44*
	TM3	+8.0%	+24.4%	+6.3%	+174.6%	+185.3%	+164.6%
	CTM-2	−9.7%	−3.6%	−3.5%	+167.9%	+165.5%	+176.3%
	TOMCAT	−70.9%	−65.3%	−71.2%	−37.0%	−32.8%	−43.7%

but the amplitude of the enhancements is significantly underestimated. Similar deficiencies in reproducing individual CO/PAN plumes were reported by Staudt et al. (2002) despite the fact that the biomass burning distribution in their model was adjusted to better reflect the actual conditions of September 1996. They suggested that this problem was caused by convective transport in the model over the biomass burning regions which tends to create a well-mixed column of enhanced CO rather than vertically confined layers of outflow as seen in the observations. In addition to the representation of individual plumes, average PAN levels are also best simulated by TM3 and CTM-2 but are strongly underestimated by TOMCAT (Table 1). PAN might be insuffi-

ciently represented in TOMCAT as a relatively simple non-methane hydrocarbon (NMHC) chemistry scheme was used. A more comprehensive scheme is currently being worked on (M. Köhler, personal communication).

3.1.2 Convective activity and lightning

Very low ozone of 20–25 ppb was observed near the northwestern corner of flight 15. Satellite images show widespread convective activity over this area associated with the SPCZ, and vertical profiles of ozone, temperature, and humidity taken from the DC-8 indicate strong vertical mixing (Fuelberg et al., 1999). The low ozone values are thus

Fig. 3. PEM-Tropics A profiles at at Tahiti, 17.5° S/149.5° W. Horizontal bars and boxes span the central 90% and 67% of the frequency distributions of the measurements grouped into individual 1 km altitude bins. Mean values are indicated by a star inside each box. Blue solid lines represent the corresponding model values. The central thick line connects the model mean values and the range between the two thinner lines represents the central 67% range calculated from the point-by-point model output. The dark grey line shows the monthly mean profile of the model grid column containing Tahiti for September 1996. This profile can be compared to the mean model profile derived from the point-by-point output. For E39/C and ULAQ, which did not provide point-by-point output, only the monthly mean profiles and standard deviations are shown.

most probably due to upward transport of air from the marine tropical boundary layer where ozone is depleted rapidly (Schultz et al., 1999). This drop in ozone is very well reproduced by all models providing some confidence in the quality of the parameterisation of convective transport near the SPCZ.

Lightning is most probably the dominant source of NO and following oxidation of HNO$_3$ in the tropical upper troposphere (Schultz et al., 1999; Staudt et al., 2002). However, positive correlations of NO and HNO$_3$ with CO on flight 17 also suggest a substantial contribution from biomass burning. Despite the same global strength the distribution of the light-

ning NO source is quite different between the models due to different parameterizations. The TM3 and LMDz-INCA models, for instance, deposit a substantial amount of lightning produced NO in the upper troposphere following the suggestions of Pickering et al. (1998). Interestingly, TM3 significantly overestimates NO and HNO$_3$ over the eastern South Pacific (Fig. 2, flights 5–10) where convection was suppressed by the large anticyclone but provides good agreement for flights 15 to 18 which were influenced by the SPCZ. It would be interesting to know what fraction of NO and HNO$_3$ in the models over these regions was attributable to lightning activity, but unfortunately the present model runs
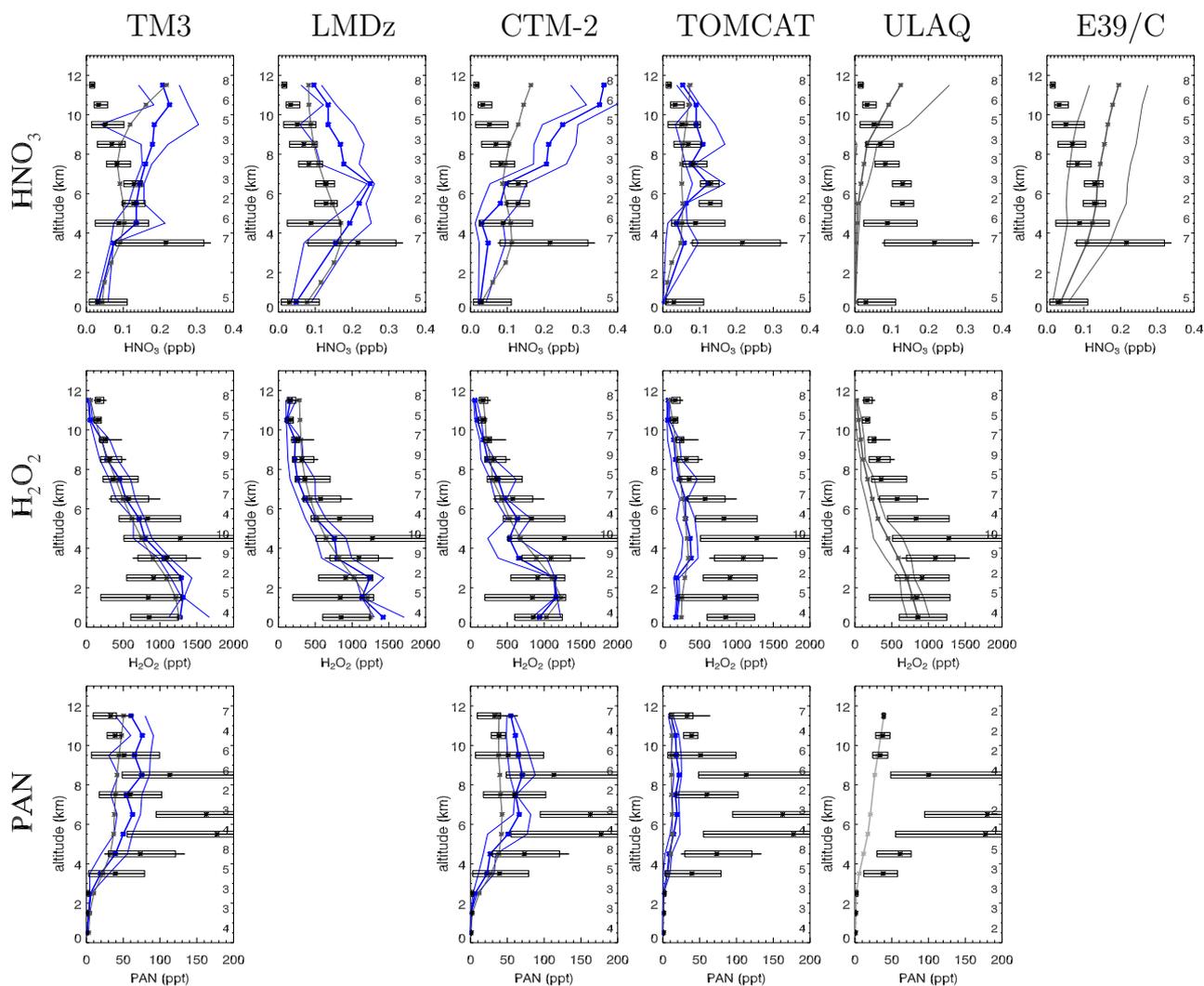
Fig. 3. Continued.

did not provide such source-specific information. The ratio of HNO$_3$:NO$_x$ over the South Pacific is clearly higher in the models than observed. This problem was also noted in previous CTM studies (Wang et al., 1998; Bey et al., 2001) and also in chemical box model calculations (Chatfield, 1994; Schultz et al., 2000). Fig. 2 shows that NO tends to be in much better agreement with the observations than HNO$_3$, which was also observed in the previous CTM studies. Several possible explanations for the tendency of models to overestimate HNO$_3$ in the tropical upper troposphere were presented in the literature including insufficient washout of HNO$_3$ (Wang et al., 1998), missing heterogeneous conversion of HNO$_3$ to NO$_x$ on sulfate aerosols (Chatfield, 1994) or on soot (Hauglustaine et al., 1996), overestimate of N$_2$O$_5$ hydrolysis which is suppressed if aerosols were mostly dry (Schultz et al., 2000), and removal of HNO$_3$ due to gravitational settling of cirrus ice crystals which is missing in the

models (Lawrence and Crutzen, 1998). However, other studies by Staudt et al. (2002) and Kuhlmann et al. (2003) did not find such large discrepancies. As long as there are such large differences among the models it is not possible to draw any firm conclusions on whether fast heterogeneous reactions of HNO$_3$ are required or not. Moreover, there is so far no experimental evidence for such fast reactions (Jacob, 2000).

3.1.3 Stratospheric influence

The eastern South Pacific was dominated by large-scale subsidence associated with the subtropical anticyclone centered near Easter Island (Fuelberg et al., 1999). Air masses over this region slowly descended from the uppermost troposphere and frequently showed influences of downward transport from the stratosphere (Fenn et al., 1999). In agreement with this picture measured and modelled ozone and HNO$_3$ exhibit a high variability (Fig. 2). Some measured ozone

peaks were associated with a drop in CO (e.g. on flight 8) suggesting a marked stratospheric influence. Enhanced PV with maximum (absolute) values around 2 PVU indeed suggest an air mass origin near the tropopause. CTM-2 even overestimates ozone significantly, in particular in air masses with elevated PV, which may point towards too strong downward transport from the stratosphere. HNO_3, another tracer with a potential stratospheric origin is also overestimated in CTM-2. TM3 also overestimates HNO_3 but performs quite well with respect to ozone. As noted before HNO_3 tends to be more strongly overestimated over this region than over the western South Pacific. Too strong downward mixing from the stratosphere is likely to contribute to this problem in some models, especially in CTM-2. However, a missing sink in HNO_3 could also explain the disagreement between models and measurements because this would affect the remote eastern South Pacific more strongly than the western South Pacific, which is closer to sources of NO.

3.1.4 Radical chemistry and ozone production

For the PEM-Tropics A conditions hydrogen peroxide (H_2O_2) has a lifetime of 1–2 days, suggesting that it is in near chemical equilibrium (Schultz et al., 1999). Since H_2O_2 is produced by recombination of the hydrogen peroxy radical (HO_2) the comparison of modelled H_2O_2 with observations provides a test for the simulated peroxy radical concentrations and hence (together with NO) for the calculated ozone production rate. Fig. 2e shows that the models tend to slightly underestimate H_2O_2 during the early flights over the eastern Pacific, but in general the agreement is very encouraging especially with respect to variability. Possible deviations in ozone production rates thus appear to be dominated by errors in NO rather than by errors in peroxy radical levels.

3.1.5 Vertical profiles at Tahiti

Vertical profiles obtained near Tahiti (17.5° S/149.5° W) are shown in Fig. 3. All flights to and from Tahiti displayed in Fig.1a contribute to this figure. Displayed are the statistics for individual 1 km altitude bins. The central blue line in each panel is the mean point-by-point model profile and the flanking thin blue lines span the central 67% of the frequency distribution, which is equal to ± one standard deviation in the case of a normal distribution. For comparison, the grey line shows the monthly mean model profile evaluated for the model grid column comprising Tahiti. Differences between these two profiles point at the limited representativity of PEM-Tropics A measurements with respect to monthly mean September 1996 conditions. Also included in the figure are monthly mean and (temporal) standard deviation profiles for September of the ULAQ and E39/C models. Since these models were run in climate mode we can not expect the same degree of agreement as for the other models. Measured CO and O_3 profiles show maxima in the altitude range between

5 and 8 km. A similar maximum in CO is missing in all models and indications of an inverse C-shaped ozone profile are only visible in LMDz-INCA and TOMCAT. The CTM-2 model quite strongly overestimates ozone in the 10–12 km altitude range consistent with the findings of Sect. 3.1.3 which suggested a too strong stratospheric contribution. The model ozone profiles are generally close to the observations below 5 km altitude, except for ULAQ and E39/C which underestimate the concentrations by about 10 ppb here. Simulated CO profiles exhibit almost no vertical structure. Only E39/C simulates increased CO in the boundary layer, which is not seen in the measurements. LMDz-INCA generally underestimates CO by some 30% whereas TM3 overestimates CO by 30% to 40%. The increased levels of measured CO, O_3 and PAN in the altitude range of 5–8 km are most probably due to numerous layers of air affected by biomass burning, which were most frequently observed at these levels (Gregory et al., 1999). Potential problems in the models concerning biomass burning were discussed earlier. With respect to HNO_3 the observed concentrations are matched the closest by TOMCAT. Most models including ULAQ and E39/C tend to overestimate HNO_3 above 7 km, and some underestimate HNO_3 values below 6 km. It is also interesting to note that the model profiles not only deviate from the observations but also differ substantially among each other. In contrast to this, H_2O_2 is well represented in the models with TOMCAT being generally too low as noted already in the previous section. Similar to CO, PAN is strongly enhanced in the 5–8 km altitude range and it is underestimated here by all models. PAN is not simulated in E39/C and LMDz-INCA. More NO_x is therefore expected to end up in the form of HNO_3 instead of PAN in these models. However, LMDz-INCA simulates comparatively little HNO_3 in the upper troposphere and the overall shape of the profile is in better agreement with the observations than in other models.

3.2 SONEX

Composites of time series of the SONEX measurement campaign are shown in Fig. 4. The data are restricted to the UT/LS region to pressures below 350 hPa (about 8–12 km). Like before, PV and Radon values (this time from the TM3 model instead of LMDz-INCA but differences are rather small) are shown in the ozone and CO panels to aid data interpretation. In addition, the fractional NO contribution from aircraft emissions calculated as the difference between two TOMCAT model runs with and without aircraft emissions is displayed in Fig. 4c. In agreement with the study by Meijer et al. (2000) this contribution is often exceeding 50%. Mean tracer concentrations and model biases for the two campaigns are summarized in Table 1. A detailed overview of the meteorological situation during individual flights including a trajectory based analysis of air parcel histories has been presented by Fuelberg et al. (2000). This information will further be used in the following discussion.
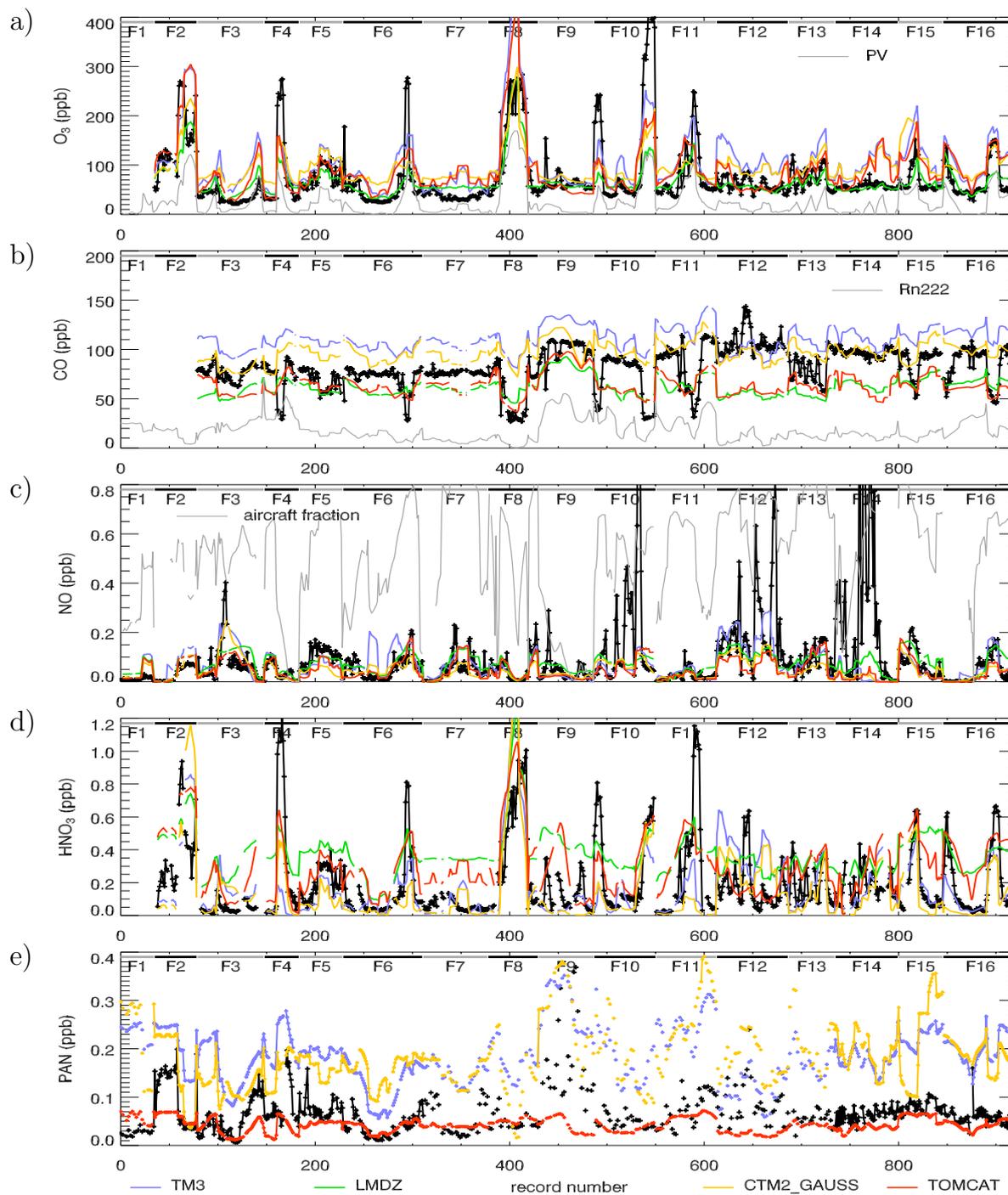
Fig. 4. Composites of SONEX time series. Only measurements between 350 and 200 hPa are included. Legend as in Fig. 2. Radon values in panel b) are from the TM3 model. The grey line in panel c) is the fraction of NO which is contributed by air traffic emissions calculated using two TOMCAT model runs with and without including air traffic emissions ("full run" – "no aircraft")/"full run".
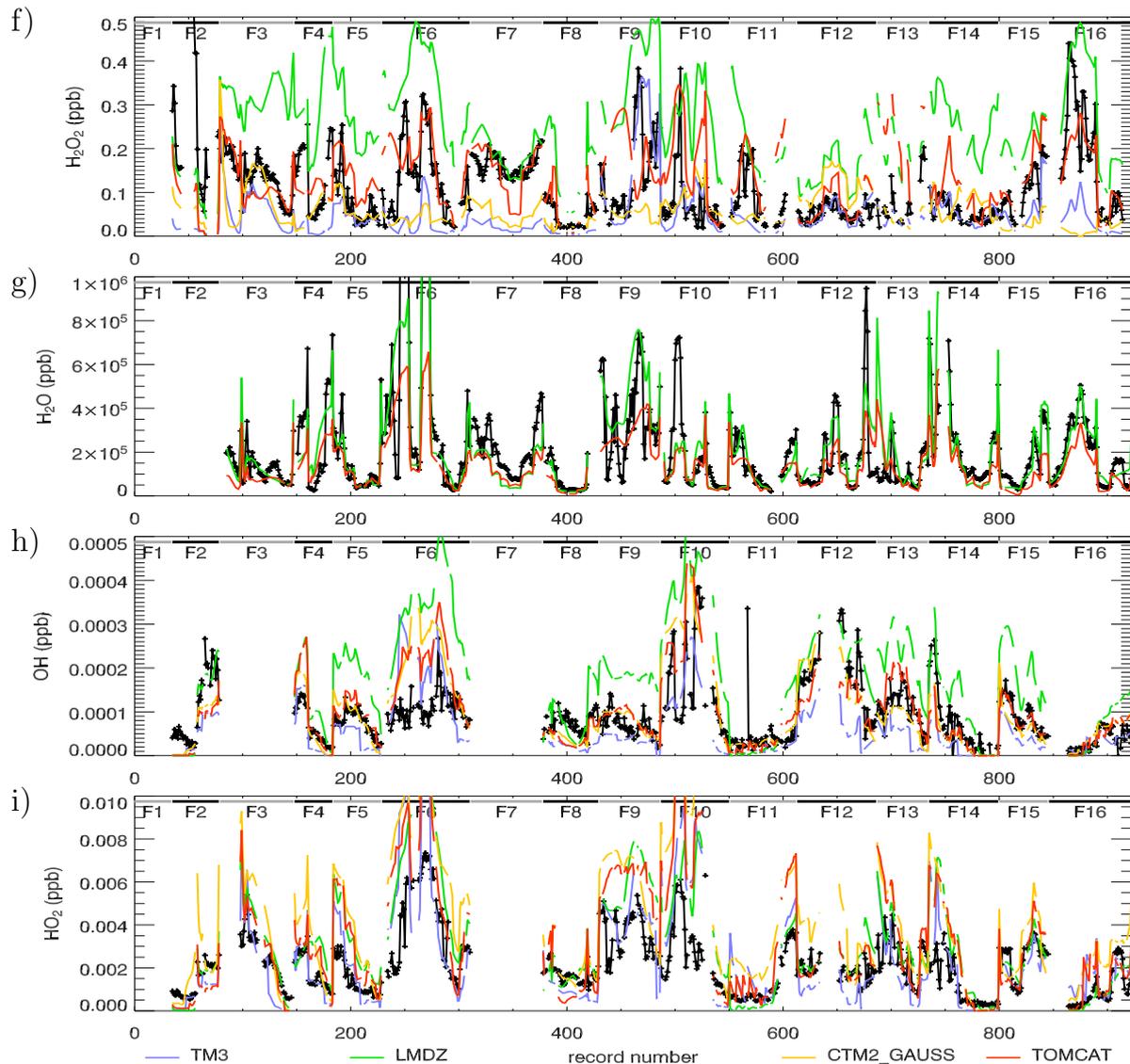
Fig. 4. Continued. In panel g) the red line showing TOMCAT is also representative of the TM3 and CTM-2 Gauss models because they all rely on ECMWF humidity fields.

3.2.1 Stratospheric influence

Ozone values often exceeded 100 ppb and reached up to 400 ppb showing that air samples were frequently collected in the lowermost stratosphere. These samples were characterized by low CO and enhanced HNO_3 concentrations and are qualitatively reproduced by the models. However, the amplitude of the CO reduction is mostly underestimated by the models and along with this increases in ozone and HNO_3 are usually underestimated, in particular within narrow features. This apparent mixing of tropospheric and stratospheric air which reduces the steep tracer gradients across the tropopause seen in the observations is evidence for insufficient model resolution and/or numerical diffusion. However,

inaccuracies in tracer transport either due to the numerical formulation of advection or errors in the ECMWF analyses may contribute as well (Bregman et al., 2001). It is quite revealing to analyse the behavior of simulated Radon during encounters of stratospheric air. On flight 4, for instance, Radon slightly increases during the sharp measured ozone peak in contrast to what would be expected for stratospheric air, and simulated ozone increases are much less pronounced than observed. Correspondingly, CO shows no drop at all in the models but rather follows the Radon increase. This suggests that continentally influenced air of high CO concentrations had mixed with stratospheric air in the models, which is in strong contrast to the measurements. Similar problems are

seen on flights 6 and 10 where the observed ozone increases are underestimated by all models. Radon drops to very low levels only during the broad structure of stratospheric air observed on flight 8 where the ozone increase is well represented by the models. TOMCAT and TM3 even overstate the ozone maximum here in agreement with the general conclusion of Brunner et al. (2003) that these models tend to overestimate O_3 in the lowermost stratosphere.

3.2.2 Transport from the lower tropical marine boundary layer

Background concentrations of ozone and CO in the upper troposphere show remarkable differences between the early flights 3 to 7 and the later flights 9 to 16. The low concentrations during the early flights are consistent with the trajectory analysis of Fuelberg et al. (1999) showing frequent advection of air from the tropical marine boundary layer. Grant et al. (2000) described these cases in more detail. Exceptionally low ozone concentrations during this time period were also reported from the NOXAR measurements from a commercial airliner crossing the North Atlantic about twice daily (Brunner et al., 2001). TM3 and CTM-2 show a similar trend between early and late flights with somewhat reduced tropospheric background CO levels during flights 3 to 7. TOMCAT and LMDz-INCA also do simulate low CO concentrations during these flights but the values are not significantly lower than during the later flights. These two models generally underestimate CO significantly during flights 9 to 16 during which advection of air from midlatitudes prevailed. This suggest that meridional gradients in CO and ozone concentrations between tropical and midlatitude regions are underestimated by the models, particularly by TOMCAT and LMDz-INCA. The most pronounced case of a simultaneous drop in both CO and ozone observed on flight 3, however, is reproduced qualitatively by all models.

### 3.2.3 Aircraft emission signatures

A variety of nitrogen oxides sources were identified during SONEX including aircraft emissions, lightning, and upward transport from continental sources (Thompson et al., 1999). A stagnant anticyclone located over the eastern North Atlantic during October 20–23 provided ideal conditions to study the impact of air traffic emissions (flights 6 and 7). Air masses had followed multiple looping paths in the North Atlantic flight corridor during the previous 3-4 days (Fuelberg et al., 2000). Enhanced NO concentrations are seen towards the end of flight 6 and in the center of flight 7. In both cases the fraction of NO emitted by air traffic (grey line in Fig. 4c) as calculated with TOMCAT reaches up to 80% and higher. Similar enhancements are not seen in tracers of continental pollution like CO and PAN providing clear evidence for a dominating source from air traffic in these cases. The model results suggest that several of the observed NO plumes are almost entirely due to air traffic emissions. In a similar analysis (Meijer et al., 2000) showed that good agreement with observed NO on these flights could only be achieved when aircraft emissions were included in the TM3 model. We conclude that the air traffic emission source is well represented in the models both in terms of its spatial distribution and in terms of its absolute strength since the observed NO signatures are followed nicely in cases where this source is dominating.

3.2.4 Lightning and surface sources

Lightning was another very important source of NO_x during SONEX and signatures of lightning activity were observed on many flights, especially those out of Bangor and the Azores (Fuelberg et al., 2000; Jeker et al., 2000). The most prominent examples are seen on flights 10, 12, and 14 with many narrow spikes in NO of 0.4 ppb and more. These peaks are mostly missed by the models and in general the signatures are strongly underestimated. The absence of narrow spikes in the models is likely due to insufficient grid resolution missing small-scale but intense convection. However, some of the enhancements have horizontal dimensions of more than one model grid box. The largest plumes on flight 10 and 12 spread over about 500 km along the flight track and the size of the plume on flight 14 approaches 1000 km. TM3 simulates the strongest enhancements on flight 12 but it fails during flight 10 and 14. The contributions from aircraft emissions and lightning as simulated by TM3 was also analysed previously by Meijer et al. (2000). However, their study only included flights 4, 6, and 7 with low and moderate lightning contributions for which reasonable agreement was found. Only LMDz-INCA calculates a marked increase on flight 14 yet still substantially smaller than observed. A detailed investigation of flight 14 by Jeker et al. (2000) showed that the NO enhancement was caused by strong lightning activity associated with a cold front over the North Atlantic off the coast of the United states. Observations from space by the Optical Transient Detector (OTD) (http://thunder.msfc.nasa.gov/data/OTDsummaries/gifs/1997_world_son.gif) show substantial lightning activity over the warm waters of the Gulf Stream off the US coast in fall 1997. Our results suggest that this lightning in marine convection is underestimated substantially by the parameterizations used in current models. This appears to be the main reason for an underestimation of NO averaged over flights 9–16 of 40 to 60% by the models (see Table 1).

Contributions from surface sources can be identified as simultaneous increases in CO and PAN. The highest CO concentration was measured on flight 12 and was associated with increased PAN. Air parcel trajectories for this air mass do not show significant vertical motion (Fuelberg et al., 2000) suggesting that surface pollution had been transported to the upper troposphere by convection, which is not resolved by the trajectory model. The models simulate only minor increases
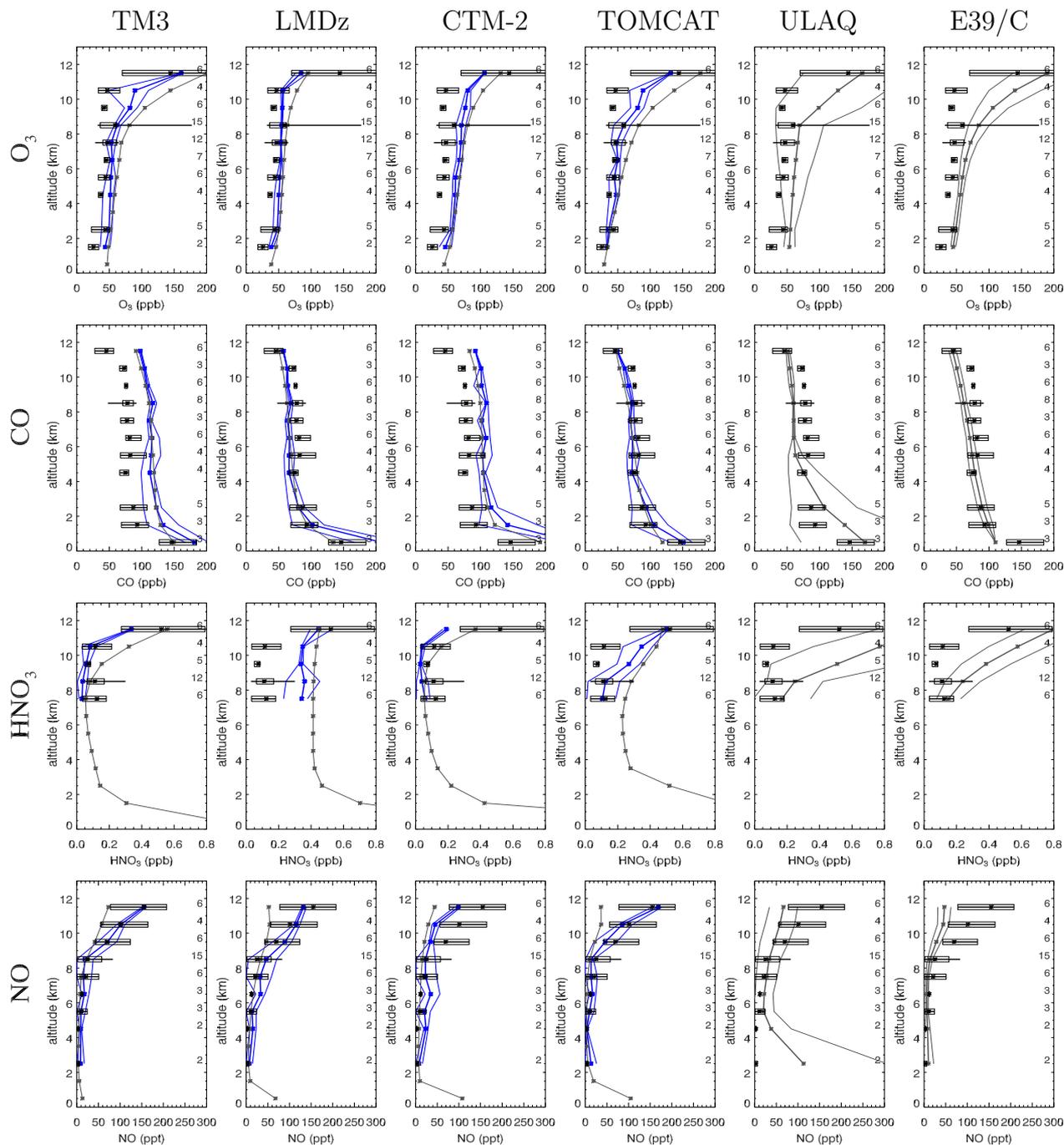
Fig. 5. SONEX profiles over Shannon, Ireland. See Fig. 3 for explanation of symbols.

in CO, PAN and Radon here, well below the amplitude of other events. Enhancements in CO and PAN observed on flights 9 and 11, however, are captured quite well by the models. Notably, although flight 9 was carried out over the eastern North Atlantic it experienced significant pollution outflow from the United States as suggested by the high Radon and CO values as well as by the trajectory analysis.

It may be concluded that plumes of surface pollutants in the upper troposphere are well represented if the upward transport occurred on scales resolved by the model winds but less so if convective transport was involved. However, clearly more case studies are needed to confirm this limited analysis. PAN concentrations are on average strongly overestimated by TM3 and CTM-2 (Fig. 4e). Except for air masses
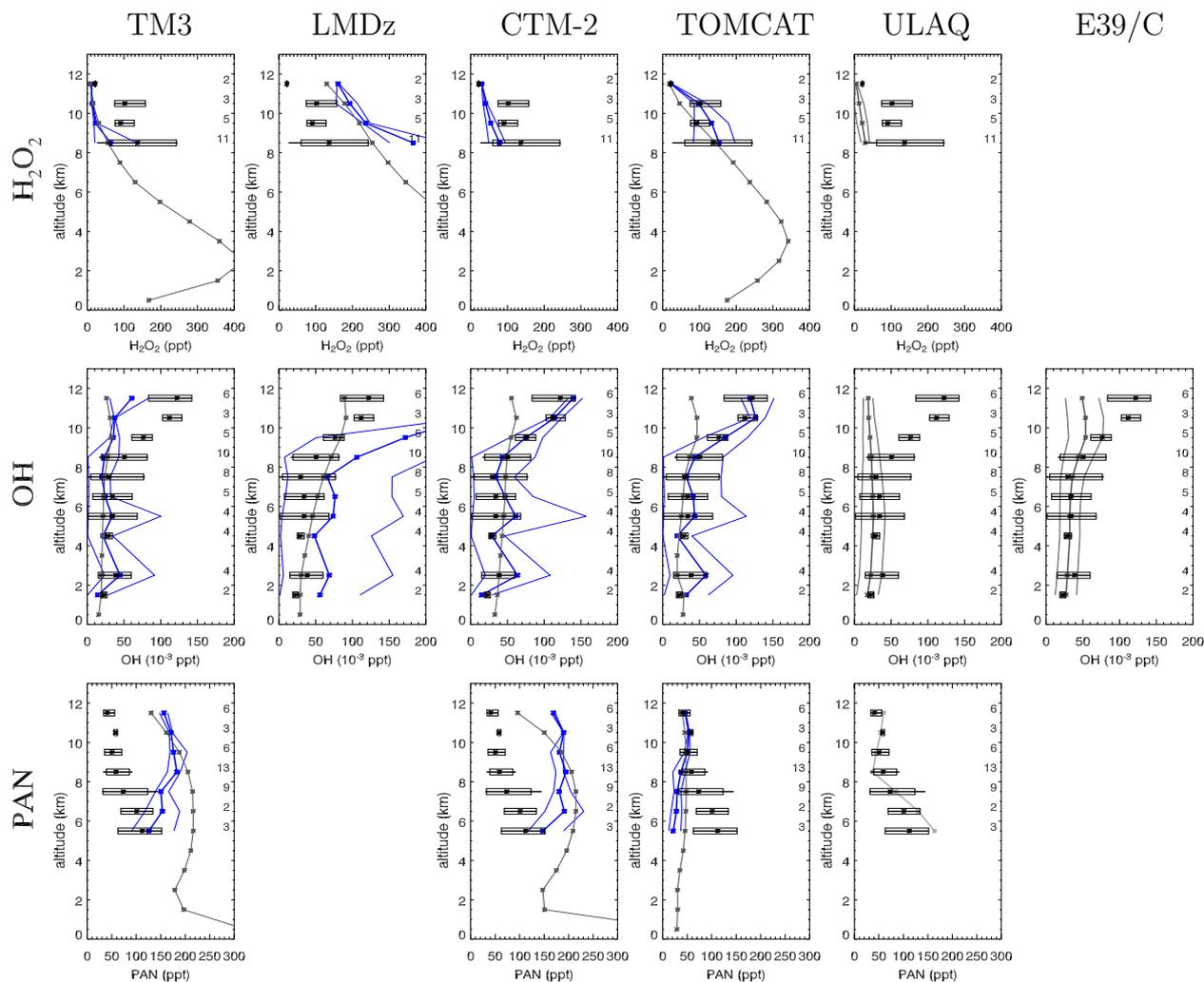
Fig. 5. Continued.

with strong continental signatures PAN is well represented in TOMCAT, which on the other hand was low in PAN over the South Pacific.

3.2.5 Radical chemistry and ozone production

The HO_x radicals, hydrogen peroxide and water vapour are shown in panels f)-i) of Fig. 4. Water vapour fields in TM3, TOMCAT and CTM-2 are taken from the driving ECMWF model. Only values for TOMCAT and the LMDz-INCA GCM are therefore shown in Fig. 4g. The LMDz-INCA model is "wetter" than ECMWF but the simulated values are in excellent agreement with observed values. The ECMWF model tends to underestimate water vapour in the upper troposphere probably because the model does not allow for supersaturation with respect to ice. The primary production of the hydroxyl radical (OH) from reaction of O^1D with H_2O is therefore underestimated in TM3 as already noted by Mei-

jer et al. (2004), and the same is most likely true for TOM-CAT and CTM-2. The overall abundance of HO_x radicals ($OH+HO_2$) is controlled by their sources and sinks whereas the ratio $OH:HO_2$ is largely determined by the HO_x cycling through the reaction of NO with HO_2 and the reaction of CO and NMHCs with OH (Wennberg et al., 1998).

Simulated OH and HO_2 concentration levels generally agree well with the observations and the variability is very well reproduced. TM3 tends to simulate too low OH concentrations in the UT whereas HO_2 is in good agreement with the observations. Accordingly the $OH:HO_2$ ratio is lower than observed (not shown) which is most likely a result of the elevated CO concentrations in TM3. The $OH:HO_2$ ratio is also too low in CTM2-Gauss probably for the same reason. Different from TM3, however, OH levels in CTM2-Gauss are close to the observed ones whereas HO_2 is overestimated. The LMDz-INCA model overestimates the abundance of OH which may be partly due to missing NMHC

chemistry since oxidation with NMHCs would reduce the ratio OH:HO$_2$. However, HO$_x$ levels are too high in this model as a whole and the reasons for this are not quite clear. Primary HO$_x$ production through photolysis of O$_3$ should be well represented since both simulated O$_3$ and H$_2$O are close to the observed values. In agreement with the elevated OH values HNO$_3$ is too high and CO too low in LMDz-INCA because of excessive oxidation of NO$_2$ and CO by OH, respectively. The absence of PAN chemistry is another likely reason for the elevated HNO$_3$ since too much NO$_x$ is going into HNO$_3$ instead of other reservoirs such as PAN. Both the OH and HO$_2$ levels in TOMCAT, on the other hand, are close to the observations and hence provide no clue for the low CO values in this model.

Hydrogen peroxide, which forms through the reaction of HO$_2$ with HO$_2$, is higher in LMDz-INCA than in the other models and in the observations (Fig. 4h). Note that the agreement was much better for PEM-Tropics A. TOMCAT, which tended to underestimate H$_2$O$_2$ over the South Pacific shows the best agreement with the SONEX observations, whereas CTM-2 and in particular TM3 are substantially too low. Differences in convective transport of H$_2$O$_2$ (Jaeglé et al., 2001) as well as differences in photolysis rates and washout of H$_2$O$_2$ are likely dominating the large variations between the models, since differences in OH (reaction with OH provides another sink for H$_2$O$_2$) and in HO$_2$ levels can not explain the results, or are rather contrary to what would be expected. Photolysis of peroxides provides an important source of HO$_x$ in the upper troposphere (Jaeglé et al., 2001), and hence the enhanced H$_2$O$_2$ levels may be rather the origin than the consequence of the elevated HO$_x$ in the LMDz-INCA model. A similar chicken and egg problem concerns the CO observations of the TM3 model. Without a more thorough investigation it is not clear whether an underestimation of OH causes too high CO or whether an overestimation of CO causes too low OH in this model.

3.2.6 Vertical profiles at Shannon

Vertical profiles obtained at Shannon, Ireland, are shown in Fig. 5. The shape of the ozone and CO profiles are generally well reproduced. Deviations from the monthly mean profile as seen by the models are also evident in the measured profiles. HNO$_3$ and ozone are too high in ULAQ and E39/C in the the upper troposphere. However, the stratospheric influence at these altitudes was clearly below average during the SONEX flights as suggested by the differences between point-by-point and monthly mean profiles in the other models. The PAN profile is only reproduced successfully by the ULAQ model whereas TM3 and CTM-2 overestimate PAN in the upper troposphere by at least a factor of three. The H$_2$O$_2$ profile over Shannon is best represented in TOM-CAT which, notably, performed poorly over Tahiti. TM3 and ULAQ simulate too little OH in the upper troposphere, possibly because of missing acetone chemistry which could be

an important source of upper tropospheric OH (Jaeglé et al., 2001). If we assume that monthly mean OH concentrations (averaged over day and night time) are about a factor of two lower than the values measured by SONEX only during daytime, then the E39/C values in the upper troposphere are probably quite close to the observations. OH profiles simulated by CTM-2 and TOMCAT, on the other hand, match the observations exceptionally well. As noted before, the agreement with measured H$_2$O$_2$ is generally less good for SONEX than for PEM-Tropics A, and differences between the models are quite large. The ULAQ model underestimates H$_2$O$_2$ similar to TM3 and CTM-2.

3.3 Quantitative analysis of the PEM-Tropics A and SONEX results

Figures 6 and 7 show Taylor diagrams corresponding to the PEM-Tropics A and the SONEX data presented in the previous section, respectively. The four panels in each figure correspond to a) the whole data set of measurements at cruise altitude, b) and c) two distinct subparts of each campaign as discussed in the previous sections, and d) the vertical profile results at Tahiti and Shannon, respectively. The diagrams allow comparison of the skill scores of the models and to show quantitatively which trace species can be better simulated than others given the different conditions encountered during the campaigns. Grey contours in the panels are isolines of skill score as defined by Eq. (1). The angle between a point in a diagram and the vertical axis indicates the correlation coefficient, and the distance from the origin (lower left corner) is the normalized standard deviation. The point denoted as "Ref" indicates the optimal point with correlation coefficient=1 and normalized standard deviation=1. The linear distance from this reference point is proportional to the pattern RMS error. The formula for the skill score (Eq. 1) includes a term for the maximum attainable correlation R_0. In principle this number would be different for each species because it depends, among other factors, on the measurement precision. Correspondingly, different skill score lines would have to be drawn for each species separately. However, since we only use 5-min averages for which the limited measurement precision is usually of minor importance, we only account for the limitations induced by the coarse model resolution and by errors in the meteorological fields driving the CTMs. As a rough measure for this we used the correlation between measured and modelled temperature as already done in Brunner et al. (2003). R_0 derived in this way is typically of the order of 0.95.

Results for subsets of the PEM-Tropics A data over the eastern and western South Pacific, respectively (Figs. 6b and c), show significant differences among each other but also when compared to the complete data set (Fig. 6a). Individual points of the four models often form clusters for a given tracer. As an example, points representing the performance of TM3 (T3), CTM-2 (C1), and TOMCAT (TC) with
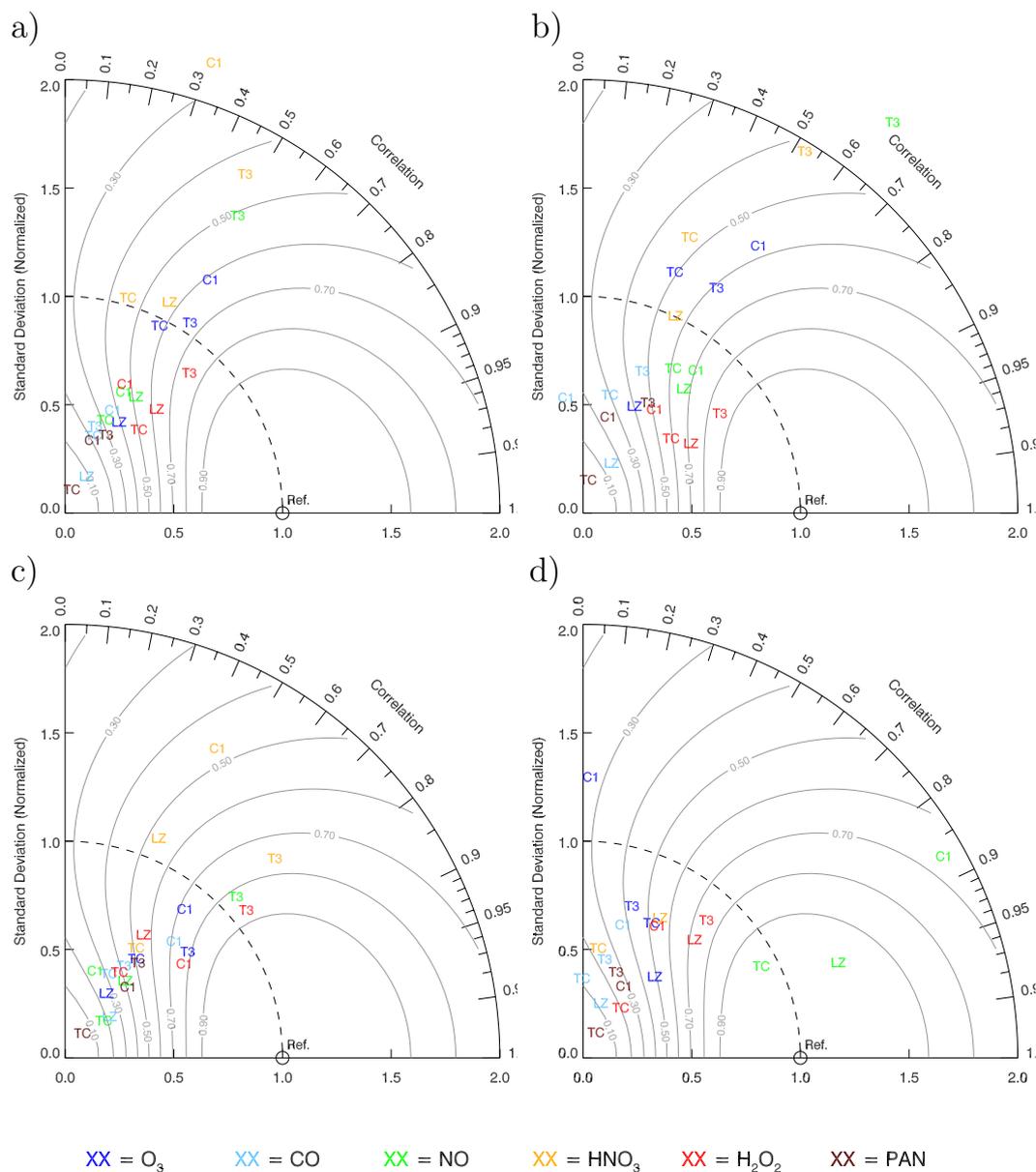
Fig. 6. Taylor diagrams of model performance for the PEM-Tropics A campaign. **a)** All measurements at p <400 hPa and between 0 and 35° S latitude (same data as in Fig. 2), **b)** flights 5–10 over eastern South Pacific only, **c)** flights 15–18 over western South Pacific only, **d)** vertical profiles at Tahiti (same data as in Fig. 3). All points of a given model are represented by a specific label: T3=TM3, C1=CTM-2 version 1, TC=TOMCAT, and LZ=LMDz-INCA. Different colors indicate different tracers (see legend at the bottom). Grey contours are isolines of model skill as defined in Eq. (1).

respect to ozone (dark blue) are clustered together in panels (a–c) with similar correlation coefficients and standard deviations, and hence similar scores. The LMDz-INCA model underestimates ozone over the South Pacific the strongest (see Table 1) and also shows a reduced performance in terms of ozone variability. However, with respect to NO (green) the LMDz-INCA model forms a cluster together with CTM-2 and TOMCAT, whereas TM3 shows a similar correlation but a much higher variability. In the area of the subtropical

anticyclone TM3 strongly overestimates both the NO variability (Fig. 6b) and the mean concentration (Table 1). Conversely, over the western South Pacific, which is frequently influenced by marine convection and moderate lightning activity, TM3 shows by far the best results in terms of NO (Fig. 6c). Carbon monoxide (light blue) which was apparently dominated by biomass burning sources during PEM-Tropics A, is rather poorly simulated by all models with the exception of CTM-2 over the western South Pacific. The

highest correlations are generally seen for hydrogen peroxide (red) but the variability is mostly underestimated. Simulated PAN concentrations compare poorly with the measurements in terms of skill score. Mean concentration levels are well reproduced only by CTM2 (see Table 1), while they are generally overestimated in TM3 and strongly underestimated in TOMCAT. A good representation of PAN, however, is quite important because long-range transport of PAN from biomass burning regions in the middle and upper troposphere followed by subsidence and subsequent decomposition dominates the supply of NO_x in the lower troposphere (Schultz et al., 1999; Staudt et al., 2002). Except for LMDz-INCA all models show a reduced performance in terms of HNO_3 over the eastern South Pacific, and the concentrations are strongly overestimated by all models by 100% to 280%. Similar differences over the South Pacific were reported previously for other models (Wang et al., 1998; Bey et al., 2001) and possible reasons for this discrepancy were discussed in the previous section. Simulated HNO_3 generally compares much better with observed values over the western South Pacific.

Figure 6d) representing the profile measurements at Tahiti indicates how well the different shapes in the vertical profiles are simulated. The high correlations for NO in all models basically reflects the fact that the observed increase in NO with increasing altitude is well represented in the models. However, the amplitude of this increase is quite strongly overemphasized in LMDz-INCA, CTM-2, and particularly in TM3 (point lies outside of figure domain). The missing maximum in simulated CO at mid-levels as observed from the DC-8 results in generally poor skill scores. Trends in the HNO_3 profiles are opposite to the observations in CTM-2 and TM3 which leads to negative correlations for these models (points would appear to the left outside of the figure).

Results for SONEX are shown in Fig. 7. Again, Figs. 7b) (early flights 3–7) and c) (late flights 9–16) show the results for subsets of the entire campaign and the corresponding mean measured concentrations and model biases are presented in Table 1. There are some similarities with PEM-Tropics A results but also striking differences.

For instance, the performance with respect to CO is again poor in all models, yet probably for a different reason. Over the SONEX domain the stratospheric influence explains much of the CO variability, and, as discussed earlier, the sharp concentration gradients across the tropopause appear not to be well reproduced by the models. While CTM-2 and TOMCAT were grouped together in terms of mean CO bias during PEM-Tropics A (Table 1), it is now the LMDz-INCA and TOMCAT models that produce very similar results at northern mid-latitudes (cf. Figs. 2 and 4). Skill scores for H_2O_2 tend to be lower during SONEX than during PEM-Tropics A, and the scores for NO are even dramatically reduced. This is due to the frequent encounters of lightning-produced NO peaks during the later flights, which are not captured adequately by the models. For the early flights 3–7 (Fig. 7b), however, which showed little lightning in-

fluence and were dominated by air traffic emissions, model skills in terms of NO are comparable to PEM-Tropics A results. Measured HNO_3 concentrations were on average 3 to 4 times higher during SONEX as compared to PEM-Tropics A. In contrast to PEM-Tropics A there is no general tendency among the models in terms of HNO_3 bias (Table 1). Some models are high compared to the measurements while others are low. It is interesting to note that these deviations are not mirrored by similar deviations in NO. NO and HNO_3 were poorly correlated in the upper troposphere during SONEX and ratios of NO to HNO_3 were usually well above the values expected from photochemical equilibrium (Jaeglé et al., 1999), indicating the presence of significant fresh NO sources to the North Atlantic such as aircraft emissions and lightning activity. HNO_3, on the other, has a sufficiently long lifetime to be transported over long distances and hence to be influenced by sources remote from the North Atlantic. Furthermore, downward transport from the stratosphere makes an important contribution to upper tropospheric HNO_3 but only very little to NO (Meijer et al., 2000). Differences between the models in their treatment of convective transport, lightning activity, HNO_3 washout and stratospheric influx are all likely to contribute to the differing signatures and mean concentrations of NO and HNO_3. Aircraft emissions, which are treated identically in the models, constituted a significant fraction of total NO observed during several SONEX flights (Meijer et al., 2000). This likely explains why the models behave much more similarly with respect to NO than with respect to HNO_3 (see Fig. 4), especially on the early flights with little lightning contribution. On these flights (3-7), the model performance is better in terms of NO than in terms of HNO_3 whereas the opposite is true for the later flights 9–16.

Figure 7 underscores the earlier notion that the variability in OH radical concentrations are well reproduced since skill scores for OH are often the highest of all species. Model biases with respect to the observations (see Table 1) were already discussed earlier. Interestingly, modelled to measured OH ratios (see Table 1) are generally much larger during the first part of the campaign (flights 3–7) where transport from low latitudes was more frequently observed than during the second part (flights 9–16). This is probably related to the significant change in mean CO levels observed between these two periods which is not well captured by the models. In addition, NO concentrations were often underestimated during the second part of the campaign. Due to the reaction of NO with HO_2 additional NO tends to shift the OH:HO_2 ratio towards higher OH. Thus, if the models had simulated higher NO values during flights 9–16 they would likely have overestimated OH in a similar way as on the earlier flights.

Crowther et al. (2002) noted that large-scale models tend to overestimate the primary OH production (from photolysis of O_3 and subsequent reaction of O^1D with H_2O) due to small-scale variability of O_3 and H_2O not resolved in the models, because these species are strongly anticorrelated in
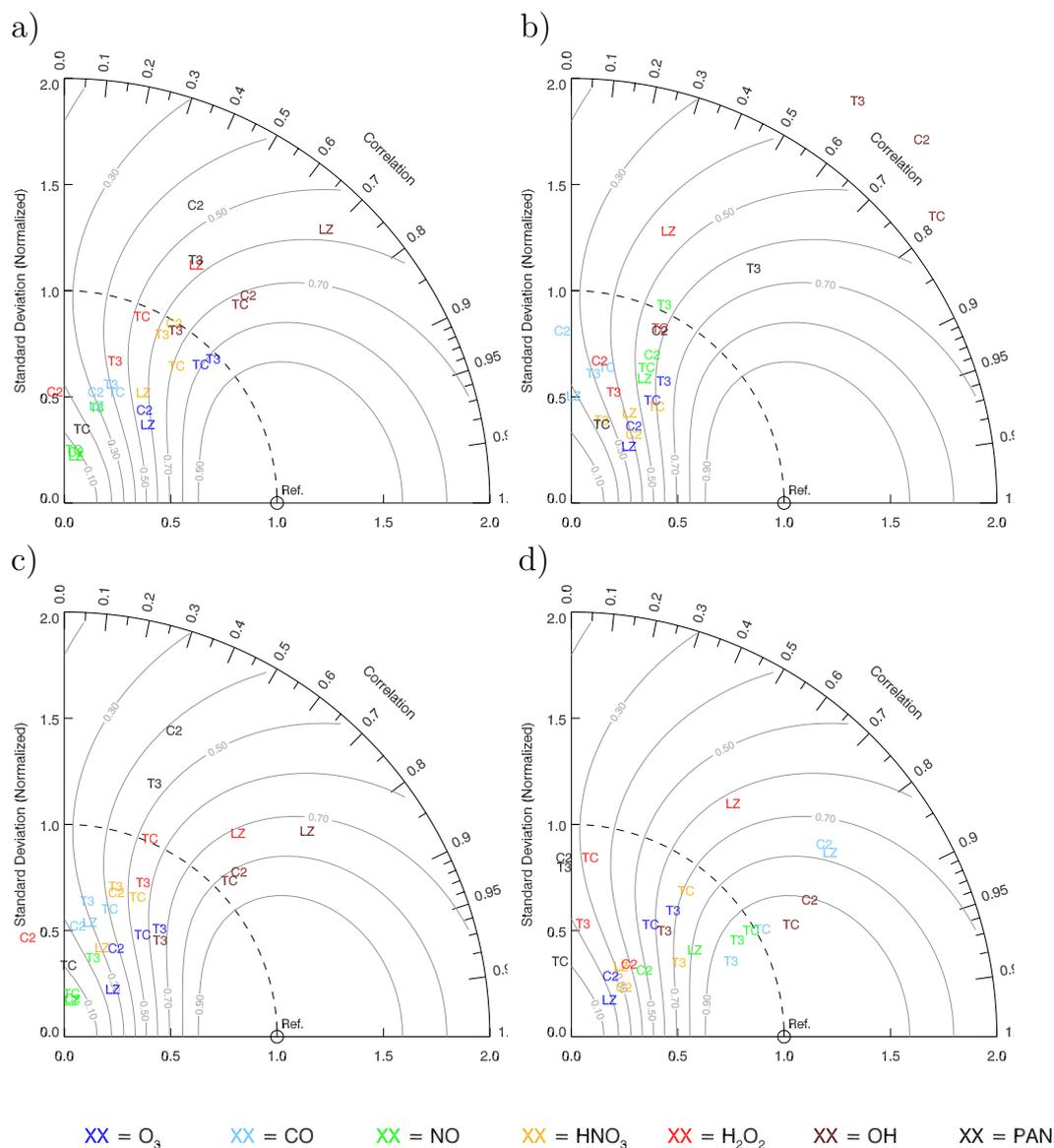
Fig. 7. Taylor diagrams of model performance for the SONEX campaign. **a)** All measurements at p <350 hPa (same data as in Fig. 4), **b)** early flights 3–7 only, **c)** late flights 9–16 only, **d)** vertical profiles at Shannon (same data as in Fig. 5). For further explanations refer to Fig. 6.

the tropopause region. Water vapour concentrations, on the other hand, tend to be somewhat too low in the ECMWF model and hence in TM3, TOMCAT and CTM-2 (see Fig. 4i) which acts in the opposite way. From our results we can not draw any general conclusion on whether current CTM models tend to under- or overestimate HO_x levels because the differences between the models were found to be too large to make such a claim.

Differences in mean H_2O_2 concentrations are significant between the models and are difficult to understand as discussed in section 3.2.5. In general, the skill scores for H_2O_2 are significantly lower for SONEX than for PEM-Tropics A,

probably because H_2O_2 was close to a photochemical steady state over the remote tropical South Pacific (Schultz et al., 1999) and hence relatively easy to simulate for the models.

Skill scores with respect to the shape of the vertical profiles at Shannon (Fig. 7d) show a generally good performance of the models with the exception of H_2O_2 and PAN. Results for OH, NO and CO mostly show very high skill scores.

4 Conclusions

A rigorous evaluation of the performance of a number of chemistry transport and chemistry-climate models was

performed by comparing interpolated model fields "point-by-point" with their measured counterparts. In this second part of the study a detailed analysis was carried out using in-situ observations from the PEM-Tropics A campaign over the remote tropical Pacific and the SONEX campaign over the more polluted North Atlantic.

1. Conclusions drawn from the comparison with PEM-Tropics A:

Simulated NO concentrations agree reasonably well with observed values in general. The TM3 model, however, substantially overestimates NO over the remote eastern South Pacific (flights 5–10) where large-scale descent in a quasi-stationary anticyclone was prevailing, but TM3 performed best over the western South Pacific near fresh lightning sources (flights 15–18) possibly owing to the special lightning parametrization in TM3 (Meijer et al., 2001). HNO_3 levels are very different between the models and are generally too high except for TOMCAT. Excessive downward transport from the stratosphere likely contributes to this problem over the eastern South Pacific, in particular in CTM-2. CO levels are generally too high in TM3 and too low in LMDz-INCA but they are very close to the observations in CTM-2 and TOMCAT. The low CO in LMDz-INCA is probably related to the too high OH concentrations. Most biomass burning plumes visible in the observations as concurrent enhancements in CO, PAN and O_3 are missing in the models. However, the excellent correlation between simulated Radon and measured CO in some of these cases indicates that the transport of continentally influenced air is well simulated but that there are deficiencies in the representation of actual biomass burning sources in the models, partly because the simulations were based on climatological biomass burning emissions. Hydrogen peroxide is generally very well simulated suggesting a good representation of HO_x radicals in the models. TOMCAT shows the largest differences with too low values especially over the eastern South Pacific. PAN concentrations were too low in TOMCAT but close to the observed levels in TM3 and CTM-2.

2. Conclusions drawn from the comparison with SONEX:

NO concentrations were often dominated by air traffic emissions. The good correspondence between simulated and measured NO for these cases suggests an adequate representation of the distribution and strength of this source in the models. Lightning plumes encountered on flights 10, 12, and 14 are almost completely missing in the models. This source appears to be significantly underestimated for the conditions of the SONEX measurements which were often affected by oceanic convection over the warm Gulf Stream off the US east coast (Jeker et al., 2000). Sharp tracer gradients across the tropopause are usually not well represented in the models, in particular in the case of narrow stratospheric intrusions. Excessive mixing between tropospheric and stratospheric air is typically seen in these cases which may lead to systematic errors in the representation of photochemistry near the tropopause because of strong correlations among the trace gases such as O_3 and H_2O (Crowther et al., 2002). Events of elevated CO and PAN indicating vertical transport from continental sources could be well reproduced by the models except for one case which appears to have been dominated by sub-grid scale convective transport. H_2O_2 levels differ strongly between the models and the differences are difficult to understand in terms of simple considerations of the chemical sources and sinks of H_2O_2. Different from PEM-Tropics A H_2O_2 appears to be much less in a chemical steady state during SONEX but is likely influenced by recent convective transport and washout, which are simulated differently by the models. Model simulated OH and HO_2 radical concentrations, on the other hand, agree well with the observations except for TM3 simulating generally too low and LMDz-INCA too high OH levels in the upper troposphere.

Interestingly, models performing best with respect to a given species over the South Pacific often showed a comparatively poor performance over the North Atlantic and vice versa. This is true for instance for PAN and H_2O_2 simulated by TOMCAT which were both significantly too low during PEM-Tropics A but very close to the observations during SONEX.

3. Performance of the ULAQ and E39/C models: Only a limited evaluation of these two models was performed here based on the comparison between monthly mean model and campaign averaged profiles at Tahiti and at Shannon. Both models show the same tendency as the other models in overestimating HNO_3 over the South Pacific ocean. H_2O_2 and PAN profiles of the ULAQ model compare well with the observations except for the missing PAN maximum in the middle troposphere which is a common feature of all model simulations. In both models the CO levels are in good agreement with the observations except for the enhanced CO values below 2 km altitude seen only in E39/C but not in the other models and the observations. Both models also tend to overestimate HNO_3 in the upper troposphere over the North Atlantic probably due to a too strong stratospheric influence. In the case of E39/C missing NMHC chemistry may also contribute to this problem since too much NO_x is likely going into HNO_3 instead of other reservoirs such as PAN. CO levels are again close to the observations with a tendency to slightly underestimate the CO abundance in the upper troposphere. The observed PAN profile over Shannon is well represented in ULAQ but both OH and H_2O_2 are significantly too low in the upper troposphere. Conversely, upper tropospheric OH levels are well matching the observed abundances in E39/C.

The point-by-point approach allows for a much more specific evaluation of model performance with respect to the discussion of individual processes than comparisons of climatological distributions, because different situations in which one or another process is dominating can be analysed separately. In order to better support the interpretation of the differences between models and observations it would be highly desirable to combine such a point-by-point comparison with

sensitivity studies in which, for instance, individual emission sources are switched off or parameterizations are varied. The TOMCAT simulations with and without air traffic emissions is the only such example used in the present study but it clearly demonstrates the large potential of such analyses. More sensitivity studies are planned to be undertaken in the upcoming EU project QUANTIFY.

References

Berntsen, T. K., Myhre, G., Stordal, F., and Isaksen, I. S. A.: Time evolution of tropospheric ozone and its radiative forcing, J. Geophys. Res., 105, 8915–8930, 2000.

Bey, I., Jacob, D. J., Yantosca, R. M., Logan, J. A., Field, B. D., Fiore, A. M., Li, Q., Liu, H. Y., Mickley, L. J., and Schultz, M. G.: Global modelling of tropospheric chemistry with assimilated meteorology: Model description and evaluation, J. Geophys. Res., 106, 23073–23096, 2001.

Board, A. S., Fuelberg, H. E., Gregory, G. L., Heikes, B. G., Schultz, M. G., Blake, D. R., Dibb, J. E., Sandholm, S. T., and Talbot, R. W.: Chemical characteristics of air from differing source regions during the Pacific Exploratory Mission-Tropics A (PEM-Tropics A), J. Geophys. Res., 104, 16181–16196, 1999.

Brasseur, G. P., Cox, R. A., Hauglustaine, D., Isaksen, I., Lelieveld, J., Lister, D. H., Sausen, R., Schumann, U., Wahner, A., and Wiesen, P.: European scientific assessment of the effects of aircraft emissions, Atmos. Envir., 32, 2329–2418, 1998.

Bregman, A., Krol, M. C., Teyssèdre, H., Norton, W. A., Iwi, A., Chipperfield, M., Pitari, G., Sundet, J. K., and Lelieveld, J.: Chemistry-transport model comparison with ozone observations in the midlatitude lowermost stratosphere, J. Geophys. Res., 106, 17479–17496, 2001.

Brenninkmeijer, C., Crutzen, P., Fischer, H., Güsten, H., Hans, W., Heintzenberg, J., Hermann, M., Immelmann, T., Kersting, D., Maiss, M., Nolle, N., Pitscheider, A., Pohlkamp, H., Scharffe, D., and Wiedensohler, A.: CARIBIC – Civil aircraft for global measurement of trace gases and aerosols in the tropopause region, J. Atmos. Ocean. Technol., 16, 1373–1383, 1999.

Brunner, D., Staehelin, J., and Jeker, D.: Large-scale nitrogen oxide plumes in the tropopause region and implications for ozone, Science, 282, 1305–1309, 1998.

Brunner, D., Staehelin, J., Jeker, D., Wernli, H., and Schumann, U.: Nitrogen oxides and ozone in the tropopause region of the Northern Hemisphere: Measurements from commercial aircraft in 1995/1996 and 1997, J. Geophys. Res., 106, 27673–27699, 2001.

Brunner, D., Staehelin, J., Rogers, H. L., et al.: An evaluation of the performance of chemistry transport models by comparison with scientific aircraft observations. Part 1: Concepts and overall model performance., Atmos. Chem. Phys., 3, 1609–1631, 2003, **SRef-ID: 1680-7324/acp/2003-3-1609**.

Chatfield, R. B.: Anomalous HNO_3/NO_x ratio of remote tropospheric air: Conversion of nitric acid to formic acid and NO_x?, Geophys. Res. Lett., 21, 2705–2708, 1994.

Chipperfield, M. P.: Multiannual simulations with a three-dimensional chemical transport model, J. Geophys. Res., 104, 1781–1805, 1999.

Crowther, R. A., Law, K. S., Pyle, J. A., Bekki, S., and Smit, H. G. J.: Characterising the effect of large-scale model resolution upon calculated OH production using MOZAIC data, Geophys. Res. Lett., 29, 1613, doi:10.1029/2002GL014660, 2002.

Emmons, L. K., Carroll, M. A., Hauglustaine, D. A., Brasseur, G. P., Atherton, C., Penner, J., Sillman, S., Levy II, H., Rohrer, F., Wauben, W. M. F., van Velthoven, P. F. J., Wang, Y., Jacob, D., Bakwin, P., Dickerson, R., Doddridge, B., Gerbig, C., Honrath, R., Hubler, G., Jaffe, D., Kondo, Y., Munger, J. W., Torres, A., and Volz-Thomas, A.: Climatologies of NO_x and NO_y: A comparison of data and models, Atmos. Envir., 31, 1851–1904, 1997.

Emmons, L. K., Hauglustaine, D. A., Müller, J.-F., Carroll, M. A., Brasseur, G. P., Brunner, D., Staehelin, J., Thouret, V., and Marenco, A.: Data composites of airborne observations of tropospheric ozone and its precursors, J. Geophys. Res., 105, 20497–20538, 2000.

Fenn, M. A., Browell, E. V., Butler, C. F., Grant, W. B., Kooi, S. A., Clayton, M. B., Gregory, G. L., Newell, R. E., Zhu, Y., Dibb, J. E., Fuelberg, H. E., Anderson, B. E., Bandy, A. R., Blake, D. R., Bradshaw, J. D., Heikes, B. G., Sachse, G. W., Sandholm, S. T., Singh, H. B., Talbot, R. W., and Thornton, D. C.: Ozone and aerosol distributions and air mass characteristics over the South Pacific during the burning season, J. Geophys. Res., 104, 16197–16212, 1999.

Fuelberg, H. E., Newell, R. E., Longmore, S. P., Zhu, Y., Westberg, D. J., Browell, E. V., Blake, D. R., Gregory, G. L., and Sachse, G. W.: A meteorological overview of the Pacific Exploratory Mission (PEM) Tropics period, J. Geophys. Res., 104, 5585–5622, 1999.

Fuelberg, H. E., Hannan, J. R., van Velthoven, P. F. J., Browell, E. V., Bieberach, G. J., Knabb, R. D., Gregory, G. L., Pickering, K. E., and Selkirk, H. B.: A meteorological overview of the Subsonic Assessment Ozone and Nitrogen Oxide Experiment (SONEX) period, J. Geophys. Res., 105, 3633–3651, 2000.

Grant, W. B., Browell, E. V., Butler, C. F., et al.: A case study of transport of tropical marine boundary layer and lower tropospheric air masses to the northern midlatitude upper troposphere, J. Geophys. Res., 105, 3757–3770, 2000.

Gregory, G. L., Westberg, D. J., Shipham, M. C., Blake, D. R., Newell, R. E., Fuelberg, H. E., Talbot, R. W., Heikes, B. G., Atlas, E. L., Sachse, G. W., Anderson, B. A., and Thornton, D. C.: Chemical characteristics of Pacific tropospheric air in the region of the Intertropical Convergence Zone and South Pacific Convergence Zone, J. Geophys. Res., 104, 5677–5696, 1999.

Grewe, V., Brunner, D., Dameris, M., Grenfell, J. L., Hein, R., Shindell, D., and Staehelin, J.: Origin and variability of upper tropospheric nitrogen oxides and ozone at northern mid-latitudes, Atmos. Envir., 35, 3421–3433, 2001.

Hauglustaine, D. A. and Brasseur, G. P.: Evolution of tropospheric ozone under anthropogenic activities and associated radiative forcing of climate, J. Geophys. Res., 106, 32337–32360, 2001.

Hauglustaine, D. A., Ridley, B. A., Solomon, S., Hess, P. G., and Madronich, S.: HNO_3/NO_x ratio in the remote troposphere during MLOPEX2: Evidence for nitric acid reduction on carbonaceous aerosols, Geophys. Res. Lett., 23, 2609–2612, 1996.

Hauglustaine, D. A., Hourdin, F., Jourdin, L., Filiberti, M.-A., Walters, S., Lamarque, J.-F., and Holland, E. A.: Interactive chemistry in the Laboratoire de Meteorologie Dynamique general circulation model: Description and background tropospheric chemistry evaluation, J. Geophys. Res., 109, D04 314, doi:10.1029/2003JD003 957, 2004.

Hein, R., Dameris, M., Schnadt, C., Land, C., Grewe, V., Köhler, I., Ponater, M., Sausen, R., Steil, B., Landgraf, J., and Brühl, C.: Results of an interactively coupled atmospheric chemistry-general circulation model: Comparison with observations, Ann. Geophys., 19, 435–457, 2001,
SRef-ID: 1432-0576/ag/2001-19-435.

Hoell, J. M., Davis, D. D., Jacob, D. J., Rodgers, M. O., Newell, R. E., Fuelberg, H. E., McNeal, R. J., Raper, J. L., and Bendura, R. J.: Pacific Exploratory Mission in the tropical Pacific: PEM-Tropics A, August–September 1996, J. Geophys. Res., 104, 5567–5583, 1999.

Horowitz, L. W., Walters, S., Mauzerall, D. L., Emmons, L. K., Rasch, P. J., Granier, C., Tie, X., Lamarque, J.-F., Schultz, M. G., Tyndall, G. S., Orlando, J. J., and Brasseur, G. P.: A global simulation of tropospheric ozone and related tracers: Description and evaluation of MOZART, version 2, J. Geophys. Res., 108, doi:10.1029/2002JD002 853, 2003.

Houghton, J. T., Ding, Y., Griggs, D. J., Noguer, M., van der Linden, P. J., and Xiaosu, D. (Eds): IPCC Third Assessment Report: Climate Change 2001: The Scientific Basis, Cambridge University Press, United Kingdom, 2001.

Jacob, D. J.: Heterogeneous chemistry and tropospheric ozone, Atmos. Envir., 34, 2131–2159, 2000.

Jacob, D. J., Prather, M., Rasch, P., et al.: Evaluation and intercomparison of global atmospheric transport models using Rn-222 and other short-lived tracers, J. Geophys. Res., 102, 5953–5970, 1997.

Jaeglé, L., Jacob, D. J., Brune, W. H., Faloona, I. C., Tan, D., Kondo, Y., Sachse, G. W., Anderson, B., Gregory, G. L., Vay, S., Singh, H. B., Blake, D. R., and Shetter, R.: Ozone production in the upper troposphere and the influence of aircraft during SONEX: Approach to NO_x saturated conditions, Geophys. Res. Lett., 26, 3081–3084, 1999.

Jaeglé, L., Jacob, D. J., Brune, W. H., and Wennberg, P. O.: Chemistry of HO_x radicals in the upper troposphere, Atmos. Envir., 35, 469–489, 2001.

Jeker, D. P., Pfister, L., Thompson, A. M., Brunner, D., Boccippio, D. J., Pickering, K. E., Wernli, H., Kondo, Y., and Staehelin, J.: Measurements of nitrogen oxides at the tropopause: Attribution to convection and correlation with lightning, J. Geophys. Res., 105, 3679–3700, 2000.

Kraabol, A. G., Berntsen, T. K., Sundet, J. K., and Stordal, F.: Impacts of NO_x emissions from subsonic aircraft in a global three-dimensional chemistry transport model including plume processes, J. Geophys. Res., 107, 4655, doi:10.1029/2001JD001 019, 2002.

Kuhlmann, R. v., Lawrence, M. G., Crutzen, P. J., and Rasch, P. J.: A model for studies of tropospheric ozone and non-methane hydrocarbons: model evaluation of ozone related species, J. Geo-

phys. Res., 108, 4729, doi:10.1029/2002JD003 348, 2003.

Lamarque, J.-F., Brasseur, G. P., and Hess, P. G.: Three-dimensional study of the relative contributions of the different nitrogen sources in the troposphere, J. Geophys. Res., 101, 22 955–22 968, 1996.

Law, K. S., Plantevin, P. H., Thouret, V., et al.: Comparison between global chemistry transport model results and measurements of ozone and water vapor by airbus in-service aircraft (MOZAIC) data, J. Geophys. Res., 105, 1503–1525, 2000.

Lawrence, M. G. and Crutzen, P. J.: The impact of cloud particle gravitational settling on soluble trace gas distributions, Tellus, Ser. B, 50, 263–289, 1998.

Lelieveld, J. and Dentener, F.: What controls tropospheric ozone?, J. Geophys. Res., 105, 3531–3551, 2000.

Levy II, H., Moxim, W. J., Klonecki, A. A., and Kasibhatla, P. S.: Simulated tropospheric NO_x: Its evaluation, global distribution and individual source contributions, J. Geophys. Res., 104, 26 269–26 306, 1999.

Marenco, A., Thouret, V., Nedelec, P., et al.: Measurement of ozone and water vapor by Airbus in-service aircraft: The MOZAIC airborne program, An overview, J. Geophys. Res., 103, 25 631–25 642, 1998.

Meijer, E. W., van Velthoven, P. F. J., Thompson, A. M., Pfister, L., Schlager, H., Schulte, P., and Kelder, H.: Model calculations of the impact of NO_x from air traffic, lightning, and surface emissions, compared with measurements, J. Geophys. Res., 105, 3833–3850, 2000.

Meijer, E. W., van Velthoven, P. F. J., Brunner, D., and Kelder, H.: Improvement and evaluation of the parameterisation of nitrogen oxide production by lightning, J. Phys. Chem. Earth (C), 26, 577–583, 2001.

Meijer, E. W., van Velthoven, P. F. J., Segers, A., Bregman, B., and Brunner, D.: Improved mass fluxes in a global chemistry-transport model: implications for upper-tropospheric chemistry. Proc. of the European conference on Aviation, Atmosphere and Climate, Friedrichshafen, Germany, June 30 to July 3 2003, European Commission, Air pollution research report, 83, 128–133, 2004.

NASA: Atmospheric Effects of Aviation, A Review of NASA's Subsonic Assessment Project, National Academy Press, Washington, D.C., 1999.

Olson, J. R., Baum, B. A., Cahoon, D. R., and Crawford, J. H.: Frequency and distribution of forest, savanna, and crop fires over tropical regions during PEM-Tropics A, J. Geophys. Res., 104, 5865–5876, 1999.

Penner, J. E., Lister, D., Griggs, D., Docken, D., and MacFarland, M. (Eds): IPCC Special Report on Aviation and the Global Atmosphere, Cambridge University Press, New York, 1999.

Pickering, K., Wang, Y., Wei-Kuo, T., Price, C., and Müeller, J.-F.: Vertical distributions of lightning NO_x for use in regional and global chemical transport models, J. Geophys. Res., 103, 31 203–31 216, 1998.

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

Schultz, M. G., Jacob, D. J., Bradshaw, J. D., Sandholm, S. T., Dibb, J. E., Talbot, R. W., and Singh, H. B.: Chemical NO_x budget in the upper troposphere over the tropical South Pacific, J. Geophys. Res., 105, 6669–6679, 2000.

Schultz, M. G., Jacob, D. J., Wang, Y. H., et al.: On the origin of tropospheric ozone and NO_x over the tropical South Pacific, J. Geophys. Res., 104, 5829–5843, 1999.

Singh, H. B., Thompson, A. M., and Schlager, H.: The 1997 SONEX aircraft campaign and coordinated POLINAT 2 activity: Overview and accomplishments, Geophys. Res. Lett., 26, 3053–3056, 1999.

Staudt, A. C., Jacob, D. J., Logan, J. A., Bachiochi, D., Krishnamurti, T. N., and Poisson, N.: Global chemical model analysis of biomass burning and lightning influences over the South Pacific in austral spring, J. Geophys. Res., 107, 4200, doi:10.1029/2003JD004 284 2002.

Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, J. Geophys. Res., 106, 7183–7192, 2001.

Thompson, A. M., Sparling, L. C., Kondo, Y., Anderson, B. E., Gregory, G. L., and Sachse, G. W.: Perspectives on NO, NO_y, and fine aerosol sources and variability during SONEX, Geophys. Res. Lett., 26, 3073–3076, 1999.

Wang, Y. and Logan, J. A.: Anthropogenic forcing on tropospheric ozone and OH since preindustrial times, J. Geophys. Res., 103, 31 123–31 135, 1998.

Wang, Y., Logan, J. A., and Jacob, D. J.: Global simulation of tropospheric O_3-NO_x-hydrocarbon chemistry - 2. Model evaluation and global ozone budget, J. Geophys. Res., 103, 10 727–10 755, 1998.

Wei, C. F., Kotamarthi, V. R., Ogunsola, O. J., Horowitz, L. W., Walters, S., Wuebbles, D. J., Avery, M. A., Blake, D. R., Browell, E. V., and Sachse, G. W.: Seasonal variability of ozone mixing ratios and budgets in the tropical southern Pacific: A GCTM perspective, J. Geophys. Res., 108, 8235, doi:10.1029/2001JD000 772, 2003.

Wennberg, P. O., Hanisco, T. F., Jaeglé, L., Jacob, D. J., Hintsa, E. J., Lanzendorf, E. J., Anderson, J. G., Gao, R.-S., Keim, E. R., Donnelly, S. G., Del Negro, L. A., Fahey, D. W., McKeen, S. A., Salawitch, R. J., Webster, C. R., May, R. D., Herman, R. L., Proffitt, M. H., Margitan, J. J., Atlas, E. L., Schauffler, S. M., Flocke, F., McElroy, C. T., and Bui, T. P.: Hydrogen radicals, nitrogen radicals, and the production of O_3 in the upper troposphere, Science, 279, 49–53, 1998.