Available online at www.sciencedirect.com

ScienceDirect

Atmospheric Environment 41 (2007) 173–188

ATMOSPHERIC ENVIRONMENT

www.elsevier.com/locate/atmosenv

Evaluation and intercomparison of Ozone and PM10 simulations by several chemistry transport models over four European cities within the CityDelta project

R. Vautard[a,b], P.H.J. Builtjes[c], P. Thunis[d], C. Cuvelier[d,*], M. Bedogni[e], B. Bessagnet[f], C. Honoré[f], N. Moussiopoulos[g], G. Pirovano[h], M. Schaap[c], R. Stern[i], L. Tarrason[j], P. Wind[j]

[a]*Laboratoire de Météorologie Dynamique, IPSL, Ecole Polytechnique, 91128 Palaiseau, France*
[b]*LSCE/IPSL laboratoire CEA/CNRS/UVSQ, Gif sur Yvette, 91191 Cedex, France*
[c]*TNO Built Environment and Geosciences, Apeldoorn, The Netherlands*
[d]*European Commission-DG Joint Research Centre, Institute for Environment and Sustainability, I-21020 Ispra, Italy*
[e]*Mobility & Environment Agency, via Beccaria 19–20122 Milano, Italy*
[f]*INERIS, Parc Halata, Verneuil en Halatte, France*
[g]*Laboratory of Heat Transfer and Environmental Engineering, Aristotle University, 54 124 Thessaloniki, Greece*
[h]*CESI Ricerca S.p.A., via Rubattino 54–20134 Milano, Italy*
[i]*Free University Berlin, Inst. of Meteorology, Carl-Heinrich-Becker-Weg 6-10, D-12165 Berlin, Germany*
[j]*Norwegian Air Pollution Institute, Air Pollution Section/Research Department, P.O. Box 43, Blindern, NO-0313 Oslo, Norway*

Received 7 March 2006; received in revised form 9 July 2006; accepted 19 July 2006

Abstract

The CityDelta project Cuvelier et al. [2006. CityDelta: a model intercomparison study to explore the impact of emission reductions in European cities in 2010. Atmospheric Environment] is designed to evaluate the air quality response of several emission abatement scenarios for 2010 at the scale of the European continent, and specifically in the areas where most people live: the cities. Before evaluating this response, the model simulations in a control case must be evaluated against observations in order to understand their main strengths and weaknesses. In this article six different models are used to simulate a full year (1999) of air quality pollutant concentrations over domains encompassing a large area around four major European cities: Berlin, Milan, Paris and Prague. Three models are used both at large-scale (typically 50 km) and small-scale resolution (5 km). The intercomparison of the simulation results for ozone and particles smaller than 10 microns (PM10) leads to the following conclusions:

(i) Models capture fairly well the mean, daily maxima and variability of ozone concentrations, as well as the time and intercity variability. However, a significant overestimation of ozone in city centres is found especially for large-scale models.
(ii) PM10 simulation skill is generally poor, and large-scale models underestimate their mass. The difference between Milan (highly polluted) and the other cities is not reproduced. All models have difficulties in capturing the observed seasonal variations.

*Corresponding author. Tel.: +39 0332 785075; fax: +39 0332 786291.
E-mail address: kees.cuvelier@jrc.it (C. Cuvelier).

1352-2310/$ - see front matter
doi:10.1016/j.atmosenv.2006.07.039

(iii) The fine scale models show higher PM10 and lower ozone concentrations in urban areas, which are closer to the observations than are the large-scale models.

Keywords: Air quality; Ozone; PM; Model intercomparison; Emission–reduction scenarios

1. Introduction

Despite the important efforts in reducing pollutant emissions in Europe in the last decades, especially in road transports and in the industry, there are still only modest signs of improving concentration levels for some pollutants like ozone and fine particulate matter. The nonlinearity of atmospheric chemistry, together with the multiplicity of diffuse pollutant sources, including biogenic sources which can amplify concentrations, may partly explain this behaviour. The rising background levels, of ozone in particular (Simmonds et al., 2005), associated with intercontinental transport (Derwent et al., 2004) also have a potentially significant role, and are a major concern since they increase the part of pollution that cannot be regulated at the European scale.

Nevertheless the regional emission contribution to air quality in Europe will remain dominant within the next decade and countries must carry on their efforts in reducing primary pollutant emission in a cost-effective manner. The clean air for Europe (CAFE) programme of the European Commission aims at developing the technical knowledge and tools for a long-term strategy and integrated policy on emission regulation in order to prevent significant negative effects of air pollution on the human health and the environment. As ozone, particulate matter and NO_2 have adverse effects on health and ecosystems, the decrease of the levels for these pollutants is of particular emphasis in CAFE.

The only way to evaluate the impacts of air quality policies at the scale of Europe in detail is through numerical air quality modelling. During the last two decades or so several chemistry-transport models have been developed in Europe and elsewhere. Given hourly emissions, meteorological fields and chemical boundary conditions, these models describe the time evolution of several tens of chemical compounds and aerosols. Today the increase in computer power and memory make models able to perform long simulations of several months or a year in a reasonable time.

There is a wide spectrum of options when building a model: the horizontal and vertical resolutions, the chemical mechanism, the method of resolution, the aerosol description and physics, as well as many other physical and chemical processes. While all these options are generally based on validated and published individual processes, parameterizations and methods, their integration into a chemistry-transport model leads to possible accumulation or compensation of errors. These problems may spoil the simulation of air quality as well as the numerical evaluation of the impact of emission changes on air quality.

Modelling uncertainty can hardly be handled using one single model. That is why within CAFE it has been recognized that the spectrum of responses of several models to given emission scenarios has a clear added value relative to the response of a single model. The basic underlying hypothesis is that the spread of responses provides the uncertainty of these responses. The CityDelta project, whose overview is given in Cuvelier et al. (2006), aims at giving the range of model responses to realistic emission scenarios for 2010. It is an open project in which 16 models participated. Since long-range transboundary pollution problems in Europe are usually tackled with regional scale models such as the EMEP Eulerian model (http://www.emep.int), with resolution of the order of 50 km, another key objective of CityDelta was to determine the amplitude and sign of the simulation errors generated by the lack of resolution in and around cities, where concentration gradients are expected to be large. Several models were therefore used at different resolutions and simulated a long-term period, the Year 1999 over several European cities and their neighbourhood.

Before being used for the evaluation of the impact of emission scenarios, the participating models or model versions have to be evaluated against observations. The main purpose of this article is to assess the ability of the ensemble of models used in CityDelta to represent the actual present concentrations of pollutants. This validation stage is part of the CityDelta project but is not its main purpose. Several questions are addressed in this article:

(1) What is the skill of state-of-the art models in simulating air quality within European cities and in their vicinity?
(2) How is model skill dependent on the city and how large is its variability among models?
(3) How poorer is the skill of regional models, with a typical resolution of 50 km, as compared to city-scale models, with a typical resolution of 5 km.
(4) Do models have sufficient skill to be used for the evaluation of emission regulation policies at the scale of the continent and its main cities?

In order to answer these questions we perform a statistical analysis of the comparison between model simulations and observations over the validation year (1999) over four city areas: Berlin, Milan, Paris and Prague. Only results from models participating to the so-called second phase of the CityDelta project were considered, where the main focus was on PM, which was not considered in the first phase of City Delta. This choice is motivated by the more complete and homogeneous model results available. Some models participating to the first phase (Cuvelier et al., 2006) did not participate to the second phase because they could not be operated over very long periods or were in a too early development stage especially concerning the ability to calculate PM10 concentrations and its chemical composition. In this article, for the sake of concision we only evaluate ozone and particulate matter.

Model intercomparisons have already been carried out with various objectives. In the course of the EUROTRAC European programme, a regional-scale model intercomparison for ozone simulation has been carried out over the year 1995 and the responses to emission changes (Roemer et al., 2003). The evaluation of the ability of models to simulate inorganic aerosol compounds has also been evaluated in the framework of EUROTRAC (Hass et al., 2003). A group of regional models have also been evaluated over a season in the context of daily forecast of ozone (Tilmes et al., 2002). Recently, a model intercomparison has been performed in the framework of the Unified EMEP model (van Loon et al., 2004). However, an intercomparison over several cities with several models and resolutions, over one full year has never been performed. The completeness of the intercomparison study in CityDelta is therefore unique.

Section 2 describes the experimental setup: models, input data, observations and simulations. Section 3 gives the results obtained for ozone and Section 4 the results for particulate matter smaller than 10 μm (PM10). Section 5 contains a synthesis and a brief discussion.

2. Experimental setup

This section gives technical details about the numerical experiments that were performed in the second phase of CityDelta. Since this article focuses on the model validation, only the control simulation is considered, the analysis of the scenario results being left for another article (Thunis et al., 2006).

2.1. Cities and model horizontal grids

Four cities were retained in the second phase of CityDelta: Berlin, Milan, Paris and Prague. This choice was motivated by the large size of the cities, therefore subject to a significant pollution, the representativeness of various European climates, the availability of emission inventories, and also the experience of the modelling teams.

For each city a minimal model simulation domain of 300 km $\times$ 300 km around the city was imposed on modelling groups. These domains are among those represented in Fig. 1 of Cuvelier et al. (2006). For each city, modelling teams were asked to provide results on a 5 km $\times$ 5 km grid, whatever the model resolution actually was. The model resolutions differ from one case to another, but can be basically classified into two categories: large-scale (LS) models with a resolution of about 50 km and fine-scale (FS) models with resolutions in the order of 5 km. Some models are used at both resolutions. Table 1 summarizes the characteristics of each city urban area, the emission annual totals, the model simulations that were performed for that city and the actual model resolution used.

2.2. Emissions

The model emissions used in CityDelta are described in Cuvelier et al. (2006) to which the

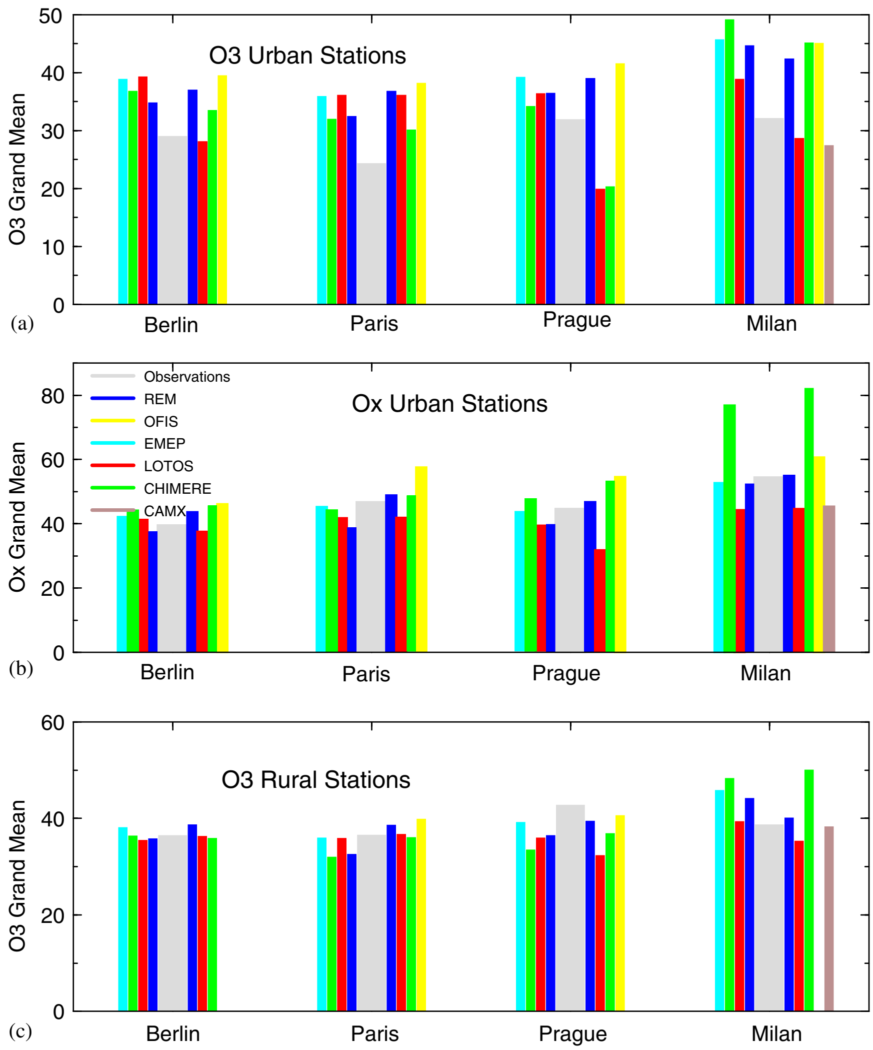

Fig. 1. Top panel (a): Observed (central gray bar) grand mean ozone concentrations for each city, together with the values for each model of CityDelta (coloured bars). Left (resp. right) to the observations are the results for large-scale (resp. small scale) models. The colour is kept the same for a given model and different resolutions. Concentrations are in ppb. Middle panel (b) shows the same as (a) for the O_x concentrations instead of ozone. Bottom panel (c) shows the same as top panel (a) for the selected rural stations.

reader is referred for further details. Here we briefly recall the main characteristics of the inventories. All models used the same emission inventories, adapted to their case. For model simulations at continental scale, emissions are based on EMEP/TNO (Vestreng, 2003). This inventory is provided by EMEP and contains annual totals for Year 2000 on a 50 km × 50 km grid covering Europe. Emissions for 2000 were used despite model simulations used 1999 as the meteorological year, because 1999 summer climate was more representative than summer 2000 (windy and cool, with only a few ozone episodes).

The EMEP inventory cannot be used as such in chemistry-transport models. Several operations have to be performed beforehand. The annual emissions have to be transformed into hourly emissions. That is done by applying, for each SNAP level 1 activity sector, seasonal, weakly and daily time profiles provided by the University of Stuttgart. Then emissions have to be distributed into model species. For NO_x this generally consists in assuming a fraction f_{NO_x} of emissions as NO and the rest as NO_2. LOTOS uses $f_{NO_x} = 0.95$, REM uses $f_{NO_x} = 0.95$, In CHIMERE, $f_{NO_x} = 0.90$, the remaining 10% are distributed as 9.2% of NO_2 and

Table 1
Some characteristics for each city population (source: www.demographia.com), annual total of NO_x (in thousands of tons of equivalent NO_2), non methanic volatile organic compound (NMVOC), and effective model resolutions used, in km or degree

City	Berlin	Milan	Paris	Prague
Population of area (M)	3.9	3.7	9.6	1.2
Inventory resolution (km)	2	5	3	5
NO_x annual	301	453	377	341
NMVOC annual	298	656	477	263
PM10 annual	67	92	96	68
CHIMERE-LS	$\frac{1}{2}$degree	Same	Same	Same
EMEP	50 km	Same	Same	Same
LOTOS-LS	0.25×0.5 latlong	Same	Same	Same
REM3-LS	0.25×0.5 latlong	Same	Same	Same
CHIMERE-FS (km)	4	5	6	5
CAMx (km)	—	5	—	—
LOTOS-CITY	$5 \times 5\,km^2$	Same	Same	Same
REM3-CITY (km)	4	4	4	4
# of O_3urban sites	13	7	10	0
# of O_3 suburban sites	0	3	3	3
# of O_3rural sites	2	3	1	1
# of PM10 sites	3	4	4	8

Note: These are the numbers for all stations we received data for (excepted street canyon ones).

0.8% of HONO. EMEP uses $f_{NO_x} = 0.95$, for CAMx $f_{NO_x} = 0.96$ and for OFIS $f_{NO_x} = 0.90$.

The non-methane volatile organic compound (NMVOC) emissions need to be adapted to the different chemical schemes used by the models. Models using CBIV-based mechanisms directly apply a speciation matrix allocating the specific VOC-species based on their carbon bonds to the CB-species. Other models using the chemical schemes EMEP (Hov et al., 1985), MELCHIOR (Lattuati, 1997), generally proceed in two steps: (i) Total NMVOC mass is speciated into a certain number of real individual compounds. This was done by using the Passant (2002) speciation, which depends on SNAP level 1 activity categories, and (ii) real NMVOCs are distributed in model NMVOC lumped classes using the Middleton et al. (1990) procedure. Finally the emissions are projected onto the model grid.

For city-scale emissions, high resolution inventories were prepared by local participants. Milan emissions, based on the Regione Lombardia inventory, gridded by TERRARIA cover the whole Lombardy region with a resolution of 5 km. PM2.5 and PM10 estimates are based on three different sets of emission factors sources (TNO-CEPMEIP, IIASA, Lombardy region). Paris emissions are described in Vautard et al. (2003) and were prepared by the AIRPARIF air quality monitoring organization. Initially PM10 and PM2.5 emissions were not available in the inventory. As a first guess both species emissions were estimated by using the SNAP level1-detailed EMEP/TNO PM10/NO_x emission ratios averaged over the emission area, and applying these ratios to city-scale NO_x emissions in order to have spatially and temporally distributed emissions for PM10 and PM2.5. The Berlin emission inventory covers Berlin, Brandenburg, Sachsen and Sachsen-Anhalt with a spatial resolution of approximately 2 km whereas the Prague inventory was constructed from the data provided over the entire Czech Republic with a spatial resolution of 5 km. For Berlin, simple scaling factors were used between PM10 and PM2.5 or TSP depending on the activity sector while, for Prague, PM estimates have been scaled to EMEP values and spatially allocated according to NO_x. For some cities the inventory did not cover the whole $300\,km \times 300\,km$ required simulation area. In such cases EMEP emissions were used as a complement. Double counting of the emissions was avoided by subtracting city emissions from EMEP cell totals.

For some species and some cities, city-scale and regional-scale were not consistent (Maffeis et al., 2006, manuscript in preparation), especially for NMVOC and particulate matter, due to differences in the building methodologies. In this second phase of the CityDelta project, this problem was circumvented by scaling the local inventories to their EMEP counterparts (see Cuvelier et al., 2006 for more details). The scaled emission annual totals are given for each city in Table 1.

2.3. Observations

Air quality observations come from surface monitoring automatic stations operated by local air quality organizations (Berlin: German Federal Environment Agency; Milan: Regione Lombardia; Paris: AIRPARIF; Prague: Czech Hydrometeorological Institute). Monitoring sites used in this model validation are classified as background urban, suburban or rural sites. Traffic monitoring stations are not considered here due to their poor representativeness at the models resolution. Monitoring data are available for the whole year 1999 on an hourly basis for O_3, NO, NO_2 and PM_{10}.

For ozone simulation evaluation, we perform all statistical analysis over two types of sites: rural sites and urban (urban background or suburban types) sites. All available rural monitoring sites where ozone is measured are used, while we select urban sites to lie within the central area of the considered city, and not in the suburbs. The number of sites of each set for each city is given in Table 1. Averages and other statistics are calculated separately over each set of sites in order to distinguish model skill close to emissions and in the city plume. This distinction was not possible for PM10 due to the lack of rural monitoring stations. Only urban sites are considered. For each city and for each group of stations, skill statistics are calculated over all stations data taken together. As there is a large spread in the number of stations in each group, statistics themselves have uncertainties varying from one city to another.

2.4. Models and input data

There are six participating models in the second phase of CityDelta, and on total nine different model configurations. These models are CHIMERE (Schmidt et al., 2001; Bessagnet et al., 2004), EMEP (http://www.emep.int), LOTOS (Schaap et al., 2004 and references therein), REM-CALGRID (Stern et al., 2003), OFIS (Moussiopoulos and Douros, 2005) and CAMx (ENVIRON, 2004). CHIMERE, LOTOS and REM were used at the two spatial scales, while EMEP was used only at regional scale and both CAMx and OFIS only at city scale. Note that the CAMx simulations only covered the city of Milan. All participating models simulate both gas phase and aerosol phase chemistry and physical processes with state-of-the art parameterizations at the time of the project (most runs were carried out in 2004. Both the number of gas-phase species and the description of aerosols are of the same order of magnitude, with a few tens of gas species, a few aerosol modes or sections and 5–7 aerosol chemical compounds (see Cuvelier et al., 2006). However,- both the chemical mechanisms and aerosol processes differ from one model to another, and the reader is referred to the model description articles for details.

Vertical resolutions vary from 4 to 6 layers for models describing the lower troposphere (CHIMERE, LOTOS, REM3) to 20 layers for EMEP. This leads to an equivalent boundary-layer resolution in all cases.

There are significant differences in meteorological forcing: unlike in subsequent studies, in CityDelta CHIMERE uses the ECMWF 3-hourly short-term (3–6 h) forecasts with time and space interpolations, as explained in Schmidt et al. (2001). EMEP uses the Norwegian meteorological analyses. LOTOS and REM use the diagnostic meteorological fields provided by the Free University of Berlin based on analysis and optimal interpolation of the ECMWF observations.

It should be noted that most models used, for the city calculations, only interpolated meteorological data from the LS information (REM used meteorological data based on available local land use and topographical information). This severely hampers the possible increase in performance of the FS versus the LS-models. Meteorlogical forcing for CAMx is an analysis of meteorological variables was carried out using the ALADIN fields as background together with about 40 ground level stations and 1 radiosounding, using the CALMET interpolator.

Boundary conditions of the LS models are given by observed Logan climatological data base for ozone for EMEP. LOTOS and REM use the Logan climatologies. CHIMERE uses the monthly climatologies of the global-scale MOZART model (Horowitz et al., 2003).

At city scale, CHIMERE, LOTOS and REM are using their own regional-scale concentrations at the boundaries while CAMx and OFIS are forced at the boundaries by gas phase and aerosol species of the EMEP model. In each case a one-way nesting (no feedback from fine-scale to LS models) procedure is used.

2.5. Simulations and procedure

Model simulations are carried out over the full 1999 year. For ozone only the summer season (April–September) and for PM10 the whole year are considered in the statistical analysis of skill. The enormous amount of data produced by the modelling groups was first processed before being sent to Ispra Joint Research Centre. Whatever the model resolution, concentrations fields were all interpolated to a $5\,\mathrm{km} \times 5\,\mathrm{km}$ grid for each city. Then concentrations in a grid were simply compared to the station lying in that grid. In order for simulated values to be comparable to observations the LOTOS model assumes a concentration profile near the surface based on the constant flux-approach/deposition which enables to calculate concentrations at a specific height above the surface (3.9 m in this case, the height of the monitoring stations) while other models directly provide the value as averaged over the lowest layer.

3. Simulation of ozone

3.1. Model biases

As shown in Fig. 1a, models generally capture the right order of magnitude of the ozone grand mean for each city in the city centres, but most of them have a positive bias. The bias is observed along the diurnal cycle for most models (not shown). Only for Prague, LOTOS and CHIMERE FS versions exhibit a significant negative bias. Despite the general positive ozone overestimations models exhibit small bias for the $O_x = O_3 + NO_2$ mean concentrations. Nitrogen dioxide concentrations are therefore generally underestimated in the city centres, and the two biases generally compensate each other, which is a strong indication that ozone titration is insufficient in models.

This behaviour could result from a systematic overestimation of model turbulent mixing near the ground, or a too thick lowest layer and a too coarse resolution near the sources, leading to too much diluted NO_x concentrations. However, models use different meteorological drivers for dispersion in the boundary layer and ozone bias occurs all day long (not shown).

For Milan city, CHIMERE has a markedly different behaviour with a significant overestimation of ozone and O_x. Photochemical ozone production is exaggerated. One explanation is that LS ECMWF analyses and short-term forecasts have too weak winds and therefore dispersion in the Po valley, as also shown by Minguzzi et al. (2005). The meteorological origin of this problem is corroborated by its disappearance when CHIMERE was later forced by another model as the MM5 mesoscale model (experiments performed later than the exercise).

Note that model biases are larger than intercity means differences, even for O_x. Models therefore are not accurate enough to simulate the small intercity differences in city centres for mean concentrations. Finally ozone overestimations are absent when rural stations are considered, confirming the sensitivity of model errors to titration in city centres only.

LS models have a too coarse horizontal resolution to represent the city-scale high nitrogen oxides. As shown in Fig. 2 ozone LS model biases are all positive and larger than the biases of their fine scale versions, except for the REM model which may result from the different meteorology used in the FS case. For O_x there is no such difference between FS and LS models showing that the lower ozone produced by FS models is specifically due to ozone titration, and not to the lack of ozone production due to the presence of high NO_x. Thus the increase

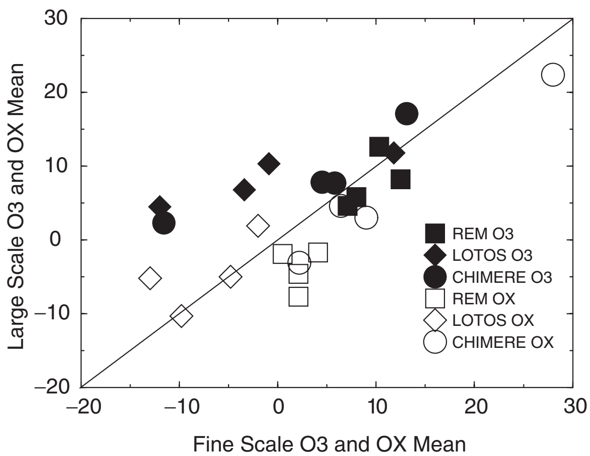

Fig. 2. Grand mean ozone and O_x biases (in ppb) of the large-scale model versions versus biases of their fine-scale versions, for the three models with simulations at the two resolutions: CHIMERE, LOTOS and REM. Each point of the graph represents stands for a city and a model.

of resolution generally helps improving the representation of ozone titration, the distribution of O_x between ozone and NO_2 being biased toward ozone in the LS case.

3.2. Daily average variability

The ability of models to simulate the day-to-day ozone variability is examined using Taylor diagrams (Taylor, 2001). In these diagrams time correlation **C** between observations and simulations is represented together with the standard deviation σ_{mod} of simulations in polar coordinates. The angle from horizontal axis is the arccosine of the correlation and the radius is the standard deviation. The exact plot represents $y = \sigma_{\mathrm{mod}}\sqrt{1 - C^2}$ as a function of $x = \sigma_{\mathrm{mod}}C$. The nice property of this graph type is that the standard deviation of the error, σ_{err}, is the distance between the point (x, y) and the point $(\sigma_{\mathrm{obs}}, 0)$, σ_{obs} being the standard deviation of the observations. Therefore, in the absence of model bias Taylor diagrams simultaneously show root mean square error (RMSE), correlation and standard deviations.

In Fig. 3 Taylor plots of the daily ozone averages in city centres are shown, all stations being considered together for the calculation of statistics. Correlations range between 0.3 and 0.8. There is no particular superiority of any model for all cities in terms of correlations. In Berlin LOTOS (LS),

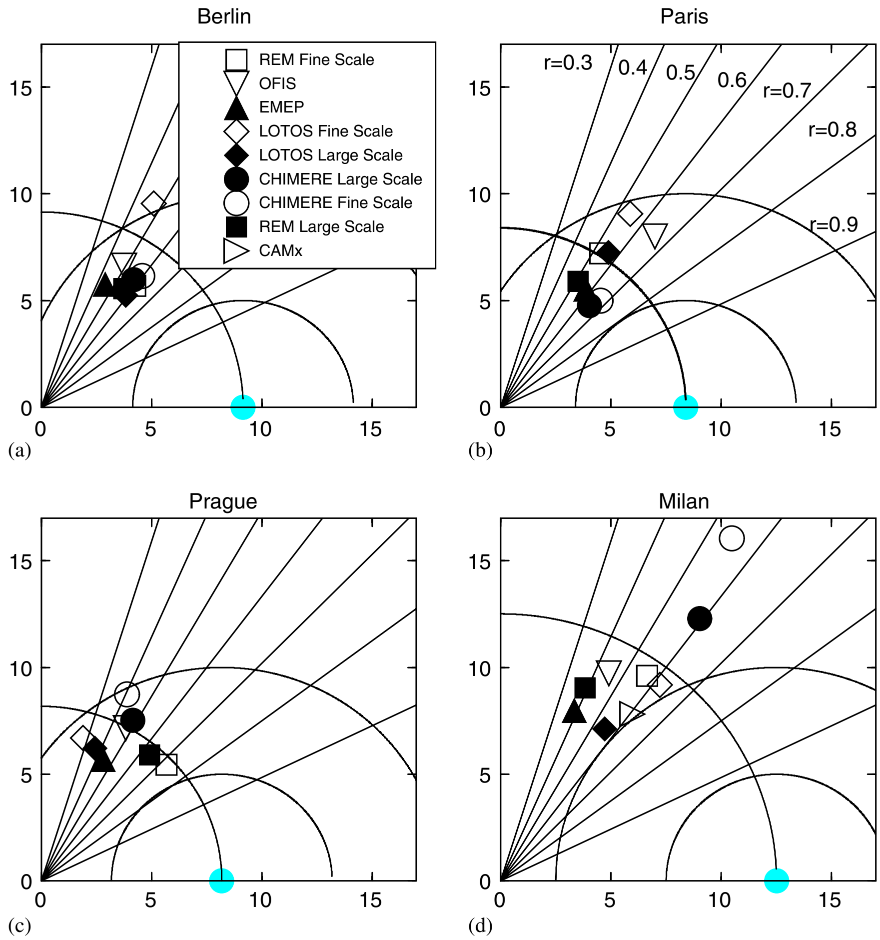

Fig. 3. Taylor diagrams of the daily ozone averages for all cities. Each graph stands for a city, and each point in a graph represents a model variability skill measure. Values along the axes are in ppb. The plot represents, for each model, the standard deviation of simulated values (radius) and the time correlation between simulated and observed values (cosine of the angle to the horizontal axis). For each city the standard deviation of the observations is shown as the point on the horizontal axis. According to the Taylor plot property circles centred on this point represent points of equal simulation error standard deviation (or RMSE in the absence of bias). As shown by Fig. 3 error standard deviations are smallest for models which exhibit highest correlations, and range between 5 and 10 ppb.

CHIMERE and REM correlations are leading, with very similar values near 0.6. CHIMERE has highest correlations in Paris together with OFIS (about 0.65–0.7). In Prague REM clearly leads (0.7). In Milan LOTOS correlations are highest but other models compete with similar correlations around 0.6: CAMx, REM and CHIMERE.

LS simulated standard deviations are usually smaller than their observed counterparts, showing that models generally underestimate the daily average variability. This deficiency results at least partly from lack of resolution since it is much less pronounced in fine scale models. Variability underprediction occurs along the day and night (not shown). CHIMERE variability in Milan is much higher than observed. Again, the excessive stagnation of ECMWF winds in the Po valley can explain the exaggerated contrast between overestimated concentrations in sunny episode days and cloudy days, enhancing the model variability.

Although variability is improved in fine-scale models, the disappointing result here is that these latter do not exhibit systematic higher correlations than their LS counterparts for CHIMERE, LOTOS and REM. As shown in the Taylor diagrams, a spatial resolution increase enhances the variability of the concentrations, but also the amplitude of the errors due to meteorology, vertical resolution, land use or other factors.

For all cities but Milan, the correlation coefficient obtained for O_x is very similar to, and in most cases slightly greater than, that of ozone (Fig. 4). In northern cities ventilation makes the daily average variability of O_x essentially driven by the daily average variability of ozone advected into the city. Thus, the skill of models to predict O_x, and to a large extent ozone, is largely dominated by the skill of simulated boundary conditions. This is consistent with a previous analysis of forecast skill over Paris (Vautard et al., 2001). The slight superiority of O_x correlation to ozone correlation is an indication of the difficulty of models to correctly simulate local titration.

For Milan the O_x correlation is always poorer than the ozone correlation for all models but CAMx. Due to the Alps dynamical barrier, the city is less influenced by regional-scale advection, and undergoes frequent stagnation. Therefore, O_x variability is due to local meteorological fluctuations and the skill strongly relies on the model ability to simulate complex-terrain flows. The bad skill of O_x daily averages is hard to explain, and may come from a difficulty in simulating NO_2 and other primary species.

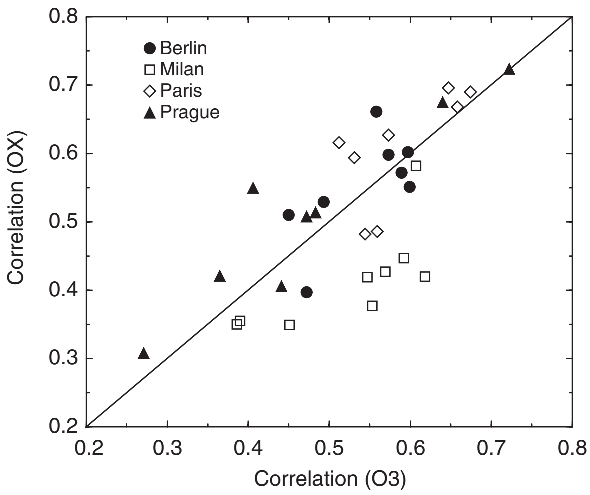

Fig. 4. Model-observation daily average time correlations for O_x vs. correlations for ozone. Each symbol represents a different city but all models use the same symbol for each city.

For rural stations the Taylor diagrams (not shown) display relatively similar structures, but the differences between large and fine scale models becomes smaller. Except for a few models in Milan, the simulated variability is weaker than the observed variability.

3.3. Daily maxima

As shown by Fig. 5, ozone daily maxima simulations in city centres are also overestimated, but less than ozone daily averages for Berlin, Paris and Prague. The first reason is that models have a slightly lower bias during daytime than during nighttime. Another reason may be that observed hourly concentrations result from the sum of a signal representative of the 5 km × 5 km resolution, plus a non representative noise mostly due to subgridscale heterogeneities. Observed daily maxima are maxima of the sum of the two signals, while simulated maxima only stand for the first representative signal. Therefore there is an inherent methodological bias, due to finite resolution, when comparing simulated and observed daily maxima: finite-resolution unbiased models should in principle underestimate observed individual-station daily maxima. Even though this effect remains to be quantified, the reduction of bias found here could therefore result from error compensation. For Milan most models overestimate the daily maxima.

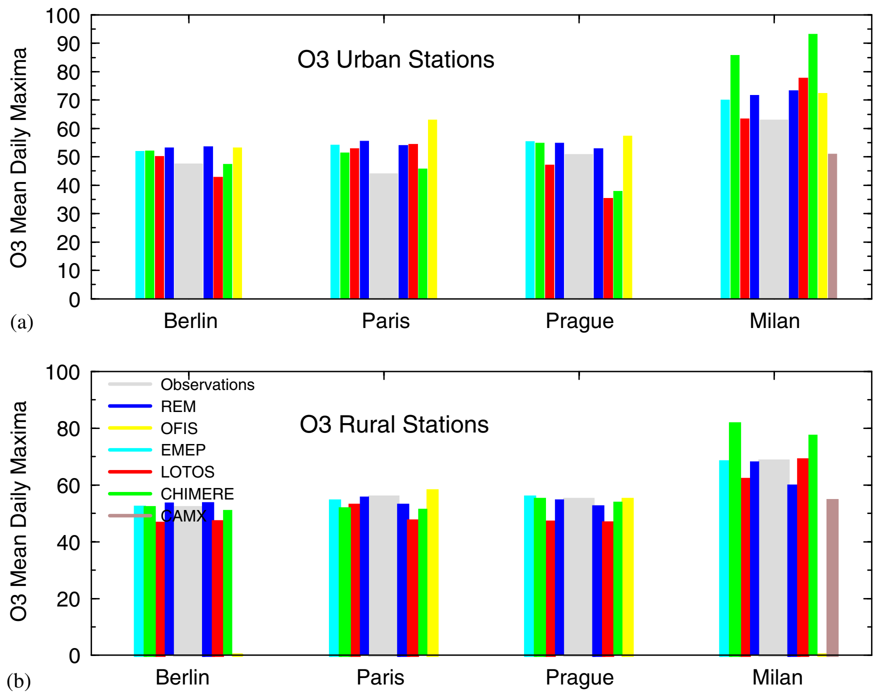

Fig. 5. Same as Fig. 1a for the ozone daily maxima instead of daily averages, for city centre stations (top panel) and rural stations (bottom panel).

As for daily averages CHIMERE has a large positive bias, and CAMx a negative bias. The concentration difference between Milan and the other cities is reproduced by the models. In rural areas (Fig. 5, bottom panel), biases are less pronounced, and the differences between LS and FS models are small.

Taylor diagrams for the daily maxima are shown in Fig. 6. Correlations between observations and simulations are generally higher than for daily averages, meaning that models generally better capture the daytime variability of ozone than its nighttime one. Correlations rise to about 0.75–0.85 for CHIMERE in Berlin, for OFIS, CHIMERE and EMEP in Paris, for REM in Prague. LOTOS and CAMx have the highest correlation in Milan. The correlation is not systematically higher in FS models than in LS models. For Prague the LS correlation is even systematically higher than the FS one.

Standard deviations are generally underestimated in all cities, as for daily averages. However, they are not systematically higher for FS models than for LS models unlike for daily averages. This is an indication of the smaller effect of downscaling for ozone during daytime than during nighttime, due to a larger nighttime titration.

The ability of models to simulate ozone episodes is evaluated by the comparison between simulated and observed 95th percentile of ozone daily maxima (Fig. 7). In general, models faithfully reproduce the 95th percentile, for Berlin, Paris and Prague. In Berlin the largest difference is the OFIS overestimation (about 20 ppb) and in Prague the LOTOS underestimation (about 20 ppb also). In Milan, CHIMERE again significantly overestimates the percentile and CAMx underestimates it. There is a large degree of variability between models, especially at fine scale, showing that the ability of models to simulate acute episodes is strongly variable in the Po valley. The differ ence between Milan and the other cities is well reproduced.

We do not observe any systematic improvement from LS to FS models in the 95th percentile. By contrast, the fifth ozone percentile (Fig. 7, bottom) is systematically overestimated by LS models, while FS models better simulate these values.

In rural areas (not shown) the 95th percentile is also generally well reproduced, with slight underestimation for most models. Ozone largest maxima generally correspond to sharp ozone pollution plume developing downwind of the

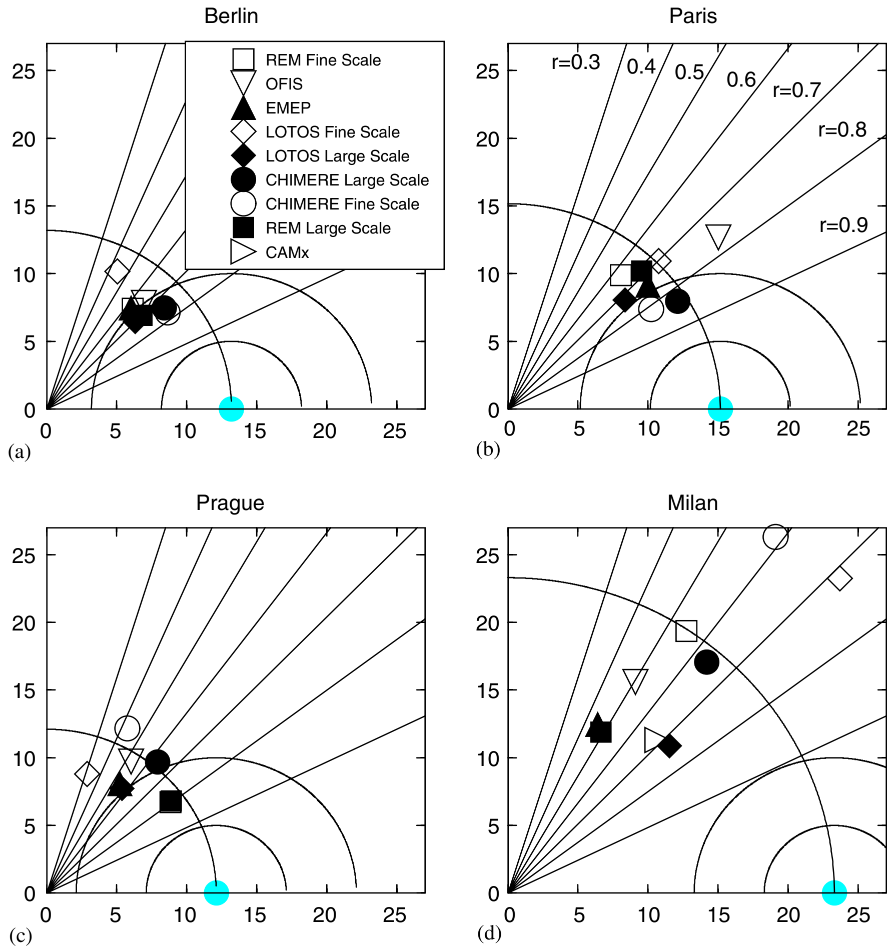

Fig. 6. Taylor diagrams as in Fig. 3 but for ozone daily maxima.

city. The expected improvement due to higher resolution in such cases is not found in the results. The fifth percentiles are mostly unbiased for all models, unlike in the city centres. This confirms the exclusive role of titration in the city centre biases, NO_x inhibition being much lower in rural areas.

3.4. The SOMO35 indicator

The statistical link between acute ozone episodes and short-term premature mortality was established in several studies (Bell et al., 2004; Gryparis et al., 2004). One critical parameter relating ozone to mortality is the daily maximal 8 h-average concentration. As no effect could significantly be detected for values of this average below 35 ppb IIASA decided to use the integrated daily excess of ozone 8 h-average to 35 ppb as a health indicator, called SOMO35 (Amann et al., 2005). Here SOMO35 is calculated over the six summer months only and is expressed in ppb.day.

As for other ozone statistics, LS models overestimate the SOMO35 indicator in city centres (Fig. 8). The overestimation factor can be quite as large as 2–3, such as for Paris. The overestimation is less systematic in FS models where titration can be better taken into account. REM exhibits a systematic overestimation with a minor improvement for the FS version. For Milan CHIMERE exhibits a strong overestimation by a factor of 2. The error in the SOMO35 behaves as that in the ozone grand mean, with amplification of the relative overestimation as only excesses to 35 ppb are considered. In rural areas of northern cities both LS and FS models but LOTOS simulate the SOMO35 to within 20%. This latter model systematically underestimates SOMO35. For Milan CHIMERE exhibits a large overestimation while the other models behave successfully.

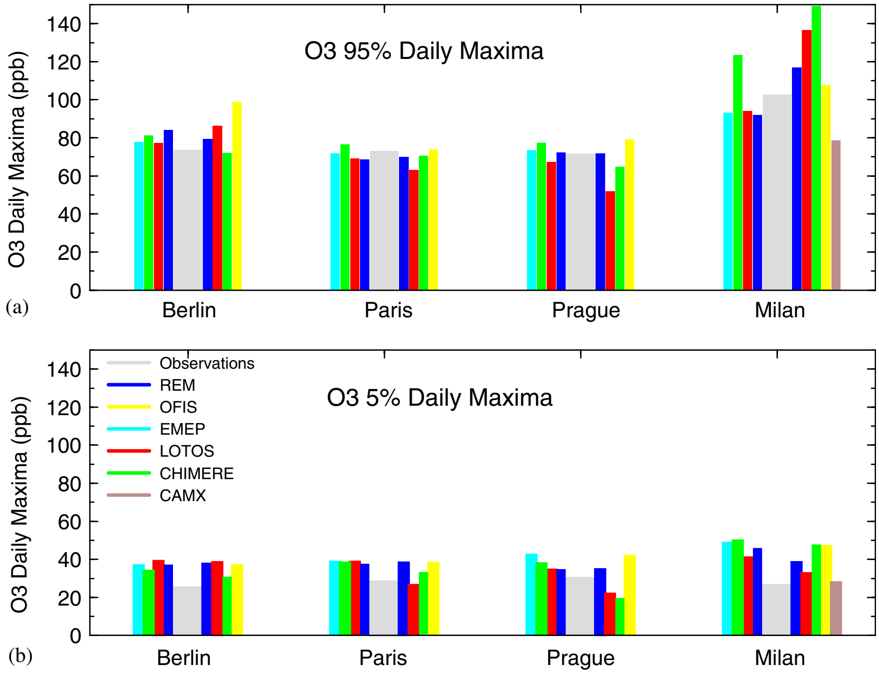

Fig. 7. 95th percentile of the ozone daily maxima in the city centres (top panel) and 5th percentile of the same quantity (bottom panel). Colors are as in Fig. 1. Units are ppb.

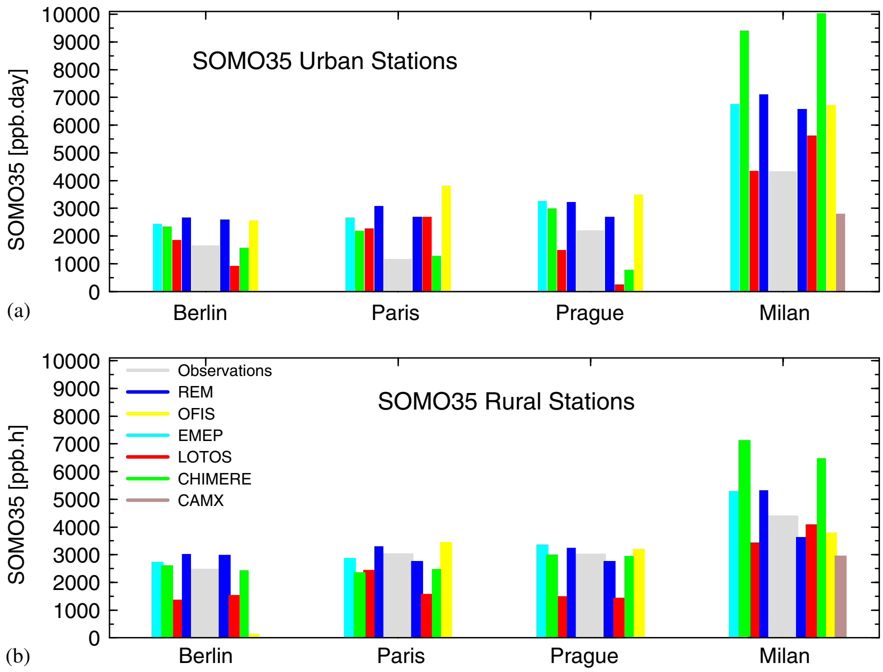

Fig. 8. Comparison of the simulated and observed SOMO35 indicator (in ppb.day) values averaged over the city-centre stations (top panel) and the rural stations (bottom panel). Colours are as in Fig. 1.

4. Simulation of PM10

For PM10 we only consider daily averages as health criteria are based on average quantities. As PM10 measurements are often missing in rural areas we only consider here city-centre stations. Statistics are now carried out over the whole 1999 year. In the analysis of PM10, some model outputs are missing, such as for LOTOS in Paris and Prague. Not all aerosol compounds are included in PM10. In general models include primary and inorganic particulate matter, but not all models have organic compounds. In the CityDelta project models do not include mineral dust.

Except for CHIMERE in Milan, LS models underestimate the city-centre concentrations of PM10 by factors sometimes quite large (Fig. 9). FS models exhibit concentrations 20–100% larger, because the higher resolution allows a better account of larger concentrations near the sources.

In Northern cities PM10 mass concentration lies in the 20–30 $\mu g\,m^{-3}$ while for Milan it is twice larger. This intercity variability is poorly represented by the models: Except for CHIMERE for which the contrast between Milan and the other cities is overestimated, the other models show small intercity variability. In particular the Milan concentrations are too small.

The difference between the behaviour of CHIMERE and other models is difficult to explain without a careful examination of simulated meteorology, which is beyond the scope of this article. Nevertheless, the overstagnation of air masses in ECMWF analyses in the Po valley, as mentioned earlier, may explain the very large values obtained for CHIMERE.

The skill of models in simulating the day-to-day variability of PM10 is examined using Taylor plots (Fig. 10). Correlations are generally weaker than for ozone daily averages or maxima (Figs. 3 and 6). Largest values are obtained for Paris and Berlin, cities having least stagnant weather conditions. Models have difficulties in correctly simulating the PM10 variability in Milan, as correlations do not exceed 0.5.

Quite surprisingly FS model versions have systematically poorer correlation and larger standard deviation than their LS counterpart. The higher variability in FS models is due to better resolution of local gradients, which also leads to errors of larger amplitude and a decrease in correlation.

For REM the variance is generally underestimated, with an improvement in the FS cases. For CHIMERE the LS variability is closer to reality than the FS variability, which is generally too high. In this model version the processing of ECMWF analyses, which do not account for urbanization effects in the surface forcing produced at times very thin boundary layers in the winter leading to very high PM10 values. Similar problems were also found later with MM5 forcing and were tentatively attributed to the lack of anthropogenic heat fluxes leading to overestimated atmospheric stability (Hodzic et al., 2005). Hence the fair standard deviations for the LS version are due to error compensation between low resolution, which decreases variability and underestimated dispersion which increases variability. This problem is particularly acute in Milan where PM10 concentrations are unrealistically high during some winter days. Other models behave in an opposite manner, with a fairly underestimated variability in this city.

The seasonality of PM10 concentrations is not well reproduced (not shown). In city centres models exhibit higher average concentrations in winter than in summer, while this seasonal difference is only found for Milan in simulations. Again, this could result for the unrealistic wintertime stability due to the lack of anthropogenic heat fluxes in meteorological models.

5. Conclusion and discussion

The skill of several models for the simulation of the concentrations of ozone and particulate matter

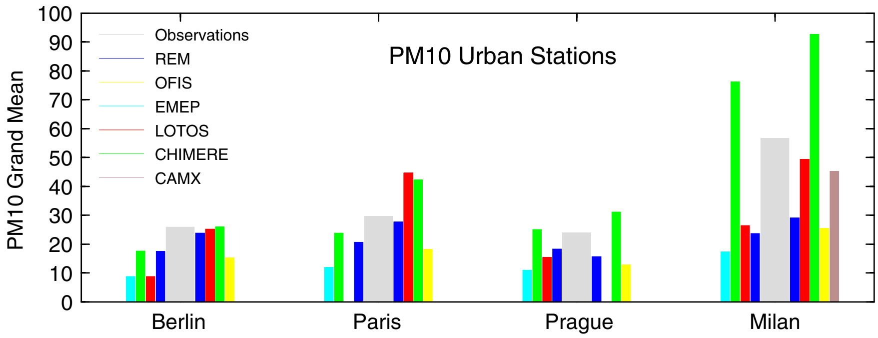

Fig. 9. Yearly average of PM10 as observed and simulated by all models, in $\mu g\,m^{-3}$. Colours and plot settings are as in Fig. 1.

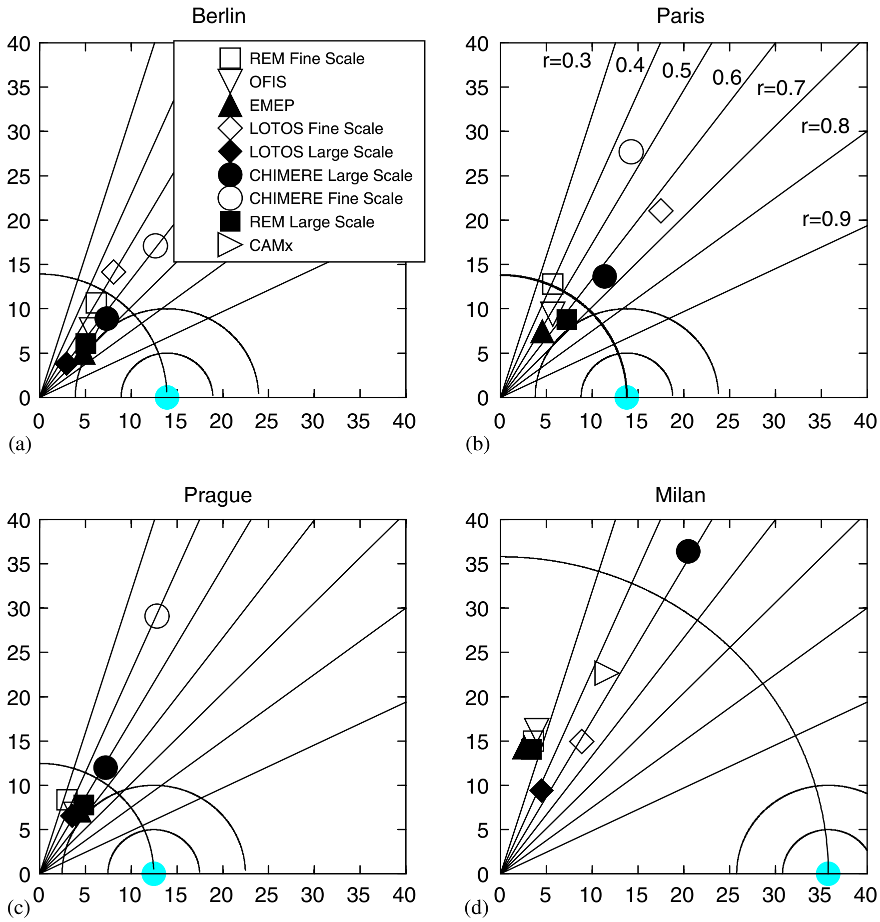

Fig. 10. Taylor plots of PM10 daily averages for the four cities under study. Symbols as in Figs. 3 and 6.

over 4 European cities has been evaluated and intercompared, within the framework of the City-Delta project, as a part of the DG Environment CAFE programme.

For ozone in the three northern cities, it is quite clear that all models have reasonable skill in the simulation of concentrations in both city centres and rural areas located nearby. They also simulate with success the ozone daily maximal concentration, 95% percentile and SOMO35 differences between Milan and the Northern-Europe cities. The ozone in rural areas is simulated with better skill than in urban areas where rapid ozone inhibition due to NO_x makes it very sensitive to the representation of mixing near sources and to errors in the meteorological parameters such as boundary layer height or wind. Most models overestimate ozone mean concentrations in city centres while no specific bias is observed for $O_x = O_3 + NO_2$ concentrations. This difference between O_x results and ozone results shows that ozone titration, insufficiently strong in LS models, having resolutions of about 50 km. LOTOS and CHIMERE fine scale versions better capture this titration effect than their LS counterparts.

The bias is also present in ozone daily maxima, to a smaller extent, but part of this improvement may be due to error compensation between representativeness of peaks calculated over individual stations and model bias.

The variability of both simulated daily averages and maxima is generally underestimated but models capture well the relative variations, as the time correlations can be as high as 0.8. The ability of models to simulate extremes is evaluated by comparing predicted and observed fifth and 95th percentiles of daily maxima. The simulation of high extremes is generally unbiased but the lowest values are overestimated in city centres. The general overestimation of daily averages and maxima is therefore due to the difficulty of models to simulate the lowest ozone values, occurring during cloudy or rainy days when lack of mixing leads to high NO_x concentrations. Finally the simulation of the health-effect indicator SOMO35 has the same strengths and weaknesses as the average ozone but the 35 ppb threshold effect amplifies the deficiencies.

The simulation of PM10 concentrations by the models is poor. LS models largely underestimate the fine particles total mass, a bias that is reduced in the FS models, which shows the importance of resolution in correctly simulating particle levels. The strong PM10 mean concentration differences between Milan and the Northern-Europe cities is not reproduced. The skill in the simulation of variability is poorer than for ozone, with correlations of 0.7 at best.

A clear overall conclusion from this study is that the models used in the CityDelta project are skilful for the simulation of ozone but fail to provide an accurate prediction of PM10 levels. Whether this failure is due to uncertainty in our knowledge of the sources, dynamical or chemical processes is not yet known. PM10 concentrations are more sensitive to urban boundary layer dynamical processes than ozone concentrations. Our knowledge in emission processes is yet very lacunar, especially concerning suspension and resuspension of deposited particles.

Another conclusion of this study is that fine-scale models show better performance for PM10 in the cities, and treat the titration effect better than LS models. However, these improvements are limited because models do not generally use small-scale meteorology, and still have limited vertical resolution. A further increase in horizontal resolution, down to 1 km $\times$ 1 km instead of 5 km $\times$ 5 km may be necessary to increase further the skill.

No given model is superior to the others. The spread of simulated mean values, variabilities is relatively large, presumably encompassing the range of response to the range of uncertainties in input model parameters, chemistry, physics or numerics. It is hypothesized here that this ensemble of models gives a representation of the uncertainty in our knowledge. In a future study the statistical properties of this ensemble approach will be quantified. However, we cannot neglect the effect of other numerical processes such as pre- or post-processing of models simulations, performed independently by each modelling group that probably yields additional irrelevant uncertainty.

Going back to the main objectives of CityDelta and this intercomparison exercise (Cuvelier et al., 2006), there are good reasons to believe that this ensemble of models is able to simulate the ozone response to emission scenarios, as both the time variability for each city and the inter-city variability are correctly reproduced for ozone. However, one has to keep in mind that LS models overestimate ozone in the city centres due to underestimated titration, and that most small-scale models also suffer from this deficiency. Depending on the chemical regime the effect of this bias on a possible bias in the concentration differences (the "deltas") is hardly predictable. For PM10, if one assumes that model biases are due to meteorology or resolution, the resulting deltas should be linked to the bias of the concentrations themselves. Since most models underestimate the concentrations, scenarios should also give an underestimated response.

As fine-scale models do not clearly exhibit superior skill than LS models, one should not expect a strong improvement in the accuracy of deltas simulated by the small-scale models. This disappointing result should however be moderated by the fact that, at the time simulations were achieved, most of the models used were still in a stage of development. The emission inventories also were quite young and many progresses have been achieved since then and will be in the near future. Thus, there are hopes that better resolved models might show increased performance models in the future.

References

Amann, M., Bertok, I., Cofala, J., Gyarfas, F., Heyes, C., Klimont, Z., Schöpp, W., Winiwater, W., 2005. Clean Air For Europe (CAFE) Programme final report, Laxenburg, Austria.

Bell, M.L., McDermott, A., Zeger, S.L., Samet, J.M., Dominici, F., 2004. Ozone and short-term mortality in 95 US urban communities, 1987–2000. Journal of the American Medical Association 292, 2372–2378.

Bessagnet, B., Hodzic, A., Vautard, R., Beekmann, M., Cheinet, S., Honoré, C., Liousse, C., Rouil, L., 2004. Aerosol modelling with CHIMERE: preliminary evaluation at the continental scale. Atmospheric Environment 38, 2803–2817.

Cuvelier, C., Thunis, P., Vautard, R., Amann, M., Bessagnet, B., Bedogni, M., Berkowicz, R., Brocheton, F., Builtjes, P., Denby, B., Douros, G., Graf, A., Honoré, C., Jonson, J., Kerschbaumer, A., de Leeuw, F., Moussiopoulos, N., Philippe, C., Pirovano, G., Rouil, L., Schaap, M., Stern, R., Tarrason, L., Vignati, E., Volta, L., White, L., Wind, P., Zuber, A., 2006. CityDelta: a model intercomparison study to explore the impact of emission reductions in European cities in 2010. Atmospheric Environment, submitted for publication.

Derwent, R.G., Stevenson, D.S., Collins, W.J., Johnoson, C.E., 2004. Intercontinental transport and the origins of the ozone observed at surface sites in Europe. Atmospheric Environment 38, 1891–1901.

ENVIRON International Corporation, 2004. User's guide to the Comprehensive Air Quality Model with Extensions Ver 4.0. Technical Report, Novato CA, USA, January 2004.

Gryparis, A., Forsberg, B., Katsouyanni, K., Analitis, A., Touloumi, G., Schwartz, J., Samoli, E., Medina, S., Anderson, H.R., Niciu, E.M., Wichmann, H.-E., Kriz, B., Kosnic, M., Skorkovsky, J., Vonk, J.M., Dörtbudak, Z., 2004. Acute effects of ozone on mortality from the "Air Pollution and Health: A European Approach" project. American Journal of Respiratory and Critical Care Medicine 170, 1080–1087.

Hass, H., van Loon, M., Kessler, C., Stern, R., Matthijsen, J., Sauter, F., Zlatev, Z., Langner, J., Foltescu, V., Schaap, M., 2003. Aerosol modelling: results and intercomparison from European Regional-scale modelling systems Special Report EUROTRAC -2 ISS, Garmisch Partenkirschen.

Hodzic, A., Vautard, R., Bessagnet, B., Lattuati, M., Moreto, F., 2005. On the quality of long-term urban particulate matter simulation with the CHIMERE model. Atmospheric Environment 39, 5851–5864.

Horowitz, L.W., Walters, S., Mauzeralles, D.Z., Emmonds, L.K., Rasch, P.J., Granier, C., Tie, X., Lamarque, J.F., Schultz, M.G., Brasseur, J.P., 2003. A global simulation of tropospheric ozone and related tracers: description and evaluation of MOZART, version 2. Journal of Geophysical Research 108 (D24), 4784.

Hov, O., Stordal, F., Eliassen, A., 1985. Photochemical oxidant control strategies in Europe: a 19 day case study using a Lagrangian model with chemistry NILU TR5/95.

Lattuati, M., 1997. Contribution à l'étude du bilan de l'ozone troposphérique à l'interface de l'Europe et de l'Atlantique Nord: Modélisation lagrangienne et mesures en altitude. Ph.D. Thesis, Université Paris 6, Paris.

van Loon, M., Roemer, M., Builtjes, P., 2004. Model intercomparison in the frame work of the review of the Unified EMEP model, TNO-MEP Report, R2004-282, Apeldoorn, The Netherlands.

Middleton, P., Stockwell, W.R., Carter, W.P., 1990. Aggregation and analysis of volatile organic compound emissions for regional modelling. Atmospheric Environment 24, 1107–1133.

Minguzzi, E., Bedogni, M., Carnevale, C., Pirovano, G., 2005. Sensitivity of CTM simulations to meteorological input. International Journal of Environment and Pollution 24 (1/2/3/4), 36–50.

Moussiopoulos, N., Douros, I., 2005. Efficient calculation of urban scale air pollutant dispersion and transformation using the OFIS model within the framework of CityDelta. International Journal of Environment and Pollution 24, 64–74.

Passant, N.R., 2002. Speciation of UK emissions of NMVOC, AEAT/ENV/0545 report, London.

Roemer, M., Beekmann, M., Bergström, R., Boersen, G., Feldmann, H., Flatoy, F., Honoré, C., Langner, J., Jonson, J. E., Matthijsen, J., Memmesheimer, M., Simpson, D., Smeets, P., Solberg, S., Stern, R., Stevenson, D., Zandveld, P., Zlatev, Z., 2003. Ozone trends according to ten dispersion models. Special Report. EUROTRAC-2 ISS, Garmisch Partenkirschen.

Schaap, M., van der Gon, H.A.C.D., Dentener, F.J., Visschedijk, A.H.J., van Loon, M., ten Brink, H.M., Puteaud, J.P., Guillaume, B., Liousee, C., Builtjes, P.H.J., 2004. Anthropogenic black carbon and fine aerosol distribution over Europe. Journal of Geophysical. Research 109, D18201.

Schmidt, H., Derognat, C., Vautard, R., Beekmann, M., 2001. A comparison of simulated and observed ozone mixing ratios for the summer of 1998 in Western Europe. Atmospheric Environment 36, 6277–6297.

Simmonds, P.G., Derwent, R.G., Manning, A.L., Spain, G., 2005. Significant growth in surface ozone at Mace Head, Ireland, 1987–2003. Atmospheric Environment 38, 4769–4778.

Stern, R., Yamartino, R., Graff, A., 2003. Dispersion Modelling within the European Community's air quality directives: Long term modelling of O_3, PM10 and NO_2. 26th ITM on Air Pollution Modelling and its Application. May 26–30, 2003, Istanbul, Turkey.

Taylor, K.E., 2001. Summarizing multiple aspects of model performance in a single diagram. Journal of Geophysical Research 106 (D7), 7183–7192.

Thunis, P., Rouil, L., Cuvelier, C., Stern, R., Kerschbaumer, A., Bessagnet, B., Schaap, M., Builtjes, P.H.J., Tarrason, L., Douros, J., Moussiopoulos, N., Pirovano, G., Bedogni, G., 2006. Analysis of large and fine scale model responses to emission-reduction scenarios within the CityDelta project. Atmospheric Environment, Submitted for publication.

Tilmes, S., Brandt, J., Flatoy, F., Bergrstrom, R., Flemming, J., Langner, J., Christensen, J.H., Frohn, L.M., Hov, O., Jacobsen, I., Reimer, E., Stern, R., Zimmermann, J., 2002. Comparison of five Eulerian air pollution forecasting systems for the summer of 1999 using the German ozone monitoring data. Journal of Atmospheric Chemistry 42, 91–121.

Vautard, R., Beekmann, M., Roux, J., Gombert, D., 2001. Validation of a hybrid forecasting system for the ozone concentrations over the Paris area. Atmospheric Environment 35, 2449–2460.

Vautard, R., Martin, D., Beekmann, M., Drobinski, P., Friedrich, R., Jaubertie, A., Kley, D., Lattuati, M., Moral, P., Neininger, B., Theloke, J., 2003. Paris emission inventory diagnostics from ESQUIF airborne measurements and a chemistry transport model. Journal of Geophysical Research 108 (D17), 8564.

Vestreng, V., 2003. Review and revision of Emission data reported to CLRTAP, EMEP Status report July 2003.