MITLibraries | DSpace@MIT

MIT Open Access Articles

Evaluation of WRF mesoscale simulations and particle trajectory analysis for the MILAGRO field campaign

Massachusetts Institute of Technology

Atmospheric
Chemistry
and Physics

Evaluation of WRF mesoscale simulations and particle trajectory analysis for the MILAGRO field campaign

B. de Foy[1], **M. Zavala**[2,3], **N. Bei**[2,3], **and L. T. Molina**[2,3]

[1]Department of Earth and Atmospheric Sciences, Saint Louis University, MO, USA
[2]Molina Center for Energy and the Environment, CA, USA
[3]Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, MA, USA

Abstract. Accurate numerical simulations of the complex wind flows in the Mexico City Metropolitan Area (MCMA) can be an invaluable tool for interpreting the MILAGRO field campaign results. This paper uses three methods to evaluate numerical simulations of basin meteorology using the MM5 and WRF models: statistical comparisons with observations, "Concentration Field Analysis" (CFA) using measured air pollutant concentrations, and comparison of flow features using cluster analysis. CFA is shown to be a better indication of simulation quality than statistical metrics, and WRF simulations are shown to be an improvement on the MM5 ones. Comparisons with clusters identifies an under-representation of the drainage flows into the basin and an over-representation of wind shear in the boundary layer. Particle trajectories simulated with WRF-FLEXPART are then used to analyse the transport of the urban plume and show rapid venting and limited recirculation during MILAGRO. Lagrangian impacts were identified at the campaign super-sites, and age spectra of the pollutants evaluated at those same sites. The evaluation presented in the paper show that mesoscale meteorological simulations are of sufficient accuracy to be useful for MILAGRO data analysis.

1 Introduction

The analysis of data from field campaigns can be greatly assisted by using numerical meteorological simulations to identify wind transport patterns. When doing this, it is important to estimate the confidence that can be placed in these simulations and the reliability of the conclusions they suggest. This paper seeks to evaluate the results of the Weather Research and Forecast (WRF) model for interpreting data from the MILAGRO field campaign. Model results are then used to quantify mixing in the basin, venting of the urban plume and Lagrangian transport past the campaign super-sites.

1.1 Basin-scale meteorology

The MILAGRO field campaign was designed to study the regional and global impact of megacities by studying the case of the Mexico City Metropolitan Area (MCMA). It consists of four components ranging from the local to the inter-continental scale: MCMA-2006, MAX-MEX, MIRAGE and INTEX-B. MCMA-2006, coordinated by the Molina Center for Energy and the Environment, focused on understanding urban emissions and boundary layer concentrations in the basin in order to assist MCMA policy makers.

The MCMA is a large sub-tropical city experiencing high pollutant concentrations from a combination of large emissions, weak winds and intense sunshine (Molina and Molina, 2002). It is located in a basin at 2240 mMSL surrounded by mountains on three sides with an opening to the Mexican Plateau to the north and a small gap in the basin rim to the southeast, see Fig. 1. Weak synoptic forcing and intense solar radiation lead to pronounced mountain-valley and urban-induced wind patterns. There has been continuous air pollution monitoring in the basin since the mid-1980's by the "Red Automática de Monitoreo Atmosférico" (RAMA). Numerous additional measurements have taken place (Raga et al., 2001) along with several international field campaigns. A review of meteorological research in the MCMA is presented
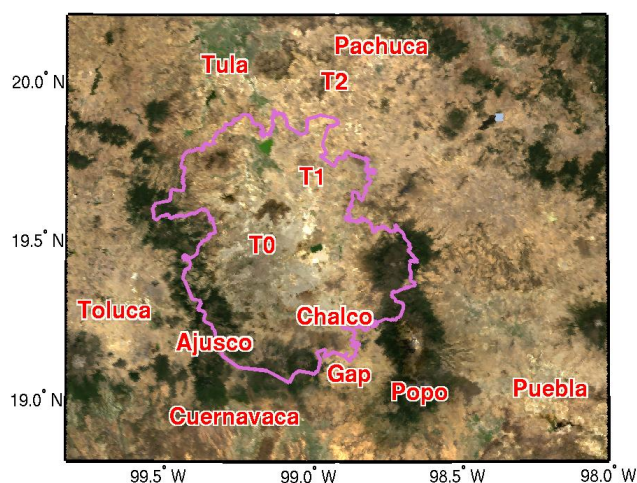
Fig. 1. Map of the MCMA and surrounding areas showing the T0, T1 and T2 supersites and surrounding cities (Puebla, Toluca, Cuernavaca, Pachuca and Tula). Terrain features shown are the Popocatepetl volcano, the Ajusco mountain and the gap in the southeast basin rim near Chalco. Political border of the MCMA as of 2003 in pink, background map is a MODIS true colour image from March 2006.

in Fast et al. (2007) on which we base the following summary.

During the MCMA-2003 field campaign (Molina et al., 2007), days were classified into three types of wind transport using observations and simulations (de Foy et al., 2005), (de Foy et al., 2006c). O3-South days experienced winds from the north during the day. Gap winds from the southeast led to an afternoon convergence zone and an ozone peak in the south of the city. O3-North days had similar daytime winds but an earlier and stronger gap flow in the afternoon leading to a convergence line, urban venting and high ozone in the north of the MCMA. Cold Surge days experienced cold surface winds from the north leading to cloudiness, rain and reduced mixing heights. The form and location of the convergence zone was found to be an important determinant of the ozone peak in the MCMA (de Foy et al., 2006a). Rapid ventilation of the urban plume was found in the basin (de Foy et al., 2006c) with short residence times and little multi-day accumulation. Limited reentry of urban pollution after it left the basin indicated that recirculation does not play a significant role in the MCMA.

This was consistent with the wind patterns observed during the IMADA-AVER Boundary Layer Experiment during February and March 1997 (Doran et al., 1998). Residence times of particle releases in the basin were found to be short but complex convergence patterns influenced the location of the ozone peak (Fast and Zhong, 1998). These findings were further corroborated for different time periods by Jazcilevich et al. (2005). The Mexico City Air Quality Research Ini-

tiative (MARI, Streit and Guzman, 1996) in February 1991 highlighted the contrast between clean sky days with strong venting and polluted days due to weak, local, thermal flows with trapping of tracer particles in night-time drainage flows (Bossert, 1997).

A description of the meteorological conditions during the MILAGRO field campaign and of the meteorological data available is presented in Fast et al. (2007). The basin-scale meteorology during MILAGRO was analysed using cluster analysis (de Foy et al., 2008). This identified wind patters during the field campaign and related them to long term climatological conditions in the basin. Clusters of surface winds, radiosonde observations and radar wind profilers were used to develop a six-category circulation model for each day of MILAGRO. The campaign started with South-Venting days with strong, direct winds flushing the urban plume to the south, leading to clean skies. These are different from O3-South days as there is no afternoon convergence zone and hence much lower pollutant loadings in the atmosphere. This was followed by O3-South and O3-North days similar to those experienced during MCMA-2003. Early morning winds towards the south meet with an afternoon gap flow leading to a wind shift and basin venting towards the north. The timing of the shift determines whether the ozone peak is located in the north (early shift) or the south (late shift) of the urban area. There were three Cold Surge events on 14, 21 and 23 March. After these, conditions were dominated by afternoon convection and rain. These days were separated into two categories depending on whether the convergence leading to the main rain fall was in the north or the south of the basin.

1.2 Model evaluation

Many model evaluation studies rely on case studies and on statistical metrics (Seaman, 2000) and (Chang and Hanna, 2004). Case studies provide valuable information on the detailed simulation of an episode, but it can be difficult to generalise the results. Statistical metrics provide objective methods that extend readily to long time periods, but it can be difficult to interpret the results in terms of inaccuracies in actual atmospheric transport. Furthermore, there can remain a tendency to perform model validation or verification studies as distinct from a model evaluation. By taking the Ptolemaic system of astronomy as an example, Oreskes (1998) questioned whether a validated model is, in fact, valid. As an alternative to validation, she suggested evaluating models in terms of their ability to answer specific scientific questions.

As an example, Pielke and Uliasz (1998) examine the problem of evaluating gridded results with point measurements. They highlight the importance of considering the uncertainties in the meteorology when determining the maximum performance that could be obtained from the model. They consider the spatial scales that can be represented by a particle trajectory model and compare those simulations

to tracer studies. The importance of the specific scientific question used in model evaluation is further demonstrated by Mass et al. (2002). They looked at the accuracy of precipitation forecasts from different spatial resolution grids, and found that higher resolution simulations do not have improved statistical performance scores even though they produce more realistic results. This suggests that the evaluation method used will depend on the application of the simulations. Dabberdt et al. (2004) highlighted the difference between the modelling needs of air quality simulations and those of severe weather forecasting. For dispersion analysis, wind direction and mixing heights during episodes of weak synoptic forcing are more important than the development of storms and quantitative precipitation forecasts.

Statistical metrics can be expanded to address specific scientific questions by considering the flow features that need to be represented. Gilliam et al. (2006) complemented a statistical evaluation of model performance with an evaluation of several months of data filtered by time-scale. This provided information on model performance for intra-day as well as multi-day phenomena. Further evaluation was carried out by comparing pseudo-trajectories from radar wind profiler data and from simulated wind fields to provide an estimate of model errors in terms of wind transport. Rife et al. (2004) analysed model performance for a low-level wind case in the Salt Lake Valley. Standard verification statistics were not able to discriminate between simulations with a large range of horizontal resolutions. Alternative metrics based on the spectral power in the wind speed signal and the spatial variance of the winds did show improvements with finer scale model results. An alternative metric that identifies the dominant feature of the flow can be used for statistical comparisons. Case et al. (2004) devised a metric to identify a sea breeze front and use that in their model evaluation. de Foy et al. (2006a) calculated a metric for the displacement of the convergence line in the MCMA basin from surface measurements that could be compared with model results.

When dealing specifically with air quality episodes, the dispersion of air pollutants is the main criteria for model evaluation. Bao et al. (2008) used the WRF model to simulate fine scale wind flows in California's Central Valley. The model was evaluated by comparing individual episodes with radar wind profiler data, and the model is found to correctly represent the different features of the flow. Detailed analysis of the errors can then lead to model diagnosis, in this case that errors could be due the initial and boundary conditions and to limitations in the land surface model. A particularly stringent test of meteorological model performance is to compare the concentrations of secondary pollutants with measured values. Otte (2008) compared predicted ozone levels with measurements in order to evaluate the performance of different model configurations. In this way, air quality measurements are used to show that observation nudging in the meteorological simulations leads to improved model accuracy. Bei et al. (2008) used ozone simulations to evalu-

ate three-dimensional variational data assimilation in MM5. Model improvements are demonstrated using plume position and peak ozone timing and concentrations in addition to statistical comparisons with meteorological observations.

As mentioned above (Pielke and Uliasz, 1998), particle trajectories are a direct representation of simulated atmospheric transport. Tracer studies, where available, provide a rigorous evaluation of the simulations for the episodes studied (Warner et al., 2008), (Nachamkin et al., 2007). In the absence of tracer studies, large plumes can serve as natural experiments for model evaluations. For example, a large nickel and vanadium plume was measured in the urban area during MCMA-2003 (Johnson et al., 2006). This was large enough to suggest a single distant combustion plume of fuel oil, a finding which was corroborated by back-trajectories pointing directly to a power plant. Forward simulations can then complement back-trajectories in evaluating the flow, for example by examining vertical dispersion and comparisons with surface measurements (de Foy et al., 2007). This case by case approach can be expanded by using "Concentration Field Analysis" (Seibert et al., 1994). Back-trajectories are used in combination with point measurements of an air pollutant over a long period of time to identify potential source regions. If the sources are known, then this method can be used to evaluate the simulated trajectories by comparing the potential source regions with the actual emission inventory. Results with carbon monoxide (CO) showed that MM5-FLEXPART provided realistic simulations of the wind transport during MCMA-2003 (de Foy et al., 2007). The method was then used with sulfur dioxide (SO_2) measurements to show the impacts of industrial point sources on the urban air quality. Instead of providing a validation of the model, this method yields an evaluation where the potential source fields contains information on the model strength and weaknesses by identifying known, unknown and spurious transport directions.

1.3 Outline

This paper evaluates model performance from one set of simulations with MM5 and two sets of simulations with WRF using three evaluation methods during the MILAGRO field campaign. The fate of the urban plume is then analysed with the best case simulations. We first describe the measurements in Sect. 2 and the models in Sect. 3. The evaluation of the model with statistical metrics is presented in Sect. 4, followed by evaluation with the Concentration Field Analysis in Sect. 5. The most accurate simulations are then analysed in greater detail with cluster analysis in Sect. 6. Having evaluated the model, we analyse the wind transport of the urban plume in the MCMA basin in Sect. 7.
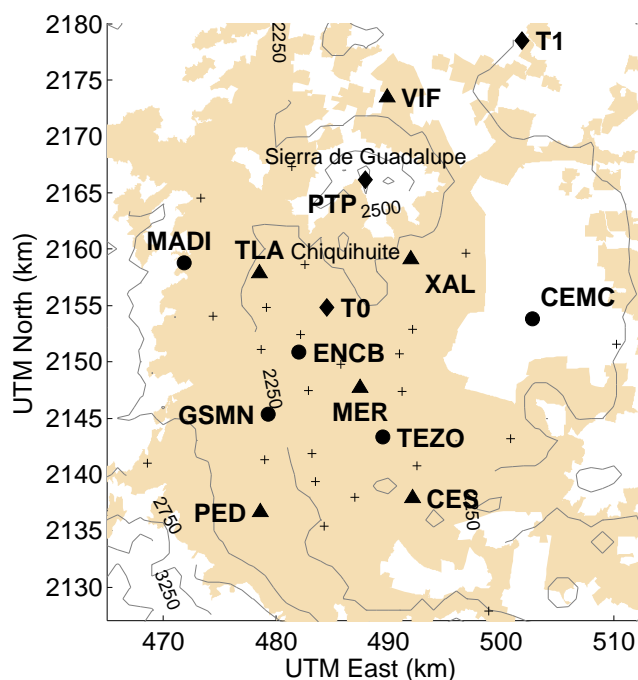
Fig. 2. Map of the MCMA showing the Meteorological Service stations (circles), the RAMA stations used in this study (triangles) and others (crosses). Also shown are the T0 and T1 supersites. Urban area of the MCMA shown in beige, terrain contours every 250 m.

2 Measurements

Figure 2 shows the location of the measurement sites used in this study. A detailed description of the meteorological data collected during the campaign can be found in Fast et al. (2007) and in de Foy et al. (2008).

The Mexican National Weather Service (SMN) carried out radiosonde observations at its headquarters (GSMN) at 00:00 UTC and 12:00 UTC. The SMN also operates a network of automated surface meteorological stations (EMA) with five stations located in the basin: GSMN, ENCB, TEZO, CEMC and MADI. The Ambient Air Monitoring Network (Red Automática de Monitoreo Atmosférico, RAMA) operates a network of surface stations measuring meteorological parameters and criteria pollutants throughout the city. 1-h average data are available online (http://www.sma.df.gob.mx/simat/) along with detailed information on the station location and surroundings. The General Direction of the National Center for Environmental Research and Training (CENICA) of the Mexican National Institute of Ecology stationed a mobile laboratory at the T1 supersite for the duration of the campaign. 1-min measurements of CO were averaged to one hour for model comparisons in this study.

Radar wind profilers were installed at T0, T1 and T2. These were 915 MHz models manufactured by Vaisala. They

**Table 1.** Model Options for the three cases: MM5, WRFa and WRFb.

Case	MM5	WRFa	WRFb
Model	MM5 v3.7.3	WRF v3.0	WRF v3.0.1
Grids	111×87, 139×85, 94×94	50×40, 64×55, 61×61	111×87, 139×85, 94×94
Resolution	27, 9, 3 km	36, 12, 3 km	27, 9, 3 km
Vertical Levels	41	35	41
Time Step	Fixed	Fixed	Adaptive
Microphysics Param.	Simple Ice	Lin	WSM6
Cumulus Scheme	Kain-Fritsch 2, d12	Kain-Fritsch 2, d12	Kain-Fritsch 2, d123
Boundary Layer Scheme	MRF	YSU	YSU
SW Scheme	Dudhia	Dudhia	Dudhia
LW scheme	Cloud Radiation	RRTM	RRTM
Land Surface Scheme	NOAH	NOAH	NOAH
W damping	–	0	1
6th Order Diffusion	–	0	1
Urbanisation	None (removed)	None (removed)	None (removed)
Modifications	Sea Surface Heat Fluxes	–	–
	Constant Soil Moisture	–	–

were operated in a 5-beam mode with nominal 192-m range gates. As described in Doran et al. (2007), the NCAR Improved Moment Algorithm was used to obtain 30-min average consensus winds. Mexico City is in the Central Standard Time zone (CST), which is 6 h behind UTC. All times in the paper are reported as CST unless marked otherwise.

3 Modelling

Mesoscale meteorological simulations were carried out with both the Pennsylvania State University/National Center for Atmospheric Research Mesoscale Model (MM5, Grell et al., 1995) version 3.7.3 and the Weather Research and Forecast model version 3.0.1 (WRF, Skamarock et al., 2005).

Table 1 shows the model options for the three cases analysed in this study. All simulations use three nested grids with one-way nesting. The initial and boundary conditions were taken from the Global Forecast System (GFS) at 1° spatial resolution and 6 h temporal resolution. The MM5 case ("modified case" in de Foy and Molina, 2006) and the second WRF case ("WRFb") use 41 vertical levels and the same domains (albeit with a slight difference due to an adjustment of projection parameters) see Fig. 3 for a map of the domains used. The first WRF case ("WRFa") uses smaller domains with coarser resolution in the horizontal for domains 1 and 2, and in the vertical for all 3 domains. Because we are using one-way nesting, it was possible to use a nesting ratio of 4 from domain 2 to domain 3 in the WRFa case. This does not pose any numerical problems but enables a larger domain that includes the Pacific and Gulf coasts. All cases calculate diffusion in coordinate space for domains 1 and 2, and in physical space for domain 3 (Zängl et al., 2004). The MRF boundary layer scheme was used in MM5 (Hong and Pan, 1996), and the YSU boundary layer scheme (Hong et al., 2006) in WRF. The Kain-Fritsch convective parameterisation (Kain, 2004) was used in both MM5 and WRF. These options for the boundary layer scheme and for the convective parameterisations were found to be the best choice during prior model testing. Overall, we sought to keep as close a match
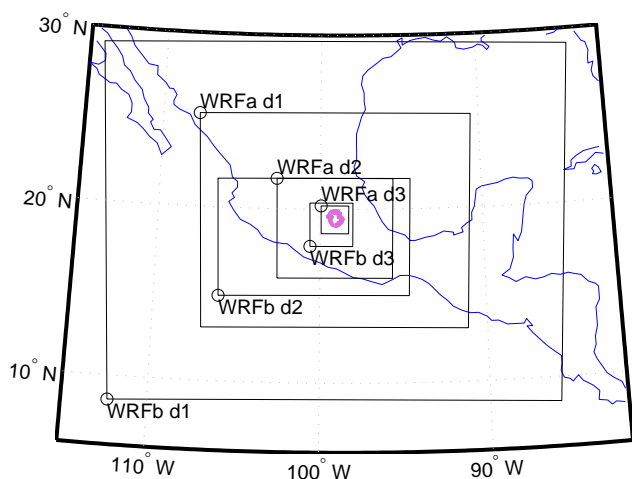
Fig. 3. Map of Mexico showing the 3 domains for the WRFa and WRFb simulations (Note: MM5a are the same as WRFb). Outline of the MCMA shown in bold pink.

as possible in model options between the cases. The intention is to compare model performance for the options that would typically be used, rather than testing for model differences due to changes in coding and advection schemes. We therefore did not select options that were included for backward compatibility, but rather selected the natural equivalent in WRF, for example by selecting the YSU scheme instead of the MRF scheme. Data assimilation was not included in the present work but could be used in the future to get additional improvements in model performance.

High resolution satellite remote sensing was used to improve the land surface representation in the NOAH land surface model for domains 2 and 3 as described in de Foy et al. (2006b). Figures 4 and 5 show the comparison of the default and modified input fields for landuse, surface albedo, vegetation fraction and land surface temperature for domain 3. The default landuse comes from AVHRR (Advanced Very High Resolution Radiometer) data taken in 1992. The growth of the MCMA can be clearly seen along with that of the neighbouring cities of Toluca, Puebla and Cuernavaca. There is more mixed shrub and grassland to the north, and there are are no longer spurious cells of deciduous broadleaf on the urban outskirts. The surface albedo is lower and varies more gradually in the MODIS data. In particular, the values for the urban area vary depending on the neighbourhood. The vegetation fraction has an impact on this, with leafier parts of the city having lower albedos. The MODIS fields clearly show the forests on the surrounding mountains and the barren areas on the plateau to the north as a result of the dry season. The deep soil temperature is substantially higher in the MODIS fields because the default option is to apply an altitude correction to the GFS data which introduces too much cooling. Surface roughness length for urban areas was set

to 25 cm, down from the default 80 cm, to account for the low, flat buildings predominantly found in the MCMA. Rooting depth was increased from 1 to 3 layers, and the radiation stress function parameter and the vapour deficit function parameter were set to the same values as the surrounding areas (100 and 40, respectively). The NOAH land surface scheme in both MM5 and WRF contains some hard-coded modifications to soil type parameters for urban grid cells that were removed. Overall, the impact of the default setup was to reduce the latent heat flux in the urban area to zero. With the current changes, the latent heat fluxes are higher, although they are still lower than in surrounding areas.

Chen et al. (2005) found that latent and sensible heat fluxes were too high over water in MM5 and proposed a correction that can reduce latent heat fluxes by up to a factor of three and sensible heat fluxes by up to a factor of two. This correction was applied to the MM5 simulation, but was not needed for WRF which had heat flux values of comparable magnitude to the post-correction MM5 simulations. One final change to the MM5 code was to keep the soil moisture constant rather than update the fields based on rainfall rates as this had a tendency to corrupt the surface heat fluxes in the basin (de Foy et al., 2006a).

The WRFa case is included in the present study to show the impact of using smaller domains with coarser resolutions for domains 1 and 2. There are some differences in the damping schemes that helped stabilise the solutions for cases with strong convection but do not otherwise affect the results. In particular, there was intense convection in the basin towards the end of the field campaign that resulted in very strong updrafts and downdrafts. As suggested by Deng and Stauffer (2006), the convection scheme was turned on for domain 3 of the WRFb case in order to reduce the vertical velocities in the model.

Stochastic particle trajectories were calculated with FLEXPART (Stohl et al., 2005) using the modified version 3.1, developed by G. Wotawa, for MM5 and using WRF-FLEXPART otherwise (Fast and Easter, 2006), (Doran et al., 2008). Vertical diffusion coefficients were calculated based on the mixing heights and surface friction velocity from the mesoscale models. Sub-grid scale terrain effects were turned off and a reflection boundary condition was used at the surface to eliminate all deposition effects. Particle locations were output every hour for analysis.

For the urban plume, forward trajectories were calculated based on urban CO emissions. These were based on the 2006 official emissions inventory for the MCMA (Comisión Ambiental Metropolitana, 2008). CO emissions are mainly from mobile sources. Gridded emissions were therefore obtained by projecting fine scale road infrastructure network maps of the city onto a 2.25 km grid. This was combined with updated data of vehicle counts and diurnal profiles (Lei et al., 2007). The grids were scaled to have a daily average of 500 particles per hour released in the bottom 50 m of the boundary layer. Figure 6 shows the resulting grid of particle
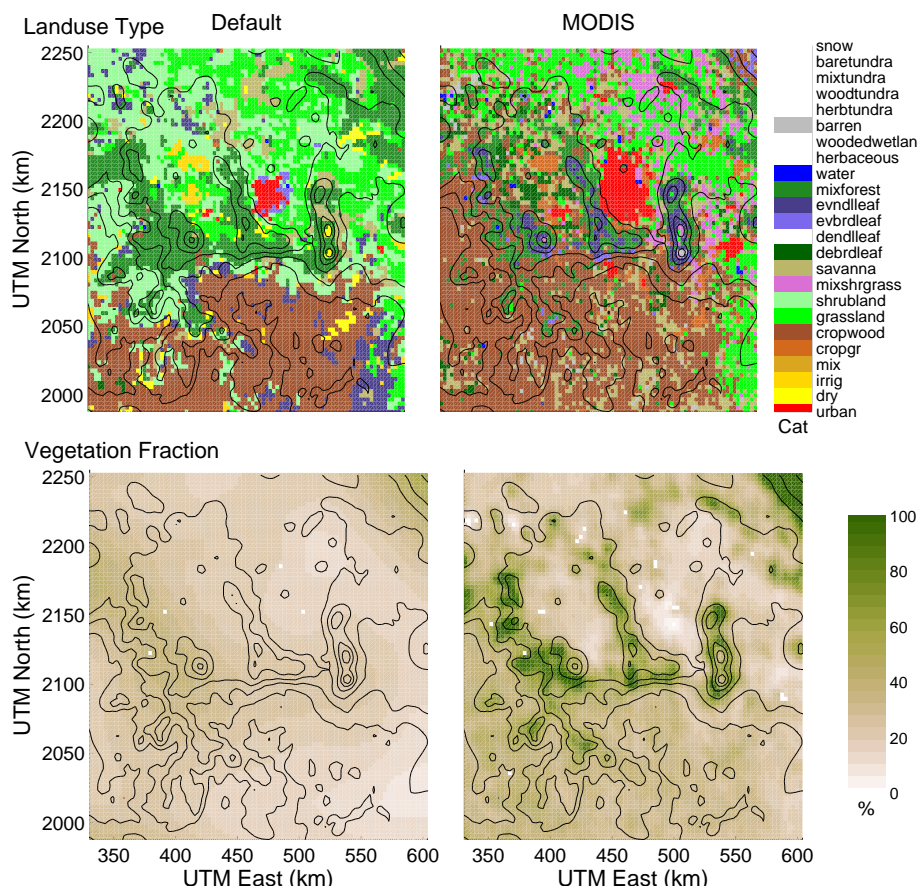
Fig. 4. Comparison of default WRF fields and modified MODIS fields for landuse categories and vegetation fraction for the fine domain. Terrain contours every 500 m.

releases at noon. Particle trajectories were integrated for 48 h.

Concentration Field Analysis (CFA, Seibert et al., 1994) is a method of determining potential source regions by combining simulated back-trajectories with point measurements of air pollutant concentrations. Backward trajectories were calculated by simulating 100 particles every hour at each measurement site. These were released at random in the bottom 50 m of the boundary layer and tracked for 48 h, storing their position every hour. From these, Residence Time Analysis (RTA, Ashbaugh et al., 1985) are calculated by counting the number of particle positions within each cell of a 6 km grid covering the entire basin for each release time. This is the equivalent of a time exposure photograph of the trajectory, indicating the time that the air mass spent in each grid cell before arriving at the release site. These grids were then scaled by the pollutant concentration at the release site for each release time during the entire campaign. By summing these scaled grids and dividing by the sum of the unscaled RTA grids, CFA grids are obtained that indicate potential source regions, see de Foy et al. (2007) for more details.

4 Model evaluation I – statistics

Table 2 shows the bias (Eq. 1, M are model results, O observations), standard deviation of errors (centred root mean square error, RMSEc, Eq. 2) and Pearson's correlation coefficient comparing results from the three model cases with observations. For the surface observations and radiosonde profiles this includes temperature, water vapour mixing ratio, wind speed and wind direction. For the radar wind profiler profiles this includes wind speed and direction. The model parameters used for comparison are the 2-m temperature and humidity and the 10-m winds at the closest grid point.

$$\text{BIAS} = \overline{M} - \overline{O} \tag{1}$$

$$\text{RMSEc} = \left(N^{-1} \sum_{i=1}^{N} \left((M - \overline{M}) - (O - \overline{O}) \right)^2 \right)^{0.5} \tag{2}$$

For the surface observations, the results are presented for each of the five SMN stations in the basin. These were calculated based on 739 data points out of a maximum of 744 h
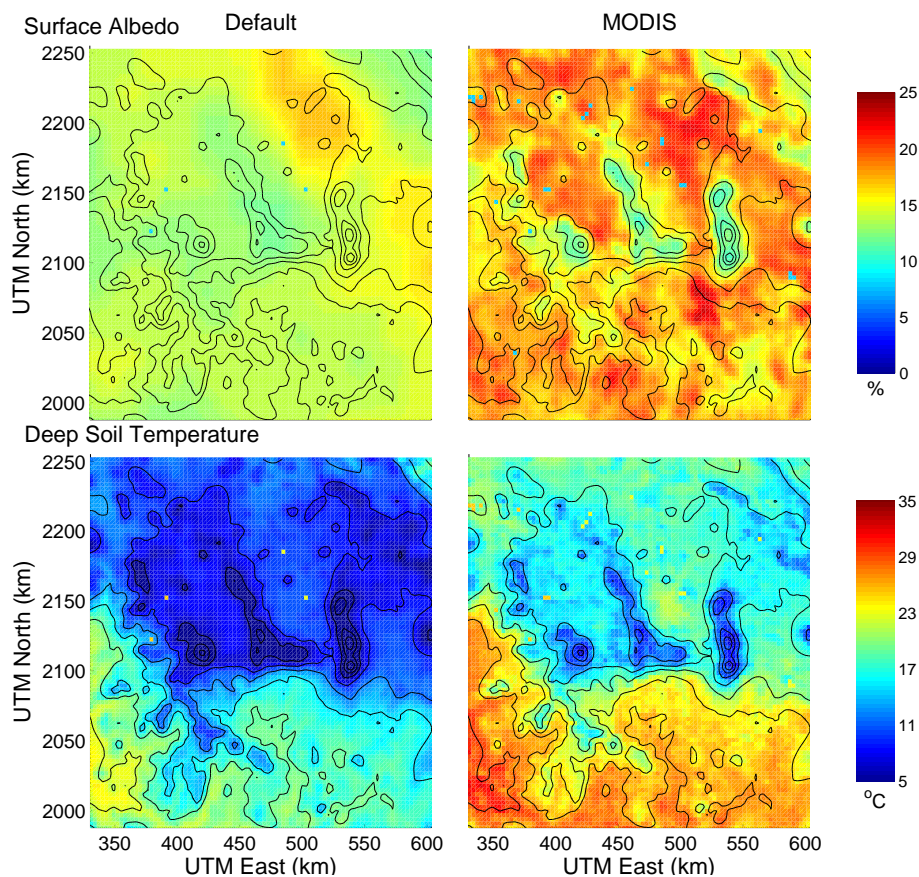
Fig. 5. Comparison of default WRF fields and modified MODIS fields for surface albedo and deep soil temperature for the fine domain. Terrain contours every 500 m.

in the month. Note that comparisons of water vapour mixing ratio are not included for MADI because the data is not representative of a model grid cell due to its location next to a small reservoir. For the radiosonde observations, the statistical indicators were calculated for each of the 111 profiles available. During the 31 days of the month, there were 13 missing soundings out of a maximum 124 possible soundings. As we are interested in the performance of the model within the basin, the comparison was limited to the lower 4000 m which includes the boundary layer. The median and inter-quartile range of the statistical indicators is reported in the table. For the radar wind profilers, the data was limited to the lower 1500 m of the boundary layer based on a trade-off between data availability versus vertical coverage. This allowed comparisons with 409 profiles from T0 and 495 profiles from T1.

Figure 7 shows statistical diagrams comparing the bias and RMSEc with the standard deviation of the observations and of the simulations (de Foy et al., 2006b). Ideally, both the bias and the RMSEc should be close to zero, and the standard variation of the observations and of the simulations should be equal to each other. The standard variation of the data is in-

cluded in order to provide context for the magnitude of the errors. As all of the indicators on the diagram have the same units as the observations themselves, the errors should be to the left and below the standard variation. Ellipses are drawn centred on the average of each indicator for a given model case, with semi-axes given by the standard deviation of the points on the diagram. Small ellipses for the errors suggest a higher degree of consistency in the model behaviour. Ellipses of the standard deviations should lie along the $y=x$ line and be as thin as possible. Their elongation is a function of station variability. In terms of network design, the longer the better although this is outside the control of the analyst.

For temperature, we see an average cold bias of around 1 K in MM5 both for the surface and for the vertical profiles, with variations between 0 and 2 K depending on the station or the profile. The bias is reduced for both WRF cases. WRFb has the smaller bias at the surface and WRFa has the smaller bias for the vertical profiles. One noticeable feature is that there is a cold bias at the urban sites (GSMN, TEZO, ENCB) but a warm bias at the sites on the periphery (MADI and CEMC) (see Table 2). MADI is next to a small reservoir and CEMC is located at an agricultural college on the edge of the MCMA.
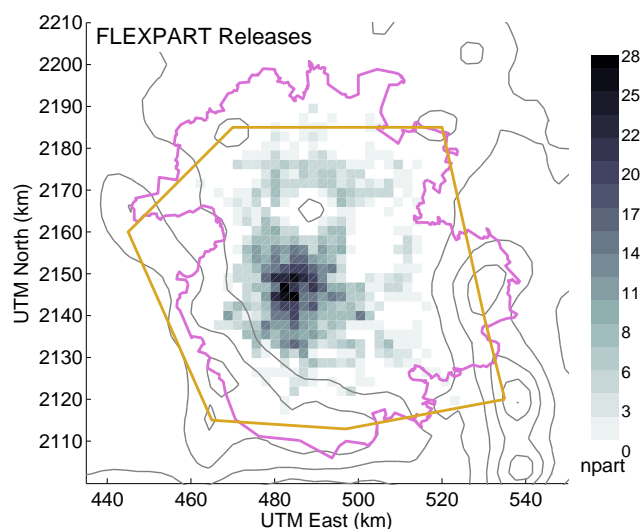
Fig. 6. Number of Particles emitted per grid cell for the urban plume tracer at noon. The basin boundary used for residence time calculations is shown in yellow. Political border of the MCMA as of 2003 in pink, terrain contours every 500 m.

This suggests the presence of an Urban Heat Island that is not fully resolved in the model. In all cases, the model variation is similar to that of the observations and larger than both the bias and the errors.

For water vapour, we see a moist bias both at the surface and in the vertical profiles. The largest bias, with values above 1 g/kg, is for the WRFb case. WRFa, which has the smallest domain and coarsest resolution, has smaller biases with values below 0.5 g/kg. The MM5 case lies between the two WRF cases but is closer to WRFb, with which it shares the same domain. It should be noted that the MM5 bias was larger before the adjustment for sea surface heat fluxes (Chen et al., 2005). The fact that a significant bias remains that is domain dependent suggests that further exploration of the sea surface fluxes is warranted. The relationship with humidity transport is not a linear one. The difference in heating between the land and the sea controls the strength of the sea breeze which controls the extent of air mass transport in the model. Dry conditions over land can therefore lead to a stronger sea breeze and hence more moisture transport and an increase in humidity. Despite the bias, the model variability is similar to that of the observations and the RMSEc values smaller than the data variability.

For wind speed, we see a slow bias in model values. This is particularly pronounced for MM5 in both surface and profile observations. WRF has a reduced bias, although part of this is because the stations outside the city have a fast bias which could be related to the lower surface roughness of the surrounding grid cells. Due to the variation between surface stations and vertical profiles, it is difficult to separate the two WRF cases. WRFa has slightly faster winds and more vari-

Table 2. Model Statistics for the three cases for surface observations, radiosonde observations and radar wind profiler data as well as for CO at T0 and T1. Bias and standard deviation of errors (RMSEc) are defined in the text, Pearson's correlation coefficient is used.

		Bias			RMSEc			Correlation Coeff.		
		MM5	WRFa	WRFb	MM5	WRFa	WRFb	MM5	WRFa	WRFb
Temperature (K)										
Sfc	ENCB	−1.4	−0.9	−0.5	1.8	1.6	1.7	0.94	0.96	0.97
	GSMN	−1.8	−1.3	−1.0	1.8	1.7	1.8	0.94	0.95	0.96
	TEZO	−1.6	−1.0	−0.6	1.7	1.5	1.7	0.95	0.96	0.96
	CEMC	−0.3	0.7	1.1	2.5	3.0	3.3	0.94	0.93	0.92
	MADI	−0.1	0.5	0.8	2.9	3.0	3.1	0.93	0.93	0.93
Raob	MEDIAN	−1.0	−0.4	−0.7	1.1	1.0	0.9	1.00	1.00	1.00
	LQR	−1.8	−1.3	−1.7	0.8	0.7	0.7	0.99	1.00	0.99
	UQR	−0.3	0.2	0.2	1.4	1.3	1.3	1.00	1.00	1.00
Water Mixing Ratio (g/kg)										
Sfc	ENCB	0.8	0.3	1.0	1.7	1.4	1.6	0.69	0.79	0.72
	GSMN	0.9	0.4	1.2	1.7	1.4	1.6	0.70	0.80	0.73
	TEZO	0.3	0.0	0.6	1.5	1.4	1.5	0.69	0.75	0.71
	CEMC	0.7	0.5	1.3	1.6	1.4	1.6	0.66	0.77	0.68
Raob	MEDIAN	0.5	0.3	0.9	0.8	0.8	0.9	0.92	0.92	0.92
	LQR	0.0	−0.2	0.3	0.6	0.6	0.6	0.86	0.86	0.84
	UQR	1.2	0.7	1.4	1.2	1.1	1.2	0.95	0.95	0.96
Wind Speed (m/s)										
Sfc	ENCB	−0.4	−0.1	−0.2	1.4	1.3	1.3	0.64	0.68	0.66
	GSMN	−0.3	0.0	−0.1	1.4	1.3	1.2	0.57	0.68	0.68
	TEZO	−0.7	−0.5	−0.5	1.5	1.3	1.4	0.69	0.70	0.68
	CEMC	−0.5	0.3	0.2	1.6	1.6	1.7	0.66	0.67	0.66
	MADI	0.6	0.7	0.7	1.5	1.2	1.4	0.56	0.66	0.63
Raob	MEDIAN	−0.9	−0.4	−0.6	2.0	2.2	2.1	0.68	0.72	0.65
	LQR	−1.9	−1.1	−1.5	1.5	1.5	1.5	0.28	0.34	0.20
	UQR	0.0	0.7	0.5	2.7	2.7	2.5	0.84	0.87	0.83
RWP	T0 MEDIAN	−0.3	0.1	−0.1	1.0	1.2	1.2	0.52	0.52	0.49
	T0 LQR	−1.2	−0.9	−1.1	0.7	0.8	0.8	−0.13	−0.26	−0.34
	T0 UQR	0.7	1.0	0.9	1.6	1.8	1.8	0.83	0.84	0.88
	T1 MEDIAN	0.0	0.5	−0.1	1.3	1.4	1.3	0.41	0.40	0.39
	T1 LQR	−1.2	−0.7	−1.3	0.8	0.9	0.8	−0.24	−0.31	−0.40
	T1 UQR	1.7	1.7	1.3	1.9	2.1	2.0	0.77	0.84	0.82
Wind Direction (degrees)										
Surface	ENCB	−40	−29	−20	78	70	73	0.32	0.38	0.36
	GSMN	−46	−20	−26	75	66	71	0.38	0.44	0.40
	TEZO	−25	−29	−21	89	65	84	0.30	0.47	0.33
	CEMC	−7	−83	−39	73	84	88	0.37	0.31	0.28
	MADI	11	3	−6	73	67	71	0.40	0.43	0.38
Raob	MEDIAN	−4	0	−9	41	34	36	0.53	0.55	0.51
	LQR	−38	−29	−44	23	15	21	0.37	0.38	0.34
	UQR	40	27	37	60	57	53	0.75	0.74	0.68
RWP	T0 MEDIAN	−6	−2	−4	26	24	25	0.63	0.63	0.63
	T0 LQR	−61	−42	−59	8	8	8	0.44	0.41	0.46
	T0 UQR	23	31	29	48	48	46	0.76	0.75	0.75
	T1 MEDIAN	−12	−4	0	16	17	14	0.61	0.63	0.62
	T1 LQR	−51	−37	−29	6	4	4	0.42	0.44	0.42
	T1 UQR	27	21	30	35	34	31	0.74	0.75	0.76
Carbon Monoxide (ppm)										
	T0	0.12	−0.14	−0.10	0.99	0.88	0.93	0.53	0.51	0.43
	T1	−0.03	−0.13	−0.16	0.46	0.43	0.36	0.29	0.15	0.33

ability in the vertical. If the bias were just a function of surface roughness, then we could expect the bias to be stronger at the surface than in the vertical. This suggests that other factors are at work, which could include terrain shielding as described in Sect. 6. Overall, the RMSEc's are slightly smaller than the standard deviations of the data and correlation coefficients are above 0.65. For the wind profiler data, there is a similar slow bias for MM5 but very little bias for WRF. The errors can be larger and the correlation coefficients correspondingly lower than for the radiosonde profiles.

Wind direction is the most variable and most difficult parameter to simulate. Overall, there is a net bias at the surface of 20 degrees or more, but very little net bias in the vertical profiles. RMSEc's can be quite large. While the values suggest that WRF performs better than MM5, there is little difference between WRFa and WRFb. If anything, WRFa seems to have better statistics than WRFb.

To summarise, all the cases have a cold bias and a moist bias accompanied by a slight slow bias. WRF has reduced biases and errors overall compared with MM5. In terms of
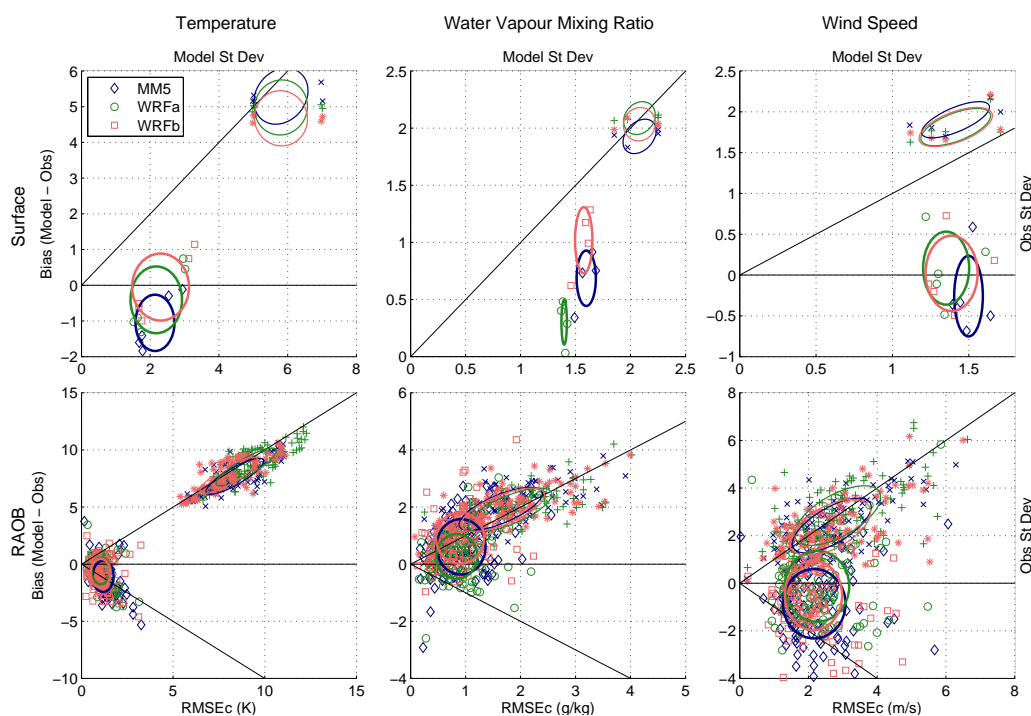
Fig. 7. Statistics diagrams for temperature, water vapour mixing ratio and wind speed comparing the MM5, WRFa and WRFb cases at five surface sites for the duration of the campaign and at one radiosonde site for each profile available during the campaign. Open symbols show the model bias versus the centred root mean square error (RMSEc). Closed symbols show the model standard deviation versus that of the observations. Ellipses are centred on the mean with semi-axes given by the standard deviations of the metrics.

statistics, the smaller domains with coarser resolutions for domains 1 and 2 have improved indicators especially for moisture transport. While this is consistent with Mass et al. (2002), the differences in the statistical metrics are small compared to the magnitude of the errors. Furthermore, the aggregate data for the entire campaign do vary between days from different episode types, as can be seen from the scatter in the radiosonde metrics (Fig. 7). This motivates the use of alternative evaluation methods that provide information about actual wind transport.

5 Model evaluation II – particle trajectories

Having evaluated model statistics, this section focuses on model evaluation using simulated trajectories and Concentration Field Analyses. Comparisons of simulated CO concentrations with measurements serve to evaluate the forward dispersion in the model. Comparisons of CFA results serve to evaluate the backward trajectories by determining if the model can identify known emission sources and known areas without sources. These methods are not independent, but they provide a way of evaluating the model in terms of the same simulated meteorological transport from two different points of view. Because transport is the final product of the simulations, this method of evaluation is able to evalu-

ate model performance in terms of its actual applications and thereby establish the degree of usefulness of the simulations.

5.1 CO Concentrations

Figure 8 shows the time series of measurements and simulated CO at the T0 and T1 supersites for the time period from 10 to 31 March (inclusive). Concentrations were obtained by counting the number of particles in a 9 by 9 km by 200 m high grid cell and scaling by the mass of CO represented by each particle. Table 2 shows the corresponding statistical metrics. Biases are small, and RMSEc's are below 1 ppm at T0 and below 0.5 ppm at T1. The correlation coefficients are around 0.5 at T0 and 0.3 at T1.

The time series show that the model performed worst during the last 4–5 days of the campaign when strong convection took place. Conversely, the best performance took place earlier in the campaign when there was little cloudiness, as has been observed during the MCMA-2003 campaign (de Foy et al., 2006a). At T0 we have a clear diurnal component due to its urban location with some inter-day variations due to episode types. T1 in contrast has much lower concentrations with more variation between "clean" and "dirty" days. Overall, the models are able to represent many of the diurnal variations and the day-to-day variations of the measurements.
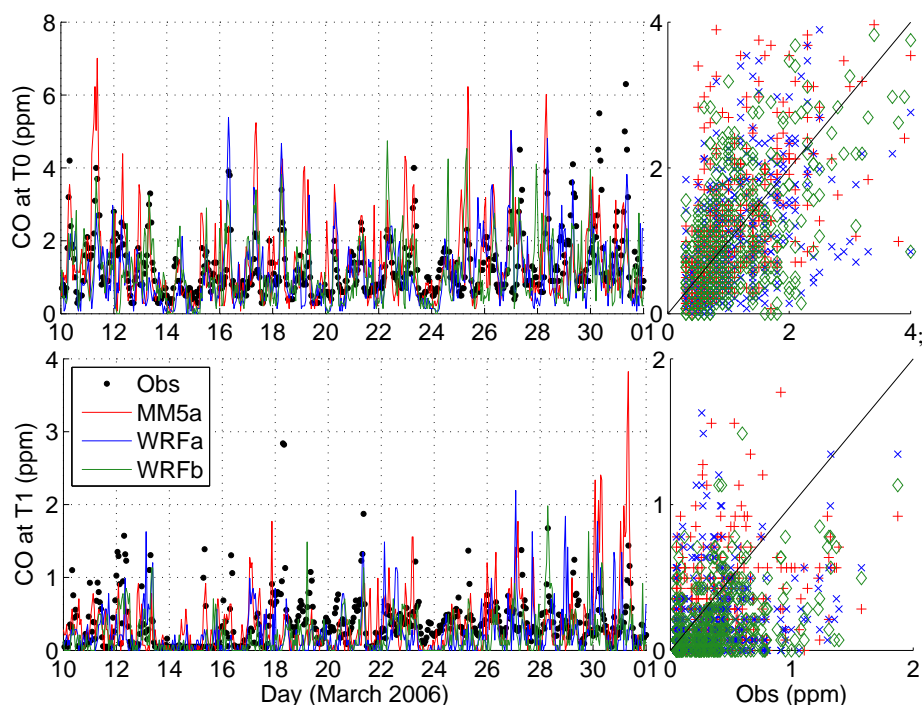
Fig. 8. CO timeseries at T0 and T1 for observations and three model cases and scatter plot of model (y-axis) versus observations (MM5a (+), WRFa (x), WRFb (◇)). Simulated CO obtained from FLEXPART particle count in a 9 by 9 km by 200 m high grid cell around the sites.

There is considerable variation in the simulations of individual peak concentrations between the models, with improved results from the WRFb case on a number of events. Nonetheless, there is little difference between the models in terms of statistical metrics.

5.2 Concentration Field Analyses

CO is emitted mainly by vehicular traffic and has well-characterised sources in the urban area (Zavala et al., 2006). SO_2 is emitted mainly by large point sources such as industrial complexes or volcanoes (Grutter et al., 2008). Backward trajectories were calculated from the RAMA sites with CO and SO_2 measurements. CFA was then calculated for the duration of the campaign. A detailed description of the method and the description of the sources for both CO and SO_2 can be found in de Foy et al. (2007). Figures 9 and 10 show the results for the 3 model cases at PED, in the southwest of the city, and VIF to the north of the MCMA for CO and SO_2. The CFA method works by identifying trajectories associated with polluted air, but has no way of determining when an emission occurred along a given trajectory. It is therefore better at resolving direction of transport than distance from the receptor site. This means that source regions from further afield will be identified as streak lines along the prevailing wind directions, extending from beyond the source towards the receptor site.

At PED, wind transport is dominated by drainage flows from the west during the night. A wind shift at sunrise leads to weak transport from the north in the morning turning to easterly during the day and back to westerly at night. For CO, CFA at PED should identify potential sources from the urban area that lies north to east of the site, and should identify clean air when there are drainage flows from the mountains to the west. For MM5, CFA identifies sources to the north as we expect, but there is a spurious potential source region in the uninhabited area to the northwest of the MCMA. This is even more pronounced with WRFa where it is stronger than the urban signal itself. For WRFb, this signature is much reduced, and the main potential source region now points to the urban area to the east and to the north. This suggests that WRFb has a much better representation of CO transport. Transport in MM5 and WRFa on the other hand is contaminated by a signature from the northwest which is probably due to an inaccurate representation of morning northerly winds from the edge of the basin rather than from the urban area.

At VIF, wind transport is from the northwest at night turning gradually to northeasterly during the day. On some days there is an afternoon shift to southerly winds. There are few sources of CO to the north, and so CFA should be able to identify the urban area to the south despite the predominantly northerly winds. There are mixed results for MM5 with some potential source regions to the south but a significant signature to the north. WRFa is a considerable improvement with
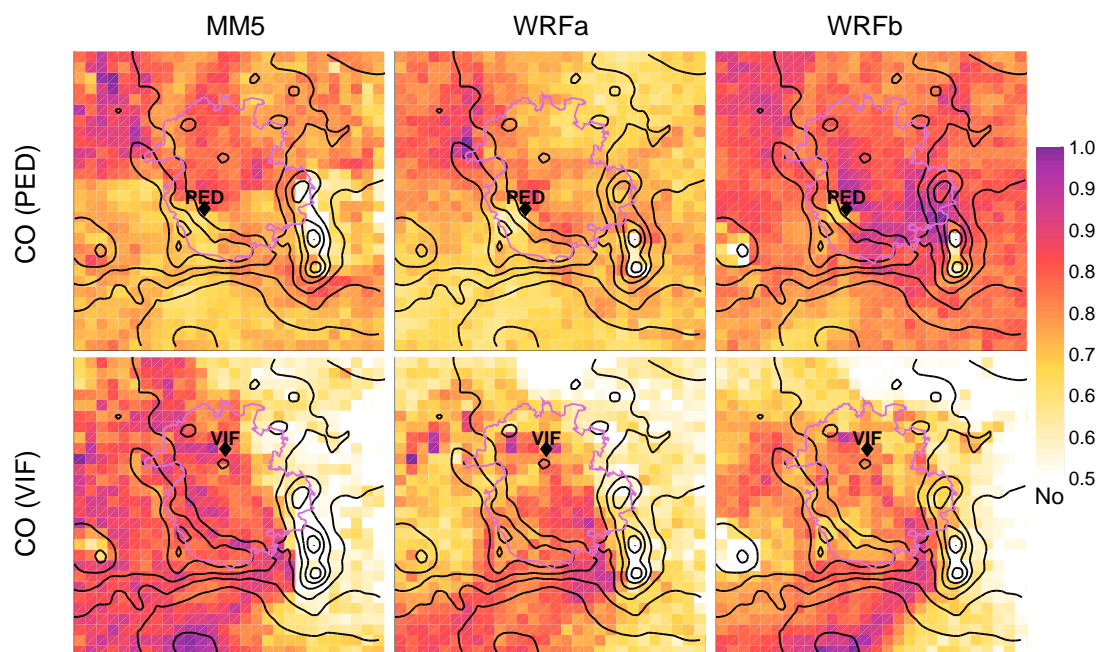
Fig. 9. Concentration Field Analysis of CO based on measured concentrations and simulated back-trajectories at PED and VIF for the three model cases. Back-trajectories were calculated every hour of March 2006. High non-dimensional number (purple) indicates possible source regions, low numbers (white and yellow) indicate areas with low emissions.
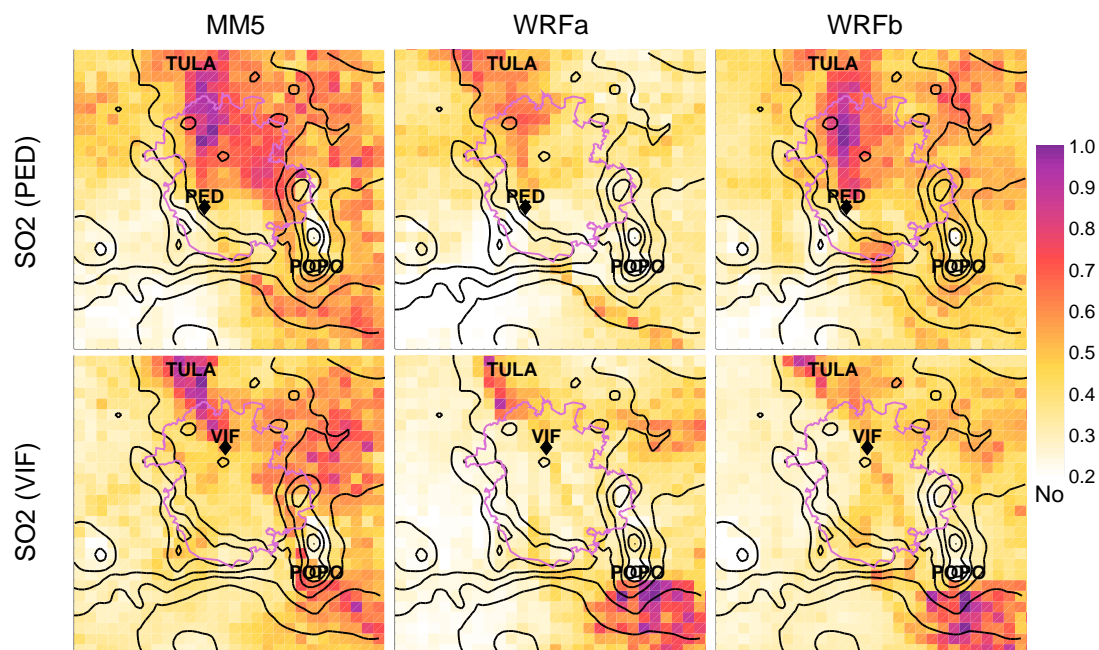
Fig. 10. Concentration Field Analysis of SO$_2$ based on measured concentrations and simulated back-trajectories at PED and VIF for the three model cases. Back-trajectories were calculated every hour of March 2006. High non-dimensional number (purple) indicates possible source regions, low numbers (white and yellow) indicate areas with low emissions.

signatures to the south, although there is also a signature to the west. This is most likely the same, incorrect, representation of morning drainage flows that occurred at PED. WRFb has the clearest CFA signature pointing to the MCMA, suggesting that it is the most accurate. Note that all three simulations have a strong signature coming from the gap flow. The method is not able to identify where along a given trajectory the emission took place, so this can be due either to emissions south of the basin, or to emissions along the way when there are strong gap flows. Studies of basin dynamics have highlighted the importance of the gap flow in forming a convergence line and flushing the urban air mass to the north (de Foy et al., 2006a, Doran and Zhong, 2000). The CFA signature therefore provides further evidence of this feature.

For SO_2, we see that all cases and both sites point to the Tula industrial site to the northwest of the MCMA, as discussed in de Foy et al. (2007). There are notable differences however between the cases, with MM5 pointing to areas to the northeast of the MCMA for both PED and VIF. As there are few known sources of SO_2 in this direction, these are believed to be model artefacts. WRF has more focused CFA signatures indicating that it is better able to resolve the wind flows associated with the SO_2 peaks. Of particular interest, CFA at PED sees little impact from the volcano, but at VIF there is a significant potential source area. Preliminary analysis suggests that during MILAGRO there were indeed more volcanic impacts on SO_2 levels in the basin. Future analysis will need to look in detail at the individual episodes to quantify the volcano impacts. For this paper, it is sufficient to see that all three cases do identify SO_2 transport from known sources, and that WRF performs better than MM5. In contrast to some of the results from the statistical analyses, CFA suggests that the WRFb case represents an improvement over the WRFa case. As described above, CFA is an integral test of model performance in the same terms as those in which the model will be used. We therefore conclude from the present section that the WRFb case is preferable to WRFa and will restrict the remaining analysis to WRFb.

6 Model evaluation III – comparison with cluster analysis

Cluster analysis has been used for meteorological analysis of conditions leading to poor air quality, see de Foy et al. (2008) for a review of existing work and application to the MILAGRO campaign. In this section, we seek to extend this work by evaluating the performance of the model in terms of the representation of the flow features identified by cluster analysis. de Foy et al. (2008) identified 8 surface wind clusters and 6 radiosonde observation clusters based on observations from a 10 and 7 year period respectively (that included MILAGRO) as well as 12 radar wind profiler clusters were identified during the field campaign itself. To compare the simulations with the observations, model data was aggregated

according to the existing clusters. Wind roses and median profiles were then used to see if the simulations represent the features shown in the observations.

With multiple stations in the basin, comparison of surface wind clusters provide information on the spatial detail of the surface transport. The results for the northerly flow and southerly flow clusters (Fig. 11) show very good agreement for the daytime sweeping flows that cross the basin, from the north, south or east.

Figure 12 shows a comparison of surface wind roses for the three drainage clusters. For the first drainage cluster (Drain1), we can see that the northwesterly flow is well represented by the model at the three stations to the north and in the center. In the south however the observations show strong down-slope flows from the west and south whereas the simulations have persistent northwesterly flow. Drain1 gives way to a stronger, later drainage cluster (Drain2). In the observations it has very persistent downslope flow at each station with easterly flow in the urban center. In the simulations however the winds are more variable and have a weaker downslope component. This is particularly noticeable at TLA and PED at the base of the western basin rim. The last drainage cluster, occurs just before sunrise (Drain3). This shows the same behavior of under-representing the drainage flow. The net result is variable winds for the eastern stations in the model compared to more persistent easterlies and southeasterlies in the observations. Overall, we therefore see an under-representation of the drainage flows in the model for the stations on the basin rim, but a representative characterisation of flows otherwise.

Figure 13 shows the median vertical profile for each of the 5 radiosonde clusters for both observations and the WRFb simulations at 18:00 UTC. As expected, we see the cold and moist bias detected by the statistical metrics and the extent of the slight slow wind speed bias in the winds. Overall, the variability of the data is well represented by the variability in the model results. Despite the poor statistical metrics, wind direction features are captured by the simulations when considering median profiles. This may explain why the simulated transport is better than would be predicted from correlation coefficients and standard deviation of errors alone.

The most striking discrepancy is in the vertical evolution of horizontal wind speeds. The model clusters have very similar profiles with stronger winds near the surface, decreasing wind speeds to a height of around 2000 m and then increasing speeds aloft. This is in contrast to the observations that have less vertical structure in the individual profiles, but more variation between the clusters. This suggests that there is too much shear in the model with too strong winds in the surface layer. There also seems to be insufficient variability aloft which could be due to the boundary and initial conditions from the GFS model. The wind shear is likewise found to be over-represented in the radar wind profiler clusters despite representative simulations of the surface transport, please see the supplementary material for fur-
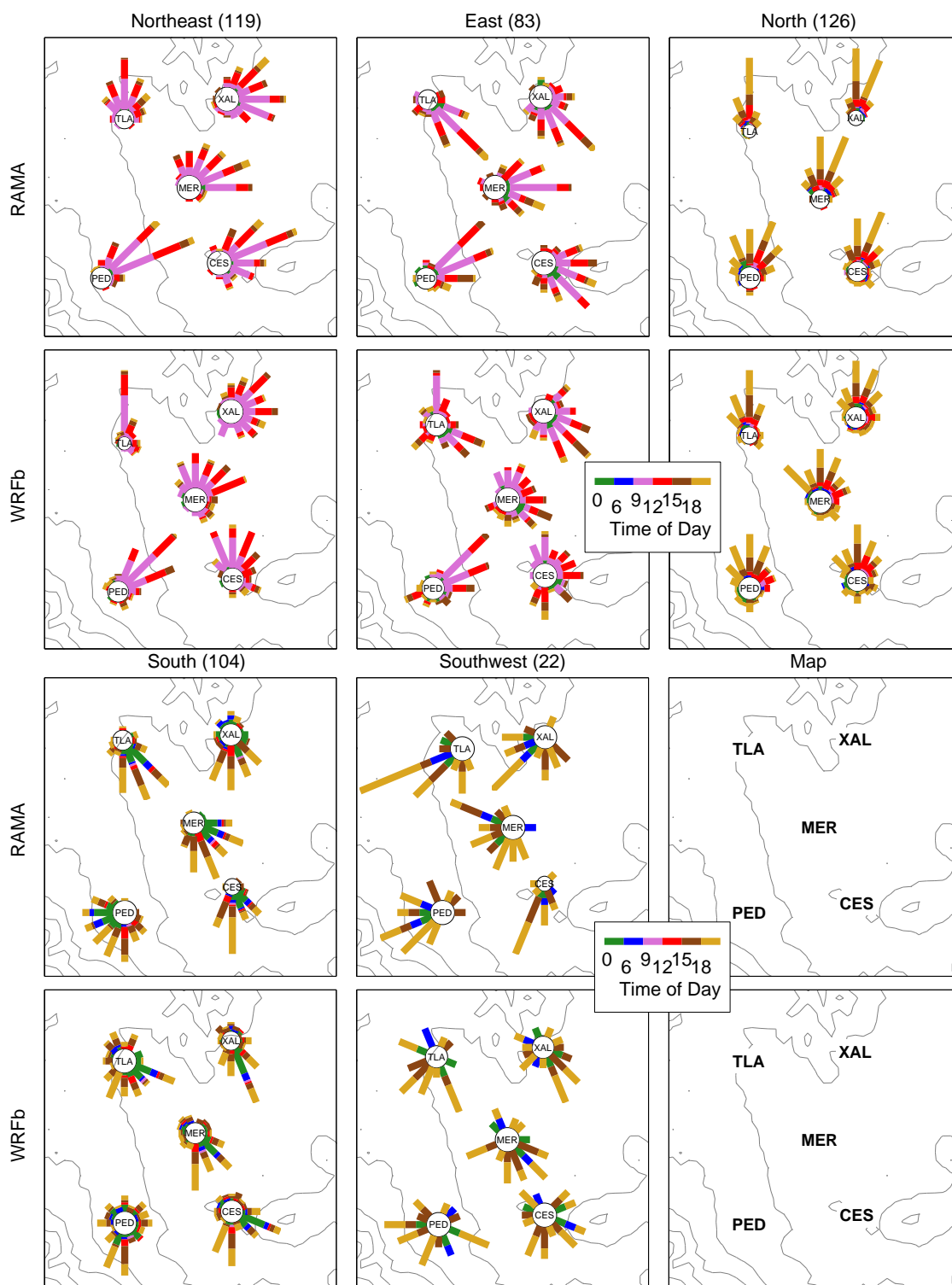
Fig. 11. Comparison of surface wind clusters from RAMA measurements and the WRFb case for the north and south flow clusters occurring during MILAGRO. Station data from TLA and XAL (north), MER (centre), PED and CES (south), see Fig. 2 for station locations. Wind rose categories by time of day. Terrain contours every 250 m.
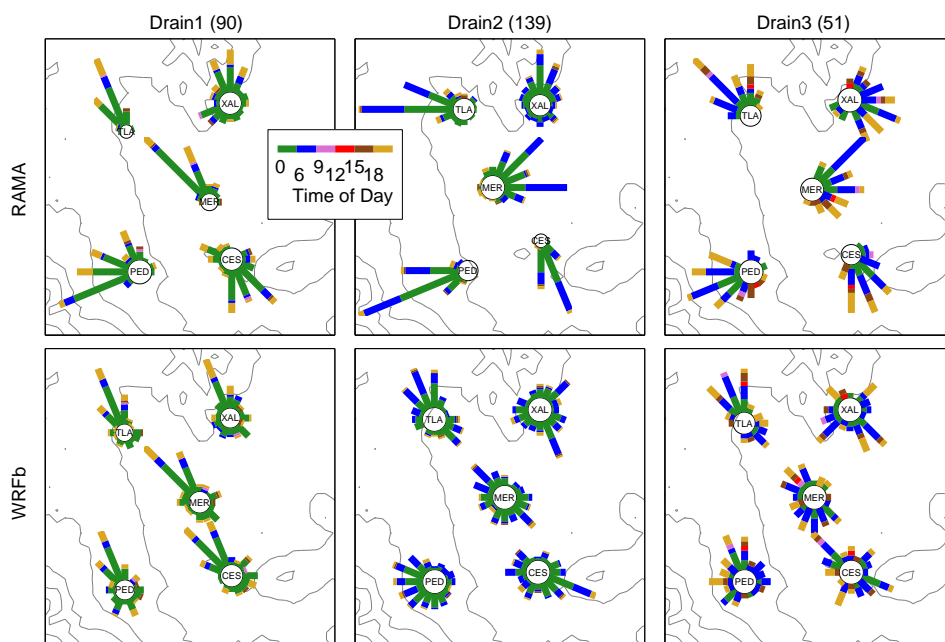
Fig. 12. Comparison of surface wind clusters from RAMA measurements and the WRFb case for the three drainage flow clusters occurring during MILAGRO. Station data from TLA and XAL (north), MER (centre), PED and CES (south), see Fig. 2 for station locations. Wind rose categories by time of day. Terrain contours every 250 m.
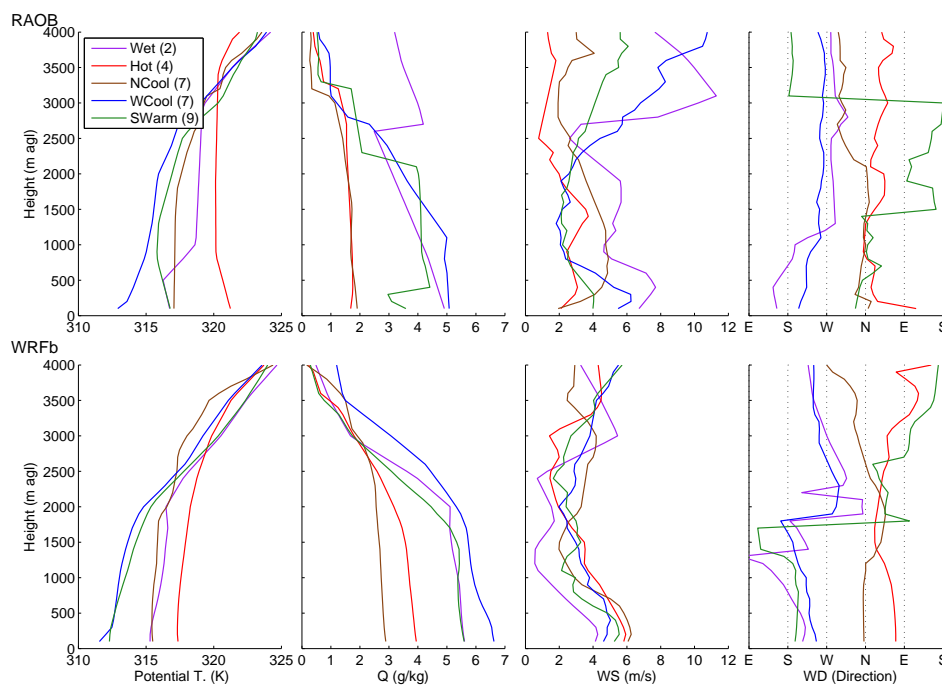
Fig. 13. Median radiosonde profiles at 00:00 UTC, GSMN, from observations and from the WRFb simulations for the five climatological clusters occurring during MILAGRO. In brackets, the number of profiles used.
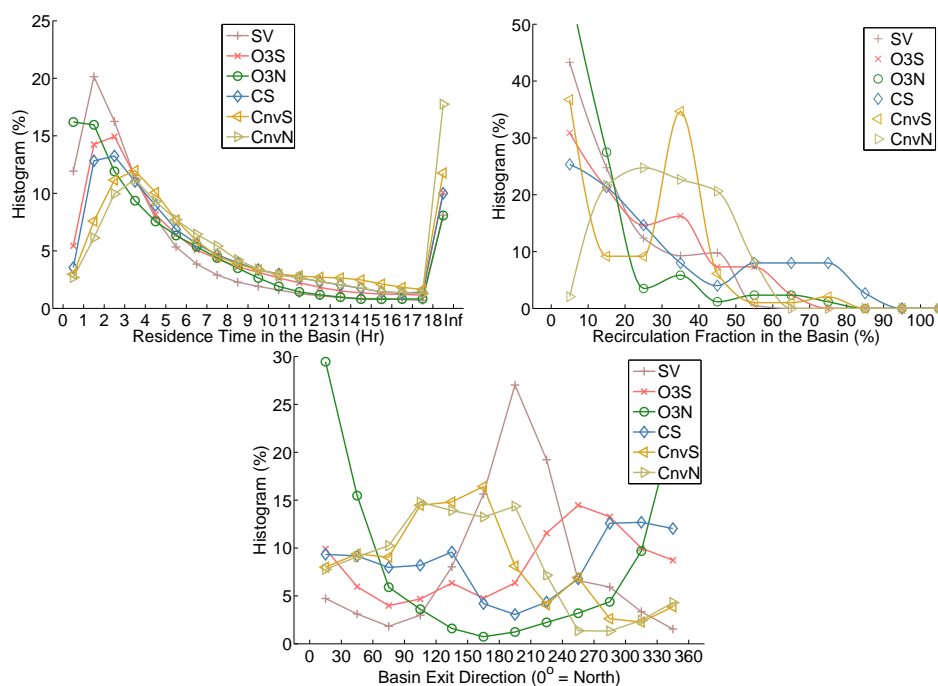
Fig. 14. Histograms of residence time, recirculation fractions and basin exit direction of urban plume tracer particles for the domain shown in Fig. 6. Results for particle releases every hour of March 2006 are classified by episode type.

ther information (http://www.atmos-chem-phys.net/9/4419/2009/acp-9-4419-2009-supplement.pdf).

7 Urban plume

Having established the usefulness of the simulations for analysing wind transport in the MCMA basin, this section evaluates the residence times, age spectra and impact factors at the supersites and compares them with existing studies.

The forward CO tracers were used to evaluate the fate and transport of the MCMA plume within the basin. Figure 14 shows histograms of particle ages, recirculation fractions and basin exit direction separated by episode type. These are calculated for the basin limits, shown in Fig. 6. This shows that the residence time in the basin are shorter than 6 h most of the time, with limited variability between the episode types. As expected, South-Venting has the shortest transport times with 50% of tracers leaving the MCMA in 3 h or less. The Convection-South and Convection-North days have slightly longer residence times with an age peak at 4 h and an increase in particles in the 12 to 18 h range.

Analysis of the recirculation fraction tells a consistent story, with little to no recirculation for half of the time periods during South-Venting and O3-North. Longer residence times of the Convection episodes are a result of the time periods with high rates of recirculation (30 to 40%). Cold surge events had higher recirculation rates during MILAGRO than

during MCMA-2003. This can be traced to stronger gap flows which cause a wind reversal into the basin.

In terms of exit direction, South-Venting has the expected clear signature of southward flow. Likewise, O3-North vents clearly to the north. In between these two cases lies O3-South which vents mainly to the southwest but has a more varied signature, indicating the impact of the afternoon wind shift. Cold Surge events vent in the southwest to north to southeast direction with less southward venting. In contrast to MCMA-2003 it does not vent through the gap flow, which is a function of the stronger gap flow and more southwesterly flows experienced during MILAGRO. This explains the larger recirculation fraction as air initially vented to the south was pushed back northwards. Finally, the Convection episodes have a broad distribution that is mainly to the southeast, as a result of the moist westerlies aloft.

Of particular interest for the analysis of the MILAGRO campaign is the urban impact at T1 and T2. Figure 15 shows a time series of particle tracers in the lower 500 m above ground level for a 9 by 9 km grid cell around the sites. At T1, there is a clear diurnal pattern because it lies on the edge of the MCMA urban emissions. During episodes with winds from the north, there are small morning peaks due to local emissions. Plume transport from the south can lead to high peaks during the night or to lower peaks with longer time duration during the day. T2 is further away from the emission sources and so only has peaks due to longer range transport. The morning peaks at T1 due to local impacts are not
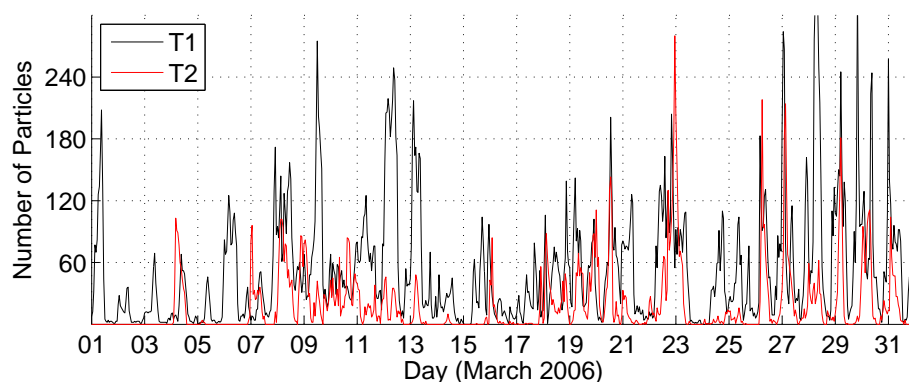
Fig. 15. Urban plume impacts at the supersites T1 and T2. Number of particles are counted for each hour of the campaign in a 9 by 9 km by 500 m high grid cell around each site.

associated with peaks at T2, but urban plume transport at T1 often result in T2 peaks.

Similar plume analysis was performed based on upper level winds in Fast et al. (2007), pseudo-trajectories from radar wind profilers in Doran et al. (2007) and WRF-FLEXPART trajectories in Doran et al. (2008). Table 3 in Fast et al. (2007) classifies the T0-T1-T2 transport as "Unlikely", "Possible" and "Likely" for each day of the campaign. This is based on the synoptic flow conditions and shows very good agreement between the "Likely" days and the present analysis. Differences are to be expected however due to both small scale circulation effects and greater time resolution of the model results. In addition, there may be cases were the simulated plume does not hit a site, but passes very close it. Small model discrepancies can make the difference between having an impact or not. This can be resolved by comparing case by case the measurements and the plume simulations.

17 March ("Possible") has simulated impacts late at night, but no impacts during the day when the simulated plume vented to the northwest and missed T2. On 24 and 25 March ("Likely"), the plume was transported to the northeast and (narrowly) missed the site for most of the day although there were limited impacts at night and early in the morning of the 25th. The actual impacts are determined by the competing effects of the westerlies aloft and mountain blocking at the surface along with the simulation of convection, accounting for the discrepancy. For "Unlikely" days, the present analysis has transport events on 4, 7, 8, 13 and 16 March. Some of these are limited night-time peaks due to nocturnal venting to the north that takes place due to decoupling of surface winds, as described above for the "Likely" days. Finally, 26 March ("Possible") experienced a strong plume hit. Convection on that day led to weak transport and the development of an ageing plume in the basin which hit T2 as it meandered around. This type of event is highly sensitive to the representation of the convection cells in the model. Measurements at

T2 should be used to confirm the event on this day, and determine if similar events took place on other convection days that could have been missed by the model.

While both the present work and Doran et al. (2008) use WRF-FLEXPART, the model domains and configurations are different and the latter uses Four Dimensional Data Assimilation (FDDA). We use a high resolution emission inventory for just the MCMA using the 2006 inventory whereas Doran et al. (2008) simulate regional emissions and biomass burning. The results are in agreement showing clear impacts on 10, 18, 19, 20 and 22 March. As discussed above, we have much smaller impacts on 24 March. Intense convection events are one of the most sensitive features in the model, and so it is to be expected that differences between the runs should be greatest on these days.

One of the goals of the field campaign is to analyse plume ageing. Figure 16 shows particle ages at the three supersites categorised by time of day. At T0, we see a very fresh plume. In the afternoon and evening, tracers are less than 4 h old. This is followed by ageing during the night leading to a bimodal distribution during rush hour with a majority of fresh emissions but up to 10% of aged emissions from the previous evening. This pattern is more pronounced at T1 with a majority of fresh particles during the day followed by ageing at night and a bimodal distribution during rush hour that lasts into the late morning with more than 30% of particles older than 12 h. At T2, we now clearly see the transport of fresh emissions during the afternoon taking 1 to 4 h, with the tracers getting progressively older during the night. By the morning, the plume impacts are at least 12 h old and this gives way to an airmass that is a mixture of fresh (2 to 6 h) and aged (12 to 16 h) tracers.

Doran et al. (2008), Fig. 9, show median particle ages of up to 80 h at the supersites. These events tend to be hits of aged plumes with low concentrations. We carried out 6-day simulations and found very good agreement between the results with similar periods of aged particles. Compared to
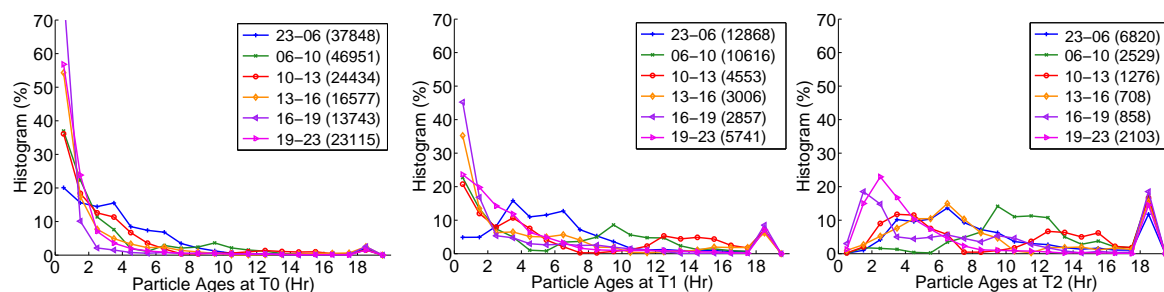
Fig. 16. Age spectra of the urban plume at T0, T1 and T2 categorised by time of day for the entire duration of the campaign. Histograms show the fraction of particles in each age category (1 h increments) for all the urban plume tracer particles present in the grid cell around the site (see Fig. 15). In brackets, the number of particles used for each histogram.

Fig. 16, these aged events correspond to the small fraction of particles with lifetime longer than 18 h.

8 Summary

We have evaluated simulations of wind transport in the MCMA basin with three complementary methods: statistical comparisons with observations, Concentration Field Analysis using air pollutant concentrations, and comparison of flow features using cluster analysis. Having evaluated the numerical model, we used a particle trajectory model to evaluate the transport times and recirculation in the basin as well as the fate of the urban plume and the Lagrangian impacts at the campaign supersites.

Statistical comparisons provide useful information on the biases in the simulations and the magnitude of the errors. Diagrams of bias versus standard deviation of errors (de Foy et al., 2006b), similar to those of Taylor (2001), were shown to be a convenient method of displaying the large amount of data generated and the scatter between different data sources (eg. observation types, station locations, profile times).

A model cold and moist bias was found to vary most with variations in domain choice. This suggests that the impact of the land and sea surface fluxes is different in the mesoscale model than in the global model used for the initial and boundary conditions. Future work will need to investigate the surface fluxes from the NOAH land surface schemes, including the fields of soil moisture which contain large uncertainties. Sea surface fluxes were found to be very high in MM5 but have been corrected in WRF. This suggests one reason for differences in model performance which should be taken into account when considering differences between WRF and MM5, especially for simulations with weak synoptic forcing.

Point comparisons do not evaluate the model in the terms in which it will be used however. Concentration Field Analysis was proposed as an additional evaluation method that takes into account the integrated transport behaviour of the wind simulations over the entire time period of the field cam-

paign. Despite the large statistical errors in winds, CFA shows that the model is able to adequately represent pollutant transport in the basin. The discrepancy between these observations may be because the statistical metrics are sensitive to hourly variability in the winds whereas the wind transport integrates over longer time periods for which the models are better able to capture the flow features (Pielke and Uliasz (1998), Stein and Wyngaard (2001)). This suggests that the simulations are of sufficient accuracy for their intended use. This method provides an efficient analysis using readily available tools that could be seen as a preliminary for more sophisticated modelling techniques (Carmichael et al., 2008).

Cluster analysis of the field campaign was used to compare observations and model results categorised by type of flow. This combines the detail available from episode by episode comparisons with the statistical confidence available from comparing multiple simulations covering a month-long period. In addition, it focuses the comparison on the representation of specific flow features observed in the observations. This analysis shows that WRF represented many of the basin flow features One finding that emerges from the comparisons, is that the drainage flows were under-represented. Because there are few observations that could yield greater detail on this feature in the MCMA, it would be instructive to look at similar cases where such measurements exist, for example in Salt Lake City and Phoenix (Fast and Darby, 2004), (Grossman-Clarke et al., 2008). Another finding is the over-representation of wind shear in the basin. This suggests that there may be more mixing in the MCMA basin than is currently simulated in WRF. This is similar to Thomsen and Smith (2008), who found that schemes with countergradient fluxes gave improved simulations of low-level convergence lines in MM5. It should be emphasised that the YSU scheme has been found to perform well in this complex case, but that cluster analysis can be used to identify potential improvements.

Particle trajectory simulations of the urban emission tracers showed that basin venting was rapid and direct during

MILAGRO, as has been found in prior campaigns. Venting was found to be fastest on South-Venting and O3-North days where winds are fairly uniform and direct. Venting is slowest on days with convection, where convergence zones in the basin lead to recirculation and longer life times. Impacts of the urban plume at T1 and T2 were in agreement with previous findings and identified days on which Lagrangian transport took place along the axis of the supersites. Age spectra of particles at the sites indicated a mixture of fresh emissions and aged emissions from the previous evening. T0 has mainly fresh particles, but this decreases with transport away from the city. At T2, the fresh particles are now 1 to 4 h old, and there is a significant fraction of aged particles. Particles were the freshest during morning rush hour and had an increasing fraction of aged particle as the day progressed and into the following night.

As modelling texts like to quote, "It is the mark of the instructed mind to rest satisfied with the degree of precision which the nature of the subject permits and not to seek exactness where only an approximation of the truth is possible" (Aristotle, 1891). Following Oreskes (1998), we propose that mesoscale meteorological models should be tested to see if their accuracy is sufficient for the purposes at hand. An effective test of this "Aristotelian Accuracy" for air quality simulations is provided by Concentration Field Analysis when there are known area or point sources of air pollutants. In this context, the present work shows that WRF is of sufficient accuracy for analysing transport in the Mexico City basin during the MILAGRO field campaign.

The simulations presented can now be used for Concentration Field Analyses of MILAGRO measurements of other species and/or instruments. This has already been carried out for the ATOFMS (Moffet et al., 2008), for mercury measurements (Rutter et al., 2009) and for PAH measurements (Thornhill et al., 2008). Further analysis can be carried out on the lifetime of the urban plume and the extent of mixing as it is vented outside of the basin. This would complement the chemical clocks described in Kleinman et al. (2008) and the particle ageing described in Doran et al. (2008). Impact areas can be analysed for different sources, including the potential impacts on airborne measurements in and around the MCMA basin. Furthermore, individual plume events can be analysed on a case by case basis to evaluate possible sources and possible atmospheric transformation mechanisms.

References

Aristotle: Nichomachean Ethics, chap. Book I, Chapter III, Longmans, Green and Co., 1891.

Ashbaugh, L. L., Malm, W. C., and Sadeh, W. Z.: A residence time probability analysis of sulfur concentrations at grand-canyon-national-park, Atmos. Environ., 19, 1263–1270, 1985.

Bao, J. W., Michelson, S. A., Persson, P. O. G., Djalalova, I. V., and Wilczak, J. M.: Observed and WRF-simulated low-level winds in a high-ozone episode during the Central California Ozone Study, J. Appl. Meteorol. Climatol., 47, 2372–2394, 2008.

Bei, N., de Foy, B., Lei, W., Zavala, M., and Molina, L. T.: Using 3DVAR data assimilation system to improve ozone simulations in the Mexico City basin, Atmos. Chem. Phys., 8, 7353–7366, 2008, http://www.atmos-chem-phys.net/8/7353/2008/.

Bossert, J. E.: An investigation of flow regimes affecting the Mexico City region, J. Appl. Meteorol., 36, 119–140, 1997.

Carmichael, G. R., Sandu, A., Chai, T., Daescu, D. N., Constantinescu, E. M., and Tang, Y.: Predicting air quality: Improvements through advanced methods to integrate models and measurements, J. Comp. Phys., 227, 3540–3571, 2008.

Case, J. L., Manobianco, J., Lane, J. E., Immer, C. D., and Merceret, F. J.: An objective technique for verifying sea breezes in high-resolution numerical weather prediction models, Weather Forecast., 19, 690–705, 2004.

Chang, J. C. and Hanna, S. R.: Air quality model performance evaluation, Meteorol. Atmos. Phys., 87, 167–196, 2004.

Chen, C. S., Beardsley, R. C., Hu, S., Xu, Q., and Lin, H.: Using MM5 to Hindcast the Ocean Surface Forcing Fields over the Gulf of Maine and Georges Bank Region, J. Appl. Meteorol. Climatol., 22, 131–145, 2005.

Comisión Ambiental Metropolitana: Inventario de Emisiones de la Zona Metropolitana del Valle de México, Tech. Rep. (Web), Secretaría del Medio Ambiente, Gobierno de México, México, 2008.

Dabberdt, W. F., Carroll, M. A., Baumgardner, D., Carmichael, G., Cohen, R., Dye, T., Ellis, J., Grell, G., Grimmond, S., Hanna, S., Irwin, J., Lamb, B., Madronich, S., McQueen, J., Meagher, J., Odman, T., Pleim, J., Schmid, H. P., and Westphal, D. L.: Meteorological research needs for improved air quality forecasting - Report of the 11th prospectus development team of the US

Weather Research Program, B. Am. Meteor. Soc., 85, 563–586, 2004.

de Foy, B. and Molina, L. T.: MODIS Land Surface Parameters for Improved MM5 Simulations in the Mexico City basin during the MILAGRO Field Campaign, in: 7th WRF User's Workshop, Boulder, CO, USA, 2006.

de Foy, B., Caetano, E., Magaña, V., Zitácuaro, A., Cárdenas, B., Retama, A., Ramos, R., Molina, L. T., and Molina, M. J.: Mexico City basin wind circulation during the MCMA-2003 field campaign, Atmos. Chem. Phys., 5, 2267–2288, 2005, http://www.atmos-chem-phys.net/5/2267/2005/.

de Foy, B., Clappier, A., Molina, L. T., and Molina, M. J.: Distinct wind convergence patterns in the Mexico City basin due to the interaction of the gap winds with the synoptic flow, Atmos. Chem. Phys., 6, 1249–1265, 2006a.

de Foy, B., Molina, L. T., and Molina, M. J.: Satellite-derived land surface parameters for mesoscale modelling of the Mexico City basin, Atmos. Chem. Phys., 6, 1315–1330, 2006b.

de Foy, B., Varela, J. R., Molina, L. T., and Molina, M. J.: Rapid ventilation of the Mexico City basin and regional fate of the urban plume, Atmos. Chem. Phys., 6, 2321–2335, 2006c.

de Foy, B., Lei, W., Zavala, M., Volkamer, R., Samuelsson, J., Mellqvist, J., Galle, B., Martinez, A. P., Grutter, M., Retama, A., and Molina, L. T.: Modelling constraints on the emission inventory and on vertical dispersion for CO and SO2 in the Mexico City Metropolitan Area using Solar FTIR and zenith sky UV spectroscopy, Atmos. Chem. Phys., 7, 781–801, 2007, http://www.atmos-chem-phys.net/7/781/2007/.

de Foy, B., Fast, J. D., Paech, S. J., Phillips, D., Walters, J. T., Coulter, R. L., Martin, T. J., Pekour, M. S., Shaw, W. J., Kastendeuch, P. P., Marley, N. A., Retama, A., and Molina, L. T.: Basin-scale wind transport during the MILAGRO field campaign and comparison to climatology using cluster analysis, Atmos. Chem. Phys., 8, 1209–1224, 2008, http://www.atmos-chem-phys.net/8/1209/2008/.

Deng, A. and Stauffer, D. R.: On Improving 4-km Mesoscale Model Simulations, J. Appl. Meteorol. Climatol., 45, 361–381, 2006.

Doran, J. C. and Zhong, S.: Thermally driven gap winds into the Mexico City basin, J. Appl. Meteorol., 39, 1330–1340, 2000.

Doran, J. C., Abbott, S., Archuleta, J., Bian, X., Chow, J., Coulter, R. L., de Wekker, S. F. J., Edgerton, S., Elliott, S., Fernandez, A., Fast, J. D., Hubbe, J. M., King, C., Langley, D., Leach, J., Lee, J. T., Martin, T. J., Martinez, D., Martinez, J. L., Mercado, G., Mora, V., Mulhearn, M., Pena, J. L., Petty, R., Porch, W., Russell, C., Salas, R., Shannon, J. D., Shaw, W. J., Sosa, G., Tellier, L., Templeman, B., Watson, J. G., White, R., Whiteman, C. D., and Wolfe, D.: The IMADA-AVER boundary layer experiment in the Mexico City area, B. Am. Meteor. Soc., 79, 2497–2508, 1998.

Doran, J. C., Barnard, J. C., Arnott, W. P., Cary, R., Coulter, R., Fast, J. D., Kassianov, E. I., Kleinman, L., Laulainen, N. S., Martin, T., Paredes-Miranda, G., Pekour, M. S., Shaw, W. J., Smith, D. F., Springston, S. R., and Yu, X. Y.: The T1-T2 study: evolution of aerosol properties downwind of Mexico City, Atmos. Chem. Phys., 7, 1585–1598, 2007, http://www.atmos-chem-phys.net/7/1585/2007/.

Doran, J. C., Fast, J. D., Barnard, J. C., Laskin, A., Desyaterik, Y., and Gilles, M. K.: Applications of Lagrangian dispersion modeling to the analysis of changes in the specific absorption of elemental carbon, Atmos. Chem. Phys., 8, 1377–1389, 2008,

http://www.atmos-chem-phys.net/8/1377/2008/.

Fast, J. D. and Darby, L. S.: An evaluation of mesoscale model predictions of down-valley and canyon flows and their consequences using Doppler lidar measurements during VTMX 2000, J. Appl. Meteor., 43, 420–436, 2004.

Fast, J. D. and Easter, R.: A Lagrangian Particle Dispersion Model Compatible with WRF, in: 7th WRF User's Workshop, Boulder, CO, USA, 2006.

Fast, J. D. and Zhong, S. Y.: Meteorological factors associated with inhomogeneous ozone concentrations within the Mexico City basin, J. Geophys. Res.-Atmos., 103, 18927–18946, 1998.

Fast, J. D., de Foy, B., Rosas, F. A., Caetano, E., Carmichael, G., Emmons, L., McKenna, D., Mena, M., Skamarock, W., Tie, X., Coulter, R. L., Barnard, J. C., Wiedinmyer, C., and Madronich, S.: A meteorological overview of the MILAGRO field campaigns, Atmos. Chem. Phys., 7, 2233–2257, 2007, http://www.atmos-chem-phys.net/7/2233/2007/.

Gilliam, R. C., Hogrefe, C., and Rao, S. T.: New methods for evaluating meteorological models used in air quality applications, Atmos. Environ., 40, 5073–5086, 2006.

Grell, G. A., Dudhia, J., and Stauffer, D. R.: A Description of the Fifth-Generation Penn State/NCAR Mesoscale Model (MM5), Tech. Rep. NCAR/TN-398+STR, NCAR, 1995.

Grossman-Clarke, S., Liu, Y. B., Zehnder, J. A., and Fast, J. D.: Simulations of the urban planetary boundary layer in an arid metropolitan area, J. Appl. Meteorol. Climatol., 47, 752–768, 2008.

Grutter, M., Basaldud, R., Rivera, C., Harig, R., Junkerman, W., Caetano, E., and Delgado-Granados, H.: SO2 emissions from Popocatepetl volcano: emission rates and plume imaging using optical remote sensing techniques, Atmos. Chem. Phys., 8, 6655–6663, 2008, http://www.atmos-chem-phys.net/8/6655/2008/.

Hong, S. Y. and Pan, H. L.: Nonlocal boundary layer vertical diffusion in a Medium-Range Forecast Model, Mon. Weather Rev., 124, 2322–2339, 1996.

Hong, S. Y., Noh, Y., and Dudhia, J.: A new vertical diffusion package with an explicit treatment of entrainment processes, Mon. Weather Rev., 134, 2318–2341, 2006.

Jazcilevich, A. D., Garcia, A. R., and Caetano, E.: Locally induced surface air confluence by complex terrain and its effects on air pollution in the valley of Mexico, Atmos. Environ., 39, 5481–5489, 2005.

Johnson, K. S., de Foy, B., Zuberi, B., Molina, L. T., Molina, M. J., Xie, Y., Laskin, A., and Shutthanandan, V.: Aerosol composition and source apportionment in the Mexico City Metropolitan Area with PIXE/PESA/STIM and multivariate analysis, Atmos. Chem. Phys., 6, 4591–4600, 2006, http://www.atmos-chem-phys.net/6/4591/2006/.

Kain, J. S.: The Kain-Fritsch convective parameterization: An update, J. Appl. Meteorol., 43, 170–181, 2004.

Kleinman, L. I., Springston, S. R., Daum, P. H., Lee, Y. N., Nunnermacker, L. J., Senum, G. I., Wang, J., Weinstein-Lloyd, J., Alexander, M. L., Hubbe, J., Ortega, J., Canagaratna, M. R., and Jayne, J.: The time evolution of aerosol composition over the Mexico City plateau, Atmos. Chem. Phys., 8, 1559–1575, 2008, http://www.atmos-chem-phys.net/8/1559/2008/.

Lei, W., de Foy, B., Zavala, M., Volkamer, R., and Molina, L. T.: Characterizing ozone production in the Mexico City Metropoli-

tan Area: a case study using a chemical transport model, Atmos. Chem. Phys., 7, 1347–1366, 2007, http://www.atmos-chem-phys.net/7/1347/2007/.

Mass, C. F., Ovens, D., Westrick, K., and Colle, B. A.: Does increasing horizontal resolution produce more skillful forecasts? The results of two years of real-time numerical weather prediction over the Pacific northwest, B. Am. Meteor. Soc., 83, 407–430, 2002.

Moffet, R. C., de Foy, B., Molina, L. T., Molina, M. J., and Prather, K. A.: Measurement of ambient aerosols in northern Mexico City by single particle mass spectrometry, Atmos. Chem. Phys., 8, 4499–4516, 2008, http://www.atmos-chem-phys.net/8/4499/2008/.

Molina, L. T. and Molina, M. J. (eds.): Air Quality in the Mexico Megacity: An Integrated Assessment, Kluwer Academic Publishers, 2002.

Molina, L. T., Kolb, C. E., de Foy, B., Lamb, B. K., Brune, W. H., Jimenez, J. L., Ramos-Villegas, R., Sarmiento, J., Paramo-Figueroa, V. H., Cardenas, B., Gutierrez-Avedoy, V., and Molina, M. J.: Air quality in North America's most populous city – overview of the MCMA-2003 campaign, Atmos. Chem. Phys., 7, 2447–2473, 2007, http://www.atmos-chem-phys.net/7/2447/2007/.

Nachamkin, J. E., Cook, J., Frost, M., Martinez, D., and Sprung, G.: Evaluation of dispersion forecasts driven by atmospheric model output at coarse and fine resolution, J. Appl. Meteorol. Climatol., 46, 1967–1980, 2007.

Oreskes, N.: Evaluation (not validation) of quantitative models, Environ. Health Persp., 106, 1453–1460, 1998.

Otte, T. L.: The impact of nudging in the meteorological model for retrospective air quality simulations. Part I: Evaluation against national observation networks, J. Appl. Meteorol. Climatol., 47, 1853–1867, 2008.

Pielke, R. A. and Uliasz, M.: Use of meteorological models as input to regional and mesoscale air quality models – Limitations and strengths, Atmos. Environ., 32, 1455–1466, 1998.

Raga, G. B., Baumgardner, D., Castro, T., Martinez-Arroyo, A., and Navarro-Gonzalez, R.: Mexico City air quality: A qualitative review of gas and aerosol measurements (1960–2000), Atmos. Environ., 35, 4041–4058, 2001.

Rife, D. L., Davis, C. A., Liu, Y. B., and Warner, T. T.: Predictability of low-level winds by mesoscale meteorological models, Mon. Weather Rev., 132, 2553–2569, 2004.

Rutter, A. P., Snyder, D. C., Stone, E. A., Schauer, J. J., Gonzalez-Abraham, R., Molina, L. T., Márquez, C., Cárdenas, B., and de Foy, B.: In situ measurements of speciated atmospheric mercury and the identification of source regions in the Mexico City Metropolitan Area, Atmos. Chem. Phys., 9, 207–220, 2009, http://www.atmos-chem-phys.net/9/207/2009/.

Seaman, N. L.: Meteorological modeling for air-quality assessments, Atmos. Environ., 34, 2231–2259, 2000.

Seibert, P., Kromp-Kolb, H., Baltensperger, U., Jost, D. T., and Schwikowski, M.: Trajectory analysis of high-alpine air pollution data, in: Air Pollution Modelling and its Application X, edited by: Gryning, S.-E. and Millan, M. M., Plenum Press, New York, USA, 595–596, 1994.

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Wang, W., and Powers, J. G.: A Description of the Advanced Research WRF Version 2, Tech. Rep. NCAR/TN-468+STR, NCAR, 2005.

Stein, A. F. and Wyngaard, J. C.: Fluid modeling and the evaluation of inherent uncertainty, J. Appl. Meteorol., 40, 1769–1774, 2001.

Stohl, A., Forster, C., Frank, A., Seibert, P., and Wotawa, G.: Technical note: The Lagrangian particle dispersion model FLEX-PART version 6.2, Atmos. Chem. Phys., 5, 2461–2474, 2005, http://www.atmos-chem-phys.net/5/2461/2005/.

Streit, G. E. and Guzman, F.: Mexico City air quality: Progress of an international collaborative project to define air quality management options, Atmos. Environ., 30, 723–733, 1996.

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

Thomsen, G. L. and Smith, R. K.: The importance of the boundary layer parameterization in the prediction of low-level convergence lines, Mon. Weather Rev., 136, 2173–2185, 2008.

Thornhill, D. A., de Foy, B., Herndon, S. C., Onasch, T. B., Wood, E. C., Zavala, M., Molina, L. T., Gaffney, J. S., Marley, N. A., and Marr, L. C.: Spatial and temporal variability of particulate polycyclic aromatic hydrocarbons in Mexico City, Atmos. Chem. Phys., 8, 3093–3105, 2008, http://www.atmos-chem-phys.net/8/3093/2008/.

Warner, S., Platt, N., Urban, J. T., and Heagy, J. F.: Comparisons of transport and dispersion model predictions of the joint urban 2003 field experiment, J. Appl. Meteorol. Climatol., 47, 1910–1928, 2008.

Zängl, G., Chimani, B., and Häberli, C.: Numerical Simulations of the Foehn in the Rhine Valley on 24 October 1999 (MAP IOP 10), Mon. Weather Rev., 132, 368–389, 2004.

Zavala, M., Herndon, S. C., Slott, R. S., Dunlea, E. J., Marr, L. C., Shorter, J. H., Zahniser, M., Knighton, W. B., Rogers, T. M., Kolb, C. E., Molina, L. T., and Molina, M. J.: Characterization of on-road vehicle emissions in the Mexico City Metropolitan Area using a mobile laboratory in chase and fleet average measurement modes during the MCMA-2003 field campaign, Atmos. Chem. Phys., 6, 5129–5142, 2006, http://www.atmos-chem-phys.net/6/5129/2006/.