



## Turbulent transport, emissions and the role of compensating errors in chemical transport models

P. A. Makar<sup>1</sup>, R. Nissen<sup>2</sup>, A. Teakles<sup>2</sup>, J. Zhang<sup>1</sup>, Q. Zheng<sup>1</sup>, M. D. Moran<sup>1</sup>, H. Yau<sup>2</sup>, and C. diCenzo<sup>2,\*</sup>

<sup>1</sup>Air Quality Research Division, Environmental Science and Technology Directorate, Environment Canada, Toronto, Ontario M3H 5T4, Canada

<sup>2</sup>Air Quality Science Unit, Pacific and Yukon Region, Environment Canada, Vancouver, B.C., Canada

\*retired

Correspondence to: P. A. Makar (paul.makar@ec.gc.ca)

Received: 22 September 2013 – Published in Geosci. Model Dev. Discuss.: 7 November 2013

Revised: 5 March 2014 – Accepted: 21 March 2014 – Published: 27 May 2014

**Abstract.** The balance between turbulent transport and emissions is a key issue in understanding the formation of O<sub>3</sub> and particulate matter with diameters less than 2.5 μm (PM<sub>2.5</sub>). Discrepancies between observed and simulated concentrations for these species have, in the past, been ascribed to insufficient turbulent mixing, particularly for atmospherically stable environments. This assumption may be simplistic – turbulent mixing deficiencies may explain only part of these discrepancies, and as turbulence parameterizations are improved, the timing of primary PM<sub>2.5</sub> emissions may play a much more significant role in the further reduction of model error. In a study of these issues, two regional air-quality models, the Community Multi-scale Air Quality model (CMAQ, version 4.6) and A Unified Regional Air-quality Modelling System (AURAMS, version 1.4.2), were compared to observations for a domain in north-western North America. The air-quality models made use of the same emissions inventory, emissions processing system, meteorological driving model, and model domain, map projection and horizontal grid, eliminating these factors as potential sources of discrepancies between model predictions. The initial statistical comparison between the models and monitoring network data showed that AURAMS' O<sub>3</sub> simulations outperformed those of this version of CMAQ4.6, while CMAQ4.6 outperformed AURAMS for most PM<sub>2.5</sub> statistical measures. A process analysis of the models revealed that many of the differences between the models' results could be attributed to the strength of turbulent diffusion, via the choice of an a priori lower limit in the magnitude of vertical diffusion coefficients, with AURAMS using 0.1 m<sup>2</sup> s<sup>-1</sup> and CMAQ4.6 using 1.0 m<sup>2</sup> s<sup>-1</sup>.

The use of the larger CMAQ4.6 value for the lower limit of vertical diffusivity within AURAMS resulted in a similar performance for the two models (with AURAMS also showing improved PM<sub>2.5</sub>, yet degraded O<sub>3</sub>, and a similar time series as CMAQ4.6). The differences between model results were most noticeable at night, when the higher minimum turbulent diffusivity resulted in an erroneous secondary peak in predicted night-time O<sub>3</sub>. A spatially invariant and relatively high lower limit in diffusivity could not reduce errors in both O<sub>3</sub> and PM<sub>2.5</sub> fields, implying that other factors aside from the strength of turbulence might be responsible for the PM<sub>2.5</sub> over-predictions. Further investigation showed that the magnitude, timing and spatial allocation of area source emissions could result in improvements to PM<sub>2.5</sub> performance with minimal O<sub>3</sub> performance degradation. AURAMS was then used to investigate a land-use-dependant lower limit in diffusivity of 1.0 m<sup>2</sup> s<sup>-1</sup> in urban regions, linearly scaling to 0.01 m<sup>2</sup> s<sup>-1</sup> in rural areas, as employed in CMAQ5.0.1. This strategy was found to significantly improve mean statistics for PM<sub>2.5</sub> throughout the day and mean O<sub>3</sub> statistics at night, while significantly degrading (halving) midday PM<sub>2.5</sub> correlation coefficients and slope of observed to model simulations. Time series of domain-wide model error statistics aggregated by local hour were shown to be a useful tool for performance analysis, with significant variations in performance occurring at different hours of the day. The use of the land-use-dependant lower limit in diffusivity was also shown to reduce the model's sensitivity to the temporal allocation of its emissions inputs. The modelling scenarios suggest that while turbulence plays a key role in O<sub>3</sub> and PM<sub>2.5</sub> formation

in urban regions, and in their downwind transport, the spatial and temporal allocation of primary PM<sub>2.5</sub> emissions also has a potentially significant impact on PM<sub>2.5</sub> concentration levels. The results show the complex nature of the interactions between turbulence and emissions, and the potential of the strength of the former to mask the impact of changes in the latter.

## 1 Introduction

Several studies within the last decade have shown the value of the comparison of multiple air-quality models to a common suite of observations. Two studies made use of data collected during the International Consortium for Atmospheric Research on Transport and Transformation/New England Air-Quality Study (ICARTT/NEAQS). McKeen et al. (2005) used seven air-quality forecast models to show that ensemble O<sub>3</sub> forecasts based on the seven-member mean and the seven-member median had better temporal correlation to the observed daily maximum 1 h average and maximum 8 h average than any individual model, for a domain covering the eastern USA and south-eastern Canada. The usefulness of uncorrected ensembles was shown to be limited by positive biases in O<sub>3</sub> inherent to all seven ensemble members. The best method of bias correction was found to be model dependent. In a subsequent examination for the same region and period using the same models, McKeen et al. (2007) found that particulate matter with diameters less than 2.5 µm (PM<sub>2.5</sub>) forecasts had similar correlation, lower bias and better skill compared to the ozone forecasts. A feature of this work was an analysis of diurnal variability – most models failed to reproduce the observed diurnal cycle of PM<sub>2.5</sub> concentrations at urban and suburban monitor locations. This error in the predicted diurnal cycle was most pronounced in the transition period between night and early morning. Four of the models showed greater diurnal PM<sub>2.5</sub> variability than was observed, with differences in emissions inventories, the Planetary Boundary Layer height (PBL) parameterizations employed and the timing of the predicted morning growth of the PBL all postulated as factors affecting the model performance. The work also identified insufficient model nocturnal mixing as a key factor in low surface sulfate predictions (due to insufficient vertical turbulent transport of sulfate aloft to the surface) and in excessively high predicted surface elemental carbon and NO<sub>x</sub> predictions (due to insufficient turbulent transport of these species emitted at the surface to higher model levels). These effects were noted for the on-line Weather Research and Forecasting – Chemistry (WRF-CHEM) model, which *de facto* makes use of the turbulence parameterizations inherent in the driving meteorology. Subsequent model ensemble work for the second Texas Air Quality Study (TexAQS II; McKeen et al., 2009) showed that the relationship between model emissions levels and concentration difference ratios was approximately

linear (to within 25 %), with improvements to emissions inventories through the use of continuous emissions monitors and updated mobile emissions resulting in better agreement with observations. The study also noted that despite ratios of PM<sub>2.5</sub> to NO<sub>y</sub> matching observations, underpredictions of PM<sub>2.5</sub> organic carbon suggested that this might be a result of compensating errors, with excessive model primary PM<sub>2.5</sub> making up for the absence of sufficient model secondary organic aerosol formation.

Multiple model intercomparisons were expanded to include both North American and European domains in the Air-Quality Model Evaluation International Initiative (AQMEII; Galmarini and Rao, 2011), with 23 modelling groups providing annual simulations on either or both of these domains. In a further refinement of ensemble forecasting techniques, researchers participating in this study found that full ensemble mean predictions could be outperformed by subset ensembles of model members selected for an optimal set of error characteristics (Solazzo et al., 2012a). Predictions of PM were also investigated; Solazzo et al. (2012b) showed that all of the models employed underestimated PM<sub>10</sub>, with better estimates for PM<sub>2.5</sub>, though “no model was found to consistently match the observations for all locations throughout the entire year,” (p. 76) – in the abstract, with North American correlation coefficients for PM<sub>2.5</sub> ranging from 0.34 at a 99 % confidence level to less than 0.10 at a 10 % confidence level. While anthropogenic emissions were prescribed as part of the intercomparison, differences in natural emissions of some PM components such as sea salt were shown to have a significant impact on some model results. The member of the ensemble which made use of a different emissions inventory from that prescribed in the study protocol was shown to have significantly different (factor of four lower) PM<sub>10</sub> emissions than the other models, showing the potential importance of emissions inventory accuracy on PM<sub>10</sub> predictions. Large differences in particulate deposition rates despite similar theoretical approaches to deposition were attributed to differences in the characterization of surface properties and near-surface meteorology, with the fractional bias of the PM<sub>10</sub> seasonal concentration varying by up to 60 % depending on which deposition module was used within a single model (Nopmongkol et al., 2012). Models with the highest deposition rates of PM<sub>2.5</sub> were also found to have the most significant negative biases in PM<sub>2.5</sub> concentrations. Model performance for PM<sub>10</sub> was better in the summer months than in winter, with difficulties in the accurate simulation of very stable boundary layers in the winter being a possible cause of model prediction errors. Most models underestimated the amplitude of the diurnal cycle of PM<sub>10</sub> as well as being biased low. The inorganic PM<sub>2.5</sub> components were better simulated than the organics, demonstrating the ongoing problem with accurate simulations of organic aerosol. While the study did not examine or compare the models’ individual chemical process parameterizations in detail, a conclusion of the work was that the details of those

parameterizations play a pivotal role in model performance, despite similarities in the overall schemes employed.

A two-model intercomparison attempted to eliminate some of the sources of model prediction variability by prescribing additional model inputs aside from the meteorology. Smyth et al. (2009) used the same emissions inventory, emissions processing system, meteorological driver, North American domain and map projection to eliminate these factors as sources of possible differences between the two models compared (the Community Multiscale Air-Quality model; CMAQ4.6, and A Unified Regional Air-quality Modelling System; AURAMS1.4.2). Despite these similarities, some significant differences in model performance were noted. AURAMS had a normalized mean bias (NMB) for hourly  $O_3$  that was less than half of that for CMAQ (21 % versus 46 %, respectively), while both models had similar normalized mean errors (NME, 47 versus 54 %). The larger NMB errors for CMAQ were shown to be related to its inability to predict the observed night-time  $O_3$  minima. Both models'  $PM_{2.5}$  predictions were biased low (AURAMS:  $-10$  % and CMAQ  $-65$  %, respectively), though both had similar NME  $PM_{2.5}$  scores, with much of the reduced  $PM_{2.5}$  bias in AURAMS being the result of high sea salt predictions in this model. Both models underpredicted the organic fraction of  $PM_{2.5}$ . The study noted the potential difficulties in the systematic assessment of individual chemical and physical processes on the model results, due to the complexity and interconnected nature of those processes.

The above body of work demonstrates both the value of model evaluation and intercomparisons and the corresponding difficulties. Conducting multi-model studies requires a considerable investment in the preparation and evaluation of model fields. A common finding of studies such as those described above is that differences in the performance between the different models lie in their process parameterizations, yet process-level studies comprise an additional level of complexity, and are consequently not always part of large-scale multi-model comparisons with observations. However, when significant differences are found between models employing harmonized input fields, process-level evaluations may provide valuable information on the reasons underscoring model performance. In the work that follows, we describe a process-level comparison of the CMAQ and AURAMS models on a more limited regional domain. The emissions inventories, emissions processing system, model domain, map projection and the driving meteorology were held in common for the two models, allowing two key factors in model performance to be identified: the accuracy of inputs used to create model emissions, and the models' parameterizations and assumptions regarding turbulent diffusion.

## 2 Model description

A detailed description of the two models may be found in Smyth et al. (2009), and updates to the AURAMS model (Gong et al., 2006) subsequent to that time may be found in Kelly et al. (2012). CMAQ v4.6 is also described in Pleim et al. (2006). Here, we note some of the main features of the two models and the framework used for comparison, with reference to Table 1.

Both models made use of meteorology from the Global Environmental Multiscale weather forecast model (GEM, v3.2.2; Côté et al., 1998); GEM simulations were carried out on a rotated latitude/longitude grid, for the period 15 July through 15 August 2005, in a series of overlapping 30 h simulations for a North American domain, with 0.1375 degree horizontal grid spacing (approximately 15 km), starting from model analysis files at 16:00 local standard time (00:00 UTC) on each day. The first six hours of each of these simulations were discarded as “spin-up” in order to allow the model's cloud variables to reach a steady state. The remaining hours (6 through 30) were retained for use as a continuous sequence of air-quality model meteorological input. Two days of spin-up time was employed in the meteorological models – this being sufficient for air parcels starting at the upwind boundary to cross the downwind boundary of this relatively small simulation domain.

The meteorological files were interpolated to the 12 km grid spacing air-quality model domain (Fig. 1a, inset white region, and Fig. 1b, which also shows the air-quality model grid along with observation network stations). The domain encompasses the coastal north-western USA and coastal south-western Canada. Several unique features of this domain should be noted, as described in more detail in previous work by Steyn et al. (2013) and Ainslie et al. (2013): (1) unlike many locations in North America, the upwind boundary condition of the domain consists of relatively “clean” air associated with trans-Pacific transport; (2) the terrain is mountainous – previous work (Brook et al., 2004) suggests recirculation events in which aged air carried aloft with upslope flow is returned to the surface over the ocean, allowing accumulation of pollutants; and (3) the boundaries between  $NO_x$  and volatile organic compound sensitivity in the region have been changing over time, indicating that the region contains markedly different chemical regimes, depending on location (Ainslie et al., 2013).

While the models employ the same horizontal domain map projection and grid, they differ in their vertical coordinate and number of levels. AURAMS uses a Gal-Chen coordinate system with 27 layers and a model top at 30 km, while CMAQ uses a sigma coordinate system with 15 layers and a model top at approximately 15 km. The thickness of the model layers differs, but tests in which the AURAMS layer thicknesses were imported into CMAQ had a negligible impact on CMAQ performance. Both models make use of their default boundary conditions; for AURAMS1.4.2, these vary

**Table 1.** Comparison of the main features of the CMAQ and AURAMS models.

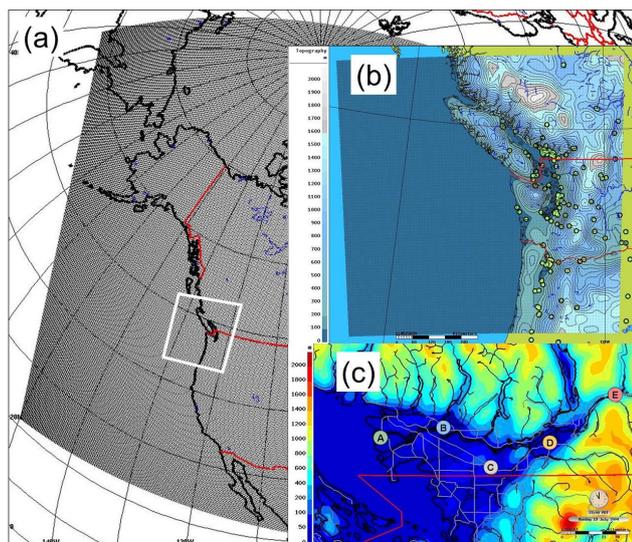
Model	AURAMS	CMAQ
Version	1.4.2	4.6
Horizontal projection	Polar stereographic true at 60° N, 93X93 grid points; 12 km grid spacing	
Emissions inventory	Anthropogenics: 2006 Canadian; 2005 US; processed using the Sparse Matrix Operating Kernel Emissions processing system Biogenics: BEIS3.0.9, processed using model-predicted temperatures and Photosynthetically Active Radiation (PAR) values.	
Particle size distribution	Sectional approach, 12 size bins	Modal approach, three modes
Vertical diffusion	Laasonen numerics, diffusion coefficients from GEM, internal eddy diffusivity minimum of 0.1 m <sup>2</sup> s <sup>-1</sup>	Internal calculation of eddy diffusivity, with internal minimum of 1.0 m <sup>2</sup> s <sup>-1</sup>
Number of vertical levels	27	15
Plume rise	Calculated on-line	Pre-calculated in SMOKE
Dry deposition	Gases: resistance parameterization based on Wesley (Zhang et al., 2002) Particles: (Zhang et al., 2001)	Resistance scheme linked to the Pleim-Xiu Land Surface Model (PX LSM) (Pleim et al., 2001; Pleim and Ran, 2011)
Driving meteorology	Global Environmental Multiscale (GEM) model, version 3.2.2, overlapping 30 h simulations starting at 0Z, initial 6 h spin-up discarded; final 24 h used for air-quality model input	
Simulation period	15 July–15 August 2005	

with season for some species, and for O<sub>3</sub> an adjustment of climatological boundary conditions in response to the local tropopause height is employed (Makar et al., 2010). While the upwind boundary conditions of the models differ, they both describe relatively clean conditions, as is appropriate for the upwind condition of the domain. A comparison of the PM<sub>2.5</sub> and O<sub>3</sub> predictions of the models over the Pacific (upwind boundary condition) relative to the urban regions shows that the changes associated with urban local chemistry and dynamics are an order of magnitude greater than the variations that may be observed in the upwind boundary region of the model. The effects described below are thus the result of local changes in the models' respective responses to the emissions, rather than to upwind boundary conditions.

Gas-phase dry deposition velocities for CMAQ4.6 are computed using an electrical resistance analog model (M3Dry, version 1.8; Pleim et al., 2001) in a pre-processing step in CMAQ's Meteorology–Chemistry Interface Processor (MCIP) (Otte and Pleim, 2010), using the MCIP-based stomatal conductance parameterization. The use of this parameterization may degrade results relative to the use of WRF's stomatal parameterization (Otte and Pleim, 2010). A modified version of the MCIP preprocessor (Smyth et al., 2005, 2006a) was used to convert GEM meteorological

files for CMAQ4.6 input. AURAMS gas-phase dry deposition also uses an electrical resistance analog, and is calculated according to Zhang et al. (2002). CMAQ uses the particle dry deposition scheme of Giorgi (1986); AURAMS uses the scheme of Zhang et al. (2001), which is also based on Giorgi (1986).

CMAQ4.6 incorporates particle deposition as a boundary condition on vertical diffusion (Binkowski and Roselle, 2003; Binkowski and Shankar, 1995), with the species mass in each of three modes being deposited separately. As noted in Binkowski and Roselle (2003), “the impaction term is omitted for coarse mode particles in the moment dry deposition velocities” (page AAC 3-6), though this has been corrected in more recent CMAQ implementations (J. E. Pleim, personal communication, 2014). CMAQ does not incorporate particle settling between particle layers, though a settling term is included in the lowest layer deposition velocity calculation (J. E. Pleim, personal communication, 2014). AURAMS1.4.2 calculates particle settling for each of the 12 size bins of the model. In our AURAMS1 (base case) simulations the settling and deposition velocities are used to calculate mass fluxes between layers, with the settling (or deposition) velocity being used to determine the destination layer of the falling particles. In our remaining AURAMS simulations, a



**Figure 1.** (a) GEM 15 km domain with boundary of CMAQ and AURAMS 12 km domain shown as inset; (b) 12 km Pacific and Yukon Region Domain, observation stations shown as green dots, background contours elevation; (c) four stations (out of 20 total) in the Lower Fraser Valley, elevation contours. A = Vancouver International Airport; B = Pitt Meadows; C = Abbotsford Airport; D = Chilliwack; E = Hope Airport.

1-dimensional semi-Lagrangian advection approach is taken, with the settling and deposition velocities being used to determine the mass transport and new vertical distribution of the particle mass. The latter modification had a minor impact on model results.

Both models made use of the same land-use parameters provided by the GEM model, and both models made use of the same emissions inventories. Environment Canada 2006 and US Environmental Protection Agency (EPA) 2005 anthropogenic inventories were combined for this work, and both models made use of the Biogenic Emissions Inventory System (BEIS3.0.9) biogenic emissions algorithms and Biogenic Emissions Landuse Database version 3 (BELD3) land-use data (US EPA, 2007). The gas-phase chemical mechanism employed in AURAMS is the Acid Deposition and Oxidant Model, version 2 (ADOM-II) mechanism (Stockwell and Lurmann, 1989), while CMAQ4.6 made use of the 1999 version of the Statewide Air Pollution Research Center (SAPRC-99) mechanism (Carter, 2000a, b). The models have a similar particulate-matter chemical speciation; however, the particle size distribution in AURAMS makes use of a 12-bin sectional approach while CMAQ uses a three-mode modal approach. The emissions in both models thus had to be speciated for that model's chemical mechanism and particle size distribution.

Emissions for both models were generated using the Sparse Matrix Operator Kernel Emissions processing system (SMOKE; Houyoux et al., 2000; CEP, 2003). Emissions

processing systems such as SMOKE make use of input emissions inventories which usually comprise annual emissions totals for different sources over a geopolitical region such as state/province/county/municipality. These annual values are distributed within the geopolitical region using spatial disaggregation data – gridded maps of the expected spatial distribution of pollutants, derived from surrogate fields believed to reflect the distribution of the emitting activities. The emissions are also required by the models on an hourly basis; hence the annual emissions must also be distributed over time. Temporal allocations are required to split the annual emissions into month-of-year, day-of-week within each month and hour-of-day within each day. The accuracy of the gridded emissions used as model input will depend on the extent to which these assigned spatial and temporal fields accurately reflect the true temporal and spatial distributions, as well as on the annual total geopolitically distributed emissions. Unfortunately, the available spatial surrogate fields are severely outnumbered by the number of emitting activities, with thousands of source types typically being represented by a few hundred surrogates (here, a total of 170 surrogates were used). Similarly, the temporal profiles used for emitting activity are often best-guess approximations which are not based on observed monthly/day-of-week/hour-of-day emissions for any given emitting activity to which they are assigned. The assignments for spatial and temporal disaggregation of annual emissions are of crucial importance in determining the resulting model accuracy for circumstances when the spatial and temporal distribution of emissions have a significant impact on local concentrations (i.e. close to the sources as opposed to further downwind). The impact of the choice of spatial and temporal disaggregation data is examined in several scenario simulations in Sect. 4.2.

The models differ in the approach taken for vertical diffusion. AURAMS uses diffusion coefficients for heat and moisture from the driving meteorological model along with a fully implicit Laasonen approach for the discretization of the diffusion equation (cf. Richtmyer, 1994). CMAQ4.6 calculates diffusion coefficients based on the driving meteorological model's values for the temperature, wind speed, total liquid water content, specific humidity, surface pressure, friction velocity and height of the boundary layer (Pleim, 2007a, b). The underlying boundary layer model (Asymmetric Convective Model, version 2, ACM2, Xiu and Pleim, 2001) includes a nonlocal transport component for unstable conditions (J. E. Pleim, personal communication, 2014). Numerical solution of the diffusion equation is carried out in CMAQ using the Crank–Nicolson discretization (cf. Richtmyer, 1994). AURAMS also includes a Crank–Nicolson algorithm option – its use did not significantly affect the AURAMS results. Both models subsequently employ a lower limit to their diffusion coefficients, with this “floor” in diffusion in AURAMS in our initial simulations being set to  $0.1 \text{ m}^2 \text{ s}^{-1}$ , and to  $1.0 \text{ m}^2 \text{ s}^{-1}$  in CMAQ4.6. The choice of a specific lower limit has a significant impact on

**Table 2.** Description of model scenarios.

Scenarios	Description
CMAQ1	Base case CMAQ
AURAMS1	Base case AURAMS
CMAQ2	First level of emissions upgrades applied to CMAQ
AURAMS1b	AURAMS code improvements applied to AURAMS, no emissions changes
AURAMS2	As in AURAMS 1b, with the first level of emissions upgrades
AURAMS3	AURAMS 2 + second level of emissions upgrades
AURAMS4	AURAMS 3 + third level of emissions upgrades
AURAMS5	AURAMS4 + use of diffusivity cut-off of $0.6 \text{ m}^2 \text{ s}^{-1}$
AURAMS6	AURAMS4 + renormalization of non-mobile area source emissions
AURAMS7	AURAMS4 + use of CMAQ5.0.1 linear interpolated diffusivity lower limit ( $1.0 \text{ m}^2 \text{ s}^{-1}$ urban, $0.01 \text{ m}^2 \text{ s}^{-1}$ rural)
AURAMS8	AURAMS4 + renormalization of non-mobile area source emissions and CMAQ5.0.1 linear interpolated diffusivity lower limit

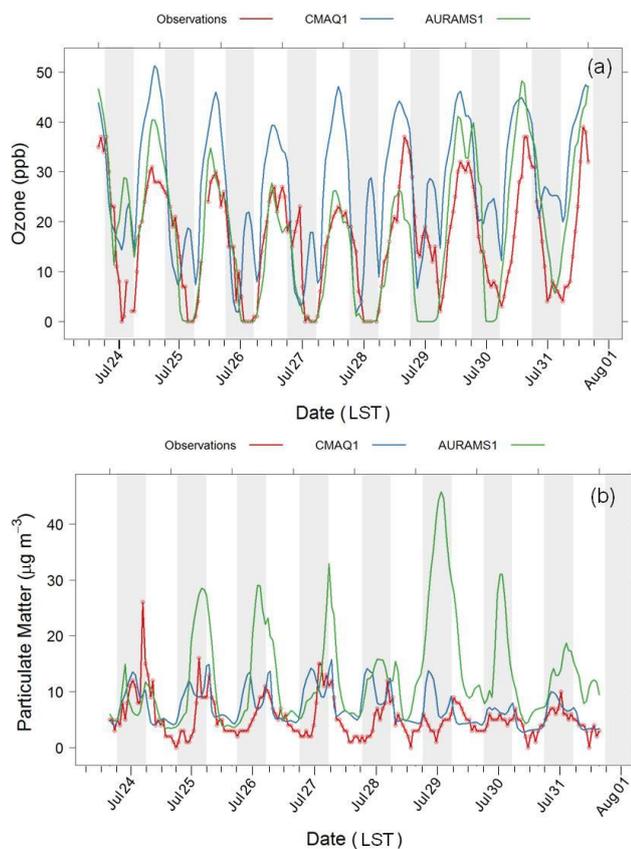
model performance (cf. Community Modelling and Analysis (CMAS), [http://www.cmascenter.org/cmaq/documentation/4.6/ADVECTION\\_DIFFUSION.txt](http://www.cmascenter.org/cmaq/documentation/4.6/ADVECTION_DIFFUSION.txt), 2006). Other options available for the use of this version of CMAQ include using higher values of the diffusion coefficient lower limit over urban areas ( $2.0 \text{ m}^2 \text{ s}^{-1}$ ) and lower values over rural areas ( $0.5 \text{ m}^2 \text{ s}^{-1}$ ). More recent versions of CMAQ (5.0.1) use a linear function of urban land-use area fraction for the lower limit of eddy diffusivity, with  $0.01 \text{ m}^2 \text{ s}^{-1}$  employed for entirely rural areas and  $1.0 \text{ m}^2 \text{ s}^{-1}$  for entirely urban areas. These recent changes to CMAQ were implemented in recognition of the fact that the rural minimum should not exceed that of the driving meteorological model, whereas the use of a higher urban minimum may be necessary if the driving meteorological model is not capable of accounting for the turbulence-enhancing effects of the urban environment. Here, the impact of these strategies was investigated in a set of scenario simulations.

The model results were evaluated using hourly  $\text{O}_3$  and  $\text{PM}_{2.5}$  data from four monitoring networks (Air Quality System, AQS; Canadian Air and Precipitation Monitoring Network, CAPMoN; Clean Air Status and Trends Network, CASTNET; and National Air Pollution Surveillance program, NAPS). Model values are hourly averages in the case of CMAQ, while AURAMS output is hour-ending averages of 15 min output. Station locations are shown in Fig. 1b, with five stations in the Lower Fraser Valley in Fig. 1c. The Lower Fraser Valley contains a large proportion of the population of the Canadian province of British Columbia; portions of our

analysis examine model performance in this sub-region in detail. AURAMS output was available on a 15 min timestep, while CMAQ output was hourly averages; the AURAMS values were averaged to create hourly values for comparison to the observations. An analysis package using the R programming language (R Development Core Team, 2010) was created for model evaluation making use of the “open-air” R package (Carslaw and Ropkins, 2011). The output package of AURAMS 1.4.2 includes output at station locations during model run time, while CMAQ output was derived from output netCDF files using the work of Pierce (2010). Visualization packages utilized in creating the graphical display of analysed fields included hexbin (Carr et al., 2010) and Lattice (Deepayan, 2008).

### 3 Model simulations

Eleven model simulations were carried out, in order to evaluate the impact of improvements to model algorithms, improvements and sensitivity to emissions inputs, and the impact of changes to the value of the lower limit for eddy diffusivity (Table 2). The first two of these are unmodified CMAQ4.6 and AURAMS1.4.2 simulations; the “base case” scenarios (CMAQ1 and AURAMS1). As will be noted below, these scenarios showed a marked difference between the models with regards to their performance for  $\text{PM}_{2.5}$  and  $\text{O}_3$ . The base case scenarios are followed by several process and emissions input related scenarios: AURAMS1b – a



**Figure 2.** Comparison between observations, CMAQ and AURAMS for (a)  $O_3$  and (b)  $PM_{2.5}$  at Vancouver Airport (station A in Fig. 1). Nights (6 p.m. to 6 a.m., local standard time (Pacific Standard Time)) shown as shaded regions.

scenario in which several process improvements were added to the AURAMS model and evaluated as a package; CMAQ2 and AURAMS2, in which the impact of improved emissions data were evaluated using both models; and six subsequent AURAMS simulations (AURAMS3 through AURAMS8), which investigated the AURAMS model sensitivity to further emissions changes and different strategies for the use of a lower limit in diffusivity than was used in the base case model. These scenarios and the rationale for their execution will be described below.

## 4 Results

### 4.1 Initial comparison and analysis

The statistical measures used in our analysis are presented in Table 3. The resulting analyses of the base case  $O_3$  and  $PM_{2.5}$  simulations from each model are summarized in the first two columns of Table 4. Table 4a, b shows the statistical scores for the entire grid, and Table 5 shows the  $PM_{2.5}$  scores for the five stations in the Lower Fraser Valley. The initial results

showed a substantial difference in model performance: AURAMS1.4.2 outperformed CMAQ4.6 for hourly ozone for the entire grid statistics (Table 4a) for all Canadian stations, aside from tying with CMAQ4.6 for correlation coefficient (not shown) and for the majority of the statistical metrics for the Lower Fraser Valley (not shown). This is in contrast to the earlier North American domain comparison by Smyth et al. (2009), where AURAMS outperformed CMAQ for  $O_3$  mean bias, normalized mean bias, mean error and normalized mean error, but CMAQ outperformed AURAMS for correlation coefficient. Previous work with CMAQ for simulations in the Lower Fraser Valley region for a 12-day period in August 2001 had significantly better  $O_3$  performance for normalized mean bias (NMB) and normalized mean error (NME) than found here (Smyth et al., 2006b: 13 % and 51 %, respectively, versus 75 % and 82 % in the current work). CMAQ simulations by Steyn et al. (2013) for the region for specific short episodes in 2006, 2001, 1995 and 1985 reported NME values ranging from 43 to 79 % (compare to 53 to 81 % in the different simulations of the current work) and NMB from  $-12$  to 64 % (compare to 31 to 75 %). Different meteorological drivers, emissions inventories and domains were used for these studies compared to the more recent work, and these may account for some of the differences in statistics, as well as the shorter time periods used in these earlier studies.

$PM_{2.5}$  scores in the current work were mixed, with CMAQ4.6 outperforming AURAMS1.4.2 across the grid (Table 4b) for minimum, y intercept, correlation coefficient, mean absolute error, mean squared error, root mean squared error and normalized mean error, and AURAMS outperforming CMAQ for mean, maximum, slope, mean bias and normalized mean bias. CMAQ outperformed AURAMS for  $PM_{2.5}$  at Canadian stations for all scores aside from maximum and slope (not shown), while the Lower Fraser Valley performance (Table 5) was mixed, with scores split between the models.

An examination of time series of  $O_3$  and  $PM_{2.5}$  at the Vancouver International Airport station (Fig. 2 depicts a portion of the total time series for clarity; the depicted model behaviour occurs throughout the simulation period) shows the marked differences between the models in comparison to observations, as well as providing a potential physical and chemical explanation for the differences. CMAQ4.6 tended to overpredict daytime  $O_3$  maxima and invariably created a night-time secondary maximum in  $O_3$  that is absent in the observations (Fig. 2a). AURAMS'  $O_3$  time series more closely followed observations than those of CMAQ, though night-time minima were sometimes lower in the model than in the observations. The relative performance of the models is clearly reversed for  $PM_{2.5}$  (Fig. 2b), with both models usually capturing the timing of the night-time peak  $PM_{2.5}$  levels, but AURAMS greatly overestimated their magnitude relative to CMAQ.

**Table 3.** Statistical measures of model performance.  $N$  is the number of paired observed-model values,  $\bar{O}$  is the mean observed value,  $\bar{M}$  is the mean model value.

Statistical Measure	Description	Formula
$R$	Pearson correlation coefficient	$R = \frac{N \sum_{i=1}^N (O_i \cdot M_i) - \sum_{i=1}^N (M_i) \sum_{i=1}^N (O_i)}{\sqrt{N \sum_{i=1}^N (M_i \cdot M_i) - \sum_{i=1}^N (M_i) \cdot \sum_{i=1}^N (M_i)} \sqrt{N \sum_{i=1}^N (O_i \cdot O_i) - \sum_{i=1}^N (O_i) \cdot \sum_{i=1}^N (O_i)}}$
$a$	Intercept of observations vs. model best-fit line	$a = \bar{M} - b \cdot \bar{O}$
$b$	Slope of observations vs. model best-fit line	$b = \frac{\sum_{i=1}^N [(O_i - \bar{O})(M_i - \bar{M})]}{\sum_{i=1}^N [(O_i - \bar{O})^2]}$
MB	Mean bias	$\text{MB} = \frac{1}{N} \sum_{i=1}^N (M_i - O_i)$
MAE	Mean absolute error	$\text{MAE} = \frac{1}{N} \sum_{i=1}^N  M_i - O_i $
MSE	Mean square error	$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (M_i - O_i)^2$
RMSE	Root mean square error	$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (M_i - O_i)^2}$
NMB	Normalized mean bias	$\text{NMB} = \frac{\sum_{i=1}^N (M_i - O_i)}{\sum_{i=1}^N O_i} \cdot 100$
NME	Normalized mean error	$\text{NME} = \frac{\sum_{i=1}^N  M_i - O_i }{\sum_{i=1}^N O_i} \cdot 100$

The timing of the two models' respective positive biases in ozone and particulate matter helps explain these results. Both the CMAQ secondary ozone maxima and the AURAMS PM<sub>2.5</sub> over-predictions occur *at night*. In an urban region at night, the dominant ozone chemical process is usually the destruction of ozone through titration by NO. The predicted surface concentrations of NO were higher in AURAMS than in CMAQ. The composition of PM<sub>2.5</sub> at night in an urban location can be expected to be dominated by the primary components of particulate matter, given that the oxidation processes that lead to secondary aerosol formation dominate during the day. This was confirmed via a check of the time series for AURAMS' speciated PM<sub>2.5</sub> for the same period as Fig. 2 (not shown): the primary PM<sub>2.5</sub> species dominated the PM<sub>2.5</sub> mass during these periods of high positive PM<sub>2.5</sub> bias. Given that emissions levels of primary PM<sub>2.5</sub> and NO were the same for both models, these results in turn implied that a difference in transport was the cause of the model differences.

While both models make use of the same wind fields, the two models differ significantly in their approach to vertical diffusion. Different numerical methods and lower limits for

diffusivity are used, as noted above. The diffusion coefficients used by the model prior to the application of the lower limits in diffusivity also differed: AURAMS1.4.2 makes use of the diffusion coefficients provided by the driving meteorological model GEM, while CMAQ4.6 recalculates diffusion coefficients internally using other fields from the driving meteorology. The diffusion coefficients generated by the CMAQ algorithm prior to the lower-limit truncation were found to produce values similar in magnitude to the GEM weather forecast model's values in other work (Kelly et al., 2012). The main remaining difference between the two base models was thus the magnitude of the assumed lower limit for the diffusivity coefficients and the strategy used for assigning those lower limits.

An AURAMS sensitivity test was conducted to determine the impact of the magnitude of the lower limit in diffusivity on the model results, with  $1.0 \text{ m}^2 \text{ s}^{-1}$  being used in AURAMS, for one selected day during the study period. The results of this test were dramatic and are shown in Fig. 3. The use of the higher diffusion coefficient cut-off halved the AURAMS NO<sub>x</sub> and PM<sub>2.5</sub> maxima, and resulted in higher night-time O<sub>3</sub> levels: the test confirmed that the main cause

**Table 4a.** O<sub>3</sub> statistics, entire grid (ppbv). Third and fourth columns: regular and bold fonts correspond to model with worse and better performance, respectively. Subsequent columns: regular font, italics and bold italics correspond to unchanged, worse and better performance, respectively, than the same model in the original comparison.

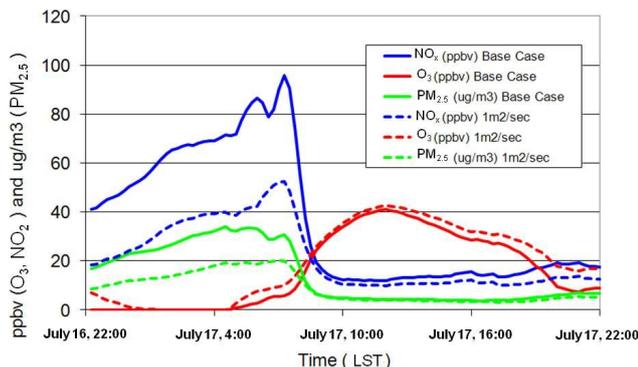
	CMAQ 1	AURAMS 1	AURAMS 1b	AURAMS 2	AURAMS 3	AURAMS 4	AURAMS 5	AURAMS 6	AURAMS 7	AURAMS 8
O <sub>3</sub> statistics	base case	base case	code improvements	code improvements + emissions 1	code improvements + emissions 1, 2	code improvements + emissions 1, 2, 3	diffusion cut-off = 0.6 m <sup>2</sup> s <sup>-1</sup>	renormalize non-mobile area sources	(AURAMS 4 + diffusion cut-off = 1.0 m <sup>2</sup> s <sup>-1</sup> urban, 0.01 m <sup>2</sup> s <sup>-1</sup> rural)	(AURAMS 4 + diffusion cut-off = 1.0 m <sup>2</sup> s <sup>-1</sup> urban, 0.01 m <sup>2</sup> s <sup>-1</sup> rural)
Number of pairs	41 846	41 846	41 846	41 846	41 846	41 846	41 846	41 846	41 846	41 846
Mean	22.67	39.79	<b>31.16</b>	<b>38.62</b>	<b>29.89</b>	<b>30.78</b>	33.05	<b>30.06</b>	<b>29.91</b>	<b>29.92</b>
Maximum	100	100.48	<b>100.21</b>	78.74	102.73	101.49	100.93	101.06	101.45	101.45
Minimum	0	1.26	<b>3.7 × 10<sup>-5</sup></b>	<b>0.36</b>	<b>6.10 × 10<sup>-5</sup></b>	<b>5.20 × 10<sup>-5</sup></b>	<b>5.20 × 10<sup>-5</sup></b>	<b>4.94 × 10<sup>-5</sup></b>	<b>9.01 × 10<sup>-5</sup></b>	<b>9.01 × 10<sup>-5</sup></b>
Y intercept (a) of observations vs. model line	31.11	<b>15.32</b>	<b>32.23</b>	<b>32.23</b>	<b>13.59</b>	<b>14.2</b>	<b>13.17</b>	<b>13.17</b>	<b>13.43</b>	<b>12.78</b>
Slope (b) of observations vs. model line	0.38	<b>0.70</b>	0.28	0.73	<b>0.73</b>	<b>0.74</b>	0.66	<b>0.73</b>	<b>0.76</b>	<b>0.76</b>
Correlation coefficient (R)	0.58	<b>0.64</b>	0.53	<b>0.66</b>	<b>0.66</b>	<b>0.66</b>	0.64	<b>0.66</b>	<b>0.67</b>	<b>0.67</b>
Mean bias	17.11	<b>8.48</b>	<b>15.95</b>	<b>15.95</b>	7.22	8.11	10.37	7.39	7.24	7.24
Mean absolute error	18.52	<b>12.53</b>	<b>18.01</b>	<b>18.01</b>	<b>12.05</b>	<b>12.26</b>	13.1	<b>11.92</b>	<b>11.62</b>	<b>11.62</b>
Root mean square error	21.25	<b>16.17</b>	<b>20.62</b>	<b>20.62</b>	<b>15.71</b>	<b>15.86</b>	16.78	<b>15.51</b>	<b>15.15</b>	<b>15.15</b>
Normalized mean bias (%)	75.42	<b>37.41</b>	<b>70.32</b>	<b>70.32</b>	31.84	35.75	45.73	32.58	31.94	31.94
Normalized mean error (%)	81.63	<b>55.26</b>	<b>79.45</b>	<b>79.45</b>	<b>53.13</b>	<b>54.09</b>	57.77	<b>52.61</b>	<b>51.25</b>	<b>51.25</b>

Table 4b. PM<sub>2.5</sub> statistics, entire grid ( $\mu\text{g m}^{-3}$ ). Font description as in Table 4a.

O <sub>3</sub> statistics	Observations	CMAQ 1 base case	AURAMS 1 base case	CMAQ 2 emissions 1	AURAMS1b code improvements + emissions	AURAMS 2 code improvements + emissions 1	AURAMS 3 code improvements + emissions 1, 2	AURAMS 4 code improvements + emissions 1, 2, 3	AURAMS 5 diffusion cut-off = $0.6\text{ m}^2\text{ s}^{-1}$	AURAMS 6 renormalize non-mobile area sources	AURAMS 7 (AURAMS 4 + diffusion cut-off = $1.0\text{ m}^2\text{ s}^{-1}$ urban, $0.01\text{ m}^2\text{ s}^{-1}$ rural)	AURAMS 8 (AURAMS4 + AURAMS6 emissions)
	Number of pairs	27 236	27 200	27 236	27 236	27 236	27 236	27 236	27 236	27 236	27 236	27 236
Mean	7.76	4.70	<b>10.58</b>	4.08	10.86	<b>10.56</b>	11.39	11.63	<b>8.92</b>	10.87	<b>9.26</b>	<b>9.26</b>
Maximum	519.0	44.49	<b>69.99</b>	28.06	<b>77.3</b>	<b>80.87</b>	<b>84.07</b>	<b>85.34</b>	49.4	<b>74.47</b>	62.13	61.78
Minimum	0.0	<b><math>1.2 \times 10^{-4}</math></b>	0.17	0.00064	0.2	0.2	0.21	0.22	0.22	0.21	0.22	0.22
Y intercept (a) of observations vs. model line		<b>3.64</b>	7.95	<b>3.36</b>	8.10	<b>7.66</b>	8.20	8.35	<b>7.10</b>	<b>7.91</b>	<b>7.38</b>	<b>7.37</b>
Slope (b) of observations vs. model line		0.14	<b>0.34</b>	0.092	<b>0.36</b>	<b>0.37</b>	<b>0.41</b>	<b>0.42</b>	0.23	<b>0.38</b>	0.24	0.24
Correlation coefficient (R)		<b>0.25</b>	0.23	0.18	0.23	<b>0.25</b>	<b>0.26</b>	<b>0.26</b>	0.23	<b>0.26</b>	0.23	0.23
Mean bias		-3.07	<b>2.82</b>	-3.69	3.1	<b>2.79</b>	3.62	3.87	<b>1.15</b>	3.11	<b>1.50</b>	<b>1.50</b>
Mean absolute error		<b>4.61</b>	6.77	4.99	7.12	6.78	7.38	7.47	<b>5.42</b>	6.92	<b>5.52</b>	<b>5.52</b>
Root mean square error		<b>7.2</b>	10.53	7.59	11.07	<b>10.41</b>	11.2	11.36	<b>8.19</b>	<b>10.47</b>	<b>8.29</b>	<b>8.29</b>
Normalized mean bias (%)		-39.51	<b>36.29</b>	-47.51	39.92	<b>35.98</b>	46.66	49.82	<b>14.82</b>	40.03	<b>19.31</b>	<b>19.28</b>
Normalized mean error (%)		<b>59.33</b>	87.14	64.3	91.75	87.35	95	96.24	<b>69.84</b>	89.17	<b>71.04</b>	<b>71.06</b>

Table 5. PM<sub>2.5</sub> statistics, Lower Fraser Valley stations. Font description as in Table 4a.

O <sub>3</sub> statistics	Observations	CMAQ 1		AURAMS 1		CMAQ 2		AURAMS1b		AURAMS 2		AURAMS 3		AURAMS 4		AURAMS 5		AURAMS 6		AURAMS 7		AURAMS 8	
		base case	3808	3813	3813	3813	3813	3813	3813	3813	3813	3813	3813	3813	3813	3813	3813	3813	3813	3813	3813	3813	3813
Number of pairs	3813	3808	3813	3813	3813	3813	3813	3813	3813	3813	3813	3813	3813	3813	3813	3813	3813	3813	3813	3813	3813	3813	3813
Mean	7.5	4.75	<b>8.70</b>	<b>8.43</b>	<b>8.27</b>	<b>8.5</b>	<b>8.64</b>	<b>8.37</b>	<b>8.5</b>	<b>8.5</b>	<b>8.5</b>	<b>8.5</b>	<b>8.5</b>	<b>8.5</b>	<b>8.5</b>	<b>8.5</b>	<b>8.5</b>	<b>8.5</b>	<b>8.5</b>	<b>8.5</b>	<b>8.5</b>	<b>8.5</b>	<b>8.5</b>
Maximum	49	31.95	<b>53.94</b>	58.72	<b>50.32</b>	42.86	40.42	58.72	42.86	40.42	42.86	40.42	42.86	40.42	42.86	40.42	42.86	40.42	42.86	40.42	42.86	40.42	42.86
Minimum	0	<b>6.1 × 10<sup>-4</sup></b>	0.52	<b>0.44</b>	<b>0.50</b>	0.47	0.49	<b>0.44</b>	0.47	0.50	0.47	0.49	0.47	0.50	0.47	0.49	0.47	0.50	0.47	0.49	0.47	0.49	0.47
Y intercept (a) of observations vs. model line		<b>3.69</b>	5.36	<b>5.31</b>	5.53	5.58	5.59	<b>5.31</b>	5.58	5.53	5.58	5.59	5.58	5.53	5.59	5.58	5.59	5.58	5.59	5.58	5.59	5.58	5.59
Slope (b) of observations vs. model line		0.14	<b>0.45</b>	0.42	0.37	0.39	0.41	0.42	0.39	0.37	0.39	0.41	0.39	0.37	0.41	0.39	0.41	0.39	0.41	0.39	0.41	0.39	0.41
Correlation coefficient (R)		0.19	<b>0.28</b>	0.27	0.26	0.29	0.31	0.27	0.29	0.26	0.29	0.31	0.29	0.26	0.31	0.29	0.31	0.29	0.31	0.29	0.31	0.29	0.31
Mean bias		-2.76	<b>1.2</b>	<b>0.93</b>	<b>0.77</b>	1.00	1.14	<b>0.93</b>	1.00	<b>0.77</b>	1.00	1.14	1.00	<b>0.77</b>	1.14	1.00	1.14	1.00	1.14	1.00	1.14	1.00	1.14
Mean absolute error		<b>4.57</b>	5.28	5.26	5.14	4.95	4.91	5.26	4.95	5.14	4.95	4.91	4.95	5.14	4.91	4.95	4.91	4.95	4.91	4.95	4.91	4.95	4.91
Root mean square error		<b>6.05</b>	7.72	7.62	7.19	6.82	6.74	7.62	6.82	7.19	6.82	6.74	6.82	7.19	6.74	6.82	6.74	6.82	6.74	6.82	6.74	6.82	6.74
Normalized mean bias (%)		-36.73	<b>15.95</b>	<b>12.38</b>	<b>10.23</b>	13.32	15.18	<b>12.38</b>	13.32	<b>10.23</b>	13.32	15.18	13.32	<b>10.23</b>	15.18	13.32	15.18	13.32	15.18	13.32	15.18	13.32	15.18
Normalized mean error (%)		<b>60.89</b>	70.4	70.06	68.59	66.05	65.42	70.06	66.05	68.59	66.05	65.42	66.05	68.59	65.42	66.05	65.42	66.05	65.42	66.05	65.42	66.05	65.42



**Figure 3.** Comparison of AURAMS results, Vancouver Airport, using default diffusion cut-off of  $0.1 \text{ m}^2 \text{ s}^{-1}$  (solid lines) and CMAQ value of  $1.0 \text{ m}^2 \text{ s}^{-1}$  (dashed lines).

of the differences between the models was the use of a higher value for the minimum diffusion coefficient in CMAQ.

The use of a higher level of diffusion than predicted by meteorological models is intended to compensate for specific inadequately modelled aspects of turbulence and transport, such as subgrid-scale flows through complex urban topography and turbulence induced by urban heat islands. Recent work with CMAQv5.0.1 using a linearly interpolated cut-off (between  $0.01 \text{ m}^2 \text{ s}^{-1}$  in rural areas and  $1.0 \text{ m}^2 \text{ s}^{-1}$  in urban areas) showed marginally worse performance when the high cut-off values in urban areas were removed (Pleim and Gilliam, 2012). Our above analysis suggests that the use of a higher-than-realistic diffusivity lower limit to describe subgrid-scale turbulent mixing may result in degraded and unrealistic ozone performance at night and may influence positive biases in the ozone concentration on the following day. The use of a relatively high value for the diffusivity lower limit allows greater vertical mixing to occur at night, allowing the emitted NO to be distributed over a larger vertical volume, reducing  $\text{O}_3$  titration and allowing more  $\text{O}_3$  to be mixed downwards into the lower part of the model. These effects allow the morning ozone production on the subsequent day to start from a higher concentration than would otherwise be the case, which may in turn allow  $\text{O}_3$  to reach higher concentrations by the late afternoon. While this change in initial morning  $\text{O}_3$  levels may contribute to the difference in the model results for  $\text{O}_3$ , it should be noted that this is not always the most significant factor, in that Fig. 2a shows that AURAMS and CMAQ sometimes have similar daytime  $\text{O}_3$  peak levels despite having very different  $\text{O}_3$  morning minima.

Given the difficulty in achieving good performance for both  $\text{O}_3$  and  $\text{PM}_{2.5}$  via the use of a large lower limit in diffusivity, our focus for the next stage of our analysis became the emissions. Most of the night-time  $\text{PM}_{2.5}$  predicted by the models is primary in origin (i.e. directly emitted), hence potential errors in emissions magnitude, timing or spatial distribution may also play a critical role in setting night-time

$\text{PM}_{2.5}$  concentrations. Consequently, below, we examine the emissions for our domain in some detail and conduct several tests to determine the impact of improvements to the emissions and of model sensitivity to emissions changes, in addition to the use of a lower limit in diffusion coefficient values. We then compare the results of the above tests to additional simulations making use of more recent methodologies for diffusivity, applying the same procedure for lower limits on diffusivity in AURAMS as is applied in CMAQ5.0.1.

## 4.2 Scenarios

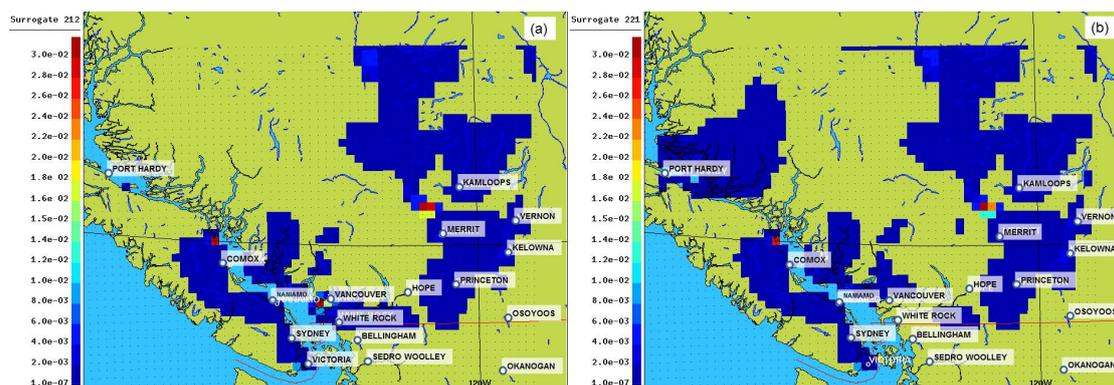
The above work led to three levels of analysis and revisions to the emissions, with a focus on the Canadian emissions data with which the authors have the greatest familiarity. The first level (“Emissions 1”) identified the top 20 emitting sources for  $\text{PM}_{2.5}$  and NO on the Canadian side of the domain. The temporal and spatial surrogate assignments for these sources were reviewed in detail to identify possible sources of  $\text{PM}_{2.5}$  positive biases (the sensitivity to the annual totals in the emissions inventories was not directly examined here). This identified errors in both spatial and temporal fields, described below, which were consequently corrected. The second level (“Emissions 2”) repeated the above analysis, but for the top 50 emitters in the four grid squares comprising the urban core of the city of Vancouver. The reasoning underlying this second analysis was that many large sources of  $\text{PM}_{2.5}$  occur outside the urban core, hence the analysis of Emissions 1 may miss spatial and temporal allocation errors important for the urban regions where the errors have the greatest impact on the model positive biases. The third level (“Emissions 3”) was to examine the impact of improving stack parameter information for primary  $\text{PM}_{2.5}$  emissions, for the specific sources in the four urban Vancouver grid squares. The details of these three stages of analysis are described below.

### 4.2.1 First-level emissions analysis: totals on the Canadian portion of the grid

Upon examining the top 20 annual sources of Canadian emissions, several deficiencies in temporal and spatial allocation were identified.

#### Temporal allocation

The links to these sources’ monthly, weekly and diurnal temporal allocation fields were used to construct grid-total time series of emissions of  $\text{PM}_{2.5}$  and NO for the summer period simulated here, allowing the relative importance of the different sources on the Canadian side of the domain during the day to be determined. The temporal profiles for on-road mobile emissions were updated based on new measurement data (Zhang et al., 2011). The temporal profiles of 21 other activities were found to be inappropriate upon review. For example, charcoal grilling (residential and commercial) was assumed to have a “flat” profile, unchanging with month, day



**Figure 4.** Comparison of spatial surrogates (a) 212 (used previously for mining activities) and (b) 221 (used in Emissions 1,2,3 scenarios). Note high values of mining activity assumed in urban Vancouver in (a), absent in (b).

of week or hour, despite the seasonality of the residential portion of this activity and the absence of this activity in late night and early morning hours. This source was the second to fourth largest source of primary  $\text{PM}_{2.5}$  at night (depending on the hour), due to this flat profile. Wood stoves and furnace boilers, and fireplaces were found to have a time-independent monthly profile (despite reduced heating energy needs in the summer, the time of the simulations of interest). Several activities (e.g. fertilizer application, land-spreading of manure, agricultural tractors, agriculture production) made use of simple sinusoidal diurnal temporal profiles with a positive offset from zero – hence late-night and early morning emissions of these daytime activities were non-zero. The temporal profile for fugitive dust emissions from paved and unpaved roads did not follow the known activity levels associated with mobile emissions (and the profile used for the former resulted in higher night-time emissions levels than that used for the latter). Marine vessel emissions were assumed to follow the temporal profile for railways. These inappropriate temporal allocation links were corrected:

1. The diurnal profile for charcoal grilling was revised to take commercial and residential activity levels into account, with zero emissions late at night and in the early morning hours.
2. The monthly profile for woodstoves/furnace boilers and for fireplaces was modified to take seasonal energy use into account.
3. The agricultural temporal profile was modified from a sinusoid with trough value greater than zero to a sinusoid which reached zero levels in the late evening/early morning.
4. Fugitive dust emissions from paved and unpaved roads were assumed to follow the same diurnal profile as mobile emissions activities.

5. Marine vessels were assumed to have a constant diurnal profile (this was a relatively minor change; the railway profile used earlier having been almost constant as well).

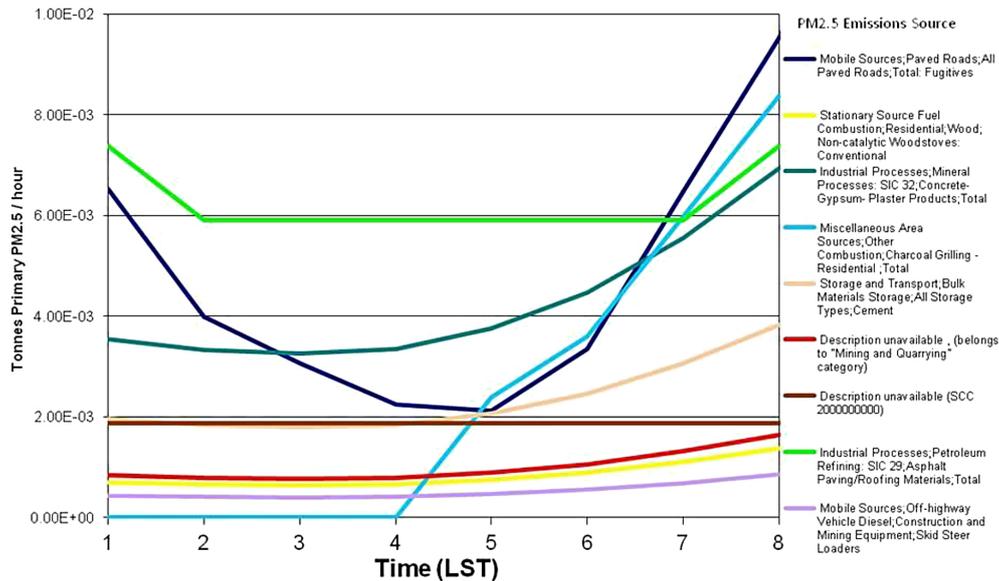
### Spatial allocation

Six new spatial surrogates were generated for mobile emissions (Zhang et al., 2011). Four activities associated with the mining industry were found to be linked to spatial surrogates that had maxima in urban regions – these linkages were switched to an existing “total mining” surrogate which better reflected the location of actual mining activities in the domain (the original surrogate included mining head offices as “mining activities”, resulting in emissions being allocated in urban Vancouver instead of the actual mining locations, see Fig. 4). Twenty-five spatial surrogates were improved through the incorporation of new Graphical Information System (GIS) fields for the Lower Fraser Valley.

#### 4.2.2 Second-level emissions analysis, Vancouver urban grid squares

### Temporal allocation

In a manner similar to the first-level analysis, a list of the top 50 annual emitters corresponding to four downtown Vancouver grid cells was generated. These were linked to monthly, weekly and diurnal temporal profiles and the resulting time series examined for accuracy with respect to the emitting activities. The resulting total emission time series for the nine largest of these sources is shown in Fig. 5. Four activities were found to be linked to profiles with no or minimal expected diurnal variation. Emissions from “other industry” were assumed to be time-invariant, despite the diurnal nature of most human activities. Asphalt paving and roofing was assumed to take place on an almost time-invariant diurnal profile (the same as used for railway



**Figure 5.** Temporal allocation of Primary PM<sub>2.5</sub> from top nine sources at night in downtown Vancouver.

emissions). “Concrete/gypsum/plaster products” and “bulk materials storage; all storage types; cement” were assumed to make use of the sinusoidal profile offset from zero mentioned above. All four of these sources were linked to a new diurnal profile which zeroed emissions during the night between 22:00 and 05:00 local standard time.

### Spatial allocation

Two spatial allocation fields, “coal industry – coal cleaning” and “mining industry crawler/tractors” were found to have maxima in urban regions – a revised linkage to the new total mining surrogate was used to take into account the actual location of mining industries.

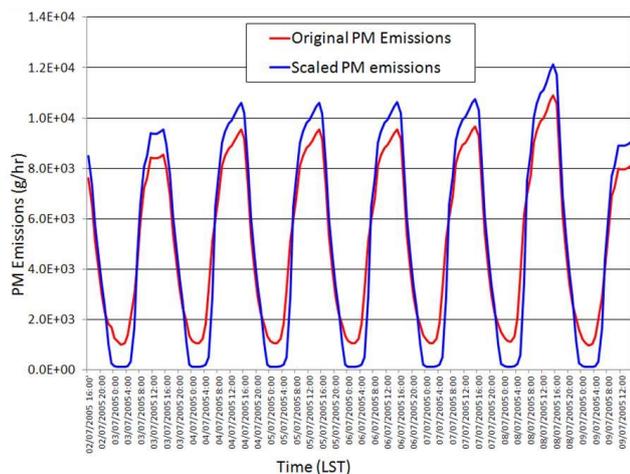
#### 4.2.3 Third level of emissions investigation, specific point sources

For one of the grid squares in urban Vancouver, minor point sources dominate as a group for PM<sub>2.5</sub> emissions, compared to major point, non-mobile area sources and mobile area sources. Only the operators of point sources with stack heights greater than or equal to 50 m are required under Canadian legislation to report stack parameters (height, diameter, exit temperature, exit velocity) associated with emissions to the Canadian National Pollutant Release Inventory (NPRI). Consequently, all stacks with elevations less than 50 m are treated as surface area sources within the Canadian portion of the domain, and the absence of plume rise in the subsequent vertical distribution of emissions may result in surface-level over-predictions of particulate matter. Point sources in the USA are available at lower heights, but a cut-off of 30 m is usually used to reduce the number of sources for which plume rise calculations are required.

Municipal-level reporting of stack parameters is, however, required for all sources in the Metro Vancouver jurisdiction. For the four largest of these facilities, the original PM<sub>2.5</sub> emissions totals (NPRI, treated as area sources) were replaced with Metro Vancouver data that included stack parameters, allowing vertical redistribution of emissions to take place, as a sensitivity test on the predicted local PM<sub>2.5</sub> levels.

#### 4.2.4 Sensitivity simulations: temporal emissions allocation versus two strategies for setting a minimum in diffusivity

A further analysis of PM<sub>2.5</sub> emissions subsequent to the above changes examined urban diurnal profiles on the basis of four main emissions categories: major point sources, minor point sources, mobile area sources and non-mobile area sources. Non-mobile area sources dominated primary PM<sub>2.5</sub> emissions (particularly in US cities where the above Canadian emissions changes were not applied). While the temporal and spatial allocations of the largest of these sources were reviewed and improved in the above analysis, this was only carried out for the Canadian side of the grid. Also, many other area sources not included in the above analysis contribute to total emissions, and the non-mobile area sources in the USA were unaffected. The default diurnal profiles of total area source emissions from the processed emissions data typically showed either a time-invariant or an offset sinusoidal shape (i.e. a temporal profile with a positive offset diurnal sinusoidal variation contributed to the bulk of the non-mobile area source emissions). In order to examine the relative importance of the diurnal variation in emissions from these sources at night, a sensitivity simulation was carried out: the PM<sub>2.5</sub> emissions from all non-mobile area sources



**Figure 6.** Comparison of total PM emissions across model domain, original versus scaled (AURAMS6 scenario, see text).

were modified using a smoothed square-wave function which reduced the emissions during the night and increased them during the day. The total mass emitted was preserved, yet proportionately less was emitted at night and more during the day. This scenario investigates the possibility that the emissions totals are correct, but that the timing of the emissions may overestimate the night-time component. Figure 6 compares the time series of grid total emissions of PM from these sources before and after this change. The intent of the simulation is to investigate the extent to which diurnal emissions behaviour of non-mobile area sources may impact the resulting concentration predictions. This in turn highlights the relative importance of accurate temporal allocation information towards the model accuracy.

The impact of two different strategies for setting a lower limit for diffusivity was also examined. In the first of these, a lower limit for diffusivity of  $0.6 \text{ m}^2 \text{ s}^{-1}$  was applied throughout the AURAMS domain. In the second, the CMAQ 5.0.1 strategy was employed, wherein the lower limit was set using a linear interpolation in land-use fraction, with values ranging between  $1.0 \text{ m}^2 \text{ s}^{-1}$  for completely urban grid squares and  $0.01 \text{ m}^2 \text{ s}^{-1}$  for completely rural grid squares. The land fractions were derived from the same BELD3 database used for biogenic emissions data (US EPA, 2007).

In a final simulation, the changes in emissions temporal allocation described at the start of this subsection were combined with the CMAQ5.0.1 diffusivity minimum strategy. All of these sensitivity studies made use of the third level of emissions improvements described above as their starting point (AURAMS4).

#### 4.2.5 Upgrades to AURAMS

Ongoing improvements to AURAMS during the course of this study included changing from the AURAMS default operator splitting setup (one step forward operator splitting) to

centred operator splitting, eliminating an additional source of differences between CMAQ and AURAMS. This was found to have a significant impact on sea salt aerosol production, significantly reducing levels offshore. In addition, the particle dry deposition algorithm was upgraded to treat particle settling and deposition in a semi-Lagrangian approach, and conservation of column mass was enforced in the vertical diffusion algorithm through separation of the area emissions, diffusion and gaseous deposition into three different operators. A separate test of this suite of changes was conducted in order to determine their impact on model performance (AURAMS1b in the subsequent discussion).

#### 4.3 Quantitative comparison of the impacts of the changes to the model and emissions

The above analysis led to nine model simulations in addition to the original base case. These scenarios are outlined in Table 2, with statistical results in Tables 4 and 5.

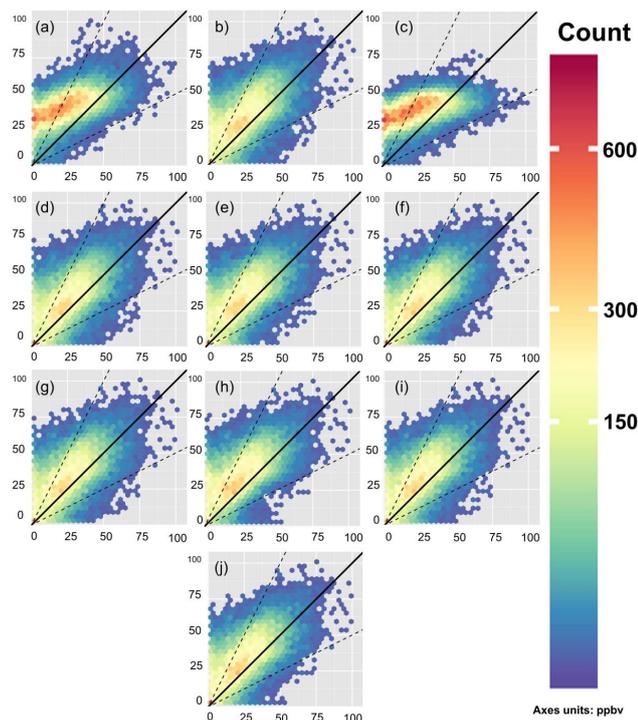
The hourly  $\text{O}_3$  and  $\text{PM}_{2.5}$  predictions from the above simulations were compared to observations as described above; summary tables of the statistical results for the entire grid are shown in Table 4a ( $\text{O}_3$ ) and b ( $\text{PM}_{2.5}$ ). The second column of the table shows observed mean, maximum and minimum values. The third and fourth columns show the results of the initial base case comparison with observations, with normal font showing the model with the lower score and bold font showing the model with the higher score. In the subsequent columns, the model results are compared to their respective base case simulation. Figures 7 and 8 show binned scatterplots of the model simulations of  $\text{O}_3$  and  $\text{PM}_{2.5}$  versus observations for the runs analysed in Table 4a and b.

##### 4.3.1 Impact of AURAMS code improvements

The changes to AURAMS' code improved statistical scores for all  $\text{O}_3$  measures (Table 4b) except for the maximum and minimum  $\text{O}_3$ , which saw a slight decrease, and correlation coefficient, which was unchanged. Comparison of Fig. 7b and d shows a relatively minor impact on the overall scatter between observations and model values for these changes, with a more pronounced difference visible between the two models (e.g. Fig. 7a vs. b). Conversely,  $\text{PM}_{2.5}$  scores became worse with the exception of maximum  $\text{PM}_{2.5}$  and the slope: Fig. 8b and d suggest a slight increase in  $\text{PM}_{2.5}$  values. Despite the statistical differences noted, the impact of the model improvements on the visual appearance of the scatterplots was minor.

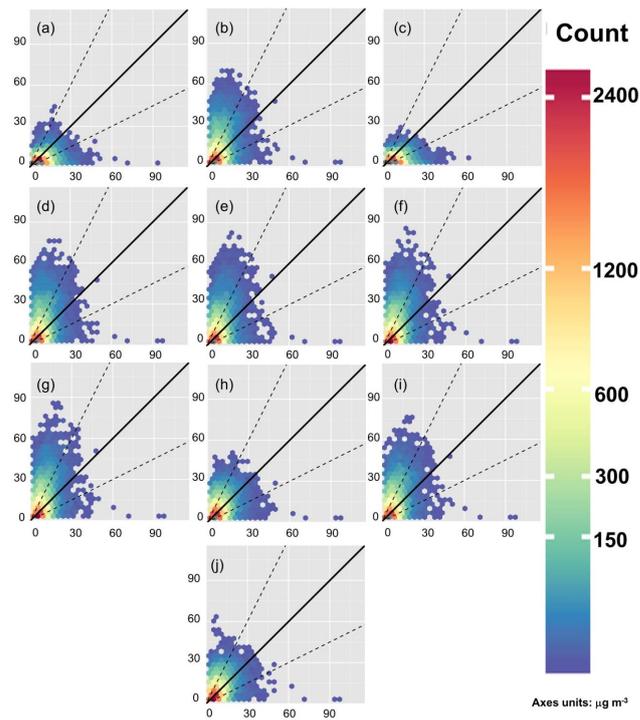
##### 4.3.2 Impact of first-level emissions improvements

For CMAQ4.6, the improvements to the emissions had a mixed effect on the model results. Ozone scores for the mean, mean bias, mean absolute error, mean squared error, root mean square error, normalized mean bias and normalized mean error all improved relative to the base case, while



**Figure 7.** Scatterplot hourly  $O_3$  comparisons of each model run versus observations. (a), CMAQ1, (b) AURAMS1, (c) CMAQ2, (d) AURAMS1b, (e) AURAMS2, (f) AURAMS3, (g) AURAMS4, (h) AURAMS5, (i) AURAMS6, (j) AURAMS7. 1 : 1 line is shown as solid line, 1 : 2 and 2 : 1 lines as dotted lines. Colour-bar scale is count frequency, showing the number of model/observation pairs falling within the given hexagon.

performance was degraded for maximum, minimum, y intercept, slope and correlation coefficient. Figure 7a and c show the lower slope and increased y intercept noted in the table. CMAQ4.6 tended to underpredict the maximum  $O_3$  values (lower values on the y axis in Fig. 7c compared to a). All CMAQ4.6  $PM_{2.5}$  scores were degraded with the use of the improved emissions, with the exception of the y intercept. Comparing Fig. 8a and c suggests that one impact of the stage 1 emissions change was to decrease CMAQ4.6's ability to simulate  $PM_{2.5}$  maxima, which is reflected in the statistics. For AURAMS, the use of the first level of emissions improvements resulted in improvements for all  $O_3$  statistics except maximum and minimum. Figure 7d and e are broadly similar: the improvements to AURAMS'  $O_3$  predictions do not result in a substantially different scatter distribution. The statistical measures for AURAMS'  $PM_{2.5}$  with the stage 1 emissions improved relative to the base case with the exception of the minimum, mean absolute error and normalized mean error, all of which showed a slight degradation of performance. Differences in  $PM_{2.5}$  scatter for the stage 1 emissions are minor: a slight shift of the distribution to the right (compare Fig. 8b, d, e). Comparing the columns in Table 4a for AURAMS simulations to isolate the impact of



**Figure 8.** Scatterplot hourly  $PM_{2.5}$  comparisons of each model run versus observations. (a) CMAQ1, (b) AURAMS1, (c) CMAQ2, (d) AURAMS1b, (e) AURAMS2, (f) AURAMS3, (g) AURAMS4, (h) AURAMS5, (i) AURAMS6, (j) AURAMS7.

the emissions improvements alone on that model, it can be seen that the  $O_3$  scores for slope and correlation coefficient have improved, while the other scores have degraded, but for  $PM_{2.5}$  all statistics with the exception of the minimum  $PM_{2.5}$  have improved.

The relative success of the first level of improved emissions data thus appears to be species and model dependant. The revised emissions had a mixed impact on CMAQ4.6's  $O_3$  performance, and degraded CMAQ4.6's  $PM_{2.5}$  performance over most statistics. For AURAMS (considering the impact of emissions alone),  $O_3$  performance was degraded slightly, while  $PM_{2.5}$  performance generally improved.

#### 4.3.3 Impact of second- and third-level emissions improvements

The second level of emissions improvements (applied only to AURAMS; "AURAMS3" columns of Table 4a and b) results in further improvements to most  $O_3$  statistics, despite a reduction in performance for  $PM_{2.5}$  for statistics other than maximum, slope and correlation coefficient. The differences relative to the first level of emissions changes are difficult to distinguish visually (Figs. 7 and 8e, f).

The third level of emissions improvements (applied only to AURAMS; "AURAMS4") showed no impact on  $O_3$  (as expected, since the final level of improvements was a sensitivity

test applied only to primary  $\text{PM}_{2.5}$  emissions; hence Fig. 7f and g are identical). Changes to the  $\text{PM}_{2.5}$  statistics across the grid were relatively minor due to this test (as might be expected given that the emissions were modified in only four grid squares in urban Vancouver). However, differences in the outer envelope of the corresponding scatterplot (Fig. 8f and g) can be observed: the third-level emissions scenario changes the distribution for cases of high model over-prediction.

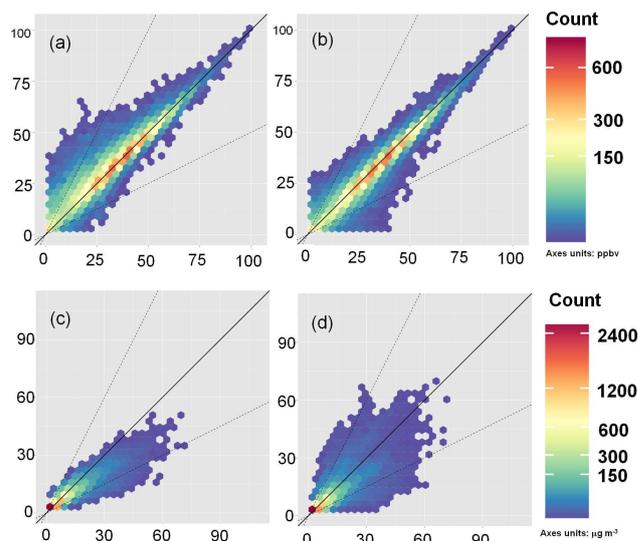
#### 4.3.4 Sensitivity simulation 1: impact of a domain-wide diffusivity cut-off

The application of a diffusion cut-off of  $0.6\text{ m}^2\text{ s}^{-1}$  (“AURAMS5”) resulted in a degradation of AURAMS’  $\text{O}_3$  performance for all scores except for the correlation coefficient, while improving AURAMS’s  $\text{PM}_{2.5}$  performance for all scores except for maximum, minimum, slope and correlation coefficient. The scatterplots for this simulation, Figs. 7 and 8h, are significantly different from the other scatterplots for AURAMS. For  $\text{O}_3$  (Fig. 7h), more of the points are clustered in the centre of the distribution, reflecting the improvement in statistics such as the RMSE. However, there are also many points along the y axis which are now in the hotter colours in Fig. 7h, indicating instances where the observed  $\text{O}_3$  was close to zero, while the modelled  $\text{O}_3$  was sometimes as high as 30 ppbv. These points correspond to cases of nighttime underprediction of NO titration of  $\text{O}_3$ , described earlier. The scatter for  $\text{PM}_{2.5}$  improved significantly, with the removal of many of the high values and a better distribution about the one-to-one line than any of the other simulations. As before,  $\text{PM}_{2.5}$  improvements via this approach came at the cost of  $\text{O}_3$  performance degradation.

#### 4.3.5 Sensitivity simulation 2: impact of temporal renormalization of non-mobile area source emissions

Renormalizing the  $\text{PM}_{2.5}$  non-mobile area sources so that less non-mobile area source emissions of these species occur at night (“AURAMS6”) maintained  $\text{O}_3$  performance for all scores (similar to AURAMS4), while improving all scores for  $\text{PM}_{2.5}$  aside from the minimum and the slope (which was unchanged). The corresponding scatterplots (Figs. 7 and 8i) show some of the same behaviour as the previous run (“AURAMS5”, Figs. 7h and 8h), and the number of  $\text{PM}_{2.5}$  points with very high over-predictions has decreased and the distribution about the one-to-one line has improved, though not to the same extent as diffusion cut-off simulation.

The two simulations above are compared relative to the base case AURAMS1 simulation in Fig. 9. One impact of using a higher diffusion cut-off for  $\text{O}_3$  (Fig. 9a) is an increase in the number of counts close to the y axis (i.e.  $\text{O}_3$  minima are increasing), while the temporal redistribution of emissions (Fig. 9b) results in both increases and decreases in



**Figure 9.** Scatterplot comparison of  $\text{O}_3$  and  $\text{PM}_{2.5}$ . (a)  $\text{O}_3$ , AURAMS5 versus AURAMS1; (b)  $\text{O}_3$ , AURAMS6 versus AURAMS1; (c)  $\text{PM}_{2.5}$ , AURAMS5 versus AURAMS1; (d)  $\text{PM}_{2.5}$ , AURAMS6 versus AURAMS1.

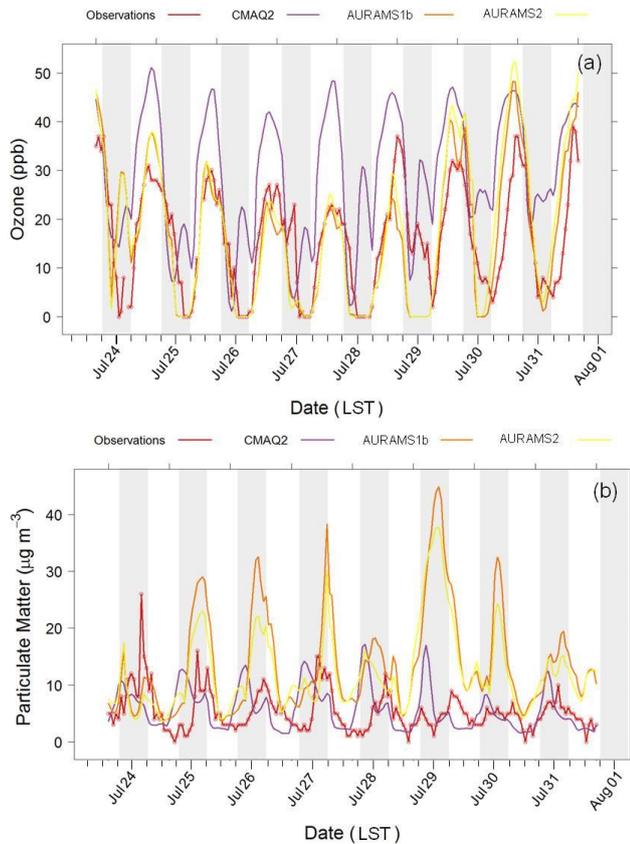
low-level  $\text{O}_3$  predictions. The higher value for the lower limit in diffusivity causes  $\text{PM}_{2.5}$  to trend downward relative to the base case (Fig. 9c), while the redistribution of emissions has a more uniform distribution across the one-to-one line, with slightly greater counts below the line (Fig. 9d).

#### 4.3.6 Sensitivity simulation 3: impact of a land-use-dependent diffusivity lower limit

The adoption of the land-use-dependent lower limit in diffusivity strategy within AURAMS (“AURAMS7”) resulted in improvement to all  $\text{O}_3$  statistics, with the exceptions of the model maximum and mean (see Table 4a), relative to both the original model runs (AURAMS1) and the simulations with improved emissions data (AURAMS4). Similar improvements in  $\text{PM}_{2.5}$  performance were also seen for the mean, mean bias, mean absolute error, root mean square error, normalized mean bias, normalized mean error and y-intercept. Model performance was degraded for the  $\text{PM}_{2.5}$  slope and correlation coefficient. Figures 7j and 8j show the corresponding scatterplots: the ozone positive bias at low concentrations has been reduced, relative to the spatially invariant lower limit in diffusivity simulations (Fig. 7h). The use of this strategy reduces the positive biases associated with the previous runs for  $\text{PM}_{2.5}$  (Fig. 8j), though the correlation along the one-to-one line has not improved.

#### 4.3.7 Sensitivity simulation 4: combined emissions and land-use-dependent diffusivity lower limits

The land-use-dependent lower limit in diffusivity was combined with the renormalized non-mobile area source

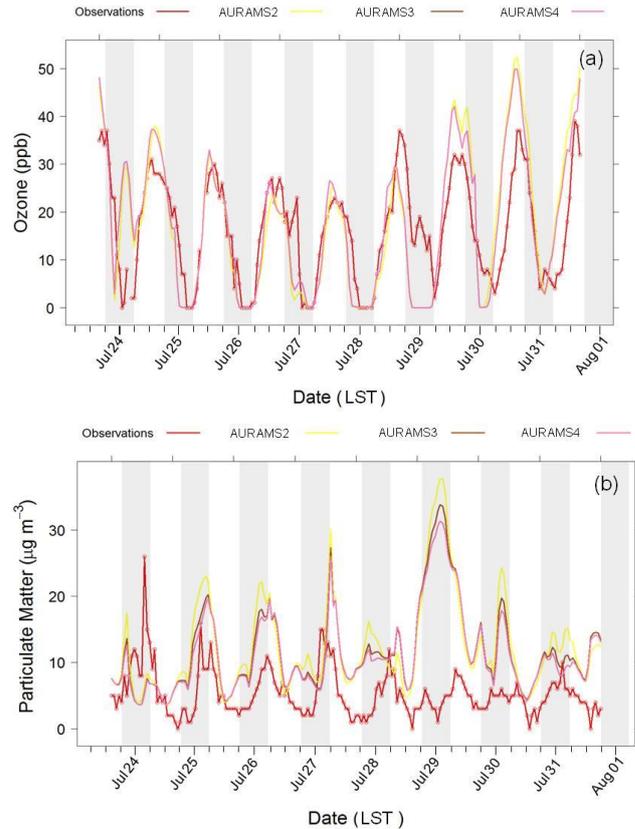


**Figure 10.** Revised stage 1 emissions and model code compared to observations for (a)  $O_3$  and (b)  $PM_{2.5}$  at Vancouver Airport. Compare to Fig. 2.

emissions scaling scenario in this simulation. The performance for this last simulation was found to be almost identical to that with the land-use-dependent lower limit in diffusivity alone; the performance improvements from the two scenarios were not additive. The scenario results imply that, while the temporal allocation of emissions can have a significant impact on model results (AURAMS6), that impact is decreased with increasing turbulence strength (AURAMS7 very similar to AURAMS8).

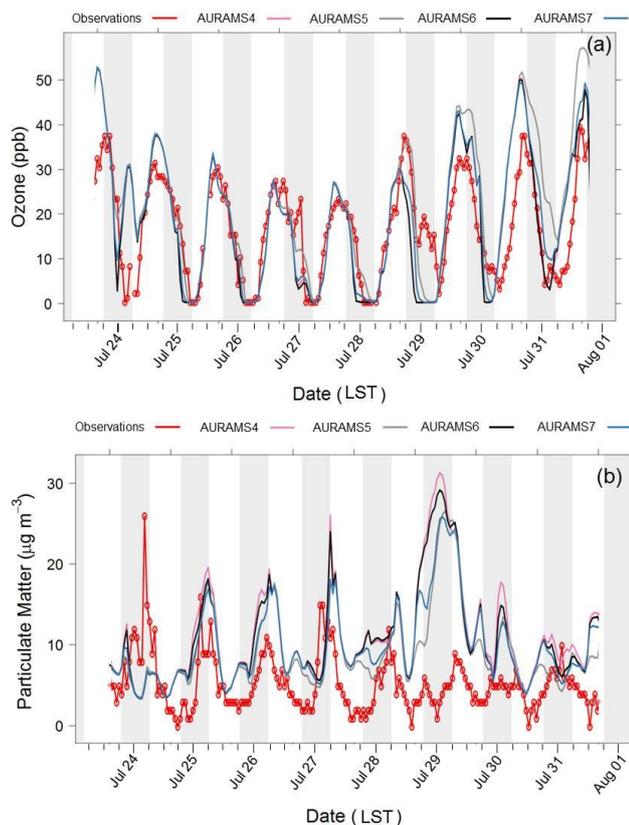
#### 4.3.8 Model performance in the Lower Fraser Valley

The performance of the models for  $PM_{2.5}$  across the five Lower Fraser Valley stations is shown in Table 5. Here, the base case performance of the two models was mixed, with each model outscoring the other for 7 out of 14 statistical measures. The first level of emissions upgrades has degraded CMAQ's performance as seen in the across-grid statistics of Table 4. The introduction of the  $0.6 \text{ m}^2 \text{ s}^{-1}$  diffusivity lower limit and the renormalizing of non-mobile area source emissions have a similar impact on model results as noted above, while the spatially invariant diffusivity lower limit degrades  $O_3$  performance for all measures except maximum and correlation coefficient (not shown).



**Figure 11.** Revised stage 1, stage 2 and stage 3 emissions compared to observations for (a)  $O_3$  and (b)  $PM_{2.5}$  at Vancouver Airport. Compare to Figs. 2, 8. Note that AURAMS3 is overplotted by AURAMS4 in (a).

Example model time series for  $O_3$  and  $PM_{2.5}$  at a station in the Lower Fraser Valley are compared to observations in Figs. 10 through 12. The degradation in CMAQ4.6's  $O_3$  performance with the use of the first level of emissions upgrades is noticeable as increases in night-time  $O_3$  levels (compare, e.g., Fig. 2, minima on the night of 30 July). AURAMS'  $O_3$  maxima increase with the use of the first-level emissions change, while AURAMS'  $PM_{2.5}$  levels decrease, sometimes substantially (cf. night of 26 July, Fig. 10b). The subsequent levels of emissions changes have relatively little impact on  $O_3$  (Fig. 11a), though local reductions in  $PM_{2.5}$  continue (Fig. 11b). Figure 12 shows the local impact of a cut-off in diffusion of  $0.6 \text{ m}^2 \text{ s}^{-1}$  on that of a reduction in non-mobile area source emissions at night (AURAMS5). Here, night-time  $O_3$  levels are erroneously increased, and night-time  $PM_{2.5}$  levels are decreased. The shift in the timing of non-mobile area source emissions of  $PM_{2.5}$  (AURAMS6) has a minimal effect on  $O_3$ , while night-time levels of  $PM_{2.5}$  decrease slightly (compare to AURAMS4). The use of a spatially varying diffusivity lower limit (AURAMS7) improves ozone performance in this urban location relative to a spatially invariant lower limit of  $0.6 \text{ m}^2 \text{ s}^{-1}$  (AURAMS5) and results in reductions in  $PM_{2.5}$  levels intermediate between the

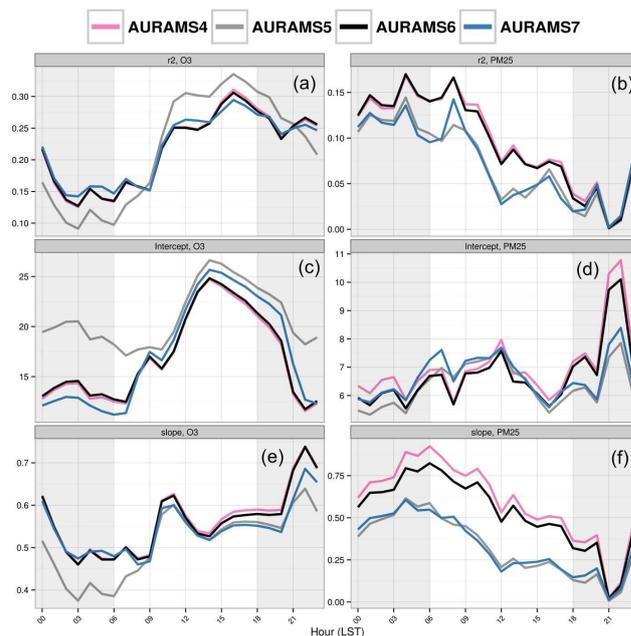


**Figure 12.** Revised stage 3 emissions, diffusion cut-off of  $0.6 \text{ m}^2 \text{ s}^{-1}$  and temporally scaled non-mobile area source emissions compared to observations at Vancouver Airport for (a)  $\text{O}_3$  and (b)  $\text{PM}_{2.5}$ . Compare to Figs. 2, 8, and 9.

other scenarios in this figure and the simulation employing a spatially invariant lower limit in diffusivity (AURAMS4). As noted in the above statistical analysis the AURAMS8 simulations were very similar to those of AURAMS7 and have not been plotted here.

#### 4.4 Time series of model statistics by hour

The last five scenarios all made use of the stage 3 emissions as their starting point and are examined here in more detail. For these simulations, the hourly grid statistics at each local hour (across all days of the simulation) were calculated and plotted as time series (Figs. 13 and 14). From Fig. 13, the use of a spatially invariant lower limit in diffusivity of  $0.6 \text{ m}^2 \text{ s}^{-1}$  (AURAMS5, grey line) reduces the  $\text{O}_3$  correlation coefficient during the night while increasing it during the day (Fig. 13a); furthermore, it increases the  $\text{O}_3$  intercept at all hours (particularly at night, Fig. 13b), decreases the  $\text{O}_3$  slope at night (Fig. 13c) and increases the night-time  $\text{O}_3$  mean bias (Fig. 14a), mean absolute error (Fig. 14c), normalized mean bias (Fig. 14e), normalized mean error (Fig. 14g) and root mean square error (Fig. 14i). The  $\text{PM}_{2.5}$  correlation coefficient for this simulation has also decreased (sometimes

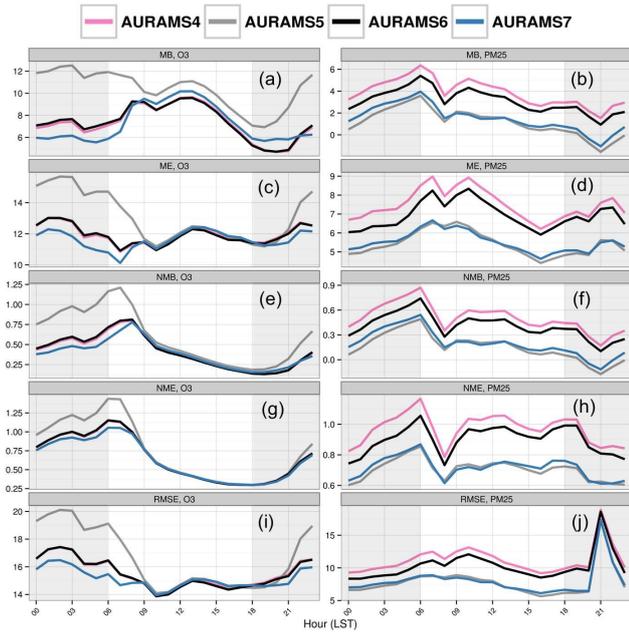


**Figure 13.** Model statistics by hour (LST) for AURAMS 4, 5, 6 and 7:  $R^2$ , intercept and slope.

halved, Fig. 13b), while the  $\text{PM}_{2.5}$  intercept improves in the early evening hours (Fig. 13d), and the slope has decreased by about 0.25 for most of the day (Fig. 13e). The use of the spatially invariant lower limit in diffusivity does, however, improve all the mean  $\text{PM}_{2.5}$  statistics throughout the day (Fig. 14b, d, h, j). Despite the improvements to mean  $\text{PM}_{2.5}$  statistics, the use of a larger spatially invariant diffusivity lower limit reduces  $\text{O}_3$  performance significantly, particularly at night, as well as reducing  $\text{PM}_{2.5}$  correlation coefficient and slope performance.

Temporally renormalizing non-mobile area source emissions of  $\text{PM}_{2.5}$  (AURAMS6, black line) has a relatively small impact on  $\text{O}_3$  performance (AURAMS4 and AURAMS6 overlap).  $\text{PM}_{2.5}$  correlation coefficients decrease very slightly (Fig. 13b), while the value of the intercept has improved (Fig. 13f), while the  $\text{PM}_{2.5}$  slope has a slight decrease, relative to the AURAMS4 or AURAMS7 simulations. The AURAMS6 results also show decreases in  $\text{PM}_{2.5}$  mean bias, mean absolute error, normalized mean bias, normalized mean error and root mean square error. The shift of primary  $\text{PM}_{2.5}$  emissions from night to day has resulted in an improvement of many of the statistics throughout the day, decreasing slope values slightly, while improving intercepts, and leaving correlation coefficients unchanged.

The AURAMS 7 and AURAMS 8 lines overlap to the extent to be indistinguishable for most of the time series; AURAMS7 is shown in these figures (pink line). The night-time  $\text{O}_3$  correlation coefficient has increased slightly, while the daytime correlation coefficient decreases (Fig. 13a); the night-time  $\text{O}_3$  intercept improves slightly while the daytime



**Figure 14.** As for Fig. 13, for mean bias, mean absolute error, normalized mean bias, normalized mean error and root mean square error.

values become worse (Fig. 13c), and a slight decrease in the O<sub>3</sub> slope occurs in the early evening (Fig. 13e). The O<sub>3</sub> nighttime mean bias, mean absolute error, normalized mean bias, normalized mean error and root mean square error all improve relative to the AURAMS4 case, with a slight degradation of daytime O<sub>3</sub> performance for mean bias. The PM<sub>2.5</sub> results are similar to the AURAMS5 (spatially invariant diffusivity minimum) case, with correlation coefficient and slope decreasing and mean statistics improving. However, compared to the spatially invariant and higher magnitude diffusivity lower limit, the ozone performance is sometimes improved rather than degraded.

## 5 Discussion

The work described above suggests the following:

1. The choice of a larger magnitude and spatially invariant minimum cut-off in diffusivity may sometimes lead to insufficient titration of ozone at night, and/or mixing of higher-level ozone downwards, creating erroneously high O<sub>3</sub> predictions at night and potentially resulting in higher O<sub>3</sub> predictions during the day. When a higher cut-off in diffusivity was tested within AURAMS, PM<sub>2.5</sub> scores were improved, but at the expense of degrading O<sub>3</sub> scores, particularly at night. If model PM<sub>2.5</sub> emissions are erroneously high, the use of a high diffusivity cut-off may compensate for these errors, lowering PM<sub>2.5</sub>. This suggests that hourly ozone performance should be used as another means of

ensuring that compensating errors of this nature are not taking place.

2. The hypothesis that at least some of the PM<sub>2.5</sub> prediction errors may result from errors in the emissions inputs has some merit. A series of tests to model emissions in which temporal and spatial allocation errors were corrected and changes in diurnal profiles were investigated showed a similar improvement to a spatially invariant diffusion cut-off approach, without degrading O<sub>3</sub> performance or even causing it to improve. This indicates that model performance may under some circumstances be as sensitive to the level of accuracy of the magnitude and spatial and temporal allocation of the driving emissions data as to the parameterization of vertical mixing. The sensitivity of the model to the temporal allocation of emissions will also depend on the strength of vertical diffusivity, with the sensitivity decreasing with increasing diffusivity strength.
3. The use of a land-use-dependent lower limit in diffusivity similar to that employed by CMAQ5.0.1 resulted in improvements to night-time O<sub>3</sub> and many PM<sub>2.5</sub> statistics, though the slope of model PM<sub>2.5</sub> versus observations was decreased, as was the correlation coefficient (which was sometimes half of its previous value, depending on the time of day). One concern about this approach was that the lower limit in diffusivity was being applied in AURAMS throughout the atmospheric column, hence possibly resulting in excessive diffusive mixing in the free troposphere and upper atmosphere. Two further sensitivity runs were carried out in which the upper extent of the region of enhanced “urban” diffusivity was limited to 2.17 km and 285 m, respectively, the former based on urban boundary layer simulations for New York City in Makar et al. (2006), and the latter based on Vancouver measurements from the same reference and from more recent observations and 250 m-resolution simulations of the mixing height in Vancouver (Leroy et al., 2014). Above those heights, the diffusivity minimum was set to 0.01 m<sup>2</sup> s<sup>-1</sup> in these additional simulations. The statistical performance of the model at the surface was identical to two or more decimal places to the AURAMS7 run for these simulations, similar to the comparison between AURAMS7 and AURAMS8. This suggests that the diffusivity minimum need only be applied to the typical height of the mixed layer above urban regions; a typical maximum altitude can be employed without degrading surface performance, while avoiding increasing diffusivity in the free troposphere and above.
4. There are other factors which may act to reduce PM<sub>2.5</sub> concentrations aside from temporal and spatial allocation. For example, fugitive emissions of PM<sub>2.5</sub> are subject to land-use-dependant reduction factors to

account for the very local-scale uptake of  $\text{PM}_{2.5}$  to vegetation, sometimes resulting in significant reductions from the inventory emissions levels for fugitive sources (cf. Pace, 2005). Similar local reduction/local availability factors may be worth considering for other  $\text{PM}_{2.5}$  sources.

5. We note that the accuracy of the relative magnitude of the emissions of different species is also important. For example, if the  $\text{NO}_x$  emissions alone are currently underestimated, then the negative impact of a high value for the minimum diffusivity on  $\text{O}_3$  performance would be decreased.
6. At least some of AURAMS  $\text{PM}_{2.5}$  over-predictions may still reside in vertical mixing, emissions or other issues: model values were still biased positively over all emissions improvement and sensitivity runs performed here, indicating that other processes are required to reduce  $\text{PM}_{2.5}$  levels.
7. It should be noted that the current work is limited, in that only emissions and diffusivity approaches were examined in detail as a cause for differences between the two model results. The model errors in general may also be reduced through adopting a higher resolution to better simulate the complex topography and urban turbulence in the region. For example, the models make use of different deposition parameterizations, and Nopmongkol et al. (2012) found that models with relatively high deposition rates for  $\text{PM}_{2.5}$  were biased low for their overall performance. Leroyer et al. (2014) found that circulation over the Vancouver urban area was best simulated at resolutions of 250 m – this resolution allowing the model to resolve urban up- and downdraughts. While changes to the timing of primary emissions of  $\text{PM}_{2.5}$  were shown to potentially account for much of the difference between the two models, changes to the particle deposition velocity algorithms may account for the remaining positive bias in AURAMS and negative bias in CMAQ for  $\text{PM}_{2.5}$ . This should be examined in future work. Also, while we have focussed on the Lower Fraser Valley in some of our analyses, the relative importance of the different processes may differ in other parts of the model domain.
8. Our work has focussed on the differences between the two models, but has important implications for the broader issue of explaining the causes for the formation of  $\text{O}_3$  and  $\text{PM}_{2.5}$  in urban and downwind environments and the relative importance of turbulence and emissions. Our results suggest that the temporal allocation of emissions may be more important in stable atmospheres than previously expected, but also that this sensitivity is reduced with increasing turbulence, in urban regions. Our results suggest that discrepancies between simulated

and observed night-time chemistry cannot be explained via increases in turbulence alone, in that  $\text{PM}_{2.5}$  correlation coefficients and slopes are still degraded with the best of the diffusivity lower-limit procedures tested here.

## 6 Conclusions

The CMAQ (version 4.6) and AURAMS (version 1.4.2) models were compared, using a common horizontal map projection and grid spacing, a common set of meteorological inputs, and a common emissions inventory and emissions processing system, for a domain on the north-west coast of North America, for a one-month simulation for the summer of 2005. The initial model results were markedly different, with AURAMS having significantly better performance for  $\text{O}_3$  than CMAQv4.6, while CMAQ's performance for  $\text{PM}_{2.5}$  was better than that of AURAMS. One of the main factors leading to the differences was found to be the magnitude of the assumed lower limit in the coefficient of vertical diffusivity employed in each model, with the adoption of a higher value in AURAMS resulting in performance more like that of CMAQv4.6. Improvements in  $\text{PM}_{2.5}$  performance associated with a larger value of a spatially invariant minimum in eddy diffusivity were also associated with significantly degraded performance for  $\text{O}_3$ . A subsequent investigation of emissions through improvements to spatial and temporal allocations and sensitivity tests showed that  $\text{PM}_{2.5}$  performance could be improved through emissions improvements, without degrading  $\text{O}_3$  performance. The use of a land-use-dependent lower limit in vertical diffusivity (similar to that used in CMAQ5.0.1) was found to improve night-time  $\text{O}_3$  performance and also  $\text{PM}_{2.5}$  performance for statistics other than the correlation coefficient and the slope (both of which were sometimes halved when this approach was adopted, depending on the time of day). The model results were shown to have a similar level of sensitivity to emissions' spatial and temporal allocation as to lower limits in vertical mixing for lower levels of turbulence. However, when urban vertical diffusivity was modified using the spatially varying lower limit, the model's sensitivity to emissions temporal allocation was greatly reduced.

The findings have important implications for our understanding of  $\text{O}_3$  and  $\text{PM}_{2.5}$  in urban environments. A spatially invariant lower limit in diffusivity was shown to be insufficient to explain the discrepancies between observations and simulations for these species. However, the choice of a lower limit on diffusivity must be made with care. A spatially varying lower limit in diffusivity improved several statistical scores, implying that accurate portrayal of urban turbulence is critical for model performance. However, higher levels of minimum diffusivity in urban areas also resulted in decreases in  $\text{PM}_{2.5}$  correlation coefficients and slopes and may also mask errors in spatial and temporal allocation of

PM<sub>2.5</sub> primary emissions. We have found that the heretofore inadequately resolved timing and spatial allocation of PM<sub>2.5</sub> primary emissions, particularly from the non-mobile area source sector, may have a considerable influence on PM<sub>2.5</sub> concentrations. We therefore recommend improvements to both area source primary PM<sub>2.5</sub> emissions data and urban turbulence parameterizations as foci for future measurement and modelling work.

These results should not be taken to imply that improvements to the model representations of turbulent mixing and/or other factors should be ruled out as a line of investigation for achieving improved model performance. Both emissions (timing, spatial distribution and magnitude) and the magnitude of turbulent diffusion were shown to be of potential importance here. Our results suggest that both processes are complementary routes for further model improvements. Model performance for both O<sub>3</sub> and PM<sub>2.5</sub> should be simultaneously evaluated in future work to ensure that improvements in one predicted species are not offset by degraded model performance in the other.

*Acknowledgements.* The authors would like to acknowledge the assistance of Bruce Ainslie of Environment Canada's Pacific and Yukon Region for helpful discussions on the CMAQ model.

Edited by: C. Jacobi

## References

- Ainslie, B., Steyn, D. G., Reuten, C., and Jackson, P. L.: A retrospective analysis of ozone formation in the Lower Fraser Valley, British Columbia, Canada. Part II: influence of emissions reductions on ozone formation, *Atmos. Ocean*, 51, 170–186, 2013.
- Binkowski, F. S. and Roselle, S. J.: Models-3 Community Multiscale Air Quality (CMAQ) model aerosol component 1. Model description, *J. Geophys. Res.*, 108, 4183, doi:10.1029/2001JD001409, 2003.
- Binkowski, F. S. and Shankar, U.: The regional particulate model 1. Model description and preliminary results, *J. Geophys. Res.*, 100, 26191–26209, 1995.
- Brook, J. R., Strawbridge, K., Snyder, B. J., Boudries, H., Worsnop, D., Anlauf, K., Sharma, S., Lu, G., and Hayden, K.: Towards an understanding of the fine particle variations in the LFBV: integration of chemical, physical and meteorological observations, *Atmos. Environ.*, 38, 5775–5788, 2004.
- Carr, D., Lewin-Koh, N., and Maechler, M.: hexbin: Hexagonal Binning Routines, R package version 1.24.0, available at: <http://CRAN.R-project.org/package=hexbin> (last access: 17 October 2013), 2010.
- Carslaw, D. and Ropkins, K.: openair: Open-source tools for the analysis of air pollution data, R package version 0.4-15, 2011.
- Carter, W. P. L.: Documentation of the SAPRC-99 Chemical Mechanism for VOC Reactivity Assessment, Final Report to California Air Resources Board Contract 92-329 and Contract 95-308, available at: <http://www.engr.ucr.edu/~carter/reactdat.htm> (last access: 17 October), 2013, 2000a.
- Carter, W. P. L.: Implementation of the SAPRC-99 Chemical Mechanism into the Models-3 Framework, Report to the United States Environmental Protection Agency, available at: <http://www.engr.ucr.edu/~carter/reactdat.htm>, (last access: 17 October 2013), 2000b.
- CEP: Carolina Environmental Program, Sparse Matrix Operator Kernel Emission (SMOKE) modelling system, University of North Carolina, Carolina Environmental Programs, Chapel Hill, NC, available at: <http://www.smoke-model.org/index.cfm> (last access: 17 October 2013), 2003.
- Côté, J., Gravel, S., Méthot, A., Patoine, A., Roch, M., and Staniforth, A.: The operational CMC-MRB Global Environmental Multiscale (GEM) model. Part 1: Design considerations and formulation, *Mon. Weather Rev.*, 126, 1373–1395, 1998.
- Deepayan, S.: Lattice: Multivariate Data Visualization with R. Springer, New York, ISBN 978-0-387-75969-2, 2008.
- Galmarini, S. and Rao, S. T.: The AQMEII two-continent Regional Air Quality Model evaluation study: Fueling ideas with unprecedented data, *Atmos. Environ.*, 45, p. 2464, 2011.
- Giorgi, F.: A particle dry deposition scheme for use in tracer transport models, *J. Geophys. Res.*, 91, 9794–9806, 1986.
- Gong, W., Dastoor, A. P., Bouchet, V. S., Gong, S., Makar, P. A., Moran, M. D., Pabla, B., Menard, S., Crevier, L.-P., Cousineau, S., and Venkatesh, S.: Cloud processing of gases and aerosols in a regional air quality model (AURAMS), *Atmos. Res.*, 82, 248–275, 2006.
- Houyoux, M. R., Vukovich, J. M., Coats Jr., C. J., and Wheeler, N. J. M.: Emission inventory development and processing for the Seasonal Model for Regional Air Quality (SMRAQ) project, *J. Geophys. Res.*, 105, 9079–9090, 2000.
- Kelly, J., Makar, P. A., and Plummer, D. A.: Projections of mid-century summer air-quality for North America: effects of changes in climate and precursor emissions, *Atmos. Chem. Phys.*, 12, 5367–5390, doi:10.5194/acp-12-5367-2012, 2012.
- Leroyer, S., Belair, S., Husain, S. Z., and Mailhot, J.: Sub-kilometer numerical weather prediction in an urban coastal area: a case study over the Vancouver metropolitan area, *J. Appl. Meteorol. Clim.*, online first, doi:10.1175/JAMC-D-13-0202.1, 2014.
- Makar, P. A., Gravel, S., Chirkov, V., Strawbridge, K. B., Froude, F., Arnold, J., and Brook, J.: Anthropogenic heat flux, urban properties, and regional weather, *Atmos. Environ.*, 40, 2750–2766, 2006.
- Makar, P. A., Gong, W., Mooney, C., Zhang, J., Davignon, D., Samaali, M., Moran, M. D., He, H., Tarasick, D. W., Sills, D., and Chen, J.: Dynamic adjustment of climatological ozone boundary conditions for air-quality forecasts, *Atmos. Chem. Phys.*, 10, 8997–9015, doi:10.5194/acp-10-8997-2010, 2010.
- McKeen, S., Wilczak, J., Grell, G., Djalalova, I., Peckham, S., Hsie, E.-Y., Gong, W., Bouchet, V., Menard, S., Moffet, R., McHenry, J., McQueen, J., Tang, Y., Carmichael, G. R., Pagowski, M., Chan, A., Dye, T., Frost, G., Lee, P., and Mathur, R.: Assessment of an ensemble of seven real-time ozone forecasts over eastern North America during the summer of 2004, *J. Geophys. Res.*, 110, D21307, doi:10.1029/2005JD005858, 2005.
- McKeen, S., Chung, S. H., Wilczak, J., Grell, G., Djalalova, I., Peckham, S., Gong, W., Bouchet, V., Moffet, R., McHenry, J., Tang, Y., Carmichael, G. R., Mathur, R., and Yu, S.: Evaluation of several PM<sub>2.5</sub> forecasts using data collected during

- the ICARTT/NEAQS 2004 field study, *J. Geophys. Res.*, 112, D10S20, doi:10.1029/2006JD007608, 2007.
- McKeen, S., Grell, G., Peckham, S., Wilczak, J., Djalalova, I., Hsie, E.-Y., Frost, G., Peischl, J., Schwarz, J., Spackman, R., Holloway, J., de Gouw, J., Warneke, C., Gong, W., Bouchet, V., Gaudreault, S., Racine, J., McHenry, J., McQueen, J., Lee, P., Tang, Y., Carmichael, G. R., and Mathur, R.: An evaluation of real-time air quality forecasts and their urban emissions over eastern Texas during the summer of 2006 Second Texas Air Quality Study field study, *J. Geophys. Res.*, 114, D00F11, doi:10.1029/2008JD011697, 2009.
- Nopmongkol, U., Koo, B., Tai, E., Jung, J., Piyachaturawat, P., and Sportisse, B.: Modeling Europe with CAMx for the air quality model evaluation international initiative (AQMEII), *Atmos. Environ.*, 53, 60–74, 2012.
- Otte, T. L. and Pleim, J. E.: The Meteorology-Chemistry Interface Processor (MCIP) for the CMAQ modeling system: updates through MCIPv3.4.1, *Geosci. Model Dev.*, 3, 243–256, doi:10.5194/gmd-3-243-2010, 2010.
- Pace, T. G.: Methodology to estimate the transportable fraction (TF) of fugitive dust emissions for regional and urban scale air quality analyses (8/3/2005 Revision), Internal report, US EPA, Research Triangle Park, North Carolina, 12 pp., available at: [http://www.epa.gov/ttn/chief/emch/dustfractions/transportable\\_fraction\\_080305\\_rev.pdf](http://www.epa.gov/ttn/chief/emch/dustfractions/transportable_fraction_080305_rev.pdf) (last access: 17 October 2013), 2005.
- Pierce, D.: ncdf: Interface to Unidata netCDF data files, R package version 1.6.3, available at: <http://CRAN.R-project.org/package=ncdf> (last access: 17 October 2013), 2010.
- Pleim, J. E.: A combined local and nonlocal closure model for the atmospheric boundary layer. Part I: model description and testing, *J. Appl. Meteorol. Climatol.*, 46, 1383–1395, 2007a.
- Pleim, J. E.: A combined local and nonlocal closure model for the atmospheric boundary layer. Part II: application and evaluation in a mesoscale meteorological model, *J. Appl. Meteorol. Clim.*, 46, 1396–1409, 2007b.
- Pleim, J. and Gilliam, R.: Simple urban parameterization for WRF-CMAQ, 2012 CMAS conference, available at: [https://www.cmascenter.org/conference/2012/slides/pleim\\_simple\\_urban\\_2012.pptx](https://www.cmascenter.org/conference/2012/slides/pleim_simple_urban_2012.pptx) (last access: 15 May 2014), 2012.
- Pleim, J. and Ran, L.: Surface flux modeling for air quality applications, *Atmosphere*, 2, 271–302, 2011.
- Pleim, J. E., Xiu, A., Finkelstein, P. L., and Otte, T. L.: A coupled land-surface and dry deposition model and comparison to field measurements of surface heat, moisture, and ozone fluxes, *Water Air Soil Poll.: Focus*, 1, 243–252, 2001.
- Pleim, J. E., Roselle, S. J., Bhawe, P. V., Bullock, O. R., Hutzell, W. T., Luecken, D. J., Nolte, C. G., Sarwar, G., Schere, K. L., Young, J., Godowitch, J., and Appel, K. W.: The 2006 CMAQ Release and Plans for 2007, Extended abstract for the 5th CMAS Conference, Chapel Hill, NC, 16–18 October 2006, available at: <http://www.cmascenter.org> (last access: 17 October 2013), 2006.
- R Development Core Team: R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0, available at: <http://www.R-project.org> (last access: 17 October 2013), 2010.
- Richtmyer, R. D.: Difference methods for initial value problems, Krieger Pub. Co., Malabar, Fla., 405 pp., 1994.
- Smyth, S., Yin, D., Roth, H., and Jiang, W.: A study of the impact of GEM and MM5 meteorology on CMAQ modelling results in eastern Canada and the northeastern United States, Institute for Chemical Process and Environmental Technology Tech. Rep., PET-1561-04S, 77 pp., 2005.
- Smyth, S. C., Yin, D., Roth, H., Jiang, W., Moran, M. D., and Crevier, L.-P.: The impact of GEM and MM5 meteorology on CMAQ air quality modeling results in eastern Canada and the northeastern United States, *J. Appl. Meteorol. Clim.*, 45, 1525–1541, 2006a.
- Smyth, S. C., Jiang, W., Yin, D., Roth, H., and Giroux, E.: Evaluation of CMAQ O<sub>3</sub> and PM<sub>2.5</sub> performance using Pacific 2001 measurement data, *Atmos. Environ.*, 40, 2735–2749, 2006b.
- Smyth, S. C., Jiang, W., Roth, H., Moran, M. D., Makar, P. A., Yang, F., Bouchet, V. S., and Landry, H.: A comparative performance evaluation of the AURAMS and CMAQ air-quality modelling systems, *Atmos. Environ.*, 43, 1059–1070, 2009.
- Solazzo, E., Bianconi, R., Vautard, R., Appel, K. W., Moran, M. D., Hogrefe, C., Bessagnet, B., Brandt, J., Christensen, J. H., Chemel, C., Coll, I., van der Gon, H. D., Ferreira, J., Forkel, R., Francis, X. V., Grell, G., Grossi, P., Hansen, A. B., Jericevic, A., Kraljevic, L., Miranda, A. I., Nopmongkol, U., Prio-vano, G., Prank, M., Riccio, A., Sartelet, N., Schaap, M., Silver, J. D., Sokhi, R. S., Vira, J., Werhahn, J., Wolke, R., Yarwood, G., Zhang, J., Rao, S. T., and Galmarini, S.: Model evaluation and ensemble modelling of surface-level ozone in Europe and North America in the context of AQMEII, *Atmos. Environ.*, 53, 60–74, 2012a.
- Solazzo, E., Bianconi, R., Prio-vano, G., Matthias, V., Vautard, R., Moran, M. D., Appel, K. W., Bessagnet, B., Brandt, J., Christensen, J. H., Chemel, C., Coll, I., Ferreira, J., Forkel, R., Francis, X. V., Grell, G., Grossi, P., Hansen, A. B., Miranda, A. I., Nopmongkol, U., Prank, M., Sartelet, K. N., Schaap, M., Silver, J. D., Sokhi, R. S., Vira, J., Werhahn, J., Wolke, R., Yarwood, G., Zhnag, J., Rao, S. T., and Galmarini, S.: Operational model evaluation for particulate matter in Europe and North America in the context of AQMEII, *Atmos. Environ.*, 53, 75–92, 2012b.
- Steyn, D. G., Ainslie, B., Reuten, C., and Jackson, P. L.: A retrospective analysis of ozone formation in the Lower Fraser Valley, British Columbia, Canada. Part I: Dynamical model evaluation, *Atmos. Ocean*, 51, 153–169, doi:10.1080/07055900.2013.781940, 2013.
- Stockwell, W. R. and Lurmann, F. W.: Intercomparison of the ADOM and RADM Gas-Phase Chemical Mechanisms, Electrical Power Research Institute Topical Report, EPRI, 3412 Hillview Avenue, Palo Alto, CA, 254 pp., 1989.
- US EPA: U.S. Environmental Protection Agency Biogenic Emissions Landcover Database, available at: <http://www.epa.gov/ttnchie1/emch/biogenic/> (last access: 2 March 2014), 2007.
- Xiu, A. and Pleim, J. E.: Development of a land surface model, Part 1: Application in a mesoscale meteorology model, *J. Appl. Meteorol.*, 40, 192–209, 2001.
- Zhang, J., Zheng, Q., Moran, M. D., Gordon, M., Liggio, J., Makar, P., Stroud, C., and Taylor, B.: Improvements to SMOKE Processing of Canadian On-Road Mobile Emissions, 2012 International Emission Inventory Conference, Tampa, Florida – 13–16 August 2012, available at: <http://www.epa.gov/ttn/chief/conference/ei20/session1/jzhang.pdf> (last access: 17 October 2013), 2011.

Zhang, L., Gong, S.-L., Padro, J., and Barrie, L.: A size-segregated particle dry deposition scheme for an atmospheric aerosol module, *Atmos. Environ.*, 35, 549–560, 2001.

Zhang, L., Moran, M. D., Makar, P. A., Brook, J. R., and Gong, S.-L.: Modelling gaseous dry deposition in AURAMS – a unified regional air-quality modelling system, *Atmos. Environ.*, 36, 537–560, 2002.