**Atmospheric
Chemistry
and Physics**

Development of a fast, urban chemistry metamodel for inclusion in global models

J. B. Cohen[1,*] and R. G. Prinn[1]

[1]Joint Program of the Science and Policy of Global Change, Massachusetts Institute of Technology,
Cambridge, MA 02139, USA
*now at: Singapore-MIT Alliance for Research and Technology (SMART) Centre, Singapore

Abstract. A reduced form metamodel has been produced to simulate the effects of physical, chemical, and meteorological processing of highly reactive trace species in urban areas, which is capable of efficiently simulating the urban concentration, surface deposition, and net export flux of these species. A polynomial chaos expansion and the probabilistic collocation method have been used to develop the metamodel, and its coefficients, so that it is applicable under a broad range of present-day and future conditions. The inputs upon which this metamodel have been formed are based on a combination of physical properties (average temperature, diurnal temperature range, date, and latitude), anthropogenic properties (patterns and amounts of emissions), and the nature of the surrounding environment (background concentrations of species). The metamodel development involved using probability distribution functions (PDFs) of the inputs to run a detailed parent chemical and physical model, the Comprehensive Air Quality Model with Extensions (CAMx), thousands of times. Outputs from these runs were used in turn to both determine the coefficients of and test the precision of the metamodel, as compared with the detailed parent model. It was determined that the deviations between the metamodel and the parent mode for many important species (O_3, CO, NO_x, and black carbon (BC)) were found to have a weighted RMS error less than 10 % in all cases, with many of the specific cases having a weighted RMS error less than 1 %. Some of the other important species (VOCs, PAN, OC, and sulfate aerosol) usually have their weighted RMS error less than 10 % as well, except for a small number of cases. In these cases, the complexity and non-linearity of the physical, chemical, and meteorological processing is too large for the third order metamodel to give an accurate fit. Finally, sensitivity tests have been performed, to observe the response of the 16 metamodels (4 different meteorologies and 4 different urban types) to a broad set of potential inputs. These results were compared with observations of ozone, CO, formaldehyde, BC, and PM_{10} from a few well observed urban areas, and in most of the cases, the output distributions were found to be within ranges of the observations. Overall, a set of efficient and robust metamodels have been generated which are capable of simulating the effects of various physical, chemical, and meteorological processing, and capable of determining the urban concentrations, mole fractions, and fluxes of species, important to human health and the global climate.

1 Introduction

Urban regions have high concentrations of species which are harmful to human health, have a direct or indirect impact on the atmosphere's radiative flux balance, and alter the land's ability to uptake carbon. Furthermore, urban regions account for a large and increasing fraction of the Earths total population and anthropogenic emissions. However, modeling the effects of urban areas on the processing and export of anthropogenic emissions is not straightforward. Urban areas are located in regions of diverse geography and meteorology, they have non-constant emissions which are based on technological, economic, and political factors, and they exhibit strongly non-linear processing of primary anthropogenic pollutants. For these reasons, urban areas account for a large amount of the variability and uncertainty in the global atmospheric spatial and temporal distributions of primary and secondary anthropogenic pollutants.

Correspondence to: J. B. Cohen
(jasonbc@smart.mit.edu)

Those substances having a large percentage of their global emissions, production, or destruction occurring in urban regions in addition to having a large impact on the global radiative balance or concentrations sufficient to affect human health, are the focus in this paper. These species are typically heterogeneously distributed over space and time within single urban regions, as well as between different urban regions, due to non-linear chemical and physical processing, differences in local and regional meteorology, and differences in local emissions. To address this level of complexity, properties of urban areas relating to geography, physics, chemistry, and human activity will have to be addressed at both the local and global scale.

Global chemistry and climate models, in general, use a spatial resolution which is much too coarse to resolve the spatial scales of real urban regions. This in turn requires that these models use, compute, or predict aggregated data or data on large spatial and temporal scales, and then use this data to approximate the desired variables on the urban temporal and spatial scales. Therefore, physical variables which control the system, the concentrations of trace species, and human factors such as primary anthropogenic emissions pertaining to the urban system are derived, for the entire urban area and its environs (a "dilution" approach). Because of this, many of the variables provided at the grid scales of global models are approximations that are not valid or appropriate for use on urban spatial and temporal scales.

In order for a parameterization of urban processing to be used in a global general circulation or chemical transport model, it must be capable of computing the concentrations of important trace species within a given urban area, and the fluxes of these species to the coarser global scale grids from the urban scale grids and back again. Additionally, the parameterization must be computationally efficient and yet still be flexible enough to simulate the highly variable emissions, upwind conditions, geography, and relevant economic and human factors found in urban areas, both for the present and into the near future. Some of the specific variables to consider include: the emissions of critical chemical species, the specific temporal and spatial distributions of the emissions, the time of the year, the geographic location, the surface conditions, the elevation, the amounts of rainfall and cloudiness, the horizontal and vertical circulation, the local temperature, the upwind concentrations of the species interacting within the urban area, the amount of sunlight, the relative humidity, and the atmospheric liquid water content. At a minimum, such variables must be known both in the urban area and at its boundaries, as a function of the time of day. In addition to this, since export from one urban area can greatly impact a neighboring urban area's properties, it is important to know where urban areas are located in relation to each other.

The parameterization must be capable of capturing the non-linear chemical and physical processes which actually occur within the urban area, because these processes can cause results which are lower than, higher than, or differ-

ently distributed in the horizontal, vertical, or temporal domains, compared with the large-scale averaging or dilution approach. Urban scale processing causes certain species to have a positive net production (chemical production minus chemical, physical, and depositional loss) due to the inclusion of urban-scale processing. Three examples are CO (where production from VOC oxidation far outweighs loss due to reaction with OH), NO_2 (from photochemical processing), and certain lower molecular weight VOCs (from the oxidation of larger molecular weight VOCs). Other species are formed nearly exclusively in urban areas as secondary products, and involve significant non-linear processing, such as peroxyacetyl nitrate (PAN), secondary organic carbon aerosol (OC), nitrate aerosol (NO_3), and sulfate aerosol (SO_4). Other species have a negative net production due to processing in urban areas, such as some large VOCs (oxidized to smaller VOCs), NO (oxidized to NO_2), SO_2 (oxidized to sulfuric acid and sulfate aerosol), highly water soluble species (entering into the aqueous phase), and primary aerosols (removed through coagulation, rain out, and deposition). A third subset of species can have either a positive or negative net production, such as: O_3, OH, some VOCs, and some optically important aerosols. This subset of species has behavior that depends on many factors, such as the concentrations of other species in the urban area, whether it is raining or dry, the strength of vertical advection and mixing, and the time of the day.

Due to this non-linear chemical and physical processing, the simple dilution approach tends to either underestimate or overestimate the concentration in the urban area and the flux from the urban area to the global system, depending on the species. Furthermore, the simple dilution approach does not capture the vertical, horizontal, and temporal characteristics occuring at the urban scale. For example, modeling the processing occurring due to a small region of strong uplift, subsidence, or rainfall, over a heterogeneously distributed concentration field, can yield substantially different results when compared with averaging the effects of the vertical air motion or rainfall. This short-coming has been demonstrated in a study using ozone as an example, in which incremental improvements in the spatial resolution of the models' emissions, chemistry and physics made the results compare more closely with measurements (Wild and Prather, 2006). However, even such efforts have only looked on horizontal spatial scales on the order of a degree, which are still far too coarse to precisely model processes occurring on the urban scale. There are many other efforts have used both models and measurements to look at the variability within an urban area due to changing factors such as model resolution, differences in background concentrations due to different prevailing large-scale meteorologic conditions, and sharp changes in emissions profiles (see, e.g., Qian et al., 2010; See et al., 2006; Huang et al., 2010). A few common theme of these studies is that they are limited to very specific urban areas and their immediate surroundings, they are not readily generalizable

over different temporal scales and geographic regions, they do not quantify the import and export across the boundaries of their regions of study, and they do not consider changes in climate conditions either from the past or into the future.

In this paper, we will describe the production of a metamodel (or a model of a model) which can simulate urban scale processing. The point of this metamodel is to be capable of interacting in a two-way fashion as a component in a larger global-scale modeling system. The metamodel reproduces fairly accurately the underlying parent model simulations of the urban concentrations, deposition fluxes, and lateral and vertical mass fluxes to the global scale, under typical present day and potential future conditions. Sensitivity tests and comparisons of the parent model and with observations from different urban areas reveal that the metamodel is quite successful in simulating the species of interest.

2 Prior reduced form urban process models

An early attempt at forming a parameterization of urban scale processing was made by Calbo et al. (1998). They used the California Institute of Technology urban model, driven by idealized (non-divergent, single directional, non-evolving) meteorology, and a simplified version of fast NO_x and VOC photochemistry which in turn drove ozone production for their gas-phase chemistry driver. The model was run using five layers in the vertical, with initial conditions being zero for many species outside the bottom two layers. The latitudes for which this reduced form model was developed included the regions from 60° S to 60° N latitude. The average surface temperature and mixed layer height were scaled by a predefined function over time of the day. The fractional cloud cover was constant in both space and time. The residence time of an air parcel in the urban area was determined from the wind speed across the urban box model, and was constant over time. All emissions were assumed to occur within fixed geographical regions of the urban box model (near the center), at the same fixed ratios to one another, and as a fixed function of distance from the center of the urban area. The emissions of VOC and NO_x were defined in such a way as to be correlated with the emissions of CO, not allowing for conditions in which there was a strong reduction in just one or two of these three species. Furthermore, the PDFs of the emissions of the input species were defined by beta distributions, and therefore had zero chance of being outside of the defined upper and lower boundaries. Finally, the initial conditions and boundary conditions were given using Air Quality Indexes, which are defined based upon generally clean upwind conditions.

The authors produced their parameterization by using the probabilistic collocation method (PCM) (Tatang et al., 1997), which is a method by which their model's response space was approximated by a set of orthonormal chaos polynomials, which had their polynomial coefficients computed based on sampling the PDFs of the input variables. All of their input parameters were fit using second order polynomials for each of the 14 input variables, which included all first order, second order, and cross combinations of the input variables. Using this method, they showed that there was a reasonable fit between the metamodel and the parent model for the mass fluxes of gas-phase species from the urban area (the parameterized model being within a few to 40 % of the parent modeled variable). The major exception was ozone.

A second attempt at an urban process PCM-based parameterization was made by Mayer et al. (2000), using the same parent model, chemical routines, and approach to idealized meteorology, as in Calbo et al. (1998). Some of the important differences between these two approaches are described below. All five layers in the vertical were initialized with non-zero initial conditions, the urban area was 200×200 km, the emissions only occurred in a core area of about 150×150 km, and the air entering was assumed to be clean, exclusively matching values found in remote locations. One further difference is that Mayer et al. (2000) treated the upwind concentration of NO_x as two separate input variables, one for the upwind concentration of NO and the other for the upwind concentration of NO_2. A third difference came from the variation of emissions as a function of time, with no diurnal variation of SO_2 emissions, and a constant and identical diurnal variation assumed for all other emitted gasses.

The most significant differences, however, came from how the uncertainties were treated. The emission PDFs were based on fits to a normal function, and therefore they had a non-zero probability of being any finite positive value. Furthermore, the emission PDFs were formed independently of each other, allowing for higher or lower values of each emitted species to be modeled, with all emissions values spread further than the beta PDF case, but still tightly concentrated around their median values. This change allowed for more regions of the parameter space to be explored by the metamodel. Finally, the probabilistic collocation method employed in this case used some selected third order terms in addition to the complete set of second order terms in their polynomial chaos expansion. These additional terms were added specifically to look at higher order effects on the net flux of ozone from the urban region. In general, the results from this effort showed a more reasonable fit than Calbo et al. (1998) for all mass fluxes of gas species of interest from the urban area, including ozone.

A significant difference between our new approach and prior efforts is inclusion of the impacts of liquid water, and its associated chemical and physical processes which do not occur under the dry conditions in prior efforts. These effects include: aqueous chemistry, uptake of soluble gases from the gas to the aerosol phase (including sulfuric acid), and wet deposition.

Another significant difference is inclusion of regions of strong vertical advection. The underlying meteorology fields are based on results computed throughout the troposphere

and therefore already include the effects of convection. These regions allow for efficient transport of trace species through the boundary layer, thereby altering their rates of chemical and physical transformation, surface deposition, and hence their lifetimes within the urban area. These effects can be especially important for photolysis reactions, temperature dependent reactions, or for species having strong liquid water or aerosol uptake potential. These effects can substantially increase or decrease the amounts of species exported from an urban area.

A third significant difference comes from the consideration of temporal and spatial variability in both atmospheric properties and surface emissions. For example, changes in the height of the boundary layer affect the net transfer in the vertical of trace species, and whether these species are shielded from or exposed to sunlight. Furthermore, emissions occurring at different times of the day lead to a difference in the local concentrations near the surface. These properties lead to considerable changes in mass fluxes and to non-linear chemical feedbacks. By specifically addressing these processes, our current modeling effort is an improvement over our previous modeling efforts, which used nonvarying, uniform, and non-divergent winds which flow only from West to East. A further improvement is the use of multiple cases of meteorology. This enables studies of the effects of differences in rain and wind fields within the urban region.

A few other attempts to look at the influence of detailed urban-scale processing on the larger-scale have also been made. Qian et al. (2010) looked at the variation due to changes in chemical and dynamical grid resolution, land surface resolution, and emissions resolution, over a small number of conditions. All of these studies were performed in an already well characterized urban area. This study did not include variations in emissions intensities, distributions, background concentrations, or meteorology over and beyond the differences between using different grid resolutions. Butler and Lawrence (2009) uses a global CTM to compute the differences in the global concentrations and loadings of species of interest. This is done by assuming differnet net emissions from urban areas and generalizing them across a small number of urban areas. There is no connection made between changes of emissions at the source within an urban area and how this impacts the overall net urban emissions changes used. In conclusion, these other attempts, while interesting in their own right, are not similar enough to this current work to make a direct comparison with, and therefore, this work is the first major piece of scholarship to continue to move forward on this specific topic.

3 Parent urban chemical transport model

Regional scale models of atmospheric chemistry and physics can effectively simulate the processing of emissions on spatial and temporal time scales resembling those on the urban scale, allowing for both the time varying concentrations within and the time varying export from a specific urban area to be computed. The model chosen to perform these calculations is the Comprehensive Air Quality Model with extensions (CAMx) (www.camx.com), which is an Eulerian model that solves the terrain following continuity equation, for the concentrations and fluxes of trace species. CAMx accounts for the emissions, vertical and horizontal transport and diffusion, gas and aerosol phase chemistry, and the wet and dry deposition of trace species. Additionally, CAMx takes into consideration how the properties of the Earth's surface, the given atmospheric conditions, and the amount of incident solar radiation as a function of space and time, further affect the concentrations and distributions of trace species. One advantage of this modeling system is that it is freely open and available and has a large user-community. One disadvantage is that it does not dynamically update the meteorology and hence cannot look at the impacts of the chemistry on the dynamics in a coupled manner. However, much of the recent peer-reviewed literature relating to urban and regional air quality has relied on CAMx (see, e.g., Russell and Allen, 2005).

The specific way in which CAMx accounts for these processes is by solving for each of the terms separately in Eq. (1):

$$\frac{\partial c}{\partial t} = -\nabla(cv) + \frac{\partial(c\eta)}{\partial z} + c\frac{\partial}{\partial z}\frac{\partial h}{\partial t} + \nabla(\rho K_v \nabla \frac{c}{p})$$
$$+ \frac{dc}{dt}\Big|_{\text{chemistry}} - \frac{dc}{dt}\Big|_{\text{deposition}} - \frac{dc}{dt}\Big|_{\text{removal}} \qquad (1)$$

In this equation c is the concentration (moles or mass per unit volume) of a given species, v is the horizontal wind velocity, η is the vertical wind velocity, h is the vertical layer height, ρ is the atmospheric density, and K_v is the turbulent exchange diffusion coefficient. The equation states that the net change in the concentration of a given species is the sum of: the convergence of the advective flux in the horizontal and vertical, and the diffusive flux; the chemical production and destruction, emissions, wet and dry deposition at the surface, and other physical removal processes (such as capture by cloud particles).

However, solving this equation is neither straight forward nor simple, requiring many assumptions. Firstly, all processes are treated as though they are uniformly distributed through each Eulerian grid box in which they occur; therefore emissions are diluted through the grids adjacent to the surface, physical and meteorological variables are assumed to have a single average value over a grid box, and tracers are considered to have a constant concentration throughout a grid box. The advection routine is solved in mass-conserving flux form, driven by realistic assimilated meteorology from the 1995 OTAG (http://capita.wustl.edu/otag/) campaign at four different sites (Vukovich, 1997).

In specific, the meteorology data was choosen to cover typical ranges for rainfall, cloud cover, and the net mass

Table 1. Physical descriptions of the four meteorological scenarios used to create the four metamodels.

Meteorology Case	Rainfall [mg m^{-3}]	Cloudiness [%]	Air Flux [10^9 kg s^{-1}]
R241-C63-W46	241.0	62.8	4.56
R000-F00-W44	0.00	0.00	4.38
R021-F19-W57	21.5	19.3	5.70
R002-F02-W16	1.72	1.75	1.61

Table 2. Average pressure at the top of each vertical layer of the urban modeling domain.

Vertical Layer	Pressure [hPa]
1	990
2	980
3	970
4	960
5	950
6	940
7	930
8	910
9	890
10	860
11	830
12	780
13	710

flux of air integrated across all five boundaries of the urban area. This was done by selecting the four sets of data, over a continuous 48-h period and spatial region corresponding to a 108 km square, that minimized and maximized the rainfall and mass flux of air. The amount of liquid water in the form of rain, the amount of cloud cover, and the mass flux of air (integrated over all four sides and the top of the urban area) through the boundaries of the urban area, are given in Table 1. The dominant features of these four sets of meteorology, as given in the table respectively, are: heavy rain and clouds at all times, dry and cloud free at all times, long residence time, and short residence time with intermittent heavy rain. In this way, some extreme conditions which are likely to be found in generalized urban areas can be simulated, based on actual meteorology found from within this week-long dataset, with less extreme conditions lying somewhere in between.

The vertical velocity is computed by integrating the density conservation equation. Wet removal occurs through Henry's Law processes, physical mixing, aqueous phase chemistry, and impaction by precipitation. Dry removal occurs through first order surface resistance removal schemes for gases and aerosols, and gravitational settling for aerosols. Gas phase chemistry is based on the Carbon Bond 4 approach (Gery et al., 1989), with a newer and more detailed representation of terpenes, low volatility organic species, and improved night time nitrogen chemistry (Sarwar et al., 2008). Aerosol phase chemistry includes explicit inorganic aqueous phase chemistry, inorganic thermodynamics, and formation of secondary organic and inorganic aerosol (Chang et al., 1987; Strader et al., 1999; Nenes et al., 1999; Koo et al., 2003). This is solved using the CMU scheme, over 10 equally spaced size bins (in log space) from 10 nm up to 5000 nm. The advantage of this technique is that it provides a more reasonable approximation of the size evolution of the aerosol, which is important if the user is looking to analyze climate effects of the aerosols. The main disadvantage of this technique is that it does not conserve both mass and number at the same time, but CAMx does not have a full two-moment scheme available for aerosol processing. The photolysis scheme uses a lookup table based on the TUV model, and accounts for reducing fluxes of incoming solar radiation as a function of overhead cloud thickness and reflection, and

for surface reflection. And finally, the precipitation processes include rain, snow, and ice (all of which are internally computed, based on the temperature and the strength of the vertical convection in the region).

The details of the Eulerian modeling framework used here include a uniform horizontal grid spacing of 4×4 km in the North/South and East/West directions, over a total region of 108×108 km. In the vertical, 13 layers were chosen with pressure coordinates, from the surface up to the free troposphere, as given in Table 2. CAMx was specifically integrated forward with the time step only allowed to vary between 1 and 3 min, so that the numerical solution of the equations would always remain stable. Finally, to make sure that any initialization assumptions were not affecting the results, the urban model was integrated for 96 h, with the initial 72 h being treated as a spin up and the final 24 h being the result used.

4 Probabilistic collocation method

The computational expense of running a detailed urban model, such as CAMx, is too large to individually simulate the largest hundreds or thousands of different urban areas (there are presently more than 200 urban areas with populations over 3 million people), over the temporal scales (multiple decades to centuries) required in the context of global climate modeling. In addition to these direct computational expenses, there are further computational costs, such as two way interactions between the global model and the urban model relating to the physical and chemical variables which drive the temperature, rainfall, horizontal wind, and chemical concentrations at the boundaries between the two models. To form a computationally efficient parameterization of the processes contained in CAMx, we follow Calbo et al. (1998) and Mayer et al. (2000) and use the probabilistic collocation

method (Tatang et al., 1997). One of the main strengths of this technique is that it produces an polynomial response surface that has a high degree of reliability in regions of high input parameter probability, while the two main weaknesses are that the method performs poorly in regions of low probability and that it takes an extremely large number of parent model runs to form.

Specifically, given a set of k input parameters x_j used to drive a model, $\{x_1, x_2, ..., x_k\}$, there are M output responses y_j predicted by the model that are functions of the x_j values; $\{y_1, y_2, ..., y_M\} = f(\{x_1, x_2, ..., x_k\})$. In this case the y_j are the physical concentrations, mass fluxes, and deposition fluxes of species that we are interested in approximating. And since there is a range of possible values which each input parameter can take, each input variable is considered independent and has its statistics defined by its PDF over this range. Since all of the input variables are considered random, the output variables also are considered random variables.

The response surface is being fitted specifically by a set of recursively defined orthonormal polynomials (P_{k_i}) where i ($= l$ or m below) is the order of the polynomial, g is the PDF of the random input variable x_j, and δ_{lm} is the Kronecker delta:

$$P_{k_{-1}} = 0$$
$$P_{k_0} = 1$$
$$\int g(x_j) \quad P_{k_l}(x_j) P_{k_m}(x_j) dx = \delta_{lm} \qquad (2)$$

Through these polynomials (Eq. 2), the independent random variables can be written as $x_j = x_{j_0} + x_{j_1} \cdot P_{k_1}$, and the dependent variable y_j can be approximated by the polynomial chaos expansion (Eq. 3):

$$y_j = y_{j_0} + \sum_{i=1}^{N} y_{j_i} P_{k_i} \qquad (3)$$

where y_{j_i} is a coefficient to be fit based on the parent model's predicted value for y_j at the given set of inputs $\{x_j\}$, and N is the order of the polynomial fit (in our case, full cubic).

In addition to forming the basis for the polynomial chaos expansion, the orthonormal polynomials are also used to help select the set of parameter values which are used to initialize the parent model. This set of input values, called collocation points, are solved for by finding the $N + 1$ roots of the $N + 1$ order polynomial corresponding to each input parameter x_j. These roots are from the high probability regions of each input parameter, and therefore the approximation of y_j is particularly good within the most probable range of values of the input parameters. In addition to this, a set of test points are generated from the solution of the $N + 2$ roots of the $N + 2$ order polynomial corresponding to each input parameter x_j. This second set of points form an excellent test for the metamodel in that they take into account what the next higher order of estimation would yield as a better set

of points for sampling the probability space spanned by the input parameters.

5   Metamodel inputs and outputs

The smallest possible set of input variables capturing the effects of urban chemical and physical processing must be derived in order to form a reduced form model which is as compact as possible. This set of inputs needs to be flexible enough to be applicable to the many variations of the properties of urban areas found throughout the world both presently and in the future out to 2100. Specifically, the variables need to span the differences in geography, location, time of the year, atmospheric temperature, cloudiness, amount and type of precipitation, circulation, time tendency of emissions, spatial tendency of emissions, the amount of each type of emitted species, and the upwind concentrations of species of interest, as a function of space and time. Specifically, there are 18 uncertain input variables which are used to derive the model, each of which has its uncertainty based on a wide range of both present day conditions and those conditions expected out to 2100 (derived through the running of a set of climate policy and no climate policy economic scenarios using the MIT Joint Program's EPPA model; Paltsev et al., 2005). These input variable PDFs are defined and described in Table 3 using the respective equations for Uniform PDFs (Eq. 4) (where $a \leq x \leq b$), Beta PDFs (Eq. 5) (where $a \leq x \leq b; p, q > 0$), and Lognormal PDFs (Eq. 6) (where $x, m, \sigma > 0$), where $a, b, p, q, m,$ and σ are the parameters, and x is the input value.

$$f(x) = \frac{1}{(b-a)} \qquad (4)$$

$$f(x) = \frac{(x-a)^{p-1}(b-x)^{q-1}}{(\int_0^1 t^{p-1}(1-t)^{q-1}dt)(b-a)^{p+q-1}} \qquad (5)$$

$$f(x) = \frac{\exp\left(-\frac{[\ln\frac{x}{m}]^2}{2\sigma^2}\right)}{x\sigma\sqrt{2\pi}} \qquad (6)$$

The first set of input variables in Table 3 are for the time, location, emission spatial distribution, and temperatures as discussed later. The second set of input variables are fluxes for those species that are directly emitted in the urban area. The directly emitted species are CO, NO_x (95 % emitted as NO and 5 % emitted as NO_2), VOCs, SO_2, BC (primary black carbon aerosol), and OC (primary organic carbon aerosol). The CO and BC emissions have been fitted by lognormal probability distribution functions, see Fig. 1, based on the results of 250 policy and 250 no policy runs of the MIT EPPA Model (Paltsev et al., 2005; Cossa, 2004).

The remaining primary emitted species are chosen so that they linearly scale with the emissions of either CO or BC (these linear coefficients are listed in Table 3). One reason

Table 3. Descriptions of and coefficients for the PDFs used to drive the CAMx model. The PDFs are defined based on Eqs. 4, 5, and 6.

Variable	Type of PDF	China	India	Developed	Developing
Day of the Year	Uniform (a, b)	1.000 365.0	1.000 365.0	1.000 365.0	1.000 365.0
Latitude of Urban Areas [°]	Beta (p, q, a, b)	3.663 3.897 22.7 44.3	1.736 1.694 9.70 31.4	5.802 9.842 34.7 51.3	3.309 2.625 −25.70 53.3
Temporal Weight	Uniform (a, b)	0.000 1.000	0.000 1.000	0.000 1.000	0.000 1.000
Spatial Weight [km]	Uniform (a, b)	21.6 93.2	21.6 93.2	21.6 93.2	21.6 93.2
Daily Average Surface Temperature [K]	Beta (p, q, a, b)	3.924 1.583 251.9 303.3	8.483 2.810 267.4 309.4	7.827 3.446 261.7 310.7	2.637 2.006 263.2 301.8
Daily Average Diurnal Temperature [K]	Beta (p, q, a, b)	3.841 4.125, 3.061 15.06	1.741 1.976, 5.318 18.76	4.114 3.368, 2.073 16.45	2.532 3.438, 4.037 21.24
CO Emissions [ton day^{-1}]	Lognormal (m, σ)	3162.0 1.908	2398.0 1.563	7171.0 2.651	5713.0 1.995
VOC Emissions [ton day^{-1}]	CO Emissions ×	0.0814	0.2194	0.2180	0.1891
NO$_x$ Emissions [ton day^{-1}]	CO Emissions ×	0.3558	0.2462	0.2815	0.1721
BC Emissions [ton day^{-1}]	Lognormal (m, σ)	88.56 2.133	39.88 1.761	75.45 2.387	68.15 2.459
SO$_2$ Emissions [ton day^{-1}]	BC Emissions ×	0.7646	0.3526	1.966	1.481
NH$_3$ Emissions [ton day^{-1}]	BC Emissions ×	1.104	3.468	0.9779	0.6038
OC Emissions [ton day^{-1}]	BC Emissions ×	1.615	1.586	2.587	3.890
Boundary Ozone [ppb]	Lognormal (m, σ)	26.23 1.162	28.57 1.165	24.64 1.182	32.83 1.458
Boundary CO [ppb]	Lognormal (m, σ)	81.63 1.473	96.69 1.489	75.11 1.475	105.6 1.628
Boundary NO$_x$ [ppt]	Lognormal (m, σ)	44.40 1.299	54.96 1.192	33.42 1.521	46.60 1.680
Boundary SO$_2$ [ppt]	Lognormal (m, σ)	182.1 1.355	254.9 1.472	149.7 1.316	339.3 2.223
Boundary Isoprene [ppt]	Lognormal (m, σ)	373.2 2.293	373.2 2.293	373.2 2.293	373.2 2.293

for doing this is that the sources of the gas-phase species CO, NO$_x$, and VOCs tend to be similar and those of the aerosol phase species and their precursors BC, OC, SO$_2$, and NH$_3$ also tend to be similar. Secondly, since the probabilistic collocation method samples the space best in regions which are of high probability, and since there is local correlation between these emissions they are not independent variables, a more reliable result is obtained using this method. The best fit linear relations between the emissions of these species are given in Table 4 using Eq. (7), where P is the emissions of
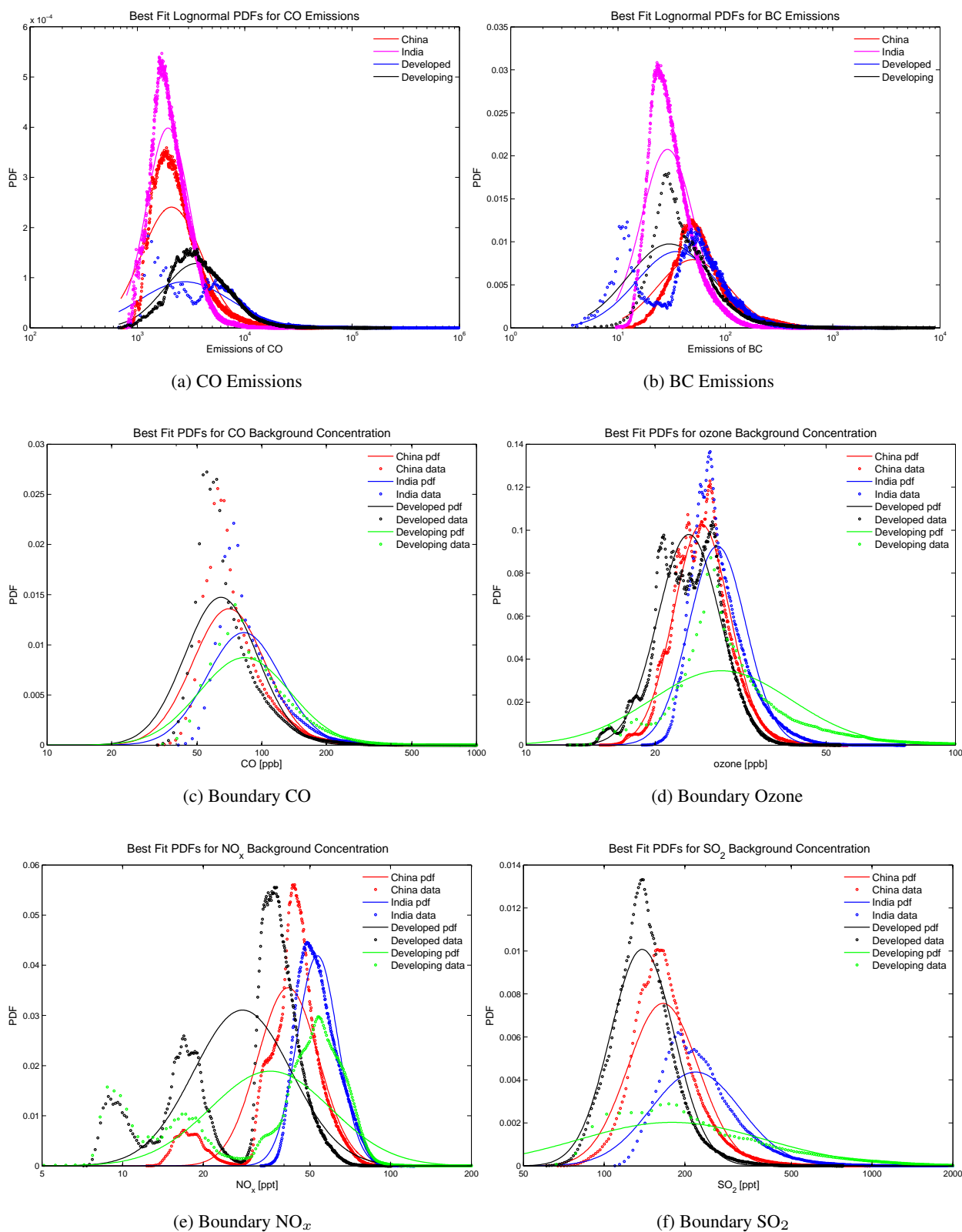
Fig. 1. PDFs of CO and BC emissions ton day^{-1}, and the boundary mole fractions of CO, ozone, NO$_x$, and SO$_2$, for each metamodel type. The dots represent the aggregated, binned, and normalized data points, while the lines represent the best fit lognormal PDFs of the respective data sets. The coefficients for these best fit lognormal PDFs are given in Table 3.

Table 4. Best fit emissions correlation statistics, based on Eq. (7). VOC and NO_x emissions are linearly related to CO emissions while OC, SO_2, and NH_3 emissions are linearly related to BC emissions.

Emitted species	Model location	Best fit slope β	Best fit intercept α
VOC	China	0.0814	0.0426
VOC	India	0.2194	−0.0190
VOC	Developed	0.2180	0.3250
VOC	Developing	0.1891	0.0571
NO_x	China	0.3558	−0.0933
NO_x	India	0.2462	−0.0711
NO_x	Developed	0.2815	0.4149
NO_x	Developing	0.1721	−0.0928
OC	China	1.6145	0.0063
OC	India	1.5864	0.0149
OC	Developed	2.5874	0.0044
OC	Developing	3.8897	−0.0179
SO_2	China	0.7646	0.1138
SO_2	India	1.3526	−0.0366
SO_2	Developed	1.9656	0.1718
SO_2	Developing	1.4809	0.1310
NH_3	China	1.1041	0.0441
NH_3	India	3.4816	0.0414
NH_3	Developed	0.9779	0.0457
NH_3	Developing	0.6038	0.0712

the parent species in $g\,day^{-1}$, X is the emissions of the subordinate species in $g\,day^{-1}$, and β and α are the best fit coefficients for the slope and intercept, respectively.

$$X = \beta P + \alpha \tag{7}$$

Furthermore, this technique employing only two independent emissions species, CO and BC, has been found to be superior to the technique employed by Mayer et al. (2000) in the case where the emissions of all or most primarily emitted species are all elevated. The superiority results from the fact that when seven different partly dependent variables are simultaneously sampled from low probability portions of their PDFs, the metamodel's reliability is inferior to when only 2 variables are sampled from low probability portions of their PDFs.

The third set of input variables in Table 3 are the mole fractions of trace species along the boundaries (the four sides and the top) of the urban area that impact the chemical and physical processing inside the urban area. The trace species that will be considered in this analysis are CO, NO_x, O_3, VOCs (represented by isoprene), and SO_2. To compute the mole fractions of these species, the aforementioned set of 250 policy and no-policy runs emissions were used for all species, except for isoprene. These were then used to drive the MIT Integrated Global Systems Model (IGSM) (Sokolov et al., 2005) to simulate the global concentrations of the relevant species, from the time period between 2000 and 2100. The

simulated concentrations of the relevant species were aggregated over time, in the vertical up to the top of the urban modeling domain, and across all relevant latitude bands. These aggregated results were then fitted by lognormal probability distribution functions using a least-squares method. Since isoprene is not simulated by the IGSM, concentration data was taken from present and past measurements (Houweling et al., 1998; Yokouchi, 1994). The results of the aggregated data along with the best fit PDFs are given in Fig. 1.

The main problems with using the results from multiple runs of the IGSM to produce these various PDFs are that the IGSM does not produce results that provide a full probabilistic sampling of polluted upwind air and the MIT IGSM does not predict certain species which could be important, such as specific VOC species. However, using the IGSM gives a better indication of how these species will change over time, and since all are fit with lognormal PDFs, values which are considerably larger are better sampled. One way to improve upon this would be to consider incorporating boundary conditions of anthropogenic and natural aerosols, such as BC, OC, sulfate, nitrate, mineral dust, and sea salt.

As noted in Table 3 the first two of these variables are the day of the year and the latitude of the urban region. These are both needed for computing the ultraviolet radiative flux. The day of the year is assigned a uniform PDF from 1 to 365. A beta fit of the distribution of latitudes of each urban area has been made for each region, making the assumption that each urban area is of equal importance. The latitudinal PDF for each region is given in Fig. 2.

The atmospheric temperature in the urban area and its range are important variables for determining the rates of many chemical reactions, Henry's Law partitioning, gas/aerosol phase partitioning, the concentration of liquid water in the urban atmosphere, and the size and total amount of precipitation. For the purposes of determining their global distribution, historical temperature data (Jones et al., 1999) has been weighted by the latitudinal beta PDFs, given in Fig. 2, for each urban area, and the resulting data fitted by a beta function. The average temperature of each of the 13 urban model vertical layers is computed assuming a linear decline with height with a standard linear lapse rate of $6.5\,K\,km^{-1}$. The spatial and temporal deviations from these layer averages in the temperature of each grid box of the urban domain are taken from the meteorology chosen for that particular urban region. The average daily surface temperature and average daily diurnal surface temperature PDFs are given in Fig. 2.

The final physical inputs to consider are rainfall and cloudiness, both of which impact the radiative fluxes, the uptake of soluble gases, and the removal rate of aerosols. After extensive testing, it has been found that treating these inputs as separate variables using the PCM approach does not yield reasonable results, due to the extremely non-linear impact these variables have on the system. Therefore, separate

(a) Average Surface Temperature

(b) Diurnal Surface Temperature


(c) Location of Urban Areas
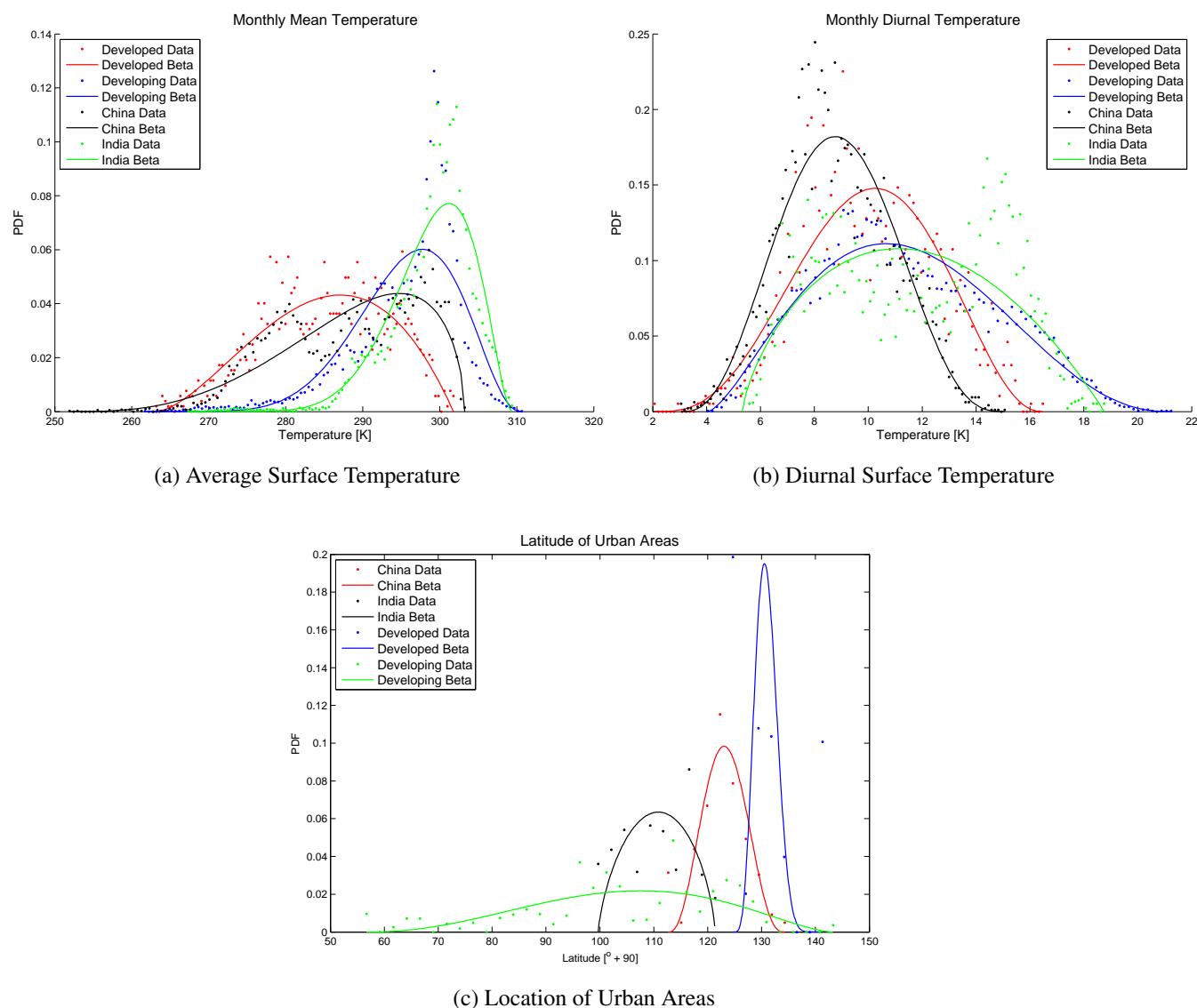
Fig. 2. PDFs of the number of degrees latitude North, from the South Pole, for the urban areas, where the dots are the fraction of urban areas in each latitude bin (urban area location data are from Center for International Earth Science Information Network, 2005); and the PDFs of the associated average daily surface temperature and the average daily diurnal temperature variation (daily high minus daily low) for each urban area, where the dots are the data for the monthly averaged respective values. In all cases, the lines represent the best beta function fits to the data.

metamodels were formed for each of the four meteorological conditions.

Another set of driving variables are designed to simulate the transportation and habitation choices people make that have an impact on the processing of species in urban regions. The first variable represents the temporal distribution of emissions. It is commonly found that emissions in urban areas have a time profile which is doubly peaked, with the peaks occurring around the times of the morning and evening rush hours. Furthermore, the middle of the day is found to have a plateau with a considerably higher amount of emissions than the nighttime plateau, as shown in Fig. 3. To ac-

count for this, an input variable *wt* is defined that is uniformly distributed from 0 to 1, and is the weight given to this double peak temporal emissions spectrum when it is linearly added to a time invariant emissions profile. Therefore, for any given value of *wt* the weights assigned to the double-peaked distribution is wt and the weight assigned to the time invariant emissions distribution is $1-wt$ (Yang et al., 2005). A second input variable relates to the spatial distribution of emissions in the urban region. Such a distribution must consider that urban areas vary greatly in terms of their density of people, activity, and thus emissions. Since most emissions are related to the population in urban centers, the emissions of both CO
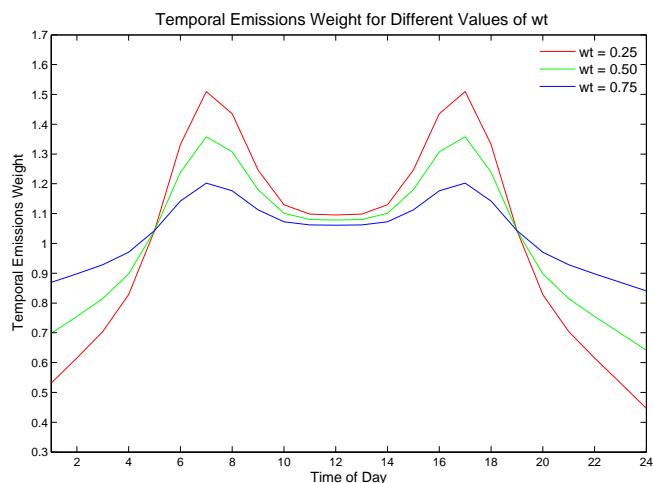
Fig. 3. The temporal weight of emissions (normalized to unity), as a function of time, for an urban area. Specifically, this is the emission time distribution obtained for the input variable $wt = \{0.25, 0.50, 0.75\}$.

and BC are considered to be spatially correlated. These spatial distributions are fitted by a 2-D Gaussian function whose standard diameter has a uniform distribution ranging from the assumed minimum size of a present world megacity of 21.6 km (Shenzhen, China) to the assumed maximum size of a present world megacity of 93.2 km (New York City Metro Area, USA). Two examples are given in Fig. 4, clearly showing the impact of this variable on the distribution of emissions.

Another input is required to address whether VOC emissions consist of a larger fraction of light hydrocarbons, corresponding to a more developed economy, or whether they have a larger fraction of heavier hydrocarbons, corresponding to a developing economy. Developed countries are assumed to have their VOC speciation follow the IPCC Third Assessment Report guidelines for Developed Nations emissions speciation of VOCs, while the rest of the world follows the IPCC Third Assessment Report guidelines for Developing Nations emissions speciation of VOCs (Prather et al., 2001). Since this final decision determines whether there are zero or non-zero values of many VOCs, the metamodels are expected to produce different results in the Developed and Developing regions of the world.

As previously mentioned, the impacts of the circulation, water content, and temperature on the processing in urban areas must be considered in detail. To address this issue, four different realistic sets of meteorology have been used to drive the urban modeling system. The point of employing these widely different cases is to numerically analyze the impact of adopting different types of realistic meteorology. This also allows for the feedback of rainfall in the global scale parent model to influence the processing at the urban scale, by providing a way to apply the same amount of rainfall, in a non-uniform manner.

Table 5. List of all species simulated by the urban metamodels. Unless otherwise stated as being a surface variable or having a specific temporal averaging, all species are simulated as a daily average value at each vertical level.

Ozone	Xylene	Black carbon mass
CO	C_2H_4	Organic carbon mass
NO	C_2H_6	Nitrate mass
NO_2	Peroxyacetyl nitrate	Surface PM_{10}
N_xO_y	H_2O_2	Surface $PM_{2.5}$
HNO_3	NH_3	Surface 8-h ozone
SO_2	Sulfate number	Surface 6-h ozone
H_2SO_4	Black carbon number	Surface ozone
Formaldehyde	Organic carbon number	Surface NO_x
CH_3CHO	Nitrate number	Surface SO_2
Toluene	Sulfate mass	

The specific outputs from CAMx which were fit include chemical mole fractions/concentrations of trace species in the urban area (ppm for gases and $\mu g\,m^{-3}$ for aerosols), the net mass flux of trace species through the boundaries of the urban area $kg\,day^{-1}$, the mass flux of trace species deposited to the surface $kg\,day^{-1}$, and the concentrations of 7 specific trace species, of interest to human health and policy, over the bottom three vertical layers of the urban area (under 100 m) (ppm for gases and $\mu g\,m^{-3}$ for aerosols). The simulated trace gases and aerosols were averaged over space and time to produce 24 h averaged concentrations and had their export fluxes, deposition fluxes, and chemical production terms integrated over space and time to produce 24 h total values for each of the the 17 output trace gases and 8 output aerosol species. The exception to 24 h averaging is for those species specific to specific human health input calculations, which have been averaged over the appropriate time span for each given species. The specific species simulated by the reduced form models are given in Table 5 for both the climate and human health related outputs.

The parameterizations formed in this work are based on a full third order polynomial chaos expansion, which also includes all third order cross terms, all degenerate second order terms and cross terms, and all degenerate first order terms, for all 18 input variables. The fits were performed using the probabilistic collocation technique. Since the majority of the predicted species have non-linear gas-phase, liquid phase, or heterogeneous chemical or physical processing, their concentrations on the urban scale are not accurately predictable using typical global model grid sizes. The non-linear processing, production, and transformation lifetimes are shorter than or similar to the timescales of large scale advection, mixing, and chemical processes found in typical global models. Therefore, the point of using a full third order expansion was to obtain a good fit to the highly non-linear processing.
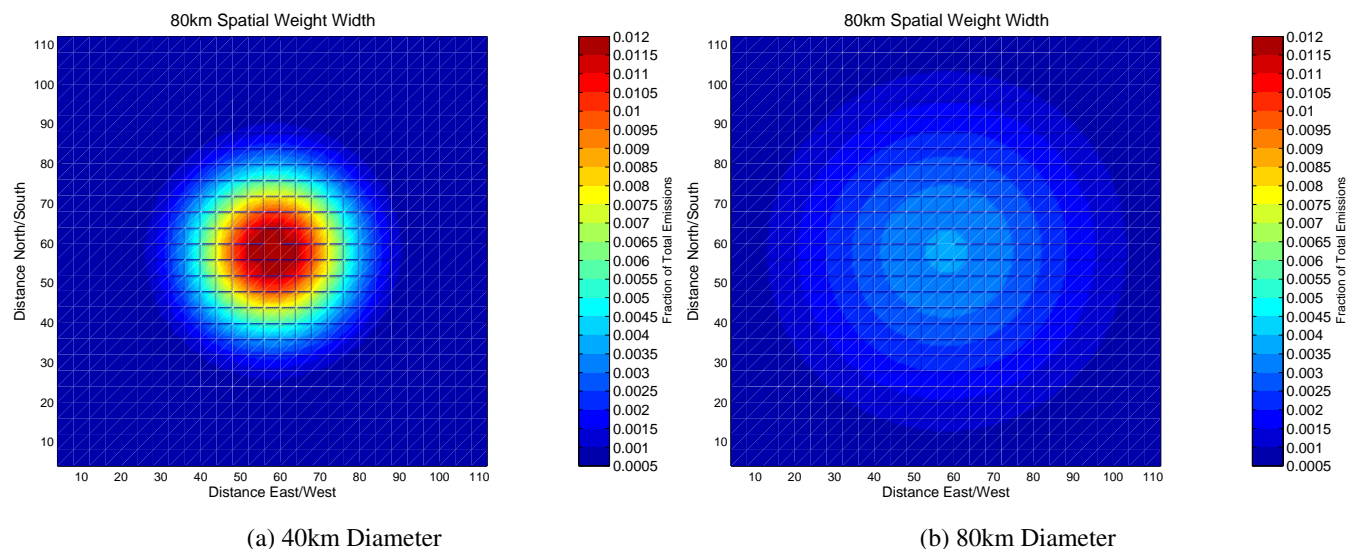
Fig. 4. Example of how the geospatial weighting of emissions impacts the distribution of urban emissions, for two example cases. Frame **(a)** corresponds to a gaussian standard diameter of 40 km and frame **(b)** to a gaussian standard diameter of 80 km. The color scale is identical for both plots.

6 Concentrations, mass fluxes, and deposition

For the parent CAMx model runs, we have examined the results for the urban concentrations, mass fluxes from the urban area, and mass fluxes deposited to the surface in the urban area, for each trace species. When looking at the results of the concentrations in mole fraction form, regions of different mole fractions must have a change in chemical production (production minus loss) of the species. Furthermore, for a species to have a mass flux from the urban area, which is not equal to its emissions, there must be a net amount of chemical processing, deposition, or convergence or divergence in the continuity Eq. (1). A convenient metric is the ratio of the mass exported from the urban areas, to the mass emitted into the urban areas:

$$\frac{\text{NetFlux}}{\text{Emiss}} = \frac{(\text{Emiss} + \text{Chem} - \text{Dep})}{\text{Emiss}} = 1 + \frac{\text{Chem}}{\text{Emiss}} - \frac{\text{Dep}}{\text{Emiss}} \quad (8)$$

Here the NetFlux is the net mass flowing out through the 5 boundaries of the urban region g day^{-1}, accounting for the total budget within the urban area; Emiss is the net mass flowing into the urban area at the surface g day^{-1}; Chem is the net mass undergoing chemical production or destruction within the urban region g day^{-1}; and Dep is the net mass deposited to the surface of the urban region g day^{-1}. Using this metric, if the net flux term is positive and larger than the emissions, then the in-situ chemical net production must be larger than the in-situ losses due to deposition. Conversely, if a species has a net flux term which is positive but smaller than its emissions, its losses due to deposition must be greater than its net chemical production. Furthermore, for a species to have a negative net flux term, then it must have an in-situ

net chemical loss large enough to consume not only all of the mass emitted at the surface, but also some of the mass transported through the boundaries into the urban area.

This net flux to emissions ratio therefore provides two efficient ways to test the validity of the model results (see Sect. 7). Firstly, any species which has a clean (zero) up-wind boundary condition must have a net flux to emissions ratio larger than or equal to zero. Secondly, a species which has no atmospheric chemical production sources in the urban area, such as Black Carbon, must have a flux to emissions ratio smaller than or equal to 1.0.

The essential test of the reliability of the reduced form model's precision and accuracy is to see how its outputs compare with the parent model. This was determined by analyzing the concentrations, mass fluxes, and surface deposition computed from running the parent model and the metamodels at all of the third and fourth order collocation points. While this test does not determine the accuracy and precision of the underlying CAMx model, we emphasize that testing of CAMx has already been carried out by many others (see Sect. 3). Scatter plots of these results, with the parent model value on the x-axis, the metamodel value on the y-axis, the ideal fit line (parent model = metamodel), and the associated RMS error, are given for BC mass, OC mass, sulfate mass, ozone, CO, formaldehyde, NO$_2$, and PAN) for each of the metamodels is given in the Supplement. The RMS and normalized RMS errors are computed using Eqs. (9)–(10), given that X_i is the value computed by the metamodel, X_{i*} is the value computed by the CAMx model, and n is the number of points analyzed.

Table 6. Normalized fractional RMS errors for species mole fractions/concentrations.

Species	China R241-F63-W46	China R000-F00-W44	China R021-F19-W57	China R002-F02-W16
BC mass	1.05×10^{-2}	1.31×10^{-2}	1.43×10^{-2}	3.05×10^{-3}
OC mass	1.05×10^{-2}	1.31×10^{-2}	1.43×10^{-2}	3.08×10^{-3}
Sulfate mass	5.23×10^{-3}	1.16×10^{-3}	1.43×10^{-3}	1.27×10^{-2}
Ozone	9.73×10^{-5}	1.23×10^{-4}	9.29×10^{-5}	3.90×10^{-4}
CO	1.06×10^{-4}	1.00×10^{-4}	1.15×10^{-4}	1.55×10^{-4}
Formaldehyde	3.77×10^{-3}	3.44×10^{-3}	2.95×10^{-3}	9.69×10^{-3}
NO$_2$	9.20×10^{-3}	6.49×10^{-3}	1.04×10^{-2}	6.94×10^{-3}
PAN	1.85×10^{-2}	1.20×10^{-2}	1.39×10^{-2}	2.12×10^{-2}

Species	India R241-F63-W46	India R000-F00-W44	India R021-F19-W57	India R002-F02-W16
BC mass	1.24×10^{-3}	4.65×10^{-3}	7.42×10^{-3}	8.46×10^{-3}
OC mass	1.20×10^{-3}	4.69×10^{-3}	7.42×10^{-3}	7.98×10^{-3}
Sulfate mass	1.88×10^{-3}	6.74×10^{-4}	3.32×10^{-3}	2.66×10^{-3}
Ozone	1.73×10^{-4}	1.42×10^{-4}	1.69×10^{-4}	3.40×10^{-4}
CO	4.49×10^{-5}	4.41×10^{-5}	4.81×10^{-5}	6.53×10^{-5}
Formaldehyde	1.91×10^{-3}	1.44×10^{-3}	1.11×10^{-3}	4.18×10^{-4}
NO$_2$	3.77×10^{-4}	3.47×10^{-4}	4.86×10^{-4}	5.10×10^{-4}
PAN	9.14×10^{-3}	3.92×10^{-3}	4.36×10^{-3}	1.07×10^{-3}

Species	Developed R241-F63-W46	Developed R000-F00-W44	Developed R021-F19-W57	Developed R002-F02-W16
BC mass	2.79×10^{-2}	3.56×10^{-3}	2.00×10^{-2}	3.02×10^{-3}
OC mass	2.74×10^{-2}	9.10×10^{-2}	1.62×10^{-1}	2.93×10^{-3}
Sulfate mass	1.08×10^{-2}	5.27×10^{-2}	8.59×10^{-2}	2.94×10^{-2}
Ozone	6.32×10^{-4}	7.09×10^{-4}	1.99×10^{-3}	1.03×10^{-3}
CO	5.86×10^{-4}	6.41×10^{-4}	1.28×10^{-3}	4.52×10^{-4}
Formaldehyde	7.12×10^{-2}	4.12×10^{-2}	1.77×10^{-2}	1.64×10^{-2}
NO$_2$	2.48×10^{-2}	2.06×10^{-2}	2.36×10^{-2}	1.53×10^{-2}
PAN	9.08×10^{-2}	9.20×10^{-2}	4.31×10^{-2}	1.39×10^{-2}

Species	Developing R241-F63-W46	Developing R000-F00-W44	Developing R021-F19-W57	Developing R002-F02-W16
BC mass	1.33×10^{-2}	5.21×10^{-3}	2.27×10^{-2}	1.43×10^{-3}
OC mass	1.32×10^{-2}	5.03×10^{-3}	2.26×10^{-2}	1.37×10^{-3}
Sulfate mass	1.45×10^{-2}	1.85×10^{-2}	3.17×10^{-2}	2.47×10^{-2}
Ozone	1.87×10^{-4}	2.69×10^{-4}	2.51×10^{-4}	6.00×10^{-4}
CO	2.66×10^{-4}	2.90×10^{-4}	2.87×10^{-4}	3.74×10^{-4}
Formaldehyde	1.36×10^{-3}	3.57×10^{-3}	3.16×10^{-3}	3.26×10^{-3}
NO$_2$	5.17×10^{-3}	6.60×10^{-4}	3.66×10^{-3}	1.90×10^{-3}
PAN	3.34×10^{-3}	1.90×10^{-3}	1.68×10^{-3}	4.09×10^{-3}

$$\text{normalized RMS} = \frac{\sqrt{\sum_{i=1}^{n} \frac{(X_{i*} - X_i)^2}{n}}}{\sqrt{\sum_{i=1}^{n} \frac{X_i^2}{n}}} \qquad (9)$$

$$\text{RMS} = \sqrt{\sum_{i=1}^{n} \frac{(X_{i*} - X_i)^2}{n}} \qquad (10)$$

The RMS error associated with the third order data points is representative of how well the metamodel corresponds to the data that was used to fit it. The error associated with the

aggregated set of third and fourth order data points is representative of how well the metamodel performs when used under realistic modeling conditions, at which the inputs are constrained only by their input PDFs. The results are given, for the same species as above, in Tables 6 and 7 for the aggregated set of third and fourth order data points.

The first conclusion is that the parent and metamodel outputs when analyzing only the third order input points, fit almost perfectly. A second conclusion is that there are a small number of outliers generated by the metamodel when the model is tested using the fourth order input points, under some model and meteorological conditions. It is these few significant outliers which contribute to almost all of the RMS

Table 7. Normalized fractional RMS errors for species mass fluxes through the boundaries of the urban area.

Species	China R241-F63-W46	China R000-F00-W44	China R021-F19-W57	China R002-F02-W16
BC mass	1.03×10^{-3}	5.82×10^{-4}	4.34×10^{-4}	6.36×10^{-4}
OC mass	1.02×10^{-3}	6.16×10^{-4}	4.53×10^{-4}	6.80×10^{-4}
Sulfate mass	7.01×10^{-3}	3.80×10^{-3}	1.59×10^{-3}	4.72×10^{-3}
Ozone	7.38×10^{-3}	5.53×10^{-3}	5.95×10^{-3}	4.91×10^{-3}
CO	1.63×10^{-4}	1.69×10^{-4}	1.69×10^{-4}	1.94×10^{-4}
Formaldehyde	5.78×10^{-3}	3.60×10^{-3}	5.23×10^{-3}	9.42×10^{-3}
NO$_2$	8.87×10^{-3}	5.83×10^{-3}	8.46×10^{-3}	6.39×10^{-3}
PAN	2.09×10^{-2}	1.27×10^{-2}	1.73×10^{-2}	2.07×10^{-2}

Species	India R241-F63-W46	India R000-F00-W44	India R021-F19-W57	India R002-F02-W16
BC mass	1.52×10^{-4}	1.82×10^{-4}	1.90×10^{-4}	1.60×10^{-4}
OC mass	1.40×10^{-4}	1.15×10^{-4}	1.30×10^{-4}	1.38×10^{-4}
Sulfate mass	5.61×10^{-4}	4.33×10^{-4}	3.41×10^{-4}	2.49×10^{-4}
Ozone	1.19×10^{-3}	1.11×10^{-3}	1.03×10^{-3}	1.07×10^{-3}
CO	1.15×10^{-4}	1.17×10^{-4}	1.22×10^{-4}	1.10×10^{-4}
Formaldehyde	2.77×10^{-3}	1.24×10^{-3}	2.09×10^{-3}	6.13×10^{-4}
NO$_2$	4.47×10^{-4}	5.53×10^{-4}	4.38×10^{-4}	7.21×10^{-4}
PAN	1.06×10^{-2}	3.02×10^{-3}	5.87×10^{-3}	1.34×10^{-3}

Species	Developed R241-F63-W46	Developed R000-F00-W44	Developed R021-F19-W57	Developed R002-F02-W16
BC mass	1.47×10^{-3}	4.64×10^{-4}	2.38×10^{-4}	1.01×10^{-3}
OC mass	7.40×10^{-3}	7.47×10^{-2}	1.27×10^{-1}	6.26×10^{-3}
Sulfate mass	1.59×10^{-2}	8.20×10^{-2}	9.70×10^{-2}	2.20×10^{-2}
Ozone	1.99×10^{-2}	1.22×10^{-2}	3.56×10^{-2}	1.16×10^{-2}
CO	7.05×10^{-4}	8.53×10^{-4}	1.57×10^{-3}	5.15×10^{-4}
Formaldehyde	4.91×10^{-2}	3.90×10^{-2}	2.37×10^{-1}	1.45×10^{-2}
NO$_2$	1.77×10^{-2}	1.16×10^{-2}	2.35×10^{-2}	1.21×10^{-2}
PAN	1.03×10^{-1}	9.26×10^{-2}	2.73×10^{-1}	1.83×10^{-1}

Species	Developing R241-F63-W46	Developing R000-F00-W44	Developing R021-F19-W57	Developing R002-F02-W16
BC mass	7.89×10^{-4}	7.50×10^{-4}	7.60×10^{-4}	1.40×10^{-3}
OC mass	7.78×10^{-4}	8.07×10^{-4}	8.02×10^{-4}	1.27×10^{-3}
Sulfate mass	6.64×10^{-3}	4.62×10^{-3}	9.83×10^{-3}	2.82×10^{-2}
Ozone	7.98×10^{-3}	6.39×10^{-3}	6.18×10^{-3}	6.00×10^{-3}
CO	6.35×10^{-4}	6.32×10^{-4}	6.35×10^{-4}	6.37×10^{-4}
Formaldehyde	1.06×10^{-3}	2.90×10^{-3}	1.82×10^{-3}	1.46×10^{-3}
NO$_2$	1.37×10^{-3}	3.14×10^{-3}	1.53×10^{-3}	3.58×10^{-3}
PAN	3.35×10^{-3}	2.06×10^{-3}	2.60×10^{-3}	5.05×10^{-3}

differences from the ideal fit. All of these outlying points are due to having at least one input which is outside of the space spanned by the third order collocation points. This is not an unexpected result.

The results of the statistical analysis of the third order collocation points are nearly perfectly consistent for the concentrations, mass fluxes, and deposition fluxes of all species, under all of meteorological conditions, using all of the metamodels. The value of the normalized RMS error is always less than 1.8×10^{-5} for the China metamodel, 1.6×10^{-5} for the India metamodel, 1.6×10^{-4} for the developed metamodel, and 4.5×10^{-5} for the developing metamodel. This shows that the metamodels behave precisely and accurately,

in relation to the parent model, at input values close to the third order collocation points. Furthermore, since the input values are based on data which is significant to only two or three decimal places, the errors are smaller than the significance levels of the input parameters, and are therefore effectively close to perfect.

The results of the analysis of the RMS errors of the concatenation of the third and fourth order collocation points is always less than 10 % under all meteorological conditions, for all metamodels, and for all quantities being modeled for ozone, CO, NO, NO$_2$, H$_2$O$_2$, BC mass, and BC number.

Sulfur is produced based on multiple pathways, leading to it being predicted less well than the above species. Under low

OH concentrations, the major pathway involves uptake by liquid water and non-linear processing in the aqueous phase, but under cases with a heightened OH concentration, the usually slow gas-phase production mechanisms can become important. Generally the daytime OH concentration is between $1.0 \times 10^7 \, molec \, cm^{-3}$ and $5.0 \times 10^7 \, molec \, cm^{-3}$, which is 1 to 2 orders of magnitude larger than the concentrations found at the global scale. The maximum OH concentration can exceed $9.0 \times 10^7 \, molec \, cm^{-3}$ under certain conditions. This is shown by the fact that the normalized RMS error for SO_2, sulfate aerosol mass, and sulfate aerosol number are less than 10 %, except in the case of the developed city metamodel. This case has an overall lower OH concentration (on order of 20 % to 50 % less), yet a larger spread in the OH concentration between different input cases, with the maximum OH concentration levels similar to the maximum in the other cases. This difference in OH is likely due to two reasons: firstly, there are significant amounts of light VOCs emitted in the Developed regions, and secondly, the Developed regions have a very different ratio of NO_x to VOC emissions. Further compounding the issue is the fact that the meteorology case which is modeled the least well (R021-F19-W57) has the most extreme horizontal and temporal gradients of cloud cover and liquid water content. Under such conditions, the numerics of the model do not fare as well when looking at the contribution to an average net flux from an urban areas, or average concentration within the urban area. As a results of these causes, the normalized RMS errors for the developed metamodel is less than 10 % for all species and all cases with the exception of meteorological scenario R021-F19-W57 which for SO_2 is less than 15 % and for sulfate aerosol number is less than 11 %.

VOCs and OC are less well predicted than the other gas phase species and BC. Part of the reason is due to the variation in the OH fields. However, the different VOC emissions profiles have a further impact, since some species have negligible emissions in the developed city metamodel cases. The resulting concentrations are so small that the fits are neither precise or relevant to the net fluxes from the urban area. Specifically, the normalized RMS errors for formaldehyde, acetaldehyde, toluene, xylene, ethene, ethane, OC mass and OC number are always less than 10 % except for the developed metamodel. The normalized RMS errors for formaldehyde and acetaldehyde are always less than 10 %, except for the deposition value in the R021-F19-W57 meteorological case, where they are less than 24 % and 14 %, respectively. The normalized RMS errors for toluene and xylene are always less than 10 %, except for meteorological case R021-F19-W57, where they are always less than 28 %. The normalized RMS errors for ethene and ethane for the developed metamodel are unacceptably large except for meteorology cases R002-F02-W16 (errors always less than 10 %) and R000-F00-W44 (errors always less than 29 %). This is because these two cases have longer air residence times in the urban area, less variation in rain, and less variation in so-

lar insulation, and so there is more chance for the chemical processing to come to something closer to a pseudo-steady state. Finally, due to the fact that some of the OC production is based on secondary processing of heavy VOCs, the normalized RMS error for OC is not as good as that of BC for the developed city metamodel. However, only in scenario R021-F19-W57 is the normalized RMS error greater than 10 %, although it is still less than 20 %. This is because in the high rain scenario, the effect of wet deposition dominates while in the low rain scenarios, secondary chemical processing is what dominates.

Ammonia is well predicted, except for the case of the developed city metamodel. Here the deposition of ammonia is imprecise because the amount of ammonia deposited is extremely small compared with the amount emitted and chemically processed. In this case, the majority of ammonia is converted into aerosol. However, since deposition accounts for only a very small amount of the loss of ammonia from the urban environment, the concentration and mass fluxes are both always modeled with a normalized RMS error of less than 10 %.

Finally, PAN is reasonably well modeled except in the developed city metamodel. This logically follows from the above analysis and the fact that PAN is based on the chemistry of both the nitrogen (NO_2) and VOC (peroxy acetyl radicals) cycles. The deposition of PAN is found to have a normalized RMS error of less than 28 %, the mass flux is found to have a normalized RMS error of less than 13 %, and the normalized RMS error of the concentration is always less than 10 %.

These results demonstrate that the meteorology plays a significant role for many species. In rainy meteorological conditions, much of the chemistry is dominated by the aqueous phase and wet removal. In dry meteorological conditions, the results are influenced by greater UV, different amounts of vertical advection, limited wet removal, and considerable dry aerosol processing. In addition, meteorological scenarios which are more variable tend to produce large spatial and temporal gradients, causing net urban variables to behave less linearly. Finally, the time scale over which the species remain in the urban area is very important, with the processing likely to be more complete, and hence easier to predict, the longer the residence time of air in the urban area.

7 Sensitivity and observational tests

The sensitivity of the response of these metamodels to different input parameters has been investigated, to determine their reliability and efficacy under many different input conditions. The results from the polynomial fits should be robust under input conditions which are in the high probability region of their distribution, based on the type of urban regions, and changes in time. To accomplish this investigation, each metamodel was run using the same set of 50 000 independently

and randomly sampled numbers chosen by selecting a random number between 0.15 and 0.85 for each input variable. This number was then used as the CDF (cumulative distribution function), thus defining a choice for the input variable. PDFs of each of the input variables so generated, for each of the different metamodels, are given in Fig. 5. This process enables testing of the metamodels at input values that evenly favor both the highly probable and less probable regions of the input variables.

One important test to tie a realistic assessment of urban areas today with the limitations imposed for this sensitivity study. Considering the world's largest 215 urban areas (those with more than 3 million people), none of the input values for emissions, background conditions, or meteorology (as derived from re-analysis data) from the period 2002–2006 fall outside of this central 70 % range used above. However, under potential future scenarios where the size of urban areas or their emissions increase too much, one may have to split single urban areas into adjacant seperate urban areas, to prevent them from growing too large and having emissions which are too high. This can be accomplished in such a way as to compute the upwind first and then use the updated background conditions to compute the downwind area second. On the opposite side, attempting to use this method to address processing of a city which is too small will possibly cause the input emissions to be too small, and in this case, the user would want to not use the metamodel in the first place.

A further test of the efficacy of the metamodels sensitivity is to compare the results with measured urban mole fractions and concentrations. Although these are not exactly the same thing, since the measured values from urban areas are usually point measurements taken near the surface, the orders of magnitude should at least be comparable. For species with a large surface source, it is expected to have a modeled volume-averaged value lower than the actual measured values, and conversely in the case of species with high destruction near the surface.
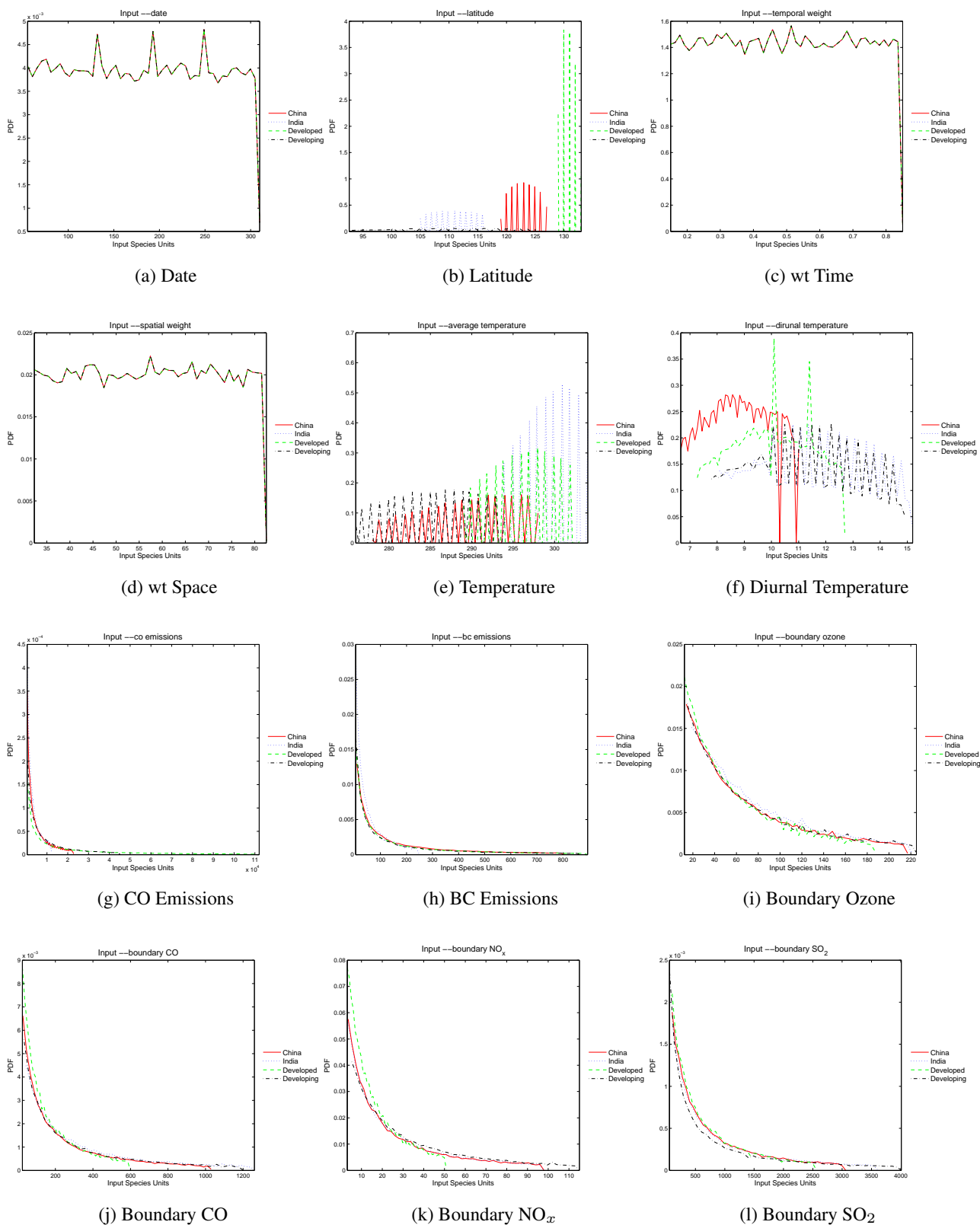
The first way in which these comparisons are made was to comparing the actual set of CAMx runs (corresponding to the points at which the metamodel exactly equals the parent model) against a set of very well characterized measurments. This has been done by comparing the lowest 100 m average urban concentrations from CAMx against long-term surface measurements from campaigns in Beijing (Huang et al., 2010) and Singapore (See et al., 2006). The reason why the lowest 100 m, daily average urban-wide concentration has been used is because it is the closest thing to a surface concentration that the metamodel computes. The reason why these two field campaigns have been choosen is that they were conducted over long and continuous periods of time, they have well characterized concentrations of aerosol species modeled by the metamodel, and that the campaigns measured under period which contain vastly different characterizations of the urban area under consideration. The comparisons have been made with the Beijing measurements for periods before

(BO), during (DO), and after (AO) the 2008 Olympic games, when a combination of different emissions, background, and meteorological scenarios produced a widely variable set of concentrations at the same measurement site. The comparisons have been made with the Singapore measurements for periods described as free from long-range forest fire smoke transport (clear) or containing long-range forest fire smoke (hazy), when a combination of different meteorological and environmental conditions were tested again at the same measurement site.

The specific way in which these comparisons were made is that the entire set of CAMx runs was analyzed so that the daily urban-average 100 m and lower concentration was computed for the following species: BC, OC, sulfate, nitrate, and ammonium. These five values were all then compared with the ranges given for the Beijing and Singapore campaigns. The CAMx model run was considered a match only if all 5 species fell within the daily average range given by the corresponding measurements. It turns out that each of the parent model runs, with the exception of the very slow wind or long residence time case, had certain input conditions which made it simultaneously simulate all 5 species, and hence to match the results. With the exception of the two periods which were characterized as being very clean (DO and clear), there were a large number of the CAMx runs which were capable of simulating the ranges described by the data. Although, as expected due to wet removal playing a large role in lowering concentrations, the two very clean times were still able to be simulated during the two rainy meteorological cases, including the only intermittent rain case. The matching sets of CAMx runs reasonably capture the spread of OC for all Beijing cases, tend to estimate BC at the high end of the measurement range and tend to estimate sulfate, nitrate, and ammonium at the low end of the measurement range. This makes sense if the site is situated within a portion of the urban area which has fewer direct emissions of BC and OC and more time to have had inorganic aerosol form. On the other hand, the matching sets of CAMx runs reasonably catpure the spread of measurements for both Singapore periods of time, with them slightly overestimating the amount of OC. The results of this comparison are summarized in Table 8, and a PDF of the normalized number of matching models on the y-axis to the modeled surface concentrations on the x-axis, for each of the species and at each time are given in Fig. 6 for the Beijing case and Fig. 7 for the Singapore case.

Now that it has been shown that the parent model, and hence the metamodel at the collocation points, can do a good job at simulating surface concentrations under variable conditions as found in Beijing and Singapore, the final test is to see how well the metamodel can simulate, on average, more poorly characterized urban areas. Six of the species that are important both on the urban and global scale, and have measurements readily available in urban areas are ozone, CO, formaldehyde, BC, OC, and sulfate, the latter

(a) Date

(b) Latitude

(c) wt Time

(d) wt Space

(e) Temperature

(f) Diurnal Temperature

(g) CO Emissions

(h) BC Emissions

(i) Boundary Ozone

(j) Boundary CO

(k) Boundary NO_x

(l) Boundary SO_2

Fig. 5. These are the PDFs of 12 of the 13 required input species for 50 000 runs of the metamodel. Each input was randomly chosen to correspond to the inverse CDF of each PDF in the range from 0.15 through 0.85. The underlying PDFs, their units, and their best fit parameters are defined in Table 3. The reason that isoprene is not included is because it is the same as the underlying PDF given in Fig. 1.

Table 8. A comparison of CAMx parent model runs that successfully simulate the mass concentrations of (BC, OC, sulfate, nitrate, and ammonium) within the corresponding measured daily average ranges, based on campaigns in Beijing (1) (Huang et al., 2010) and Singapore (2) (See et al., 2006). The range of measured values given is the spread from minimum to maximum measured daily average concentration under the conditions given under the measurement source column. The number of CAMx model runs corresponds to the total number of runs in which all 5 chemical species were simultaneously modeled within the bounds given by the measurements. The four numbers correspond to the meteorological types, in the same order as given in Table 1. The PDFs of the underlying distribution as a function of model concentration are are given in Figs. 6 and 7.

| Species	Measurement Source	Range of Measured Values	Number of CAMx Runs
BC [$\mu g\,m^{-3}$]	China BO	0.6–4.2	204,30,113,0
OC [$\mu g\,m^{-3}$]	China BO	7–45	204,30,113,0
sulfate [$\mu g\,m^{-3}$]	China BO	3–66	204,30,113,0
nitrate [$\mu g\,m^{-3}$]	China BO	2–42	204,30,113,0
ammonium [$\mu g\,m^{-3}$]	China BO	2–42	204,30,113,0
BC [$\mu g\,m^{-3}$]	China DO	0.4–2.4	2,0,4,0
OC [$\mu g\,m^{-3}$]	China DO	6–36	2,0,4,0
sulfate [$\mu g\,m^{-3}$]	China DO	2–31	2,0,4,0
nitrate [$\mu g\,m^{-3}$]	China DO	1–25	2,0,4,0
ammonium [$\mu g\,m^{-3}$]	China DO	2–19	2,0,4,0
BC [$\mu g\,m^{-3}$]	China AO	0.6–4.3	263,32,136,0
OC [$\mu g\,m^{-3}$]	China AO	7–58	263,32,136,0
sulfate [$\mu g\,m^{-3}$]	China AO	3–45	263,32,136,0
nitrate [$\mu g\,m^{-3}$]	China AO	2–28	263,32,136,0
ammonium [$\mu g\,m^{-3}$]	China AO	2–21	263,32,136,0
BC [$\mu g\,m^{-3}$]	Singapore Clear	0.28–6.24	5,1,2,0
OC [$\mu g\,m^{-3}$]	Singapore Clear	0.24–7.63	5,1,2,0
sulfate [$\mu g\,m^{-3}$]	Singapore Clear	0.20–16.69	5,1,2,0
nitrate [$\mu g\,m^{-3}$]	Singapore Clear	0.06–2.97	5,1,2,0
ammonium [$\mu g\,m^{-3}$]	Singapore Clear	0.04–1.55	5,1,2,0
BC [$\mu g\,m^{-3}$]	Singapore Hazy	0.48–10.44	382,556,852,0
OC [$\mu g\,m^{-3}$]	Singapore Hazy	1.84–23.05	382,556,852,0
sulfate [$\mu g\,m^{-3}$]	Singapore Hazy	2.97–16.42	382,556,852,0
nitrate [$\mu g\,m^{-3}$]	Singapore Hazy	0.06–12.23	382,556,852,0
ammonium [$\mu g\,m^{-3}$]	Singapore Hazy	0.30–4.85	382,556,852,0

three contributing to the total urban particulate mass concentration. The results for the concentrations of each of these species is shown in Table 9 respectively for the China, India, Developed, and Developing metamodels, where the statistics of the 50 000 runs are compared with the statistics of the measurements on the same plots. In general, the results of this sensitivity analysis compare reasonably well with the actual measured mole fractions and concentrations of ozone, CO, formaldehyde and BC in each simulated area.

The maximum ozone concentrations predicted by the metamodel are slightly higher than observed for the maximums, while the medians predicted by the metamodel compare reasonably well with measured median concentrations. These metamodel medians and maxima are larger than the observed mean monthly average value of 23 ppb and maximum monthly average of 110 ppb found in the Guadala-

jara, Mexico urban area (Ramirez et al., 2009). However, the metamodel medians are quite reasonable when compared with the annual column averaged concentration measurements (from the surface to 3000 m) found over Beijing, New York, Paris, and Tokyo, which respectively are 49–53 ppb, 44 ppb, 40 ppb, and 47 ppb (Ding et al., 2008). Furthermore, since these are annual averages, higher concentration events in the upper part of the boundary layer, especially during the summer months, are less likely to be observed, as shown in the adjacent measurements for summertime averaged ozone over Beijing (Ding et al., 2008). As expected, the model results are higher than the measured values, because the model is giving the total vertically averaged ozone concentration in the urban area, which includes the upper regions which have no surface deposition, less titration by fresh NO emissions, and further time for the NO_2/NO ratio to increase.
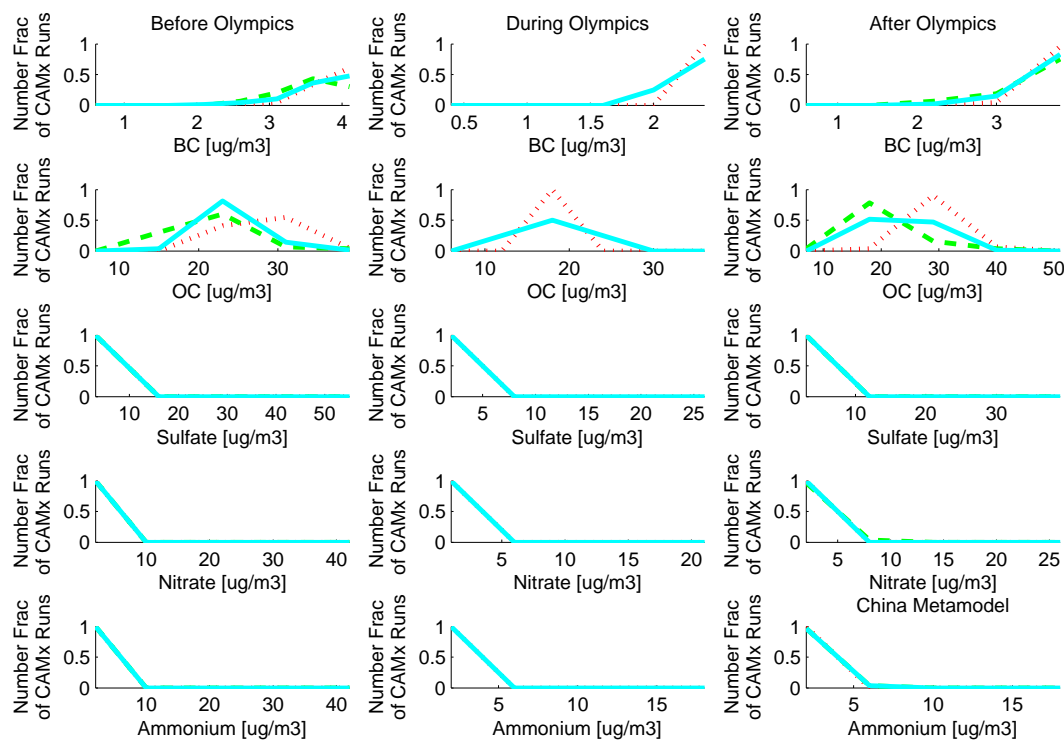
Fig. 6. PDFs of the number of CAMx runs matching the entire set of concentrations for the Beijing site. The x-axis is the concentration of each respective species (BC, OC, sulfate, nitrate, and ammonium), during each time period considered at the Beijing site (BO, DO, and AO), while the y-axis is the normalized number count in terms of the number of CAMx runs matching all 5 species (Table 8). The different colors refer to the different meteorologies for each metamodel.
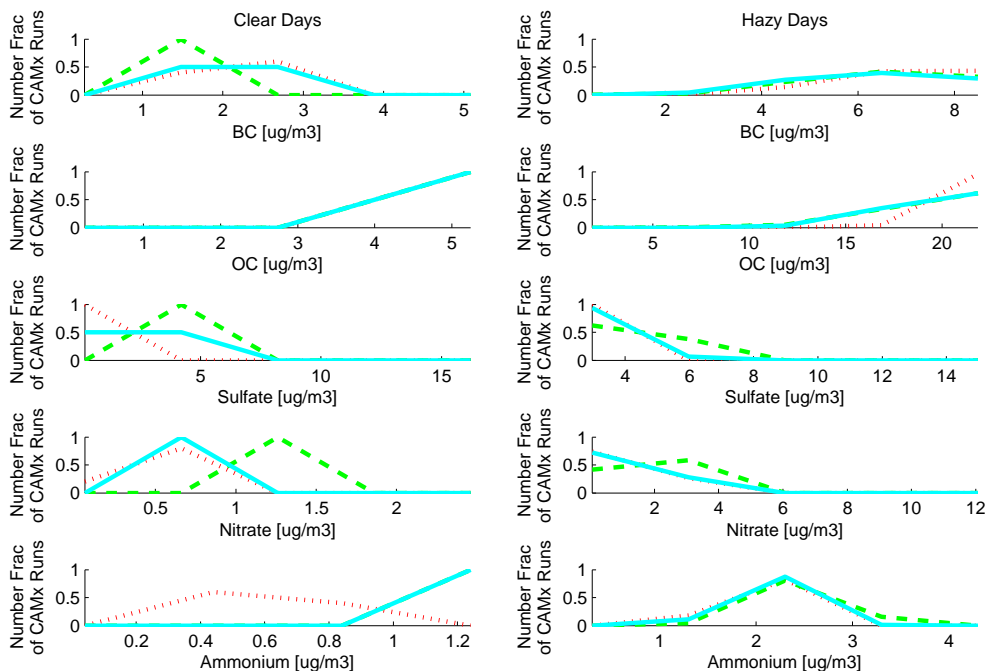
Fig. 7. PDFs of the number of CAMx runs matching the entire set of concentrations for the Singapore site. The x-axis is the concentration of each respective species, during each time period considered at the Singapore site, while the y-axis is the normalized number count in terms of the number of CAMx runs matching all 5 species (Table 8). The different colors refer to the different meteorologies for each metamodel.

Table 9. Results of metamodel simulated urban area concentration medians (50 % of results are above this value and 50% of results are below this value) and maxima, for each meteorology type, based on 50000 randomly selected input parameters. Where mentioned in the text, measurement values are included in parenthesis, and the citations are given according to the following list: [a] Ding et al. (2008), [b] Ramirez et al. (2009), [c] Dutta et al. (2009), [d] Lei et al. (2009), [e] Huang et al. (2010), [f] Badarinath et al. (2009), [g] Beegum et al. (2009), [h] Dutkiewicz et al. (2009), [i] Husain et al. (2007), [j] See et al. (2006), [k] Godoy et al. (2009).

Species	Metamodel	Range of Metamodel Median Values	Range of Metamodel Maximum Values
Ozone [ppb]	China	88.4–113 (49–53[a])	326–418
Ozone [ppb]	India	118–154	311–436
Ozone [ppb]	Developed	58.4–84.3 (40–47[a])	293–1520
Ozone [ppb]	Developing	108–129 (33[b])	381–437 (110[b])
CO [ppm]	China	0.265–0.705	1.82–5.64
CO [ppm]	India	0.240–0.620	1.12–3.40
CO [ppm]	Developed	0.522–1.55	8.49–27.9
CO [ppm]	Developing	0.452–1.34 (1.9[b])	3.86–12.8 (9.2[b])
Formaldehyde [ppb]	China	0.332–1.73	6.82–18.1
Formaldehyde [ppb]	India	0.685–5.53 (10–19[c])	9.40–38.5 (46[c])
Formaldehyde [ppb]	Developed	0.746–107	23.5–250
Formaldehyde [ppb]	Developing	0.968–10.7 (4–9[d])	10.1–129 (35[d])
BC [μg m^{-3}]	China	2.60–20.1 (0.4–4.3[e])	27.8–208
BC [μg m^{-3}]	India	1.24–9.20 (12[f], 5[g], 10[h], 2[h], 1[h], 5–21.7[i])	8.97–74.1 (35[g], 45[g], 15[h], 110[i])
BC [μg m^{-3}]	Developed	2.19–16.3 (0.28–10.44[j])	29.6–229
BC [μg m^{-3}]	Developing	2.14–17.5 (1.4–3.3[k])	35.8–258
OC [μg m^{-3}]	China	4.22–32.8 (6–58[e])	45.2–337
OC [μg m^{-3}]	India	1.98–16.0	14.3–113
OC [μg m^{-3}]	Developed	5.75–28.8 (0.24–23.05[j])	77.0–4630
OC [μg m^{-3}]	Developing	8.53–73.6	140–1010
Sulfate [μg m^{-3}]	China	3.17–24.6 (2–66[e])	89.3–468
Sulfate [μg m^{-3}]	India	4.10–28.0	15.9–798
Sulfate [μg m^{-3}]	Developed	3.53–18.2 (0.20–16.69[j])	27.6–835
Sulfate [μg m^{-3}]	Developing	4.40–30.9	80.7–335

The order of magnitude of the CO concentrations predicted by the metamodel is reasonable, with the ranges of the maximum and median concentrations also given in Table 9. These results are somewhat lower than the average observed monthly average concentration of 1.9 ppm and the observed maximum monthly average concentration of 9.2 ppm found in the Guadalajara urban area (Ramirez et al., 2009). As expected, the modeled values are slightly lower than the measured values, since the background CO concentration is usually lower than the surface CO concentration due to surface emissions.

The formaldehyde concentration predicted by the metamodel shows ranges of the maximum and median concentrations that are given in Table 9. The metamodel results are low or comparable to (a) the observed average monthly concentration of 4–9 ppb and observed maximum average monthly concentration of up to 35 ppb in Mexico City (Lei et al., 2009), and (b) the observed mean daily values of 10–19 ppb and maximum daily value of 46 ppb for Kolkata India (Dutta

et al., 2009). The modeled concentrations are expected to be lower than the surface measurements, both because the background formaldehyde in the parent model run is zero, and surface emissions enhance the surface concentrations.

The ranges of the maxima and medians of the BC mass concentration, as predicted by the metamodels are shown in Table 9. These values are generally lower than the August 2006 average BC concentration in Hyderabad, India of 12 μg m^{-3} (Badarinath et al., 2009); a March to May monthly average high and low concentration of BC in Hyderabad of 5–35 μg m^{-3} and in Delhi, India of 5–45 μg m^{-3} (Beegum et al., 2009); a November 2006 to February 2007 monthly average BC concentration in Karachi, Pakistan of 10 μg m^{-3}, a June to September monthly average Karachi BC concentration of 2 μg m^{-3}, and a daily mean Karachi BC concentration in the range from 1–15 μg m^{-3} (Dutkiewicz et al., 2009); a Lahore, Pakistan average Winter BC concentration of 21.7 μg m^{-3}, with the range over any given day from 5–110 μg m^{-3} (Husain et al., 2007); and a Rio de Janeiro mean

annual BC concentration of 1.4–$3.3\,\mu g\,m^{-3}$ (Godoy et al., 2009). Since the background boundary concentration of BC in the model is assumed to be zero, and since the maximum concentrations are near the surface source, the average BC concentration in the model should be lower than the measured values.

The ranges of the maxima and medians of the OC and sulfate mass concentration, as predicted by the metamodels for each meteorological scenario are further shown in Table 9. Adding the mass concentrations of BC, OC, and sulfate provide the total anthropogenic component of the PM value as provided by the metamodel. Although this is not the equivalent to PM, since it does not include anthropogenic nitrate aerosol, or non-anthropogenic aerosols such as dust, sea salt, and forest fire aerosols, it can provide a lower bound estimation, particularly in areas which do not have large natural upwind sources of aerosols. These metamodel values are lower than the observed total PM_{10} monthly average concentration of $51\,\mu g\,m^{-3}$ and a maximum monthly average concentration of $265\,\mu g\,m^{-3}$ found in the Guadalajara urban area (Ramirez et al., 2009).

Furthermore, the Flux to Emissions ratio for each of these species has been investigated and compared with the ratios expected to result from the chemical and physical processing of the species. The graphs of these ratios are given in Fig. 8 for the China metamodel, Fig. 9 for the India metamodel, Fig. 10 for the Developed metamodel, and Fig. 11 for the Developing metamodel, for BC, OC, sulfate aerosol, CO, NO_2, and formaldehyde. The median value of the ratio in the case of CO is close to 1.0 in all metamodel cases, indicating that the effects of deposition and chemistry are not significant when compared with emissions, over the 24 h timescale of the urban metamodel run. This means that a significant fraction of the VOC may not be fully oxidized to CO by the time it has been exported from the urban area, since the deposition of CO is negligible.

The Flux to Emissions ratio for formaldehyde can be used to investigate the extent of the VOC emissions oxidized before being exported from the urban area. In the case of formaldehyde there are a few competing factors. First, in cases of large rainfall or cloudiness, less formaldehyde is produced through photochemistry and more is removed through wet deposition. In these cases, the median value of the ratio is found to be small, often under 0.2, whereas in cases of little or no rainfall and thin or no cloud cover, the median of the ratio, is up around 0.4. Second, in the cases of low molecular weight VOC emissions typical in developed urban regions, the median value of the ratio can be as high as 0.8 depending on the meteorology, due to the more rapid oxidation of light VOCs into formaldehyde.

The Flux to Emissions ratio for NO_2 is useful for determining what the expected ozone production will be downwind from the urban area. Since only 5 % of the emissions of NO_x is in the form of NO_2, any ratio which is larger than 0.05 indicates an increase in the export of NO_2, with respect

to the simple dilution approach. The results show that the median value of the ratio actually ranges from 0.1 to 0.4, depending on the meteorology scenario.

The Flux to Emissions ratio for BC should be and always is in the range from 0.0 to 1.0. It is also a very strong function of the amount of rainfall, with the median value being as low as 0.35 in the case of high rainfall and the median value being as high as 0.95 in the case of no rainfall. This further indicates that dry deposition is much less important than wet deposition in the case of BC.

The Flux to Emissions ratio for OC should always be the same as or larger than BC, since the sources of OC are emissions, which are correlated linearly to those of BC, plus a small amount of secondary production due to oxidation and condensation of high molecular weight VOCs. This results in the median ratio of OC always being about equal to that of BC in the cases of high rainfall (in which case these low vapor pressure secondary products are also more efficiently removed, and less efficiently oxidized due to a lessened actinic flux), and from 1 % to 5 % higher in the case of low rainfall.

Finally, the Flux to Emissions ratio for SO_2 should always be larger than or equal to 0.0, indicating how much of the gas is converted to sulfate aerosol as a result of urban processing. The ratio is a less strong function of the rainfall for SO_2 than for BC and OC. This results from the fact that an important production mechanism for sulfate aerosol requires the presence of liquid water (although the same removal mechanisms are at play for all three aerosol types). What is particularly interesting is that the median of the SO_2 ratio in the case of the India metamodel is from 0.05 to 0.10 larger than in the other three metamodel cases, for each of the meteorological scenarios. This is caused at least in part by the higher average temperature in Indian cities increasing the oxidation efficiency of SO_2 in both the gas and aqueous phases.

8   Conclusions

We have developed and tested a computationally efficient metamodel approach for analyzing the impact of the nonlinear physical and chemical processing of the emissions of gases and aerosols in urban regions worldwide, which can be incorporated into a global chemical model. Since the majority trace gas and aerosol emissions occur in hundreds of urban centers, and since the effects of their processing have both local and global impacts on the direct and indirect radiative forcing of the climate, this is an important problem to address. Our reduced form metamodels have been formed to simulate these effects for urban areas in diverse geographical regions, for multiple realistic meteorological conditions, and for a range of human-induced patterns and distributions of anthropogenic emissions. These metamodels are specifically designed to efficiently simulate the urban concentration, surface deposition, chemical production, and

(a) BC EF

(b) OC EF

(c) Sulfate EF

(d) CO EF

(e) Formaldehyde EF

(f) NO$_2$ EF
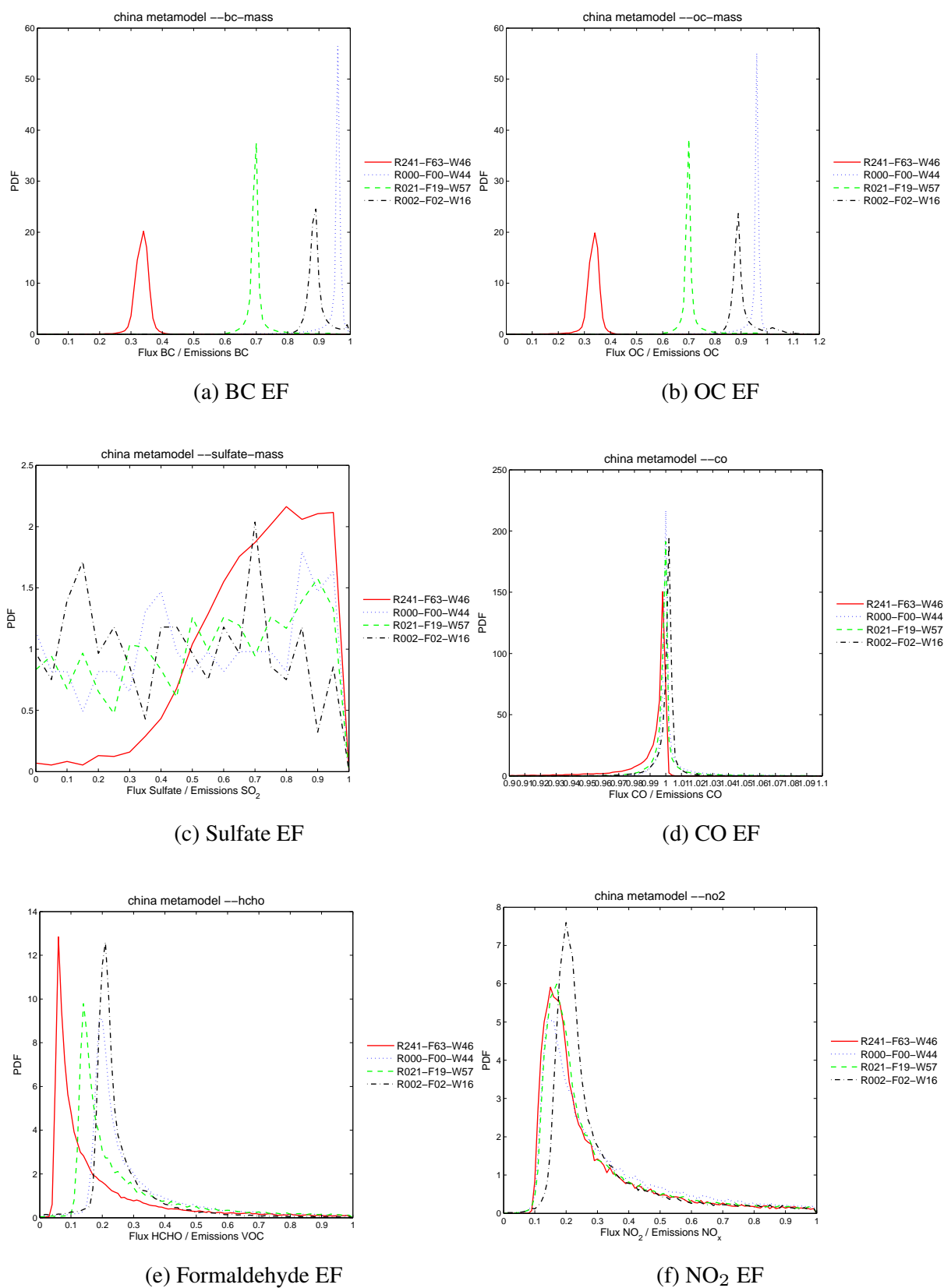
Fig. 8. These are the PDFs of the total urban area EFs (Eq. 8), from the China metamodel. For inputs, please refer to Fig. 5.

(a) BC EF

(b) OC EF
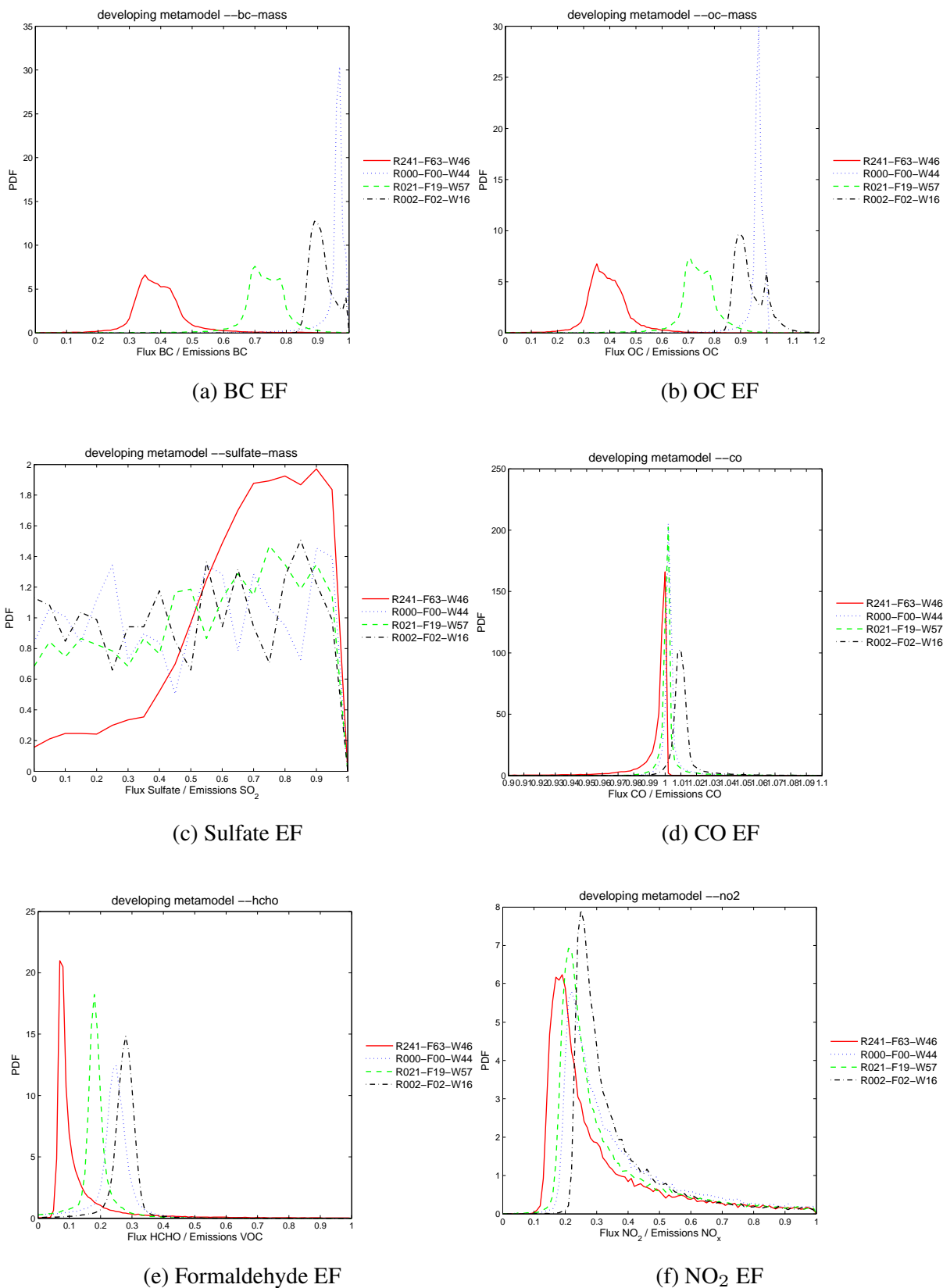
(c) Sulfate EF

(d) CO EF

(e) Formaldehyde EF

(f) NO$_2$ EF
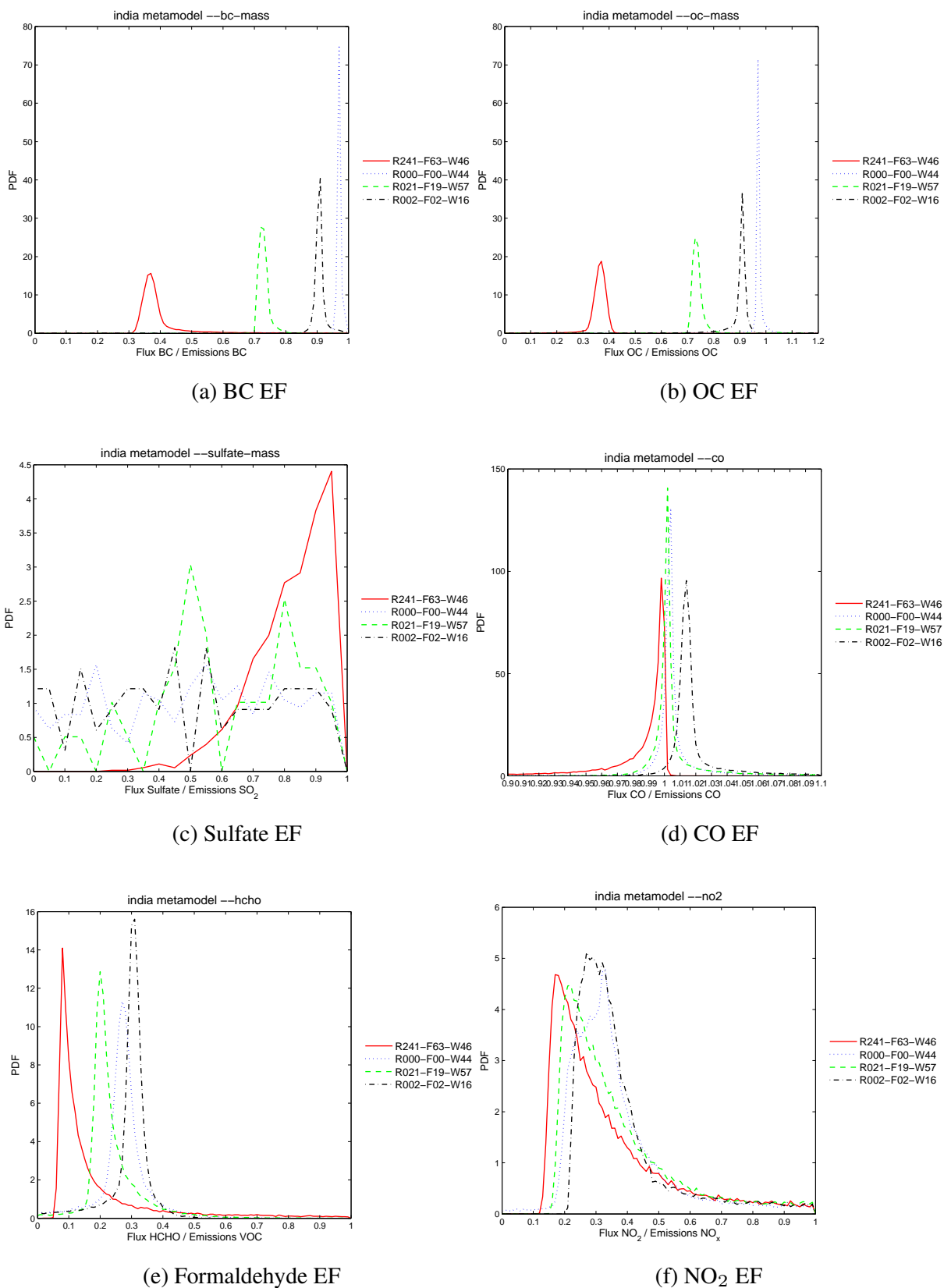
Fig. 9. These are the PDFs of the total urban area EFs (Eq. 8), from the India metamodel. For inputs, please refer to Fig. 5.

(a) BC EF

(b) OC EF

(c) Sulfate EF

(d) CO EF

(e) Formaldehyde EF

(f) NO$_2$ EF
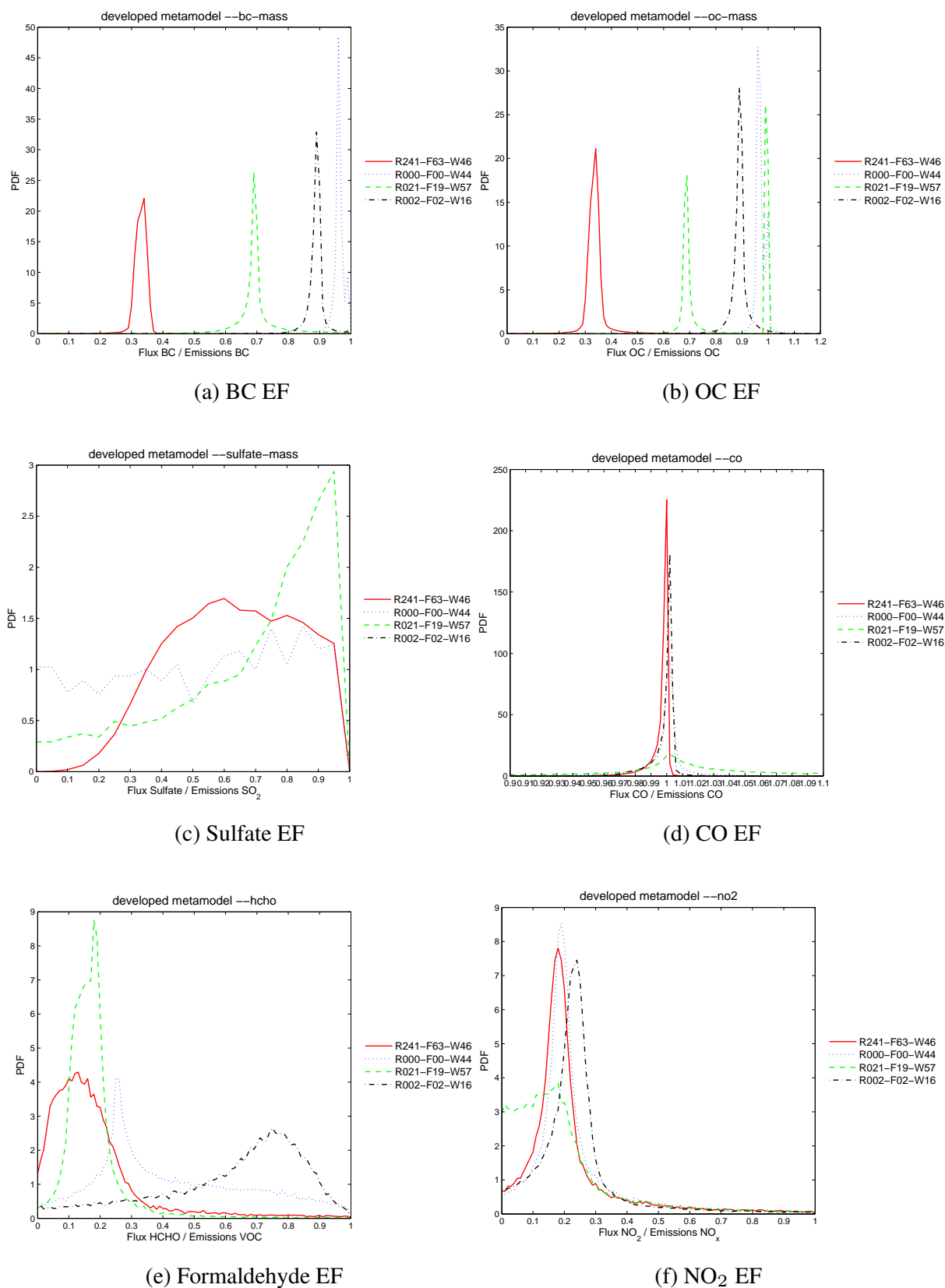
Fig. 10. These are the PDFs of the total urban area EFs (Eq. 8), from the Developed metamodel. For inputs, please refer to Fig. 5.

(a) BC EF

(b) OC EF

(c) Sulfate EF

(d) CO EF

(e) Formaldehyde EF

(f) NO$_2$ EF

Fig. 11. These are the PDFs of the total urban area EFs (Eq. 8), from the Developing metamodel. For inputs, please refer to Fig. 5.

net mass flux of species important to human health and climate.

These metamodels have been designed to be fast enough, so that they can be integrated into a global scale modeling platform and thus better capture the concentrations and mass fluxes found in real urban areas, as compared with the process of dilution of urban emissions into much larger grid volumes that the large-scale models currently use. Polynomial chaos expansions have been used to create and fit these metamodels based on a broad range of conditions, applicable both to the present and hypothetical future world, so that they remain applicable both for current conditions and for studies which look at the chemistry of urban areas in the future. The various outputs are based on a set of 18 inputs, which include physical properties, such as the local temperature, daily diurnal temperature range, the day of the year, and the geographic location. Secondly, they include the anthropogenic properties of urban emissions, such as the temporal and spatial weighting of emissions, and the magnitude of emissions of many relevant anthropogenic species. Thirdly, the remaining inputs include the upwind (or background) concentrations of trace species, both anthropogenic and natural in nature, which have an impact on the processing in the urban area. These inputs have been gathered from multiple sources, and used to generate a set of PDFs of their potential values in current and future urban areas. In the final application, these inputs will come from the global model grid boxes containing the urban area.

These PDFs were then used to create the chaos polynomials in a recursive process and to determine inputs at a set of thousands of collocation points at which to run the detailed parent chemical and physical model, CAMx. Another set of higher order points were also generated from these input PDFs, which were used to test the fit to the parent CAMx model. An additional benefit derives from the fact that using a higher order polynomial fit requires that more specific points be sampled from each input PDF. Hence the metamodel is more precise for more extreme values of the input, while still allowing for the polynomial fit to be more reliable near the center of the input PDFs.

The deviations between the metamodel and the parent model were computed in terms of a normalized RMS error. Many important species, such as ozone, CO, NO_x, and BC were found to have a normalized RMS error less than 10 % for all of the metamodels, under all meteorological conditions, with many of the species having a normalized RMS error less than 1 %. Some of the other important species, such as VOCs, PAN, OC, and sulfate aerosol are usually fit well, except for a few meteorological cases in just one of the metamodel regions, in which they are fit less well. This is associated with the highly non-linear source and sink processes of these chemical species, and the geographic areas and meteorological scenarios of choice. The reason for the less good fit in each of these cases is largely explained in terms of the complexity of this physical, chemical, and meteorological processing. Therefore, using third order fits is a significant improvement when compared with previous efforts, which only used second order fits.

A set of sensitivity and observational tests have been performed to compute the response and accuracy of the various metamodels to a very broad set of potential inputs different from those used to produce the fits. The point of this testing was to determine if the metamodel could accurately handle multiple inputs from low probability regions of the input PDFs, and this was generally determined to be the case. Furthermore, these results were compared with observations of ozone, CO, formaldehyde, BC, and PM_{10} from a few urban areas where they were available. In most of the cases, the output distributions were found to be similar to the observations, especially given that the metamodel predicts average urban concentrations, and not point surface measurements for these species.

There are some effects, however, which are not included but could be important to address in the future. The inclusion of meteorology for regions which have complex topography (such as Mexico City or Chongqing) or for regions located in Equatorial latitudes (such as Singapore), would further provide a more realistic treatment in these urban regions. Furthermore, inclusion of realistic meteorology as a function of different seasons of the year would also provide a more realistic treatment of urban processing with respect to annual cycles. Other improvements could include dynamical feedbacks such as: consideration of urban scale topography, online calculation of absorption and scattering, and urban hydrological processes.

Overall, it appears that these metamodels can efficiently and robustly simulate the urban concentrations, mole fractions, and fluxes of species, important to human health and the global scale climate. Hopefully, further investigations in this general area could lead to improvements of our understanding of the emissions of gasses, aerosols, and their precursors; the processing of aerosols at both large and urban grid scales; the modeling of non-linear effects that occur on scales smaller than the grid-scale; and the impact that subgrid-scale urban processes have on the larger atmosphere.

Supplementary material related to this article is available online at: http://www.atmos-chem-phys.net/11/7629/2011/ acp-11-7629-2011-supplement.pdf.

References

Badarinath, K., Kharol, S., Reddy, R., Gopal, K., Narasimhulu, K., Reddy, L., and Kumar, K.: Black carbon aerosol mass concentration variation in urban and rural environments of India – a case study, Atmos. Sci. Lett., 10, 29–33, doi:10.1002/asl.206, 2009.

Beegum, S., Moorthy, K., Babu, S., Satheesh, S., Vinoj, V., Badarinath, K., Safai, P., Devara, P., Singh, S., DUC, V., and Pant, P.: Spatial distribution of aerosol black carbon over India during pre-monsoon season, Atmos. Environ., 43, 1071–1078, doi:10.1016/j.atmosenv.2008.11.042, 2009.

Butler, T. and Lawrence, M.: The influence of megacities on global atmospheric chemistry: a modelling study, Environ. Chem., 6, 219–225, 2009.

Calbo, J., Pan, W., Webster, M., Prinn, R., and McRae, G.: Parameterization of urban subgrid scale processes in global atmospheric chemistry models, J. Geophys. Res.-Atmos., 103, 3437–3451, 1998.

Center for International Earth Science Information Network (CIESIN), Columbia University; and Centro Internacional de Agricultura Tropical (CIAT): Gridded Population of the World Version 3 (GPWv3): Population Grids. Palisades, NY: Socioeconomic Data and Applications Center (SEDAC), Columbia University, available at: http://sedac.ciesin.columbia.edu/gpw, 2005.

Chang, J. S., Brost, R. A., Isaksen, I. S. A., Madronich, S., Middleton, P., Stockwell, W. R., and Walcek, C. J.: A 3-Dimensional Eulerian Acid Deposition Model – Physical Concepts and Formulation, J. Geophys. Res.-Atmos., 92, 14681–14700, doi:10.1029/JD092iD12p14681, 1987.

Cossa, P.: Uncertainty analysis of the cost of climate policies, Master's thesis, Massachusetts Institute of Technology, http://dspace.mit.edu/handle/1721.1/30074, 2004.

Ding, A. J., Wang, T., Thouret, V., Cammas, J.-P., and Nédélec, P.: Tropospheric ozone climatology over Beijing: analysis of aircraft data from the MOZAIC program, Atmos. Chem. Phys., 8, 1–13, doi:10.5194/acp-8-1-2008, 2008.

Dutkiewicz, V., Alvi, S., Ghauri, B., Choudhary, M., and Husain, L.: Black carbon aerosols in urban air in South Asia, Atmos. Environ., 43, 1737–1744, doi:10.1016/j.atmosenv.2008.12.043, 2009.

Dutta, C., Som, D., Chatterjee, A., Mukherjee, A., Jana, T., and Sen, S.: Mixing ratios of carbonyls and BTEX in ambient air of Kolkata, India and their associated health risk, Environ. Monit. Assess., 148, 97–107, doi:10.1007/s10661-007-0142-0, 2009.

Gery, M., Whitten, G., Killus, J., and Dodge, M.: A photochemical kinetics mechanism for urban and regional scale computer modeling, J. Geophys. Res.-Atmos., 94, 12925–12956, 1989.

Godoy, M., Godoy, J., Roldao, L., Soluri, D., and Donagemma, R.: Coarse and fine aerosol source apportionment in Rio de Janeiro, Brazil, Atmos. Environ., 43, 2366–2374, doi:10.1016/j.atmosenv.2008.12.046, 2009.

Houweling, S., Dentener, F., and Lelieveld, J.: The impact of non-methane hydrocarbon compounds on tropospheric photochemistry, J. Geophys. Res.-Atmos., 103, 10673–10696, 1998.

Huang, X.-F., He, L.-Y., Hu, M., Canagaratna, M. R., Sun, Y., Zhang, Q., Zhu, T., Xue, L., Zeng, L.-W., Liu, X.-G., Zhang, Y.-H., Jayne, J. T., Ng, N. L., and Worsnop, D. R.: Highly time-resolved chemical characterization of atmospheric submicron particles during 2008 Beijing Olympic Games using an Aerodyne High-Resolution Aerosol Mass Spectrometer, Atmos. Chem. Phys., 10, 8933–8945, doi:10.5194/acp-10-8933-2010, 2010.

Husain, L., Dutkiewicz, V., Khan, A., and Ghauri, B.: Characterization of carbonaceous aerosols in urban air, Atmos. Environ., 41, 6872–6883, doi:10.1016/j.atmosenv.2007.04.037, 2007.

Jones, P., New, M., Parker, D., Martin, S., and Rigor, I.: Surface air temperature and its changes over the past 150 years, Rev. Gephys., 37, 173–199, 1999.

Koo, B. Y., Ansari, A. S., and Pandis, S. N.: Integrated approaches to modeling the organic and inorganic atmospheric aerosol components, Atmos. Environ., 34, 4757–4768, doi:10.1016/j.atmosenv.2003.08.016, 2003.

Lei, W., Zavala, M., de Foy, B., Volkamer, R., Molina, M. J., and Molina, L. T.: Impact of primary formaldehyde on air pollution in the Mexico City Metropolitan Area, Atmos. Chem. Phys., 9, 2607–2618, doi:10.5194/acp-9-2607-2009, 2009.

Mayer, M., Wang, C., Webster, M., and Prinn, R.: Linking local air pollution to global chemistry and climate, J. Geophys. Res.-Atmos., 105, 22869–22896, 2000.

Nenes, A., Pandis, S. N., and Pilinis, C.: Continued development and testing of a new thermodynamic aerosol module for urban and regional air quality models, Atmos. Environ., 33, 1553–1560, doi:10.1016/S1352-2310(98)00352-5, 1999.

Paltsev, S., Reilly, J., Jacoby, H., Eckaus, R., McFarland, J., Sarofim, M., Asadoorian, M., and Babiker, M.: The MIT emissions prediction and policy analysis (EPPA) Model: Version 4, The MIT Joint Program on the Science and Policy of Global Change Report Series 125, Massachusetts Institute of Technology, http://globalchange.mit.edu/files/document/MITJPSPGC_Rpt125.pdf, 2005.

Prather, M., Ehhalt, D., Dentener, F., Derwent, R., Dlugokencky, E., Holland, E., Isaksen, I., Katima, J., Kirchhoff, V., Matson, P., Midglet, P., and Wang, M.: Atmospheric chemistry and greenhouse gases, in: Climate change 2001: The scientific basis. Contribution of working group I to the third assessment report of the Intergovernmental Panel on Climate Change, Cambridge University Press, 2001.

Qian, Y., Gustafson Jr., W. I., and Fast, J. D.: An investigation of the sub-grid variability of trace gases and aerosols for global climate modeling, Atmos. Chem. Phys., 10, 6917–6946, doi:10.5194/acp-10-6917-2010, 2010.

Ramirez, S., Hermes, U., Andrade, G., Maria, D., Bejaran, R., Garcia, G., Mario, E., Wallo, V., Pompa, T., Ana, C., and delaTorreVillasenor, O.: The spatial-temporal distribution of the atmospheric polluting agents during the period 2000–2005 in the Urban Area of Guadalajara, Jalisco, Mexico, J. Hazard. Mater., 165, 1128–1141, doi:10.1016/j.jhazmat.2008.10.127, 2009.

Russell, M. and Allen, D.: Predicting secondary organic aerosol formation rates in southeast Texas, J. Geophys. Res.-Atmos., 110, 7S17–7S30, 2005.

Sarwar, G., Luecken, D., Yarwood, G., Whitten, G., and Carter, W.: Impact of an updated carbon bond mechanism on predictions from the CMAQ modeling system: Preliminary assessment, J. Appl. Meteorol. Clim., 47, 3–14, doi:10.1175/2007JAMC1393.1, 2008.

See, S. W., Balasubramanian, R., and Wang, W.: A study of the physical, chemical, and optical properties of ambient aerosol particles in Southeast Asia during hazy and nonhazy days, J. Geophys. Res.-Atmos., 111, 0148–0227,

doi:10.1029/2005JD006180, 2006.

Sokolov, A., Schlosser, A., Dutkiewicz, S., Paltsev, S., Kicklighter, D., Jacoby, H., Prinn, R., Forest, C., Reilly, J., Wang, C., Felzer, B., Sarofim, M., Scott, J., Stone, P., Melillo, J., and Cohen, J.: MIT integrated global system model (IGSM) version 2: Model description and baseline evaluation, The MIT Joint Program on the Science and Policy of Global Change Report Series 124, Massachusetts Institute of Technology, 2005.

Strader, R., Lurmann, F., and Pandis, S. N.: Evaluation of secondary organic aerosol formation in winter, Atmos. Environ., 33, 4849–4863, doi:10.1016/S1352-2310(99)00310-6, 1999.

Tatang, M., Pan, W., Prinn, R., and McRae, G.: An efficient method for parametric uncertainty analysis of numerical geophysical models, J. Geophys. Res.-Atmos., 102, 21925–21932, 1997.

Vukovich, F.: PAMS data analysis: An investigation of local meteorological effects on ozone during the OTAG 1995 episode and the weekday/weekend differences in the Northeast Corridor, Tech. rep., United States Environmental Protection Agency, http://capita.wustl.edu/otag/reports/pamsinv2/pamsinv2.html, 1997.

Wild, O. and Prather, M.: Global tropospheric ozone modeling: Quantifying errors due to grid resolution, J. Geophys. Res.-Atmos., 111, 11305–11318, doi:10.1029/2005JD006605, 2006.

Yang, K., Ting, C., Wang, J., Wingenter, O., and Chan, C.: Diurnal and seasonal cycles of ozone precursors observed from continuous measurement at an urban site in Taiwan, Atmos. Environ., 39, 3221–3230, doi:10.1016/j.atmosenv.2005.02.003, 2005.

Yokouchi, Y.: Seasonal and diurnal variation of isoprene and its reaction products in a semirural area, Atmos. Environ., 28, 2651–2658, 1994.