

## Parameterization of urban subgrid scale processes in global atmospheric chemistry models

Josep Calbó,<sup>1</sup> Wenwei Pan,<sup>2</sup> Mort Webster, Ronald G. Prinn, and Gregory J. McRae

Massachusetts Institute of Technology Joint Program on the Science and Policy of Global Change, Cambridge

**Abstract.** We have derived a parameterization consisting of a set of analytical expressions that approximate the predictions by the California Institute of Technology - Carnegie-Mellon University (CIT) Urban Airshed Model for the net export to the environment (i.e., effective emissions) of several chemical species, as functions of 14 input parameters. For each species, effective emissions are a function of actual urban emissions of this and other species and of other urban domain properties such as meteorology. Effective emissions may be "aged" emissions of primary pollutants or actual production of secondary pollutants. To develop the parameterization we have applied the probabilistic collocation method, which uses the probability density functions of the inputs to generate a set of orthogonal polynomials. These polynomials are then used as the basis for a polynomial chaos expansion that approximates the actual response of the CIT model to its inputs. We assume that seasonal variations can be represented by sinusoidal functions. The parameterization provides a computationally very efficient simulation of the actual model behavior. We have compared the outputs of the parameterization with the outputs of the CIT model, and we conclude that it gives a quite good approximation for effective emissions, at least in the regions of highest probability of the input parameters. This parameterization is applicable to detailed uncertainty and sensitivity analyses and enables computationally efficient inclusion of urban-scale processes as sub-grid scale phenomena in global-scale models.

### 1. Introduction

Atmospheric pollution has a wide range of temporal and spatial scales. Climate change associated with changes in greenhouse gases is a global long-term problem, with characteristic space and timescales about  $10^4$  km (Earth diameter) and  $10^9$  s (30 years). In contrast, characteristic scales of urban air pollution are  $10^2$  km and 1 day ( $10^5$  s).

Usually, the climate change and air pollution issues are studied independently, and little attention is focused on interactions and relationships between them. For example, most scientific and policy analyses of the greenhouse gas issue do not consider the likely influence of local air pollution processes and controls on global emissions, or on the cost of alternative greenhouse gas control measures. This lack of consideration of local pollution occurs in part because the climate science and policy issues are great analytical challenges by themselves; it also occurs in part because the analytical models used for assessment of the greenhouse issue cannot resolve the space and timescales of urban pollution due to computational constraints. Failure to consider the interaction between global and urban aspects of anthropogenic emissions may be a serious shortcoming in current analyses because these two aspects are linked in at least three important ways: through the

chemistry of the atmosphere, through the role of combustion in each, and through the potential interaction of local pollution policies with greenhouse gas control measures.

This paper focuses on the first of these links: specifically, the way in which the global and local aspects of atmospheric pollution are linked through tropospheric chemistry. For example, the chemistry that leads to urban photochemical smog (involving  $\text{NO}_x$ , hydrocarbons,  $\text{CO}$ ,  $\text{O}_3$  and UV light) also involves important greenhouse gases (methane and ozone). Also, aerosols resulting from  $\text{SO}_x$  and  $\text{NO}_x$  emissions contribute both to urban pollution and acid rain and to changes in the Earth's radiative balance.

Specifically, we have developed a parameterization or "reduced-form" model that allows us to estimate the effective emissions ( $EE$ ) from an urban area as functions of the urban true emissions ( $E$ ) and local meteorology ( $M$ ). The relationship between  $EE$  and  $E$  is complex because the high concentrations of gases and aerosols found in polluted urban areas, combined with nonlinearities (e.g., reactions of  $\text{NO}_x$  or  $\text{HO}_x$  species with themselves) and interactions (e.g., reaction of  $\text{NO}_x$  species with  $\text{HO}_x$  species or heterogeneous reactions on aerosols), lead to different rates of production and removal when the true emissions are confined to urban areas compared with artificially spreading the same true emissions evenly over a (usually very much larger) grid cell of a global chemistry and climate model. A similar issue is addressed in a paper by Gillani and Pleim [1996]. They studied the importance of subgrid scale effects in regional models and concluded that chemical nonlinearities lead to important differences when these effects are taken into account.

To include urban processes in a global model, we need to compute  $EE$  day-by-day for, say, the 70 urban areas of the world that will have more than 3 million inhabitants by year

<sup>1</sup>Now at Departament d'Enginyeria Industrial, Universitat de Girona, Girona, Spain.

<sup>2</sup>Now at Lawrence Livermore National Laboratory, Livermore, California.

2000 [World Health Organization-United Nations Environment Program (WHO-UNEP), 1993]. Therefore the urban process subroutine in the global model has to be integrated over more than 3 million ( $70 \times 365 \times 120$ ) days in a 120 year run of a coupled chemistry and climate model such as that of R. Prinn et al. (Integrated global system for climate policy assessment: Feedbacks and sensitivity studies, submitted to Climatic Change, 1997) (hereinafter referred to as Prinn et al. (submitted manuscript, 1997)). This effectively prevents the direct use of any of the presently available three-dimensional urban models in a global model, and it is the impetus for the parameterization discussed here.

The California Institute of Technology-Carnegie-Mellon University Urban Airshed Model [McRae et al., 1982] is the basis upon which we have built our parameterization. The CIT Urban Airshed Model can deal directly with the physical and chemical phenomena that lead to the aging of pollutants within an urban domain, since the model was designed specifically to simulate these processes. This model requires hourly fields of several meteorological variables, hourly emissions of several chemical species, initial conditions and data related to the domain. A typical 1 day integration for a domain discretized in 10,000 cells requires about 1 hour of time on a fast workstation. As noted above, we cannot use directly such a model to carry out the 3 million day integrations required in the global model.

Our parameterization of the urban model uses the probabilistic collocation method [Tatang, 1994] which produces analytical expressions relating model outputs to model inputs thus enabling uncertainty analysis of computationally expensive models at a very low computational cost.

The next section (section 2) describes the two basic components used for derivation of the parameterization: the CIT Urban Airshed Model and the collocation method. We discuss in the third section the selection of inputs and outputs and development of parameterization itself. In the fourth section, a comparison between the results obtained using the true model and the parameterization is given, followed by a sensitivity and uncertainty analysis for the parameterization. Finally, conclusions and proposals for further work are given in section 5.

## 2. Basic Components

### 2.1. CIT Urban Airshed Model

The CIT Urban Airshed Model is a photochemical air quality model, designed for predicting the spatial and temporal distribution of both inert and chemically active pollutants. This air quality model has been described in detail [McRae et al., 1982, 1992; Harley et al., 1993] and has been applied in several studies [e.g., Milford et al., 1989; Instituto Mexicano del Petroleo and Los Alamos National Laboratory (IMP-LANL), 1994]. The model solves the species continuity equations using the operator splitting method. Horizontal Cartesian and vertical terrain-following coordinates are used to discretize the domain. Diffusion is handled with explicit second-order finite differences. Horizontal advection is treated using implicit fourth-order finite element integration with spatial filtering, and vertical advection is solved by means of a first-order upstream finite difference scheme. Turbulent mixing is parameterized by applying the  $K$  theory, and turbulent diffusion coefficients are functions of the Monin-Obukhov length

and surface roughness height. Dry deposition is modeled through pollutant deposition velocities, which are also functions of the Monin-Obukhov length.

The photochemical reaction formulation employed at present in the CIT model is based on the work by Lurmann et al. [1987]. This formulation is focused on fast reactions between nitrogen oxides and volatile organic compounds that drive ozone production. Transformations from  $\text{NO}_x$  to nitrous and nitric acids, and from  $\text{SO}_2$  to  $\text{SO}_3$  (i.e., from oxides to aerosol precursors) are also considered. In contrast to a global model, the chemistry of longer-lived species, such as methane, is not included. Thirty-five species (single species or lumped classes) and 106 reactions are included. Sixteen of these species are treated as directly emitted species, and 14 of them are used to set initial and boundary conditions (see Table 1). Both area and point source emissions can be handled. Hourly, grid-resolved emission data must be provided to the model. Boundary conditions, given by the combination of surface and vertical profiles of concentrations, can also have an hourly resolution if needed.

Hourly three-dimensional fields of several meteorological variables must be provided: wind, humidity, temperature, mixing layer height, and two radiation scaling factors (one for the whole solar radiation spectrum, the other for the UV light). These fields may either be derived from observed surface and sounding data or supplied by a meteorological model. The CIT Urban Airshed Model has also been coupled interactively to the U.K. Meteorological Office Mesoscale Model [Golding, 1992] to produce the Integrated Meteorological and Air Quality System (IMAQS) [Pan, 1996]. For two reasons, the development of our parameterization is based on the original CIT Urban Airshed Model. First, many mesoscale meteorological phenomena are forced by the underlying topography, and our desire for a computationally efficient parameterization prevented us from specifically including topography or land use. Thus the power of the mesoscale meteorological model would not have been well utilized. Second, in order to develop the parameterization, the full model has to be run some hundreds of times and the CIT Urban Airshed Model is several times faster than the IMAQS model.

### 2.2. Collocation Method

The probabilistic collocation method [Tatang, 1994; Tatang et al., 1997] was a tool designed originally to simplify uncertainty analysis of computationally expensive models. The method is one of several mathematical techniques for reducing a model to a simpler form. The probabilistic collocation method approximates the response of the model by a polynomial function of the uncertain inputs. Suppose we have a "black box" model described by

$$y = f(x_1, \dots, x_k, \dots, x_K) \quad (1)$$

where  $x_k$  is one of the  $K$  input parameters to the model  $f$ , and  $y$  is the output (or "response" to the inputs) of the model. Since there is a range of possible values for each input parameter, these inputs can be treated as random variables. Thus the output of the model can also be considered as a random variable. The goal of the collocation method is to approximate the  $K$  dimensional "response surface"  $y$  by a simple polynomial function of  $x_k$ . The polynomials used by the collocation technique are orthogonal polynomials  $P_k^j(x_k)$ , defined by

$$\begin{aligned}
 P_k^{-1}(x_k) &= 0 \\
 P_k^0(x_k) &= 1 \\
 \int_{x_k} g_k(x_k) P_k^i(x_k) P_k^j(x_k) dx_k &= C \delta_{ij}, \forall i, j
 \end{aligned}
 \tag{2}$$

where  $g_k(x_k)$  is the probability density function (PDF) of the independent random variable  $x_k$ ,  $\delta_{ij}$  is the Kronecker delta, and  $C$  is a real number (if  $C=1$  the polynomials are orthonormal). For example, for a Gaussian PDF the  $P_k^j$  are Hermite polynomials. The polynomial definition procedure is implemented using ORTHPOL [Gautschi, 1994].

Using these polynomials, the independent random variables can be written as

$$x_k = x_k^0 + x_k^1 P_k^1(x_k)
 \tag{3}$$

and the dependent variable  $y$  can be approximated by a polynomial chaos expansion (PCE):

$$y = y_0 + \sum_{i_1, i_2, \dots, i_K}^{1 \leq i_1 + i_2 + \dots + i_K \leq N} y^{i_1 i_2 \dots i_K} P_1^{i_1}(x_1) P_2^{i_2}(x_2) \dots P_K^{i_K}(x_K)
 \tag{4}$$

where  $y_0, y^{i_1 i_2 \dots i_K}$  are coefficients to be determined, and  $N$  is the order of the approximation. By running the actual model a number of times we obtain the required number of sets of  $(x_k, y)$  values to solve for  $y_0$  and all  $y^{i_1 i_2 \dots i_K}$ .

The orthogonal polynomials, as noted above, are used in the polynomial chaos approximation of the model. Equally important, they are also used to generate the sets of parameter values  $x_k$  (the so-called collocation points) for solving for the approximation through (4) [Webster et al., 1996]. Specifically, the  $N+1$  roots of the  $(N+1)$ th order polynomial corresponding to each parameter  $x_k$  are used to define the collocation points.

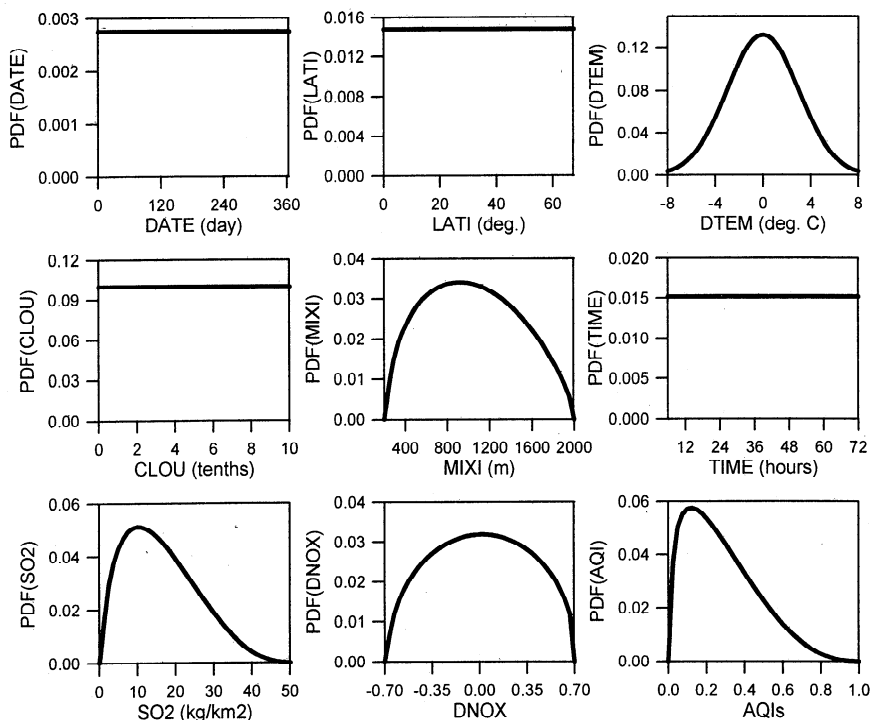
Because these roots help define the high probability region of each input parameter, we obtain an approximation of  $y$  that is particularly good within the most probable range of values of the input parameters. Moreover, the roots of the  $(N+2)$ th order polynomials are used to define another set of collocation points which can be used to estimate the error of the approximation which, again, takes account of the actual probability distribution of the parameters. The approximation defined by (4) (which we refer to here as the "parameterization") is therefore designed to perform in the region of highest probability of the inputs. Versions of (4) are developed for each output variable ( $y$ ) of interest.

### 3. Methodology

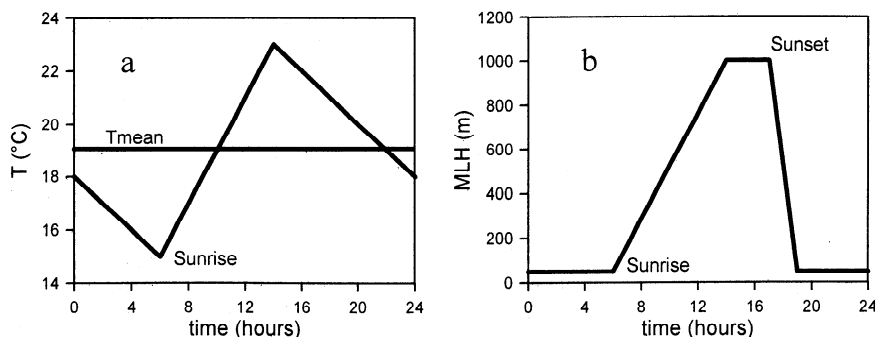
#### 3.1. Selected Inputs

The CIT Urban Airshed Model requires potentially a large amount of input data to run. To develop our numerically efficient parameterization, we have restricted the total number of input values to 14. The criteria for choosing these inputs, the range of their possible values and PDFs, and procedures for mapping single value inputs to spatial or temporal variation where required, are discussed below.

**3.1.1. Time (DATE).** The solar radiation flux driving photochemistry depends on astronomical factors (time and location) as well as atmospheric factors (reflection, absorption and scattering). The effect of time is taken into account through specifying the date in Julian days (DATE) starting with 0 on January 1, and ending with 364 on December 31. This input date (along with input latitude) is also used to define the mean air temperature for the site. The probability distribution for DATE is, of course, uniform (i.e., all days in the year are equally probable, Figure 1).



**Figure 1.** Probability density functions (PDFs) for most input parameters defined in the text. The PDF for CO is the same as that for SO2 but ranges between 0 and 500 kg/km<sup>2</sup>. The PDF for DVOC is the same as for DNOX. The PDFs for all air quality indices (AQI) are the same.



**Figure 2.** (a) Example of the assumed daily cycle of temperature. In this case, the mean temperature is 19°C. (b) Example of the assumed daily cycle of the mixing layer height (MLH). In this example, the maximum height (i.e., the parameter MIXI) has been set to 1000 m. Note that sunrise and sunset depend on latitude and date in both cases.

**3.1.2. Location (LATI).** Location is specified by the latitude of the region. Although the distribution of major cities is not uniform with latitude (most of the largest cities are placed between 20° and 50° in both hemispheres), we have also for convenience defined a uniform PDF for this parameter (Figure 1). The selected range of values for this input variable (LATI) is from 0° to 67.5° north and south since there are no major cities outside of this range. The parameterization which is formally derived for the northern hemisphere can be used for the southern hemisphere cities simply by changing the date by 6 months. This ignores the small effect of the Earth's orbital eccentricity which presently allows more sunlight incident on the southern hemisphere than the northern hemisphere.

The longitude of the urban area is not one of the selected input parameters. The reason is that the dominant effect of longitude is the difference between local time and universal time which is accounted for by having all the runs carried out in local time.

**3.1.3. Air temperature (DTEM).** Air temperature is important since rate constants for many reactions depend on temperature. Hourly fields of surface temperature are required by the CIT model. Temperature is calculated as the sum of three terms. The first is the climatic average temperature appropriate to the season and latitude of the domain which is computed using the formula proposed by *North and Coakley* [1979] for a symmetrized Earth:

$$T(\varphi, d) = T_0 - T_1 \sin(\varphi) \cos \omega(d - 32) - T_2 \left( \frac{3}{2} \sin^2(\varphi) - \frac{1}{2} \right). \quad (5)$$

Here  $T(\varphi, d)$  is the mean (in the climatic sense) temperature for latitude  $\varphi$  and day  $d$  ( $d=0$  at December 22),  $T_0=14.9^\circ\text{C}$  is the global mean temperature,  $T_1=15.5^\circ\text{C}$ ,  $T_2=28.0^\circ\text{C}$ , and  $\omega$  is the frequency related to the annual cycle. The root mean square error of this expression, relative to actual temperature measurements, is only 2°C. The expression is convenient, since it uses two existing inputs (latitude and date) and it preserves the symmetry between hemispheres.

To take account of variability around the climatic mean, we input a parameter, DTEM, which is the difference between actual temperature and the climatic mean given by (5). The PDF corresponding to this input is a normal distribution consistent with its definition (Figure 1).

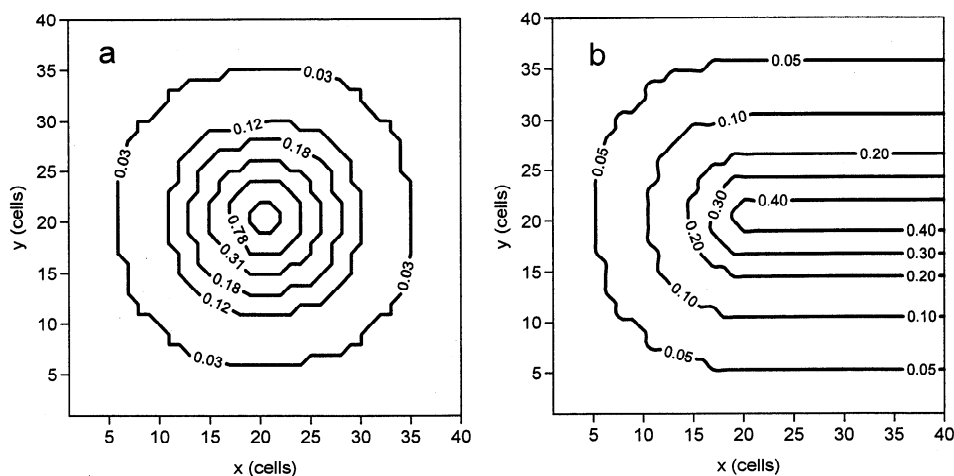
Adding the temperature from (5) and DTEM yields a daily mean temperature. To represent the daily cycle, we assume an

oscillation of 8°C with the minimum at sunrise and the maximum 2 hours after noon and linear variations between these extreme values (Figure 2). While varying with time, temperatures are assumed to be uniform in space through the urban domain.

Another atmospheric input required by the CIT model is hourly fields of specific humidity. Moisture content is chemically important since it plays a role in some gaseous and aqueous phase reactions, although no aqueous phase reactions are considered in the model runs discussed in this paper. In this study we have not included specific humidity as an input parameter. Based on some previous tests, we have instead assumed relative humidity to be the same for all runs and equal to 60%. From this relative humidity and daily average temperature, we compute the equivalent specific humidity (in g/kg). The specific humidity is then considered constant for that run, unless at the cooler nighttime temperatures it would imply a relative humidity higher than 100%. Then, specific humidity is set to the saturation level. As with the other meteorological fields, the humidity field is assumed uniform in space.

**3.1.4. Cloud cover (CLOU).** The CIT model requires inputs of hourly values for the scaling factors for total and ultraviolet solar fluxes. These fluxes are related to the altitude of the region and to the optical characteristics of the atmosphere. In this study we approximate the UV scaling factor by equating it to the scaling factor for total solar radiation. This latter factor can be estimated from cloud cover data using the expression proposed by *McRae et al.* [1992]. Cloud cover in tenths of overhead sky area (CLOU) is therefore one of the chosen inputs. The derived solar total and UV radiation scaling factors are assumed constant throughout the run and uniform in space. Possible values for cloud cover range from 0 to 10 tenths. The probability distribution function assumed is uniform (see Figure 1).

**3.1.5. Mixing layer height (MIXI).** Turbulence plays an important role in dispersion of an urban plume of pollutants. The CIT model parameterizes the turbulent characteristics of the atmosphere (using the turbulent diffusivities) based on solar radiation, wind, and most importantly, mixing layer height. We express the latter height in terms of the input parameter MIXI, defined as the maximum mixing layer height during the day. Mixing layer height has, like temperature, a daily cycle, which has been approximated, as shown in Figure 2. At night (from 2 hours after sunset to sunrise), the mixing



**Figure 3.** Examples of assumed spatial distributions of emissions and initial concentrations. (a) Emissions of CO, at 12 hours, in  $\text{ppm m}^{-1} \text{min}^{-1}$  (these are the units used by the CIT model to express emissions into the domain; they refer to the increase of concentration in the first computational layer per unit of vertical length and per unit of time) when the input parameter CO is  $100 \text{ kg km}^{-2}$ . If the domain is  $200 \times 200 \text{ km}^2$ , the urban area covers a circle of  $150 \text{ km}$  diameter, and downtown (where 50% of emissions are released) occupies a circle of  $40 \text{ km}$  diameter. (b) Initial concentrations of CO, in ppm, when the input parameter AQVOC is 0.1.

layer height is set equal to the larger of MIXI/20 or 50 m. The mixing layer height then increases linearly from sunrise to its maximum value (MIXI) 2 hours after noon, where it then remains until sunset. Finally, it decreases linearly to its nighttime height in the 2 hours following sunset. Like other atmospheric fields, a uniform space distribution has been assumed. The probability distribution of MIXI is defined by a beta distribution ranging between 200 and 2000 m. The distribution (Figure 1) is asymmetrical, and the most probable value of MIXI is about 800 m.

**3.1.6. Residence time (TIME).** The CIT model requires specification of the size of the urban region and the wind velocities and direction. However, given the assumed symmetry of our urban region (defined below) and uniformity of atmospheric fields, the time that pollutants reside inside the domain is the major chemically important parameter determined by the wind velocity. Hence we input the residence time, TIME, defined as the width of the urban region divided by the wind velocity, instead of the velocity separately. The probability distribution for TIME is uniform, between 6 hours and 3 days. For an area of  $200 \text{ km}$  side, that means a range of winds between 0.7 and  $10 \text{ m/s}$ .

When running the CIT model to generate the parameterization, a square region of  $200 \text{ km}$  side is assumed. It is discretized into  $40 \times 40 = 1600$  grid cells. The urban region is assumed flat and a radially symmetric central emitting region (the city) covers almost half of the area (Figure 3). Different surface characteristics (in particular, surface roughness) have been defined for each part of the domain. The vertical structure is the same for all runs: the top is fixed at  $2090 \text{ m}$  height, and there are five layers below that with varying depth. The height of the top corresponds approximately to the maximum mixing layer heights typical of the continental environments of many large cities.

The CIT model needs hourly fields of three-dimensional winds. For development of the parameterization, information regarding wind is reduced to one single value: the mean (in space and time) wind speed in the urban area defined as the domain width ( $200 \text{ km}$ ) divided by TIME. The wind in any

single model run is considered constant in time and uniform in space, no vertical wind is assumed, and wind direction is the same for all the runs. Although there is an expected daily variation of wind speed, related to the solar and boundary forcing and topography, it is domain specific and has not been taken into account. The assumptions of zero vertical wind and horizontally uniform wind are consistent with a flat terrain. Wind direction effects are not considered, because they have a minor contribution relative to wind speed, particularly when all the other relevant fields (terrain, emissions, boundary conditions, solar radiation, etc.) are so symmetric (Figure 3). Wind direction for all runs has arbitrarily been set perpendicular to one side of the CIT model domain.

**3.1.7. Emission of sulfur oxides (SO<sub>2</sub>).** The CIT model requires hourly fields of emissions of several pollutants (see Table 1), and it can distinguish between point sources and area sources. We have assumed for all pollutants that we have area sources only distributed throughout the emitting city area in the domain. The city area is defined as a circle with a diameter  $3/4$  of the width of the domain. Inside this circular area, we have defined six circular bands with different emissions in each of them. The two central ones (with diameters  $1/10$  and  $2/10$  of domain width) account for 50% of total emissions. The following two (with diameters  $3/10$  and  $4/10$  of domain width) account for 25% of total emissions. The most exterior rings (with diameters  $5/10$  and  $7.5/10$  of domain width) account for the residual 25% of emissions. In terms of emissions per unit area, the ratios are 1, 0.31, 0.125, 0.075, 0.05 and 0.0125 for each circular band relative to the central part of the city (see Figure 3).

There is, of course, a wide range of emission patterns. For example, for coastal cities, the pattern is more linear, or, at least, does not have radial symmetry. We have checked the sensitivity of the CIT model results to different assumed patterns of emissions. Specifically, we have defined an elongated emitting region (to simulate a coastal city) and run the CIT model for this emission pattern with the city axis either perpendicular or parallel to the wind direction. The differences between these two runs and the equivalent circular city are

**Table 1.** Main Species Included in the CIT Model

Main Species	Directly Emitted Species	Initial and Boundary Conditions
CO	yes	yes
NO	yes	yes
NO <sub>2</sub>	yes	yes
SO <sub>2</sub>	yes	yes
O <sub>3</sub>		yes
N <sub>2</sub> O <sub>5</sub>		
HNO <sub>3</sub>		
OH		
HO <sub>2</sub>		
H <sub>2</sub> O <sub>2</sub>		
Formaldehyde	yes	yes
Acetaldehyde	yes	yes
Methyl ethyl ketone	yes	yes
Alkanes (≥C <sub>4</sub> )	yes	-a
Ethene	yes	-a
Alkenes (≥C <sub>3</sub> )	yes	-a
Toluene	yes	-a
Higher aromatics	yes	-a
Methanol	yes	
Ethanol	yes	
NH <sub>3</sub>	yes	
SO <sub>3</sub>	yes	
PAN		-b
NO <sub>3</sub>		

<sup>a</sup> Initial conditions (IC) and boundary conditions (BC) for these volatile organic compounds are defined using a single number that accounts for several reactive hydrocarbons (RHC).

<sup>b</sup> IC and BC for PAN are inferred from O<sub>3</sub> concentrations, as follows: [PAN]=0.01[O<sub>3</sub>].

shown in Figure 4. For some outputs, differences may reach 80%, but the average difference is around 20% in both cases. Although these differences are noticeable, they are not larger than the errors in the derived parameterizations discussed later. Therefore the circular city approximation is considered sufficient for our purposes.

The input parameter SO<sub>2</sub> corresponds to the total mass of SO<sub>2</sub> emitted per day divided by the area of the domain. Five detailed city emission inventories have been checked to define a typical range of values for this parameter (see Table 2). These values are consistent with rough estimations of emissions in several other cities of the world [WHO-UNEP, 1993]. We have used a beta distribution, with limits of 0 and 50 kg/km<sup>2</sup>/d and a maximum probability value of 10 kg/km<sup>2</sup>/d (Figure 1) to define the PDF of SO<sub>2</sub>. The value of 10 kg/km<sup>2</sup>/d corresponds to emission of 333 kg/km<sup>2</sup>/d in the most central circular band.

Although some sulfur oxides are emitted by vehicles, the main sources of this pollutant are industries, especially power plants, which generally operate on a continuous basis. Therefore only a small amplitude daily cycle can be expected for the SO<sub>2</sub> emissions and we therefore assume these emissions are constant in time. In addition, emission of SO<sub>3</sub> is assumed to be 2% of the emission of SO<sub>2</sub>.

**3.1.8. Emission of carbon monoxide (CO).** We define the input CO as the total daily emissions of carbon monoxide divided by the domain area. The range of possible values has also been derived after investigation of several emission inventories (see Table 2). A beta distribution bounded by 0 and 500 kg/km<sup>2</sup>/d, with a most probable value of 100 kg/km<sup>2</sup>/d is

used to define the PDF for this parameter (Figure 1). Emission of CO follows the same assumed spatial pattern as SO<sub>2</sub> (Figure 3).

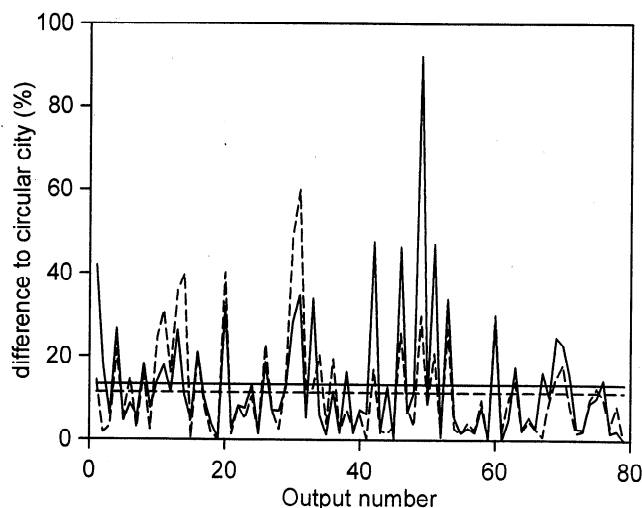
Carbon monoxide is mainly emitted by moving vehicles. Therefore its emissions show a strong daily cycle that follows the pattern of the traffic cycle. Several emission inventories for developed cities (listed in Table 2) have been studied and used to define a typical daily cycle of CO. This typical cycle has high emissions between 6 hours and 18 hours (LT) and maximum emissions at 15 hours. We realize the limitations of applying this cycle derived from developed cities to cities in developing countries, or of applying the same cycle to countries as different as the United States and Spain as far as traffic schedules are concerned. However, we believe that it is important foremost to include the fact that higher emissions occur simultaneously with higher radiation levels, and the cycle we have chosen reflects this basic correlation.

**3.1.9. Emissions of volatile organic compounds (DVOC).** Similar to other pollutants, emissions of several volatile organic compounds (VOC) need to be input into the CIT model on an hourly basis. Emissions of VOC have several possible sources, with road traffic being the main one in cities. Therefore we can expect that total VOC emissions are highly correlated with CO emissions. We have verified this correlation using the emission inventories that we have available (Table 2). The linear regression relating total VOC emissions  $E_{VOC}^{CO}$  to CO emissions  $E_{CO}$  is:

$$E_{VOC}^{CO} = 0.32 E_{CO} + 0.98. \quad (6)$$

However, the actual amount of VOC emitted in a specific city may be different from  $E_{VOC}^{CO}$ . We therefore define the parameter DVOC, which accounts for the difference between the actual total VOC emissions,  $E_{VOC}$ , and the emissions derived from (6):

$$E_{VOC} = (1 + DVOC) E_{VOC}^{CO}. \quad (7)$$



**Figure 4.** Differences between CIT model results when we approximate an elongated urban area by the circular area described in the text: solid line, the elongated domain is perpendicular to wind direction; dashed line, the elongated domain is parallel to wind direction. Horizontal lines show mean error for all the outputs in both cases.

**Table 2.** Emissions per Unit Area in Several Urban Areas

Urban Area	Reference	CO	SO <sub>2</sub>	NO <sub>x</sub>	VOC
Los Angeles	<i>Harley et al.</i> [1993]	101	1.9	13.1	26.5
Upper Rhine	<i>Schneider et al.</i> [1996]	46	8.9	17.3	25.1
Manchester	<i>Longhurst et al.</i> [1996]	15	6.8	5.6	6.1
Barcelona	<i>Costa and Baldasano</i> [1996]	220	19.5	58.6	80.2
Melbourne	G. Adamkiewicz (personal communication, 1996)	138	2.7	18.7	36
Correlation to CO emission			0.55	0.89	0.95

Units in kg/km<sup>2</sup>/d.

For the PDF of DVOC, we have chosen a symmetric beta distribution between -0.7 and 0.7 (Figure 1).

We allocate total VOC emissions among 10 different VOC species (Table 1) using factors derived from the emission inventory for Los Angeles, California [*McRae et al.*, 1992] and consistent also with those derived for Melbourne, Australia (G. Adamkiewicz, unpublished data, 1996) and Mexico City [*IMP-LANL*, 1994]. The temporal evolution and spatial pattern of VOC emissions are assumed to follow the same pattern as CO.

**3.1.10. Emissions of nitrogen oxides (DNOX).** The same procedure used to define VOC emissions has been applied to define emissions of nitrogen oxides because emissions of NO<sub>x</sub>, like VOC, are highly correlated with emissions of CO (Table 2). The NO<sub>x</sub> emissions ( $E_{NO_x}^{CO}$ ) derived from  $E_{CO}$  using linear regression are

$$E_{NO_x}^{CO} = 0.23 E_{CO}. \quad (8)$$

As for VOC, a parameter, DNOX, is defined to take account of differences between actual emissions and  $E_{NO_x}^{CO}$ . DNOX has the same PDF as DVOC. Ninety-five percent of NO<sub>x</sub> emission (by mass) is assumed to be NO, while the other 5% is assumed to be NO<sub>2</sub>. NO<sub>x</sub> emissions follow the same spatial and temporal distributions as CO and VOC.

*Sillman et al.* [1990] noted that the emission ratio  $r = E_{VOC}/E_{NO_x}$  may be used to distinguish typical urban emissions from point sources (e.g., power plants). When  $r$  is computed as atoms of C over atoms of N by volume,  $r < 1$  for power plants; and  $r > 1$  for urban areas (where typically  $r \approx 5$ ). Using the above correlations between NO<sub>x</sub>, VOC, and CO

emissions,  $r \approx 4$ . By setting the parameters DNOX and DVOC at their minimum and maximum values, we see  $1 < r < 9$ .

**3.1.11. Initial and boundary conditions (AQINOX, AQISO2, AQIVOC, and AQIOZO).** The mass of pollutants both present inside the domain when the simulation starts and entering the domain through the boundaries are very important for the photochemistry. To simplify the amount of input data required, we simply express the initial and boundary condition information using four so-called air quality indexes (AQI). These AQI, denoted AQINOX, AQISO2, AQIVOC, and AQIOZO determine initial concentrations and boundary concentrations for NO and NO<sub>2</sub>, SO<sub>2</sub>, VOC and CO, and ozone and peroxyacetyl nitrate (PAN), respectively. These AQI vary from 0 to 1 and represent fractions of maximum concentrations for each species or group of species. The maximum concentrations are appropriate for polluted atmospheres as given by *Seinfeld* [1986], *Finlayson-Pitts and Pitts* [1986], and the U.S. national ambient air quality standards (NAAQS) and other standards (see Table 3). Runs begin and initial conditions apply at midnight. To estimate maximum concentrations for formaldehyde, acetaldehyde, methyl ethyl ketone and other VOC from the total concentration of reactive hydrocarbons, RHC, we have slightly modified the factors provided in the CIT User's Manual [*McRae et al.*, 1992]. We assume a beta distribution between 0 and 1 for the PDF of each AQI (Figure 1). The most probable value is assumed to be 0.1, which corresponds to a quite clean atmosphere. The justification for the latter assumption is that boundary conditions are set through these AQI values, and we assume that the nonurban area around the city is not usually very polluted.

**Table 3.** Concentrations for Defining AQI

Pollutant	Clean Troposphere <sup>a</sup>	Polluted Troposphere <sup>a</sup>	WHO Air Quality Guideline	NAAQS	Value for AQI=1
CO	200-1000	1000-50,000	25,000 (1 hour)	35,000 (1 hour)	4000
SO <sub>2</sub>	1-10	20-200	125 (1 hour)	500 (3 hours)	100
NO <sub>2</sub>	0.1-20	50-500	190 (1 hour)	50 (annual)	300
NO	0.01-20	50-2000		500 (1 hour)	100
RHC <sup>b</sup>	100-500	500-1500		240 (3 hours)	1000 <sup>c</sup>
O <sub>3</sub>	20-80	100-500	100 (1 hour)	120 (1 hour)	250

See text; units in ppb by mol. WHO, World Health Organization; NAAQS, National Ambient Air Quality Standard. Averaging time for concentrations for each specific standard is given in parentheses.

<sup>a</sup> Range of values for so-called clean and polluted atmospheres are taken from *Seinfeld* [1986] and *Finlayson-Pitts and Pitts* [1986].

<sup>b</sup> Values for reactive hydrocarbons are given in ppb by moles of C (ppbC).

<sup>c</sup> RHC=1000 ppb means the following concentrations for each species: formaldehyde, 40 ppb; acetaldehyde, 20 ppb; methyl ethyl ketone, 22 ppb; alkanes ( $\geq C_4$ ), 95 ppb; ethene, 28 ppb; alkenes ( $\geq C_3$ ), 18 ppb; toluene, 20 ppb; higher aromatics, 8 ppb.



Initial concentrations vary throughout the domain. The values given by multiplying the maximum concentrations by the AQI are assumed to apply to city center and downwind from it (Figure 3). At different distances from the center, initial concentrations are assumed to be lower, so that at the upwind boundary, the concentration is 0.1 times the central concentration (see Figure 3). Boundary conditions are kept constant for the whole run.

### 3.2. Selected Outputs

The CIT model predicts as outputs grid-resolved concentration fields of all species considered as well as their fluxes in and out of the domain. As for the inputs, computational demands limit the amount of output information that we can consider. Mass balance in the domain for each chemical species requires

$$M_i(t) = M_i(0) + F_i^{\text{in}}(0,t) - F_i^{\text{out}}(0,t) + E_i(0,t) + R_i(0,t) \quad (9)$$

where  $M_i(t)$  is the total mass of species  $i$  in the domain after  $t$  hours;  $M_i(0)$  is the mass at the beginning of the run (initial conditions);  $F_i^{\text{in}}(0,t)$  and  $F_i^{\text{out}}(0,t)$  are the total amounts of species  $i$  that have entered and exited the domain through the boundaries, respectively, over time  $t$ ;  $E_i(0,t)$  is the true emission of species  $i$  over time  $t$ ; and  $R_i(0,t)$  is the amount of species added (positive) or removed (negative) in the domain by chemical reactions between  $t=0$  and  $t$ . The effective emission ( $EE_i(t)$ ), is defined as the sum of the true emissions and the net production or destruction due to chemistry. Since the CIT model does not provide explicitly the term  $R_i$ , we derive the effective emissions using

$$\begin{aligned} EE_i(t) &= E_i(0,t) + R_i(0,t) \\ &= M_i(t) - M_i(0) - F_i^{\text{in}}(0,t) + F_i^{\text{out}}(0,t). \end{aligned} \quad (10)$$

Since the long-term average of  $M_i(t) - M_i(0)$  (accumulation in domain) is generally much less than the long-term average for  $F_i^{\text{net}}(t) = F_i^{\text{out}}(0,t) - F_i^{\text{in}}(0,t)$  (net export from domain) then, over the long term, the effective emissions equal approximately the export of pollutants from the domain.  $EE_i$  may be aged primary emissions or production of secondary pollutants, depending upon the species. In principle,  $EE_i$ , like  $R_i$ , may have positive or negative values. We are interested in the effective emissions of (primary) species that are actually emitted ( $\text{NO}$ ,  $\text{NO}_2$ ,  $\text{CO}$ ,  $\text{SO}_2$ ,  $\text{SO}_3$  and some VOC species) and (secondary) species generated by chemistry ( $\text{O}_3$ ,  $\text{HNO}_2$ ,  $\text{HNO}_3$ ,  $\text{HNO}_4$  and PAN).

We typically run the CIT model for 1 day. This length of time is of the same order than the mean residence time of pollutants in the domain and enables resolution of the effects of the daily cycles of emissions and solar radiation. We could output the effective emissions every hour; to reduce the number of outputs, but still define the daily cycle, we output results every 6 hours.

In summary, the outputs that we include in our parameterization are the total mass inside the domain at the beginning of the simulation, and after 6, 12, 18 and 24 hours ( $M_i(0)$ ,  $M_i(6)$ ,  $M_i(12)$ ,  $M_i(18)$ ,  $M_i(24)$ ) and the net export fluxes after 6, 12, 18 and 24 hours ( $F_i^{\text{net}}(0,6)$ ,  $F_i^{\text{net}}(0,12)$ ,  $F_i^{\text{net}}(0,18)$ ,  $F_i^{\text{net}}(0,24)$ ) for all species. All these outputs are expressed per unit area. In addition, we output quantities that are useful air pollution indicators: domain-averaged and 8 hour-averaged concentra-

tions for some pollutants ( $\text{SO}_2$ ,  $\text{NO}_x$ ,  $\text{CO}$  and  $\text{O}_3$ ), total  $\text{SO}_2$  deposition every 8 hours, and maximum ozone concentration and the time when the peak is reached. These latter outputs allow us the use of the parameterization in assessment of local air pollution control policies.

### 3.3. Obtaining the Parameterizations

Having defined the inputs and their PDFs and outputs of interest, we use the program DEMMUCOM [Tatang, 1994] to obtain the desired parameterizations. DEMMUCOM defines the collocation points (i.e., specific inputs) necessary to run the CIT model as many times as required and computes the coefficients of the polynomial expansion.

Since we are potentially interested in 31 species, and nine outputs for every species, plus the 17 additional air pollution-related outputs, we have a total of 296 outputs and hence need to generate 296 parameterizations. This is not computationally demanding, since the same number of runs of the CIT model is needed for one output as for multiple outputs. The required number of runs of the CIT model is instead defined by the number of inputs and the desired level of accuracy of the approximation. For example, to obtain the 296 parameterizations with 14 inputs, the first order ( $N=1$ ) approximation requires 15 runs, the second order ( $N=2$ ) approximation requires 120 runs, and the third order ( $N=3$ ) needs 680 runs. The rapid increase in the number of runs needed with increasing order of the approximation is due to the increasing number of possible cross products which can involve up to 14 polynomials of the input parameters (see (4)).

## 4. Results and Discussion

### 4.1. Parameterization Equations

Initial applications of the above methodologies indicated that a second-order approximation provides a good balance between accuracy of results and number of CIT model runs required. Also, the third-order polynomial chaos expansion did not show a clear improvement in the accuracy of parameterization outputs when compared with true model outputs [Calbó *et al.*, 1996]. Hence we chose a second-order expansion for the work described here.

We can in addition take advantage of any a priori knowledge about the CIT model response surface to improve the approximation. In particular, we know that the response surface is periodic regarding the parameter DATE. Since a second-order polynomial expression does not have a periodic behavior, we augment the expansion by including a cosine function. The parameterization that we derive is, specifically, a second-order polynomial expansion for 13 input variables (all except DATE) multiplied by a function of the input DATE that takes account of the annual cycles in output  $y$ :

$$\begin{aligned} y &= y_0 + \left[ A(x_2, \dots, x_{14}) \cos(\omega x_1 + \delta(x_2, \dots, x_{14})) \right] \\ &\quad \times \text{PCE}(x_2, \dots, x_{14}) \end{aligned} \quad (11)$$

$$\begin{aligned} \text{PCE} &= \sum_{i=2}^{14} y_i^1 P_i^1(x_i) + \sum_{i=2}^{14} y_i^2 P_i^2(x_i) \\ &\quad + \sum_{i=2}^{14} \sum_{j=i+1}^{14} y_{ij}^{11} P_i^1(x_i) P_j^1(x_j). \end{aligned} \quad (12)$$



Here  $x_1$  is the input DATE,  $x_2, \dots, x_{14}$  are the other 13 inputs,  $A$  and  $\delta$  are the amplitude and the phase angle to be found for each set of input parameters, and  $y_0$  is the mean value for the output.

Based on physical reasoning, there are specific constraints on outputs:  $M_A(t) \geq 0$ ;  $F_i^{net}(t) \geq -F_i^{in}(t)$ ; deposition and concentration  $\geq 0$ ; and the time of peak ozone concentrations lies in the range 0 to 24 hours. Due to space limitations, the large number of coefficients that multiply each polynomial to define each output, and the orthogonal polynomials corresponding to each input are not given here. Instead, Fortran and C versions of the parameterizations are available by contacting the authors.

4.2 Parameterization Tests

DEMMUCOM includes a methodology for checking the error in expressions like (11) and (12) for the purposes of uncertainty analysis. As such, relatively small variations of an uncertain input around a central value resulting in small variations of the output around its mean value are assumed. We are interested in larger variations of inputs so we have followed another procedure. First, we have defined 100 sets of the 14 input parameters by randomly sampling within each input parameter's PDF, while demanding that 50 of these sets have all input parameters inside the range of values defined by the collocation points used to derive the parameterization. Then, we run both the CIT model and the second-order parameterization with each set of input parameters, and we compare the results.

Wilmott [1982] has suggested that several statistical indexes, such as the mean bias error, the root mean square error, the mean absolute error, the intercept and the slope of the least squares regression, among others, should be reported

when analyzing a model's performance. As a summary of all these measures, he proposed an "index of agreement"  $d$  defined by

$$d = 1 - \frac{\sum_{i=1}^P (y_T^i - y_A^i)^2}{\sum_{i=1}^P (|y_A^i - \bar{y}_T| + |y_T^i - \bar{y}_T|)^2} \tag{13}$$

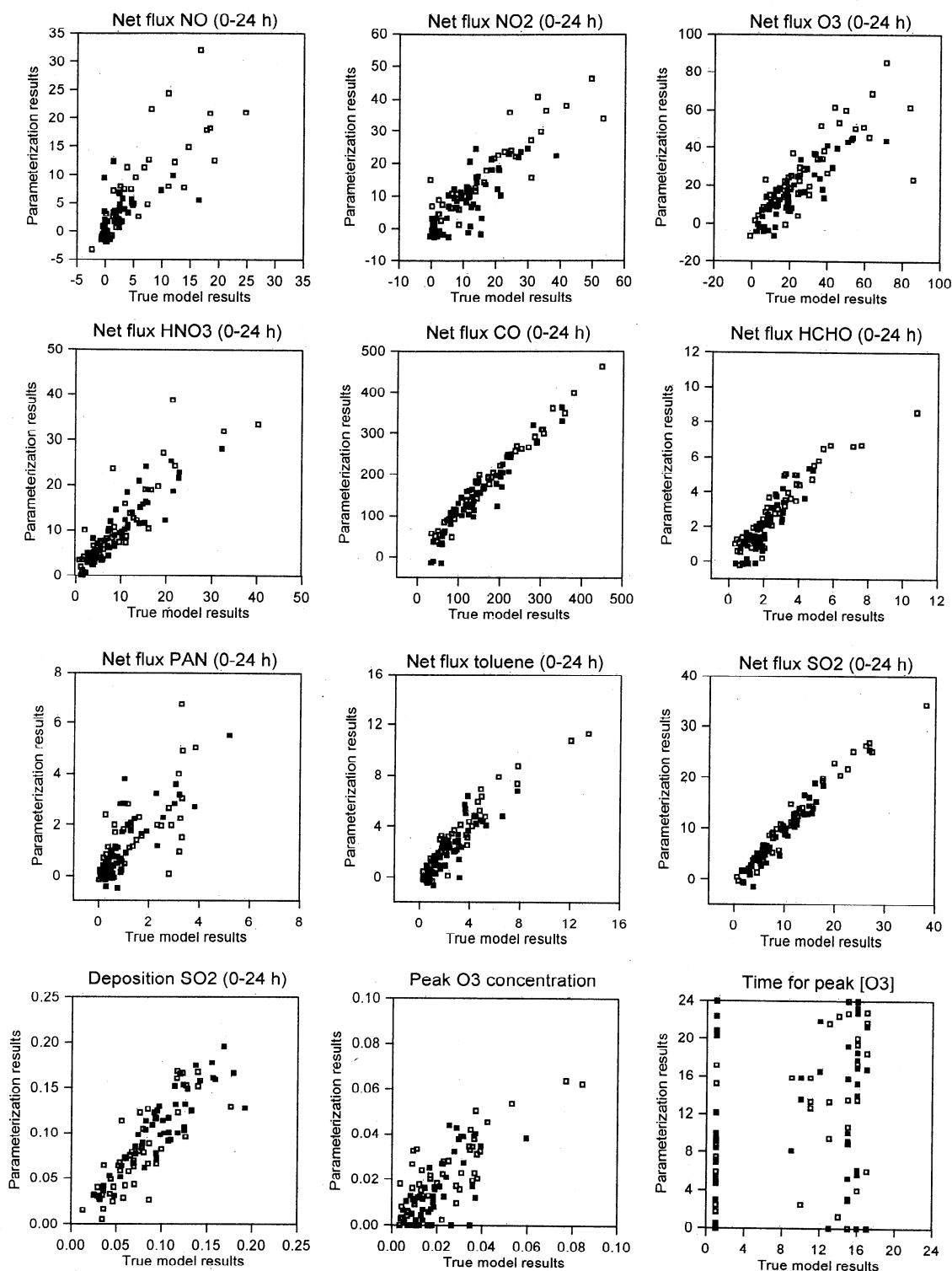
where the subscript  $T$  denotes true model results, the subscript  $A$  denotes parameterization results, and the overbar denotes the mean value.  $P$  (=100) is the number of comparisons used to check the error. With this definition, the better the model, the closer  $d$  is to unity. Wilmott [1982] notes that this index, which is both relative and bounded, is particularly useful for comparing different models with the same set of inputs and outputs. In our case, we use it to compare the second-order parameterizations to the parent CIT model.

The values for  $d$  for each output are shown in Table 4. Most  $d$  values are evidently close to unity, indicating the parameterization is a quite good approximation for most of the outputs. Only for a very few outputs is the index of agreement below 0.75. Specifically, the time when the peak ozone concentration is reached is poorly approximated ( $d = 0.5658$ ). This result could be expected, since the second-order expansion does not provide high time resolution and the time when peak ozone occurs is a nonlinear function of the input parameters.

Correlations between parameterization and CIT model outputs are generally good (Figure 5). The points in Figure 5 that are far from the perfect correlation lines generally correspond to a set of input values that is very unlikely to happen (i.e., one or more input parameters take a value close to the limit of

Table 4. Index of Agreement  $d$  for All the Outputs

Species	$M(0)$	$M(6)$	$F^{net}(0-6)$	$M(12)$	$F^{net}(0-12)$	$M(18)$	$F^{net}(0-18)$	$M(24)$	$F^{net}(0-24)$
NO	0.9997	0.9522	0.7372	0.9134	0.9256	0.8655	0.9379	0.7732	0.8944
NO <sub>2</sub>	0.9999	0.9594	0.9579	0.9641	0.9506	0.8925	0.9392	0.891	0.9315
O <sub>3</sub>	0.9999	0.9786	0.9352	0.8488	0.942	0.9491	0.9116	0.9505	0.8979
HNO <sub>3</sub>	1.	0.9084	0.9395	0.9381	0.9408	0.9468	0.9464	0.9357	0.9389
CO	0.9982	0.9884	0.9224	0.984	0.9503	0.9868	0.9726	0.9863	0.9829
HCHO	0.9982	0.9906	0.9297	0.9518	0.9451	0.9718	0.9347	0.9798	0.9519
ALD2	0.9975	0.9871	0.9471	0.9662	0.953	0.9464	0.9417	0.947	0.9545
MEK	0.9946	0.9829	0.9269	0.9764	0.9492	0.9813	0.951	0.9819	0.9634
PAN	0.9999	0.9775	0.9483	0.9171	0.9149	0.9018	0.8351	0.8925	0.8527
ALKA	0.9993	0.9909	0.9341	0.9777	0.9565	0.9606	0.9546	0.9703	0.9645
ETHE	0.9993	0.9911	0.9322	0.9566	0.9507	0.9159	0.9322	0.9452	0.9366
ALKE	0.9993	0.9521	0.8928	0.6943	0.8787	0.6032	0.8004	0.7408	0.7819
TOLU	0.9993	0.9909	0.9341	0.9697	0.9564	0.9445	0.9484	0.9635	0.9544
AROM	0.9993	0.9904	0.9249	0.7244	0.9253	0.6787	0.829	0.8417	0.7990
SO <sub>2</sub>	0.9994	0.9909	0.9465	0.985	0.9601	0.9784	0.9757	0.9859	0.9859
SO <sub>3</sub>	1.	0.991	0.6395	0.9526	0.9362	0.9375	0.9351	0.9486	0.9438
	0-8 hours	9-16 hours	17-24 hours						
SO <sub>2</sub> deposition	0.8829	0.9291	0.8851						
SO <sub>2</sub> concentration	0.8916	0.9230	0.8612						
CO concentration	0.9462	0.9521	0.8508						
NO <sub>x</sub> concentration	0.9046	0.9085	0.7465						
O <sub>3</sub> concentration	0.8921	0.9337	0.8855						
O <sub>3</sub> peak	0.8226	O <sub>3</sub> peak time	0.5658						

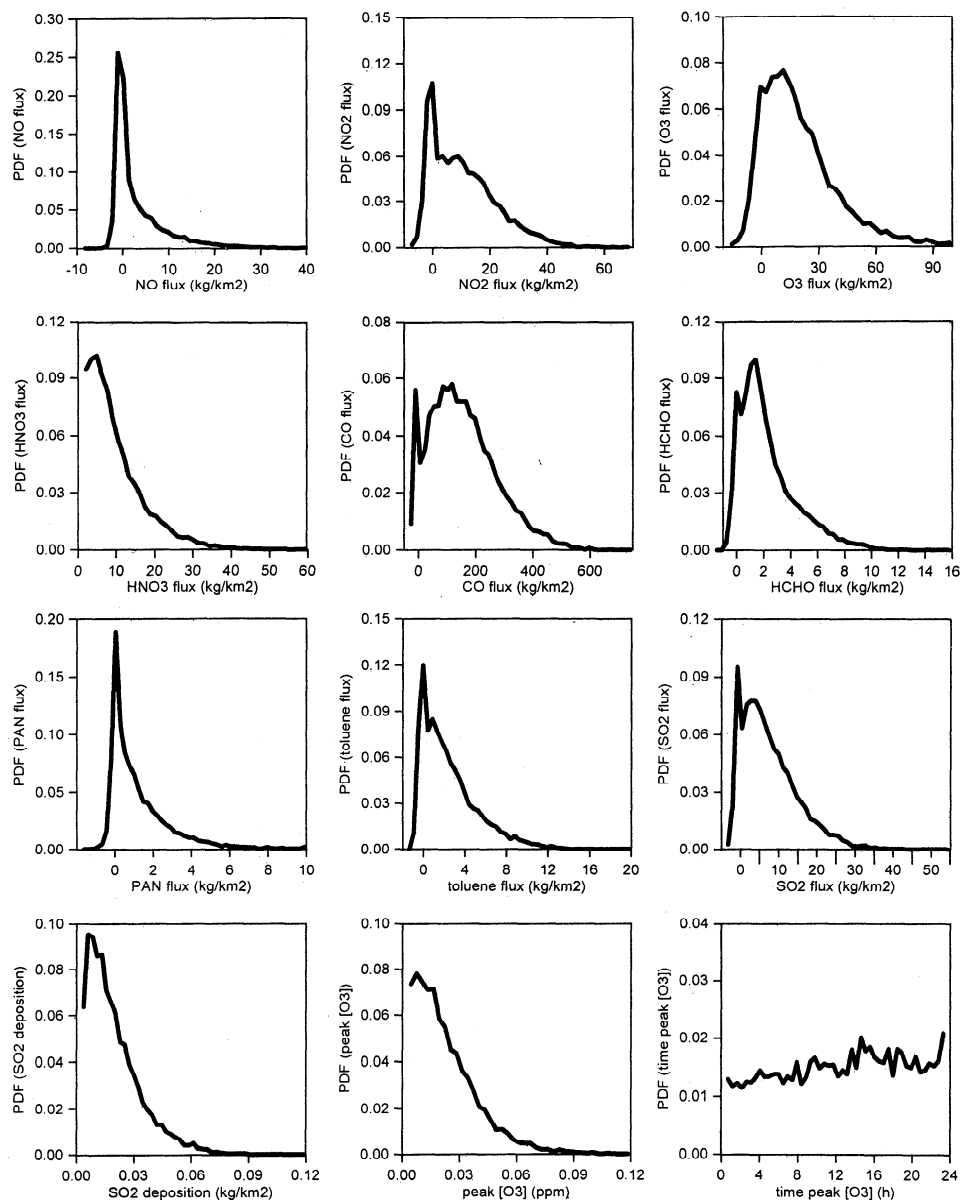


**Figure 5.** Sample of comparisons between parameterization results and true model results. We have chosen 12 out of the 296 outputs. Here a total of 100 runs using sets of input parameters generated randomly are shown. Solid squares correspond to points whose input parameters lie inside the range of sampling values used to generate the approximation. Open squares have one or more parameters outside that range.

their PDFs, represented in the figure by open squares). Again the only exception to the goodness of the parameterizations is the time for peak ozone concentration. Except for this time, linear regressions between CIT model and parameterization show intercept values close to 0, slopes close to 1, and correlation coefficients exceeding 0.6.

### 4.3. Parameterization Applications

The derived parameterization can be used for computationally efficient simulation of urban scale processes in global-scale models, for determining probability distributions of outputs, and for carrying out sensitivity analyses. To de-



**Figure 6.** Probability density functions for selected outputs obtained by running the parameterizations 10,000 times with different sets of Monte Carlo generated input parameters.

termine PDFs for each output, we have run the 296 parameterizations 10,000 times with sets of input parameters generated by Monte Carlo sampling within the ranges defined for each. Results for the total net flux of nine species through lateral boundaries in 1 day ( $F_i^{\text{net}}(0,24)$ ), sulfur dioxide deposition, peak ozone concentration and the time when this peak is reached are shown in Figure 6. Note that net fluxes are usually biased toward positive values indicating that the flux of pollutants leaving the urban region exceeds the flux of pollutants entering, as expected. For some species (e.g., ozone) the net flux can be negative indicating efficient net destruction rather than production of these species in certain circumstances. The  $\text{SO}_2$  deposition flux may reach  $0.1 \text{ kg/km}^2/\text{d}$ , although in general it is less than  $0.03 \text{ kg/km}^2/\text{d}$ . Similarly, the peak ozone concentration may reach  $0.1 \text{ ppm}$ , but usually it is well below  $0.06 \text{ ppm}$ . Although this value may seem low, note that we define the peak to be the maximum domain averaged concentration. The actual peak in a specific grid cell in

the domain may be 2 to 5 times higher than the domain average. Consistently with previously noted results, the PDF for the time of peak ozone does not have any significant pattern.

We have carried out two types of sensitivity analyses. First, DEMMUCOM provides the contribution of each input parameter to the variance of the outputs (Table 5). Note that the residence time, as well as certain air quality indices and emissions are the inputs that explain most of the variance for most of the outputs. Specifically, residence time is the most important contributor to the variance of half of the analyzed outputs. This makes sense, since the residence time is the parameter that defines how long pollutants undergo chemical reactions in the domain. The input CO is also very important since this parameter also determines emissions of  $\text{NO}_x$  and VOC through the correlations (6) and (8). The air quality indexes for  $\text{NO}_x$ , VOC, and ozone, which determine initial and boundary conditions in the simulation, are also important parameters. Note that while the  $\text{SO}_2$  emissions and  $\text{SO}_x$  air qual-

**Table 5.** Input Parameters Sorted by Their Contribution to the Variance of the Outputs

Input	Explains Most of the Variance for This Number of Outputs	Is the Least Important for This Number of Outputs
TIME	67	0
AQIVOC	18	0
CO	23	3
AQINOX	11	0
LATI	2	0
AQIOZO	7	2
MIXI	5	4
NOX	2	3
CLOU	0	7
TEMP	0	4
VOC	1	16
AQISOX	4	30
SO2	5	76

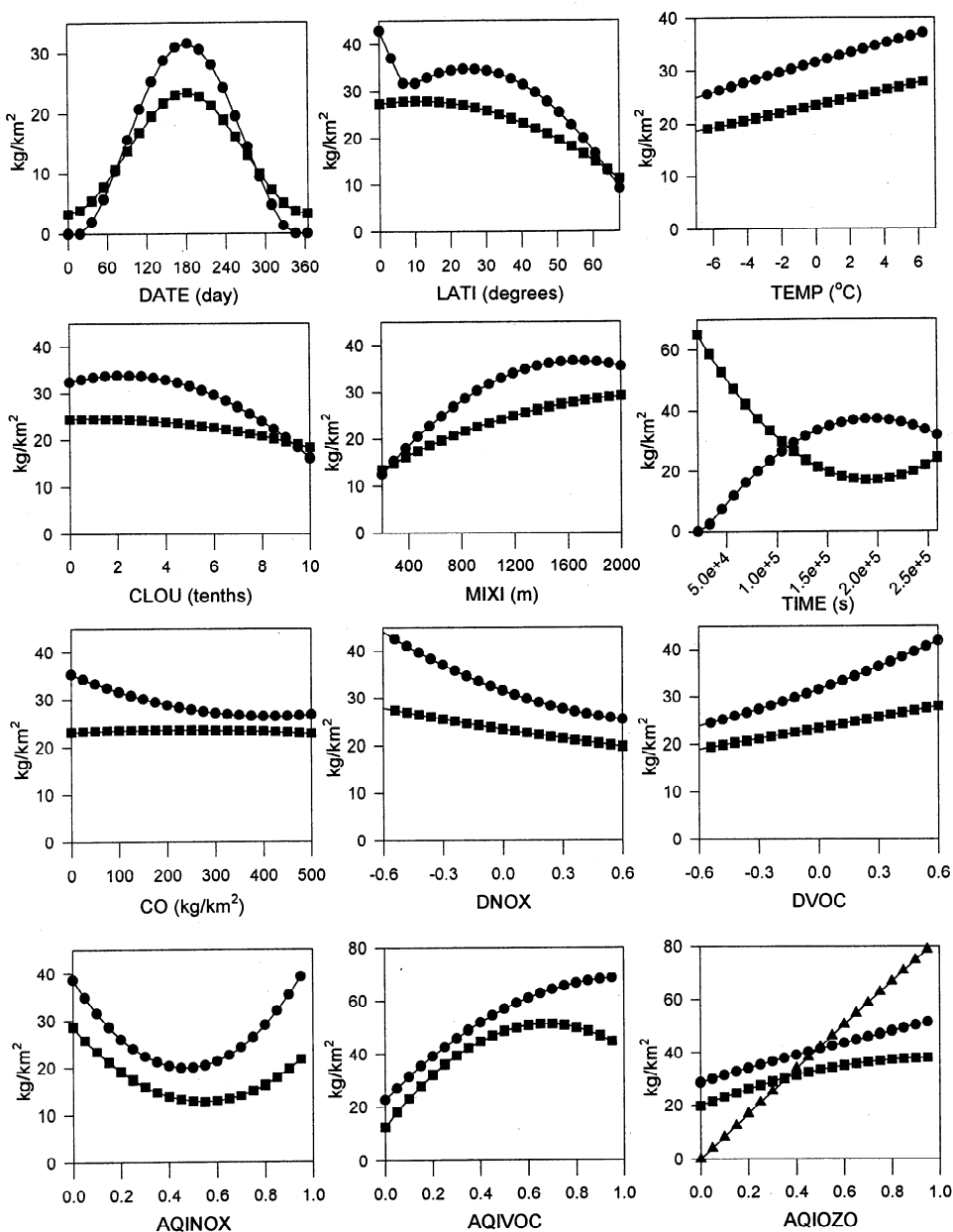
ity index appear less important, this is misleading because the number of outputs related to  $\text{SO}_x$  is much less than the number of outputs related to  $\text{NO}_x$ -VOC- $\text{O}_3$  photochemistry. The parameter LATI is quite important, since it drives (along with DATE) the incoming solar radiation and the temperature. Note that DATE does not appear in this analysis, since, as explained earlier, the seasonal effect has been considered external to the collocation method. Among meteorological inputs, the mixing layer height (MIXI) is the most important, especially for concentration related outputs. MIXI is naturally not important for outputs of the total amount of pollutants in the domain (as opposed to concentration), and it also has no effect at night when MIXI/20 is less than 50 m (its default minimum). The deviation from climatic temperature (DTEM) and cloudiness (CLOU) have in general a minor contribution to the variance of the analyzed outputs. For CLOU, this result is related simply to the fact that approximately half of the outputs correspond to nighttime (at 6 and at 24 hours) when cloud effects are not relevant.

Our second sensitivity analysis addresses the behavior of one output when changing one or more inputs. Since we have chosen 14 input parameters, each parameterization has a fourteen-dimensional (14-D) response surface. Obviously, we cannot easily represent a 14-D surface, but we can show 1-D or 2-D sections through this hypersurface. For two outputs (total mass of ozone after 1 day and net export flux of ozone during 1 day), several 1-D sections are shown in Figure 7. They have been obtained by following this procedure. First, a set of reference values for the input parameters was chosen (DATE=182, LATI=40°N, DTEM=0, CLOU=5, MIXI=1000 m, TIME = 129,600 s,  $\text{SO}_2$ =10 kg/km<sup>2</sup>/d,  $\text{CO}$ =100 kg/km<sup>2</sup>/d,  $\text{DNOX}$ =0,  $\text{DVOC}$ =0,  $\text{AQINOX}$ =0.1,  $\text{AQISOX}$ =0.1,  $\text{AQIVOC}$ =0.1, and  $\text{AQIOZO}$ =0.1). These reference values are the most probable values for nonuniform distributions and central values for uniform distributions. Then, we allowed one input parameter at a time to vary between its limits, keeping all other parameters constant. The sensitivity of  $\text{O}_3$ -related outputs to the inputs  $\text{SO}_2$  and  $\text{AQISO}_2$  is not shown, since this sensitivity is negligible.

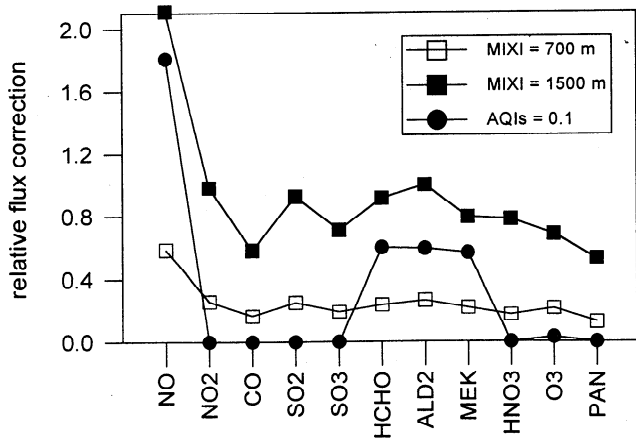
Total mass and net flux of ozone are both larger for DATE values around 200 (summer at 40°N) when both solar irradiance and temperature are higher. Highest ozone mass and net flux occur in the tropics (LATI close to 0), which is reason-

able since these latitudes receive maximum of solar radiation. The ozone mass and flux increase with increasing temperatures (DTEM) due to the dependence of key reactions rates on temperature. The effect of cloud increases is, as expected, to lower ozone generation. The deeper the mixing layer, the larger the ozone total mass and net flux. Deeper mixing layers mean the same amount of emissions enter a larger air mass, which, due to chemical nonlinearity, leads to enhanced ozone generation. The longer the residence time, the larger the mass accumulated in the domain and the lower the net flux of ozone. Conversely, for short residence times, ozone generation is small and ozone is advected rapidly out of the domain. The sensitivity of ozone to the parameter CO is apparently small. Recall that CO refers not only to carbon monoxide emissions, but it also defines emissions of  $\text{NO}_x$  and VOC. The highest accumulation of  $\text{O}_3$  is reached with low CO (i.e., low CO,  $\text{NO}_x$  and VOC emissions), but we caution that the approximation in the limit of very high emissions may be poor, since the beta function PDF of CO has small probability in that limit. The effect of the input parameters  $\text{DNOX}$  and  $\text{DVOC}$  (i.e., differences of the emissions of  $\text{NO}_x$  and VOC from those defined by CO) is also shown in Figure 7. The amount of ozone created is lower when the emissions of  $\text{NO}_x$  relative to CO are higher. On the other hand, the amount of ozone generated increases when the emissions of VOC, relative to CO, increase. In summary, the higher the ratio of VOC to  $\text{NO}_x$  emissions, the higher the generation of ozone. Recall that this result is for a given reference set of the other input parameters. Finally, the effect of the indexes AQI (which determine both initial and boundary conditions) is also very important. The results for AQINOX show that higher ozone creation occurs with either clean or polluted  $\text{NO}_x$  atmospheres. In contrast, the generation of ozone is lower in clean VOC atmospheres. The effect of increasing AQIOZO is, of course, to add more ozone at the beginning of the simulation. This results in more ozone at the end, and also in a larger flux through the boundaries.

A major proposed application of the parameterization developed here is to estimate effective emissions of pollutants from large urban areas using (10) in global-scale chemical models. Hence this parameterization has been derived in such a way that it can be run sequentially at 6, 12, or 24 hours time steps for each polluted urban region in the global model. The relevant parameterization outputs are the net fluxes through the boundaries during the 6-24 hour run and the total masses of pollutants in the urban domain after each run. The masses at the end of one run are then used to set the initial conditions for the next run. However, some corrections must be made. First, at the end of one run, we find amounts of pollutants present in the upper layers, depending on the maximum extent of the mixing layer during that run. However, the initial conditions for the next run assume that the two upper layers contain no pollutants. Therefore some of the mass present inside the domain at the end of one run is not going to be present at the beginning of the next run and must be accounted for as an extra (upper ventilation) flux of pollutants from the urban area into the global model. Second, initial concentrations of several pollutants are set through the same air quality index (e.g., NO and  $\text{NO}_2$  through AQINOX). This means that initial concentrations always show the same relative amounts of such pollutants (in the case of  $\text{NO}_x$ , the ratio is 1 ppm of NO for 3 ppm of  $\text{NO}_2$ ). This initial ratio, of course, changes during the run due to the urban emissions and chemistry. Again, for some

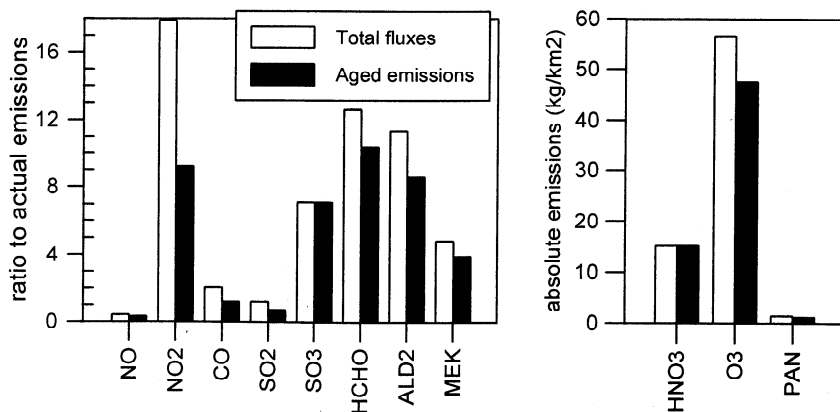


**Figure 7.** Sensitivity of the total mass of ozone in the domain after 1 day (circles) and net flux through the lateral boundaries during 1 day (squares) to variations in the indicated inputs. In the bottom right-hand plot, the triangles refer to the initial mass in the domain. Read  $5.0e+4$  as  $5.0 \times 10^4$ .



pollutants some mass has to be removed from the urban domain, and accounted for as another extra (AQI inconsistency) flux into the global model. In Figure 8 upper ventilation and AQI inconsistency fluxes for some specific cases are shown. Note that the upper ventilation flux depends strongly on mixing layer height, while the AQI inconsistency flux varies widely depending on the species. The effective emissions to be used by the global model are therefore the sum of three

**Figure 8.** Examples of fluxes to be added to lateral fluxes due to upper ventilation (squares) and AQI inconsistency (circles) (see text). Fluxes are added to ensure mass conservation. Here they are plotted as values relative to the corresponding net lateral fluxes.



**Figure 9.** Examples of total fluxes out of an urban domain into a global-scale model for several species. (left) For primary species, fluxes are shown relative to their corresponding actual (surface) emissions. (right) For secondary species, absolute values are given.

fluxes: the net flux through the boundaries (given as one of the parameterization outputs), the upper ventilation flux, and the AQI inconsistency flux. In this way we ensure mass conservation for all species.

## 5. Discussion and Conclusions

We have presented in this paper a methodology for developing a parameterization of urban-scale processes that affect export rates of pollutants (effective emissions) from the urban region. The effect of these processes on emissions is important, as can be seen in Figure 9. The ratio of effective emissions to actual emissions is shown for several pollutants for a specific case defined by a reference set of data describing the actual emissions, the meteorology and the initial and boundary concentrations. Two quantities are given in Figure 9: the flux for each species that is to be input in a global-scale model and the fraction of this flux that is strictly due to chemistry (that is, discounting the effect of pollutants already present in the urban area). Some species are consumed by the urban-scale chemistry (NO, SO<sub>2</sub>) while effective emissions of other species are several times bigger than actual emissions (NO<sub>2</sub>, SO<sub>3</sub>, several oxidized hydrocarbons). The importance of urban-scale chemistry in creating secondary species such as ozone is also obvious from Figure 9 where their absolute emissions are shown.

The principal result of our work is the definition of a set of polynomial expressions that approximate the predictions by the CIT Urban Airshed Model for effective emissions of several species, as functions of 14 input parameters. Specifically,

$$EE_i^t = f_i^t(E_j, M) \quad (14)$$

where effective emissions of species  $i$  at time  $t$  ( $EE_i^t$ ) are a function of actual emissions of several species ( $E_j$ ) and of other urban domain properties ( $M$ ) such as meteorology. The total mass inside the urban domain and the net fluxes out of the domain of NO, NO<sub>2</sub>, HNO<sub>2</sub>, HNO<sub>3</sub>, HNO<sub>4</sub>, O<sub>3</sub>, CO, PAN, SO<sub>2</sub>, SO<sub>3</sub>, and several VOC species, among other species, after 6, 12, 18 and 24 hours of photochemical activity are estimated. The domain-averaged and time-averaged concentrations of SO<sub>2</sub>, NO<sub>x</sub>, CO, and O<sub>3</sub>, and the total deposition of SO<sub>2</sub> are also estimated through other parameterizations. All these parameterizations need 14 input values: the latitude of the

city, the date of the simulation, the temperature, cloud cover and mixing layer depth (one value for each meteorological parameter), the estimated residence time (defined as the ratio between length of domain and mean wind), the emissions of CO, NO<sub>x</sub>, VOC and SO<sub>2</sub>, and finally four air quality indexes describing the initial pollution in the urban area of interest and the environment surrounding this area.

To develop the parameterization, we applied the probabilistic collocation method, which uses the probability density functions of the inputs to generate a set of orthogonal polynomials which form the basis for a polynomial chaos expansion that approximates the actual response of the CIT model to its inputs. To simulate the seasonal cycle, the polynomial expansion is augmented by a cosine function. The parameterization provides a computationally very efficient simulation of the behavior of the actual model. By comparing the outputs of the parameterizations with the outputs of the true model, we conclude, in general, that the second-order polynomial expansion is adequate for our purposes.

One use of these parameterizations is to facilitate inclusion of urban-scale processes as subgrid scale phenomena in global-scale models. Corrections to the parameterizations to ensure mass conservation in this application are given. Another use, demonstrated in this paper, is for uncertainty and sensitivity analyses. Both of these uses are based on the capability of carrying out hundreds or thousands of runs of the parameterizations in a very short time.

Certain improvements and extensions are suggested for the future based on this work. First, the probabilistic collocation methodology could be used to obtain separate parameterizations for different types of large cities (defined by their shape, location, topography, etc.). Alternatively, a set of parameterizations could be used as a basis to simulate any specific city as a linear combination of the members of the set. Second, a detailed study of the quality of the approximations, involving not only comparisons with true model results but also with observations, is needed. For this purpose, there are some cities for which detailed emission inventories and meteorological and air quality data are available. Third, a more detailed sensitivity analysis for all the outputs that we have parameterized (including simultaneous changes in two or more inputs) is desirable and will be the object of a later paper. Fourth, the implementation of the parameterizations in global chemistry models is of course the ultimate goal. The use of this parame-

terization in the Massachusetts Institute of Technology (MIT) global chemistry and climate model (Prinn et al., submitted manuscript, 1997) is underway and will also be addressed in a future report. Finally, the use of this parameterization as a fast tool for real-time urban air pollution control is also possible.

**Acknowledgments.** This work was carried out during the first author's postdoctoral visit to MIT which was funded by the Comissió Interdepartamental de Recerca i Innovació Tecnològica (CIRIT), Catalunya, Spain. This research was supported by the U.S. National Science Foundation grant 95-23616 to the MIT Joint Program on the Science and Policy of Global Change. Discussions and suggestions from Jean Fitzmaurice and Henry Jacoby, of the MIT Joint Program, were greatly appreciated.

## References

- Calbó, J., W. Pan, M. Webster, R.G. Prinn, and G.J. McRae, Parameterization of urban-scale photochemistry in global chemistry models (abstract), *Eos Trans. AGU*, 77(46), Fall Meet. Suppl., F84, 1996.
- Costa, M., and J.M. Baldasano, Development of a source emission model for atmospheric pollutants in the Barcelona area, *Atmos. Environ.*, 30(2), 309-318, 1996.
- Finlayson-Pitts, B.J., and J.N. Pitts, *Atmospheric Chemistry: Fundamentals and Experimental Techniques*, 1098 pp., John Wiley, New York, 1986.
- Gautschi, W., Algorithm 726: ORTHPOL - A package of routines for generating orthogonal polynomials and Gauss-type quadrature rules, *ACM Trans. Math. Software*, 20(1), 21-62, 1994.
- Gillani, N.V., and J.E. Pleim, Sub-grid scale features of anthropogenic emissions of NO<sub>x</sub> and VOC in the context of regional Eulerian models, *Atmos. Environ.*, 30(12), 2043-2059, 1996.
- Golding, B.W., An efficient non-hydrostatic forecast model, *Meteorol. Atmos. Phys.*, 50, 89-103, 1992.
- Harley, R.A., A.G. Russell, G.J. McRae, G.R. Cass, and J.H. Seinfeld, Photochemical modeling of the southern California air quality study, *Environ. Sci. Technol.*, 27, 376-388, 1993.
- Instituto Mexicano del Petroleo and Los Alamos National Laboratory (IMP-LANL), Mexico City Air Quality Research Initiative, Volume III: Modeling and Simulation, *Rep. LA-12699, UC-902*, 122 pp., 1994. (Available as DE94015933INZ from Natl. Tech. Inf. Serv., Springfield, Va., and at LANL website at <http://lib-www.lanl.gov/la-pubs/00285762.pdf>)
- Longhurst, J.W.S., S.J. Lindley, D.E. Conlan, D.J. Rayfield, and T. Hewison, Air quality in historical perspective: A case study of the Greater Manchester conurbation, in *Urban Air Pollution*, vol. 2, edited by H. Power and N. Moussiopoulos, pp. 69-109, Comp. Mech., Southampton, England, 1996.
- Lurmann, F.W., W.P. Carter, and L.A. Coyner, A surrogate species chemical reaction Mechanism for urban-scale air quality simulation models, vol. 1, Adaptation of the mechanism, Final report to the U.S. EPA under contract 68-02-4104, ERT Inc. and Statewide Air Pollut. Res. Cent., Univ. of Calif., Riverside, 1987.
- McRae, G.J., W.R. Goodin, and J.H. Seinfeld, Development of a second generation mathematical model for urban air pollution, I, Model formulation, *Atmos. Environ.*, 16, 679-696, 1982.
- McRae, G.J., A.G. Russell, and R.A. Harley, *CIT Photochemical Airshed Model: Data Preparation Manual*, 127 pp., Carnegie-Mellon Univ., Pittsburgh, Pa., and Calif. Inst. of Technol., Pasadena, 1992.
- Milford, J.B., A.G. Russell, and G.J. McRae, A new approach to photochemical pollution control: Implications of spatial patterns in pollutant responses to reductions in nitrogen oxides and reactive organic gases, *Environ. Sci. Technol.*, 23(10), 1290-1301, 1989.
- North, G.R., and J.A. Coakley Jr., Differences between seasonal and mean annual energy balance model calculations of climate and climate sensitivity, *J. Atmos. Sci.*, 36, 1189-1204, 1979.
- Pan, W., The role of aerosols in the troposphere: Radiative forcing, model response, and uncertainty analysis, *Rep. 43*, 259 pp., Cent. for Global Change Sci., Mass. Inst. of Technol., Cambridge, 1996.
- Schneider, C., C. Kessler, and N. Moussiopoulos, Influence of emission input data on ozone level predictions for the Upper Rhine valley, in *Urban Air Pollution*, vol. 2, edited by H. Power and N. Moussiopoulos, pp. 1-39, Comput. Mech., Southampton, England, 1996.
- Seinfeld, J.H., *Atmospheric Chemistry and Physics of Air Pollution*, 731 pp., John Wiley, New York, 1986.
- Sillman, S., J.A. Logan, and S.C. Wofsy, A regional scale model for ozone in the United States with subgrid representation of urban and power plant plumes, *J. Geophys. Res.*, 95, 5731-5748, 1990.
- Tatang, M.A., Direct incorporation of uncertainty in chemical and environmental engineering systems, Ph.D. thesis, Mass. Inst. of Technol., Cambridge, 1994.
- Tatang, M.A., W. Pan, R.G. Prinn, and G.J. McRae, An efficient method for parametric uncertainty analysis of numerical geophysical models, *J. Geophys. Res.*, 102, 21, 925-21, 932, 1997.
- Webster, M., M.A. Tatang, and G.J. McRae, Application of the probabilistic collocation method for an uncertainty analysis of a simple ocean model, *Rep. 4*, 21 pp., Mass. Inst. of Technol. Joint Program on the Sci. and Policy of Global Change, Cambridge, 1996.
- World Health Organization-United Nations Environment Program (WHO-UNEP), *Urban Air Pollution in Megacities of the World*, 230 pp., Blackwell, Cambridge, Mass., 1993.
- Wilmott, C.J., Some comments on the evaluation of model performance, *Bull. Am. Meteorol. Soc.*, 63(11), 1309-1313, 1982.
- J. Calbó, Departament d'Enginyeria Industrial, Universitat de Girona, Avda. Lluís Santaló, s/n, 17071 Girona, Spain. (e-mail: [jcalbo@speedy.udg.es](mailto:jcalbo@speedy.udg.es))
- G. J. McRae, R. G. Prinn, M. Webster, MIT Joint Program on the Science and Policy of Global Change, MIT Building E40-271, Cambridge, MA 02139. (e-mail: [mcrac@mit.edu](mailto:mcrac@mit.edu); [rprinn@mit.edu](mailto:rprinn@mit.edu); [mort@mit.edu](mailto:mort@mit.edu))
- W. Pan, Lawrence Livermore National Laboratory, Livermore, CA 94550. (e-mail: [wpan@alum.mit.edu](mailto:wpan@alum.mit.edu))

(Received July 9, 1997; accepted September 10, 1997.)