

**MODELING AIR QUALITY IN URBAN AREAS:
A CELL-BASED STATISTICAL APPROACH†**

by

Hag-Yeol Kim

Center for GIS
Seoul Development Institute
Seoul 100-250, S. Korea

Jean-Michel Guldmann

Department of City and Regional Planning
The Ohio State University
Columbus, Ohio 43210, U.S.A.

† Financial support from the Urban Affairs and Urban Assistance Program, Ohio Board of Regents, is gratefully acknowledged. Many thanks are due to Phil Viton, Steve Gordon, and three anonymous referees of this journal for helpful comments and suggestions on earlier drafts.

Correspondence: Professor Jean-Michel Guldmann, Department of City and Regional Planning, The Ohio State University, 190 W 17th Avenue, Columbus, Ohio 43210, U.S.A.
Tel: 614/292-2257, Fax: 614/292-7106, Email: guldmann.1@osu.edu

ABSTRACT

Statistical regression models are presented, that explain the observed variations, across urban areas, in the concentrations of two major pollutants, ozone and carbon monoxide. Model specification and estimation are based on an explicit and new spatial framework derived from the theoretical concept of well-mixed cells, whereby the basic Fickian system of diffusion equations is integrated over the regional space partitioned into a grid of large cells. The concentration in each cell results from the balance of pollutant flows into and out of this cell and of pollutant emissions and removal within that cell, and is expressed as the sum of two concentration contributions: (1) the local effect, dependent upon pollution-related factors around the measuring station, and (2) the regional effect, dependent upon pollutant flows originating outside the local area. A large data base is developed, making extensive use of GIS technology, to spatially relate such data as pollution measurements, meteorological factors, land-use characteristics, Census socio-economic data, and major highway network characteristics. The results confirm the appropriateness of the well-mixed cell framework, are in line with general knowledge regarding the determinants of ozone and carbon monoxide concentrations, and clarify the role of transportation, residential fuel-use, economic activities, natural environments, and meteorological factors such as temperature and solar radiation. About 50% of the variations in concentrations are explained by these models. Several areas of further research are outlined.

1. INTRODUCTION

The enactments of the U.S. Clean Air Acts (1963, 1970) and subsequent Amendments (1977, 1990) have led, over 1970-1999, to a 31% decrease in aggregate pollution emissions, including decreases of 31% for carbon monoxide (CO), 37% for sulfur dioxide (SO₂), 71% for particulate matter (PM₁₀), 98% for lead (Pb), and 42% for volatile organic compounds (VOCs), and an increase of 17% for nitrogen oxides (NO_x). More striking, these improvements have occurred while the population grew by 33%, the vehicle miles traveled (VMT) by 140%, and the Gross Domestic Product (GDP) by 147%. However, this aggregate assessment masks the fact that, in 1999, 150 Million tons of air pollution were released into the air, and approximately 62 Million people lived in counties with monitored concentrations above the primary standards (those designed to protect public health) for one or more of the six principal pollutants, particularly ozone in the Northeast, California, some rural areas, and some national parks. Also, despite a 60-80% decline in per-car emissions since the 1960s, total emissions from mobile sources have not decreased proportionately, because more frequent and longer trips are generated by more people and cars (EPA, 2000a; EPA, 2000b).

Air quality management and planning are therefore still necessary, and even more so in the future due to rapid urban development, vehicular traffic congestion, and growing energy consumption. Quantitative models relating urban and public policy decisions to the resulting air quality are thus still needed. However, a review of the literature points to an extremely heterogeneous body of modeling methodologies and empirical studies. Most of them are site-specific and capture only a specific aspect of the pollution problem (e.g., role of economic output, or traffic, or meteorology), and few distinguish between local, regional and national effects. Building on the well-mixed cell theoretical approach to pollution diffusion analysis, this paper develops a completely new spatial framework for the statistical

estimation, through regression analysis, of more comprehensive air quality models, that relate the concentrations of air pollutants in an urban area to (1) local urban factors, including meteorological, socioeconomic, land-use, and transportation characteristics, and (2) background pollution flows originating out of the urban area and contributed by regional and national sources. Recent developments in Geographical Information Systems (GIS) and dramatic increases in data availability through the Internet, provide much of the empirical infrastructure of this research. An extensive database is built for a large cross-section of urban areas across the U.S. for the year 1990, and the estimated models successfully separate local and background effects, and identify major local pollution factors, thus allowing for a quantitative assessment of various possible local policy decisions.

The methodology is applied to ground-level ozone (O_3) and carbon monoxide (CO). O_3 is the primary constituent of photochemical smog. It results from the chemical reaction of NO_x and VOCs, in the presence of heat and sunlight. VOCs are emitted by motor vehicles, chemical plants, refineries, other industrial sources, and commercial and consumer products. NO_x , a group of highly reactive gases containing nitrogen and oxygen in varying amounts, are emitted by motor vehicles (35%), power plants (29%), and other sources of combustion (EPA, 2000a). O_3 negatively affect the human respiratory system, as well as agricultural and forest yields. It is a pervasive problem throughout the U.S. Because O_3 and its precursors can be transported by winds over large areas (hundreds to thousands miles), it is clearly a regional/national rather than a local problem. CO is a dangerous asphyxiant, that reduces the blood ability to carry oxygen to cell tissues, and is mainly produced by the incomplete combustion of carbon and its compounds. The dominant source of CO emissions is motor vehicle exhaust (60% nationwide and 95% in cities). Other sources of CO include industrial processes, non-transportation fuel combustion, and wildfires (EPA, 2000a). The spatial impact of CO is more localized than that of O_3 , with high concentrations in areas with heavy traffic congestion. Thus, in addition to their intrinsic importance,

these two pollutants are well contrasted in terms of sources, formation processes, and spatial concentration patterns, and provide an opportunity for contrasted analyses.

The paper is organized as follows. The literature review is presented in Section 2. The modeling methodology is discussed in Section 3. The data used in the empirical analyses are described in Section 4. The results are presented in Section 5. Conclusions are outlined in Section 6.

2. LITERATURE REVIEW

Air quality modeling involves the development of quantitative relationships between pollution emissions from point and area, fixed and mobile sources, and the resulting concentrations at different receptor sites over different time scales (e.g., hourly, 24-hour, annual average). Several modeling methodologies have evolved over the last 30 years and are discussed below.

2.1. Dispersion Models

Dispersion models directly relate source emissions to receptor ambient concentrations through mathematical representations of the physical and chemical processes that govern the transport and removal/chemical transformation of the pollutants, once they have left the emission points. The Gaussian exponential plume formula has been extensively applied to chemically stable pollutants, with meteorological inputs such as wind speed and direction, atmospheric stability, and mixing height. The formula is used to compute, for any averaging time, transfer coefficients a_{ij} linking the emission of source i , Q_i , to the concentration at receptor j , C_j , through the linear relationship: $C_j = \sum_i a_{ij} Q_i$. (For a review of the early literature, see Guldmann and Shefer, 1980, Chapter 2). More complex models involve the discretization of basic fluid dynamics and chemical reactions equations, over a trajectory in the case of long-distance transport of SO_2 and its chemical conversion to sulfates (e.g., the ATMOS model of acid deposition over Asia, Arndt et al., 1997), or over a three-dimensional grid covering the region under study in the case of

ozone. (Seinfeld, 1988). Dispersion models have been used to assess the impacts of pollution emission abatement strategies.

2.2. Time-Series Models

Time-series models have been developed to take advantage of the availability of time-linked (hourly, daily, monthly), site-specific concentration data. The concentration at time t is related to past concentrations at the same site, and to present and past meteorological conditions and source strengths. AR (Autoregressive) and ARIMA (Autoregressive Integrated Moving Average) models have been developed by Chock (1975), Roy and Pellerin (1982), Milionis and Davis (1994a, 1994b), and Shi and Harrison (1997). The AR approach is well illustrated by Roy and Pellerin (1982), who relate monthly average SO₂ concentrations over 1968-1977 at 11 stations in Montreal to past concentrations and to changes in the regulation of the sulfur content of heating and industrial oil. A 12th order AR model is used to make the series stationary. Standard regression models have also been used. Annand and Hudson (1981) estimate a nonlinear model relating daily SO₂ concentrations in the Manchester area to a carry-over of the previous day's concentration, mixing height, wind speed, and heating degree-days. Cooke and Wadden (1981) use stepwise regression to explain 67% of the variance of daily sulfate concentration in Chicago as a function of total suspended particulate (TSP), relative humidity, and O₃ levels.

A related approach involves calculating, over different time frames (12-hour to 10-day), the air back-trajectory of a pollutant plume in order to determine its origin and atmospheric pathway. Wolff and Lioy (1978) use regression analysis to relate the maximum afternoon O₃ concentrations at New Jersey sites to meteorological, emission, and concentration variables along a 24-hour upwind air parcel trajectory, including the maximum upwind O₃, hydrocarbon (HC) and NO_x emissions, temperature in the previous day, and the current day's temperature, wind speed, absolute humidity, and HC and NO_x emissions. The

strong dependence of local O_3 upon upwind O_3 is linked to the relatively long lifetime of O_3 overnight above the inversion layer. Note that the upwind O_3 level also incorporates the emission history of the air parcel prior to the previous day. Lioy et al. (1982) use principal component analysis to relate average summertime sulfate concentrations to upwind particulate and SO_2 emissions, local and upwind O_3 concentrations, local temperature and humidity, and regional wind speed, explaining 50% of the sulfate variance. Ashbaugh (1983) identifies the origins and transport pathways of air masses arriving in the Grand Canyon National Park, and determines the existence of clean and polluted air transport corridors by using a trajectory clustering technique. Moy et al. (1994) uses back-trajectories to show that the variability of CO , O_3 , and NO_x concentrations in the Shenandoah National Park is related to the variability in synoptic-scale circulation: the highest levels of O_3 are not related to air stagnation and local emissions, but rather to emissions from industrialized areas in the Northwest quadrant, containing the Ohio River Valley and the Great Lakes Region. Moody and Samson (1989) use cluster analysis (CA) to relate the chemical variability in precipitation to differences in atmospheric transport along 18-hour back-trajectories. Brankov et al. (1998) use three-day back-trajectories and CA to examine the relationship between synoptic-scale atmospheric transport patterns and O_3 concentrations at several locations in the Northeastern U.S., revealing that high O_3 levels are likely to happen when air trajectories originate from the Southeast and Midwest regions.

2.3. Urban Statistical Models

2.3.1. *Aggregate models*

In this category of models, spatial diffusion is either neglected or highly simplified. Early such models have been developed by Gifford and Hanna (1973), and generalized by Horie (1974) and de Nevers and Morris (1975). They treat urban areas as uniform pollution sources, with no account for specific sources nor for the directionality of the diffusion process. Gifford and Hanna estimate the

following model: $C=kQ/u$, with C =pollution concentration, Q =source emission per unit area, u =average wind speed, and k =parameter (function of city size). McMullen et al. (1968) use correlation and regression analyses over a sample of 66 Metropolitan Statistical Areas (MSA) to relate 16 air quality measures (particulates, VOCs, sulfates, SO_2 , etc.) to 13 community variables, showing that lead concentration is primarily a function of gasoline sales, SO_2 concentration of SO_2 emissions, and sulfate concentration of ambient SO_2 . Cleveland et al. (1977) analyze O_3 concentrations at 41 urban sites in the Northeastern U.S. and show, through regression analysis, that the ratio of nighttime to daytime concentrations increases with distance from New York City, as a result of the transport of high diurnal concentrations in the New York City region into Connecticut and Massachusetts at nighttime, where relatively low local O_3 concentrations are found at daytime. Glen et al. (1996) estimate a nonlinear regression model relating the monthly mean CO concentrations in 15 cities over the period 1984-1991 to motor vehicle emissions, mixing height and surface wind speed. Some urban models focus on local economic and regulatory variables. Broder (1986), using regression analysis and data for a sample of 93 MSAs over the period 1973-77, relates changes in ambient particulate (TSP) concentrations to changes in manufacturing employment, additions to the pollution abatement capital stock, changes in wind speed, heating degree-days, precipitation, and changes in the ash content of fuel used by power plants. Kahn (1997), using 1981-1989 particulate data, regresses their annual concentration on county manufacturing employment and plant survival rates, precipitation, and a dummy variable for each year, to detect specific annual effects. Manufacturing decline through relocation of firms to less regulated areas appears to be the major determinant of the decrease in the observed concentrations. Powell (1997) regresses particulate and O_3 annual concentrations on manufacturing activity, power generation, vehicle-miles traveled, population, and compliance costs, using data for the cities of Cleveland, Pittsburgh, and Baltimore over 1972-1992. Particulate concentrations are best explained by manufacturing shipments, particulate control

costs, and coal use at power plants. O₃ concentrations are best explained by temperature, vehicle-miles traveled, and O₃ control costs.

2.3.2. *Intra-Urban Models*

Some studies focus on the spatial distribution of concentrations, their distance decay, and their relation to population density. Hewitt (1991) shows that there is an apparent distance decay of NO₂ concentrations in the urban area of Lancaster, U.K., from the city center's main road to each of 49 receptors, that there is a much larger spatial variation in annual average NO₂ concentrations in the urban core than in the suburbs, resulting from stop-start driving conditions, the immediate proximity of the sampling sites to the main road, and the influence of vehicle-induced dispersion, whereas the much lower variability in the suburban area is due to the contribution of background NO₂ concentrations. Suh et al. (1995) analyze the variations in Philadelphia of sulfate and ammonia concentrations with population density and distance/direction from the city center, showing that sulfate concentration is fairly uniform, whereas ammonia concentration varies with population density, temperature, and wind speed. A related study of particulates by Burton et al. (1996) indicates that small particles (PM_{2.5} - particulate matter less than 2.5 microns in aerodynamic diameter) are transported into Philadelphia from regional sources to the West and South, while coarse particles are generated primarily from local sources, with the highest concentrations on weekdays and in areas with higher population density.

Geographical Information Systems (GIS) models belong to the same model category, but make use of a GIS to interrelate various spatial data sets. Anderson and Greenberg (1981) use discriminant analysis to assess the impacts of 21 land-use categories on particulate concentrations at 30 monitoring sites in New Jersey over 1973-1977. The land-use data are collected in 3 concentric rings (0-1, 1-5, 5-10 miles) around the sampling site, and the rings are divided into 8 sectors. The results show that concentrations increase with residential land use (heating, incineration, local transport) and with

industrial/commercial land uses, but decrease with evergreens (pollution filtering). More recently, Briggs et al. (1997), making use of detailed land-use and vehicular traffic data for the cities of Huddersfield (UK), Prague, and Amsterdam, delineate circular buffers of 300 meters radius around 40 O₃ and NO₂ monitoring stations, and estimate regression models explaining average concentrations as functions of traffic volumes, industrial and high-density housing land uses, altitude, and proximity to various types of roads. The models are validated by comparing computed and observed concentrations at 10 reference sites, with excellent results.

2.4. Summary

The above review points to a heterogeneous body of models, each focusing on a specific aspect of the air pollution problem. In general, there is little comprehensive research across multiple spatial scales (local, regional, and national). Many of the studies involve site- or episode-specific air quality analyses. Except for the air back-trajectory analyses, none of the studies distinguish between local and background contributions to air pollution within a coherent spatial framework. No modeling approach could be found, that provides an empirically-based general relationship between concentrations and the multitude of factors - emissions, meteorology, atmospheric chemistry, pollution cleansing - influencing air quality at various geographical scales. It is the purpose of this paper to provide the theoretical basis and empirical testing of such a relationship.

3. **METHODOLOGY**

3.1. Theoretical Framework

At the most fundamental physical and chemical level, the process of convection, diffusion, and transformation of air pollutants is represented by a system of partial differential equations, known as the Fickian system, representing the mass balance of each pollutant species in any elementary volume

$dv = dx dy dz$ in which the pollutant species appears. This system must be solved under appropriate initial and boundary conditions. The basic equation

$$\frac{\partial C_p}{\partial t} = -\nabla(\mathbf{u}C_p) - \nabla q_p + E_p + R_p - S_p, \quad (1)$$

where C_p is the concentration of pollutant p in volume dv , simply states that the accumulation of pollutant p in dv during time dt ($= \partial C_p / \partial t$) is related to the rates of pollutant flow into and out of this volume $[-\nabla(\mathbf{u}C_p) - \nabla q_p]$, to the emission rate (E_p) from pollution sources located in dv , to the production and destruction rates of the pollutant by chemical reaction (R_p), and to the rate of pollutant removal (S_p) through natural processes. The inflow and outflow of pollutants take place through two processes: advection $[\nabla(\mathbf{u}C_p)]$ and diffusion (∇q_p), where q_p is the mass flux of pollutant p due to turbulent diffusion. Because the tendency is for mass to move from regions of higher concentrations to regions of lower concentrations, the rate of mass transfer due to diffusion is generally assumed to be proportional to the concentration gradient, with $q_p = -D\nabla C_p$. The diffusivity matrix D has components that depend on the wind field and vertical temperature gradients, and on the earth surface roughness. The chemical production rate R_p involves the production of pollutant p through the chemical interactions of other pollutants, with

$$R_p = R_p(C_1, \dots, C_j, \dots, C_N) \quad j \neq p. \quad (2)$$

R_p is generally taken as a polynomial in the concentrations, where the coefficients are the reaction rate constants obtained from laboratory experiments. Examples of such atmospheric reaction include the creations of ozone and acid rain. The emission rate E_p is a function of the characteristics X of the sources (e.g., number of housing units by type of home-heating fuel, energy needed and type of fuel used in different industries, numbers of cars by model type and vintage year), with

$$E_p = E_p(X). \quad (3)$$

The case of ozone (O₃) involves three interrelated systems of equations: one for O₃, where the emission rate is always zero, and one each for VOCs and NO_x, where the emission rates may be positive. The three systems are linked through the ozone chemical interaction equation (2). Finally the removal rate S_p is a function of atmospheric meteorological factors M (e.g., precipitation) and land-use factors Z (e.g., forest canopy) in the volume dv , with

$$S_p = S_p(M, Z). \quad (4)$$

Note that the vector Z appears in Eq. (4) only if the volume (cell) dv is adjacent to ground level.

In general, the focus is on steady-state conditions (i.e., when $\partial C_p / \partial t = 0$), and not on transitory conditions. The general approach is to seek numerical solutions by transforming Eq. (1) into a system of difference equations that permit a step-by-step numerical calculation of the concentrations at the center of each cell of a three-dimensional grid superimposed on the area analyzed.

A related approach is based on the integration of the diffusion equation over an arbitrary fixed volume, as proposed initially by Seinfeld and Kyan (1971). The airshed is divided into an array of L cells, of equal or unequal volumes. Each cell is considered a well-mixed reactor with permeable walls and movable lid, within which the pollutant is evenly distributed, hence the term “well-mixed cells model”. Consider cell k , as illustrated in Figure 1, with concentration C_{pk} for pollutant p . For the sake of simplicity, assume that the bottom side is the ground, and the upper side the base of the inversion layer. Therefore, all airflows take place through the lateral cell walls. Consider next the adjacent cell j where the pollutant concentration is C_{pj} . Let q_{jk} be volumetric rate of airflow from cell j to cell k , and q_{kj} the same rate from cell k to cell j . The total inflow of pollutant p into cell k from all other cells ($j=1 \text{ @ } L$) is

then: $\sum_{j=1}^L \mathbf{q}_{jk} C_{pj}$. Likewise, the total outflow of pollutant p from cell k to all other cells ($j=1 \text{ @ } L$) is:

$\sum_{j=1}^L \mathbf{q}_{kj} C_{pk}$. Under the assumption of steady-state conditions, Eq. (1) becomes

$$\left(\sum_{j=1}^L \mathbf{q}_{jk} C_{pj} \right) - \left(\sum_{j=1}^L \mathbf{q}_{kj} C_{pk} \right) + E_{pk} + R_{pk} - S_{pk} = 0, \quad (5)$$

which simply states that the net balance of pollutant p flows between cell k and all the adjacent cells is equal to the net pollutant accumulation within cell k due to source emissions, chemical reactions, and natural removal. Let $\mathbf{q}_k = \sum_{j=1}^L \mathbf{q}_{kj}$ be the total rate of airflow out of cell k . Equation (5) is rewritten as

$$C_{pk} = \sum_{j=1}^L \left(\frac{\mathbf{q}_{jk}}{\mathbf{q}_k} \right) C_{pj} + \frac{1}{\mathbf{q}_k} (E_{pk} + R_{pk} - S_{pk}) \quad . \quad (6)$$

The first term in Eq. (6) represents the external contributions to C_{pk} , or “regional” effect, C_{pkR} . The second term represents the internal contributions to C_{pk} , or “local” effect, C_{pkL} . Then:

$$C_{pk} = C_{pkR} + C_{pkL} \quad (7)$$

3.2. Modeling Approach

Equation (6) characterizes steady-state conditions during a given period of time (say, one hour), when meteorological conditions are not changing (e.g., constant wind speed and direction). This equation can be used in a simulation mode if all the parameters that characterize its components are known for every time period t ($=1 \rightarrow \Theta$). The length of the basic period must be small enough so that it is always characterized by steady-state conditions. As widely recognized in the literature, there is much uncertainty about these parameters, which of course translates into uncertainty about the output - the computed concentrations. The approach adopted here is to use this equation to derive an empirically testable statistical relationship (Eq. 15), which, when estimated with actual data (see Section 5), should provide, in

an aggregate and indirect fashion, estimates of the fundamental parameters of Eq. (6).

Consider first the “regional” effect component of Eq. (6), which is indexed by the time period t , while deleting the pollutant index p to simplify the presentation, with:

$$C_{ktR} = \sum_{j=1}^L \left(\frac{q_{jkt}}{q_{k,t}} \right) C_{jt} = \sum_{j=1}^L F_{jkt} C_{jt} . \quad (8)$$

During any given steady-state period t , the wind blows into only one direction. If it blows from cell j to cell k , then necessarily the two flows q_{jkt} and $q_{k,t}$ are identical, and their ratio $F_{jkt} = 1$. Otherwise, $F_{jkt} = 0$. The average annual “regional” effect is then:

$$C_{kR} = \frac{1}{\Theta} \sum_{t=1}^{\Theta} \sum_{j=1}^L F_{jkt} C_{jt} . \quad (9)$$

Equation (9) is rewritten in terms of the annual average concentration in cell j , C_j , and the annual frequency of winds in the direction $j-k$, F_{jk} . Let ΔC_{jt} be the concentration deviation from its mean at time t . Then:

$$C_{kR} = \sum_{t=1}^{\Theta} \sum_{j=1}^L \frac{F_{jkt}}{\Theta} (C_j + \Delta C_{jt}) = \sum_{j=1}^L F_{jk} C_j + \frac{1}{\Theta} \sum_{t=1}^{\Theta} \sum_{j=1}^L F_{jkt} \Delta C_{jt} \quad (10)$$

For a given cell j , the sum $\mathbf{e}_{kj} = (1/\Theta) \sum_{t=1}^{\Theta} F_{jkt} \Delta C_{jt}$ includes both positive and negative terms, which are likely to cancel each other to some extent if the wind blows from j to k with similar frequencies during periods when the concentration deviation is positive and during those when it is negative. However, even if this were not the case for a specific cell j (e.g., $\mathbf{e}_{kj} > 0$) because of a high correlation between the wind direction $j \rightarrow k$ and positive concentration deviations, this same correlation would have the opposite effects on cell m symmetrical to j with respect to k . Indeed, all cells around k are likely to be characterized, at any given time, by either negative or positive concentration deviations, as these background concentrations

result from the same large-scale air movements. It is therefore reasonable to assume that the second component in Eq. (10) is small and represents a noise/error term, \mathbf{e}_{kR} , while the first component is a good approximation of the average annual concentration in cell k (C_{kR}). Then:

$$C_{kR} = \sum_{j=1}^L F_{jk} C_j + \mathbf{e}_{kR} . \quad (11)$$

Let C_{ktL} be the “local” effect component of Eq. (6), as indexed by t , with:

$$C_{ktL} = \frac{1}{q_{k,t}} (\mathbf{E}_{kt} + \mathbf{R}_{kt} - \mathbf{S}_{kt}) . \quad (12)$$

The average annual “local” effect is:

$$C_{kL} = \frac{1}{\Theta} \sum_{t=1}^{\Theta} C_{ktL} , \quad (13)$$

and the average annual “total” concentration at k is:

$$C_k = \sum_{j=1}^L F_{jk} C_j + C_{kL} + \mathbf{e}_{kR} . \quad (14)$$

Consider now a monitoring station O where the concentration C_O of some air pollutant is measured. This station may be viewed as the center of a circle, base of a cylindrical vertical cell up to the inversion layer, as illustrated in Figure 2. This circle is located, fully or partially, within the boundaries of an urban area. Consider next the four quadrants centered at station O , and, in each quadrant, the monitoring station closest to O . Let C_1 , C_2 , C_3 , and C_4 be the pollutant concentrations at these four stations. In each quadrant, consider the area between the dashed arc of radius d , and the arc of the circle surrounding the central station. This area may be viewed as the basis of a cell. The central circular cell and the four quadrant cells can be viewed as a well-mixed cell system, with homogeneous concentrations measured by C_O , C_1 , C_2 , C_3 , and C_4 .

The component C_{OL} represents the local contributions to concentration, due to emissions from the

residential, commercial, industrial and transportation sectors, local atmospheric chemical reactions, and pollutant removal through natural processes. Let N be a vector of population and housing variables, A a vector of economic activity variables, B a vector of transportation variables, Z a vector of land-use variables, and M a vector of meteorological variables. A generalized formulation of the local contribution function C_{OL} would then be: $C_{OL} = F(N, A, B, Z, M)$. While most of the variables in the vectors (N, A, B, Z) may be viewed as proxies for local emission variables, and those in M as proxies for local atmospheric effects, it is important to note that some of the variables in Z , in particular the amounts of forests and other natural systems, may measure the air pollution control effects of natural features, via direct absorption/filtering effects (i.e., the removal factor S_p in Eq. 4), as well as through cooling effects that may reduce energy consumption in the summer, and hence the resulting pollution generated (McPherson et al., 1998; Sailor, 1998; Cionco and Ellefsen, 1998). If F_{jO} is the annual frequency of winds blowing towards O within sector j , Eq. (14) is rewritten as:

$$C_o = \sum_{j=1}^4 F_{jO} C_j + F_L(N, A, B, Z, M) + \mathbf{e}_{OR} \quad . \quad (15)$$

Equation (15) can be either linear or nonlinear, depending upon the structure of the local function F_L . Whatever the function selected, it is very unlikely to capture all the local effects, and all the unobservable effects will sum up as a local error term \mathbf{e}_{OL} that will be added to the regional error term \mathbf{e}_{OR} . Under the assumption that a sample of observation units, made of a central and four sectoral monitoring stations, can be created, the goal of the empirical analysis is then to estimate the parameters of the function F_L . Finally, note that, with reference to the literature review, the regional effect component bears a clear connection with air back-trajectory models, and the local effect component with urban, aggregate models.

4. DATA

4.1. Air Pollution Concentrations

Originally extracted from the U.S. Environmental Protection Agency (EPA) Aerometric Information Retrieval System (AIRS), air pollution data have been downloaded from the Web site of the Center of Air Pollution Impact and Trend Analysis (CAPITA)¹, and include *hourly* concentrations (in p.p.b.: parts per billion) of O₃, NO (nitric oxide), NO₂, NO_x, SO₂, CO, and NMHC (non-methane organic compounds), over six months (April-September) in each of the years 1986-1995, at monitoring stations throughout the U.S. The data also includes the coordinates of these stations.

All the 797 monitoring stations for O₃ and 418 stations for CO, that had been active for the full 6 months of 1990, were retained, and 6-month (April-September) average concentrations of O₃ and CO were computed for each station for the year 1990. By averaging concentrations, the temporal variability over these six months is obviously lost. However, this loss is acceptable because the goal of this research is to uncover general average relationships, as outlined by Eq. (14). Approximately 90% (375) of the CO stations are inside urban areas, but only 60% (479) in the case of O₃. This pattern is, of course, a reflection of the fact that CO is primarily an urban problem, whereas O₃ concentrations have been found to be often higher in rural areas, where they must be monitored.

4.2. Meteorology

Meteorological data have been downloaded from the Solar and Meteorological Surface Observation Network (SAMSON) at the CAPITA Web site, for 232 meteorological stations over the period 1961-1992. This database contains discrete daily values for 37 atmospheric variables, including wind speed and direction, solar radiation, sky cover, temperature, humidity, precipitation, visibility, and ceiling height. The variables retained are those that have a clear bearing on the processes of pollutant formation, dispersion, and transport (wind direction, wind speed, solar Direct Normal Radiation - DNR, and

temperature), and only for the 225 stations active in 1990. Another solar radiation measure available in SAMSON is the Global Horizontal Radiation (GHR). However, only the DNR includes the ultra-violet component, which plays an active role in the formation of O₃. While both the DNR and GHR are highly correlated (0.92), the DNR is retained in the following analyses. Note also that only 3% of the DNR data are measured, the rest being modeled.

Since weather stations are generally not collocated with pollutant monitoring stations, the nearest weather station is assigned to each monitoring station so as to approximate meteorological conditions around the monitoring site. In the final samples used to estimate equations (24) and (25), the median distances between weather and air quality monitoring stations are 32 km for O₃ (n=319) and 23 km for CO (n=117). The corresponding 75% quantiles are 51 km and 41 km, and the 90% quantiles are 76 km and 60 km. These relatively small distances suggest that the meteorological data are representative of the conditions at the air quality station, and therefore should not bias the results of the forthcoming analyses. To be consistent with the quadrant-cell approach (Section 3.2), wind direction at each station is represented by its frequency within each of the 4 quadrants. Average values for wind speed, temperature, and solar radiation, and wind direction frequencies are computed with the discrete daily values over the period April-September 1990, in concurrence with the average pollution concentrations.

4.3. Land Use

Digital maps of the Land Use and Land Cover (LULC) database have been downloaded from the U.S. EPA Web site². Based on NASA high-altitude aerial photo-coverage and survey, the LULC has been created by the U.S. Geological Survey during the late 1970's and early 1980's, to provide land use and land cover information for the whole U.S territory. Despite an extensive search, more recent digital maps (e.g., for 1990) with the same comprehensiveness could not be found, and most likely do not exist, although they might become available with the release of formerly classified past satellite data.

Land uses are classified into 9 major categories: urban and built-up, agricultural, range, forest, water, wetlands, barren, tundra, and perennial snow ice. Each major category is further divided into 2-digit sub-categories (see Anderson et al., 1976). Urban features are captured with the resolution of 10 acres, and the remainder with a 40-acres resolution. Given these limits, the coverage does not include undetectable objects like very narrow or long land uses. Although the LULC data are somewhat outdated to describe the fringes of urban areas in 1990, they are most likely adequate to depict land uses patterns in the core of those areas, and in smaller places that have not experienced growth during the 1980s.

The LULC data are organized by quadrangles. As the area around a monitoring station may spread across two or more quadrangles, it may be necessary to edge-match and join adjacent quadrangles. However, joining all U.S. quadrangles to maximize sample size is a difficult task in terms of computer capacity and time limitations. Therefore, quadrangles with a high density of stations, in particular in California, the Northeast, and the Great Lakes Region, are selected as major sampling areas. In order to maintain sample diversity, quadrangles covering lightly populated areas with monitoring stations are also selected, in parts of Alabama, Arizona, Colorado, Florida, and Texas. As a result, 345 O₃ stations and 247 CO stations are retained in the database.

4.4. Transportation Network

Based on the 1995 TIGER/Line³ files, an electronic map of all U.S. major roads is available in the form of an ArcView⁴ Shape file in a sample data set from ESRI. These roads belong to three classes. The roads in class A1 include interstate and some toll highways, divided by a median strip and accessible by way of ramps. Those in class A2 are nationally and regionally important highways with unlimited access. Those in class A3 are secondary and connecting roads. Only roads in class A1 and A2, categorized as primary roads, are used in this study, because of computing limitations. However, this omission is not likely to be a problem, because the intersections between roads in any two categories are

also retained. Pollution levels are likely to be high if a monitoring station is located close to a road and an intersection. Thus, several parameters characterizing this proximity are computed, using GIS techniques, including: (1) the shortest distances from a central station to a primary road and to an intersection, (2) the total number of intersections within a given radius, and (3) the sum of the distances to all these intersections. In order to perform sensitivity tests of the local effects from roads, the radii considered are 5, 10, and 15 km.

4.5. Socioeconomics

Population, housing, employment and commuting data are drawn from the 1990 Census of Population and Housing. Population and housing variables are taken as proxies for residential pollution emissions, including (1) numbers of persons and households, and (2) numbers of housing units and type of space-heating fuel (utility gas, fuel oil, electricity, coal). Employment data include the numbers of people employed in major economic sectors (agriculture, mining, construction, manufacturing, etc.). Although these employment data characterize the place of residence, they are taken as proxies for production output, itself linked to pollution emission. Commuting variables related to travel time and modes of transportation are taken as proxies for pollution emissions from transportation. While pollution is also produced by other types of trips (shopping, recreation), no comprehensive data could be found on the flows of these trips, which are assumed proportional to commuting flows.

5. **RESULTS**

5.1. Regional Cell Size

The first empirical issue is the determination of the quadrant cell size, as measured by the radius d of the outer boundary circle (the radius of the inner boundary circle is always 15 km - see Figure 2). In order to determine the maximum value of d that maintains the homogeneity of well-mixed cells, a sensitivity

analysis has been conducted with d varying over the ranges [40 - 450 km] in the case of O_3 , and [50 - 300 km] in the case of CO, while estimating the following model:

$$C_o = \mathbf{b}_o + \mathbf{b}_c \left[\sum_{j=1}^4 F_{jo} C_j \right] + \mathbf{e} \quad (16)$$

The lower bounds of the two ranges (40 km, 50 km) have been selected to allow for the creation of a minimum-size sample of central stations surrounded by quadrant stations within distance d . In the case of O_3 , there are 49 observations with quadrant stations between 15 and 40 km, and, in the case of CO, 24 observations with quadrant stations between 15 and 50 km. As C_o is the total (observed) concentration (6-month average) at station O , the intercept \mathbf{b}_o represents the average value of the local effect (C_L), and the error term \mathbf{e} includes the deviations of both local and regional effects. The coefficient \mathbf{b}_c should, based on Eq. (15), be set equal to 1 (constrained regression). However, it is allowed to be endogenously determined, and represents the rate of survival of the pollutant on its way from the regional cells to the central cell. Equation (16) is estimated by OLS regression.

Table 1 presents the estimates of \mathbf{b}_o and \mathbf{b}_c in the O_3 case, with distances varying by 10 km increments. \mathbf{b}_c varies within a narrow range [0.85 - 1.16], and is highly significant and increasingly so until d reaches 370 km. Beyond this distance, the significance of \mathbf{b}_c decreases monotonously. The coefficient of determination (R^2) declines very gradually with d , explaining 34% to 50% of the variability of the concentration. When $d \geq 180$ km, the constant term \mathbf{b}_o becomes significant. As the distance d increases from 40 km to 450 km, the number of central stations surrounded by a station within each quadrant (i.e., the sample size) increases from 49 to 640. The high values of \mathbf{b}_c (close to but below 1) confirm that there is little decay of O_3 when transported in the atmosphere. Finally, the error term \mathbf{e} was found to be uncorrelated with the regional effect variable ($\sum_{j=1}^4 F_{jo} C_j$), which indicates that the estimates of \mathbf{b}_o and

b_c are unbiased, and that all the omitted variables and the regional effect have separate and additive influences on C_O .

With respect to CO, when $d \geq 60$ km, b_c is positive and significant, supporting the expectation that C_O is positively influenced by regional effects (Table 2). The CO model explains 15% to 20% of the variability of the concentration. When $d \geq 240$ km, the R^2 declines to 0.15. The intercept b_0 is always significant. When d increases from 50 km to 300 km, the sample size increases from 24 to 197. Note that the value of b_c lies mostly within the range [0.6-0.7], which points to a much higher rate of decay than in the case of O_3 .

In order to further verify that Eq. (16) is appropriate for estimating regional effects, it is tested by considering the following hypotheses: Hypothesis 1: the wind speed within sector i and the distance between stations i and O have an effect on the concentration; and Hypothesis 2: each quadrant concentration has a distinct regional effect (i.e., different coefficients apply to the 4 quadrants). Rejection of both hypotheses would further support the well-mixed cell theory that underlies this empirical analysis. Acceptance of either hypothesis would undermine it. Also, the results of these two analyses should further help identifying the boundary of the regional cells, provided that the theory is deemed acceptable.

To test Hypothesis 1, Eq. (16) is modified to account for the distance d_j between stations j and O , and the average wind speed u_j within sector j , with:

$$C_O = b_0 + b_c \sum_{j=1}^4 d_j^a u_j^b F_{jO} C_i + e \quad (17)$$

Equation (17) has been estimated through nonlinear regression. The results, available in Kim (1999), show that the parameters a and b in both the O_3 and CO models are insignificant at all distances.

Testing Hypothesis 2 implies two successive steps. The first step is to estimate Eq. (18) where distinct coefficients b_j apply to the four quadrants, with:

$$C_o = \mathbf{b}_o + \sum_{j=1}^4 \mathbf{b}_j F_{j0} C_j + \mathbf{e} . \quad (18)$$

The second step is to test whether the four coefficients are identical, or whether having separate \mathbf{b}_j 's significantly improves the explanatory power of the model. In order to test the null hypothesis $\mathbf{b}_1 = \mathbf{b}_2 = \mathbf{b}_3 = \mathbf{b}_4$, the general F -test statistics is calculated, based on the error sum of squares of the detailed model SSE(F) and the error sum of squares of the reduced model SSE(R). The detailed results are available in Kim (1999). In the case of O₃, when $d \leq 370$ km, the two models are not significantly different. When $d \geq 380$ km, the quadrant-specific model turns out to be better. As indicated earlier, at $d = 380$ km, the R^2 starts decreasing with d , after its stabilization within the distance range [240-370 km] (Table 1). This result suggests that the well-mixed regional cell model is valid up to 370 km, which is selected as the appropriate regional cell boundary for O₃. With regard to CO, there is no significant difference between the two models for any distance. As discussed earlier (Table 2), the pattern of R^2 when $d \leq 230$ is different from that when $d \geq 240$. To maximize the sample size without much reduction in the R^2 , the distance of 230 km is selected as the regional cell boundary for CO. Overall, the results of testing Hypotheses 1 and 2 support the well-mixed cell approach to represent regional pollution effects.

5.2. Model Estimation

5.2.1 *Overview*

Combining Eqs. (15) and (16) provides the general form of the model to be estimated:

$$C_o = \mathbf{b}_c REFF + F_L(N, A, B, Z, M) + \mathbf{e} , \quad (19)$$

where $REFF = \sum_{j=1}^4 F_{j0} C_j$. We assume that the function F_L can be expressed as follows:

$$F_L(N, A, B, Z, M) = G_L(M) H_L(N, A, B, Z) , \quad (20)$$

where G_L represents meteorological effects, such as chemical transformations and dispersion, and H_L

represents the aggregate effect of pollution emissions from the residential sector (N), the economic sector (A), the transportation sector (B), as modified by the cleansing effects of the environment, measured by specific land-use variables (Z). The variables that make up the meteorological vector M include the average (over April-September 1990) temperature, T , direct normal solar radiation, S , and wind speed, U . In the case of O_3 , the primary factors are T and S , and the function G_L is taken as

$$G_L^{O_3} = S^g T^d \quad (21)$$

where g and d are positive parameters to be determined. Equation (21) implies that, for a given pattern of O_3 precursor emissions, the resulting concentration increases with S and T . In the case of CO, meteorological conditions influence concentrations in indirect ways. Lower temperatures cause motor vehicles to produce more CO than at higher temperatures. Also, wind speed increases the dispersion of CO. Hence, the function G_L is taken as

$$G_L^{CO} = T^l U^m, \quad (22)$$

where l and m are negative parameters to be determined. The mathematical structure of the function H_L is unknown, but can be approximated by a polynomial development. In the following, we focus on a linear development, with

$$H_L(N, A, B, Z) = b_N N + b_A A + b_B B + b_Z Z \quad (23)$$

Second-order (quadratic) developments have also been considered, but their estimation turned out to be unsuccessful, primarily because of multicollinearity.

The next issue is the selection of the specific variables making up Eq. (23). The database includes 36 variables in the socioeconomic category (N, A, B), 11 variables in the transportation network category (B), and 30 variables in the land-use category (Z). The search for the “best” model is a difficult task, due

to the large number of possible variables, and much exploratory analysis is therefore necessary. The final selection of variables accounts for both the overall explanatory power of the model, as measured by its R^2 , the importance of a variable from a theoretical viewpoint, and its significance. Because of the inherent variable correlations (e.g., residential area vs. population), special care was taken to minimize multicollinearity, which yields imprecise estimates (Gujarati, 1988). In order to identify quasi-linearity between any two variables, Pearson correlations were computed. After entering a variable into a regression, variation inflation factors (VIF) were computed, that measure the inflation of the estimated regression coefficients, as compared to when the explanatory variables are not linearly related (Neter et al. 1996). The VIF values of all the selected variables have been kept below 2, thus making multicollinearity a negligible problem. Details on the exploratory analyses can be found in Kim (1999).

5.2.2. *Ozone Model*

The variables retained in the function H_L are:

POP: total population of the place;

ELEC: percentage of housing units using electricity for space heating;

HGFUEL: percentage of housing units using oil, wood, or coal for space heating;

CONSTR: percentage of the workforce employed in the construction sector;

PUBT: percentage of workers using public transportation on the journey-to-work;

FRST: percentage of deciduous and mixed forest land.

Because central stations are constrained to be located within urban places, the final sample is reduced to 303 observations. Descriptive statistics are presented in Table 3. Equation (19) has been estimated through OLS regression while varying exogenously the parameters (and * (Eq. 21) over the range [0-3].

The model with the highest R^2 is obtained with (=1.49 and *=2.05, and is presented below:

$$C_o^{O_3} = 8.81 + 0.665REFF + [1.46 * 10^{-13}POP + 5.04 * 10^{-6}CONSTR - 4.25 * 10^{-8}PUBT$$

$$(5.02) \quad (11.42) \quad (2.19) \quad (8.47) \quad (4.87)$$

$$- 5.08 * 10^{-9}ELEC + 8.38 * 10^{-9}HGFUEL - 6.49 * 10^{-9}FRST] * S^{1.49} * T^{2.05}$$

$$(4.48) \quad (2.11) \quad (2.60)$$

$$R^2 = 0.565 \quad F=54.6 \quad (24)$$

The t-statistics are presented in parentheses below each estimated coefficient. The model explains about 56% of the variability of C_o . There is no correlation between the error (residuals) and each explanatory variable, pointing to the unbiasedness of the estimated coefficients. All the coefficients are statistically significant, and have the expected signs. Indeed, a larger population generates more chemical precursors necessary for O_3 formation, hence an increase in population should have a positive impact on C_o . Similarly, an increase in $CONSTR$ is expected to have a positive impact on C_o , since the construction industry is an emission source of VOCs. Important VOC sources include motor vehicles, gasoline vapors, and chemical solvents, related to traffic congestion taking place around construction sites, and to the need by construction activities for a lot of petroleum-related products (paints, coatings, etc.). In contrast, an increase in $PUBT$ is expected to have a negative influence on C_o , because a higher public transportation ridership reduces the total number of private vehicles on the roads, leading to a decrease in VOC and NO_x emissions. An increase in $HGFUEL$ is expected to have a positive impact on C_o , because fuel combustion increases the production of NO_x and VOCs. However, an increase in $ELEC$ is expected to have a negative impact, as electric heating substitutes for fuel combustion within the residential sector. An increase in $FRST$ is expected to have a negative impact on C_o due to the uptake of VOCs and NO_x by vegetation/forests. The two meteorological variables S and T are known to have positive effects on C_o , since O_3 formation is activated by intense solar radiation and high temperature. However, these effects

are nonlinear, as increasing radiation and temperature accelerate the formation of O₃.

5.2.3. Carbon Monoxide Model

The variable retained in the function H_L are:

- TTD**: total commuting distance (miles) by workers (computed by using the average travel time and a traffic flow speed of 40 miles/hour);
- AGR**: percentage of agricultural land within the 15 km circle boundary;
- AVDND5**: average distance (meters) from the CO monitoring station to all intersections within a 5-km radius.

Equation (19) has been estimated through OLS regression while varying the parameters δ and γ : (Eq. 22) over the range [-1,0]. The highest R² is obtained when $\delta = \gamma = 0$, and therefore the meteorological function G_L^{CO} (Eq. 22) is deleted from the model. The final sample includes 117 stations located within urban places. Descriptive statistics are presented in Table 4. The final equation, with t-statistics in parentheses below each estimated coefficient, is:

$$C_O^{CO} = 1135.962 + 0.473 \mathbf{REFE} + 1.30 * 10^{-5} \mathbf{TTD} - 3.332 \mathbf{AGR} - 0.192 \mathbf{AVDND5} \quad (25)$$

(4.60) (3.64) (5.57) (-1.71) (-2.71)

R² = 0.43 F=21.54

About 43% of the variations of C_O are explained by the model. The signs of all the coefficients are consistent with the initial expectations. The higher **TTD**, the more traffic and CO emissions, and the higher the CO concentrations. The variable **AGR** is negatively related to C_O , pointing to the cleansing effect of the natural environment. With respect to the variable **AVDND5**, the closer road intersections are to the station, the higher the CO concentration. Finally, the fact that the concentration is a summertime average may explain why both temperature and wind speed have no noticeable effects. Indeed, low winter temperatures are known to force vehicle engines to consume more gasoline and generate more pollution, whereas the range of summer temperatures may have little effect on engines.

Also, wind speeds are generally lower in summer, with frequent calm situations, again little conducive to concentration reduction. Thus, using summer-average CO concentrations may hide the effects of temperature and wind speed, which might become apparent when using winter-average concentrations (unfortunately unavailable from CAPITA). However, there is no reason to assume that the effects of the other variables, as indicated in Eq. (25), would be any different in winter.

5.3. Elasticity Analyses

While the explanatory power of the estimated models (24) and (25) precludes their use for predictive purposes due to the magnitude of the error terms, their coefficients are precisely estimated and can be used for impact analyses. In order to assess the importance of the independent variables in determining concentration levels, the elasticities of concentrations with regard to each variable have been computed. The elasticity of C_O with regard to a specific variable V is measured by

$$e = (\partial C_O / \partial V) / (C_O / V) , \quad (26)$$

and indicates the percentage increase in C_O resulting from a 1% increase in V , all other variables being held constant. These elasticities are more appropriate to measure the magnitudes of impacts than the regression coefficients themselves, which are unit-dependent and thus are not directly comparable. The elasticities have been computed for each variable set equal to its sample mean value, and two other values, one standard deviation above and below this mean (1+STD, 1-STD). Whatever the value selected for variable V , all the other variables are set equal to their sample mean values.

5.3.1. *The Case of Ozone*

The elasticity functions for the O₃ model are presented in Table 5, and the values of the elasticities in Table 6. At the sample mean, a 1% increase in the background pollution flow, *REFF*, leads to a C_O increase of 0.70%. While this effect is inelastic, it is the most influential. Thus, O₃ control cannot be solely achieved at the local level, but must account for regional and large-scale effects. At one standard

deviation above the mean *REFF* value (1+STD), a 1% increase in *REFF* leads to a C_O increase of 0.73%. At one standard deviation below (1-STD), a 1% increase in *REFF* leads to a C_O increase of 0.66%.

At the sample mean, a 1% increase in solar radiation, *S*, or temperature, *T*, leads to C_O increases of 0.003% and 0.004%, respectively. While *S* and *T* cannot be controlled through policy decisions, these elasticities have implications for the design of air pollution warning systems, as well as for the control of other variables. For instance, at the (1+STD) value, a 1% increase in *S* leads to a C_O increase of 0.005%. If this increase is expected to have an adverse impact on human health and the environment, the variables which affect the elasticity of *S* (i.e. *CONSTR*, *PUBT*, *ELEC*, *HGFUEL*, and *FRST*) may be manipulated to offset the increase in C_O , i.e., to make the elasticity of C_O to *S* equal to zero (see formula in Table 5). For example, if the public transit ridership, *PUBT*, is used as a policy tool to control C_O , the city must provide transit services so that about 5.5%⁵ of the population can use public transportation. Likewise, if the forested land, *FRST*, is to be used as a control instrument, about 14.4% of the area around the monitoring station should be used as forested land. Thus, the effect of *S* can be offset by controlling other influential variables. At the lower range (1-STD), a 1% increase in *S* leads to a C_O increase of 0.002%. Likewise, a 1% increase in the temperature *T* at (1+STD) leads to a C_O increase of 0.006%, and, at the (1-STD) level, to an increase of 0.003%.

With respect to the socioeconomic variables, a 1% increase (2,959 persons) in the average total population, *POP*, results in a C_O increase of 0.010%. At the (1+STD) level, a 1% increase in *POP* leads to a C_O increase of 0.038%. At the sample mean, an increase of 1% in the share of construction industry employment, *CONSTR*, leads to a C_O increase of 0.069%. Also, a 1% increase in *CONSTR* at the (1+STD) level leads to a C_O increase of 0.100%. In contrast, at the sample mean, a 1% increase in

public transportation ridership, *PUBT*, leads to a C_O decrease of 0.051%, and at the (1+STD) level, to a decrease of 0.141%. Thus, these results confirm that increasing public transportation ridership may be an important policy tool for O_3 reduction. The house fuel consumption variables, *ELEC* and *HGFUEL*, also provide useful results in terms of their elasticities: at the sample mean, *ELEC* has an elasticity of -0.022 , and *HGFUEL*, an elasticity of 0.016. Thus, increasing the level of *ELEC* (i.e., the number of housing units using electricity for space heating) may be a more effective way to control O_3 than reducing the level of *HGFUEL* (i.e., the number of housing units using oil, wood, or coal). To reduce O_3 pollution, a policy could be implemented that encourages the construction of electricity space-heating housing units, and the shifting away from wood, coal, and petroleum burning. Finally, a 1% increase in forested land, *FRST*, leads to a C_O decrease of 0.019%. At the (1+STD) level, the decrease would be 0.047%. Thus, a policy of conserving or extending forested areas in and around a city may lead to a decrease in ozone pollution. Overall, as is apparent in Table 6, all elasticities increase with the level of the input variables.

5.3.2. *The Case of Carbon Monoxide*

The CO elasticity functions and elasticities at the three sample levels are presented in Table 7. At the sample mean, the elasticity of the CO regional component (0.41) is smaller than in the case of O_3 (0.70). CO is less responsive to background concentrations than O_3 because background CO decays more rapidly than background O_3 , which confirms that CO is a more locally-oriented problem. However, the impact of background CO is by no means negligible. At the sample mean level, a 1% increase in total travel distance, *TTD*, leads to a C_O increase of 0.077%. In an urban area with the average number of commuting workers (313,792 persons), if 20% of the workers were to reduce their work-trip distance by one mile, C_O would decrease by 0.8 p.p.b. This could be achieved by using public transportation or carpooling. At the sample mean, a 1% increase in *AGR* leads to a decrease in C_O of 0.07%. As the most

influential variable, a 1% increase in the average distance to all intersections within a 5 km radius, *AVDND5*, leads to a C_O decrease of 0.56%. Furthermore, at the (1+STD) level, C_O decreases by 0.74% for a 1% increase in *AVDND5*. Thus, the results confirm that intersections are truly important emission sources of CO. Constructing road interchanges instead of intersections, and implementing traffic signal management at intersections might be useful to reduce CO levels. This result also suggests the need for a uniform location policy for monitoring stations. As CO concentrations are very sensitive to the locations of the monitors, the latter should always be set at the same distance from the closest intersections, to ensure that comparable concentration measures are obtained and consistent air quality policy decisions are derived. Finally, retaining and expanding agricultural land within urban areas may be another important approach to CO control.

6. CONCLUSIONS

Based on a spatial framework derived from the theoretical concept of well-mixed cells, linear regression models have been estimated with cross-sectional data on U.S. urban areas in 1990, to assess the relationships between the average urban concentrations of ground-level ozone and carbon monoxide, and various urban factors. These models integrate both regional/national and local air pollution effects, while considering land use patterns, transportation networks, socio-economic factors, and meteorology, through an extensive use of Geographical Information Systems (GIS). While the overall explanatory power of these models is good, it is not sufficient for concentrations forecasting, due to the magnitude of the error term. However, because their coefficients are precisely estimated, these models can be reliably used for impact assessment and policy analysis.

The ozone model explains about 56% of the variability in O_3 concentrations, with significant regression coefficients for all explanatory variables, including the regional effect variable (background

concentrations) and the local effect variables (total population, employment in the construction industry, public transportation ridership, housing units with different space-heating fuels, and forested land). The model also accounts for solar radiation and temperature in its local component. The results show that O₃ background concentrations are the most critical factor impacting local O₃ concentrations, and confirm that O₃ is transported over long distances, with minimal decay. Heat and sunlight interact, as expected, with O₃ precursors, the emissions of which are proxied by the local variables. To reduce O₃ concentrations, the use of public transportation and the construction of electric space-heating housing units should be encouraged. In addition, the conservation of forested land may also be a tool to reduce O₃ concentrations. The carbon monoxide model explains 43% of the variations of CO concentrations. The explanatory variables include a regional effect variable (background concentrations) and local effect variables (total travel distance, agricultural land, and average distance to all intersection within a 5 km radius). The results show that CO is less a regional problem than O₃, although the relationship with background concentrations is significant. Unlike O₃, there is no interaction between local CO emissions and meteorological variables, such as temperature and wind speed, but this result may be due to the use of summertime average concentrations. This interaction might become apparent when using winter or year-round average concentrations. All the explanatory variables related to transportation are very significant, confirming that the dominant source of CO is motor vehicles. The closer road intersections are to a station and the higher the work-trip mileage, the higher the resulting CO concentration. These results imply that CO reductions may be achieved by providing intersection improvement and better transit services, but also that monitoring stations should be similarly located with respect to road intersections. In addition, the conservation of agricultural land may help control CO concentrations through the cleansing effect of the natural environment.

Further research should involve: (1) using new and updated data, in particular major point source

emissions, traffic flow and congestion measures, and more recent land-use/land cover information; (2) estimating alternative functional forms (e.g. nonlinear) of the models; and (3) expanding and estimating the models within a multi-period framework, which would make it possible to assess the effects of time-linked urban factors, such as traffic flows. Research is underway in some of these areas and will be reported in the near future.

Footnotes

1. <http://capita.wustl.edu/capita>
2. <ftp.epa.gov/pub/spdata/EPAGIRAS>
3. Topologically Integrated Geographic Encoding and Referencing data format (Bureau of the Census).
4. Desktop mapping and GIS software from Environmental Systems Research Institute (ESRI).
5. $PUBT = - (\beta_2 * POP + \beta_3 * CONSTR + \beta_5 * ELEC + \beta_6 * HGFUEL + \beta_7 * FRST) / \beta_4$ (elasticity of S = 0).

REFERENCES

Anderson R.F. and Greenberg M.R., 1981. The Associations Between Land Use and Particulate Air Quality in New Jersey, 1973-1977, Modeling and Simulation 12: 191-196.

Annand W.J.D. and Hudson A.M., 1981. Meteorological Effects on Smoke and Sulfur Dioxide Concentrations in the Manchester Area, Atmospheric Environment 15, 5: 799-806.

Arndt R.L., Carmichael G.R., Streets D.G., and Bhatti N., 1997. Sulfur Dioxide Emissions and Sectorial Contributions to Sulfur Deposition in Asia, Atmospheric Environment 31, 10: 1553-1572.

Ashbaugh L.L., 1983, A Statistical Trajectory Technique for Determining Air Pollution Source Regions, Journal of the Air Pollution Control Association 33, 11: 1096-1098.

Brankov E., Rao S.T., and Porter P.S., 1998. A Trajectory-Clustering-Correlation Methodology for Examining the Long-Range Transport of Air Pollutants, Atmospheric Environment 32, 9: 1525-1534.

Briggs D.J., Collins S., Elliot P., Fischer P., Kingham S., Lebret E., Pryn K., Van Reeuwijk H., Smallbone K., and Van Der Veen D., 1997. Mapping Urban Air Pollution Using GIS: A Regression-Based Approach, International Journal of Geographical Information Science 11, 7: 699-718.

Broder I.E., 1986. Ambient Particulate Levels and Capital Expenditures: An Empirical Analysis, Proceedings of the American Statistical Association (Business and Economic Statistics Section): 288-293.

Burton R.M., Suh H.H., and Koutrakis P., 1996. Spatial Variation in Particulate Concentrations Within Metropolitan Philadelphia, Environmental Science and Technology 30, 2: 400-407.

Chock, D. P. et al., 1975. "Time-Series Analysis of Riverside, California Air Quality Data", Atmospheric Environment 9, 978-989.

Cionco R.M. and Ellefsen R., 1998. High Resolution Urban Morphology Data for Urban Wind Flow Modeling, Atmospheric Environment 32, 1: 7-17.

Cleveland, W. S. et al., 1977. "Geographical Properties of Ozone Concentrations in the

- Northeastern United States”, Journal of the Air Pollution Control Association, **27**, 4: 325-328.
- Cooke, M. J. and Wadden, R. A., 1981. “Atmospheric Factors Influencing Daily Sulfate Concentrations in Chicago Air”, Journal of the Air Pollution Control Association **31**, 11: 1197-1199.
- EPA, 2000a. National Air Pollutant Emission Trends, 1900-1998. Report EPA-454/R-00-002, U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Research Triangle Park, NC 27711 - March.
- EPA, 2000b. Latest Findings on National Air Quality: 1999 Status and Trends. Report EPA-454/F-00-002, U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Research Triangle Park, NC 27711 - August.
- Glen W.G., Zelenka M.P., and Graham R.C., 1996. Relating Meteorological Variables and Trends in Motor Vehicle Emissions to Monthly Urban Carbon Monoxide Concentrations, Atmospheric Environment **30**, 24: 4225-4232.
- Gifford F.A. and Hanna S.R., 1973. Modeling Urban Air Pollution, Atmospheric Environment **7**: 131-136.
- Gujarati, D. N., 1988. Basic Econometrics, 2nd ed., McGraw-Hill Company.
- Guldmann J.-M. And Shefer D., 1980. Industrial Location and Air Quality Control: A Planning Approach, John Wiley & Sons, New York.
- Hewitt, C. N., 1991. “ Spatial Variations in Nitrogen Dioxide Concentrations in an Urban Area”, Atmospheric Environment, **25B**, 429-434.
- Horie Y., 1974. A General Rollback Model of Regional Air Pollution Control Planning, Environment and Planning A **6**: 215-228.
- Kim, H.Y., 1999. GIS-Based Statistical Models of Urban and Regional Air Quality: The Cases of Ozone and Carbon Monoxide. Ph.D. Dissertation, The Ohio State University, Columbus, Ohio.
- Kahn M.E., 1997. Particulate Pollution Trends in the United States, Regional Science and Urban Economics **27**: 87-107.
- Lioy P.J., Mallon R.P., Lippmann M., Knelp T.J., and Samson P.J., 1982. Factors Affecting the Variability of Summertime Sulfate in a Rural Area Using Principal Component Analysis, Journal of the Air Pollution Control Association **32**, 10: 1043-1047.
- McMullen T.B., Fensterstock J.C., Faoro R.B. and Smith R., 1968. Air Quality and Characteristic Community Parameters, Journal of the Air Pollution Control Association **18**, 8: 545-549.
- McPherson E. G., Scott K.I., and Simpson J.R., 1998. Estimating Cost Effectiveness of Residential Yard Trees for Improving Air Quality in Sacramento, California, Using Existing Models, Atmospheric Environment **32**, 1: 75-84.

- Milionis A.E. and Davies T.D., 1994a. Regression and Stochastic Models for Air Pollution-I. Review, Comments and Suggestions, Atmospheric Environment 28, 17: 2801-2810.
- Milionis A.E. and Davies T.D., 1994b. Regression and Stochastic Models for Air Pollution-II. Applications of Stochastic Models to Examine the Links Between Ground-Level Smoke Concentrations and Temperature Inversions, Atmospheric Environment 28, 17: 2811-2822.
- Moody, J. L. and Samson, P. J., 1989. "The Influence of Atmospheric Transport on Precipitation Chemistry at Two Sites in the Midwestern Unites States", Atmospheric Environment 23, 10: 2117-2132.
- Moy, L. A. et al., 1994. "Relationship Between Back Trajectories and Tropospheric Trace Gas Concentrations in Rural Virginia", Atmospheric Environment 28, 17: 2789-2800.
- Neter, J. et al., 1996. Applied Linear Statistical Models, 4th ed., IRWIN.
- Nevers De N. and Morris J.R., 1975. Rollback Modeling: Basic and Modified, Journal of the Air Pollution Control Association 25, 9: 943-947.
- Powell M.R., 1997. Three-City Air Study, Discussion Paper 97-29, Resource for the Future, Washington DC.
- Roy, R. and Pellerin, J., 1982. "On Long Term Air Quality Trends and Intervention Analysis", Atmospheric Environment 16, 1: 161-169.
- Sailor D.J., 1998. Simulations of Annual Degree Day Impacts of Urban Vegetative Augmentation, Atmospheric Environment 32, 1: 43-52.
- Seinfeld, J. H., and Kyan, C. P., 1971. "Determination of Optimal Air Pollution Control Strategies", Socio-Economic Planning Sciences, 5, 173-190.
- Seinfeld J.H., 1988. Ozone Air Quality Models, Journal of the Air Pollution Control Association 38, 5: 616-645.
- Shi J.P. and Harrison R.M., 1997. Regression Modeling of Hourly NO_x and NO₂ Concentrations in Urban Air in London, Atmospheric Environment 31, 24: 4081-4094.
- Suh H.H., Allen G.A., Koutrakis P., and Burton R.M., 1995. Spatial Variation in Acidic Sulfate and Ammonia Concentrations Within Metropolitan Philadelphia, Journal of the Air and Waste Management Association 45: 442-452.
- Wolff G.T. and Liou P.J., 1978. An Empirical Model for Forecasting Maximum Daily Ozone Levels in the Northeastern U.S., Journal of the Air Pollution Control Association 28, 10: 1034-1038.

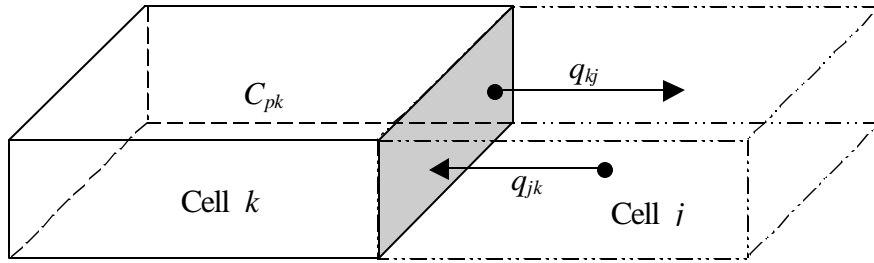


Figure 1: Well-mixed cell

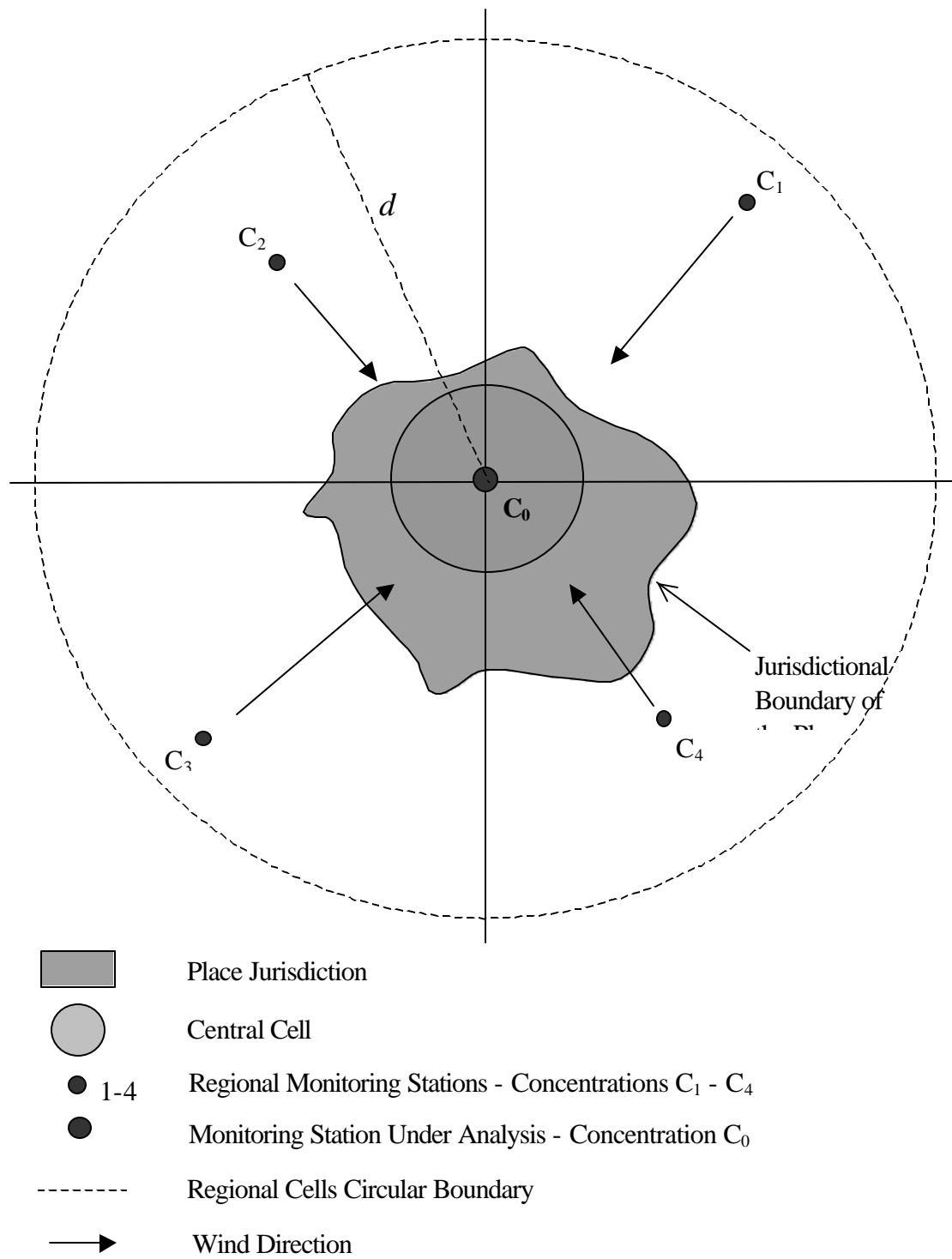


Figure 2: Circular application of a well-mixed cell system

Table 1: Regional effects regression of O₃ concentrations:
Variations with maximum distance

Max. Dist. (meter)	Obs. #	R²	Intercep t-value		b_c t-value	
			t	b₀		
450000	640	0.337	5.219	3.39	0.849	18.03
440000	640	0.337	5.219	3.39	0.849	18.03
430000	638	0.336	5.319	3.46	0.846	17.92
420000	638	0.336	5.319	3.46	0.846	17.92
410000	633	0.339	5.245	3.41	0.847	17.98
400000	630	0.340	5.211	3.39	0.847	17.97
390000	628	0.340	5.198	3.38	0.848	17.95
380000	620	0.351	4.508	2.91	0.871	18.29
370000	615	0.359	4.098	2.65	0.884	18.54
360000	611	0.358	4.125	2.65	0.885	18.45
350000	609	0.356	4.301	2.76	0.880	18.31
340000	605	0.357	4.315	2.76	0.879	18.28
330000	600	0.357	4.253	2.71	0.882	18.23
320000	593	0.357	4.193	2.65	0.883	18.12
310000	586	0.361	4.165	2.64	0.884	18.14
300000	581	0.359	4.258	2.68	0.881	17.99
290000	576	0.361	4.163	2.61	0.884	17.99
280000	569	0.359	4.239	2.64	0.882	17.82
270000	561	0.363	4.090	2.53	0.887	17.84
260000	554	0.364	3.979	2.45	0.890	17.79
250000	543	0.370	3.684	2.25	0.898	17.82
240000	538	0.370	3.632	2.20	0.900	17.73
230000	530	0.378	3.269	1.98	0.911	17.93
220000	525	0.376	3.346	2.01	0.908	17.76
210000	517	0.371	3.947	2.39	0.887	17.44
200000	512	0.371	3.951	2.38	0.887	17.34
190000	502	0.376	3.791	2.28	0.892	17.37
180000	484	0.384	3.505	2.08	0.902	17.34
170000	469	0.393	3.031	1.77	0.916	17.37
160000	453	0.391	3.220	1.86	0.910	17.01
150000	433	0.407	2.220	1.26	0.943	17.20
140000	408	0.390	3.010	1.65	0.916	16.12
130000	360	0.393	3.592	1.93	0.891	15.24
120000	324	0.396	2.299	1.13	0.931	14.53

110000	276	0.399	3.334	1.58	0.897	13.50
100000	237	0.433	1.282	0.56	0.969	13.39
90000	218	0.446	0.676	0.29	0.989	13.20
80000	197	0.450	0.372	0.15	1.002	12.64
70000	167	0.428	0.374	0.14	0.997	11.12
60000	137	0.497	-1.375	-0.49	1.047	11.55
50000	93	0.495	-2.171	-0.63	1.063	9.44
40000	49	0.521	-4.436	-0.91	1.158	7.16

Table 2: Regional effects regression of CO concentrations:
Variations with maximum distance

Max. Dist. (meter)	Obs. #	R ²	Intercept b_0	t-value	b_c	t-value
300000	197	0.15	406.6345	4.4180	0.6192	5.8970
290000	189	0.15	412.4422	4.4010	0.6121	5.7300
280000	182	0.15	414.3201	4.3710	0.6096	5.6530
270000	178	0.15	419.7955	4.4070	0.6134	5.6600
260000	174	0.16	407.6174	4.2300	0.6383	5.7760
250000	170	0.15	425.4050	4.2870	0.6242	5.4920
240000	164	0.16	424.6904	4.2460	0.6364	5.5480
230000	156	0.18	401.6714	3.9090	0.6738	5.7300
220000	153	0.17	405.5930	3.8980	0.6648	5.5960
210000	148	0.18	400.0176	3.7990	0.6821	5.6470
200000	136	0.19	398.9062	3.6060	0.7073	5.5320
190000	126	0.19	405.5787	3.5250	0.7183	5.4030
180000	119	0.16	464.7862	3.7560	0.6686	4.7380
170000	112	0.14	485.8158	3.6950	0.6421	4.3000
160000	104	0.18	441.0663	3.3360	0.6894	4.6820
150000	89	0.19	477.5792	3.8750	0.6140	4.4870
140000	83	0.20	458.4723	3.5520	0.6391	4.4950
130000	78	0.19	479.6872	3.4880	0.6247	4.1840
120000	71	0.16	562.5371	3.8600	0.5732	3.6820
110000	67	0.18	532.3607	3.5200	0.5991	3.7160
100000	63	0.18	550.9572	3.5700	0.5858	3.6070
90000	60	0.19	522.9644	3.3340	0.5987	3.6450
80000	48	0.16	594.2994	4.4200	0.4217	2.9050
70000	40	0.17	589.6022	3.9960	0.4298	2.8130
60000	35	0.11	699.2372	4.1570	0.3442	2.0520
50000	24	0.09	797.4019	4.3050	0.2631	1.4500

Table 3: Descriptive statistics for the variables used in the O₃ model

Variable	N	Mean	Min.	Max.	STD.
C_o (p.p.b.)*	303	29.8	13.8	69.6	6.97
REFF (p.p.b.)	303	30.6	15.5	51.2	5.01
POP (#)	303	295977	249	7322564	870928
CONSTR (%)	303	0.06	0.01	0.19	0.03

PUBT (%)	303	5.26	0	53.40	8.14
ELEC (%)	303	19.37	0.51	97.78	16.19
HGFUEL (%)	303	8.37	0	80.93	15.06
FRST (%)	303	12.97	0	84.06	17.67
S (watt-hour/m³)	303	496.62	321.6	823.6	141.92
T (deg. Celsius)	303	23.42	17.7	33.9	3.66

* p.p.b.: parts per billion

Table 4: Descriptive statistics for the variables used in the CO model

Variable	N	Mean	Min.	Max.	Std.
C_o (p.p.b.)*	117	994.0	71.2	4156.6	539.5
REFF (p.p.b.)	117	858.7	15.3	1813.5	340.3
TTD (mile-person)	117	5870822	25815	77455141	17245201
AGR (%)	117	19.83	0.00	91.21	22.67
AVDND5 (meters)	117	2915	1777	4594	544

* p.p.b.: parts per billion

Table 5 Elasticity functions for the O₃ model

Variable	Elasticity Function
----------	---------------------

<i>REFF</i>	$\frac{\partial C_o}{\partial REFF} * \frac{REFF}{C_o} = b_1 * \frac{REFF}{C_o}$
<i>S</i>	$\frac{\partial C_o}{\partial S} * \frac{S}{C_o} = a * S^a * T^b * (b_2 * POP + b_3 * CONSTR + b_4 * PUBT + b_5 * ELEC + b_6 * HGFUEL + b_7 * FRST) * \frac{1}{C_o}$
<i>T</i>	$\frac{\partial C_o}{\partial T} * \frac{T}{C_o} = b * S^a * T^b * (b_2 * POP + b_3 * CONSTR + b_4 * PUBT + b_5 * ELEC + b_6 * HGFUEL + b_7 * FRST) * \frac{1}{C_o}$
<i>POP</i>	$\frac{\partial C_o}{\partial POP} * \frac{POP}{C_o} = b_2 * S^a * T^b * \frac{POP}{C_o}$
<i>ONSTR</i> <i>R</i>	$\frac{\partial C_o}{\partial CONSTR} * \frac{CONSTR}{C_o} = b_3 * S^a * T^b * \frac{CONSTR}{C_o}$
<i>PUBT</i>	$\frac{\partial C_o}{\partial PUBT} * \frac{PUBT}{C_o} = b_4 * S^a * T^b * \frac{PUBT}{C_o}$
<i>ELEC</i>	$\frac{\partial C_o}{\partial ELEC} * \frac{ELEC}{C_o} = b_5 * S^a * T^b * \frac{ELEC}{C_o}$
<i>HGFUEL</i> <i>L</i>	$\frac{\partial C_o}{\partial HGFUEL} * \frac{HGFUEL}{C_o} = b_6 * S^a * T^b * \frac{HGFUEL}{C_o}$
<i>FRST</i>	$\frac{\partial C_o}{\partial FRST} * \frac{FRST}{C_o} = b_7 * S^a * T^b * \frac{FRST}{C_o}$

Table 6 Elasticities for the O₃ model

Variable	Elasticity		
	At sample mean	At 1 std. more* (1+STD)	At 1 std. less* (1-STD)

<i>REFF</i>	0.696	0.727	0.657
<i>S</i>	0.003	0.005	0.002
<i>T</i>	0.004	0.006	0.003
<i>POP</i>	0.010	0.038	**
<i>CONSTR</i>	0.069	0.100	0.036
<i>PUBT</i>	-0.051	-0.141	**
<i>ELEC</i>	-0.022	-0.042	-0.004
<i>HGFUEL</i>	0.016	0.044	**
<i>FRST</i>	-0.019	-0.047	**

*From sample mean value

**Standard deviation > mean value

Table 7 Elasticity functions and elasticities for the CO model

Variable	Elasticity			
	Elasticity function	At sample mean	At 1 std. more* (1+STD)	At 1 std. less* (1-STD)
<i>REFF</i>	$b_1 * \frac{REFF}{C_o}$	0.409	0.491	0.294
<i>TID</i>	$b_2 * \frac{TID}{C_o}$	0.077	0.246	**
<i>AGR</i>	$b_3 * \frac{AGR}{C_o}$	-0.066	-0.154	**
<i>AVDND5</i>	$b_4 * \frac{AVDND5}{C_o}$	-0.562	-0.744	**

*From sample mean value

**Standard deviation > mean value