Environmental Energy and Economic Research (2017) 1(1):23-42
DOI 10.22097/eeer.2017.46455

A Hybrid Deterministic-Statistical Model Integrating Economic, Meteorological and Environmental Variables to Air Pollution

Farhang Shahparaki[a,*], Majid Abbaspour[b], Majid Shafie-Pour[c], Mahmoud Mahmoudi[d]

[a] Islamic Azad University Sciences and Research Branch, Tehran, Iran
[b] Sharif University of Technology, Tehran, Iran
[c] Graduate Faculty of Environment, University of Tehran, Tehran, Iran
[d] Tehran University of Medical Science, Tehran, Iran

Received: 2 November 2016 /Accepted: 12 January 2017

Abstract
The following study is based on a hybrid statistical-deterministic model designed for the assessment of the daily concentration of sulfur dioxide (SO_2), carbon monoxide (CO) and particulate matter (PM_{10}) as major pollutants in the Greater Tehran Area (GTA): the capital of Iran. The model uses three available or assessable variables including economic, meteorological and environmental in the GTA for the year 2003. Economic sectors which are examined in this study are firstly traffic, secondly residential-commercial heating and thirdly industry. The model determines to what degree each of the aforementioned sectors, in accordance to their associated fuel consumption, is responsible for air pollution. The model also relates emission data from the three sectors whilst taking into consideration meteorological parameters. Thereafter, economic and meteorological parameters as independent explanatory variables opposed to the concentration of pollutants measured at the monitoring network stations which are dependent variables. All data is given in the form of time series for the year 2003 in specified areas discussed. The method adopted for the calculation of the regression coefficients of the model, is based on nonlinear least squares multiple regression analysis. The model has been tested on the available monitoring network stations for aforementioned pollutants in the GTA. Model verification has been carried out spatially in the year 2003 and temporally for the year 2005. Results show that the concentration of pollutants in the GTA can be estimated using this model. Areas of further research are outlined which indicate possible enhancement of this approach and relevant application extensions.

Keywords: Empirical hybrid model, statistical-deterministic distribution, air pollution modeling, traffic pollution

Introduction and air pollution issues in mega cities

The main goal of this study is to design a model based on a hybrid statistical- deterministic distribution for the assessment of daily concentrations of SO_2, CO and PM_{10} in the Tehran urban environment, utilizing a collection of suitable data secured through available

[*] Corresponding author E-mail: farhang_shahparaki@yahoo.com

monitoring network stations located in the GTA for the year 2003. This model can determine to what degree motor vehicle traffic, indoor heating and regional industries are responsible for the designated pollutants.

Many of the world's major cities are confronted with a series of environmental crises on a daily basis. Episodes of poor air quality are among the most perceptible and insidious of these environmental challenges. Pollution and the escalating deterioration of air quality in large cities around the globe are closely related to various economic activities. In large urban areas, there are three types of such activities namely motor vehicle traffic, indoor heating and local industries. A brief study shows that the World Health Organization (WHO) guidelines for SO_2, CO and PM_{10} are regularly exceeded in the majority of large cities.

Tehran, the capital of Iran, has a burgeoning population which has skyrocketed in the past 50 years from a 1952 census of 0.7 million to more than 10 million in 2007. The metropolis has an average altitude of 1280 meters above sea level and an area of approximately 730 square kilometers divided into 22 municipal zones.

The city is bordered in the north by the Alborz Mountains and characterized by a gentle slope from north to south. It is confined from west to east by the city of Karaj and the Damavand mountains, respectively. Tehran is the main center of economic, social and political activity and is a major hub of industry, education, housing and recreation for tens of millions of residents.

The most important cause of air deterioration in Tehran is transportation which generates emission of most common atmospheric pollutants including SO_2, CO, *PM*, *NOx* and *HC*. SO_2, CO, PM_{10} have been chosen for this study due to the availability of their measurements and because of the great concern of the Department of the Environment (DOE) about these pollutants.

The assessment of the impacts of increased road traffic on air quality helps not only to evaluate the actual situation but also to assess the possible effects of measures towards a more sustainable transport system. In addition to the dramatic acceleration in Tehran's road traffic over the last two decades, industrial activity has increased extensively in many districts, as well.

Tehran's geographical confinement on the northern, eastern and southern regions results in the accumulation of pollutants in the city. Other parameters which also play a critical role leading to a challenging weather situation include:

. A scarcity of rainfall with less than an average total of 240 mm per annum;
. An irregular distribution of rainfall throughout the year with approximately 87% occurring between mid November and mid April;
. The occurrence of temperature inversion around 60% of the year;
. An average wind velocity of less than 3 ms^{-1} for more than 70% of the year.

In conjunction with the specific climatic and topographical factors of the GTA and escalation of urban economic activities, concentration of air pollutants in the GTA remains at high levels especially after unusual events. Therefore considering weather dispersion due to specific climate and topographic factors, pollutants may tend to accumulate.

Modelling approach

Attributes of hybrid statistical-deterministic distribution

Experience has shown that studies of such environmental consequences must address numerous complicating factors in the form of meteorological effects, unusual traffic conditions and the impact of other control programs. These variables complicate the collection of suitable data that can be analyzed by a certain model in order to demonstrate a particular finding with the desired degree of statistical confidence. Different models can be used to predict air pollution, including statistical, deterministic and more recently, hybrid statistical- deterministic models.

One of the important limitations of statistical forecasting is the lack of diagnostic capabilities because the sources of pollution cannot be identified by sensitivity experiments. This presents a major drawback to evaluation and execution of pollution reduction measures in real time. Deterministic models normally require numerous accurate input data (emissions, meteorology, and land cover) that are difficult to collect in real time.

However they can be still considered usable on simple personal computers or on small workstations. Computing secondary pollutants contribution through atmospheric chemistry modeling may also be considered an effective tool in recent years. The possibility of using a deterministic model, constituting the standard approach to the modeling of atmosphere pollutant concentration is often limited (Jakeman *et al.,* 1988a). In addition, tuning of physical model parameters and validation are exhaustive in process due to the extensive computer time required. Several deterministic- based models exist to evaluate and predict the pollutant dispersion in urban areas, but the majority may be causal in nature and fail to predict the extreme concentrations (Gokhale *et al.,* 2004; Jakeman *et al.,* 1988b).

One solution would be to use statistical models presenting variants which correct certain effects in complex dispersion mechanisms which have been simplified in deterministic models. Thereby accessing a large selection of variables to function may not be required (Hecq *et al.,* 1994a). The hybrid approach allows the prediction of the distribution of air pollutant concentrations from input variables such as emission strengths and meteorological characteristics. From conceptual and practical points of view, stochastic models also have some advantages over statistical regression base models. Some suggestions are made on how elements of the two methodologies (regression and stochastic models) can be combined (Milionis *et al.* 1994).

A hybrid deterministic and statistical approach technique, also based on the use of the Kalman filter**,** has been proposed (Melli *et al.* 1981). Kalman filters have been used in air pollution problems to obtain more accurate predicted values in episode forecasting and control (Zannetti, 1990). As a result, a hybrid air quality application of the Kalman filter was developed by Zannetti and Switzer (Zannetti, *et al.,* 1979). To overcome the limitations of both statistical and deterministic approaches, an empirical model based on a hybrid statistical-deterministic distribution has been developed at Centre for Economic and Social Studies on the environment (CESSE) at the Université Libre de Bruxelles, as an econometric type model using non-linear multiple regression analysis *(*Hecq *et al.,* 1994a).

This type of model has been constructed and tested using various parameters within an urban atmosphere (Hecq *et al.,* 1994a; Favrel *et al.,* 1998).

However, assessing the relationship between the urban atmospheric pollution and economic activities in the GTA is a difficult task due to the following constraints:

- The city size and its geographic location cause relevant variation of climatic conditions throughout its extensions.
- Irregularly located and continuously growing distribution of local and background industries
- Traffic loaded areas residential-commercial heating hot spots
- Complex relations between weather phenomena and economic activities
- Lack of relevant data in time series form
- Lack of a sufficient number of monitoring stations covering 22 municipal zones of the GTA.

Basic approach

The hybrid model consists of two separate modules. The first links economic activities with emissions level and the second leads to immissions by modeling the relationships between the previously calculated emissions and various appropriate Meteorological variables. The model is based on the daily mean level of the pollutants in 56 selected grids by means of simple random sampling in the GTA for the year 2003.

Figure (1) shows the GTA city map together with the stationary and portable monitoring stations.

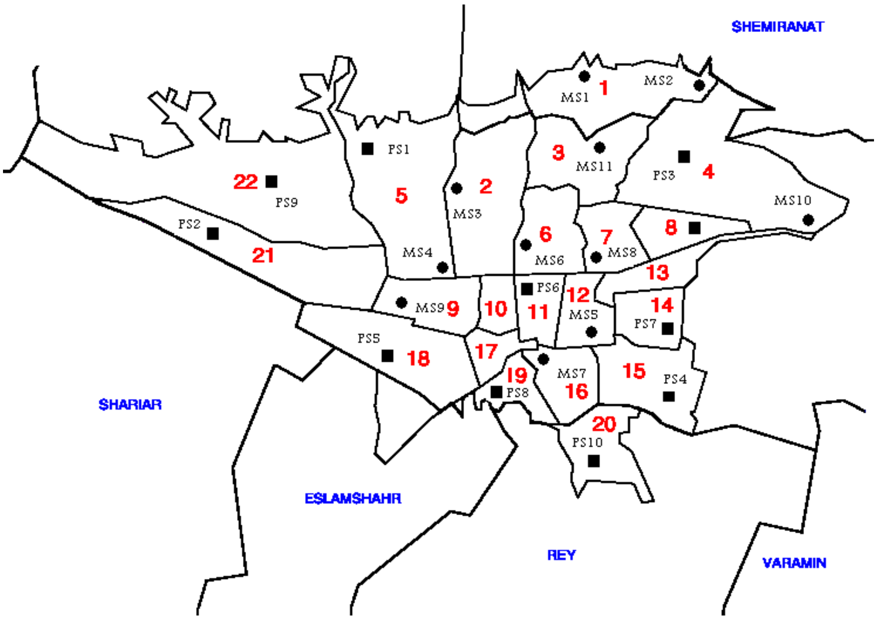

Figure 1. Great Tehran Area map and schematic monitoring site locations

Daily mean concentration of the pollutants have been measured on a regular 24 hour basis at the 11 stationary monitoring stations noted as (SM) as dependent variables.

SM1: Tajrish monitoring station
SM2: Aghdasie monitoring station
SM3: Pardisan monitoring station
SM4: Azadi monitoring station
SM5: Bazar monitoring station
SM6: Fatemi monitoring station
SM7: Bahman monitoring station
SM8: Vila monitoring station
SM9: Mehrabad monitoring station
SM10: Sorkhe hesar monitoring station
SM11: Gholhak monitoring station

In the meantime the same data has been collected for other locations pointed out in the schematic map by 10 portable monitoring stations (PM) for the years 2003 and 2005.
Meteorological variables are examined separately for SO_2 and for other pollutants such as PM_{10}, NO and NO_2 (Hecq *et al.*, 1994a), (Jakeman *et al.*,1988a; Koo et al., 1990; Favrel *et al.*, 1998).

Module: Emissions–Immissions

Pollutants including SO_2, CO and PM_{10} are modeled independently based on the daily average of observations from the monitoring stations for the year 2003.
Relationships between the variables are based on a hybrid statistical- deterministic distribution based on two types of temporal sets of data:
. A linear relationship among SO_2, CO and PM_{10} emissions and the concentration of pollutants in the atmosphere
. A non-linear relationship between the concentration of SO_2, CO and PM_{10} in the atmosphere and determinant meteorological variables
The novelty of the adopted method for the hybrid model takes into account the relationship of economic as well as meteorological factors with pollutant concentration in the ambient air.

General form of the model

The model is designed in reference to three types of variables:
. Independent explanatory variables representing daily SO_2, CO and PM_{10} emissions originating from economic activities;
. Independent explanatory variables indicating the influence of weather conditions on transport and dispersal of SO_2, CO and PM_{10};
. Dependent variables relating to the concentration of SO_2, CO and PM_{10} in the atmosphere measured at the monitoring sites.
General form of the model is as follows:

$$[SO_2] = [\alpha_1.Sp_1 + \beta_1.Mp_1 + \gamma_1].(\frac{1}{vel})^{\delta_1}.\exp(\eta_{1.}preci).\exp(\lambda_1.mixh) \quad (1)$$

$$[CO] = [\alpha_2.Sp_2 + \beta_2.Mp_2 + \gamma_2].(\frac{1}{vel})^{\delta_2}.\exp(\eta_{2.}preci).\exp(\lambda_2.mixh) \quad (2)$$

$$[PM_{10}] = [\alpha_3.Sp_3 + \beta_3.Mp_3 + \gamma_3].(\frac{1}{vel})^{\delta_3}.\exp(\eta_{3.}preci).\exp(\lambda_3.mixh) \quad (3)$$

$[SO_2]$, $[CO]$ and $[PM_{10}]$ are daily average pollutant concentrations in ppm for SO_2 and CO and μgm^{-3} for PM_{10} measured at different monitoring stations within the period considered.

Spi stands for residential-commercial heating and *Mpi* stands for motor vehicle traffic variables showing the daily emission of pollutant *i* due to stationary and mobile sources, respectively.

vel., preci. and mixh. represent the daily average of wind velocity in meter per second (m/s), daily average precipitation in millimeter (mm) and mixing height layer at midday in kilometers (Km) respectively.

αi, βi , γi, δi, ηi, λi are regression coefficients determined by means of multiple non-linear regression analysis.

Their values are estimated by means of non-linear generalized least squares with the hypothesis of non-zero contemporary covariance between the residues of the equations (Hecq *et al.*, 1994a).

The model is designed in simultaneous equations in terms of one equation per multiple stations monitoring pollutants concentration in ambient air.

Contributions from varying economic activities can be partially demonstrated based on the overall immissions according to equations (4) to (6) for concentration of pollutant *i* (Favrel *et al.*, 1998):

Pollutant *i* concentration originating from house and office heating

$$[x_i] = [\alpha_i.Sp_i].(\frac{1}{vel})^{\delta_i}.\exp(\eta_{i.}\,preci).\exp(\lambda_i.mixh) \tag{4}$$

Pollutant *i* concentration originating from motor vehicle traffic

$$[x_i] = [\beta_i.Mp_i].(\frac{1}{vel})^{\delta_i}.\exp(\eta_{i.}\,preci).\exp(\lambda_i.mixh) \tag{5}$$

Pollutant *i* concentration originating from industries as background

$$[x_i] = [\gamma_i].(\frac{1}{vel})^{\delta_i}.\exp(\eta_{i.}\,preci).\exp(\lambda_i.mixh) \tag{6}$$

The background level is a surplus to the pollution inputs from economic activities and can also be taken as any other far-distant sources particularly as the result of the long-distance transport of pollutants carried by winds (Hecq *et al.*, 1994a).

This phenomenon is very significant for PM_{10} with natural sources, especially for marginal districts of the GTA as a function of wind direction.

Model application

Module: Economic activity-Emissions

Domestic and indoor heating, traffic and industry emissions have been evaluated by the level of fuel burnt and relative emission factors.

Emissions caused by populated local industries in the GTA, while they may not be considered negligible, have been taken as background emissions due to inaccessible and unreliable daily statistics and considerable uncertainties regarding fuel level consumption. The impact is integrated as an external import together with other industries located in the GTA suburb in the model formulation.

Sample size

Most statistical methods are based on the assumption of random sampling. This simply means that every unit in the population has an equal chance of being chosen for the sample. Furthermore, the selection of random units should be independent of other units that have been sampled. A relatively simple and reliable method for randomization is to use random number generation. In order to calculate the optimum sample size in the GTA with a total area of $730km^2$, 730 equal cited grids of 10^6 square meter have been considered as statistical population. Considering a normal distribution of PM_{10} values in the GTA based on monitoring stations data and its bigger variation range (R) compared to other pollutants, particulate matter has been chosen for the determination of sample size.

Using variance formula, required sample volume based on equation (7) (Bartlett *et al.*, 2001) a total of 56 sampling units out of 730 grids will be sufficient to cover total GTA cited grids based on a random sampling procedure.

$$n = \frac{Z_{1-\alpha/2}^{2}\delta_{x}^{\ 2}}{d^{2}} \tag{7}$$

Where

n Sample size

$Z_{1-\alpha/2}^{2}$ Z-score (extracted from standard normal Z table)

$\sigma_{x}^{\ 2}$ Standard deviation

d^{2} Acceptable error

Residential and commercial heating

Annual emissions from stationary sources can be easily calculated by multiplying the consumption of each fuel by its emission factor leading to the formula (8) (Favrel *et al.*, 1998):

$$TPSj = \sum EFj,k.YCk \tag{8}$$

Where,

j Pollutant index (SO_2, CO, PM_{10});

k Fuel index;

$TPSj$ Annual emission of pollutant j due to stationary sources;

EFj,k Emission factors of pollutant j for stationary sources by fuel k;

YCk Yearly consumption of fuel k used by stationary sources;

The resulting annual emissions have been converted by some authors into daily emissions by means of degree-day (DD) in order to be integrated in the model using the (9) and (10) respectively (Favrel *et al.*, 1998):

$$DD = 16.5°C - Te \qquad \text{with DD} \geq 0 \tag{9}$$

$$Te = 0.6Tm + 0.3Tm_{-1} + 0.1Tm_{-2} \tag{10}$$

Where Tm is the average temperature of day m in °C and Te is the daily equivalent temperature that actually integrates the temperatures of the two previous days. According to National Iranian Gas Company (NIGC) statistics for the year 2003, it is assumed that stationary sources begin heating below the threshold of 15.0°C in mid-November in Tehran.

Since the density of industrial activities in the GTA cannot be considered negligible, emission from local and background industries has been evaluated as a residue from the statistical data treatment. Also, it is necessary to take into account the low height of the smoke stacks in the majority of existing industries located in the GTA.

Since the model is based on statistical time series analysis, the time cycles of industrial emissions are of high importance. However due to a lack of verifiable data for industrial activities, local and non-local industrial sources, as well as background sources, can all be rated together as annual constants. Monthly mean consumption of natural gas (NG) in each district of the GTA for the year 2003 can be calculated with equation (11) according to NIGC statistics:

$$GCi,j = \frac{GCj.FTi}{FT}, [i:(1-22), j:(1-12)] \tag{11}$$

Where

GCi,j Natural Gas consumption (m^3), district i, month j

GCj Total Natural Gas consumption (m^3), month j

FT Number of total households in the GTA

FTi Average number of households in district i

Monthly mean emission for pollutants in selected grids can be obtained by the equation (11) in conjunction with suitable emission factor by equation (12):

$$GFi,j,l = GCj,l.N.EFi,[i:(1-22), j:(1-12), l:(SO_2, CO, PM_{10})] \tag{12}$$

Where

GFi,j,l Monthly average emission in tons by pollutant l, district i, month j

N Number of residential and commercial units in each grid

EFl Emission factor for pollutant l

By dividing the mean emission resulting from equation (12) by the number of days in each month, the daily average can be simply calculated. Emission factors applied for the residential-commercial sector has been adopted from the European Environmental Agency source year 1996 (EEA, 1996). The number of residential and commercial units for 2003 has been provided by the Tehran Geographical Research Center.

Motor Vehicle Traffic

In addition to the stationary data, AQCC has established a mobile emission inventory survey in Tehran for the DOE and the Japan International Co-operation Agency (JICA) study team. According to this plan, the emission rate of the mobile sources from vehicles and trains running in the GTA in the year 2003 is calculated and estimated for the year 2009. However relevant data for the year 2003 has been calculated using correction factor.

Data base of AQCC study is prepared based on EMME2 software provided by Tehran Comprehensive Transportation and Traffic Studies Company.

Calculations were computed on a 24-hour daily basis for the year 2003 based on AQCC data sets considering the appropriate correction factor.

Almost the same procedure as econometric models may be linked with emission models such as COPERT III which allows for the calculation of pollutant emission and fuel consumption per km driven for the assessment of the annual pollutant emission and fuel consumption. Calculation of the total annual fuel consumption is a calibration parameter for estimating uncertain parameters such as the average annual mileage driven on each road class and for each vehicle category.

Emission emitted from the vehicles calculated by formula (13) has been provided by AQCC as a mobile emission inventory survey in the GTA for the DOE and the JICA study team based on formula (13) as follows:

$$EA = Ef.Kt.Tv \tag{13}$$

Where

EA Pollution emission in kg

Ef Emission factor of vehicles in Kg/Km

Kt Distance run by vehicle in Km/vehicle

Tv Traffic volume in a certain trip

The final result of *EA* is in conjunction with the following data and facts incorporated in the GTA:

- Vehicle type percentages
- Vehicle age distribution
- Fuel consumption data
- Emission factors (based on the Tehran Transport Emission Reduction Project (TERP) and the Global Environmental Facility (GEF). TERP is an international project in which AQCC has participated as a technical assistant as a joint venture with three Swedish firms (SWECO, SMHI and MTC). GEF in collaboration with the World Bank and the Tehran municipality provides the funding for the technical assistance.
- Tehran's main routes network, covers the precise length of main routes. Statistics on slope (road state including horizontal, up-hill and down-hill) based on related municipal zones using Arc view software.
- Morning rush hour factor
- Incorporation of four seasons and the time of day
- Alternate (corrected) factors for week-daily and year-monthly basis
- Study of an integrated master plan for air pollution control in the GTA

Calculation of the annual emissions from road traffic is based on the volume of traffic in vehicle-kilometers driven by different vehicle categories within 22 municipal zones of the GTA main routes with suitable emission factors for different vehicle categories circulating in the same area within the period considered. Emissions generated by road traffic in the GTA was provided by using the yearly kilometers and the representative average speed (ECE driving, average 40, 60 and 80 km/hr) of each vehicle category on each road class (horizontal, up-hill and down-hill) and adding the previously mentioned factors.

By using two suitable traffic indices, a daily index, defined as a function of the day in the week and a monthly index, defined as a function of the month in the year, daily emissions required by the model were obtained from AQCC data sets. Each main road has been broken up into multiple links and nodes and extended to a certain length. Therefore, a total of 11,155 main links and nodes of the GTA routes network with a certain length has been constructed by this procedure.

Subsequent calculations of the cumulative percentages of the main routes network in the GTA, will be written based on formula (14) as follows:

$$Si = \frac{Ri}{\sum_{i=1}^{22} Ri} \tag{14}$$

Where

Si Cumulative percentage of main routes network, district i

Ri Sum of main routes network length (m), district i

In result, total yearly emission in tons by each pollutant per km of main routes network in all districts can be calculated as by equation (15) as follows:

$$Qil = \frac{Kil.Si}{Ri}, [i:(1-22), l:(SO_2, CO, PM_{10})] \quad (15)$$

Where

Qil Total yearly emission in tons by pollutant l per km, district i

Kil Total pollution in tons for year 2003, district i, pollutant l

By using two suitable traffic indices, a daily index, defined as a function of the day in the week and a monthly index, defined as a function of the month in the year, daily emissions required by the model are obtained from AQCC data sets.

Table (1) shows the relevant monthly basis indices as an alternate factor for the months of the year. Table (2) shows the relevant daily basis indices.

Table 1. Alternate factor for the months of the year for the GTA

Month	coefficient	Month	coefficient	Month	coefficient	Month	coefficient
Jan.	1.024	Apr.	0.927	Jul.	0.902	Oct.	1.00
Feb.	1.028	May	0.915	Aug.	0.921	Nov.	1.002
Mar.	0.768	Jun.	0.923	Sep.	1.012	Dec.	0.954

Table 2. Alternate factor all the day's in a week for the GTA

Day's	Coefficient
Working day's	1.0
Thursday	0.9
Friday	0.5

It should be noted that in Iran, working days are from Saturday to Thursday with Friday considered a holiday. However, Thursday also may be regarded as a partial working day for the majority of governmental and private organizations.

Consequently, monthly average emission in tons by pollutant j, per km of main routes network in district i in the GTA could be written using equation (16) as follows:

$$Pijl = \frac{Qij.Cof_l}{\sum_{j=1}^{12} Cof_l}, [i:(1-22), j:(1-12), l:(SO_2, CO, PM_{10})] \quad (16)$$

Where

$Pijl$ Monthly average emission in tons by pollutant l, per km of main routes network in district i, month j

Cof_l Year-month basis indices coefficient

Using daily traffic indices from AQCC data sets by Table (2) as a function of the day in week mean daily emissions for each pollutant in tons have been calculated. Traffic data resulting from this method has finally been used in appropriate form in every grid in conjunction with indoor and office heating.

Meteorological data

Meteorological parameters also play an important role in the pollutant concentration measured by the monitoring stations. Some authors demonstrated that there are statistical correlations between concentrations of pollutants in the urban atmosphere, pollutant emission and some meteorological parameters like wind speed or temperature (Benarie, 1980).

For example, studies of statistical correlation show that wind speed and direction have a decisive effect on the dispersion of Sulfur dioxide (Hecq *et al.,* 1994a).

Temperature has a direct effect on pollutants concentration as increases in temperature lead to indoor heating reduction. Temperature also influences the mixing layer depth which in turn affects the pollutant concentration levels.

Certain authors also reported SO_2 wash-out effect is caused by rain (Koo *et al.,* 1980).

The most pertinent meteorological variables that will be used in this model are as follows:

Vel Daily average wind velocity (m/s)

preci Daily average precipitation (mm)

mixh Mixing height at midday (km)

Relevant meteorological data to each grid has been carefully examined after focusing on spatial considerations of adjacency of suitable meteorological site locations within the selected grids. Due to lack of any information regarding the mixing height, the "Holzworth method" has been adopted to determine mixing height based on available upper layer data provided by IRIMO for the year considered. Holzworth and others have developed objective methods to simplify and homogenize the analysis of the often complex stratification of the atmospheric boundary layer (ABL) and to estimate the mixing height under convective conditions. The basic idea of the Holzworth method is to follow the dry adiabatic conditions starting at the surface with the measured or expected (maximum) temperature up to its intersection with the temperature profile from the most recent radio sounding (Holzworth, 1964). Table (3) shows the monthly mean mixing height (MH) layer based on daily MH layer calculations by the Holzworth method at midday in km at the Mehrabad International Airport site, located in the southwest area of the GTA in 2003.

Table 3. MH layer in the GTA

Months	Average MH layer (Km)
1	4.0
2	3.4
3	3.6
4	2.4
5	2.0
6	1.0
7	1.0
8	1.0
9	1.7
10	2.0
11	2.2
12	3.3

The variation of mixing height due to diurnal variations of solar radiation, synoptic conditions and local terrain strongly affects pollutant concentrations.

In accordance with the graph-based findings of relevant pollutant concentration in the atmosphere and meteorological variables of wind speed, rainfall and mixing height layer, a non-linear relation has been adopted in this study (Hecq *et al.,* 1994a; Favrel *et al.,* 1998; Koo *et al.,* 1990).

All meteorological data, except mixing height, has been provided in the form of daily means by the Iran Meteorological Organization based on six major climatology, synoptic and rain range stations located in different parts of the GTA for the period considered.

The following sections summarize the notable results associated with emission assessment, emissions-immissions relationship and the degree to which each major economic activity is responsible for the overall pollution concentrations for the current situation in the GTA for the year 2003.

Results and discussion

The method adopted for the establishment of the coefficients is based on the least square hypothesis and application of the nonlinear multiple regression analysis method. The model verification has been carried out spatially for the year 2003 for 11 selected grids randomly chosen and temporally for the year 2005 in the GTA separately. As stated in section 2.4, the model is designed to enable estimates of daily concentrations of SO_2, CO and PM_{10} on the basis of temporal sets of economic and meteorological variables. The model so designed is a system of simultaneous equation which means one equation for all 56 monitoring stations in respect to each pollutant studied. Regression coefficients to be estimated αi, βi , γi, $\delta i, \eta i$, λi as well as terms of error e_1, e_2 and e_3.

Pollutant emissions

Due to total Vehicles Kilometer Traveled (VKT) in 2003 for the GTA 56 cited grids have been divided into three subgroups based on the VKT factor as follows:

. High Traffic (HT) subgroup: $VKT \geq 2000 \times 10^6 VKT$

. Medium Traffic (MT) subgroup: $1000 \times 10^6 VKT \leq VKT \leq 2000 \times 10^6 VKT$

. Low Traffic (LT) subgroup: $VKT \leq 1000 \times 10^6 VKT$

The above mentioned subdivisions make it possible to clearly distinguish between pollution caused by the various traffic subgroups. Figure (2) shows VKT ranges per district within the GTA in the year 2003.

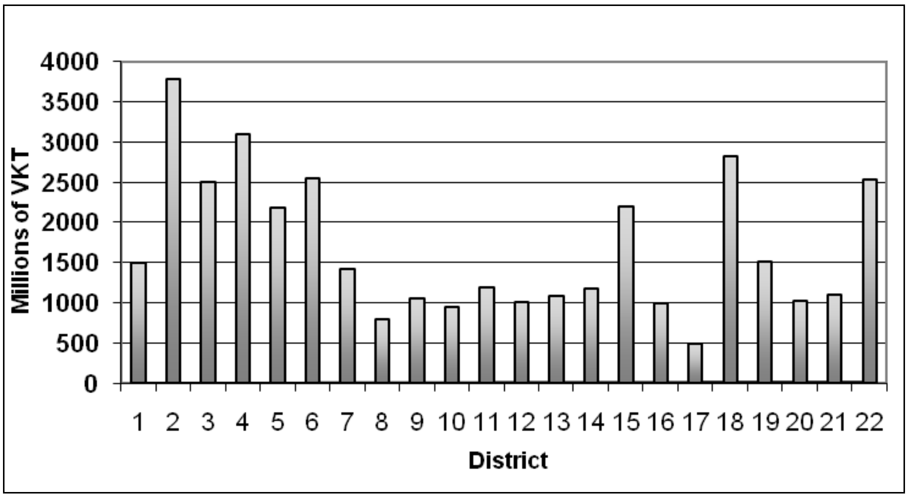

Figure 2. VKT per districts for the GTA

As a result, the HT sub-group, including districts 2,4,18,6,22,3,15 and 5 with a total of 56% emission share and the MT sub-group, including districts 19,1,7,12,11,14,21,13 and 9 with a total 23% share and the LT sub-group, including districts 16, 20, 10, 8, and 17 with a total of 21% share, has been specified, respectively. Figures (3) to (5) present a comparison between the concentrations measured and those that have been estimated by the model for SO_2, CO and PM_{10} relating to the 56 grids during the year 2003.

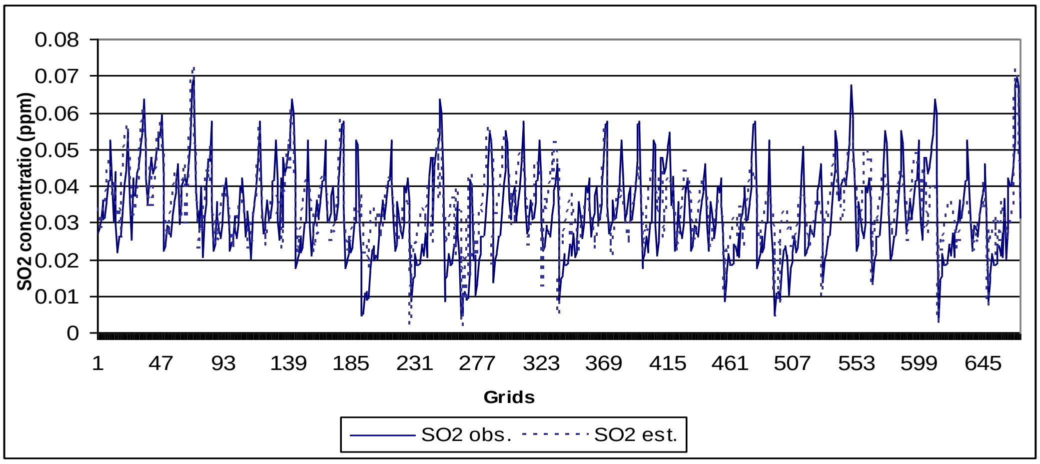

Figure 3. Comparison between SO_2 observed and estimated monthly average in the GTA

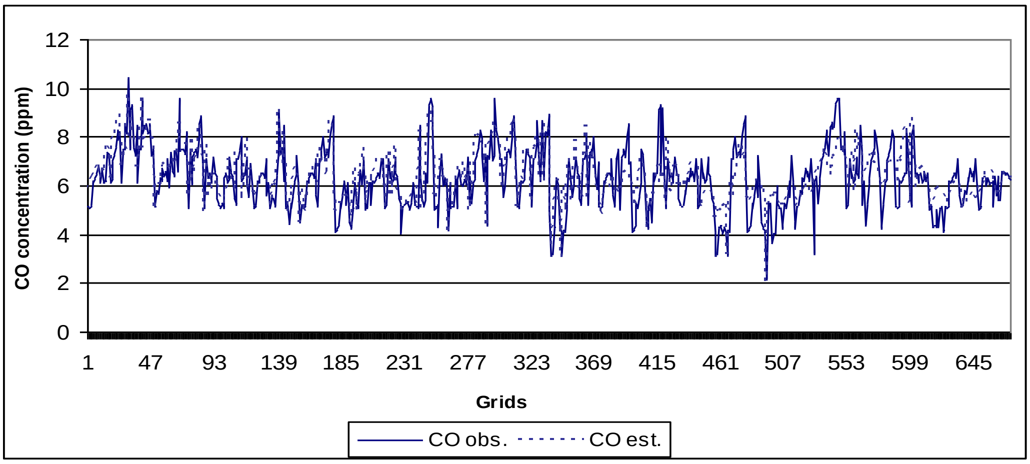

Figure 4. Comparison between CO observed and estimated monthly average in the GTA

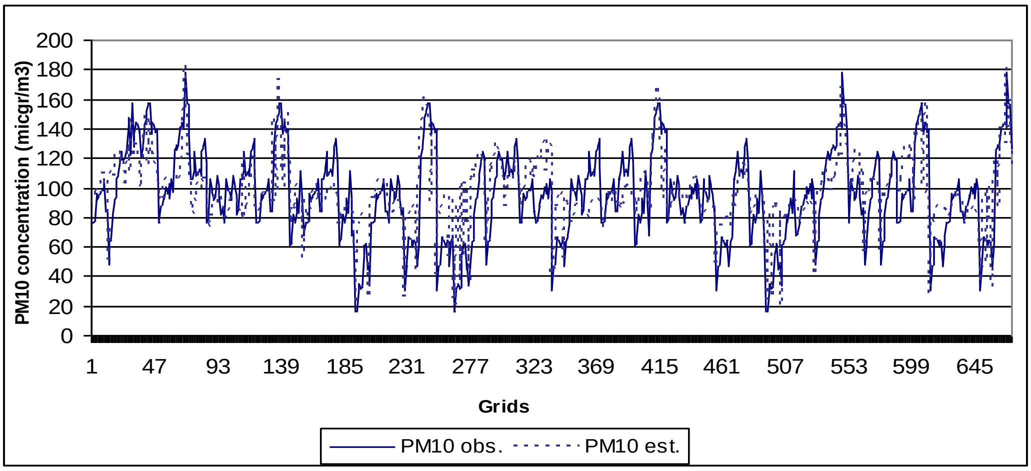

Figure 5. Comparison between PM_{10} observed and estimated monthly average in the GTA

Statistical results

Below are statistical viewpoints of the model and estimated coefficients which make it possible to ascertain the model's degree of validation:

. Acceptable R-square measures for linear regression between observed and estimated concentration of $[SO_2]$, $[CO]$ and $[PM_{10}]$, setting the Intercept at zero are 0.74, 0.70 and 0.62 respectively.
. Coefficients α and β as relevant to residential-commercial heating and motor vehicle traffic respectively are in accordance with the localization of LT, MT and HT subgroups based on VKT for the GTA.
. Coefficients δ, η and λ associated with meteorological parameters, including wind speed, precipitation and mixing height layer, show the expected sign and the *t*-test between the means, proving the important role they play in regard to pollutant levels in the ambient air.

As can be noted, the estimates arrived at by the model are consistent with the concentrations for all three pollutants measured at monitoring stations during the year 2003.

This consistency clearly demonstrates model reproducibility for average daily concentrations of the pollutants based on various economic activities in the form of time series within the year studied by taking into account, simultaneously, the most dominant meteorological parameters.

Pollutants

Sulfur Dioxide (SO_2)

Sulfur dioxide concentration in the ambient air results mainly from background levels in all three different subgroups, according to the contributions of different activities to the overall immission, using equations 16 to 18. Tables (4) and (5) represent how much the three subgroups: road traffic, residential-commercial heating and background are responsible for SO_2 in the years 2003 and 2005, respectively.

Table 4. SO_2 percent share for sub-groups in the GTA,2006

Source	SO_2 % (HT)	SO_2 % (MT)	SO_2 % (LT)
Traffic	37.0	26.5	18.5
House	6.5	7.0	8.0
Industry	56.5	66.5	73.5

Table 5. SO_2 percent share for sub-groups in the GTA,2008

Source	SO_2 % (HT)	SO_2 % (MT)	SO_2 % (LT)
Traffic	38.0	28.0	21.5
House	7.0	8.0	9.0
Industry	55.0	64.0	69.5

In table (4), the major share for SO_2 is a direct result of the background for all three sub-groups. SO_2 is closely linked to the use of fossil fuel, particularly petroleum products, therefore high levels generally occur in areas with heavy industrial activity sites which are normally categorized as LT subgroups.

Electricity generated from power plants is the main source of SO_2 pollution along with other contributing sources. Focusing on a report prepared for a Tehran power plant located in the southern sector, emission of sulfur dioxide was estimated to be around 100 metric tons per day based on World Bank Report year 2002.

Assessment of the model clearly suggests an abatement of SO_2 share emitted from residential-commercial heating from 20% to an average of 7.5% in three traffic sub-groups for the year 2003 which is a positive consequence of using natural gas as the dominant fuel source accessible to more than 90% of the population of Tehran in 2003 according to Ministry of Energy report, 2003.

The same trend exists of a negative correlation between residential-industrial and traffic responsibilities share as noted in tables (2) and (3). Of interest is that most of the agricultural activities take place in the same southern part of the GTA. The background describes immssions other than those from traffic and residential-commercial heating due to difficulties in accessing precise figures of different fuel consumptions in most of the factories located in the GTA. The resulting immission, therefore, corresponds to the natural background level plus the input from industry on both inside and outside the urban area. The model is not able to make any distinction between these two components and therefore immisions attributed to background have been noted in equation (6) with relevant immissions rated together for local and non-local industries as annual constants. In general, no single modeling approach may generate optimum results in terms of the full range of performance (Nunnari *et al.*, 2004).

Table (6) shows the various economic shares by months for Sulfur dioxide for the years 2003 and 2005.

As the figures in table (6) indicate, SO_2 level increased during the winter time and this is more apparent for the residential-commercial heating and background sectors. Traffic level is not as influenced by seasons as in two other sectors.

However, a slight increase during September and October is partially related to the commencement of school activities as well as driving habits and traffic jams within the period considered. It is important to also consider the diminishing of the mixing height for all sectors that may increase the volume of pollutants.

Table 6. SO_2 monthly distribution share by economic activities in the GTA

SO_2% - seasonal share, 2006				SO_2% - seasonal share, 2008			
Months	Background	House	Traffic	Months	Background	House	Traffic
1	6.5	9.1	8.1	1	6.4	7.8	8.0
2	6.6	7.1	6.8	2	6.7	4.6	6.9
3	5.7	4.3	5.9	3	6.1	4.0	6.2
4	6.8	4.4	7.1	4	6.9	4.6	7.1
5	6.5	5.0	6.9	5	6.6	4.5	7.0
6	9.6	5.7	9.5	6	7.8	5.1	8.1
7	9.2	3.9	8.7	7	9.1	6.3	8.6
8	10.2	7.8	9.7	8	10.4	10.1	9.9
9	10.6	12.4	10.1	9	10.9	13.8	10.4
10	9.8	17.1	9.8	10	10.1	15.3	10.1
11	9.7	15.3	9.1	11	9.8	13.6	9.1
12	8.9	8.0	8.3	12	9.1	10.1	8.4

Carbon Monoxide (CO)

Using the model outputs in conjunction with equations (5) to (6), share of responsibility for CO emitted by road traffic, residential-commercial heating and background are assessed and shown in tables (7) and (8) for the years 2003 and 2005, respectively.

Table 7. CO percent share for sub-groups in the GTA, 2006

Source	CO % (HT)	CO % (MT)	CO % (LT)
Traffic	87.0	76.0	68.0
House	1.5	2.5	3.5
Industry	11.5	21.5	28.5

Table 8. CO percent share for sub-groups in the GTA, 2008

Source	CO % (HT)	CO % (MT)	CO % (LT)
Traffic	88.5	77.0	71.5
House	2.0	3.0	3.5
Industry	9.5	20.0	25.0

As can be noted from table (7), a major share for CO pollutant is caused by traffic in almost all three subgroups as high levels of CO generally occur in areas with heavy traffic congestion. However, a decrease equal to approximately 14.5% and 28.0% may be seen for MT and LT subgroups as compared to the HT subgroup, respectively. This decrease is directly related to the grids located in designated subgroups in 22 districts of the GTA.

Residential-commercial heating has a nearly constant level, but there is a relative increase in background level from the HT to LT subgroups that can be explained by lower concentration of industrial activities in the HT subgroup.

In table (10), the same situation exists for the year 2005. According to official data from Iran’s Ministry of Industries the number of vehicles produced within the GTA during the years 2003 & 2005 were constant.

Therefore, changes in air pollution in the GTA according to the relevant emission factors for each vehicle subcategory during these years are negligible.

Table (9) shows monthly distribution share of CO as a result of various economic activities for the years 2003 and 2005.

Table 9. CO monthly distribution share by economic activities in the GTA

CO % - seasonal share, 2006				CO % - seasonal share, 2008			
Months	Background	House	Traffic	Months	Background	House	Traffic
1	9.4	14.3	8.7	1	9.3	13.3	8.6
2	8.9	7.8	8.3	2	9.0	10.2	8.3
3	6.6	11.0	8.2	3	6.5	8.1	8.1
4	7.5	7.8	7.8	4	7.6	5.4	7.8
5	7.0	5.1	7.1	5	7.2	4.9	7.4
6	7.8	4.8	8.1	6	7.8	5.3	8.0
7	7.6	3.9	8.0	7	7.6	5.3	8.0
8	8.1	4.9	8.4	8	8.1	5.4	8.3
9	9.2	7.4	8.6	9	9.0	6.4	8.5
10	9.3	6.9	8.8	10	9.3	9.3	8.8
11	9.6	10.8	9.1	11	9.6	12.4	9.2
12	9.0	15.3	8.9	12	9.1	14.0	9.0

As can be seen in table (9), traffic sources of CO show a relatively constant trend within the months, although there is a slight increase from September to January due to days with more inversion. In contrast, there is a slight decrease in summer due to less inversion occurrences as well as less traffic because of the academic summer holiday. It has been shown that daylight savings time, applied in summer months, causes a rise in evening traffic and creates an increase in fuel consumption; hence, a higher pollution level in the atmosphere.

Obviously, for residential-commercial heating, the more fuel consumed in winter time, the higher the rate of CO. Conversely, a lower rate is seen from April to August due to the higher frequencies and velocities of wind.

Measurements show the average wind velocity can increase by approximately 55% from March to July as compared to August to January in the GTA based on IRIMO data for 2003. Background share as evidenced by traffic does not fluctuate considerably within the months and is generally influenced by wind velocity and mixing height layer.

Analysis of the meteorological parameters affecting concentrations of air pollutants leads to the conclusion that mixing height facilitates the regional transport of ambient air pollutants. Mixing height or mixing depth signifies the height above the surface throughout which a pollutant such as smoke can be dispersed.

During times of surface temperature inversions, the mixing height decrease and pollutants dispersion is minimal. Transport wind signifies the average wind speed throughout the depth of the mixed layer. Ventilation rate (VR) written as formula (17) is the product of mixing height and transport wind, therefore representing the ability of the boundary layer to diminish the pollutants concentrations.

*Ventilation Rate (m*m /s) = Mixing Height (m) x Transport Wind (m/s)*

When VR values are low, there is not much mixing potential and surface air quality suffers. When VR values are consistently low (day and night), it is possible to “smoke in” large areas for several days. Average mixing height layer clearly suggests a 65% decrease during autumn and winter as compared to spring and summer, clearly caused by more frequent inversion occurrence during the colder months.

Particulate matter (PM_{10})

Tables (10) and (11) illustrate the share of responsibility for PM_{10} as emitted by road traffic, residential-commercial heating and background factors assessed in equations 4 to 6 for different subgroups separately. Same applies to PM_{10} according to tables (10) and (11).

Table 10. PM_{10} percent share for sub-groups in the GTA,2006

Source	PM_{10} % (HT)	PM_{10} % (MT)	PM_{10} % (LT)
Traffic	83.5	74.5	64.0
House	2.0	3.0	3.5
Industry	14.5	22.5	32.5

Table 11. PM_{10} percent share for sub-groups in the GTA, 2008

Source	PM_{10} % (HT)	PM_{10} % (MT)	PM_{10} % (LT)
Traffic	84.0	76.0	64.5
House	2.5	3.5	4.0
Industry	13.5	20.5	31.5

Traffic plays a major role in the GTA urban pollution with an average share of 75% among the three subgroups. Background level covers 23% on average and residential-commercial heating as in the case of CO has a minor impact.

Background level increased by 12% and 18% for the MT and LT subgroups, respectively, in parallel with a decrease in traffic level which can be explained by dominant wind direction from west to east as well as less residential housing accounting for decreased traffic in the MT and LT subgroups. Also, it should be noted that 70% of the total winds in the GTA occur below 3 ms^{-1}. Analysis of particulate matter shows that there is a negative relationship between wind velocities less than 4-5 ms^{-1} and, on the other hand, a positive relationship for velocities higher than 5 ms^{-1}. This contrasts with other pollutants which are normally dispersed by winds at any level. Such a phenomenon can considerably affect border districts of the GTA as in districts 20, 18, 4 and 15 which receive particulate matter on an erosion basis from border areas surrounding the GTA mainly covered by dry lands. Model output also confirms a more than 100% increase in particulate matter pollution for border districts. Although the HT subgroups districts 3,5,6,2 and district 12, the oldest part of the city, embracing the Bazaar (old market), show a higher rate of PM_{10} pollution.

Table (12) shows the various economic shares for particulate matter by months for the years 2003 and 2005.

Table 12. PM_{10} monthly distribution share by economic activities in the GTA

PM_{10} % - seasonal share, 2006				PM_{10} % - seasonal share, 2008			
Months	Background	House	Traffic	Months	Background	House	Traffic
1	6.6	11.3	8.2	1	6.5	8.0	8.1
2	7.5	8.0	7.8	2	7.5	5.3	7.7
3	7.0	5.2	7.1	3	7.2	4.8	7.3
4	7.8	4.9	8.0	4	7.7	5.3	7.9
5	7.5	5.7	7.9	5	7.5	5.3	8.0
6	8.1	5.1	8.4	6	8.0	5.4	8.3
7	9.1	3.8	8.6	7	9.0	6.4	8.5
8	9.3	7.0	8.8	8	9.3	9.3	8.9
9	9.6	11.0	9.1	9	9.7	12.5	9.2
10	9.0	15.5	8.9	10	9.1	14.1	9.1
11	9.4	14.6	8.7	11	9.3	13.3	8.6
12	9.0	8.0	8.3	12	9.0	10.2	8.3

An almost average 25% increase in CO and PM_{10} originating from background sources can be observed from July to December. In other words, on a seasonal basis, PM_{10} shows a maximum level during winter and spring.

This increase can be explained through more fuel consumption by industries as well as lower wind velocities for the time considered. Almost the same situation appears for residential-commercial heating. In the meantime, some authors have noted that that the highest average concentration for CO, and PM_{10} occurred at a humidity level above 80% and a temperature range between 10 to 20 C°.

Conclusions

The adopted approach, based on the development of a hybrid statistical-deterministic distribution is capable of assessing the emissions resulting from the main human activities as anthropogenic pollutants in an urban air pollution study. This approach may also determine the share of responsibility of human economic activities based on the pollutant concentration measured at the GTA monitoring stations network. The results demonstrate an acceptable consistency between the pollutant concentrations observed in the ambient air and those estimated by the model. Therefore the major purpose of this study can be summarized as follows:

. To determine the degree to which various economic activities are responsible for the existing pollutant concentrations of SO_2, CO and PM_{10} pollution in a mega city like Tehran.

. Using a Hybrid model based on available limited data and minimum number of variables compared to inherent constraints of deterministic models.

The contribution for SO_2 differs from that of the two other pollutants as road traffic is not the predominant source. Background level declined from 85% to 82%, while road traffic increased from 11.5% to 13.0% for the period considered.

Also residential-commercial heating kept its average share equal to 4%. Based on the model outputs, traffic contribution, relevant to CO immissions in the GTA region, has been determined as being very significant with an average value of 77% for the year 2003 and 79% for year 2005. In addition, background level share decreased from 21% for the year 2003 to 18% in 2005. Residential-commercial heating doesn't appear to play a significant role at all. In relation to PM_{10} pollution, relatively similar results gained as traffic share appeared around 75% for both years, while background sources decreased from 70% to 65% for the years 2003 and 2005, respectively. In contrast, residential-commercial heating, increased from 8.5% to 10%, when applied to the same period.

Significant improvement in the GTA ambient air quality in reference to CO and PM_{10} can be mainly imputed from traffic diminution on the one hand and enhancement of local transportation facilities on the other hand.

However, consumption of low sulfur content fuels in power plants and other local and non-local industrial activities, as well as buses, minibuses and heavy trucks operating within the GTA, will be of the utmost concern in pursuing any policy to reduce SO_2 in ambient air.

The model is available in computer code at Sharif University of Technology (SUT).

Acknowledgments

The authors deeply appreciate the contribution and assistance to this project of Professor Walter Hecq at the Centre for Economic and Social Studies on the Environment, Université Libre de Bruxelles. We also extend our gratitude to AQCC, DOE and IRIMO for providing invaluable data for this study.

References

Bartlett, J. E., Kotrlik, J. W., Higgis, C. C., (2001). Information Technology Learning and performance Journal, 19(1).

Benarie, M. M., (1980). The simple box model simplified. In Atmospheric Pollution, Proceedings of the 14th Intl. Colloquium, Studies in Environmental Science. 8, 49-53.

European Protection Agency (EEA), (1996). Emission inventory guide book.

Favrel, V., Hecq, W., (1998). A model for the assessment of the contribution of road traffic to air pollution in the Brussels urban area, Transport: Assessing and reducing the impacts. 27-29 Oct., Milan, Italy.

Gokhale, S., Khare, M., (2004). A review of deterministic, stochastic and hybrid vehicular exhaust emission models. International Journal of Transport Management.

Hecq, W., Borisov, Y., Debever, C., Dupierreux, J. M., (1994a). An empirical hybrid model for the assessment of daily SO2 concentrations in an urban environment. Journal of Environmental Management 42, 181-198.

Holzworth, C. G. (1964). Estimates of mean maximum mixing depths in the contiguous United States, Mon. Wea. Rev. 92, 235-242.

Jakeman, A. J., Simpson, R. W., Taylor J. A. (1988a). Modeling distributions of air pollutant concentrations, the hybrid deterministic-statistical distribution approach. Atmospheric Environment. 22, 163-174.

Jakeman, A. J., Jun, B., Taylor, J. A., (1988b). On the variability of the wind speed exponent in urban air pollution models. Atmospheric Environment. 22, No. 9, 2013-2019.

Koo, J. K., Ko, S. O. and Hong, M. S. (1990). Acid rain model considering altitudinal precipitation rate. Atmospheric Environment, 24A, 2133-2139.

Melli, P., Bolzern, P., Fronza, G., Spirito, A., (1981). Real-Time Control of Sulphur Dioxide Emissions from an Industrial Area. Atmospheric Environment. 15, 5, 653-666.

Milioins, A. E., Davies T. D. (1994). Regression and stochastic models for air pollution - I. Review, comments and suggestions. Atmospheric Environment. Vol. 28, no. 17, 2801-2810.

Nunnari G., Dorling S., Schlink U., Cawley G., Foxall R., and Chatterton T. (2004). Modeling SO2 concentration at a point with statistical approaches Environmental Modeling & Software. Volume 19, Issue 10, 887-905.

Zannetti, P., and Switzer, P., (1979). The Kalman filtering method and its application to air pollution episode forecasting. Paper presented at the APCA Specialty Conference on Quality Assurance in Air Pollution Measurement, New Orleans, Louisiana.

Zannetti, P. (1990). Air Pollution Modeling, Theories, Computational Methods and Available software, Computational Mechanics Publications. Van Nostrand, Reinhold.

Acronyms

ABL: Atmospheric Boundary Layer
APPETISE: Air Pollution Episodes: Modeling Tools for Improved Smog Management
AQCC: Air Quality Control Company
CESSE: Centre for Economic and Social Studies on the Environment
DD: Degree of Day
DOE: Department of Environment
ECE: Economic Commission for Europe
EEA: European Environmental Agency
GEF: Global Environmental Facility
GTA: Greater Tehran Area
HT: High Traffic
IRIMO: Iran Meteorological organization
JICA: Japan International Cooperation Agency
LIDAR: Light Detection and Ranging
LT: Low Traffic
MH: Mixing Height
MT: Medium Traffic
MTC: Motor Test Center
NG: Natural Gas
NIGC: National Iranian Gas Company
NIOC: National Iranian Oil Company
SMHI: Swedish Meteorological and Hydrological Institute
SODAR: Sonic Detection and Ranging
SWECO: Swedish Company
TERP: Tehran Transport Emission Reduction Project
TGRC: Tehran Geographical Research Center
UNEP: United Nation Environment Program
VKT: Vehicle Kilometer Traveled
WHO: World Health Organization