A review of deterministic, stochastic and hybrid vehicular exhaust emission models

Sharad Gokhale, Mukesh Khare *

Department of Civil Engineering, Indian Institute of Technology Delhi, Hauz Khas, New Delhi 110 016, India

Abstract

Vehicles spend more time near junctions and intersections in different driving modes, i.e., queuing, decelerating or accelerating and thus generating more pollutants than at road links [Claggett, M., Shrock, J., Noll, K.E., 1981. Carbon monoxide near an urban intersection. Atmos. Environ. 15, 1633-1642]. As a result, the receptors in these urban corridors are prone to frequent exposures of high pollutant concentrations (episodic conditions). In order to predict such 'episodes', an air quality model, capable of estimating the entire range (middle and extremes) of pollutant concentration distribution is needed. Hybrid models (combining deterministic and statistical distribution models) have demonstrated the ability to predict the entire range of pollutant concentrations in such complex dispersion situations with reasonable accuracy [Jakeman, A., Simpson, R.W., Taylor, J.A., 1988. Modelling distributions of air pollutant concentrations-III: Hybrid modelling deterministic-statistical distributions. Atmos. Environ. 22 (1) 163-174]. The present paper reviews the relevant deterministic and stochastic based vehicular exhaust emission models that may be hybridized and thus generate a hybrid model with improved prediction accuracy. The paper also describes the implications of hybrid models in formulating the Episodic-Urban Air Quality Management Plan (e-UAQMP).

Keywords: Vehicular exhausts; Statistical models; Statistical distribution models; Hybrid models; Probability distributions; Episodic conditions

1. Introduction

Air quality models are indispensable tools to assess the impact of air pollutants on human health and the urban environment. The most critical part of assessment studies is to know the present as well as future air quality levels. Several deterministic based models exist to evaluate and predict the pollutant dispersion in urban areas but majority of them are 'causal' in nature and fail to predict the 'extreme' concentrations (Khare and Sharma, 2002; Jakeman et al., 1988). The statistical distributional models that are 'non-causal' and based on the historical data overcome the above limitation and predict the 'extreme' concentrations with reasonable accuracy (Jakeman et al., 1988). However, what is needed is not only 'extreme' ranges of prediction but also the 'middle' ranges. Hybrid modelling is one of the techniques that estimate the 'entire range' of the distribution of air pollutant concentrations by combining deterministic based models with suitable statistical distributional models. In the past, Taylor et al. (1985) applied such a technique to predict the entire range of pollutant concentrations for vehicular exhaust emissions. Recently, Gokhale et al. (2003) have attempted to develop a hybrid model for one of the traffic intersections in Delhi, India, to predict the entire concentration profile of carbon monoxide (CO).

The paper first, describes the vehicular pollution studies carried out at roadways and traffic intersections. It is followed in Section 3 by a comprehensive review of the deterministic, numerical, statistical, statistical distribution and hybrid vehicular exhaust models. The
implications of hybrid model applications in traffic management are described in Section 4 with concluding remarks in Section 5.

2. Air pollution studies at traffic intersections and urban roadways

Over the last two decades, motor vehicles have emerged as one of the critical sources of urban air pollution. They are the largest source of gaseous and particulate (PM10) emissions in most of the Asian cities (Bombay, Calcutta, Delhi, Dhaka, and Karachi in South Asia and Bangkok, Beijing, Shanghai, Jakarta, and Manila in East Asia) exceeding the contributions from resuspended road dust, heavy fuel oil and coal combustion, and refuse burning (Faiz and Sturm, 2000). In these cities, pollution levels often exceed World Health Organization (WHO) air quality guidelines by a factor of three or more (World Resource Institute, 1992, 1998).

Vehicular pollution dominates all other sources in urban centers throughout the world. As such, many studies ranging from simple measurements and reporting to the sophisticated rigours of modelling exercises in complex urban environments have been reported in literature. Kondo (1973) presented the efforts of the Japan Society of Mechanical Engineering (JSME) to develop an Air Pollution Prediction System (APPS) at an intersection. Noll et al. (1974), Noll and Miller (1975a,b) and Noll et al. (1977) described the environmental impact of highways on ambient air quality. Ellis (1978) analyzed the effect of vehicle 'cold' and 'hot' operating conditions on exhaust emissions. Eskridge and Hunt (1979) discussed the prediction of traffic induced turbulence and velocity fields near urban roadways. Heinmiller (1978), Ostrouchov (1978) and Chang et al. (1980) carried out studies to investigate the impact of ambient temperature on vehicular exhaust emissions. Rao et al. (1979), Middleton et al. (1979), Brennan and McCrae (1988) and McCrae and Hickman (1989) reported vehicular air pollution studies related to characteristics of turbulence and dispersion of pollutants at complex urban intersections and roadways. Thus the air quality near urban road intersections has been a subject of a number of publications (Kunzelman et al., 1974). Dabberdt and Sandys (1978) published a procedure for calculating CO concentration near congested intersections. Clagget and Miller (1979) evaluated a line source model for CO at the urban freeway and intersection. Claggett et al. (1981) presented a methodology for identification of the air quality levels near intersections based on diffusion model predictions and provided an analysis of CO data collected near a signalized urban intersection.

A hybrid methodology (CAL3Q), based upon signalized intersection analysis and deterministic queuing theory was developed by Transport Research Board (Newell, 1982; Schattanek et al., 1990). In 1989, the USEPA commissioned a performance evaluation of several methodologies that combined emission, traffic and dispersion models to identify the modelling approach that best estimated CO concentration near congested intersections. Of the eight models tested, the CAL3QHC performed best in predicting CO concentrations in the vicinity of a congested intersection (Schattanek et al., 1990).

A number of case studies for various regions in different countries related to the impact of vehicular emissions on the urban air quality have been reported by Longhurst et al. (1996), Heida et al. (1994), Sulieman et al. (1994) and Jenkins et al. (1995). Akerdalu (1994) modeled the vehicular CO emissions at a Nigerian city traffic light controlled roadway junction. Hongchang and Daniel (2001) used CAL3QHC to predict CO and nitrogen dioxide (NO2) concentrations at one of the intersections at Shanghai (China). The model under predicted the pollutant concentrations as it failed to take into account the effects of parameters, e.g. ‘queuing’, ‘accelerating’, ‘decelerating’, ‘delays’ and ‘erratic’ traffic flows resulting from ‘heterogeneity’ in traffic. Recently, Gokhale and Patil (2004) studied the size distribution of aerosols at one of the traffic intersections in Mumbai, India. The study revealed that the fine mode of particulate, i.e., PM1.1 contributes about 30% of the PM10.

3. Models and their application

The air quality models are broadly classified into four types, viz. deterministic, statistical, statistical distribution and hybrid. The deterministic models are based on fundamental mathematical description of atmospheric processes and establish the cause and effect relationship between the output and the inputs. The statistical models, on the other hand, are based upon semi-empirical statistical relations among available data and measurements. The statistical distribution models are probability-based capable of estimating the entire range of pollutant concentration distribution. The hybrid models are the combination of deterministic and statistical distribution models.

3.1. Deterministic models

Turner (1970) developed a model from the equation for infinite line source of perpendicular case, presented by Sutton (1932), considering oblique winds when the angle between the wind and the line source is greater than 45 degrees. Later, Csanady (1972) developed a model for finite line source but it is applicable only when the wind is perpendicular to the roadway. Another Gaussian-based model is the California line source.
model (CALINE) that uses separate equations for calculating pollutant concentrations under crosswind and parallel wind conditions (Beaton et al., 1972). Calder (1973) worked on Turner's model and shows that the equations give incorrect results for oblique winds. He has derived an approximation formula, which gives comparatively accurate results for wind angles down to 15 degrees. Carpenter and Clemana (1975) developed the AIRPOL-4 model based on the technique of segmentation in conjunction with an appropriate numerical scheme to evaluate the Gaussian integral. The roadway co-ordinate, according to this model, is translated onto a receptor co-ordinate system. The two co-ordinate systems have the advantage that they permit the Gaussian equation to be directly applied to each roadway point. The model is capable of predicting concentrations for both upwind and downwind of a roadway at any desired sampling intervals. The US EPA later came out with another Gaussian based model, EPA-HIWAY-1 (Zimmerman and Thompson, 1975), which estimated the pollutant concentration from the vehicular exhausts. CALINE (Beaton et al., 1972) was further developed into a newer version, CALINE-2 (Ward et al., 1977). It followed the concept of the 'box model' and assumed that pollutants were well mixed over the roadway up to a fixed height.

The general motors (GM) experiments show that mechanical mixing plays an important role in dispersion of pollutants near the roadway (Chock, 1977a). The comparison of experimental results with the EPA-HIWAY model has revealed many limitations and deficiencies. These include (i) inadequate dispersion parameters; (ii) no treatment for plume rise; (iii) tendency towards severe 'overprediction' when winds are parallel to the road (Chock, 1977b). Chock (1978) proposed a simple line source (GM) model to describe downwind dispersion of pollutants near the roadway. This model was developed to relax the limitations and at the same time to eliminate the necessity of numerical integration in estimating the roadway pollutant dispersion.

As a part of major roadway dispersion project undertaken by the New York State Department of Environmental Conservation, a performance evaluation study was carried out for four Gaussian dispersion models viz. HIWAY-1 (Zimmerman and Thompson, 1975), AIRPOL-4 (Carpenter and Clemana, 1975), CALINE-2 (Ward et al., 1977), GM (Chock, 1978) and four numerical models viz. DANARD (Danard, 1972), MROAD-2 (Kirsh and Mason, 1975), RAGLAND (Ragland and Pierce, 1975), ROADS (Pitter, 1976). Of the models tested, the GM and HIWAY-1 models performed better compared to other Gaussian models. However, the GM model provides a better simulation, by far, for parallel wind cases than any of the other models tested in both neutral and unstable stability conditions (Sistla et al., 1979). As a result, CALINE-3, the third version of CALINE series, superseded CALINE-2 after the latter has been found to 'overpredict' substantially the observed concentration under stable and/or parallel wind conditions. Rao and Keenan (1980) evaluated the EPA HIWAY-1 model (Zimmerman and Thompson, 1975) and came out with an improved version, the HIWAY-2. The HIWAY-2 uses three stability curves (unstable, stable, neutral) out of six Pasquill-Gifford curves and gives more realistic concentration estimates, when compared to its previous version. Cohn and McRoy (1982) studied some of the features of CALINE-3; among them, are explicit considerations of averaging time, surface roughness, deposition velocity, elevated highway sections, and mixing cell volume. Green and Bullin (1982) evaluated the roadway dispersion models viz. CALINE-2, HIWAY, AIRPOL-4 and TRAPS IIM using experimental mass flux profiles and mass conservation checks. The shapes of the predicted and experimental mass flux profiles agree for only one model that is TRAPS IIM. Several other roadway dispersion models have been found to be inaccurate because they neglect the change in the wind speed with height but the TRAPS model allows the wind speed to vary with height. This model employs the gradient-transport equation and represents the actual transport characteristics. Chock (1982) studied dispersion of pollutants near roadways based on GM experiments and also reviewed experiments and models to-date. Rodden et al. (1982) evaluated CALINE-3, CALINE-2, AIRPOL-4A, HIWAY-2 and TRAPS IIM using data collected in Houston, Dallas, San Antonio, El Paso and Texas as well as the GM experiments data. CALINE-2 and CALINE-3 show nearly identical statistics for the Texas sites. There have been some improvements observed in CALINE-3 predictions for the GM data.

Hickman and Colwill (1982) described a simple and effective method of estimating pollutant concentrations around highways, which used the Gaussian dispersion theory with empirical modifications. Another method developed at TRRL, used a set of graphs constructed from computer model that take into account vehicle flow, vehicle speed, and distance of receptor from roads (Waterfield and Hickman, 1982).

CALINE-4, the most recent version of the CALINE series represents a refinement and extension of the capabilities contained in CALINE-3. The model has been verified using data from several independent field studies including at low wind speeds and parallel wind conditions (Benson, 1984, 1986, 1992). Luhar and Patil (1989) further modified the GM model and developed the general finite line source model (GFLSM), which takes into account all orientations of wind directions. The GFLSM requires the receptor location at 90 degree to the segment of road and the length of the line source (segment of the road) as three times the distance between the receptor location and the road. Kono and
Ito (1990a) developed a microscale dispersion OMG-volume source model, which has been compared with measured SF concentrations by Kono and Ito (1990b) along with HIGHWAY and JEA model. The IITCO model (Singh et al., 1990) has been used to predict the CO concentration for Kuwait traffic conditions. The results have been compared with the intersection midblock model (IMM), which outperformed the IITCO model. Burden et al. (1994) applied CALINE-4 and DMRB (The Highways Agency, 1994) models to predict NO2 concentrations at an intersection in Bristol, UK. Akeredou et al. (1994) applied CALINE-4 to forecast CO at road intersections. Dabberdt et al. (1995) evaluated and compared HIGHWAY-2 and CALINE-4 and a numerical model near one of the urban intersections. The results show that the Gaussian based models under performed. Recently, Khare and Sharma (1999) have modified GFLSM for Delhi's traffic conditions to develop the Delhi finite line source model (DFLSM).

Gramotnev et al. (2003) determined average emission factors for vehicles on a busy road. Marmure and Manmane (2003) evaluated several mobile and line source models in Israel. For oxides of nitrogen (NOx) concentration prediction, the study compared two line source models, i.e., CALINE-4 (Benson, 1992) and HIWAY-2 (Peterson, 1980) and two mobile source models, i.e., MOBILE-5B (USEPA, 1994) and COPERT-3 (Ntziachristos and Samaras, 2000; Kouridis et al., 2000; COPERT-III, TR-49, 2000). The results show that COPERT-3 generates appropriate emission factors for free flowing traffic situations. At grade road sections, CALINE-4 and HIWAY-2 perform in a similar way; for cut/depressed sections, HIWAY-2 performs better during unstable conditions while CALINE-4, is better during stable conditions.

3.2. Numerical models

Danard (1972) developed a two-dimensional Eulerian model, DANARD, to predict CO near a highway, which solves the mass-conservation equation numerically as outlined by Dufort and Frankel (1953). Ragland and Pierce (1975) developed a model, RAGLAND, to predict the pollutant concentration due to highway traffic. It solves the continuity equation for either parallel or non-parallel wind vectors using an efficient matrix inversion technique. In this model, the boundary conditions, proposed by Danard (1972) were used with more appropriate values of crosswind diffusivity. Kirsh and Mason (1975) developed the MROAD-2 model, which numerically solves the mass conservation equation. The size of the grid is specified by the user and the model allows the existence of several line sources, including elevated roadways. Pitter (1976) described ROAD model, which is a two-dimensional conservation model. The model determines the steady state concentration of pollutants by numerically solving the equations, governing atmospheric advection-diffusion and chemical reactions. Eskridge and Thompson (1982) developed the ROADWAY model, based on finite difference principle that predicts pollutant concentrations near a roadway. The ROADWAY model includes the vehicle wake theory, which was originally developed and later modified by Eskridge and Thompson (1982) and Eskridge and Rao (1983, 1986), respectively. Eskridge and Thompson (1982) further developed the ROADCHEM model incorporating the chemical reaction involving NO, NO2, and ozone (O3) as well as advection and dispersion. Other relevant numerical models are PAL (Peterson, 1978) and PALDs (Rao, 1982; Rao and Snodgrass, 1982), developed by USEPA used for predicting line source emissions.

Recently, there have been advancements in several models that are based on the k-theory approach. CAR-FMI model was developed from the GFLSM (Luhr and Patil, 1989), in which, the dispersion parameters have been modeled as functions of Monin-Obhukov length, the friction velocity and the mixing height (Harkonen et al., 1995, 1996; Karppinen et al., 2000a,b). Ketzel et al. (2000) compared the numerical street dispersion model with the measurements from wind tunnel and field measurements. Oettl et al. (2001), with the FMI team, presented an evaluation of the performance of the Gaussian line source model, CAR-FMI and a Lagrangian dispersion model, against the NOx experimental data at Elimaki. The CAR-FMI model has shown poor performance for parallel winds. The CAR-FMI results have also been compared with the measurements of NO, NO2, and O3 near major roads in Espoo (Walden et al., 1995; Harkonen et al., 1997) and further evaluated against the field measurements near a major road in Southern Finland (Kukkonen et al., 2001a). Rao (2002) recently modified ROADWAY-2 model based on US EPA’s ROADWAY model and has included the option for simplified chemical reactions for NO, NO2, and O3 as described by Eskridge and Catalano (1987). The model also incorporates atmospheric boundary layer with turbulent kinetic energy closure and surface parameterization, which derives the mean and turbulent profiles from input meteorology.

3.3. Statistical models

Statistical models have been used to establish an empirical relationship between air pollutant concentrations and meteorological parameters. They are quite useful in real time short-term forecasting. Hikaru and Tsukatani (1973) used a statistical model based on the time-series of the air quality data for oxides of sulfur (SOx) predictions in Japan. The model relates air pollutant concentrations with averaging time and frequency of occurrence. The percentiles of concentrations are functions of probability density, approximated by the
lognormal distribution, and spectrum density, approximated by the Markoffian spectrum. Hipel et al. (1975) described the formulation of an intervention analysis model using the stochastic transfer function technique. Box and Tiao (1975) applied intervention analysis to economic and environmental problems, describing the effects of interventions on a given response variable in the presence of dependent noise structure. Hipel et al. (1977a) used the Integrated Autocorrelation Function (IACF) and Integrated Partial Autocorrelation Function (IPACF) methods in developing Autoregressive Integrated Moving Average (ARIMA) model based on Box-Jenkins approach. Drufuca and Giugliano (1977) developed a prediction model for sulfur dioxide (SO\textsubscript{2}) concentrations based on theory of probability of exceedances and return period. Tiao and Hillmer (1978) analyzed CO and lead (Pb) emissions data for Los Angeles city to determine the effect of intervention as a result of the introduction of the catalytic converter. This empirical model has helped in determining trends in CO and Pb concentrations with traffic data and meteorological conditions. Roberts (1979a,b) and Shively (1990) used extreme value theory to estimate the long-term trend of O\textsubscript{3} concentrations at two monitoring sites in Houston. Finzi et al. (1980) employed time series models with pollutant and meteorological variables to single-site multiple-variables studies. They have illustrated three multivariate stochastic mathematical models of daily SO\textsubscript{2} concentration in an urban area of Italy. Multivariate time series models (ARIMAX), in which the pollutant concentrations at a certain instant are expressed as linear combinations of present and previous physical inputs, have also been described by Finzi and Tebaldi (1982). Bacci et al. (1981) developed a stochastic based model of SO\textsubscript{2} dispersion around a power plant. The model described the diurnal dynamics of a variable taken as representative of ground level pollution viz. 2-h Dosage Area Product (DAP). The basic difficulty in actual implementation of the DAP indicator is the requirements of future radiation, wind direction and speed to be forecast separately at each time step. Murray and Farber (1982) modeled the historical-visibility sulfate database statistically using time-series analysis for Salt Lake City for 3-year time period, beginning January 1971. This study examined the procedure involved in selecting the most appropriate statistical techniques for this kind of database. It also made a comparison of various types of regression models to the statistical methods. Hayas et al. (1982) used the spectral technique to atmospheric dust pollution to determine the periodicities of the time series formed by 3-h concentrations of atmospheric particulates for the winter and summer months. Hirtzel and Corotis (1982) described the effect of autocorrelation on the distribution of the maximum value in a sequence of autocorrelated variables. The results of this study were later used in developing a method for calculating the calibrated probability distribution of the maximum concentration as a function of the autocorrelation. North et al. (1984) developed a Markov-type model based on up and down crossings of threshold concentrations of series of daily CO concentration in Madrid, Spain. Chock (1984) studied the effects of autocorrelation on the extreme values. It has been found that a high positive autocorrelation significantly lowered the values of the 'extreme' and 'near extreme', while increasing their variability. Baker et al. (1984) compared the predictions of chemical trace concentrations from two stochastic models under explicit assumptions. Of these models, one is based upon a linear stochastic equation in which random removal events or storms form a Poisson process, and the other is based upon multiplicative random process, which produces a lognormal distribution. Simpson and Jakeman (1985) developed the CRES model to forecast the worst-case pollution scenarios for acid gases and suspended particulates in Newcastle, Australia. This model takes into account the effects of both long-term meteorological fluctuations and changes in emissions. Nouh (1986) introduced a methodology to determine the sufficient period of sampling CO concentrations at a particular location in Saudi Arabia using a stochastic approach.

Recently, the single monitoring site-multiple variable category has received the greatest attention due to the application of chemical mass balance regression type receptor models and also multivariate models (Henry et al., 1984). Lowenthal and Rahn (1987) used principal component and factor analysis methods in the receptor model studies. Later, in order to deal with missing observations in air pollution data sets, methods proposed by Davison and Hemphill (1987) have invariably been used to replace the missing observations. Butterman (1991) has described a procedure to estimate missing data, to determine extrema, and to derive the uncertainties for the air quality data. Hernandez et al. (1992) used state space modelling coupled with Kalman filtering and Box Jenkins (Transfer Function and ARIMA) to forecast daily air particulate iron (Fe) and Pb concentrations in Madrid, Spain. The adaptive state-space models performs significantly better than ARIMA models. A statistical time-series analysis was applied to study the interdependence between primary and secondary pollutants in the Taipei area (Hsu, 1992). Estimation using the Vector Auto Regression Model (VAR) has indicated that 2-4 h time lags are sufficient to represent the observed values at two selected sites. The impulse response functions and variance decompositions of NO, NO\textsubscript{2} and O\textsubscript{3} were derived using the vector moving average representation to examine the significance of one species on others. Influence of photochemistry and transport processes on these air pollutants at different locations has also been evaluated. Use of a different
and more complex type of real time predictor, the Kalman filtering technique has also been reported in the literature (Ng and Young, 1990; Young et al., 1991; Hernandez et al., 1991; Schlink and Herberth, 1997). Salcedo and Dias (1992) studied the time-series analysis of air pollution levels, which involve identification of long-term variations in mean (trend) and of cyclical' or 'periodic' components. Monteiga et al. (1993) presented a dynamic system, used to predict the ambient concentrations of SO2, every 5 min. For forecasting, they have used a 'mixed' model, which has a parametric and non-parametric component. Confidence intervals have also been evaluated for future observations, using bootstrap and classical techniques. Milionis and Davies (1994a) used regression and stochastic models for air pollution prediction. It has been argued that stochastic models are preferred from both the conceptual and practical points of view. Following this, Milionis and Davies (1994b) have used Box-Jenkins Transfer Function modelling to assess the relative importance of various meteorological factors on the surface smoke concentration. Box-Tiao intervention modelling has also been applied to examine the effects of extreme meteorological events on air pollution concentrations. In another study, Trier and Firinguetti (1994) have suggested the superiority of Transfer Function models over Kalman filtering technique. Sharma et al. (1999) used the extreme value theory to predict the number of violations of the national ambient air quality standards (NAAQS) for an urban road intersection in the Delhi city, India. Sharma and Khare (1999) applied intervention analysis based on the Box Jenkins approach to investigate the impact of legislation to control vehicular pollution. Sharma and Khare (2000a,b) have also presented a short term and real time forecast of the ambient air pollution levels due to vehicular sources. In addition, Sharma and Khare (2001) have carried out a comprehensive review of deterministic and statistical models for vehicular exhaust emissions. Further, stochastic ARIMA has been developed for maximum O3 concentration forecasts in Athens, Greece. For this purpose, Box-Jenkins approach has been applied to analyze nine years of air quality observation records (Sliz et al., 2002). Recently, Dirks et al. (2002, 2003) have developed semi-empirical models based on standard meteorological and traffic information. The models have been applied to investigate the effects of changes in meteorology and traffic flow on carbon dioxide (CO₂) concentrations. They have also developed a model for predicting and/or interpolating the missing CO concentrations data. In this model, the site-specific characteristics like turbulence have been incorporated.

Recently, statistical tools such as artificial neural networks (ANN) and Fuzzy Logic Theory (FLT) have been used as alternative tools in modelling pollutants from vehicular traffic. Raimondi et al. (1997) reported an air pollution model based on FLT, which takes into account model uncertainties and describe daily dynamics of a variable Dosage Area Product (DAP). A neural network based model has been developed by Drozdowicz et al. (1997) to predict the CO concentration in the urban areas of the Rosario city. Conme (1997) and Conme and Diem (1999) developed multivariate regression models for CO prediction using variables and interaction terms. These models are based on nocturnal stability as well as time series. Khare and Sharma (2001) developed a traffic volume model (TVM) using inverse modelling technique for Delhi’s traffic conditions. Nagendra and Khare (2002) presented a detailed review of vehicular exhaust emission models including ANN based models. In addition, Nagendra and Khare (2003) have further analyzed the one-year data on traffic, emission and meteorology using principal component analysis (PCA). Fisher (2003) has presented a comprehensive review on Fuzzy set theory. This theory can be applied to decision-making processes involving pollution prevention and human health assessment.

3.4. Statistical distribution models

Larsen (1971) carried out a comprehensive analysis of the data collected under the continuous air quality-monitoring program for the year 1962-1968. The data contains concentrations of seven pollutants viz. CO, NO, NO₂, NOₓ, O₃, SO₂ and hydrocarbon (HC) for eight cities. The study aimed to evaluate various common statistical distributions. Based on this, Larsen has concluded that regardless of pollutant types, site conditions, averaging time and urban concentration, frequency distributions are in the lognormal distributional form. Furthermore, Barry (1971), Scriven (1971) and Gifford (1974) conducted tracer studies that show an exponential distribution of pollutant concentrations for isolated sources. Potential applications of the extreme value theory to the air quality data have been described by Singpurwala (1972). Lynn (1974) compared the 2- and 3-parameters gamma and lognormal distributions, the normal distribution, the four parameters beta distribution, and the Pearson distribution, concluding that the two parameter gamma and lognormal models are preferred over the three parameters models. Polliaek (1975) considered the two-parameter Weibull distribution, as applicable to air quality data along with lognormal, gamma and Pearson distributions. Lemon (1975) developed maximum likelihood estimators for the 3-parameter Weibull distribution, based on various left and right-censored data situations. For the case of single censoring from the left and progressive censoring from the right, the developed estimation procedures have involved the simultaneous solution of two iterative equations to the arduous task of solving three simultaneous equations, as outlined by Harter and Moore (1965).
For the 2-parameter Weibull distribution (Cohen, 1965), the above case has resulted in an estimation procedure involving only one iterative equation. Curran and Frank (1975) show that exponential or Weibull distributions, in general, yield a better fit to air pollutant data. Ott and Mage (1976) suggested that a censored 3-parameter lognormal distribution is applicable to the air quality data. Bencala and Seinfeld (1976) observed that several common distributions could be used to fit the observed data, one of which may be the lognormal distribution. They have examined the 2- and 3-parameters lognormal, gamma and Weibull distributions and have found that the 3-parameter lognormal model performs the best. Kalpasanov and Kurchatova (1976) investigated the statistical distribution of air pollutant concentrations in Sofia, Bulgaria. It has been found that the distribution is skewed even after the logarithmic transformation. Cohen and Norgaard (1977) used the maximum likelihood method in order to estimate parameters for the censored samples using 3-parameter gamma distribution (Pearson type III distribution). Mage and Ott (1978) examined the 'straightness' of Larsen's original SO2 probability plot and have observed that rather than fitting the standard 2-parameter model, the data describe a curve that is better suited to the censored 3-parameter lognormal model. Thus, they conclude that the lognormal model may be a useful practical tool for representing the frequency distributions of air pollutant concentrations. Statistical techniques have also been applied to model wind speed data in the form of frequency distributions, which has covered the range from 'calm' to the 'maximum' observed wind speed (Mage, 1980).

Tsukatani and Shigemitsu (1980) proposed determination of Pearson system of distribution by using the first three moments of the data. About half of the data observed, show a converse curve when the cumulative frequencies are drawn on the probability paper of lognormal distribution which belong to type VI of Pearson system of distribution. The Pearson system of distribution fits a wide variety of data and is more flexible than the lognormal model. Cats and Holtslag (1980) investigated the dependence of air pollution frequency distributions on wind direction by dividing SO2 data sets into subsets. They found that the 2-parameter lognormal model provides a satisfactory estimate of the 98-percentile concentration. The frequency distribution of each subset turned out to be lognormal with the logarithmic standard deviation equal to 0.76 for hourly concentration measurements. Nieustadt (1980) performed a comparative study of the hourly and daily averaged concentrations of SO2, evaluated using the Gaussian plume model, for a period of three months at 125 receptor points. The agreement between calculated and observed data at the 50-percentiles has been found to be satisfactory with background concentration. Berger et al. (1982) analyzed 24-h SO2 data in Ghent, Belgium. The analysis shows that the 2-parameter gamma distribution model better represents the data than a lognormal distribution model. Holland and Terence (1982) developed a computer program for fitting statistical distribution to air pollution data using maximum likelihood estimation. The software made use of maximum likelihood estimation to fit the following distributions: normal, 3-parameter lognormal, 3-parameter gamma, 3-parameter Weibull, Johnson SB and 4-parameter beta distributions. The goodness-of-fit for each probability model was compared in several ways. Later, an attempt was made by Surman et al. (1982) to verify the assumptions of Larsen's model, particularly the assumption of lognormal distribution of pollutant concentrations. Similar studies have been undertaken by Barlow and Singpurwala (1972), Marcus (1973), Mage and Ott (1978), Curran and Frank (1975) to examine the validity of Larsen's model prediction

Georgopoulos and Seinfeld (1982) presented methodologies and limitations in describing air quality through statistical distribution of pollutant concentrations and have explained the use of extreme statistics in the evaluation of different forms of air quality standards as well as interpolation of the rollback calculations. Simpson and Jakeman (1984) proposed a model to estimate the effect of long-term meteorology on maximum daily acid gas levels for 10 years of data for Newcastle in Australia. The model assumed that: (i) the probability distribution function of both air pollution data and wind speed are lognormally distributed and, (ii) on average, an inverse relationship exists between air pollution levels and wind speed. The model showed that about half of the variations are directly related to fluctuations in the wind speed distribution. Simpson (1984) examined the estimates for a cumulative frequency distribution using four different types of data set: (i) continuous monitoring throughout the year (ii) continuous monitoring for one week out of every four (iii) random sampling at any time (iv) random sampling only on weekdays between 9 am and 5 pm. The results show that out of four data sets, data set (ii) has yielded good results. Simpson also used Larsen's model to reproduce cumulative frequency distribution for O3, NO2 and TSP. He has found that data set (ii) has been the most appropriate for making predictions. Simpson et al. (1984) developed a model to estimate frequency distribution of SO2 concentrations for isolated point source and maxima at different averaging times, given the maximum at one averaging time. The data sets were collected from three monitoring sites at the power station in the Upper Hunter Valley in New South Wales, Australia. The model estimated maxima quite well for the 0.5-, 1-, 3-h averaged data sets using 8-h averaged data set. Mage and Ott (1984) investigated the effectiveness of three different approaches for estimating parameters using a lognormal distribution—method of fractiles, method of moments and method.
of maximum likelihood. The results have been compared with another approach—direct empirical linear rollback. It has been found that the empirical linear rollback approach is influenced by the stochastic properties of extreme values, creating a tendency to ‘underpredict’ or ‘overpredict’. Giugliano (1985) developed empirical models to describe extreme values of SO2 concentration from the data collected over a 10 years period. Simpson et al. (1985) have evaluated the Atmospheric Turbulence Diffusion Laboratory (ATDL) model using the statistical distributions for TSP and acid gas concentration data along with wind speed data. The study has concluded that the ATDL model performs well in the 30-70 percentiles range of statistical distributions when the daily average and meteorological data are used. Taylor et al. (1986) examined the process by which a distribution model may be identified from a range of alternative models. A procedure for selecting amongst the exponential, lognormal and Weibull distribution has been applied to 24-h average suspended particulates (b-scattering), O3, CO, SO2, NOx and nitrous oxide (N2O) observations recorded in Melbourne, Australia. It was found that (a) lognormal distribution is appropriate for particulate data and the majority of the NO, NOx, SO2 data sets (b) gamma distribution is best for both CO, NOx and O3 (c) Weibull distribution is appropriate for a significant number of CO and O3 data sets.

For identifying normal (or lognormal) distributions for air quality data, it has been found that the Lilliefors test and the D’Agostino statistic are more appropriate than the \( x^2 \) (v) test for the wide range of non-normal set of data. The selection procedure was based on the work of Taylor and Jakeman (1985) who have combined the Lilliefors statistics for these distributions with the likelihood functions to identify the best model. Jakeman et al. (1986) described the procedure for the estimation of parameters for a distributional model. Monte Carlo simulation has been used to compare methods for fitting statistical distributions to such data, where the distributional form is known. Taylor et al. (1987a) used three statistical models to predict the upper percentiles of the distribution of air pollutant concentration from restricted data sets recorded over yearly time intervals. The first is an empirical quantile-quantile model. The second model represents the priori selection of a distributional model for the air quality data. The third model has employed an identification procedure on each data set. It has been found that the first and the third models yield the best results when the three highest values and the 98-percentiles of 45 year of 24-h acid gas data have been examined. As expected, the results worsen as the data sets become smaller. Taylor et al. (1986) further described the performance of methodological tools needed to develop statistical models. Ott (1990) presented a model of dilution experiments and suggested the theory of successive random dilutions. The experiments show that the lognormal distributional model represents the physical phenomena of pollutant dispersion.

Rumberg et al. (2001) extended the work of Cats and Holtslag (1980), to derive the relation between probability distribution function and SO2 data sets, using 2-parameter lognormal distribution model. It has been found that 3-parameter lognormal distribution model better represents the PM2.5 and PM8.0 concentration data. Chung and Cheng Fang (2002) estimated the frequency distribution of PM10 and PM2.5 by the statistics of wind speed at Sha-Lu, Taiwan. In this study, three theoretical distributions namely, lognormal, Weibull, and type V Pearson have been tested to fit the measured data on PM2.5, PM10 and wind speed. The results show that lognormal distributional model performs best. Hadley and Taumi (2002) carried out an assessment study to evaluate the changes in SO2 concentrations with probability distribution functions using the lognormal model. This study covers 10 monitoring sites and time periods up to 40 years. Recently, Gokhale and Khare (2003) reviewed the common methodologies to carry out statistical distribution modelling.

3.5. Hybrid models

Hybrid models combine features of both deterministic and statistical distribution models. A hybrid deterministic statistical methodology for predicting the distribution of air pollutant concentrations has been described by Jakeman et al. (1988). Simpson et al. (1983) introduced a modelling methodology to predict maximum total suspended particulates (TSP) concentrations. The methodology consists of using the ATDL model of Gifford and Hanna (1971) and the Larsen model (1971) to predict the distribution of TSP from wind speed data for the upper percentiles. The results agree with the work done by Daly and Steele (1976) for CO. This methodology works quite well for time average of 8-h or more but, in its present form, is questionable for shorter time averages. Jakeman and Taylor (1985) refined the hybrid urban model of Simpson et al. (1984) making it more flexible in application and providing a methodology for quantifying the effects of uncertainty associated with model predictions. Using Gamma distribution as the statistical component of the hybrid model, it has been found that at least 85% of the predictions of the 98-percentiles concentrations fall within 95% confidence interval. Taylor et al. (1985) applied a hybrid modelling methodology to the problems of dispersion of CO from line sources at Melbourne, Australia. The model has combined a deterministic component, the GM model (Chock, 1978) with a statistical component—the 2-parameter Weibull distribution, to produce estimates of the entire distribution of pollutant concentrations. The model predicts the pollutant concentration with
Table 1
Comparative description of vehicular pollution models

<table>
<thead>
<tr>
<th>Models</th>
<th>Pollutants and parameters</th>
<th>Applicability</th>
<th>Prediction scope</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Deterministic models</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CALINE</td>
<td>CO, NOx, SPM</td>
<td>Roadway pollution</td>
<td>Tend to over predict for parallel winds. No treatment of plume rise due to heated exhausts</td>
<td>Beaton et al. (1972)</td>
</tr>
<tr>
<td>CALINE-2</td>
<td>CO, NOx, SPM</td>
<td>Roadway pollution</td>
<td>Predicts poorly for unstable and neutral stability conditions. Over predicts for parallel winds and under predicts for oblique winds</td>
<td>Ward et al. (1977)</td>
</tr>
<tr>
<td>CALINE-3</td>
<td>CO, NOx, SPM</td>
<td>Roadway pollution</td>
<td>Over predicts for parallel winds and do not consider mechanical and thermal turbulence in dispersion coefficients</td>
<td>Benson (1979)</td>
</tr>
<tr>
<td>CALINE-4</td>
<td>CO, NOx, Aerosol</td>
<td>Roadway pollution</td>
<td>Over predicts for parallel winds</td>
<td>Benson (1992)</td>
</tr>
<tr>
<td>HIWAY-1</td>
<td>CO</td>
<td>Roadway pollution</td>
<td>Predicts poorly for low winds, over prediction for stable conditions and parallel winds and do not consider plume rise due to heated exhausts</td>
<td>Zimmerman and Thompson (1975)</td>
</tr>
<tr>
<td>HIWAY-3</td>
<td>CO</td>
<td>Roadway pollution</td>
<td>Predicts poorly for low winds, over predicts for parallel winds and no treatment for plume rise</td>
<td>Rao et al. (1980)</td>
</tr>
<tr>
<td>HIWAY-4</td>
<td>CO</td>
<td>Roadway pollution</td>
<td>Predicts poorly for low winds, over predicts for parallel winds and no treatment for plume rise</td>
<td>Rao et al. (1980)</td>
</tr>
<tr>
<td>GM</td>
<td>CO</td>
<td>Roadway pollution</td>
<td>Predicts poorly for low winds and over predict for parallel winds</td>
<td>Chock (1978)</td>
</tr>
<tr>
<td>ISCST-2</td>
<td>CO, NOx, SPM</td>
<td>Roadway pollution</td>
<td>Tendency to predict high for parallel wind conditions. No treatment for turbulence caused by heated exhausts</td>
<td>US EPA (1992)</td>
</tr>
<tr>
<td>DFLSM</td>
<td>CO</td>
<td>Roadside pollution for heterogeneous traffic condition</td>
<td>Predicts poorly for low winds</td>
<td>Khare and Sharma (1999)</td>
</tr>
<tr>
<td><strong>Statistical models</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARIMAX</td>
<td>Traffic densities, wind speed, and inversion heights</td>
<td>Traffic pollution</td>
<td>Prediction accuracy is medium</td>
<td>Tiao et al. (1975)</td>
</tr>
<tr>
<td>Regression</td>
<td>CO, length of daylight, solar radiation, inversion height and other met parameters</td>
<td>Traffic pollution</td>
<td>Medium to high prediction accuracy</td>
<td>Aron and Aron (1978); Karim and Matsui (1998); Comrie (1997); Comrie and Diem (1999)</td>
</tr>
<tr>
<td>Linear stochastic</td>
<td>Antecedent CO concentration</td>
<td>Traffic pollution</td>
<td>Low prediction accuracy</td>
<td>MacCollister and Willson (1975); Sharma and Khare (2000a,b).</td>
</tr>
</tbody>
</table>
Air pollution episodes are associated with sudden occurrences of high concentration of pollutants. Four hybrid models have been developed by Taylor et al. (1985), which predict ambient SO2 concentration accuracy better than a factor of two, over all percentiles of the concentration distribution. Four hybrid models have been developed by Taylor et al. (1985), which predict ambient SO2 concentration accuracy better than a factor of two, over all percentiles of the concentration distribution. Four hybrid models have been developed by Taylor et al. (1985), which predict ambient SO2 concentration accuracy better than a factor of two, over all percentiles of the concentration distribution. Four hybrid models have been developed by Taylor et al. (1985), which predict ambient SO2 concentration accuracy better than a factor of two, over all percentiles of the concentration distribution. Four hybrid models have been developed by Taylor et al. (1985), which predict ambient SO2 concentration accuracy better than a factor of two, over all percentiles of the concentration distribution.
are generally governed by the local meteorology and the dispersion mechanism. Urban road junctions and signalized intersections are more prone to these ‘episodic exceedances of pollutants’. As a result, the receptors in these ‘corridors’ are vulnerable to high pollutant concentration exposure. Following this, there is a need of an Episodic-Urban Air Quality Management Plan (e-UAQMP), targeting specifically to predict such exceedances and thus prevent catastrophic damages to the environment. Fig. 1 shows the e-UAQMP, based on the framework of Longhurst et al. (1996). From the figure, it is evident that the effectiveness of e-UAQMP depends as to how accurately the ‘extremes’ and the ‘probability of exceedances’ of pollutant concentrations are predicted. In such scenarios, the hybrid models that use both emission inventory (deterministic modelling component) and pollutant concentration data (statistical modelling component) predict the entire range of concentration distribution in form of probability. As a result, hybrid model predictions become more reliable in framing an e-UAQMP. Subsequently, such an e-UAQMP may assist the decision makers in designing an air pollution alert system and provide sensitive elements of the population with alternatives to take protective measures.

5. Conclusions

Deterministic models are ‘causal’ in nature and do not account for the stochasticity involved in air pollutant concentrations. Moreover, they are based on limited inputs and their prediction capability depends on the conditions fulfilling the simplifying assumptions, used in the model formulation. As a result, they do not predict the ‘entire’ distribution of concentration. The statistical models that are ‘non-causal’ in nature establish the semi-empirical relationship between available data and measurements and determine the historical trends. However, they do not take into account the emission inventory and meteorology. The hybrid model, on the other hand, effectively addresses the above limitations inherent in the deterministic and the statistical models, respectively. The model uses the ‘output’ of the deterministic model as its input (thus taking into account the influence of ‘emission’ and ‘meteorology’).
and then identifies and parameterizes the most suitable statistical distribution model (taking into account the historical statistical trend of the data). In a nutshell, it, thus combines the best attributes of the ‘deterministic’ and the ‘stochastic’ based models, which ultimately results in enhanced prediction accuracy, when compared to other prediction models.

References


Gokhale, S.B., Patil, R.S., 2004. Size distribution of aerosols (PM10) and lead (Pb) near traffic intersections in Mumbai. Environ. Monitoring Assess. 95 (1-3), 311-324.


Kirsh, J.W., Mason, B.F., 1975. Mathematical models for air pollution studies involving the oregano 11205 highway project systems, science and software Report, SSS-R-76-2744, La Jolla, CA.


Larsen, R.I., 1971. A mathematical model for relating air quality measurements to air quality standards. Publication No. AP89, USEPA, Research Triangle park, NC, USA.


Millanis, A.E., Davies, T.D., 1994b. Regression and stochastic models for air pollution-II: Application of stochastic models to examine the links between ground-level smoke concentrations and temperature inversions. Atmos. Environ. 28 (17), 2811-2822.


Schattanek, G., Kahng, J., Stratou, T., Braverman, T.N., 1990. EPA User's guide to CAL3QHC A modelling methodology for predicting pollutant concentrations near roadway intersections, USEPA.


Sharma, P., Khare, M., 2000a. Real-time prediction of extreme ambient carbon monoxide concentrations due to vehicular exhaust
Young, P.C., Ng, C.N., Lane, K., Parker, D., 1991. Recursive learning control of CO concentration. IEEE, 261-266.