Environ Monit Assess (2009) 157:105–112
DOI 10.1007/s10661-008-0520-2

Forecasts using Box–Jenkins models for the ambient air quality data of Delhi City

Pragati Sharma · Avinash Chandra · S. C. Kaushik

Received: 4 April 2008 / Accepted: 14 August 2008 / Published online: 26 September 2008
© Springer Science + Business Media B.V. 2008

Abstract The monthly maximum of the 24-h average time-series data of ambient air quality—sulphur dioxide (SO_2), nitrogen dioxide (NO_2) and suspended particulate matter (SPM) concentration monitored at the six National Ambient Air Quality Monitoring (NAAQM) stations in Delhi, was analysed using Box–Jenkins modelling approach (Box et al. 1994). Univariate linear stochastic models were developed to examine the degree of prediction possible for situations where only the past record of pollutant data are available. In all, 18 models were developed, three for each station for each of the respective pollutant. The model evaluation statistics suggest that considerably satisfactory real-time forecasts of pollution concentrations can be generated using the Box–Jenkins approach. The developed models can be used to provide short-term, real-time forecasts of extreme air pollution concentrations for the Air Quality Control Region (AQCR) of Delhi City, India.

Keywords Box–Jenkins models · Linear stochastic models · Time-series analysis · Real-time forecasting

P. Sharma (✉) · A. Chandra · S. C. Kaushik
Centre for Energy Studies, Indian Institute of Technology Delhi, Hauz Khas, New Delhi 110016, India
e-mail: pragatimails@indiatimes.com

A. Chandra
e-mail: chandra@ces.iitd.ac.in

S. C. Kaushik
e-mail: sckaushik@ces.iitd.ac.in

Introduction

The deteriorating air quality is primarily attributed to rise in motor vehicle population and industrialisation and resulting exhaust emissions in the urban regions. To implement air quality management and public warning strategies for pollutant levels, reasonably accurate forecasts of the atmospheric concentration of pollutants as function of space and time are necessary. This can be done by air pollution models. The air quality "predictor" for air pollution can be developed either by analytical or by statistical means. Analytical models are, in general, more suitable for making long-term, forecasts/planning decisions (Juda 1989; Zannetti 1989). For air pollution "episodes" characterised typically by fast dynamics, these models do not give satisfactory results (Cats and Holtslag 1980; Nieuwstadt 1980; Jakeman et al. 1988; Raimondi et al. 1997). Moreover, in the absence of additional parameters required as input, such as, wind vector, temperature, traffic characteristics (for emission factor computations), the analytical models fail to provide quantitative description of the atmospheric pollution (e.g. Chock 1978; Benson 1979, 1989; Petersen 1980, etc.). Stochastic modelling of the pollution time-series provides an alternative approach.

In the present study, an attempt has been made to determine the extent of prediction possible using data set restricted only to the past record of air quality data—SO_2, NO_2 and SPM concentration time series. For this purpose, univariate linear stochastic models based on the Box–Jenkins modelling techniques have been developed for the six (NAAQM) stations run by the Central Pollution Control Board (CPCB). The models can be utilised for supplying real-time forecasts of extreme pollutant concentrations, predicting future concentration levels on the basis of data recorded in previous periods.

Site and data description

The NAAQM network was started by the CPCB during 1984–1985 at the National level. Within the purview of this, six NAAQM stations had been established in various parts of Delhi with the basic idea to find the current status of air quality, its seasonal variations with increased industrialisation and urbanisation and to understand increase in various air pollution generation activities. The location details and description of surrounding area around various NAAQM stations are given in Table 1.

The raw, daily average pollutant concentration data, from September 1, 1987 to December 31, 2005, was collected from the NAAQM stations. However, lot of missing values were encountered for years 1987 and 1988. Prerequisite for time series analysis is to have continuous time-series data without missing values. Thus, reading for the first two years were not included in the data analysis. The 24-h monthly maximum time-series for the three pollutants was formed from the raw data. This data set was further divided into two sample groups viz. (1) "the development sample" from January, 1989 to December, 2003; and "the test sample" from January, 2004 to December, 2005. As the name suggests, the former set was used in the development of the model, while the latter for testing various models, thereby treating it (the test sample) as unobserved data set in order to compare it with the predictions made by these models. The division was done for achieving two modelling objectives as suggested by Benarie (1980), namely: (1) representation of observed data; (2) prediction that is effective for other as yet unobserved samples. The model development sample data was further standardised, so that each variable varies in the same scale, by subtracting the mean values of each series from each observation of the respective series and then dividing the result by the standard deviation of that series.

Model formulation

Theory

The Box–Jenkins modelling approach for univariate models, more commonly known as ARIMA (autoregressive integrated moving average) analysis, consists of extracting the predictable movements, trends and serial correlations from the observed data until a sequence of *white noise* (or shocks) remains. This is done by decomposing the time series into several components via the *autoregressive* (*AR*), the *integration* (*I*; difference) and the *moving average* (*MA*) operators, sometimes called *filters* (linear). Thus, for instance, to remove the trends or the non-stationarity in the time

Table 1 Location details and description of NAAQM stations

Monitoring station	Location	Type of area	Parameters monitored
Ashok Vihar	North-north west	Residential	SPM, NO_2, SO_2
Janak Puri	South-west	Residential	SPM, NO_2, SO_2
Nizamuddin	South-east	Mixed use	SPM, NO_2, SO_2
Shahadra	North-east	Industrial	SPM, NO_2, SO_2
Shahzada Bagh	North-west	Industrial	SPM, NO_2, SO_2
Sirifort	South	Residential	CO, SPM, NO_2, SO_2

series z_t (defined as a sequence of N observations equidistant in time, such as $z_1, z_2, \ldots, z_N$), the difference operators, also called the integration operators, '∇' and '∇_s', defined as

$$\nabla z_t = z_t - z_{t-1} \qquad \text{and} \qquad \nabla_s z_t = z_t - z_{t-s} \tag{1}$$

where, s denotes the period or the span, i.e. the length of the seasonal cycle, may be applied d and D times respectively (d and D being the order of regular and seasonal differencing respectively). To remove serial correlations in the series, two operators—autoregressive (AR) and moving average (MA)—can be applied. They are expressed as polynomials of the *backward shift operators, B* and B^s, defined as

$$Bz_t = z_{t-1} \quad \text{and} \quad B^s z_t = z_{t-1} \tag{2}$$

The regular and seasonal AR operators $\phi_p\,(B)$ and $\Phi_P\,(B^s)$, are respectively polynomials of order p in B and P in B^s, such that

$$\phi_p\,(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p \tag{3a}$$

and

$$\Phi_P\left(B^s\right) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \ldots - \Phi_P B^{Ps} \tag{3b}$$

They are used to express the current values of the series as a finite sum of the past values of the series. The regular and seasonal MA operators are similarly defined as:

$$\theta_q\,(B) = 1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q \tag{4a}$$

and

$$\Theta_Q\left(B^s\right) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \ldots - \Theta_Q B^{Qs} \tag{4b}$$

The MA operators express the current values of the data as a finite sum of the current and past values of the *shock* (*or random noise*), a_t.

Stepwise model building process

The Box–Jenkins method provides a unified approach for *identifying* which filter(s) are most appropriate for the series being analysed, for *estimating* the parameters describing the filters (i.e. for estimating the grid sizes of the series), for *diagnosing* the accuracy and reliability of the models that have been estimated, and finally for *forecasting*. The selection of the most appropriate model is a step-by-step procedure. The preliminary step in any time series analysis is to plot the observations against time. The plot is often a valuable part of any data analysis, since qualitative features such as trend, seasonality, discontinuities and outliers will usually be present in the data. Most of the probability theory of time series is concerned with stationary time series, and for this reason the analysis often requires to turn a non-stationary series into a stationary one in order to use this theory. Plotting the data may suggest the transformation necessary to make the series stationary. The transformations usually done for stabilising the variance in the series are "logarithmic", "square root" or "power transformations" (Mills 1991). To remove the non-stationarity caused by trend and seasonality, Box and Jenkins (1970) advocate regular and seasonal differencing transformation respectively. Once the series is made stationary by proper transformation, following iterative model building process is followed.

The *first step* involves the selection of a general class of models using the autocorrelation function (ACF) and the partial autocorrelation function (PACF). The ACF measures the amount of linear dependence between observations in a time series that are separated by lag k. The PACF plot helps to determine how many AR terms are necessary for the model. A tentative model may be specified, based on the shape of the ACF and PACF and a set of rules (Box et al. 1994; *second step*). The concept of *model parsimony* is followed, for selecting a model, i.e. a model with the smallest possible number of parameters is preferable. The general class of univariate Box–Jenkins models, denoted by ARIMA (p, d, q) × (P, D, Q)$_s$ describing the current term z_t of a time series by its own past values is obtained by combining the operators defined in Eqs. 2 to 4a and 4b as

$$\phi_p\,(B)\,\Phi_P\left(B^s\right)\nabla_s^D \nabla^d x_t = \theta_q\,(B)\,\Theta\left(B^s\right) a_t + c \tag{5}$$

The *third step* involves the estimation of the model parameters. This is done by means of an iterative non-linear least-square algorithm (Marquardt 1963) The statistical adequacy of the model is checked by performing diagnostic tests (*fourth step*). This is usually done by two criteria: primarily by the examination of the residual series for interdependence and a Chi-square statistics test based on the first twenty lagged residual autocorrelations, the Portmonteau goodness-of-fit-test (Box et al. 1994). The former is accompanied by correlation analysis through the residual ACF plots. If the residuals are not correlated, then they should be *white noise*. If any of these two criteria are not valid, the model should be refined (step 2) and re-estimation of the parameters (step 3) should be carried out. After the procedure has been applied for a given time series, a calibrated model is obtained that has encoded the basic statistical properties of the time series into some models parameters. Therefore, by taking conditional expectations at time 't' of each term of the ARIMA model in Eq. 5 and writing $w_t = \nabla^d \nabla_S^D z_t$ the minimum mean square error (MMSE) for forecasts are:

$$
\begin{aligned}
\left[w_{t+l}\right] = {} & \phi_1 \left[w_{t+l-1}\right] + \phi_2 \left[w_{t+l-2}\right] + \cdot\cdot \\
& + \phi_p \left[w_{t+l-p}\right] + \left[a_{t+l}\right] - \theta_1 \left[a_{t+l-1}\right] \\
& - \theta_2 \left[a_{t+l-2}\right] - \ldots - \theta_q \left[a_{t+l-q}\right] + c
\end{aligned} \tag{6a}
$$

and the forecasts for seasonal ARIMA model are

$$
\begin{aligned}
\left[w_{t+l}\right] = {} & \phi'_1 \left[w_{t+l-1}\right] + \phi'_2 \left[w_{t+l-2}\right] + \cdot\cdot \\
& + \phi'_{p+sP} \left[w_{t+l-p-sP}\right] + \left[a_{t+l}\right] \\
& - \theta'_1 \left[a_{t+l-1}\right] - \theta'_2 \left[a_{t+l-2}\right] - \ldots \\
& - \theta'_{q+sQ} \left[a_{t+l-q-sQ}\right] + c
\end{aligned} \tag{6b}
$$

where $l = 1, 2, \ldots$ is the lead time for the forecasts w_{t+l}; $\phi'_1, \phi'_2, \ldots$ are the generalised AR parameters defined by

$$\phi'(B) = \phi_p(B)\,\Phi_P\left(B^s\right); \tag{7a}$$

and $\phi'_1, \phi'_2, \ldots$ are the generalised MA parameters defined by

$$\theta'(B) = \theta_q(B)\,\Theta_Q\left(B^s\right) \tag{7b}$$

Thus, the Box–Jenkins iterative approach for constructing linear time-series models can be summarised in four steps:

(a) *Identification* of preliminary specifications of the model;
(b) *Estimation* of parameters of the model;
(c) *Diagnostic* checking of model adequacy; and
(d) *Forecasting* further realisation

The developed models

The four standard stages of Box–Jenkins methodology, were followed in the formulation of univariate models for each of the three pollutants. The final estimates of the parameters with corresponding standard deviation and t-ratio, which is indicative of the statistical significance of model parameters ($|t - \text{ratio}| \geq 2.0$), for the time series ARIMA models developed for SO_2, NO_2 and SPM for each of the six NAAQM stations are presented in Table 2. It may be specified here that for the SO_2-, NO_2- and SPM-series (each using 180 observations for model formulation), the autocorrelation becomes *significantly different from zero* at 95% confidence level, when in absolute it exceeds the critical value which is approximately $0.149 (= |\frac{2}{\sqrt{n}}|)$. Portmanteau or the modified Box–Pierce i.e. Q-statistics was performed at different 12, 24 and 36 lags to check the adequacy of the various formulated models. Both the tests indicate that the residuals can be considered as white noise, indicating adequate model fit. Further, stationarity, invertibility tests and the statistical significance of the model parameters were examined and were found to satisfy all model fit requirements. Finally, metadiagnosis showed that the models given in Table 2 are the best models.

Model performance

The forecast performance of the developed models has been evaluated by statistical means using "the test sample" data. There are several ways of judging forecasts, such as plotting measured and predicted value sequences and then making a visual inspection. This is the simplest method to

Table 2 Final estimates of parameters of the time series ARIMA models

Pollutant	Model	Parameter	Estimate	Standard deviation	t-ratio
Developed for the Ashok Vihar monitoring station					
SO_2	ARIMA (0, 1, 1)	θ_1	0.69	0.06	11.45
NO_2	ARIMA (0, 1, 1)	θ_1	0.76	0.06	11.98
SPM	ARIMA (1, 0, 1)	ϕ_1	0.318	0.14	2.22
		θ_1	−0.06	0.14	0.39
Developed for the Janak Puri monitoring station					
SO_2	ARIMA (0, 1, 1)	θ_1	0.75	0.05	15.28
NO_2	ARIMA (0, 1, 1)	θ_1	0.76	0.05	14.43
SPM	ARIMA (1, 0, 1)	ϕ_1	0.35	0.20	1.73
		θ_1	0.10	0.21	0.45
Developed for the Nizamuddin monitoring station					
SO_2	ARIMA (0, 1, 2)	θ_1	0.48	0.08	6.27
		θ_2	0.24	0.08	3.02
NO_2	ARIMA (2, 1, 1)	ϕ_1	0.31	0.08	4.06
		ϕ_2	0.16	0.07	2.19
		θ_1	0.93	0.03	34.56
SPM	ARIMA (1, 0, 1)	ϕ_1	0.27	0.14	1.89
		θ_1	−0.15	0.14	1.09
Developed for the Shahadra monitoring station					
SO_2	ARIMA (0, 1, 1)	θ_1	0.76	0.10	7.78
NO_2	ARIMA (0, 1, 1)	ϕ_1	0.41	0.10	4.01
		θ_1	0.86	0.06	14.32
SPM	ARIMA (1, 0, 1)	ϕ_1	0.39	0.16	2.44
		θ_1	0.08	0.16	0.48
Developed for the Shahzada Bagh monitoring station					
SO_2	ARIMA (1, 1, 1)	ϕ_1	−0.20	0.13	1.49
		θ_1	0.37	0.13	2.84
NO_2	ARIMA (0, 1, 1)	θ_1	0.75	0.09	8.36
SPM	ARIMA (0, 0, 3)	θ_1	0.44	0.06	6.84
		θ_2	0.12	0.09	1.38
		θ_3	0.34	0.08	4.35
Developed for the Sirifort monitoring station					
SO_2	ARIMA (0, 1, 1)	θ_1	0.65	0.07	8.75
NO_2	ARIMA (0, 1, 1)	θ_1	0.71	0.07	10.70
SPM	ARIMA (1, 0, 1)	ϕ_1	0.14	0.14	0.97
		θ_1	−0.28	0.13	2.12

evaluate the model performance and gives a physical feel of the forecast, as the ranges(s) within which the model is performing satisfactorily can be directly observed from the graph. However, this analysis is not free from subjectivity. To introduce objectivity in the numerical error analysis, the performance should be judged by certain statistical evaluation indices. Thus, two indicators are used in researching the predictive skill of the models developed in the present study. The root mean square error (RMSE) and its decomposed components—systematic ($RMSE_s$) and unsystematic ($RMSE_u$)—is a very useful evaluation index as suggested by Willmott (1981) and Willmott et al. (1985). The RMSE is defined as follows.

$$RMSE = \sqrt{RMSE_s^2 + RMSE_u^2} \tag{8}$$

where

$$RMSE_s = \left[\frac{1}{n}\sum_{i=1}^{n}\left(\hat{P}_i - O_i\right)^2\right]^{1/2} \tag{9}$$

$$RMSE_u = \left[\frac{1}{n}\sum_{i=1}^{n}\left(\hat{P}_i - P_i\right)^2\right]^{1/2} \tag{10}$$

where O_i and P_i are the observed and predicted values, respectively and $\hat{P}_i = a + bO_i$, a and b being the slope and intercept of the least squares regression of forecast variable on observed variable. The simple relationship (Eq. 8) indicates a useful decomposition of the total error into

Table 3 Model evaluation statistics

Monitoring station	Model evaluation statistics			Pollutants		
				SO_2	NO_2	SPM
Ashok Vihar	D	Fitted		0.8930	0.9030	0.8980
		Forecast		0.9160	0.9180	0.8690
	R^2	Fitted		0.9060	0.9065	0.9681
		Forecast		0.9400	0.9380	0.9665
		$RMSE_s$	Fitted	0.4306	0.3238	0.7614
	RMSE		Forecast	0.1865	0.2761	0.7087
		$RMSE_u$	Fitted	0.2841	0.1921	0.1178
			Forecast	0.0788	0.1119	0.1097
Janak Puri	d	Fitted		0.9090	0.9050	0.8920
		Forecast		0.8810	0.8830	0.8680
	R^2	Fitted		0.9165	0.9073	0.9828
		Forecast		0.8764	0.8738	0.9857
		$RMSE_s$	Fitted	0.4568	0.4914	0.8892
	RMSE		Forecast	0.0393	0.1046	0.6946
		$RMSE_u$	Fitted	0.2696	0.3267	0.0668
			Forecast	0.0312	0.0882	0.0357
Nizamuddin	d	Fitted		0.9050	0.9060	0.8950
		Forecast		0.8890	0.9000	0.9050
	R^2	Fitted		0.9528	0.9375	0.9612
		Forecast		0.9372	0.9356	0.9672
	RMSE	$RMSE_s$	Fitted	0.5435	0.4393	0.7063
			Forecast	0.1014	0.6547	0.7027
		$RMSE_u$	Fitted	0.1426	0.1675	0.1344
			Forecast	0.0320	0.0236	0.1170
Shahadra	d	Fitted		0.8900	0.8990	0.8960
		Forecast		0.9190	0.9120	0.8700
	R^2	Fitted		0.8845	0.9537	0.9755
		Forecast		0.9348	0.9805	0.9768
		$RMSE_s$	Fitted	0.4370	0.6311	0.7565
	RMSE		Forecast	0.2399	0.2770	1.3253
		$RMSE_u$	Fitted	0.3477	0.1529	0.0861
			Forecast	0.1138	0.0278	0.1581
Shahzada Bagh	d	Fitted		0.9040	0.8990	0.9010
		Forecast		0.8890	0.8820	0.9140
	R^2	Fitted		0.9431	0.9004	0.9437
		Forecast		0.9300	0.8693	0.9715
	RMSE	$RMSE_s$	Fitted	0.2890	0.4029	0.4688
			Forecast	0.1001	0.3798	0.8646
		$RMSE_u$	Fitted	0.0952	0.2781	0.1476
			Forecast	0.0377	0.3390	0.1343
Sirifort	d	Fitted		0.9010	0.9060	0.9000
		Forecast		0.9020	0.8840	0.8720
	R^2	Fitted		0.9260	0.9221	0.9659
		Forecast		0.9373	0.9068	0.9568
		$RMSE_s$	Fitted	0.4229	0.4198	0.6923
	RMSE		Forecast	0.1518	0.3602	0.9196
		$RMSE_u$	Fitted	0.1926	0.2180	0.1182
			Forecast	0.0558	0.2595	0.1747

systematic and unsystematic elements. Also recommended by Willmott (1981) and Willmott et al. (1985) is the index of agreement (d), defined as

$$d = 1 - \frac{\sum_{i=1}^{n} (P_i - O_i)^2}{\sum_{i=1}^{n} \left(\left| P_i - \overline{O} \right| + \left| O_i - \overline{O} \right| \right)^2} \qquad 0 \leq d \leq 1 \tag{11}$$

where $\overline{O}$ being the mean of observed values. The index d determines the extent to which magnitudes and the signs of the observed values about the $\overline{O}$ are related to the predicted deviations about $\overline{O}$, and allows for sensitivity towards differences in O and P as well as proportionality changes (Rao et al. 1985). Being dimensionless and having the limits of 0.0 (indicating no agreement) and 1.0 (indicating perfect agreement), d may be viewed as *standardised* (by the variability in the predictions and observations about the observed mean) measure of the mean square error. The index d was proposed by Willmott (1981) as an alternative to R and R^2. This index is both a relative and bounded measure, while R and R^2 are not consistently related to the accuracy of prediction. Willmott and Wicks (1980) observed that the "high", or the statistically significant values of R and R^2 may in fact be misleading, as they often are unrelated to the size of the difference between O_i and P_i.

Table 3 presents the summary of the model evaluation statistics for the various models at all the NAAQM stations for the three pollutants i.e., SO_2, NO_2 and SPM. The index of agreement (d) for the forecasts of the various models varies from 0.8693 to 0.9857, which is a very satisfactory forecast. Thus, most of the models are able to explain equal to or more than 86.93% of the potential for error, i.e., at least 86.93% of the predictions are error free.

Conclusions

The results obtained in the study reveal that linear stochastic models such as ARIMA models provide a useful quantitative description of air quality. The Box–Jenkins models, though follow a "black-box" approach, the system characteristics are intrinsically represented by the data themselves. The technique is simple and requires less computational effort to provide the forecast. The only prerequisite is the availability of sufficiently long historical data set for model formulation (to allow reliable empirical identification of the character of the data generation process); if possible, at least 50 and preferably 100 successive observations should be used (Box et al. 1994). The methodology is particularly effective for temporally autocorrelated data, such as time-series of the air pollution concentration series. However, the stochastic models being site specific, need to be used with care and cannot answer the "what-if" questions. They can be used as forecasting tool, for they provide a better estimate of air quality than the analytical models. Thus, separate models should, ideally be developed for different AQCR, hosting monitoring stations. The forecasting accuracy of stochastic models decreases with time, Therefore, a rapid availability of the pollutant data set is desirable to achieve a sufficiently accurate forecast.

To develop an efficient public warning strategy for pollutant levels, accurate predictors of air quality are required. The results obtained in the study are quite satisfactory and warrant continued work in the area of time-series modelling.

Acknowledgements The authors are grateful to the Central Pollution Control Board (CPCB), New Delhi, for providing the necessary air quality data for the AQCR.

References

Benarie, M. M. (1980). *Urban air pollution modelling*. London: The MIT Press.

Benson, P. E. (1979). CALINE 3 A versatile dispersion model for predicting air pollutant levels near highways and arterial roads. Final report. FHWA/CA/TL-79/23. California Department of Transportation, Sacramento, CA.

Benson, P. E. (1989). CALINE 4—a dispersion model for predicting air pollution concentrations near roadways. Final report. FHWA/CA/TL-84/15. California Department of Transportation, Sacramento, CA.

Box, G. E. P., & Jenkins, G. M. (1970). *Time series analysis: Forecasting and control* (2nd edn.). San Francisco: Holden Day, 1976.

Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (1994). *Time series analysis: Forecasting and control* (3rd edn.). Englewood Cliffs, NJ: Prentice Hall.

Cats, G. J., & Holtslag, A. A. M. (1980). Prediction of air pollution frequency distribution—Part 1. The lognormal distribution. *Atmospheric Environment, 14*, 255–258. doi:10.1016/0004-6981(80)90285-1.

Chock, D. P. (1978). A simple line-source model for dispersion near roadways. *Atmospheric Environment, 12*, 823–829. doi:10.1016/0004-6981(78)90019-7.

Jakeman, A. J., Simpson, R. W., & Taylor, J. A. (1988). Modelling distributions of air pollutant concentrations - III. The hybrid deterministic-statistical distribution approach. *Atmospheric Environment, 22*, 163–174. doi:10.1016/0004-6981(88)90309-5.

Juda, K. (1989). Air pollution modelling. In P. N. Cheremisinoff (Ed.), *Encyclopedia of environmental control technology, air pollution control* (Vol. 2, pp. 83–134). Houston, TX, USA: Gulf Publishing Company.

Marquardt, D. W. (1963). An algorithm for least square estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics, 11*, 431–441. doi:10.1137/0111030.

Mills, T. C. (1991). *Time series techniques for economists*. Cambridge: Cambridge University Press.

Nieuwstadt, F. T. M. (1980). Prediction of air pollution frequency distribution - II. The Gaussian plume model. *Atmospheric Environment, 14*, 259–265. doi:10.1016/0004–6981(80)90286–3.

Petersen, W. B. (1980). User's Guide for HIWAY-2, highway air pollution model, 3–17, EPA-600/8–80–018. Environmental Science Research Laboratory, Office of Research and Development US Environmental Protection Agency, Research Tringle Park, NC, 27711.

Raimondi, F. M., Rando, F., Vitale, M. C., & Calcara, A. M. V. (1997). Short-time fuzzy DAP predictor for air pollution due to vehicular traffic. In R. S. José & C. A. Brebbia (Eds.), *Measurements and Modelling in Environmental Pollution. First International Conference on Environmental Pollution MMMEP 97* (pp. 189–200). Southampton, Boston: Computational Mechanics Publications.

Rao, S. T., Sistla, G., Petersen, W. B., Irwin, J. S., & Turner, D. B. (1985). Evaluation of the performance of RAM with the regional air pollution study data base. *Atmospheric Environment, 19*, 229–245. doi:10.1016/0004-6981(85)90091-5.

Willmott, C. J. (1981). On the validation of models. *Physical Geography, 2*, 184–194.

Willmott, C. J., Ackleson, S. G., Davis, R. E., Feddema, J. J., Klink, K. M., Legates, D. R., et al. (1985). Statistics for the evaluation and comparison of models. *Journal of Geophysical Research, 90*, 8995–9005. doi:10.1029/JC090iC05p08995.

Willmott, C. J., & Wicks, D. E. (1980). An empirical method for the spatial interpolation of monthly precipitation within California. *Physical Geography, 1*, 59–73.

Zannetti, P. (1989). Simulating short-term, short-range air quality dispersion phenomena. In P. N. Cheremisinoff (Ed.), *Encyclopedia of environmental control technology, air pollution control* (Vol. 2, pp. 159–195). Houston, TX, USA: Gulf Publishing Company.