Pergamon

Atmospheric Environment Vol. 31, No. 24, pp. 4081-4094, 1997

1352-2310/97 $17.00 + 0.00

PII: S1352-2310(97)00282-3

REGRESSION MODELLING OF HOURLY NO_x AND NO_2 CONCENTRATIONS IN URBAN AIR IN LONDON

JI PING SHI and ROY M. HARRISON*

Institute of Public and Environmental Health, School of Chemistry, The University of Birmingham, Edgbaston, Birmingham B15 2TT, U.K.

(*First received* 13 *March* 1997 *and in final form* 17 *June* 1997. *Published September* 1997)

Abstract—Based on hourly measurements of NO_x, NO_2 and O_3 and meteorological data, an ordinary least squares (OLS) model and a first-order autocorrelation (AR) model were developed to analyse the regression and prediction of NO_x and NO_2 concentrations in London. Primary emissions and wind speed are the most important factors influencing NO_x concentrations; in addition to these two, reaction of NO with O_3 is also a major factor influencing NO_2 concentrations. The AR model resulted in high correlation coefficients ($R > 0.95$) for the NO_x and NO_2 regression based on a whole year's data, and is capable of predicting NO_2 ($R = 0.83$) and NO_x ($R = 0.65$) concentrations when the explanatory variables were available. The analysis of the structure of regression models by Principal Component Analysis (PCA) indicates that the regression models are stable. The results of the OLS model indicate that there was an exceptional NO_2 source, other than primary emission and reaction of NO with O_3, in the air pollution episode in London in December 1991. © 1997 Elsevier Science Ltd.

Key word index: Regression model, nitrogen dioxide, NO_x, pollution episode.

INTRODUCTION

Nitrogen dioxide (NO_2) is one of the most important air pollutants. Vehicle exhaust and other combustion emissions are the main source of nitrogen oxides NO_x ($NO_x = NO + NO_2$) in urban air. Most primary NO_x is emitted in the form of nitric oxide (NO), but most is ultimately converted to NO_2 by reaction with ozone (O_3) in the atmosphere. Both for air quality forecasting and for development of control strategies it is important to identify the factors controlling NO_x and NO_2 concentrations and to develop a function, F, which would allow the prediction of NO_2 concentration $C_{NO_2}(x, t)$ at any point in space x, and time t. Two different approaches have been adopted to identify F (Milionis and Davies, 1994): (a) atmospheric diffusion models (e.g. Hanna *et al.*, 1982; Pasquill and Smith, 1983; Panofsky and Dutton, 1984), and (b) regression models (e.g. Chock *et al.*, 1975; Wolff and Lioy, 1978; Revlett, 1978; Zanetti, 1990). The first approach is the most logical and obvious way, but it depends upon detailed (hourly) NO_x emission data distributed over the Greater London area, which is not available currently. Thus, a regression model has been developed and is described in this paper to predict hourly NO_x and NO_2 concentrations in urban air in central London.

Regression models are generally site dependent. Moreover, there is no guarantee as to the reliability of the model once it is extrapolated beyond the range of the input data used to construct it (Chock *et al.*, 1975). However, the Greater London area is quite large and air quality data analysis shows that very similar NO_x and NO_2 concentrations are measured at the central London (CLL) and West London (WL) sites (Shi, 1996) which are approximately 4 km apart. Therefore, the prediction will be representative, at least, of the central London area. Inoue *et al.* (1986a, b) have carried out regression analysis and prediction of NO_x based on data measured in a Japanese city. They used traffic, weather and some complex composite variables as explanatory variables. Regression parameters based on monthly data were obtained but no NO_2 regressions were conducted. In our study Ordinary Least Squares (OLS) and first-order autocorrelation (AR) models based on whole year's data are used to carry out regression analyses of both NO_x and NO_2. A Principal Component Analysis (PCA) method is used to characterise the structures of the regression models.

MODEL INPUT DATA AND EXPLANATORY VARIABLES

NO_x *and* NO_2 *data*

Hourly measurements of NO_x and NO_2 over two-years (June 1989–May 1990, June 1991–May 1992)

*Author to whom correspondence should be addressed.

Table 1. Details of sampling sites

Abbr.	Site Name	Grid ref.	Height (m)	Information
STE	Stevenage	TL237225	8	On the edge of a residential new town near a light industrial estate. 100 m east of A1(M) motorway. Topography flat.
SIB	Sibton	TM364719	6	Open flat cereal farmland. Woodland to the north west.
HAR	Harwell	SU474863	8	Site adjacent to AEA Harwell research laboratories which is surrounded by flat cereal fields. The busy A34 is nearby.
LH	Lullington Heath	TQ538016	120[a]	On a high plateau 5 km from the south coast. Immediate area is a NCC heathland.
CLL	Central London Laboratory	TQ292791	13	Situated facing a quite backstreet within a busy city centre located in Victoria station
BRI	Bridge Place	TQ289788	8	Second floor office overlooking backstreet near Victoria Station, central London.

[a] 120 m above sea level.

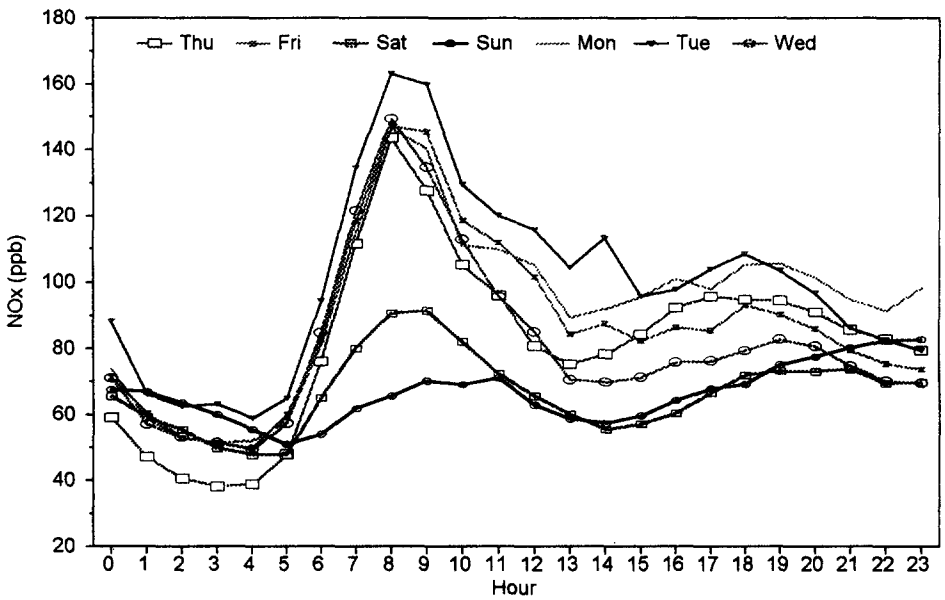

Fig. 1. Average diurnal variation of NO_x at CLL site from June 1989 to May 1990.

at central London monitoring sites (CLL/BRI, see Table 1) were used in the model. Diurnal variations of NO_x for each day of the week are shown in Fig. 1. This was derived by averaging hourly NO_x data on Sunday, Monday... Saturday, respectively, from June 1989 to May 1990. The NO_x data, given at Greenwich Mean Time (GMT), was corrected to British Summer Time (BST) during summer. Clearly, there were lower NO_x concentrations on Sunday and Saturday. Weekdays had similar profiles with a traffic related peak at 8:00 am.

O_3 *Data*

Hourly average ozone (O_3) concentrations were also used. Details of the ozone sampling sites are given in Table 1. The monitoring sites STE, SIB, LH and HAR (cf. Table 1) are distributed in four different directions outside London, shown in Fig. 2.

Meteorological data

Hourly meteorological data, measured (GMT) at London Weather Centre (Latitude 51.52N, Longitude 0.12W, Altitude 77 m) during June 1989–May 1990 and June 1991–May 1992 were obtained from the Meteorological Office. The utilised data were as follows:

- Wind Speed (WS): metres per second (m/s),
- Wind Direction (WD): tens of degrees from true North,
- Air Temperature (Temp): dry bulb temperature (tenths of a degree Celsius),
- Relative Humidity (rh): percent (%),

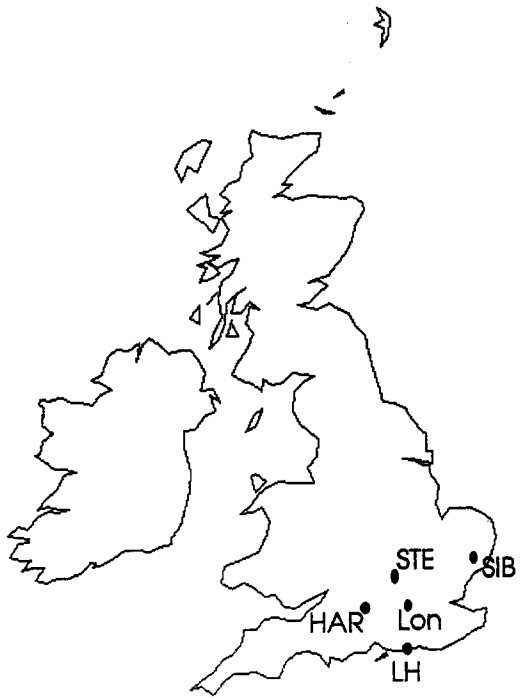

Fig. 2. Locations of ozone monitoring sites.

- Total Radiation (TotRad): Watts per square metre (W/m^2),
- Pasquill Stability Category (Stab): 1–7 during 1991–1992 (1—very unstable, 7—very stable); 1–9 during 1989–1990 (1—very unstable, 9—very stable),
- Boundary Layer Depth (BLD): metres,

Explanatory variables

Previous investigators have generally found that wind speed (WS) and mixing height (BLD) are the only readily observed meteorological parameters with a direct effect on observed concentration C. Increase of wind speed, generally speaking, increases dilution and dissipation of the pollutants. Reduction of mixing height tends to prevent upward mixing, and would be expected to cause increase of concentration at the ground (Annand and Hudson, 1981). Other meteorological parameters may have an indirect influence on ground-level NO_x concentrations.

A major problem in developing the model is the lack of information on diurnal and day-to-day variations in NO_x source strength. When air quality data are averaged over an entire year (see Fig. 1), it is clear that real systematic differences exist both between hours of the day and between weekdays and weekend which must relate to source strength, as well as meteorology. This source strength variation must be reflected in the model, and in the present work measured NO_x concentrations and a simple box model were used to estimate the NO_x emission rate. The box model was of the form:

$$Q/d = C_{NO_x} \times \mathrm{BLD} \times \mathrm{WS}$$

where Q is the hourly NO_x emission rate, d is the linear cross-wind dimension of the city. Because Greater London is approximately circular, d is assumed the same for all wind directions. This box model was used in an inverse of the manner as used by Derwent *et al.* (1995 and references therein), to calculate the source strength from measurements of instantaneous concentrations. This was carried out for each day of the week and for each hour of the day. Diurnal variations of Q/d for each day of the week from June 1989 to May 1990 were averaged and are shown in Fig. 3. A relative emission factor, q which is by definition proportional to Q/d, and has a value arbitrarily set as $q = 1$ at 5:00 on weekend

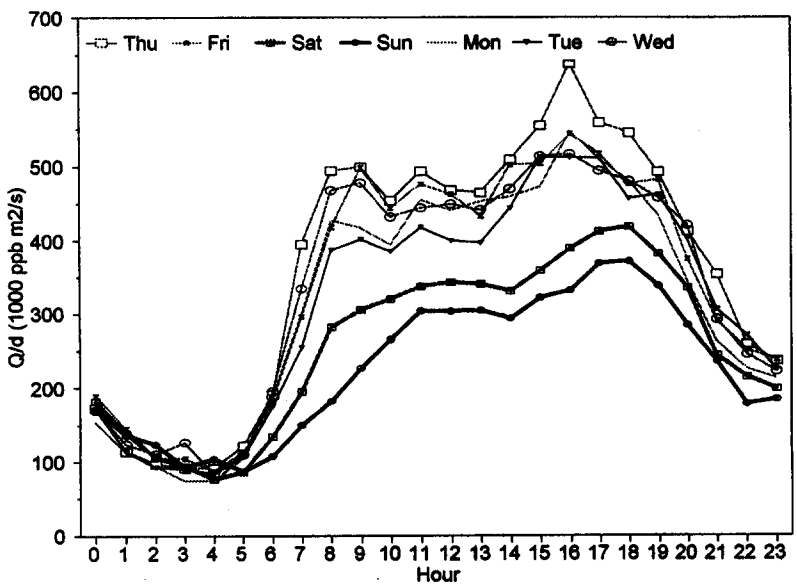

Fig. 3. Average NO_x emission rate diurnal variation over June 1989–May 1990.

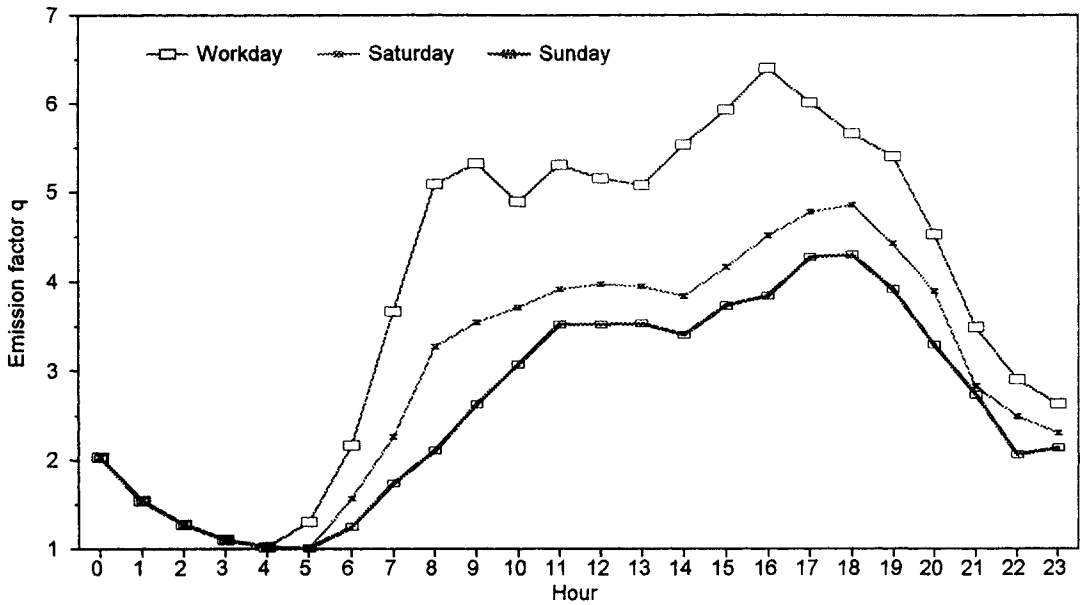

Fig. 4. Derived NO_x emission for from June 1989–May 1990.

days, as shown in Fig. 4, was defined. From Fig. 4, it may be seen that: (1) all weekdays had the same emission factors, obtained by averaging them, (2) weekdays and weekends had the same emission factors from 0:00 to 4:00, and (3) specific emission factors were used for Saturday and the Sunday, as shown in Fig. 4.

The general relationship between concentration C and explanatory variables that affect it is

$$C = F(X_1, \ldots, X_k)$$

where X_i ($i = 2, \ldots, k$) are explanatory variables. The explanatory variables chosen for C_{NO_x} based on some pre-analysis are displayed in Table 2. Explanatory variables chosen for NO_2 analysis are listed in Table 3. It is assumed that the primary emission factor of NO_2 was proportional to that of NO_x. X_8 ($MaxO_3NO$) is the product of NO (at CLL/BRI site) with the maximum O_3 concentration among the four sites STE, SIB, LH and HAR, and the contribution of NO_2 from the reaction of NO with the background O_3 was therefore considered. X_9 looks to explain the effect of photochemical production and removal of NO_2.

Table 2. Explanatory variables for NO_x regression

Variable	Name	Definition
X_1	CONST	Constant term: $X_1 = 1$
X_2	q	NO_x emission factor
X_3	WS	Wind Speed (m/s)
X_4	BLD	Boundary Layer Depth (m)
X_5	Temp	Temperature (°C)
X_6	Stab	Pasquill Stability Category
X_7	rh	Relative Humidity (%)

ORDINARY LEAST SQUARES (OLS) MODEL

Model selection

A power-law model for the function F has been accepted by most investigators (Annand and Hudson,

Table 3. Explanatory variables for NO_2 regression

Variable	Name	Definition
X_1	CONST	Constant term: $X_1 = 1$
X_2	q	NO_x emission factor
X_3	WS	Wind Speed (m/s)
X_4	BLD	Boundary Layer Depth (m)
X_5	Temp	Temperature (°C)
X_6	Stab	Pasquill Stability Category
X_7	rh	Relative Humidity (%)
X_8	$MaxO_3NO$	NO × Maximum O_3 among (STE, SIB, LH, HAR) sites
X_9	TotRad	Total Radiation (W/m^2)

Table 4. Results of NO_x regression by OLS model (data from June 1989–May 1990)

Multiple R			0.81450				
R Square			0.66340				
Adjusted R square			0.66317				
Standard error			0.40284				
Analysis of variance							
		DF	Sum of square		Mean square		
Regression		6	2722.45476		453.74246		
Residual		8512	1381.31606		0.16228		
$F = 2796.06959$		Signif $F = 0.0000$					
Variable	B	SE B	Beta	Tolerance	VIF	T	Sig T
ln(q)	0.7992	0.0088	0.6760	0.7194	1.390	91.177	0.0000
ln(WS)	−0.6969	0.0157	−0.6361	0.1917	5.217	−44.284	0.0000
ln(Temp)	−0.4285	0.0101	−0.3054	0.7640	1.309	−42.447	0.0000
ln(Stab)	0.4562	0.0231	0.2010	0.3823	2.613	19.769	0.0000
ln(rh)	0.1538	0.0208	0.0531	0.7089	1.411	7.381	0.0000
ln(BLD)	−0.0225	0.0093	−0.0351	0.1891	5.288	−2.430	0.0151
(Constant)	3.9280	0.1270				30.927	0.0000

Note. MultR is the multiple correlation coefficient; Rsq is the adjusted R square; Beta = $B_k(S_k/S_c)$, where S_k and S_c are the standard deviation of the kth explanatory variable and dependent variable, B_k is partial regression coefficient; Tolerance of a variable is a commonly used measure of colinearity, Tolerance$_i = 1 - R_i^2$, where R_i is the multiple correlation coefficient when the ith independent variable is predicted from the other independent variables; Variance Inflation Factor (VIF) is closely related to the tolerance; $T = B_i/S_{Bi}$; Sig T is significance level of T.

1981). So, C can be expressed as,

$$C = F(X_1, \ldots, X_k) = e^{B_1} X_2^{B_2} \ldots X_k^{B_k}$$

where k is the number of explanatory variables. Taking natural logarithms of both sides, thus

$$\ln(C) = B_1 + B_2 \ln(X_2) + \cdots + B_k \ln(X_k) \quad (1)$$

where B_1 is the intercept, B_2–B_k the partial slope coefficients. They will be determined by the regression model. Here, X_5 is temperature. There was only one hour during June 1989–May 1990 and less than thirty hours during June 1991–May 1992, when the hourly temperature was negative. Because the negative temperatures were in the range −1.5–0°C, and most of them between −1–0°C, it was assumed that all negative temperature = 0.1°C for the purpose of calculation.

Results and discussion

SPSS for Windows software (version 6.1) was used in this study. Listwise treatment was applied to missing data—a case would be eliminated if it had a missing value for any variables on the list. The regression results are displayed in Table 4 and 5 with clear physical meanings.

Table 5. Results of NO_2 regression by OLS model (data from June 1991–May 1992)

Multiple R			0.8182				
R Square			0.6695				
Adjusted R square			0.6691				
Standard error			0.2965				
Analysis of variance							
		DF	Sum of squares		Mean square		
Regression		8	1464.2980		183.0373		
Residual		8061	722.9422		0.0897		
$F = 2067.13377$		Signif $F = 0.0000$					
Variable	B	SE B	Beta	Tolerance	VIF	T	Sig T
ln(q)	0.2700	0.0091	0.3037	0.4150	2.410	29.570	0.0000
ln(WS)	−0.3760	0.0129	−0.4185	0.1997	5.006	−29.209	0.0000
ln(BLD)	0.0432	0.0079	0.0833	0.1778	5.623	5.486	0.0000
ln(Temp)	−0.0437	0.0056	−0.0565	0.8018	1.247	−7.930	0.0000
ln(Stab)	0.1290	0.0202	0.0657	0.3831	2.610	6.346	0.0000
ln(rh)	−0.1883	0.0182	−0.0774	0.7254	1.379	−9.327	0.0000
ln(MaxO_3NO)	0.2583	0.0047	0.5165	0.4705	2.125	54.660	0.0000
ln(TotRad)	−0.0422	0.0021	−0.2020	0.3928	2.546	−19.524	0.0000
(Constant)	2.6624	0.1095					0.0000

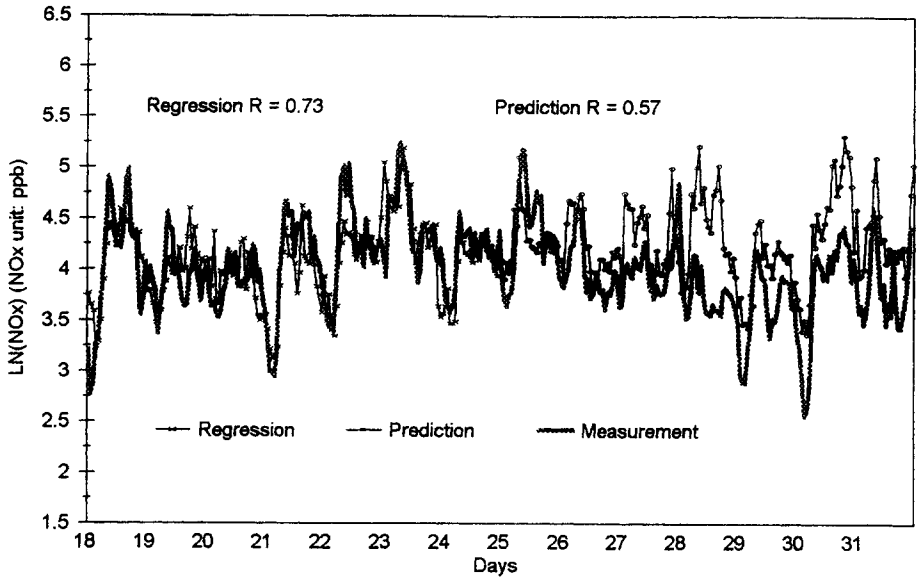

Fig. 5. Comparison of NO_x measurements with the regression and the prediction calculated by the OLS model.

(a) For NO_x, emissions and wind speed (WS) were the most important factors (cf. Beta values in Table 4); high NO_x concentrations are favoured by cold, stable and humid weather, this was especially true in winter smog episodes; boundary layer depth (BLD) seemed to be less important, one probable reason for this unexpected outcome was that BLD was itself very strongly correlated with WS (see VIF value in Table 4), the result was also consistent with that of Annand and Hudson (1981).

Prediction was also carried out. Unknown parameters in the equation (1) were obtained from the regression over the period 1st June 1989–24th May 1990 and used to predict NO_x concentrations in the last week (25–31 May). The regression (18–24 May) and prediction (25–31 May) values were compared to the measurements, shown in Fig. 5. Based on almost a whole year's data, the correlation coefficient for a regression and prediction over a week were 0.73 and 0.57 respectively, which is not as good as expected.

(b) For NO_2, the results of regression are displayed in Table 5. Primary emission of NO_2 (assumed to be proportional to NO_x) and wind speed (WS) were still the important factors, but the reaction of NO with O_3 seems the main source of NO_2. This agrees well with the fact that the primary emission contributes only a small fraction of ambient NO_2 concentrations (Harrison and Shi, 1996); most of measured urban background NO_2 comes from the reaction of NO with O_3. Temperature and stability became less important than those in the NO_x regression. NO_2 was not favoured by strong sunlight. Again, BLD was less important, the reason could be the same as that in NO_x regression. The reason that NO_2 was slightly favoured by higher BLD could be that NO_2 was formed from the reaction of NO with ozone-rich air mixed down from aloft.

As in the NO_x analysis, the regression of NO_2 was taken over the period 1 June 1991–24 May 1992 and the prediction was given for the last week (25–31 May). The regression (18th–24th May) and predicted (25–31 May) values were compared to the NO_2 measurements, shown in Fig. 6. Again, the correlation coefficients are not high. The main reason could be presence of autocorrelation which is common in time series.

Residual analysis

Rarely are assumptions not violated one way or another in regression analysis and other statistical procedures. Carrying out regressions without considering possible violations of the necessary assumptions can lead to results that are difficult to interpret and apply (SPSS, 1993).

Normality analysis shows that the residuals both for NO_x and NO_2 regression were normally distributed. The residual distribution for NO_x regression is shown in Fig. 7. Durbin–Watson value tests for NO_x and NO_2 regression suggest that the residuals are positively correlated. There is a very strong relationship between the residuals at adjacent time points, the lag 1 values (cf. Fig. 8). The autocorrelation coefficients at various lags are shown in Fig. 9. The partial autocorrelation function (shown in Fig. 10) indicates that only lag 1 autocorrelation appears to be important.

Residual analysis also shows that most of the points far away from the regression line are those cases with a small wind speed. Major error could be introduced

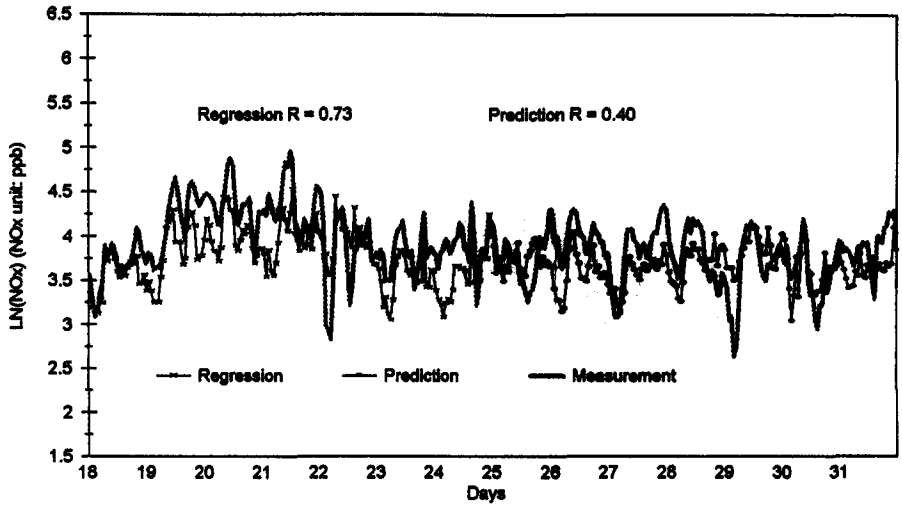

Fig. 6. Comparison of NO_2 measurements with the regression and the prediction calculated by the OLS model.

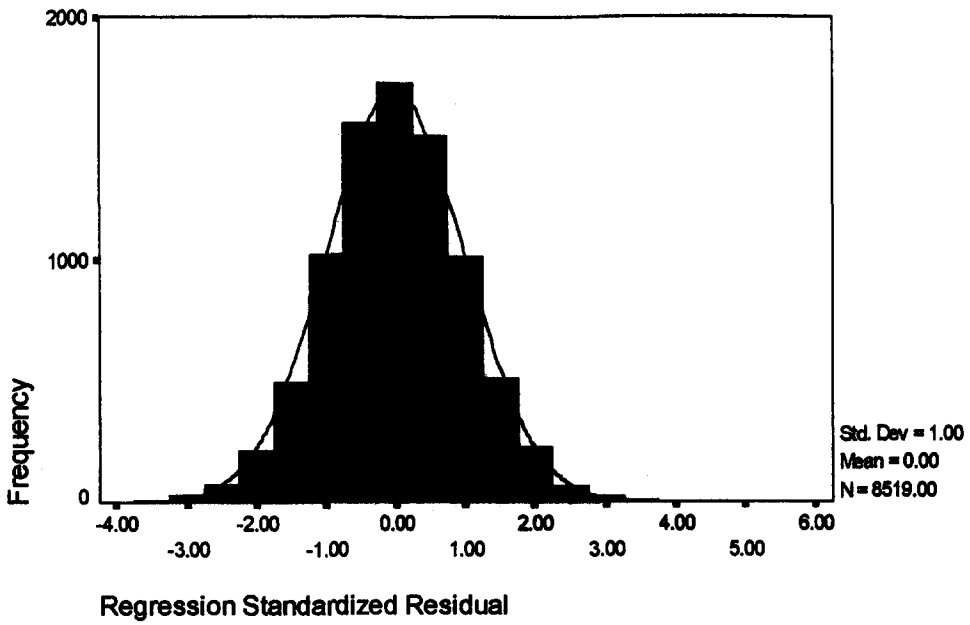

Fig. 7. Normality test for the OLS model residual of NO_x regression. A normal distribution is superimposed on a histogram of observed frequencies, indicated by the bars.

into the emission factor q when the wind speed is small. Low wind speeds are difficult to measure accurately and may be indicative of variable wind strengths across the conurbation. Both factors contribute to the uncertainty.

Because the residual had strong autocorrelation, the independence assumption was violated. The OLS coefficient estimates may no longer be efficient. Thus, an efficient model will be required to deal with this problem.

AUTOREGRESSION (AR) MODEL

Model selection

To solve the problem of autocorrelation, a first-order autoregression model is employed,

$$\ln(C_j) = B_1 + \sum_{i=2}^{k} B_i \ln(X_{ij}) + u_j$$

$$u_j = \phi u_{j-1} + e_j, \quad j = 1, \ldots, N \qquad (2)$$

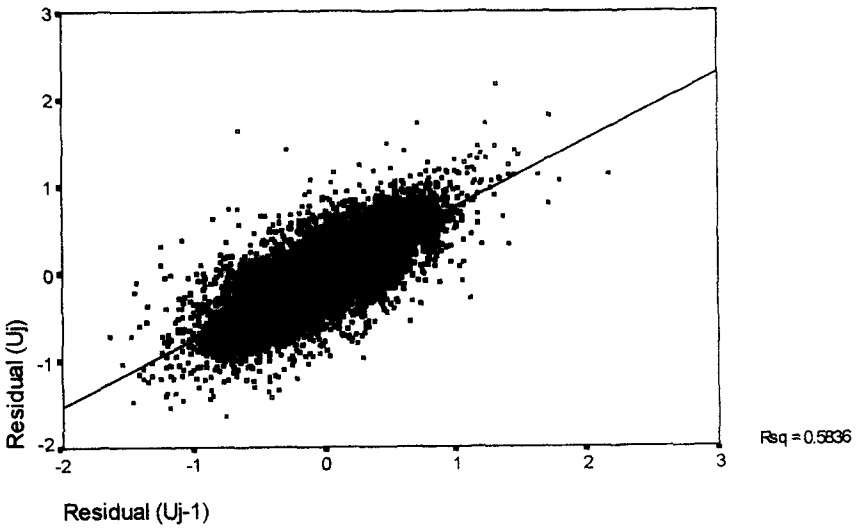

Fig. 8. Scatter plot of first-order autocorrelation for the residual of NO_x regression.

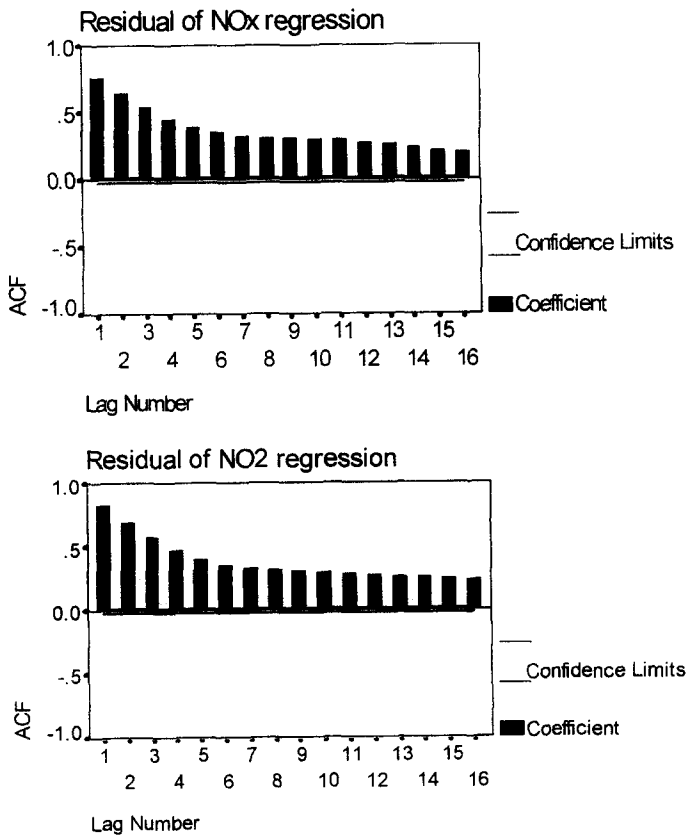

Fig. 9. Autocorrelation coefficients of the OLS model residual for NO_x and NO_2 regression.

where e_j is a white noise (independent normal value with a mean of 0 and a constant variance of σ^2). Thus,

$$\ln(C_j) = \beta_1 + \sum_{i=2}^{k} \beta_i \ln(X_{ij}) + \beta_{k+1} \text{Lag1}_j + e_j \qquad j = 1, \ldots, N \tag{3}$$

where Lag1 is a new variable,

$$\text{Lag1}_j = \ln(C_{j-1}) - B_1 - \sum_{i=2}^{k} B_i \ln(X_{ij-1})$$

Results and discussion

Regression and predictions similar to those calculated by the OLS model were carried out. The

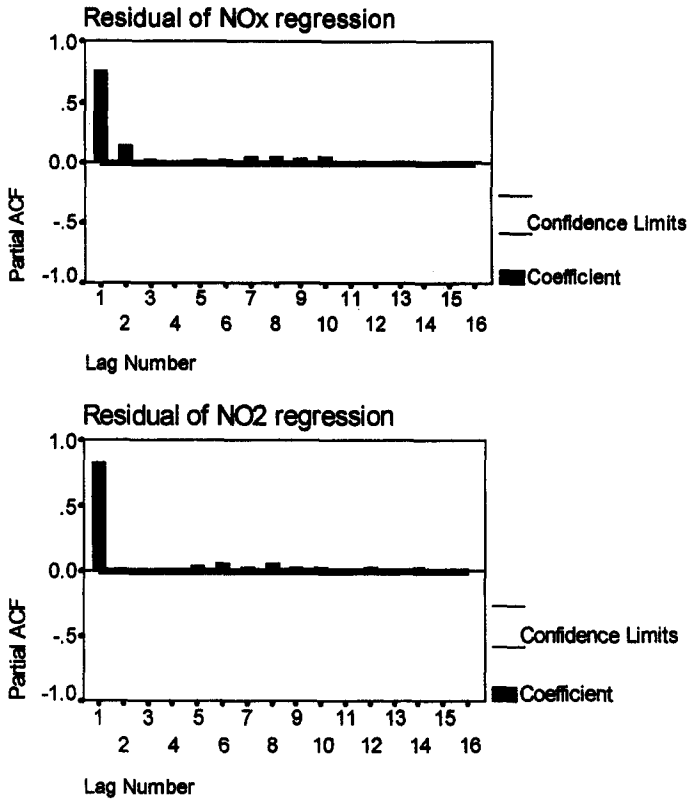

Fig. 10. Partial autocorrelation coefficients of the OLS model residual for NO_x and NO_2 regression.

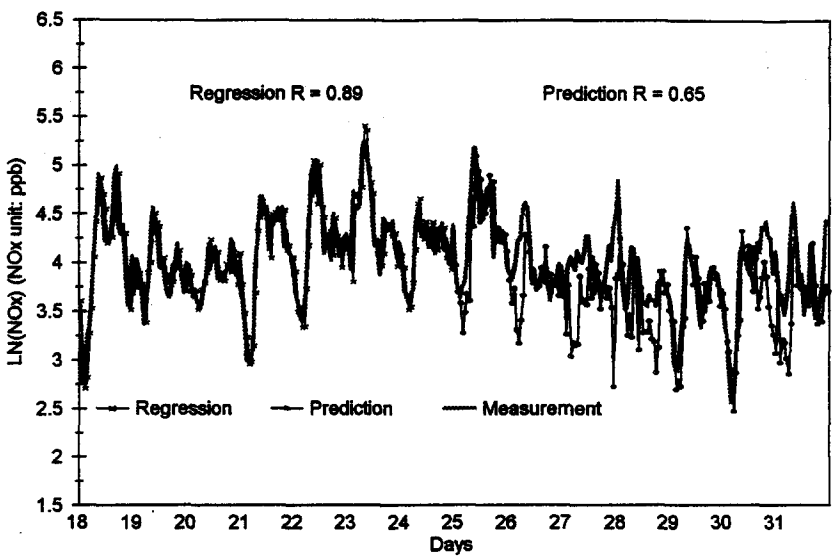

Fig. 11. Comparison of NO_x measurements with the regression and the prediction calculated by the AR model.

results are shown in Fig. 11 for NO_x and Fig. 12 for NO_2. Unknown parameters in equations (2) and (3) for NO_x were obtained from the regression over the period 1 June 1989–24 May 1990, and used to predict NO_x concentrations in the last week (25–31 May). For NO_2, the regression was taken over the period of 1st June 1991–24 May 1992 and the prediction was given for the last week (25–31 May 1992). Comparing the results of the AR model (shown in Figs 11 and 12) with those of the OLS model (Figs 5 and 6), it is

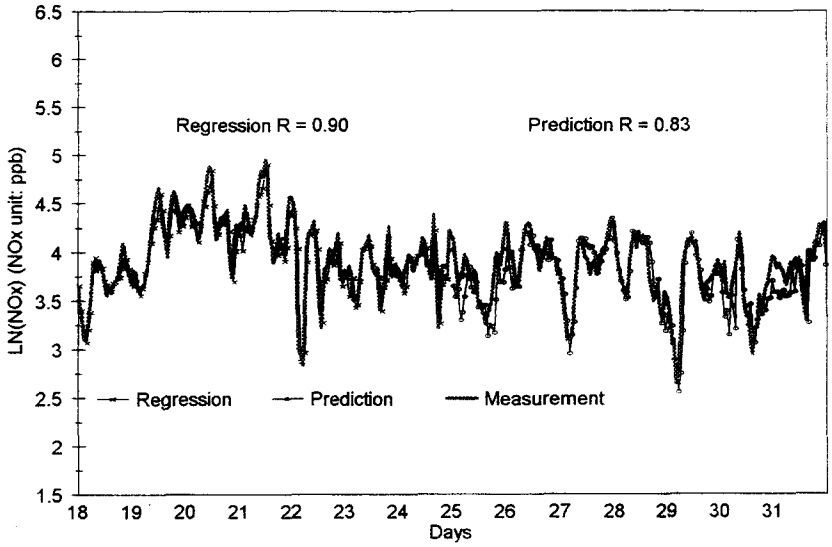

Fig. 12. Comparison of NO_2 measurements with the regression and the prediction calculated by the AR model.

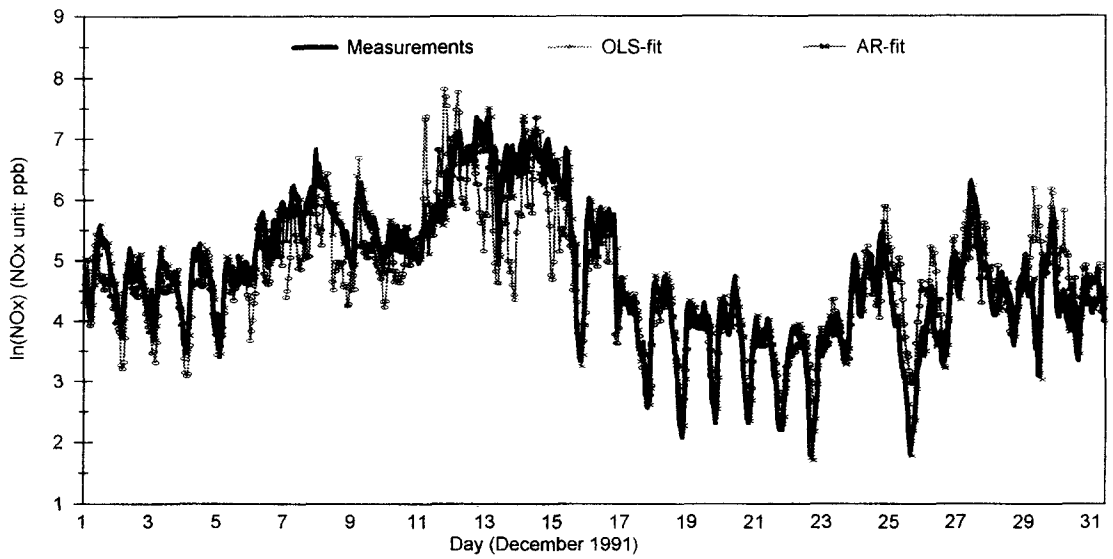

Fig. 13. Comparison of hourly NO_x measurements with the values calculated by the OLS and AR models based on the whole year data regression.

obvious that the AR model is much more efficient than the OLS model both in regression and prediction. Seen from Figs 11 and 12, the AR model has similar correlation coefficients R for NO_x and NO_2 regressions, but has a higher R for the prediction of NO_2 than that of NO_x. Generally, NO_2 is secondary pollutant and NO_x is primary ($NO_x = NO + NO_2$). Thus, it indicates that the explanatory variables chosen for NO_2 regression are reasonable.

There was a severe smog episode in London during 11–17 December 1991 (Bower *et al.*, 1994; Derwent *et al.*, 1995). Hourly NO_x measurements were compared with those calculated by the OLS and AR models, shown in Figs 13 and 14. The parameters in the OLS and AR models were taken from a regression over the period of June 1991 to May 1992. The AR model performed well both for NO_x and NO_2 regressions. The NO_x levels calculated by the OLS model reasonably matched with the measurements, but the NO_2 levels during the episode did not. Since the OLS model has a similar regression coefficient R for NO_x and NO_2, the fact that the observed NO_2 levels were consistently higher than modelled values during the episode implies that an exceptional NO_2 source was

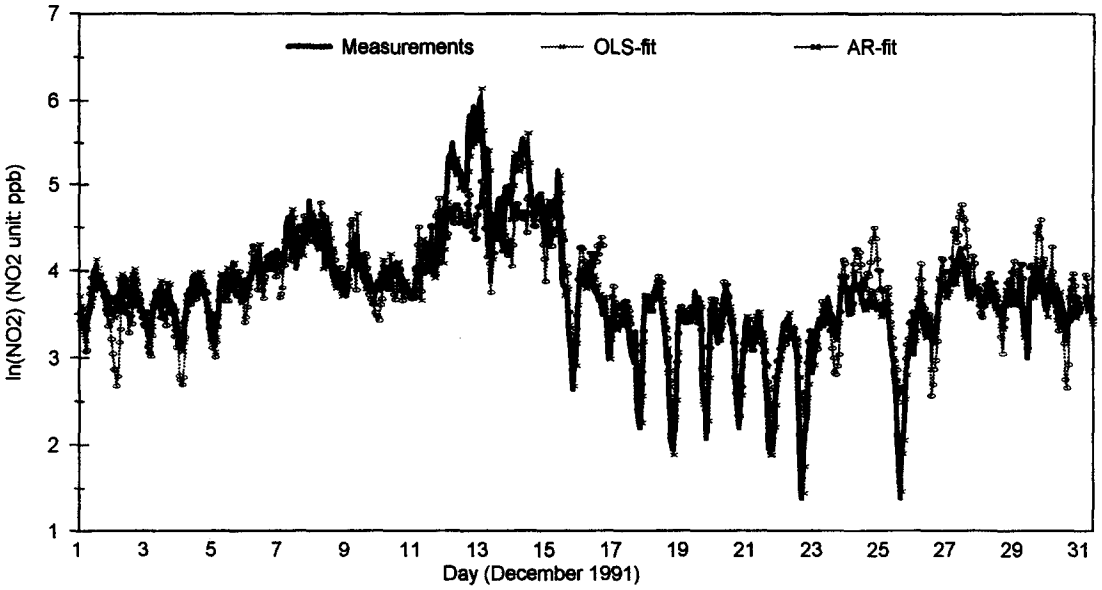

Fig. 14. Comparison of hourly NO_2 measurements with the values calculated by the OLS and AR models based on the whole year data regression.

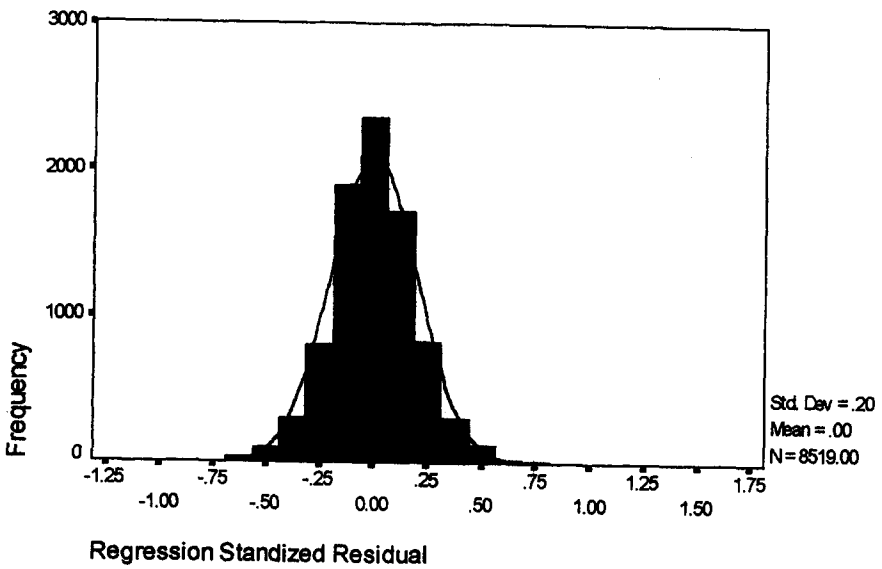

Fig. 15. Normality test for the AR model residual for NO_x regression.

present. Other research has shown that reaction of $2NO + O_2 \rightarrow 2NO_2$ was an important NO_2 source in the episode, but cannot explain the rapidity of rise of NO_2 (Bower *et al.*, 1994; Harrison *et al.*, 1997).

Residual analysis

Autocorrelation and distribution of the AR model residuals for NO_x and NO_2 were analysed. It is found that there is no significant autocorrelation in the residuals and they are close to a normal distribution, as shown in Fig. 15.

STABILITY OF THE MODEL

Figures 11 and 12 show that the AR model is able to predict with a correlation coefficient R of 0.83 for NO_2 and with 0.65 for NO_x. Further investigation was carried out to see whether the model is stable. One method to examine stability of the model is Principal Component Analysis (PCA) (Inoue *et al.*, 1986a). The PCA analysis of the variables is shown in Tables 6 and 7 for NO_x and NO_2 respectively, where only up to three values (its absolute value $\geqslant 0.5$) for

Table 6. Principal component analysis for the variables—factor loading for NO_x regression

	June 1989–May 1990			June 1991–May 1992		
Variable	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3
ln(NO_x)	−0.5057		0.7630	−0.6063		0.7173
ln(q)			0.8195			0.7566
ln(WS)	0.9563			0.9661		
ln(BLD)	0.8966			0.8731		
ln(Temp)		−0.8417			0.7253	
ln(Stab)		0.6002			−0.6483	
ln(rh)		0.7599			−0.6977	

Table 7. Principal component analysis for the variables—factor loading for NO_2 regression

	June 1989–May 1990			June 1991–May 1992		
Variable	Factor 1	Factor 2	Factor 3	Factor 2	Factor 3	Factor 1
ln(NO_2)	0.6816			0.7927		
ln(q)	0.7933			0.7130		
ln(WS)		0.9270			0.9281	
ln(BLD)		0.9083			0.8323	
ln(Temp)			0.8275			0.7304
ln(Stab)			−0.5483			−0.7112
ln(rh)			−0.7430			−0.6498
ln($MaxO_3NO$)	0.8509			0.8553		
ln(TotRad)	0.5940		0.5936			0.7494

Table 8. Comparison of the OLS and AR model parameters for NO_x regression between two years

Date	June 1989–May 1990			June 1991–May 1992		
Model	OLS		AR	OLS		AR
Variable	B	Beta	β	B	Beta	β
Lag1			0.9275			0.9428
ln(q)	0.7992	0.6760	0.7197	0.8631	0.6249	0.7385
ln(WS)	−0.6969	−0.6361	−0.1105	−0.8007	−0.5732	−0.1189
ln(Temp)	−0.4285	−0.3054	−0.4998	−0.2986	−0.2458	−0.2382
ln(Stab)	0.4562	0.2010	0.0688	0.5092	0.1658	0.0709
ln(rh)	0.1538	0.0531	0.2611	0.3731	0.0984	0.4562
ln(BLD)	−0.0225	−0.0351	−0.1378	−0.0611	−0.0751	−0.0070
(Constant)	3.9280		3.5882	3.2716		2.1862
Multiple R	0.8145		0.9573	0.8270		0.9644
Residual	8512			8431		

each factor were taken. The variables were classified into the same three groups in two years. In Table 6, the first factor corresponds to the meteorological parameters which have a direct effect on NO_x concentrations; the second factor corresponds to those which have an indirect effect on NO_x; the third factor corresponds to the emission of NO_x. The contribution of each variable to the NO_x may be seen in Table 8. In Table 7, the first group corresponds to the sources of NO_2; the second corresponds to those which have a direct effect on NO_2 concentrations; the third corresponds to those which have an indirect effect on NO_2. The contribution of each variable to the NO_2 concentration is exemplified by Table 9. Putting these together, it is quite reasonable to believe that the model is stable and can be used for prediction.

The factor order between the two years is different in Table 7. The reason could be that: (1) the eigenvalue difference between the factors was small, so the difference in the weight of factors was also small; (2) the sample population and the distribution of missing data were different in the two years, 7089 in 1989–1990 and 8070 in 1991–1992.

Table 9. Comparison of the OLS and AR model parameters for NO_2 regression between two years

Date	June 1989–May 1990			June 1991–May 1992		
Model	OLS		AR	OLS		AR
Variable	*B*	Beta	β	*B*	Beta	β
Lag1			0.9196			0.9204
ln(*q*)	0.1880	0.2702	0.1461	0.2700	0.3037	0.2680
ln(WS)	−0.3514	−0.5361	−0.0412	−0.3760	−0.4185	−0.0557
ln(Temp)	0.0371	0.0448	0.0263	−0.0432	−0.0565	−0.0471
ln(Stab)	0.1584	0.1066	0.0035	0.1290	0.0658	0.0103
ln(rh)	−0.3207	−0.1833	−0.0296	−0.1883	−0.0774	0.0343
ln(BLD)	0.0713	0.1944	0.0079	0.0432	0.0823	0.0079
ln($MaxO_3NO$)	0.1506	0.4213	0.1667	0.2583	0.5165	0.1959
ln(TotRad)	−0.0252	−0.1526	−0.0149	−0.0422	−0.2020	−0.0222
(Constant)	3.3842		2.3828	2.6624		2.0343
Multiple *R*	0.7459		0.9532	0.8182		0.9630
Residual	7080			8061		

CONCLUSIONS

An OLS model and a first-order AR model have been used for the regression and prediction of hourly NO_x and NO_2 concentrations in central London. Source strength and wind speed were the most important factors which influence NO_x concentrations; however, in addition to these two factors, $MaxO_3NO$ (the product of NO and maximum O_3 concentration from among four monitoring sites) also had a major influence on NO_2 concentrations, reflecting the secondary pollutant character of NO_2.

Residual analysis showed that lag1 autocorrelation was present in the residual serial of the OLS model. Therefore, the independence assumption was violated and the OLS coefficient estimates may not be efficient. The AR model can be used for NO_x and NO_2 regression with a high correlation coefficient ($R > 0.95$), and is capable of predicting NO_2 ($R = 0.83$) and NO_x ($R = 0.65$) when the explanatory variables are available. The comparison of regression parameters and PCA results for two separate years suggests that the model is stable.

The OLS and AR models were applied to analyze NO_x and NO_2 concentrations during a severe pollution episode in London based on a whole year's data. That NO_x concentrations, calculated by the OLS model based on the regression of a whole year's data, could simulate the observed high NO_x levels during the episode suggests that the high NO_x levels were not the result of abnormal emissions. However, although the OLS model had a similar *R* for NO_x and NO_2 regression, the NO_2 measurements in the episode were consistently higher than those simulated. This indicates that an exceptional NO_2 source, other than those explanatory variables displayed in Table 3, was present in the episode, which is confirmed by Harrison *et al.* (1997). The AR model performed well both for NO_x and NO_2 regressions.

Acknowledgements—This work was funded by the U.K. Department of the Environment under the contract No: EPG 1/3/02.

REFERENCES

Annand, W. J. D. and Hudson, A. M. (1981) Meteorological effects on smoke and sulphur dioxide concentrations in the Manchester Area. *Atmospheric Environment* **15** (5), 799–806.

Bower, J. S., Broughton, G. F. J. and Stedman, J. R. (1994) A winter NO_2 smog episode in the UK. *Atmospheric Environment* **28** (3), 461–475.

Chock, P. D., Terrell, T. R. and Levitt, S. B. (1975) Time series analysis of riverside, California, air quality data. *Atmospheric Environment* **9**, 978–989.

Derwent, R. G., Middleton, D. R., Field, R. A., Goldstone, M. E., Lester, J. N. and Perry, R. (1995) Analysis and interpretation of air quality data from an urban roadside location in central London over the period from July 1991 to July 1992. *Atmospheric Environment* **29** (8), 923–946.

Hanna, S. R., Briggs, G. A. and Hosker, R. P. (1982) *Handbook of Atmospheric Diffusion*, TIC-11223, U.S.A. Department of Energy.

Harrison, R. M. and Shi, J. P. (1996) Sources of nitrogen dioxide in winter smog episodes. In: *The 5th International Symposium: Highway and Urban Pollution; Science of Total Environment* **189/190**, 391–399.

Harrison, R. M., Shi, J. P. and Grenfell, J. L. (1997) Novel nighttime free radical chemistry in severe nitrogen dioxide pollution episodes. *Atmospheric Environment* (submitted).

Inoue, T., Hoshi, M. and Taguri, M., (1986a) Regression analysis of nitrogen oxide concentration. *Atmospheric Environment* **20** (1), 71–85.

Inoue, T., Taguri, M. and Hoshi, M. (1986b) Prediction of nitrogen oxide concentration by a regression model. *Atmospheric Environment* **20** (12), 2325–2337.

Milionis, A. E. and Davies, T. D. (1994) Regression and stochastic models for air pollution—I. review, comments and suggestions. *Atmospheric Environment* **28** (17), 2801–2810.

Panofsky, H. A. and Dutton, J. A. (1984) Atmospheric Turbulence. Wiley, New York.

Pasquill, F. and Smith, F. B. (1983) *Atmospheric Diffusion* 3rd Edn. Wiley, Chichester.

Revlett G. H. (1978) Ozone forecasting using empirical modelling. *Journal of Air Pollution Control Association* **28** (4), 338–343.

Shi, J. P. (1996) The chemistry of NO_x in urban air. Ph.D. thesis, The University of Birmingham.

SPSS (1993) SPSS for Windows, Base System User's Guide Release 6.0. Marija J. Norusis, SPSS Inc.

Wolff, G. T. and Lioy, P. J. (1978) An empirical model for forecasting maximum daily ozone levels in northeastern US *Journal of Air Pollution Control Association* **28**, 1034–1038.

Zannetti, P. (1990) *Air Pollution Modelling*. Van Nostrand Reinhold, New York.