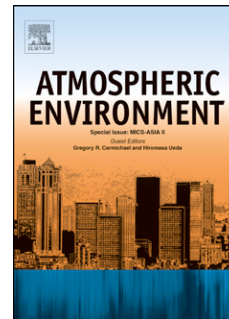


# Accepted Manuscript

Title: Prediction of Traffic-Related Nitrogen Oxides Concentrations using Structural Time Series Models

Authors: Anneka Ruth Lawson, Bidisha Ghosh, Brian Broderick



PII: S1352-2310(11)00442-0

DOI: [10.1016/j.atmosenv.2011.04.053](https://doi.org/10.1016/j.atmosenv.2011.04.053)

Reference: AEA 10411

To appear in: *Atmospheric Environment*

Received Date: 11 January 2011

Revised Date: 14 April 2011

Accepted Date: 19 April 2011

Please cite this article as: Lawson, A.R., Ghosh, B., Broderick, B. Prediction of Traffic-Related Nitrogen Oxides Concentrations using Structural Time Series Models, *Atmospheric Environment* (2011), doi: [10.1016/j.atmosenv.2011.04.053](https://doi.org/10.1016/j.atmosenv.2011.04.053)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

## Prediction of Traffic-Related Nitrogen Oxides Concentrations using Structural Time Series Models

Anneka Ruth Lawson<sup>a</sup>, Bidisha Ghosh<sup>b\*</sup>, Brian Broderick<sup>c</sup>

<sup>a</sup>Civil, Structural and Environmental Engineering Department, Trinity College, Dublin, Ireland.  
lawsona@tcd.ie

<sup>b</sup>Civil, Structural and Environmental Engineering Department, Trinity College, Dublin, Ireland.  
bghosh@tcd.ie

<sup>c</sup>Civil, Structural and Environmental Engineering Department, Trinity College, Dublin, Ireland.  
bbrodrck@tcd.ie

### ABSTRACT

Ambient air quality monitoring, modeling and compliance to the standards set by European Union (EU) directives and World Health Organization (WHO) guidelines are required to ensure the protection of human and environmental health. Congested urban areas are most susceptible to traffic related air pollution which is the most problematic source of air pollution in Ireland. Long-term continuous real-time monitoring of ambient air quality at such urban centers is essential but often not realistic due to financial and operational constraints. Hence, the development of a resource-conservative ambient air quality monitoring technique is essential to ensure compliance with the threshold values set by the standards. As an intelligent and advanced statistical methodology, a Structural Time Series (STS) based approach has been introduced in this paper to develop a parsimonious and computationally simple air quality model. In STS methodology, the different components of a time-series dataset such as the trend, seasonal, cyclical and calendar variations can be modeled separately. To test the effectiveness of the proposed modeling strategy, average hourly concentrations of nitrogen dioxide and nitrogen oxides from a congested urban arterial in Dublin city centre were modeled using STS methodology. The prediction error estimates from the developed air quality model indicate that the STS model can be a useful tool in predicting nitrogen dioxide and nitrogen oxides concentrations in urban areas and will be particularly useful in situations where the information on external variables such as meteorology or traffic volume is not available.

### KEYWORDS

Air Quality Forecast, Structural Time Series, Nitrogen Oxides, Vehicular Emission, Dublin City

---

\* Corresponding author

## 1. Introduction

Oxides of nitrogen are one of the major pollutants present in the ambient air. These oxides are harmful gaseous substances largely produced by the combustion of fossil fuels. Short-term exposure to nitrogen oxides can cause defective pulmonary function, with asthmatics being the most vulnerable, while long-term exposure can cause abnormal effects in the lungs, spleen, liver and blood. The adverse effects of nitrogen oxides (NO<sub>x</sub>) exposure include changes in cell type in tracheobronchial and pulmonary regions, lung structure, metabolism and defense against bacterial and viral infection and emphysema-like symptoms (World Health Organization, 2000). Traffic emissions are the principal source of this pollutant. With the increasing numbers of vehicular traffic on roads, the development of regulations for limiting and controlling the levels of nitrogen oxides in ambient air is of utmost importance.

To alleviate the adverse effects of NO<sub>x</sub> exposure, the level of nitrogen oxides in ambient air should be limited to an annual average of 30µg/m<sup>3</sup> and for nitrogen dioxide (NO<sub>2</sub>) the annual average should be limited to 40µg/m<sup>3</sup> in Europe (2008/50/EC Directive, Council of the European Union). Compliance to the standards set by the EU directives (2008/50/EC Directive, Council of the European Union) and World Health Organization (WHO) guidelines are a part of national legislation in Ireland (2001/81/EC Directive, Council of the European Union and European Environment Agency, 2009). However at the current levels, Ireland's production of oxides of nitrogen remains in exceedance of its limit by more than 10% (Minister for the Environment, Heritage and Local Government, 2002 & 2004). In order to reach and remain congruent with what legislation considers to be a suitable pollutant level for ambient air, an air quality management strategy must be implemented. This strategy must involve a method of predicting future pollutant levels, to allow preventative action to be taken in advance if the pollutant levels are in danger of exceeding legislative limit concentrations. It would also allow pollutant trends to be studied in order to enable implementation of long term strategies for reduction of pollutant concentrations.

There exist various theoretical and statistical techniques to model and predict air quality. The theoretical techniques, such as atmospheric dispersion modeling, aim to identify the underlying physical and chemical equations controlling the pollutant concentrations and hence require detailed emissions data and meteorological information. Atmospheric dispersion modeling techniques have been applied to model NO<sub>x</sub> concentrations by Simpson et al. (1990). On the other hand, the statistical techniques aim to identify the dynamical behavior of the air pollutant level observations and often do not require detailed emission inventories. The well-known statistical techniques used for air quality modeling are regression analysis and Artificial Neural Network (ANN) algorithms. The Autoregressive Integrated Moving Average (ARIMA) model is a sophisticated regression technique which has been applied by Sanchez et al., (1997) and Hsu (1997) to model various air pollutants, including oxides of nitrogen. Sharma et al. (2009), Ghazali et al. (2010) have used regression based techniques to model NO<sub>2</sub> concentrations, NO<sub>x</sub> concentrations and related ozone concentrations. Statistical learning algorithm based regression techniques (support vector regression) have been applied successfully to model oxides of nitrogen (Lu and Wang, 2005). These concentrations have also been modeled using various ANN based algorithms where predictions are often improved by utilizing information on other explanatory variables such as traffic volume or meteorological conditions (Lu et al., 2003, Chelani, 2005 and Hoffman, 2006, Cai et al. 2009).

Another potential statistical method to model air quality is the Structural Time Series (STS) models. This is a class of time-series models, formulated based on state-space methodology and consequently utilizes the concept of unobserved components. The different unobserved components of a STS model are trend, seasonal fluctuations, cyclical and calendar variation together with the effects of explanatory variables and interventions (outliers and sudden changes). These unobserved components have a direct interpretation in terms of the temporal variability of a time-series dataset. Due to the recursive and Markovian nature of the STS model, it is easy to handle missing values and outlier values and it is also possible to back-predict past unobserved levels. In the past, STS models have been successfully applied for assessing the effects of the introduction of the UK seatbelt legislation (Harvey and Durbin, 1986), modeling urban traffic flow (Ghosh et al., 2009) and

modeling climatological observations (Visser and Molenaar, 1995), but it is a relatively unexplored concept in the area of air quality modeling. Similar techniques have been applied by Schlink and Herbarth (1997) to predict SO<sub>2</sub> concentrations and by Zolaghadri and Cazaurang (2005) to predict PM<sub>10</sub> concentrations. However, there is no instance of the application of STS methodology in modeling the oxides of nitrogen in a traffic-dominated urban environment.

This paper will employ the STS methodology to develop an air quality model to predict concentrations of oxides of nitrogen (NO<sub>x</sub> and NO<sub>2</sub>). The originality of this application lies in the meaningful depiction of the different factors which influence variations in the concentrations of oxides of nitrogen utilizing the different components of the STS model. A case study in the city centre of Dublin in Ireland is performed to test the effectiveness of the proposed prediction model. The paper is organized in four sections. The next section describes the theory behind the STS modeling methodology. The third section discusses the developed STS air quality model and test NO<sub>x</sub> and NO<sub>2</sub> data. The fourth section concludes the paper.

## 2. Structural Time-Series Modeling

### 2.1 Theoretical Background

The STS methodology (Harvey, 1989) is a particular time-series analysis which assumes that a time-series is made up of a number of unobserved components such as (deterministic and stochastic) trend, seasonal, regression elements and disturbance terms, which have direct interpretation from the time-series data. Each component can be modeled discretely, allowing their evolution over time to be studied, and their contribution to the final predictions can be observed clearly.

A univariate STS model for the observed time-series data  $\mathbf{y}(1 \times T)$  can be described by the following general equation

$$y_t = \mu_t + \gamma_t + v_t + \psi_t + \varepsilon_t, \quad \varepsilon_t \sim \text{NID}(0, \sigma_\varepsilon^2) \quad t = 1, \dots, T \quad (1)$$

where,  $y_t$  is the observed data at an instant of time  $t$ ,  $\mu_t$  is the trend,  $\gamma_t$  is the seasonal,  $v_t$  is the slope,  $\psi_t$  is the autoregressive (AR) component, and  $\varepsilon_t$  is the observation disturbance error or the random error component at the same instant of time. The disturbance errors are assumed to be normally and independently distributed (NID) with zero mean and variance  $\sigma_\varepsilon^2$ . Equation (1) contains all possible temporal components of the STS model. In this paper, models containing different combinations of these components have been studied in order to achieve the most accurate prediction algorithm.

The first model evaluated is the Local Level Model; this contains a stochastic trend and an irregular component and formulates as

$$y_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim \text{NID}(0, \sigma_\varepsilon^2) \quad t = 1, \dots, T \quad (2)$$

The stochastic trend (level) component denotes the long-term movement of the time series which can be extrapolated into the future. A Markov model of the stochastic trend can be,

$$\mu_{t+1} = \mu_t + \xi_t, \quad \xi_t \sim \text{NID}(0, \sigma_\xi^2) \quad t = 1, \dots, T \quad (3)$$

where  $\xi_t$  is the level disturbance; the variance of this component,  $\sigma_\xi^2$  and the variance of the observation disturbance,  $\sigma_\varepsilon^2$ , are mutually uncorrelated. The level component denotes a linear trend when  $\sigma_\xi^2 = 0$ , and the local level model, described in equation (2) & (3), collapses to a Deterministic Level Model, whereby equation (2) changes to,

$$y_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim \text{NID}(0, \sigma_\varepsilon^2) \quad t = 1, \dots, T \quad (4)$$

The Local Linear Trend Model is obtained by adding a slope component the local level model. The slope component signifies the long-term change of the trend and is included in a model as follows:

$$\left. \begin{aligned} \mu_{t+1} &= \mu_t + \nu_t + \xi_t \\ \nu_{t+1} &= \nu_t + \zeta_t \end{aligned} \right\} \begin{aligned} \xi_t &\sim \text{NID}(0, \sigma_\xi^2) \\ \zeta_t &\sim \text{NID}(0, \sigma_\zeta^2) \end{aligned} \quad t = 1, \dots, T \quad (5)$$

where  $\nu_t$  is the slope component and  $\zeta_t$  is the slope disturbance. The Local Linear Trend Model will collapse into a Deterministic Linear Trend Model when the variance of the disturbance term,  $\sigma_\zeta^2$  equals to zero.

The Local Level Model with Seasonal is formulated by adding a seasonal component,  $\gamma_t$  to equation (2) in the local level model

$$y_t = \mu_t + \gamma_t + \varepsilon_t \quad \varepsilon_t \sim \text{NID}(0, \sigma_\varepsilon^2) \quad t=1, \dots, T \quad (6)$$

The seasonal is the periodic component of the time-series; it may be 12 months, 24 hours or any other fixed period  $s$ . The seasonal component is modeled using a trigonometric specification,

$$\gamma_{t+1} = \sum_{j=1}^{\lfloor s/2 \rfloor} \gamma_{j,t} \quad (7)$$

where each  $\gamma_{j,t}$  is generated by

$$\begin{bmatrix} \gamma_{j,t+1} \\ \gamma_{j,t+1}^* \end{bmatrix} = \begin{bmatrix} \cos \lambda_j & \sin \lambda_j \\ -\sin \lambda_j & \cos \lambda_j \end{bmatrix} \begin{bmatrix} \gamma_{j,t} \\ \gamma_{j,t}^* \end{bmatrix} + \begin{bmatrix} \omega_{j,t} \\ \omega_{j,t}^* \end{bmatrix}, \quad j = 1, \dots, S/2 \quad t = 1, \dots, T \quad (8)$$

where  $\lambda_j = 2\pi j/s$  is the frequency in radians and  $\omega_{j,t}$  and  $\omega_{j,t}^*$  are mutually uncorrelated random normal disturbances with zero mean and common variance  $\sigma_\omega^2$ . The superscript \* in (8) indicates the minimal realization of the state vector in the state space form (Durbin & Koopman, 2001). The seasonal component denotes a deterministic periodicity when  $\sigma_\omega^2 = 0$ . The Local Level Model with Seasonal is described by equations (3) and (6)-(8). The Deterministic Level and Seasonal Model is defined by equations (4) and (6)-(8).

As indicated in equation (1), an autoregressive component can be added to the time-series model. An Autoregressive Process of order  $p$ , AR ( $p$ ) suggests that each observation is correlated to the previous  $p$  observations in the dataset.

$$\psi_t = \phi_1 \psi_{t-1} + \phi_2 \psi_{t-2} + \dots + \phi_p \psi_{t-p} + \kappa_t, \quad \kappa_t \sim \text{NID}(0, \sigma_\kappa^2) \quad t = 1, \dots, T \quad (9)$$

where  $\phi_1, \phi_2, \dots, \phi_p$  are diagonal matrices and  $\kappa_t$  is the disturbance of the autoregressive component and the variance of  $\kappa_t$  is  $\sigma_\kappa^2$ . A Deterministic Level and Seasonal Model with AR will be defined by equation (1) (excluding the  $\nu_t$  component), equation (3) and (7)-(9). The disturbances ( $\varepsilon_t, \xi_t, \zeta_t, \omega_{j,t}, \kappa_t$ ) experienced by each component in the STS model are mutually uncorrelated and the variances of the disturbances of each of the components ( $\sigma_\varepsilon^2, \sigma_\xi^2, \sigma_\zeta^2, \sigma_\omega^2, \sigma_\kappa^2$ ) represents the degree to which each component stochastically varies with time and are termed as hyperparameters. Equations (1) to (9) are generally solved in state-space form using Kalman filter based algorithms (Harvey, 1989). The hyperparameters and the components are estimated using a maximum likelihood estimation method.

The general representation of the STS model in the state-space form is as follows:

$$\text{Observation Equation: } y_t = Z_t \alpha_t + \varepsilon_t \quad \varepsilon_t \sim \text{NID}(0, H_t) \quad (10)$$

$$\text{State Equation: } \alpha_{t+1} = T_t \alpha_t + R_t \eta_t \quad \eta_t \sim \text{NID}(0, Q_t) \quad t=1, \dots, T \quad (11) \text{ where}$$

$\alpha_t$  is the vector of unobserved components, known as the state vector;  $Z_t$  and  $T_t$  are the state system matrices;  $R_t$  is the error system matrix,  $\eta_t$  is the state disturbance matrix and  $H_t$  and  $Q_t$  are the variance matrices. Equations (10) and (11) represent the general state form of the STS model. The different models described in the previous section can be represented in this form with the different vectors and matrices having different dimensions and values. For the purpose of analysis,

the Kalman filter provides a method of updating the state vector when a new observation becomes available. Forecasts are made by extrapolating the components estimated at the end of the sample set using the Kalman filter.

## 2.2 Prediction Error Estimates

The prediction accuracy of the proposed models is estimated from the mean absolute percentage error (MAPE) and the root mean square error (RMSE) estimates.

MAPE is a measure of accuracy as a percentage and is defined as

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{x_t - \hat{x}_t}{x_t} \right| \quad (12)$$

where  $n$  is the number of data points predicted,  $x_t$  is the observed value and  $\hat{x}_t$  is the corresponding predicted value.

RMSE is a measure of precision which is defined as

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (x_t - \hat{x}_t)^2} \quad (13)$$

These error estimates permit comparison of the observed data and the forecasted data, to evaluate the prediction accuracy of the proposed models.

## 3. Modeling Nitrogen Oxides Concentrations at Dublin City Centre

### 3.1 Air Pollution Data

The proposed STS methodology has been applied to model nitrogen dioxide (NO<sub>2</sub>) and nitrogen oxide (NO<sub>x</sub>) concentrations as observed in ambient air at a road-side air quality monitoring station on Pearse Street, in Dublin City.

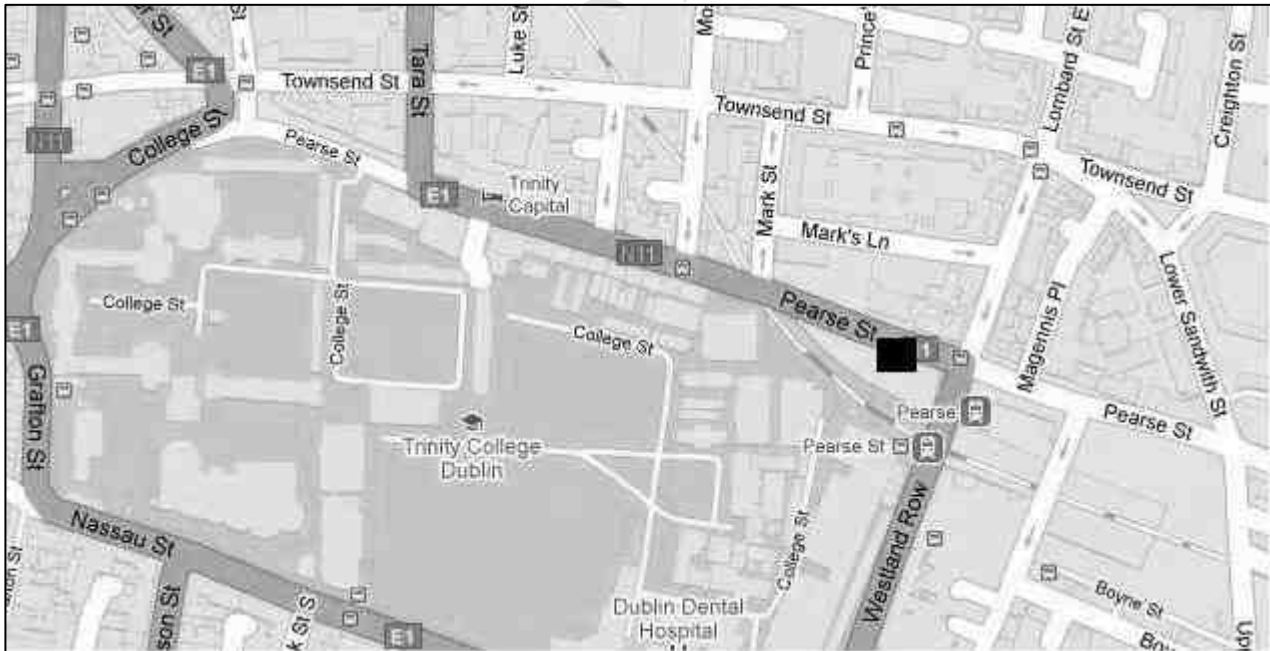


Fig. 1 Air quality monitoring station.

Pearse Street is a four-lane-one-way street situated in the centre of Dublin City, and is a major thoroughfare for traffic traversing the city centre from South to North. It has an approximate east-west orientation and carries an average daily traffic flow of 60,000 vehicles, of which 10% are HDVs. The average speed of the vehicles during the daytime is approximately 20km/hr, rising to

approximately 35km/hr at night. The street can be characterized as a street canyon, with an average building height of 16m and a width of 21m. These conditions lead to reduced pollution advection and dispersion, and to the formation of a recirculating air vortex within the canyon whenever the roof-top wind direction is perpendicular or near-perpendicular to the street orientation. Consequently, observed ambient pollution concentrations are especially sensitive to short-term variations in emissions from vehicles on the street. The air sampling point was located 3.2m from the kerbside and 1.8m above street level. The campus of Trinity College Dublin is located to the South of the monitoring site, with the remainder of the surrounding area being commercial.

Hourly concentrations in units of parts per billion (ppb) of  $\text{NO}_2$  and  $\text{NO}_x$  were observed during the summer months between May to September. In the case of an urban transport network, weekend traffic dynamics and related emission levels are inherently different from those occurring on weekdays. Hence, in this study, modeling is carried out on the data observed during weekdays only. Time-series plots of the  $\text{NO}_2$  and  $\text{NO}_x$  hourly concentration levels for 20 days are shown in Fig. 2.

Nitric oxide (NO) and  $\text{NO}_x$  concentrations were measured using an API Model 200A  $\text{NO}_x$  analyzer (Teledyne Instruments, 2008), which estimates the concentration of nitrogen dioxide ( $\text{NO}_2$ ) as the difference of these. The measurement method determines NO from the light intensity of the chemiluminiscent gas phase reaction of NO and ozone ( $\text{O}_3$ ) which produces oxygen and electronically excited  $\text{NO}_2$  molecules. The energy from these excited molecules is released as photons, whose light intensity is directly proportional to the NO concentration. Any  $\text{NO}_2$  in the sample is converted to NO by heated molybdenum and the total  $\text{NO}_x$  is measured. A sampling interval of one hour was employed.

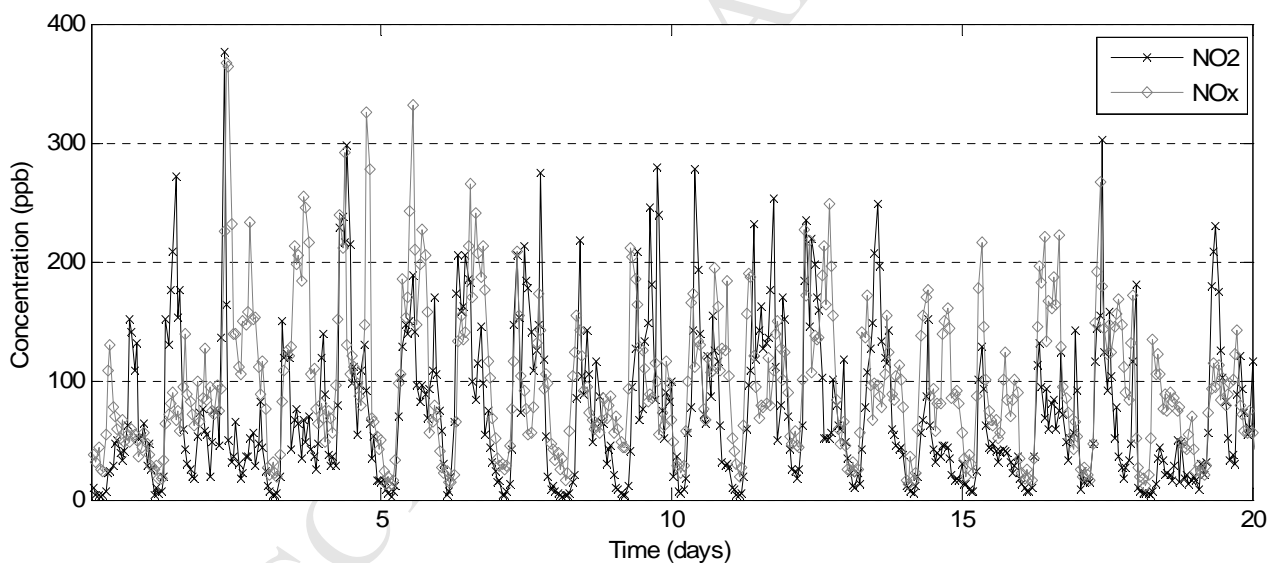


Fig. 2 Hourly concentrations of  $\text{NO}_2$  and  $\text{NO}_x$ .

Nitrogen dioxide was monitored from the 8<sup>th</sup> of May 2006 until the 7<sup>th</sup> of July 2006. 40 days of data, or 960 observations of hourly  $\text{NO}_2$  concentrations during this time-period, were chosen for modeling purposes. This allowed 39 days to be fitted to the various time-series models, with the final day being used to compare the actual observations with the forecasts. Fig. 2 shows a plot of the observed concentrations of  $\text{NO}_2$  over 20 days of the monitoring period. The mean and standard deviation of the data set are 73.4ppb and 64.4ppb respectively. According to the aforementioned EU directives (2008/50/EC Directive, Council of the European Union), the hourly concentration limit for  $\text{NO}_2$  is  $200\mu\text{g}/\text{m}^3$  or 106 ppb (annual exceedance limited to 18 hours per annum) in effect from 1<sup>st</sup> January 2010. Fig. 2 shows that multiple exceedances of this hourly limit were observed during the data collection period. Hence, it is apparent that management strategies are required to be

developed to control and limit the frequency of high nitrogen dioxide concentrations at the study site.

Nitrogen oxide concentrations were monitored from the 3rd of July 2006 until the 22<sup>nd</sup> of September 2006. 39 days were fitted to the various time-series models and the final day was used to compare the actual observations with the forecast observations. Fig. 2 shows a plot of the concentrations of NO<sub>x</sub> for 20 days during the monitoring period. The mean and standard deviation of these data are 101.5 ppb and 65.1 ppb, respectively.

The Autocorrelation Functions (ACF) of the NO<sub>2</sub> and NO<sub>x</sub> datasets are plotted in Fig. 3. It can be observed that the time-series observations of both datasets display a definitive periodicity over each 24 data-points of hourly observations, or over a day. From the plots in Fig. 3, it can also be noted that the time-series datasets are non-stationary in nature. In STS methodology, stationarity is not a requirement for modeling the time-series observations and hence unlike other existing time-series analysis techniques no further treatment is necessary to ensure stationarity of the datasets.

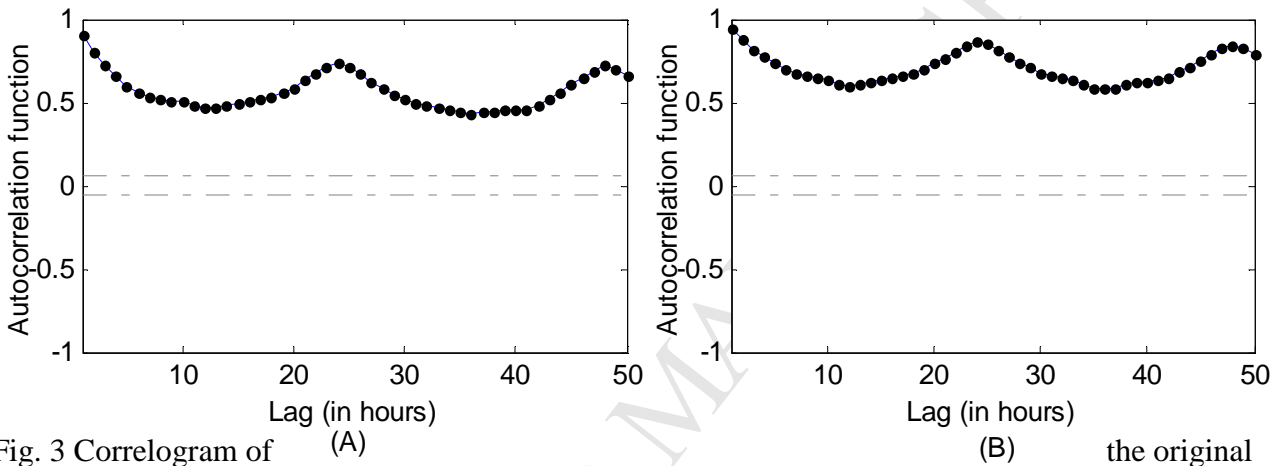


Fig. 3 Correlogram of (A) time-series observations for NO<sub>2</sub> and NO<sub>x</sub>. (B) the original

### 3.2 Modeling of NO<sub>2</sub> and NO<sub>x</sub> Concentrations (24 hours)

The time-series of NO<sub>2</sub> and NO<sub>x</sub> concentrations have been modeled using STS methodology. The use of this methodology is inspired by its successful application in traffic flow prediction (Ghosh et. al., 2009) and the fact that the measured concentrations of oxides of nitrogen were obtained at a roadside monitoring station where the pollutant levels are strongly affected by the traffic volumes. As the effect of daylight hours on atmospheric stability also influences concentration levels, a daily periodicity has been observed and modeled.

The STS models of the NO<sub>2</sub> and NO<sub>x</sub> concentration levels have been developed in a step-wise fashion. For both the datasets, the models are built up from the basic Deterministic Level Model (equation (4)) and then subsequently local level, trend, seasonal and autoregressive components are added to obtain the most suitable model (equations (2)-(3), (5)-(9)).

The elegance of the STS model lies in the meaningful depiction of the components as shown in Fig. 4. In the figure, level, seasonality and the random error components (as obtained from the Local Level with Seasonal Component Model) are shown individually for NO<sub>2</sub> concentrations in three different subplots. The local level component can be indicative of background pollution levels. Subplot (a) shows the stochastic level component as simulated and predicted by the proposed STS model. Subplots (b) and (c) of the figure show the seasonal component and the irregular component respectively, as simulated and predicted by the model. Subplot (b) shows that there exists a daily seasonality which is deterministic ( $\sigma_o^2 = 0$ ) in nature.



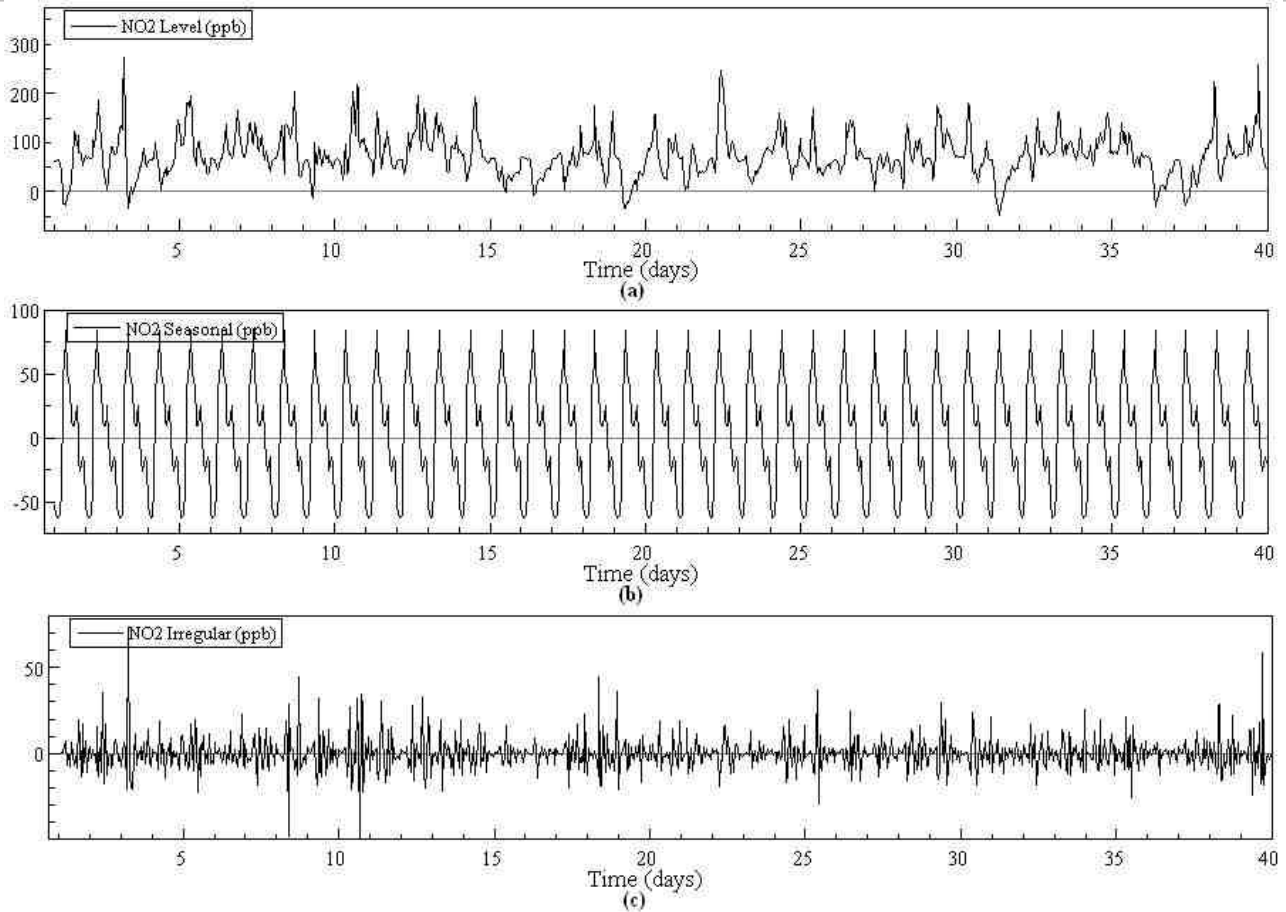


Fig. 4 (a) Stochastic level, (b) Seasonal and (c) Irregular components from the Local Level with Seasonal Model of  $\text{NO}_2$ .

At each step of the model fitting, the residuals were analyzed for any remaining correlation between the residual data-points. Apart from that, the goodness of the model fitting has been checked using the Akaike's Information Criterion. The criterion used is based on the formula

$$\text{AIC} = \log(\text{PEV}) + 2 \frac{m}{T} \quad (14)$$

where  $\text{PEV}$  is Prediction Error Variance,  $m$  is the number of stationary components in the state vector plus the number of disturbance parameters estimated and  $T$  is the number of observations in the time series. The ACF values at lag 1, 2 and 3 of the residual data-points and the AIC value at each step of the modeling are listed in Table 1 for  $\text{NO}_2$  and in Table 2 for  $\text{NO}_x$ .

Table 1 Performance Estimates of  $\text{NO}_2$  Time-Series Model

	AIC	ACF of Residuals			MAPE (%)	RMSE (ppb)
		Lag1	Lag 2	Lag 3		
Deterministic Level Model	8.34	0.59	0.29	0.18	267.78	51.07
Local Level Model	7.57	0.52	0.20	-0.12	129.23	68.71
Deterministic Linear Trend Model	8.34	0.58	0.27	0.15	262.43	51.17
Local Linear Trend Model	7.57	0.52	0.20	-0.12	129.29	68.51
Deterministic Level Model with Seasonal	7.83	0.36	0.03	0.10	43.39	27.18
Local Level Model with Seasonal	7.32	0.18	-0.09	-0.10	47.39	37.19
Deterministic Level and Seasonal Model with AR1	7.18	0.09	0.08	0.07	28.54	27.13
Deterministic Level and Seasonal Model with AR2	7.18	0.09	0.08	0.06	28.42	27.08

It can be observed from Table 1 that the AIC values for Deterministic Level and Seasonal Model with Autoregressive Components (AR1 and AR2) are the lowest making them the best-fit models to the  $\text{NO}_2$  concentration levels in ambient air. It can also be observed that the ACF values

for the residual data-points from both the models are considerably low and the residuals can be treated as white noise.

Table 2 Performance Estimates of NO<sub>x</sub> Time-Series Model

	AIC	ACF of Residuals			MAPE (%)	RMSE (ppb)
		Lag1	Lag 2	Lag 3		
Deterministic Level Model	8.36	0.68	0.46	0.40	125.69	50.88
Local Level Model	7.41	0.75	0.56	0.30	69.08	97.03
Deterministic Linear Trend Model	8.36	0.71	0.50	0.46	127.74	50.93
Local Linear Trend Model	7.41	0.75	0.56	0.30	69.32	97.30
Deterministic Level Model with Seasonal	7.51	0.16	0.07	0.01	25.43	20.92
Local Level Model with Seasonal	7.05	0.33	0.47	0.05	51.01	44.86
Deterministic Level and Seasonal Model with AR1	6.92	0.13	0.06	0.04	18.59	20.36
Deterministic Level and Seasonal Model with AR2	6.92	0.13	0.06	0.04	18.41	20.36

It can be observed from Table 2, that similar to the NO<sub>2</sub> concentration model, the AIC values for Deterministic Level and Seasonal Model with Autoregressive Components (AR1 and AR2) are the lowest making them the best-fit models for the NO<sub>x</sub> concentration levels in ambient air. It can also be observed that the ACF values for the residual data-points from both the models are considerably low and the residuals can be treated as white noise.

### 3.2.1 Prediction from STS model applied to the air pollution data (24 hours)

The NO<sub>2</sub> and NO<sub>x</sub> concentrations in ambient air are modeled using STS methodology. The effectiveness of the models is tested by comparing their prediction accuracies. For both NO<sub>2</sub> and NO<sub>x</sub>, 39 days of data are used to predict 1 day (24hours) ahead. In Fig. 5, the original observations along with the forecasts for the NO<sub>x</sub> concentration levels are shown as an example.

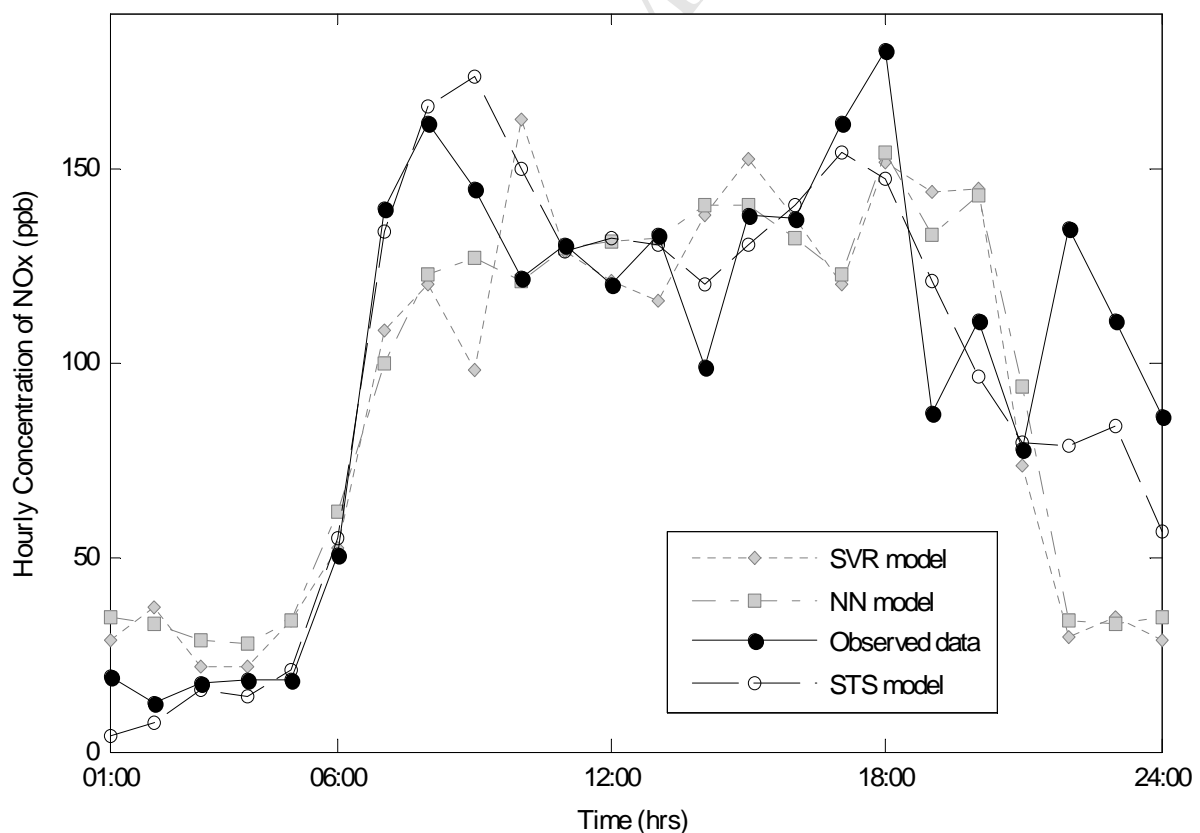


Fig. 5 Original observations and predicted values for STS model, ANN model and SVR model for NO<sub>x</sub> data.

It can be observed that the predicted values match the original observations closely. The forecasting precision of the proposed models are estimated from the MAPE and RMSE values. Tables 1 & 2 present these values for each model. Based on the MAPE, RMSE and AIC values, the Deterministic Level and Seasonal Model with Second Order Autoregression component produces the most accurate predictions for both NO<sub>2</sub> and NO<sub>x</sub>. The models produce reasonably accurate forecasts for both datasets without using any other external information such as wind speed or traffic volume, etc. The STS model developed with NO<sub>x</sub> observations provides more precise forecasts than the model developed using NO<sub>2</sub> observations. The coefficient of variation of the NO<sub>2</sub> observations is 69.2%, while for NO<sub>x</sub> observations the value is around 51%, which could be a potential reason for the lower accuracy of the STS model predictions developed for NO<sub>2</sub> observations. However, the physical phenomenon influencing this variability in precision cannot be identified from the model. As NO<sub>2</sub> is a secondary pollutant formed by photochemical reactions in the atmosphere, a more complex STS model including hourly concentrations of ozone in ambient air as an external variable might provide a more complete solution.

The predictive performance of the developed STS model has been compared with an ANN model and a Support Vector Regression (SVR) model. A back-propagation feed-forward neural network based algorithm (Cai et al., 2009) and a SVR algorithm using spline kernel (loss function:  $\epsilon$ -insensitive) (Lu and Wang, 2005) have been employed for this purpose. The prediction results from all three models are plotted in Fig. 5. Both the ANN and SVR models perform poorly compared to the STS model, especially in the off-peak periods. This could be expected because existing ANN and/or SVR models used for predicting 24 hour NO<sub>2</sub> and NO<sub>x</sub> levels often use other meteorological and traffic related information as input, which improves the accuracy of the modeled results. However, to achieve an appropriate comparison with the developed STS model, only pollutant concentration levels have been used as input to all three models. As a result, the ANN and SVR models perform less effectively. This illustrates that the STS model represents a useful technique for predicting ambient pollutant concentrations when other information is not available due to financial or operational constraints.

### 3.3 Modeling of NO<sub>2</sub> and NO<sub>x</sub> Concentrations (Peak hours)

As the highest air pollutant concentrations occur during peak traffic conditions, pollutant concentrations recorded during morning and evening peak traffic have been extracted from the same NO<sub>2</sub> and NO<sub>x</sub> data series, in an attempt to improve upon the results obtained when modeling an entire day (24 hours). Morning and evening peaks are modeled separately, with morning peak and evening peak hours considered to be from 07.00hrs to 10.00hrs and 16.00hrs to 19.00hrs, respectively. For each data set, the models are built up from the basic Deterministic Level Model (equation (4)), adding the various components to obtain an appropriate model.

For each model fitted, the residuals at each step are analyzed for any remaining correlation between data-points, and the AIC is considered to check the goodness of fit of the model. The ACF values at lags 1, 2 and 3 of the residual data-points and the AIC values for each model are listed in Tables 3-6 for the morning and evening peak of NO<sub>2</sub> and NO<sub>x</sub>.

Table 3 Performance Estimates of NO<sub>2</sub> Time-Series Model (Morning Peak)

	AIC	ACF of Residuals			MAPE (%)	RMSE (ppb)
		Lag1	Lag 2	Lag 3		
Deterministic Level Model	8.29	0.36	0.09	0.10	13.09	25.67
Local Level Model	8.26	0.17	-0.13	-0.06	11.21	16.83
Deterministic Linear Trend Model	8.28	0.36	0.10	0.11	10.68	17.08
Local Linear Trend Model	8.27	0.15	-0.15	-0.05	12.04	17.46
Deterministic Level Model with Seasonal	8.29	0.37	0.09	0.09	13.10	23.10
Local Level Model with Seasonal	8.24	0.16	-0.15	-0.08	8.91	12.23
Deterministic Level and Seasonal Model with AR1	8.10	-0.03	-0.10	0.03	12.72	22.33
Deterministic Level and Seasonal Model with AR2	8.09	-0.03	-0.10	0.03	12.99	22.69

Table 4 Performance Estimates of NO<sub>2</sub> Time-Series Model (Evening Peak)

	AIC	ACF of Residuals			MAPE (%)	RMSE (ppb)
		Lag1	Lag 2	Lag 3		
Deterministic Level Model	7.06	0.48	0.16	0.05	50.47	41.88
Local Level Model	6.99	0.17	-0.22	-0.12	29.49	29.11
Deterministic Linear Trend Model	7.07	0.48	0.15	0.05	42.19	37.76
Local Linear Trend Model	7.00	0.09	-0.23	-0.11	29.53	29.61
Deterministic Level Model with Seasonal	7.05	0.52	0.23	0.06	41.81	36.92
Local Level Model with Seasonal	6.90	0.08	-0.16	-0.10	19.74	21.20
Deterministic Level and Seasonal Model with AR1	6.71	0.02	-0.04	-0.01	23.98	26.17
Deterministic Level and Seasonal Model with AR2	6.71	0.00	-0.03	0.01	26.44	27.34

Table 5 Performance Estimates of NO<sub>x</sub> Time-Series Model (Morning Peak)

	AIC	ACF of Residuals			MAPE (%)	RMSE (ppb)
		Lag1	Lag 2	Lag 3		
Deterministic Level Model	7.65	0.54	0.27	0.21	10.30	17.36
Local Level Model	7.31	0.14	-0.20	-0.16	43.51	43.35
Deterministic Linear Trend Model	7.65	0.04	0.01	0.01	73.22	58.94
Local Linear Trend Model	7.32	0.12	-0.23	-0.18	41.68	42.20
Deterministic Level Model with Seasonal	7.62	0.60	0.42	0.23	7.32	12.21
Local Level Model with Seasonal	7.18	0.08	-0.04	-0.24	27.23	29.40
Deterministic Level and Seasonal Model with AR1	7.08	-0.06	0.06	-0.15	3.00	5.31
Deterministic Level and Seasonal Model with AR2	7.18	-0.03	-0.12	-0.11	23.58	26.37

Table 6 Performance Estimates of NO<sub>x</sub> Time-Series Model (Evening Peak)

	AIC	ACF of Residuals			MAPE (%)	RMSE (ppb)
		Lag1	Lag 2	Lag 3		
Deterministic Level Model	7.66	0.55	0.21	0.14	24.10	36.99
Local Level Model	7.46	0.14	-0.30	-0.20	29.66	52.80
Deterministic Linear Trend Model	7.62	0.56	0.22	0.16	27.76	49.79
Local Linear Trend Model	7.47	0.13	-0.31	-0.20	29.85	53.54
Deterministic Level Model with Seasonal	7.62	0.60	0.30	0.16	19.74	29.41
Local Level Model with Seasonal	7.31	0.08	-0.22	-0.21	33.65	53.70
Deterministic Level and Seasonal Model with AR1	7.14	0.04	-0.07	-0.06	21.84	31.83
Deterministic Level and Seasonal Model with AR2	7.15	0.04	-0.07	-0.02	23.85	35.11

It can be seen from Tables 3-6 that the AIC values suggest the ‘Local Level Model with Autoregressive Components (AR1 and AR2)’ to be the best fitting models to the data. The residuals obtained from these models also have considerably low ACF values and therefore these residuals can be treated as white noise.

### 3.3.1 Prediction from STS model applied to the air pollution data (peak hours)

Both AM and PM peak hour STS models for NO<sub>2</sub> and NO<sub>x</sub> concentrations in ambient air utilize 39 days of 4-hour peak period data to predict the 4 hours of peak period concentrations on the following day. The MAPE and RMSE values as shown in Tables 3-6 indicate the effectiveness of each model. Based on the MAPE and RMSE values in Tables 3 and 4, the Local Level Model with a Seasonal Component produces the most accurate predictions for the morning and evening peak period model for NO<sub>2</sub> concentrations. On the other hand, the MAPE and RMSE values suggest that the Deterministic Level Models with an Autoregressive Component produce the most accurate predictions for the morning and evening peak period model for NO<sub>x</sub> concentrations.

For illustrative purposes, the forecasts from the peak-period models for NO<sub>x</sub> concentrations have been plotted in Fig. 6 (a) and (b) along with the corresponding forecasts obtained from 24 hour models for NO<sub>x</sub> concentrations. The peak period models provide more accurate peak hour forecasts than the 24 hour models.

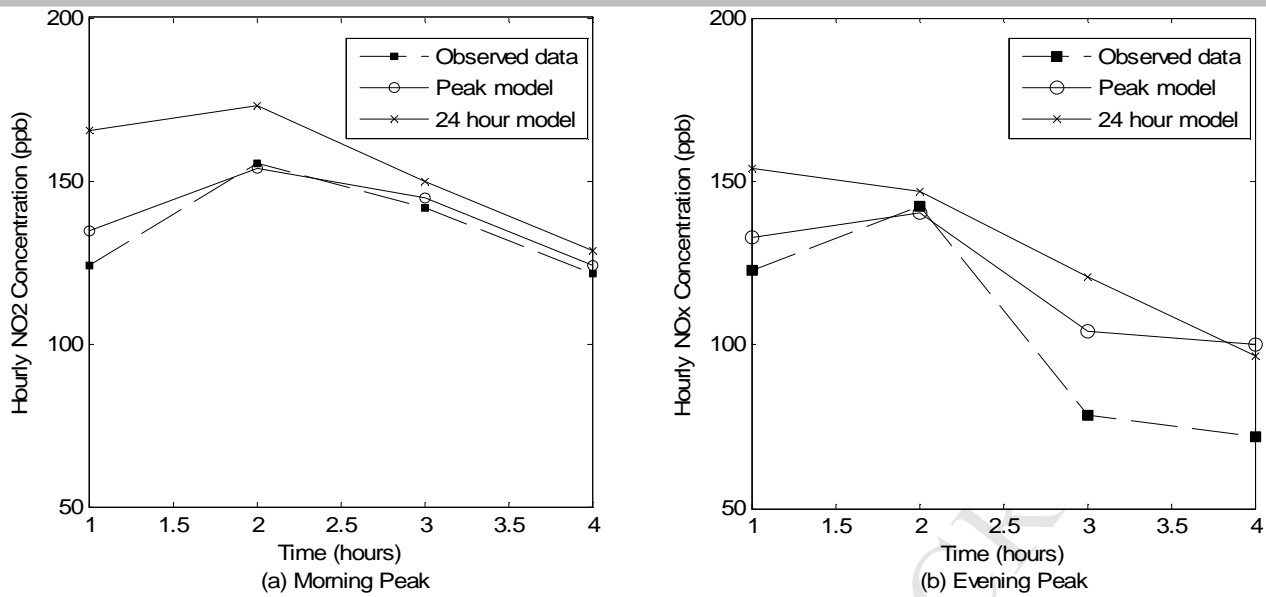


Fig. 6 Comparison of 24 hour, AM peak and PM peak models against observed  $\text{NO}_x$  peak hour data.

### 3.4 Management of Missing Values within the Air Pollution Data using STS Methodology

As problems can often occur during data collection, which may prevent data from being recorded, the modeling methodology must be capable of managing missing observations. In STS modeling using a Kalman filtering process, missing data is handled in a similar way as forecasting. To illustrate this, a 24 hour period (observations of day 33) has been removed from the  $\text{NO}_2$  concentration dataset. These 24 data points have been regarded as 'missing' observations.

Table 7 Actual Observations and Missing Observation Estimates Values for  $\text{NO}_2$

Time (Hours)	1	3	5	7	9	11	13	15	17	19	21	23
Actual Observations (ppb)	26	15.8	15.1	201	188.7	111.1	104.2	89.6	121	79.2	33	61.2
Missing Observation Estimates (ppb)	19	10.1	12.5	101	126.3	124.2	113.6	82.3	89.6	73.1	49	59.5

The missing observations have been estimated using a STS model with deterministic level, seasonal and AR2 components. The MAPE and RMSE values calculated by comparing the actual observations and missing observation estimates are 26.185% and 39.685ppb respectively.

### 3.5 Management of Outliers within the Air Pollution Data using STS Methodology

An outlier is an observation which is inconsistent with a model, usually an abnormally large value of the irregular disturbance at a specific time. STS can detect and manage outliers within the system to reduce the negative effects they can have on a forecast. As shown in Fig. 7 six outliers have been selected from the  $\text{NO}_2$  air pollution concentration time-series data. STS removes these outlier values and replaces them with more congruent values gained from applying the Kalman filtering process.

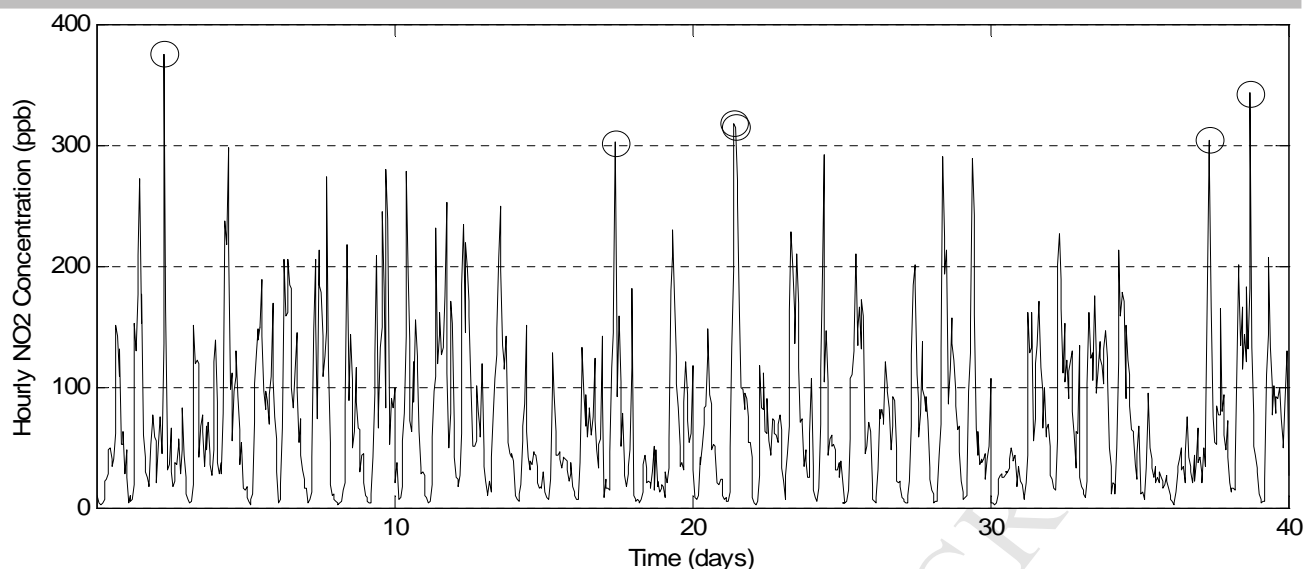


Fig. 7 Outliers within the  $\text{NO}_2$  time-series data.

#### 4. Conclusion

In this paper, a univariate STS methodology has been applied to model and predict traffic-related  $\text{NO}_x$  and  $\text{NO}_2$  hourly concentration levels at a city center road-side location. This is the first successful application of STS methodology to the prediction of ambient  $\text{NO}_x$  and  $\text{NO}_2$  concentrations. It is also the first instance of modeling 24 hour continuous air pollution data acknowledging the presence of deterministic periodicity in the data. In the developed STS model, the temporal evolution of each individual component (level, seasonality, etc.) of the pollutant concentrations can be traced separately. The STS model developed in this study has three main components, a level component, a deterministic seasonal component and an irregular component. The level component is an indicator of background concentration; the seasonal component shows the within-day dynamics of nitrogen oxide levels which are influenced mainly by the traffic flow and the day light hours. The irregular component indicates day-to-day variation which signifies or reflects the local changes in the pollutant levels. The STS model is superior to other existing statistical techniques in its meaningful depiction of the different factors which influence the variation in nitrogen oxide concentrations.

The developed model can provide one-step and multi-step ahead (a season ahead) forecasts without any major loss of accuracy. When the model considers daily seasonality, it can predict a season or 24 steps ahead into the future, owing to the fact that the forecast function solely depends on the current estimates of the seasonal effects. Apart from that, checking for stationarity is not critical for the STS models.

The model can handle missing observations and outliers without any major complications and this adds to the attractiveness of the STS model.

Overall, the analysis shows that a STS model can be a useful tool in predicting traffic-related  $\text{NO}_x$  and  $\text{NO}_2$  hourly concentrations in urban areas and is particularly useful if the information on input variables such as meteorology and emissions are not available and also if missing values or outlier values exist within the data set. The model has obvious potential benefits by offering improved air quality forecasting for environmental and public health management.

#### REFERENCES

Teledyne Instruments, 2008. Teledyne API Model 200A  $\text{NO}_x$  Analyzer Instruction Manual, 02246, Rev. G, DCN 5247.

Cai, M., Yin, Y., Xie, M., 2009. Prediction of hourly air pollutant concentrations near urban arterials using artificial neural network approach. *Transportation Research Part D: Transport and Environment*, 14, 32-41.

Chelani, A.B., 2005. Predicting chaotic time series of PM10 concentration using artificial neural network. *International Journal of Environmental Studies* 62(2), 181-191.

Council of the European Union, 2001. Directive 2001/81/EC on national emission ceilings for certain atmospheric pollutants, *Official Journal of the European Communities*.

Council of the European Union, 2008. Directive 2008/50/EC on ambient air quality and cleaner air for Europe, *Official Journal of the European Communities*.

Durbin, J. and Koopman S.J., 2001. *Time Series Analysis by State Space Methods*. Oxford Statistical Science Series, Oxford University Press.

European Environment Agency, 2009. Exceedance of air quality limit values in urban areas (version 2). European Environment Agency.

Ghazali, N.A., Ramli, N.A., Yahaya, A.S., Yusof, N.F., Sansuddin, N., Madhoun, W.A., 2010. Transformation of nitrogen dioxide into ozone and prediction of ozone concentrations using multiple linear regression techniques. *Environmental Monitoring and Assessment* 165 (1-4), 475-489.

Ghosh, B., Basu, B., O'Mahony, M., 2009. Multivariate short-term traffic flow forecasting using time-series analysis. *IEEE Transactions on Intelligent Transportation Systems* 10 (2), 246-254.

Harvey, A.C., Durbin, J., 1986. The Effects of Seat Belt Legislation on British Road Casualties: A Case Study in Structural Time-series Modeling. *Journal of the Royal Statistical Society* 149(3), 187-227.

Harvey, A.C., 1989. *Forecasting, Structural Time-series Models and the Kalman Filter*. Cambridge University Press, London.

Hoffman, S., 2006. Short-time forecasting of atmospheric NO<sub>x</sub> concentration by neural networks. *Environmental Engineering Science* 23(4), 603-609.

Hsu, K.J., 1997. Application of vector autoregressive time series analysis to aerosol studies. *Tellus*, 49B, 327-342.

Lu, W.Z. & Wang, W.J., 2005 Potential assessment of the support vector machine method in forecasting ambient air pollutant trends. *Chemosphere*, 59, 693-701.

Lu, W.Z., Wang, W.J., Wang, X.K., Xu, Z.B., Leung, A.Y.T., 2003. Using improved neural network model to analyze RSP, NOX and NO2 levels in urban air in Mong Kok, Hong Kong. *Environmental Monitoring and Assessment* 87(3), 235-254.

Minister for the Environment, Heritage and Local Government, 2002. *Air Quality Standards Regulations*. Report No. S.I. No. 271/2002, Minister for the Environment, Heritage and Local Government in Ireland.

Minister for the Environment, Heritage and Local Government, 2004. Ozone in Ambient Air Regulations. Report No. S.I. No. 53/2004, Minister for the Environment, Heritage and Local Government in Ireland.

Sanchez Gomez, M.L., Caballero Sahelices, C., Lucio Garcia, J.H, 1997. ARIMA modelling of NO, NO<sub>2</sub> and O<sub>3</sub> time series. International Conference on Measurements and Modelling in Environmental Pollution, MMEP, Proceedings, 251-260.

Schlink, U., Herbarth, O., 1997. A component time-series model for SO<sub>2</sub> data: forecasting, interpretation and modification. *Atmospheric Environment* 31(9), 1285-1295.

Sharma, P., Chandra, A., Kaushik, S.C., 2009. Forecasts using Box-Jenkins models for the ambient air quality data of Delhi City. *Environmental Monitoring and Assessment* 157(1-4), 105-112.

Simpson, D., Perrin, D.A., Varey, J.E., Williams, M.L., 1990. Dispersion modelling of nitrogen oxides in the United Kingdom. *Atmospheric Environment. Part A. General Topics* 24(7), 1713-1733.

Visser, H., Molenaar, J., 1995. Trend estimation and regression analysis in climatological time-series: an application of structural time-series models and the Kalman filter. *Journal of Climate* 8(5), 969-979.

World Health Organization, 2000. *Air Quality Guidelines for Europe, Second Edition*. WHO Regional Publications, European Series 91.

Zolghari, A., Cazaurang, F., 2005. Adaptive nonlinear state-space modelling for the prediction of daily mean PM<sub>10</sub> concentrations. *Environmental Modelling and Software* 21(6), 885-994.