ESTIMATION AND FORECASTING OF CARBON MONOXIDE EMISSION FROM MOBILE SOURCE IN METROPOLITAN AREA

Luiz Felipe Amaral
Electrical Engineering Department – Pontifical Catholic University of Rio de Janeiro, Brazil
lfelipe@ele.puc-rio.br

Reginaldo Rosa Cotto de Paula
CEFET-ES and Mechanical Engineering Department - Pontifical Catholic University of Rio de Janeiro, Brazil
rcotto@mec.puc-rio.br

Marcos Sebastião de Paula Gomes
Mechanical Engineering Department – Pontifical Catholic University of Rio de Janeiro, Brazil
mspgomes@mec.puc-rio.br

Reinaldo Castro Souza
Electrical Engineering Department – Pontifical Catholic University of Rio de Janeiro, Brazil
reinaldo@ele.puc-rio.br

Abstract. *Estimation and forescating of air quality is an important issue for environmental research due to the great impact caused by air pollution in metropolitan areas. Urban air pollution is characterized by high levels of carbon monoxide (CO) caused by transport emissions. The presence of high concentration of CO in confined spaces can cause adverse health effects. In this work, a study was performed on the vehicle emissions of CO in the Rebouças tunnel, located in the Rio de Janeiro metropolitan area, Brazil. Measurements of mean hourly CO concentrations in the tunnel were obtained on a 24-h basis for 20 days during November 2002. The goal was to apply the Box-Jenkins and Dynamic Regression models to provide an estimate and forecasting of the mean hourly CO concentration in the tunnel for investigating how people are exposed to air pollutants from mobile sources. The results of the models showed that both methodologies can be employed as a useful quantitative tool for the description of vehicular exhaust emissions of CO in the Rebouças tunnel.*

Keywords*. Vehicle emissions, Tunnel modelling, Time series model, Dynamic regression model, Forecasting.*

1. Introduction

In urban regions, motor vehicle exhaust is the major source responsible source for CO. This problem is more relevant in locations where high concentrations can be expected, such as in an urban tunnel. Tunnels are usually constructed in order to minimize traffic congestion or due to topographic constraints. However, if the ventilation system is inadequate and combined with intense vehicle traffic, the result will be elevated CO concentrations (Chan et al. 1996). Even if people do not spend very long times while driving through a road tunnel, these high concentrations can be extremely dangerous for the human health (Bellasio, 1997). For instance, CO has an affinity 220-240 times greater than oxygen to associate with hemoglobin, forming carboxyhemoglobin (COHb) (Smithline et al., 2003) in short time. The presence of COHb in the blood decreases the capacity of oxygenation of body tissues, leading to tissue hypoxia (Raub, 1999). Studies have indicated that the exposure to CO concentrations from mobile sources may be linked to health effects (Flachsbart, 1999; Atimtay et al., 2000). Very high concentrations (> 1,000 ppm) may be lethal with death resulting from asphyxiation. Low levels of concentration (several hundred ppm) after prolonged exposures may cause cardiovascular and neurological symptoms (unconsciousness, headache, fatigue, nausea, vomiting) and death.

This paper describes the forecasting of CO concentrations due to vehicle traffic inside the Rebouças tunnel, located in the metropolitan region of Rio de Janeiro, Brazil. The basic concept of forecasting is to identify a pattern or relationship between variables. After that the pattern is extended or extrapolated into the future to make a forecast. To accomplish this objective, in the present study, it was used time series models based on Box-Jenkins methodology, and causal models based in the Dynamic Regression methods. The selection and application of the proper forecast methodology can be considered an important planning and control tool for CO concentration from mobile sources (Metz and Samaras, 1994; Sharma and Khare, 2001). In this work, the idea has been to provide an analysis of hourly maximum 1-h averages of CO concentrations in the Rebouças tunnel and to establish an evidence of causal relationships between CO concentrations and vehicle flux during a condition of intense traffic. The mean hourly CO concentrations (ppm) were measured in four monitoring station and vehicle flux was measured by using loop detectors and high resolution video; both were obtained at every 1 h by recording from 6:00 h to 21:00 h. This study has the objective to analyze a quantitative tool for helping the public authorities in the development of control and prevention strategies for the CO concentration in the Rebouças tunnel during the condition of intense traffic. This is mainly due to the high CO concentration levels that were found in this site.

2. Stationarity: Unit root test

In the present work, in order to perform the forecasting using the Dynamic Regression model stationary data is required. The decision regarding stationarity in the given times series, i.e. whether or not it is possible to find a unit root in the data, has important implications in time series analysis (Dickey et al., 1986). In the literature for time series analysis, the unit root test is considered the most accurate technique to decide whether or not a time series is stationary. In the present work, the Augmented Dickey-Fuller (ADF) and Phillps-Perron (PP) tests for unit roots were applied. The ADF test (Dickey and Fuller, 1981) uses a regression model such as:

$$\nabla X_t = \alpha + \delta t + \beta y_{t-1} + \sum_{i=1}^{k} \gamma \nabla X_{t-1} + \varepsilon_t \tag{1}$$

where ∇X_{t-1} expresses the lagged first differences, ε_t adjusts the serial correlation errors and α, δ, β and γ are the parameters to be estimated. . In the present work, it was considered the ADF with intercept, i.e., in the Eq. (1), $\delta = 0$. The null hypotheses (H_o) and alternative hypotheses (H_1) for a unit root in variable X_t are: Ho: $\beta = 0$ and H_1: $\beta < 0$. The PP test (Phillps and Perron, 1988) is based on the regression:

$$X_t = \mu + \beta\left(t - \frac{T}{2}\right) + \alpha X_{t-1} + \varepsilon_t \tag{2}$$

where X_t is the variable to be tested, t and T are the time trend and the sample size, respectively, μ, β and α are parameters to be estimated and ε_t is the error process. Testing for a unit root in X_t is equivalent to testing the null hypothesis that $\alpha = 1$ against the alternative hypothesis that $\alpha < 1$. In the present work, in the Eq. (2), $\beta = 0$. If both tests reject the null hypothesis of a unit root on account of the alternative hypothesis, with determined confidence level, the series may be considered stationary. However, when the null hypothesis is accepted, the series is non-stationary.

3. Forecasting using univariate models

A time series $\{X_t\}$ is a set of observations that measures the state of some variable over time t. A time series method uses only the historical data for a variable to develop a model for predicting future values. In this work, in order to select the best time series method for the forecasting of the CO concentration, two models were evaluated: univariate Box-Jenkins (B-J) (Box, Jenkins and Reinsel, 1994) and Exponential Smoothing (E-S) (e.g., Brokcwell and Davis, 1996).

3.1. Box-Jenkins models

The Box-Jenkins methodology comprises a family of time series forecasting models. The generalized model is called Seasonal Auto Regressive Integrate Moving Average, SARIMA $(p,d,q)\times(\mathrm{P,D,Q})_s$. It consists of various possible separate model combinations: Autoregressive, AR (p), Moving Average, MA (q), combined ARMA ($p,q,$), seasonal AR $(P)_s$, seasonal MA $(Q)_s$, etc. Time series forecasting techniques aim at finding an appropriate formula so that the residuals are as small as possible and exhibit no pattern, i.e., no autocorrelation.

The B-J methodology procedures are divided into the following steps: (1) Identification: Identifies the appropriate model by plots of the original series, of the auto-correlation function (ACF) and of the partial auto-correlation function (PACF). These results usually suggest one or more models that could be fit; (2) Estimation: Estimates parameters of the selected model; (3) Diagnostic testing: Diagnosis of the fit to past data. Test for significance of autocorrelation of the residuals (Ljung-Box test). If diagnostic test indicates lack of fit is necessary to return to step 1; (4) Forecasting: Forecast future values by the fitted model of the time series that passes the diagnostic test of step 3.

In general, a univariate time series model can be modeled as a combination of the past values, X_t and past errors ε_t (with mean zero and variance σ^2), denoted as SARIMA $(p, d, q)\times(P,D,Q)_s$, is given by general formula:

$$\phi_p(B)\Phi_P(B^s)\nabla_s^D\nabla^d X_t = \theta_q(B)\Theta_Q(B^s)\varepsilon_t \tag{3}$$

where s denotes the seasonal period, B and B_s are the backward shift operators defined as: $BX_t = X_{t-1}$ and $B_sX_t = X_t - X_{t-s}$, respectively. The difference operators, ∇ and ∇_s, are given by: $\nabla X_t = X_t - X_{t-1}$ and $\nabla_s X_t = X_t - X_{t-s}$, respectively. The regular and seasonal AR operators $\phi_p(B)$ and $\Phi_P(B^s)$, are respectively polynomials of order p and P:

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p \qquad \text{and} \qquad \Phi_P(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \ldots - \Phi_p B^{Ps} \tag{4}$$

The regular and seasonal MA operators $\theta_q(B)$ and $\Theta_Q(B^s)$, are respectively polynomials of order q and Q, is given by

$$\theta_q(B) = 1 - \theta_1 B - \theta_2 B - \ldots - \theta_q B^q \qquad \text{and} \qquad \Theta_Q(B^s) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \ldots - \Theta_Q B^{Qs} \tag{5}$$

When an appropriate model has been fitted to the data, the forecasting of future values for the stationary times series can be performed. Therefore, by taking conditional expectations at time t of each term of the SARIMA model in Eq. (3) and writing $F_t = \nabla^d \nabla_s^D X_t$, the Minimum Mean Square Error (MMSE) for the predictions may be estimated as,

$$[F_{t+h}] = \phi_1[F_{t+h-1}] + \phi_2[F_{t+h-2}] + \ldots + \phi_p[F_{t+h-p}] + \varepsilon_{t+h} - \theta_1[\varepsilon_{t+h-1}] + \theta_2[\varepsilon_{t+h-2}] + \ldots + \theta_q[\varepsilon_{t+1-q}] \tag{6}$$

and the forecast for the seasonal ARIMA model may be calculated by

$$[F_{t+h}] = \phi_1'[F_{t+h-1}] + \ldots + \phi_{p+sP}'[F_{t+h-p-sP}] + \varepsilon_{t+h} - \theta_1'[\varepsilon_{t+h-1}] + \ldots + \theta_{q+sQ}'[\varepsilon_{t+1-q+sQ}] \tag{7}$$

where h = 1,2, …, is the lead time for the forecast F_{t+h}; $\phi_1', \phi_2', \ldots$ are the generalized AR parameters and $\theta_1', \theta_2', \ldots$, are the generalized MA parameters.

3.2. Forecasting using exponential smoothing method

The smoothing models are basic methods of extrapolating time series. They refer to a class of methods in which the value of a time series at some point in time is determined by the past value of the time series.

3.2.1. Single exponential smoothing (SES)

In order to obtain the forecast for actual time t, the recursive form of the SES method is given by,

$$F_t = \alpha X_t + (1 - \alpha) F_{t-1} \tag{8}$$

where X_t is the actual value, F_t is the forecast value, α is the smoothing coefficient which ranges from 0 to 1 and e_t is the forecast error at time t.

3.2.2. Holt-Winters (H-W)

The single exponential smoothing is adequate only for stationary and non-seasonal time series with no structural change (Lim and McAleer, 2001). The Holt-Winters method (Holt, 1957; Winters, 1960) is an extrapolative method that removes the effects of the level, trend, and seasonal components of a time series regardless of the nature of the time series data being collected. Forecast at time t for period $t + k$ in the H-W model is defined by:

$$F_{t+k} = E_t + kT_t \qquad \text{with} \qquad k = 1, 2, \ldots, \tag{9}$$

where E_t is the level at time t, T_t is the trend at time t and k is the any arbitrary number of periods after period t. The level E_t is given by: $E_t = \alpha X_t + (1 - \alpha)(E_{t-1} + T_{t-1})$ and the trend T_t is given by: $T_t = \beta(E_t - E_{t-1}) + (1 - \beta)T_{t-1}$, where α and β are the smoothing coefficients.

4. Causal model: Dynamic Regression

The causal model is used to determine the dynamic relationship between the series of endogenous variable (Y_t), and exogenous variables ($X_{i,t}$), i = 1,2, …, n (n being the number of the exogenous variables), the true nature of which can be complex and is frequently unknown, can be obtained by transfer function-noise model (Box et al., 1994) expressed in most general form as

$$Y_t = \frac{\omega_i(B)\Omega_i(B^s)}{\delta_i(B)\Delta_i B^s} X_{i,t-b_i} + \frac{\theta(B)\Theta(B^s)}{\phi(B)\Phi(B^s)}\varepsilon_t \tag{10}$$

where ε_t is the white noise (mean zero and variance σ^2); B and B_s are the backward shift operators; ∇ and ∇_s are the he difference operators; $\phi_p(B)$ and $\Phi_P(B^s)$ are the regular and seasonal AR operators; $\theta_q(B)$ and $\Theta_Q(B^s)$ are the regular and seasonal MA operators and others parameters in the model are defined as follows:

$$\omega(B) = \omega_o - \omega_1 B - ... - \omega_l B^l \qquad \text{and} \qquad \delta(B) = 1 - \delta_1 B - ... - \delta_r B^r \tag{11}$$

$$\Omega(B^s) = \Omega_o - \Omega_1 B^s - ... \Omega_L B^{Ls} \qquad \text{and} \qquad \Delta(B^s) = 1 - \Delta_1 B^s - ... - \Delta_R B^{Rs} \tag{12}$$

When a fitted model is obtained, the next step is to perform the forecasting of the future values. The Minimum Mean Square Error (MMSE) for forecasts can be obtained as

$$Y_{n+h} = \delta_1^* Y_{n+h} + .. + \delta_r^* Y_{n+h-r^*} + \omega_o^* X_{n+h} - \omega_1^* X_{n+h-1} - .. - \omega_{l^*}^* X_{n+h-l^*} + \varepsilon_{n+h} - \theta_1^* \varepsilon_{n+h-1} - .. - \theta_{q^*}^* \varepsilon_{n+h-q^*} \tag{13}$$

where h = 1,2, ..., is the lead time for the forecast Y_{n+h}; δ_i^*, ω_i^* and θ_i^* are the generalized parameters

5. Forecast errors

The forecast error is important to evaluate the accuracy of the forecasting procedures. The forecast error (e_t) is defined as the difference between the actual value (X_t) and the forecast value (F_t), for the same period ($e_t = X_t - F_t$). Several measures have been used to choose among alternative forecasting techniques and/or comparing the performance of a given technique (Armstrong and Collopy, 1992). In the present study, the accuracy of the forecasting techniques was evaluated by using the following measures: Mean Absolute Deviation (MAD), Mean Absolute Percentage Error (MAPE) and Geometric Mean of the Relative Absolute Error (GMRAE). Table (1) shows these measurements.

Table 1. Techniques that were applied to measure forecast error in present study.

MAD	**MAPE**	**GMRAE**
$MAD = \frac{\sum_{t=1}^{n} \lvert X_t - F_t \rvert}{n}$	$MAPE = \frac{\sum_{i=1}^{n} \left(\lvert X_t - F_t \rvert \right) / X_t}{n} \times 100$	$GMRAE = \left[\prod_{n=1}^{n} \frac{\lvert F_{m,H,s} - X_{H,s} \rvert}{\lvert F_{rw,H,s} - X_{H,s} \rvert} \right]^{1/n}$

where, in the MAD and MAPE: X_t is the actual value, F_t is the forecast of future values and n is the sample size and in the GMRAE: $X_{m,H,s}$ is the forecast from method m for horizon H of series s, $X_{H,s}$ is the actual value at horizon H of series s, $F_{rw,H,s}$ is the forecast from random walk method for horizon H of series s.

6. Rebouças tunnel description

In the present study, the time series and causal models for prediction and forecasting of the CO concentration were applied due to the vehicle intense traffic in the Rebouças tunnel. The tunnel extension is 2.8 km with two galleries (one gallery for each direction): Lagoa / São Cristovão direction (L – SC) and São Cristovão / Lagoa direction (SC – L). Forced ventilation is provided by 96 automatic blower fans located in the galleries (57 blower fans located in the L – SC direction and 39 blower fans in the SC – L direction). The control of CO in the interior of the galleries is performed and controlled during 24 hours, aiming at the fact that the users inside the tunnel have an air quality acceptable. However, due to the lack of precise information on the operation time and period for the fans, this work did not investigate the influence of the blower fans operations on the CO concentration inside the tunnel. The vehicle traffic typically through of the Rebouças tunnel has maximum speed of 90 km h^{-1}. The vehicle distribution varies depending on time of day and day of the week. Due the height restriction in the tunnel (maximum tunnel height is 9.0 m), the heavy-duty vehicles are excluded.

The data were collected during 20 days, November, 2002. The first data point correspond to 5 Tuesday and the last one to 25 Monday (except 13 Wednesday, because poor data quality). The CO concentrations (ppm) and vehicle flux were obtained every 1-h by recording from 6:00 h to 21:00 h, this period was investigated due the condition of intense traffic and

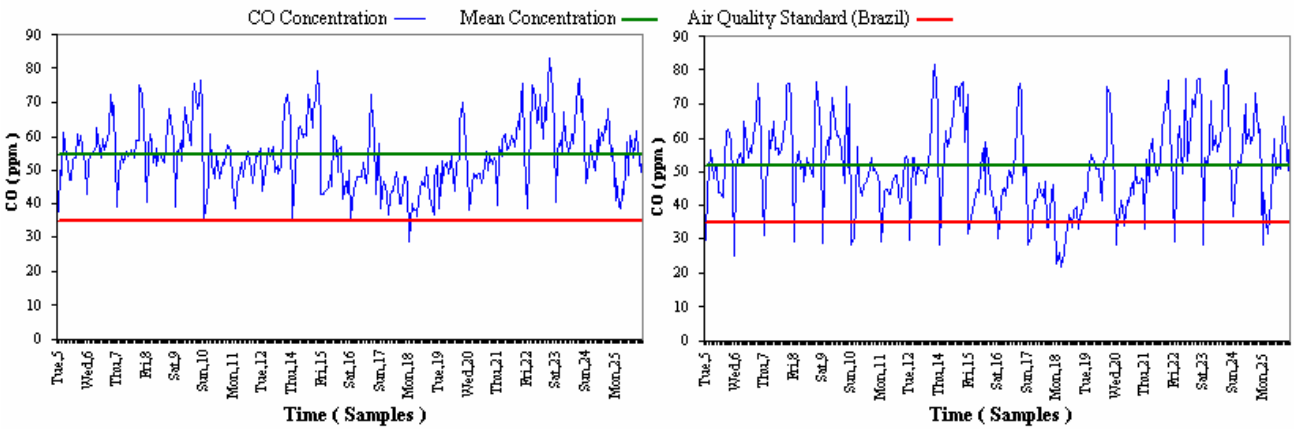
low quality of others measurements data. The CO concentrations were collected in each gallery from 4 sensors on the monitoring stations: Sensor 1 (S1) and Sensor 3 (S3) in the Lagoa / São Cristovão direction and Sensor 2 (S2) and Sensor 4 (S4) in the São Cristovão / Lagoa direction. The traffic counts were obtained from loop detectors and high resolution video observations.

7. Results and discussion

Figures (1) and (2) show the originals time series of the average CO concentrations (blue line) measurements in the sensors S1 and S3 (L - SC direction), and in the sensors S2 and S4 (SC - L direction), respectively. The green line represents the mean hourly of the CO concentrations. These results showed that the majority of the CO concentrations that were measured in the Rebouças tunnel were above of the Brazilian Air Quality Standards: 35 ppm (red line) or 4,000 $\mu g/m^3$ for 1 h average specified by CONAMA legislation. There were differences between in the magnitude of the CO concentrations observed in each gallery of the tunnel. The most significant episode was observed for the sensors S1 and S3 (L – SC direction), which the mean concentrations of the original time series were 51.81% and 48.02% greater than national standard, respectively. In this case of sensors S2 and S4 (SC – L direction) the mean concentrations of the original time series were 11.49% and 15.23% greater than national standard, respectively. Figure (3) shows the original time series of the vehicle flux collected in two galleries of the tunnel. Table (2) shows the descriptive statistics of the original time series of the vehicle flux in the Rebouças tunnel. These results of the Tab. (2) suggested that the behavior of original time series of the vehicle flux can be considered the same for both series. Therefore, the difference observed between the CO concentrations in the L - SC and SC - L directions may be attributed to the different patterns of driving and traffic conditions and, topography of the tunnel. The Rebouças tunnel in the L - SC direction has an acclivity about 4%, and this may cause congested conditions with low speeds and acceleration/deceleration events. Several published studies have demonstrated that the relations between the driver characteristics and traffic conditions can have a significant impact on vehicular exhaust emissions (Holmén and Niemeier, 1998; Lin and Niemeier, 2003). Table (3) shows the results of the ADF and PP unit root tests. The application of unit root tests suggested that all the original series are stationary, so the data series investigates can be available to modelling. This procedure was available by using of software EViews 3(1997).

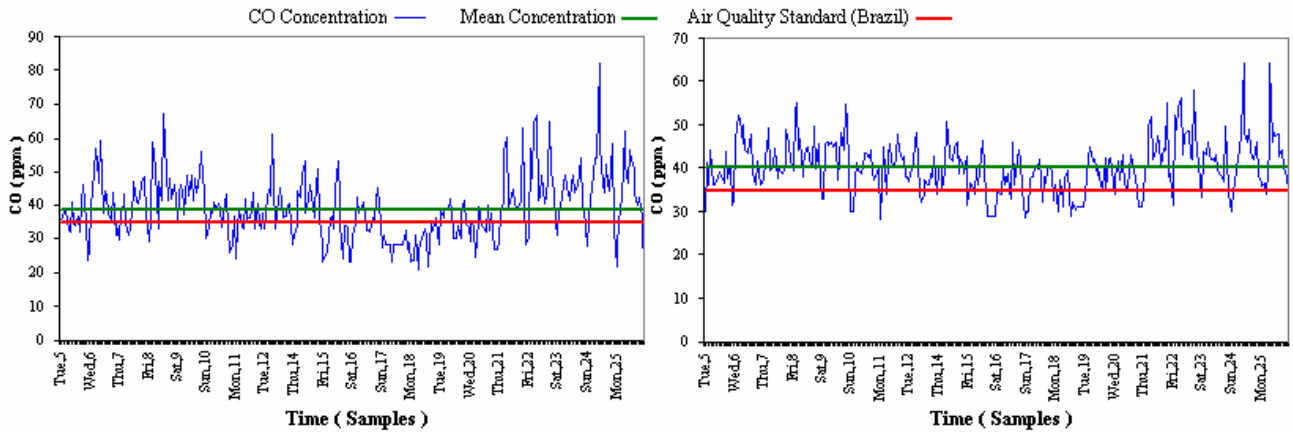
Table 2. Descriptive statistics of the original time series of the vehicle flux in the Rebouças tunnel.

VEHICLE FLUX				
Direction	**Mean**	**Maximum**	**Minimum**	**Std. Deviation**
L – SC	5112	6674	1122	1063
SC – L	5154	6701	1452	995



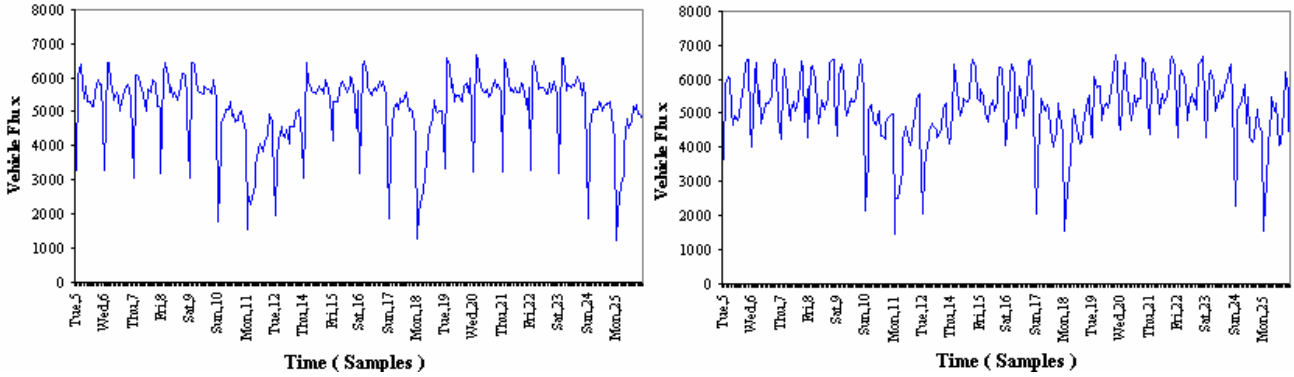
(a) (b)

Figure 1. Original time series of the CO concentrations: (a) sensor S1 and (b) sensor S3 (Lagoa / São Cristovão direction).



(a) (b)

Figure 2. Original time series of the CO concentrations: (a) Sensor S2 and (b) sensor S 4 (São Cristovão / Lagoa direction).



(a) (b)

Figure 3. Distribution of the vehicle flux in the Rebouças tunnel: (a) Lagoa / São Cristovão direction and (b) São Cristovão / Lagoa direction. Source: C.C.O. (Operational Control Center from Rebouças tunnel).

Table 3. Results of the ADF and PP unit root tests for both CO concentrations and vehicle flux time series.

CO CONCENTRATIONS SERIES						
Sensor	**t - statistic**		**Critical Value**			
	ADF	**PP**	**τ (1%)**	**τ (5%)**	**τ (10%)**	**Decision**
S1	- 5.179	- 8.432	- 3.454	- 2.871	- 2.572	H_o is rejected with 99% confidence. The series is stationary.
S3	- 5.374	- 8.217	- 3.454	- 2.871	- 2.572	H_o is rejected with 99% confidence. The series is stationary.
S2	- 5.033	- 8.618	- 3.454	- 2.871	- 2.572	H_o is rejected with 99% confidence. The series is stationary.
S4	- 5.812	- 9.486	- 3.454	- 2.871	- 2.572	H_o is rejected with 99% confidence. The series is stationary.
VEHICLE FLUX SERIES						
Direction	**ADF**	**PP**	**τ (1%)**	**τ (5%)**	**τ (10%)**	**Decision**
L – SC	- 4.847	- 9.296	- 3.454	- 2.871	- 2.572	H_o is rejected with 99% confidence. The series is stationary.
SC – L	- 4.535	- 7.804	- 3.454	- 2.871	- 2.572	H_o is rejected with 99% confidence. The series is stationary.

7.1. Results of forecasting of the CO concentration by using Box-Jenkins models

Once the stationary behaviors of the series were identified, the next step was to apply the Decomposition Analysis, in order to identify the various patterns that can appear simultaneously in the time series of CO concentrations. Table (4) shows the results of the classical Decomposition Analysis (X_t = TC + S + I), where TC, S and I are trend-cycle, seasonal and irregular components, respectively.

The analysis of the classical decomposition in Tab. (4) shows the various patterns of the CO concentrations time series. The original time series of the CO concentrations for S1 showed approximately identical pattern of trend-cycle and seasonal, 36.43(%) and 35.34(%) respectively. The CO concentrations series for S3 was more seasonal with 41.97% of the seasonality. In the case of sensors S2 and S4 the CO concentrations series were most irregulars, with 44.48% and 53.06%, respectively. Another important result of the classical decomposition is that sensors located in the same gallery of the tunnel showed approximately identical pattern in each analyzed component.

The next step was to take the decision of a choice between the Box-Jenkins (BJ) or Exponential Smoothing (ES) methods to forecast. This procedure was obtained by FPW (Forecast Pro for Windows) program. The results were: (1) Sensor S1: B-J methods outperform E-S methods by 4.499 to 5.217 in the out-of-sample MAD; (2) Sensor S3: B-J methods outperform E-S methods by 4.822 to 6.929 in the out-of-sample MAD; (3) Sensor S2: B-J methods outperform E-S methods by 9.490 to 10.312 in the out-of-sample MAD and (4) Sensor S4: B-J methods outperform E-S methods by 4.538 to 6.155 in the out-of-sample MAD.

Table 4. Classical decomposition multiplicative of original time series of the CO concentrations.

	Component (%)		
Sensor	**Trend-cycle (TC)**	**Seasonal (S)**	**Irregular (I)**
S1	36.43	35.34	28.24
S3	33.30	41.97	24.73
S2	37.99	17.53	44.48
S4	28.44	18.50	53.06

Out sample analysis consist in a way to know if the model is making a reasonable forecast, i.e., some last data (in the present work the last data were 15 observations) are (extracted) from the sample, in order to test prediction power of the model. The results of the out-of-sample forecast suggested that the model recommended to forecast were the B-J models. Therefore, E-S models were not selected for the forecast when using the FPW program. In order to reduce the heteroscedasticity (i.e., the variance of the time series isn't a constant), a log-transformation was applied for all time series of the CO concentrations. This transformation was obtained by using of the Box-Cox transformation, $T_\lambda(X_t)$ (e.g., Brocwell and Davis, 1996), where $T_\lambda(X_t) = \ln X_t$ with $X_t > 0$ and $\lambda = 0$. The following step was the identification of an appropriate model to forecast. In the present study was tested various Box-Jenkins models and several of them were found to be adequate. Table (5) shows the results of the best SARIMA $(p, d, q) \times (P,D,Q)_s$ models fitted to forecast of the CO concentrations. These models showed good significance and were considered adequate in the modeling procedure.

Once the best selected models for CO concentrations were fitted, the next step was to make analysis of the statistics using in-sample and out-of-sample statistics. The results of the in-sample MAPE showed that all models were responsible for mistakes, in the mean on the order of about 10%. The best fitted model was the sensor S1 (error of 8%) and the worst fitted model was the sensor S2 (error of 13%). The results of the out-of-sample MAPE also showed that all models were causing mistakes, in the mean on the order of about 12%. In the case, the best fitted model was the sensor S4 (error of 10%) and the worst fitted model was the sensor S3 (error of 14%). The results of MAPE can be considered reasonably well. Therefore, how the original series CO concentration showed a pattern with significant irregularity in the analyzed of classical decomposition, the results of these statistics can be considered very well, this suggested that all fitted models have good predictive capacity.

The Ljung-Box statistic Q_{LB} formulated by Ljung and Box (1978) was used to test the hypothesis that the residuals do not present autocorrelation. This statistic, for specific lags, follows a chi-square distribution with degrees of freedom equal to the order of the maximum lag minus the number of the estimated coefficients. The results for lag 1 to 18 showed that Q_{LB} and for sensors S1, S3, S2 and S4, were statistically significant as shown in Tab. (6). These results also suggested that significant autocorrelation was not found in the residuals of the selected models. Table (6) summarizes the results of the in-sample and out-sample statistics, respectively. In the Tab. (6) H is the forecast horizon and N is the forecasting quantities.

The seasonal index represents the extent of the seasonal influence for a particular segment of the series data. The assumption involves a comparison of the expected values of that period to the mean. A seasonal index indicates how much the average for that particular period tends to be above (or below) the mean. Figures (4) and (5) show the results of the seasonal index, S_i, for CO concentrations that were measured in the Rebouças tunnel during all investigated period (6:00 h to 21:00 h). In the case for sensors S1 and S3 the maximum and a minimum S_i occurred in the same period. This behavior

suggested that the critical situation occur during 18:00 h to 19:00 h. In the case of sensors S2 and S4 the maximum and a minimum S_i occurred in the same period and the critical situation occur during 10:00 h to 11:00 h.

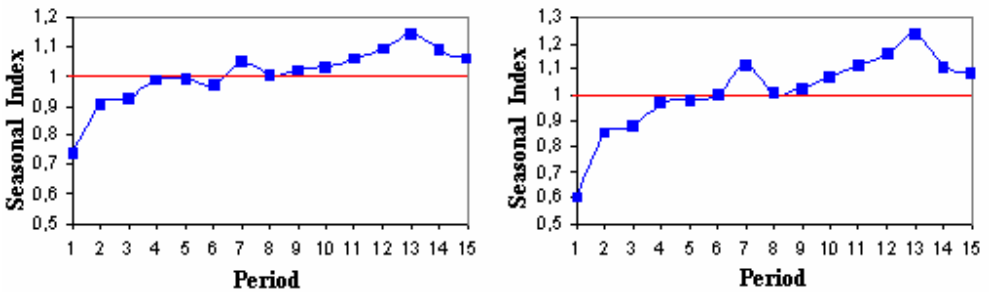
Figures (6) and (7) show the predictions using the B-J (red line) and the observed series (black line) during the selected period. These results suggested that the fifteen-step ahead forecast of the CO concentrations was in good agreement with the observed series, because fitted models capture well the trend and seasonality of the observed series. However, the predicted series tends to sub-estimate the observed series, but, the performance of the trend and seasonality of B-J models can be considered reasonably good in the forecast of CO concentrations in the Rebouças tunnel.

Table 5. Box-Jenkins models selected to perform the forecasting of the CO concentrations in the Rebouças tunnel.

CO CONCENTRATIONS				
Forecast model for Sensor 1 ⇒ SARIMA $(2,0,1)*(1,0,1)_s$				
Parameter	**Coefficient**	**Standard Error**	**t-Statistic**	**Significance**
AR [1]	1.3502	0.1076	12.5457	1.0000
AR [2]	- 0.3711	0.0972	- 3.8185	0.9999
MA [1]	0.8061	0.0800	10.0749	1.0000
AR_s [15]	0.9996	0.0000	20123.9289	1.0000
MA_s [15]	0.9270	0.0169	54.7652	1.0000
Forecast model for Sensor 3 ⇒ SARIMA $(1,0,0)*(1,0,1)_s$				
Parameter	**Coefficient**	**Standard Error**	**t-Statistic**	**Significance**
AR [1]	0.7693	0.0379	20.3169	1.0000
AR_s [15]	0.9991	0.0001	19715.2735	1.0000
MA_s [15]	0.9198	0.0221	41.5836	1.0000
Forecast model for Sensor 2 ⇒ SARIMA $(1,0,0)*(2,0,0)_s$				
Parameter	**Coefficient**	**Standard Error**	**t-Statistic**	**Significance**
AR [1]	0.5762	0.0490	11.7625	1.0000
AR_s [15]	0.1352	0.0598	2.2601	0.9762
AR_s [30]	0.2688	0.0585	4.5942	1.0000
Forecast model for Sensor 4 ⇒ SARIMA $(1,0,0)*(2,0,0)_s$				
Parameter	**Coefficient**	**Standard Error**	**t-Statistic**	**Significance**
AR [1]	0.5294	0.0500	10.5783	1.0000
AR_s [15]	0.1385	0.0584	2.3718	0.9823
AR_s [30]	0.1993	0.0586	3.4009	0.9993

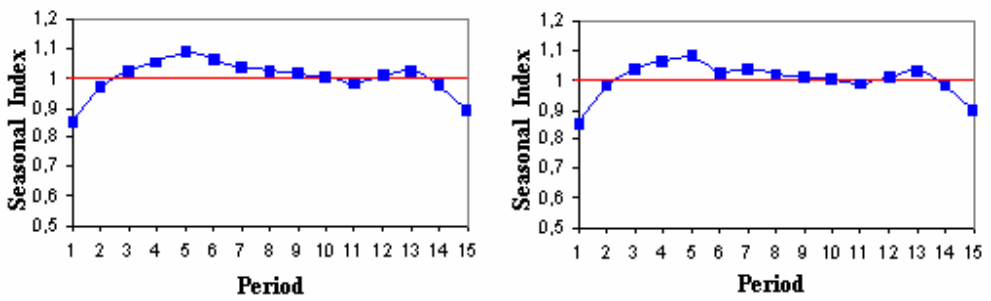
Table 6. Statistics of residuals of the B-J models that were selected to forecast of time series of the CO concentrations in the Rebouças tunnel.

	IN-SAMPLE			OUT-OF-SAMPLE			
				MAPE (%)		GMRAE	
Sensor	**MAPE (%)**	**Ljung-Box (18)**	**p - value**	**H=1 and N=15**	**Cumulative**	**H=1 and N=15**	**Cumulative**
S1	7.89	28.23	0.9413	11.1	10.9	0.739	0.662
S3	10.36	27.56	0.9310	14.2	10.7	0.818	0.377
S2	13.45	25.53	0.8891	13.2	15.6	0.378	0.388
S4	9.41	28.02	0.9383	10.0	10.4	1.082	0.718



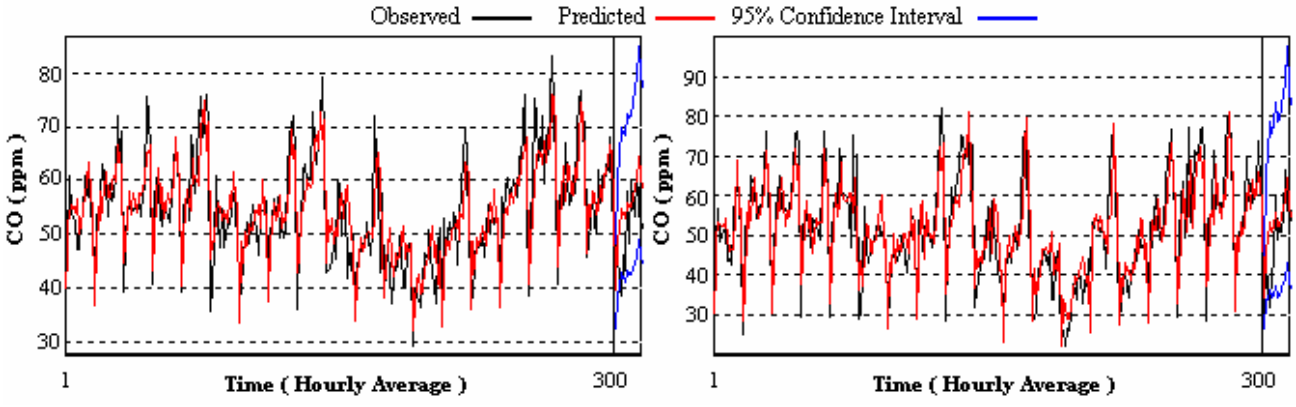
(a) (b)

Figure 4. Seasonal index: (a) Sensor S1 and (b) Sensor S3 (Lagoa / São Cristovão direction).



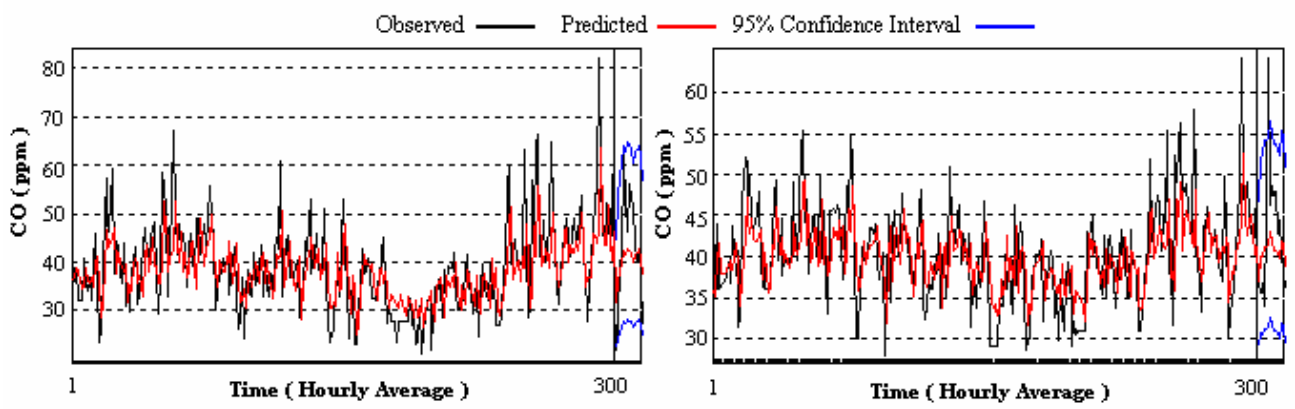
(a) (b)

Figure 5. Seasonal index: (a) Sensor S2 and (b) Sensor S4 (São Cristovão / Lagoa direction).



(a) (b)

Figure 6. Time series of observed and predicted hourly maximum CO concentrations: (a) Sensor S1 and (b) Sensor S3.



(a) (b)

Figure 7. Time series of observed and predicted hourly maximum CO concentrations: (a) Sensor S2 and (b) Sensor S4.

7.2 Results of forecasting of the CO concentrations and vehicle flux by using Dynamic Regression models

In order to perform the causal relationship between CO concentrations and vehicle flux in the Rebouças tunnel, in this work it was used the Dynamic Regression models. Table (7) shows the results of the selected Dynamic Regression models. The best model selected for sensor S4 presented 94% of significance for the DUMMY term in the t-test. It is well known that the significance levels should be 95%, however, this doesn't compromise the analysis. A dummy variable is a numerical variable used in the regression analysis to represent subgroups of the sample. These dummy variables were introduced to capture the seasonality during rush hours.

The results of the in-sample and out-of-sample statistics were similar to those obtained with the B-J models. Hence, it can be conclude that all models were responsible for mistakes in the mean on the order of about 10% on the in-sample MAPE. The best fitted model was the sensor S1 (error of 8%) and the worst fitted model was the sensor S2 (error of 13%). The results of the out-of-sample MAPE also showed that all models were causing mistakes, in mean, around 11%. The best fitted model was the sensor S4 (error of 9%) and the worst fitted model was the sensor S2 (error of 13%). The results of the Q_{LJ} suggested that the selected causal models, due to the relationship between CO concentrations and vehicle flux for S1, S3, S2 and S4, showed that a significant autocorrelation was not found in the residuals of the selected models as shown in Tab. (8). In Table (8) H is the forecast horizon and N is the forecasting quantities.

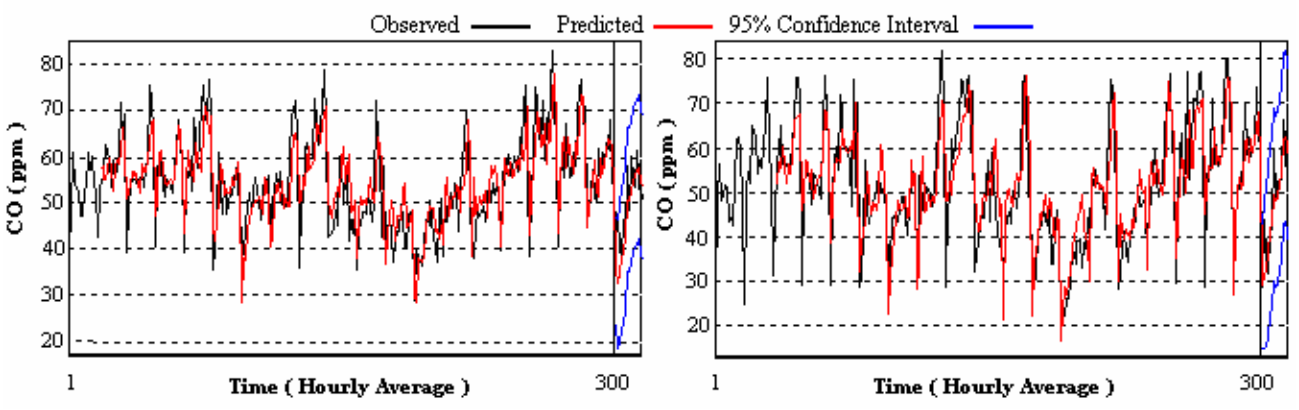
Figures (8) and (9) show the predictions using the Dynamic Regression models (red line) and the series observed (black line) during the selected period analyzed. These results showed that the fifteen-step ahead forecast of the causal relationship between CO concentrations and vehicle flux were in relatively good agreement. The fitted models in this modelling captures well the trend and seasonality of the observed series and present a good performance. These results suggested that Dynamic Regression models can be applied to forecast the relation between investigated variables in the Rebouças tunnel.

Table7. Dynamic Regression models selected to perform the forecasting of the causal relationship between CO concentrations and vehicle flux in the Rebouças tunnel.

SENSOR 1				
Parameter	**Coefficient**	**Std. Error**	**t-Statistic**	**Significance**
Vehicle traffic	0.0044	0.0005	9.7731	1.0000
CO concentration [-15]	0.2392	0.0519	4.6069	0.9999
CO concentration [-1]	0.3443	0.0451	7.4649	1.0000
Vehicle traffic [-1]	0.3096	0.0767	4.0388	0.9999
Vehicle traffic [-2]	0.1513	0.0615	2.4582	0.9860
SENSOR 3				
Parameter	**Coefficient**	**Std. Error**	**t-Statistic**	**Significance**
DUMMY [-1]	4.3940	1.4012	3.1358	0.9983
Vehicle traffic	0.0055	0.0005	10.3847	1.0000
CO concentration [-1]	0.2232	0.0451	4.9491	0.9999
CO concentration [-30]	0.2121	0.0548	3.8677	0.9998
Vehicle traffic [-1]	0.3952	0.0775	5.1018	1.0000
Vehicle traffic [-2]	0.2096	0.0654	3.2044	0.9986
SENSOR 2				
Parameter	**Coefficient**	**Std. Error**	**t-Statistic**	**Significance**
Vehicle traffic	0.0027	0.0006	4.3681	0.9999
DUMMY	2.3122	1.1010	2.1001	0.9643
Const.	25.1226	3.4674	7.2453	1.0000
Vehicle traffic [-1]	0.5323	0.0503	10.5778	1.0000
Vehicle traffic [-14]	0.2089	0.0554	3.7680	0.9998
SENSOR 4				
Parameter	**Coefficient**	**Std. Error**	**t-Statistic**	**Significance**
Vehicle traffic	0.0019	0.0004	4.9080	0.9999
DUMMY	1.2709	0.6702	1.8963	0.9421←
Const.	29.5545	2.1139	13.9808	1.0000
Vehicle traffic [-1]	0.4553	0.0592	7.6844	1.0000
Vehicle traffic [-2]	0.1658	0.0592	2.7997	0.9949

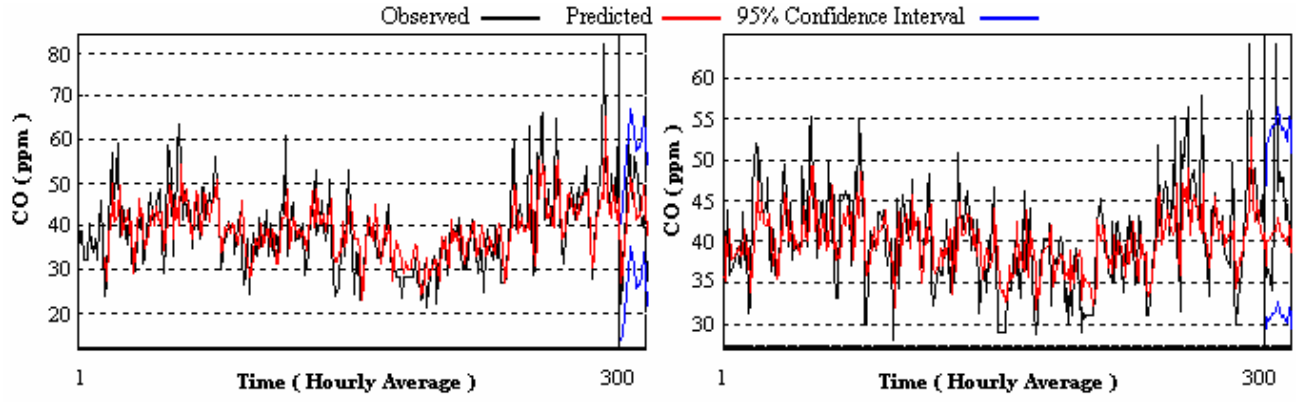
Table 8. Statistics of residuals of the causal models that were selected to forecast of CO concentrations and vehicle flux in the Rebouças tunnel.

	IN-SAMPLE			**OUT-OF-SAMPLE**			
				MAPE (%)		**GMRAE**	
Sensor	**MAPE (%)**	**Ljung-Box (18)**	**p – value**	**H=1 and N=15**	**Cumulative**	**H=1 and N=15**	**Cumulative**
S1	8.51	27.78	0.9345	11.2	7.0	0.889	0.400
S3	11.53	16.35	0.4318	10.8	7.5	0.489	0.204
S2	13.11	20.58	0.6988	12.7	16.5	0.726	0.519
S4	9.04	24.22	0.8521	9.1	11.8	0.802	0.743



(a) (b)

Figure 8. Observed and predicted causal relationship between CO concentrations and vehicle flux: (a) S1 and (b) S3.



(a) (b)

Figure 9. Observed and predicted causal relationship between CO concentrations and vehicle flux: (a) S2 and (b) S4.

8. Conclusions

The application of time series and causal models for assessing the CO vehicular exhaust emissions in an urban tunnel was conducted. CO concentrations were measured at four monitoring stations in intense traffic conditions (6:00 h to 21:00 h). The original time series showed high CO concentration levels in the Rebouças tunnel. In many situations, these concentrations were above of the national air quality standard. This can be attributed to intense vehicle traffic and reduced natural ventilation in this site. In both tunnel galleries the analysis of original time series of the vehicle flux suggested that identical behavior occurred. However, the worst-case scenario for the original time series of the CO concentrations occurred in the L – SC direction, where the national air quality standard was significantly exceeded. This behavior may be related to driving and traffic conditions and tunnel characteristics on the gallery in the L – SC direction.

The results of the predicted series using the B-J models and the original time series of the CO concentrations showed good agreement, because fitted models capture well the trend and seasonality of the observed series. However, the predicted series tends to sub-estimate the original time series of the CO concentrations. Therefore, the performance of B-J models can be considered reasonably good to forecast CO concentrations in the Rebouças tunnel.

The results of the predictions using the Dynamic Regression models and the original series showed that the causal relationship between the CO concentrations and the vehicle flux were relatively good. The fitted models in this modelling captures well the trend and seasonality of the observed series and present a good performance. However, there was no improvement in the modelling of the CO concentrations with the inclusion of vehicle flux as an input series data. This model has the ability of identifying that vehicle traffic influenced the high CO concentration levels in the Rebouças tunnel. In future works to improve the forecasting ability it could be necessary to identify other important factors that induces CO levels, such as meteorology conditions, fleet characteristics, and vehicle speed.

The results suggested that both methodologies can be employed as useful quantitative tools in the description of the CO concentrations in the Rebouças tunnel, mainly in the absence of information on the atmospheric conditions. Particularly, they may help the public authorities in the control and prevention of the high CO concentration levels from mobile sources.

9. Acknowledgement

The authors acknowledge Rui Moreira Ribeiro Filho and Wagner Santos Batista from C.C.O – Rebouças tunnel. The authors acknowledge support from CNPq and IIF (International Institute of Forecasters).

10. References

Armstrong, J.S. and Collopy, F., 1992, "Error measures for generalizing about forecasting methods: Empirical comparisons", Vol. 8, pp. 69-80.

Atimtay, A.T., Emri, S., Bagci, T. and Demir, A.U., 2000, "Urban CO exposures and Its Health Effecst on Traffice Policemen in Ankara", Environmental Research, Vol. 82, pp. 222-230.

Bellasio, R., 1997, "Modelling traffic air pollution in road tunnels", Atmospheric Environment, Vol. 31, pp. 1539-1551.

Bos, G.E.P. and Jenkins, G.M., 1976, "Time series analysis: forecasting and control", Holden Day, San Francisco.

Box, G.E., Jenkins, G.M., and Reinsel, G.C., 1994, "Time Series Analysis: Forecasting and Control", 3rd ed. Prentice Hall, Englewood Cliff.

Brockwll, P.J. and Davis, R.A., 1996, "Introduction to Time Series and Forecasting", Springer-Verlag, New York.

Chan, L.Y., Zeng, L., Oin, Y. and Lee, S.C., 1996, "CO concentration inside the Cross Harbor Tunnel in Hong Kong", Environment International", Vol. 22, pp. 405-409.

Dickey, D.A., Bell, W.R. and Miller, R.B., 1986, "Unit Root in time series models: test and implications". American Statist, Vol. 40, p. 12.

Dickey, D.A. and Fuller, W.A., 1981, "Likelihood ratio statistics for autoregressive time series with a unit root, Econometrica, Vol. 49, pp. 1057-1072.

Flachsbart, P.G., 1999, "Human exposure to carbon monoxide from mobile sources", Chemosphere – Global Change Science, Vol. 1, pp. 301-329.

Geurts, M. and Ibrahim, I., 1997, "EViews 3, Quantitative Micro Software. Irvine CA: EViews 3.

Holmén, B.A., and Niemeier, D.A., 1998, "Characterizing the effects of driver variability on real-word vehicle emissions, Transportation Research Part D: Transport and Environment, Vol. 3, pp. 117-128.

Holt, C.C., 1957, "Forecasting seasonals and trends by exponentially weighted moving averages", Carnegie Institue of Technology, Pittsburgh, PA.

Lin, J. and Niemeier, D.A., "Regional driving characteristics, regional driving cycles", Transportation Research Part D: Transport and Environment, Vol. 8, pp. 361-381.

Ljung, G.M. and Box, G.E., 1978, "On a measure of lack of fit in time series models", Biometrica, Vol. 65, pp. 297-303.

Metz, N. and Samaras, Z., 1994, "Comparative assessment of two forecasting models for road traffic emissions: a German case study", The Science of The Total Environmental, Vols. 146 and 147, pp. 339-349.

Phillips, P.C.B. and Perron, P., 1988, "Testing for a unit root in time series regression", Biometrica, Vol. 75, pp. 335–346.

Sharma, P. and Khare, M., 2001, "Short-term, real-time prediction of the extreme ambient carbon monoxide concentrations due to vehicular exhaust using transfer funciton-noise model", Transportation Research Part D: Trnasport and Environment, Vol. 6, pp. 141-146.

Smithline, H.A., Ward, K.R., Chiulli, D.A., Blake, H.C. and Rivers, E.P., 2003, "Whole body oxygen consumption and critical oxygen delivery in response to prolonged and severe carbon monoxide poisoning", Resuscitation, Vol. 56, pp 97-104.

Winters, P.R., 1960, "Forecasting sales by exponentially weighted moving averages", Management Science, Vol. 6, pp. 324-342.

11. Copyright Notice