# Real Time Monitoring of Carbon Monoxide Using Value-at-Risk Measure and Control Charting

Bersimis, Sotirios and Degiannakis, Stavros and Georgakellos, Dimitrios

Department of Statistics and Insurance Science, University of Piraeus, Greece, Department of Economic and Regional Development, Panteion University, Greece, Department of Business Administration, University of Piraeus, Greece

1 January 2017

# Real Time Monitoring of Carbon Monoxide Using Value-at-Risk Measure and Control Charting

Sotirios Bersimis[1], Stavros Degiannakis[2*], Dimitrios Georgakellos[3]

[1] *Department of Statistics and Insurance Science, University of Piraeus, Greece*

[2] *Department of Economic and Regional Development, Panteion University, Greece*

[3] *Department of Business Administration, University of Piraeus, Greece*

***Abstract***

One of the most important environmental health issues is air pollution, causing the deterioration of the population's quality of life, principally in cities where the urbanization level seems limitless. Among ambient pollutants, carbon monoxide (CO) is well known for its biological toxicity. Many studies report associations between exposure to CO and excess mortality. In this context, the present work provides an advanced modelling scheme for real time monitoring of pollution data and especially of carbon monoxide pollution in city level. The real time monitoring is based on an appropriately adjusted multivariate time series model that is used in finance and gives accurate one-step-ahead forecasts. On the output of the time series, we apply an empirical monitoring scheme that is used for the early detection of abnormal increases of CO levels. The proposed methodology is applied in the city of Athens and as the analysis revealed has a valuable performance.

**Keywords:** Air Quality Surveillance, Atmospheric Pollution, Autoregressive Conditional Heteroskedasticity modelling, Control Charts, Diag-aVECH, Multivariate Statistical Process Monitoring, Multivariate Time Series, Value-at-Risk.

[*] Coresponding author. E-mail address: *s.degiannakis@panteion.gr*

## 1. Introduction

Even if air quality has been significantly improved, air pollution currently remains one of the most important environmental health issues (Min *et al.*, 2009). Actually, urbanization, which is the result of rapid economic growth, causes serious air pollution related problems in many areas all over the world. According to a recent estimation of the World Health Organization (WHO), almost 1.4 billion urban residents in the developing countries respire exceeding air quality guidelines (Gokhale and Khare, 2005). Based on a similar estimate of the United Nations, more than 600 million people in urban areas worldwide are exposed to dangerous levels of air pollutants, most of them traffic-generated. Subsequently, the quality of the air (indoors and outdoors) affects the morbidity and mortality resulting from respiratory and cardiovascular diseases (Han and Naeher, 2006).

In mega cities, such as Bombay, Calcutta, Delhi, Dhaka, Karachi, Bangkok, Beijing, Shanghai, Jakarta and Manila, where the pollution levels often exceed the WHO air quality guidelines by a factor of 3 or 4, the mortality due to outdoor air pollution is ranging between 0.4–1.1% of the total annual deaths (Gokhale and Khare, 2005). Among common air pollutants that draw intense concerns is carbon monoxide (CO), which is known for its biological toxicity (Han and Naeher, 2006). Many studies report associations between exposure to CO and mortality and hospital admission. Population-based and susceptible panel research findings suggest that CO and other traffic related pollutants may alter cardiac autonomic regulation through limiting oxygen carrying capacity of haemoglobin (Tao *et al.*, 2011).

Therefore, increased concern over the adverse health effects of air pollution has highlighted the need for air pollution measurements, especially in urban areas, where many sources of air pollutants are concentrated (Chaloulakou *et al.*, 2003a). Specifically, ecological and environmental monitoring has become increasingly important. Monitoring usually involves sampling from several sites of a similar habitat at regular (or irregular) intervals through time. The purpose of monitoring is to determine where and when an impact may have occurred or, once detected, may still be occurring. Moreover, various statistical methods attempt to provide a way of identifying when an environmental system is going “out-of-control”, so as to employ appropriate remedial measures (Anderson and Thompson, 2004).

In this context, the present work provides an advanced forecasting scheme for real time monitoring of carbon monoxide pollution in city level. The real time monitoring scheme uses an appropriately adjusted multivariate time series model that comes from the area of financial modelling. This time series model succeeds accurate one-step-ahead forecasts. These forecasts are then feed in a control chart which early detects abnormal increases of CO levels. Early signals of abnormal increases of CO levels can be used for public protection. An application of this scheme is presented in the city of Athens, Greece.

The next section highlights the main issues of the examined problem, including a concise overview of carbon monoxide pollution in general, as well as a more specific analysis of CO exposure in the greater Athens' area. The 3rd section introduces the framework for the real time monitoring of urban pollutants in many stations across time. A multivariate ARCH model[1] specification and the VaR measure (a very popular tool in financial literature) are employed for the time series modelling of the air pollution variables. Afterwards, in the monitoring phase, a multivariate Statistical Process Monitoring technique is illustrated including two statistics for detecting possible global changes and local changes. Subsequently, section 4 examines the application, whereas section 5 presents the concluding remarks.

## 2. Problem Identification

### 2.1. Carbon monoxide pollution

Carbon monoxide is one of the main reactive trace gases in the earth's atmosphere: it influences both the atmospheric chemist O and the climate (Badr and Probert, 1994). The natural background levels of CO, in areas away from urban centres and human activities, are in the order of 60–70 ppb in the Southern Hemisphere and 120–180 ppb in the Northern Hemisphere (Georgoulis *et al.*, 2002; Choi and Chang, 2006). It is an outcome of natural tropical forest fires and oxidation of biogenic hydrocarbon from plants while ocean is known to be a natural source of CO as well (Asatar and Nair, 2010). CO is primarily generated by motor vehicle emission, which accounts for an estimated 89% of CO emissions from anthropogenic sources in developed countries. Therefore, CO can be used as a marker for the contribution of traffic to air pollution (Bel *et al.*, 2015, Potoglou and Kanaroglou, 2005). The atmospheric lifetime of CO is relatively long (3 months approximately) and it can be transported in global scale (Peng *et al.*, 2007). Furthermore, CO is an intermediary in determining the future concentrations of many environmentally important trace gases such as methane and hydrochlorofluorocarbons (US EPA, 2000). Thus, carbon monoxide monitoring and modelling are very important issues in atmospheric pollution abatement and public health protection.

Carbon monoxide is a colourless, odourless, and tasteless air toxin (e.g. Chen *et al.*, 2011). CO is one of many ubiquitous contaminants of our environment that requires prevention and control measures (Raub *et al.*, 2000). The association between CO exposure and adverse cardiovascular outcomes has been well supported by previous findings (e.g. Min *et al.*, 2009). According to the US Environmental Protection Agency (US EPA), people with cardiovascular disease, such as coronary artery disease, are most at risk (US EPA, 2009). The World Health Organization has set specific air quality guidelines for different CO exposure averaging times,

[1] ARCH models have become vital tools for financial analysts in asset pricing, i.e. Bollerslev *et al*. (1988), portfolio construction, i.e. Engle (2002), risk management, i.e. Christoffersen (2003), option pricing, i.e. Duan (1995), as well as for estimating relationships from economic theory (interest rates, i.e. Gray, 1996, inflation modelling, i.e. Engle, 1982, business cycle synchronization, i.e. Degiannakis *et al*., 2014).

which are summarized as follows: 100 $mg/m^3$ for 15-min exposure, 60 $mg/m^3$ for 30-min, 30 $mg/m^3$ for 1-h and 10 $mg/m^3$ for 8-h exposure. There is no long-term average guideline (Chaloulakou *et al.*, 2003b).

Since it has been recognized that carbon monoxide pollution constitutes a remarkable threat for the public health mainly over the densely populated cities, a number of studies have sought to identify common risk factors for carbon monoxide intoxication, generally by conducting retrospective analyses of case reports (Montoya *et al.*, 2008). Moreover, CO can be used as a tracer for pollution from biomass burning and anthropogenic activities such as traffic (Choi and Chang, 2006). Many studies have been conducted concerning mainly urban areas worldwide e.g. the city of San Diego in California (Luria *et al.*, 2005). A common goal in most of these studies is to better understand carbon monoxide pollution patterns using various air pollution models in order to estimate the spatial and/or temporal distribution of CO sources in each case. Among them, there are studies, analyses and reports concerning carbon monoxide pollution in the greater Athens area, which will be presented hereafter.

**2.2. Exposure to carbon monoxide in the Athens urban area**

Air pollution constitutes one significant environmental problem for the greater Athens area for more than 3 decades (e.g. Mirasgedis *et al.*, 2008). Central residential areas are greatly affected by the intense traffic density in the nearby commercial areas (Diapouli *et al.*, 2008). More precisely, emissions from the road transport sector are dominant, with the number of vehicles in circulation exceeding 2 million (Grivas *et al.*, 2012). Other sources of atmospheric pollution in the Athens Basin are industry and heating. The main area of concentration of air-polluting industry is along a south-west/north-east axis in the historic centre of the city and in the western suburbs. In the city of Athens almost 100% of total carbon monoxide emissions are attributed to mobile sources. The air quality standard for CO is established by the European Union at $10mg/m^3$, as a value never to be exceeded by 8h mean concentrations (Mavroidis *et al.*, 2007).

Modelling of carbon monoxide pollution in the Athens area has been the subject of several studies. Viras *et al.* (1996), after nine-year measurements of CO concentrations in one of the central air pollution monitoring sites in Athens, showed that higher levels of CO were traced during the cold period of the year while during the morning and the night hours the levels increased due to both the adverse for pollution dispersion meteorological conditions observed during those hours and to the intense traffic observed at the same time (the levels are lower during the weekends especially on Sundays). Vellopoulou and Ashore (1998) have examined commuters' exposures to CO in the greater Athens region. Flouris (2006) has examined the specific atmospheric conditions in Athens during the summer of 2004 and the Games of XXXVIII Olympiad. It was mentioned that, since Athens began introducing a new generation of more efficient public buses at the end of 90s and the new metro was established in 2000 the concentrations of CO has remained at low levels in all sectors of the greater Athens.

**2.3. Environmental monitoring**

Pollution variables exhibit high time correlation. Thus, various time series models have been studied in the literature for forecasting air-pollution data. In an effort to forecast daily air-pollution concentrations, many researchers have developed daily forecasting models. The need for accurate modelling of air pollution has driven researchers to both statistical and artificial intelligent (mainly neural networks) methods (Prybutok *et al.*, 2000). Conventional statistical models include among others linear models, SARIMA models, Kalman Filters, etc. Linear models were first fitted by Aron and Aron (1978) in order to predict CO levels. Sahu and Mardia (2005) applied a Bayesian Kriged Kalman model for short-term forecasting of air pollution levels. Kumar and Jain (2010) study forecasting methods based on ARIMA models. Donnelly *et al.* (2015) propose a real time air quality forecasting using integrated parametric and nonparametric regression techniques.

An interesting fact is that many papers are looking in both time and spatial domain. For example, Bowman *et al.* (2009) proposed a spatiotemporal model for predicting air-pollution data.

Many authors have proposed artificial intelligent techniques. Kukkonen *et al.* (2003) gave an extensive evaluation of neural networks for predicting air-pollution concentrations, compared with appropriate deterministic modelling systems. Niskaa *et al.* (2004) used neural network model for forecasting air-pollution time series using a parallel genetic algorithm. Kurt *et al.* (2008) presented an online air pollution forecasting system using neural networks. Pisoni *et al.* (2009) used polynomial NARX models for predicting ozone levels. Dıaz-Robles (2008) used a hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas of Chile. Ibarra-Berastegi (2010) focused on the prediction of hourly levels up to 8h ahead for five pollutants (SO2, CO, NO2, NO and O3) and six locations in the area of Bilbao (Spain) using neural networks (NNs).

Moreover, many researchers proposed statistical process monitoring (SPM) techniques for continuously evaluating air-pollution measurements. Vaughana and Russella (1983) proposed monitoring point sources of pollution using control charts. Fasso (1998) proposed one-sided multivariate testing techniques for environmental monitoring. Pettersson (1998) used multivariate SPM techniques (Hotelling's $T^2$) for monitoring biodiversity. A cumulative sum type of methodology for environmental monitoring was presented by Manly and Mackenzie (2000). Corbett and Pan (2002) proposed the use of CUSUM chart as a tool to monitor emissions data so that abnormal changes can be detected in a timely manner. Yoo *et al.* (2008) enhanced process monitoring for wastewater treatment systems for using control charts. Pan and Chen (2008) presented a control chart for autocorrelated data using autoregressive fractionally integrated moving-average model to monitor the long-memory air quality data. Morrison (2008) used control charts to interpret environmental monitoring data.

In the methodological section that follows, we illustrate a unified framework incorporating time series modelling and control charting for establishing a real time CO monitoring scheme. Since the paper proposes a sequel of steps, the procedure is directly transferable into different contexts of monitoring pollution variables. The proposed framework is divided in two layers. In the 1st layer we define the multivariate modelling of dynamic (across time and regions) relationship among pollution variables. Hence, we state the appropriate framework of modelling the conditional mean, variance, confidence interval, and correlation of variables under investigation. In the 2nd layer, we apply a monitoring scheme to explore the early detection of abnormal increases of CO levels either globally or locally.

## 3. A Framework for Real Time Monitoring of Urban Pollution

In the light of the aforementioned discussion, it is clear that an automated mechanism, able to signal when the forecasted next day CO levels are evaluated as high would be very useful. Thus, in this section, a proactive framework for real time monitoring is proposed, with the application of an appropriate time series modelling, along with a suitable process monitoring procedure, which will provide a model that estimates dynamically the next day's CO levels as well as their variances and covariances. The dynamic specification enhances our availability to proceed to accurate estimates of next day's confidence interval based on the most recently available information. The framework consists of both an air-pollution forecasting layer borrowed from area of finance as well as a monitoring technique that automatically signals alarms.

### 3.2 Framework layer I: Time series model

*3.2.1. Multivariate time series modelling of air pollution*

In the following paragraphs, a multivariate framework is defined, known as multivariate Autoregressive Conditional Heteroskedasticity (ARCH) modelling. The intention is to provide daily dynamic estimates of the level, variance and correlation for the air pollution variables.

For $\mathbf{x}_t = [x_{1,t} \quad \dots \quad x_{n,t}]'$ denoting the $(n \times 1)$ vector with the $n$ variables on a daily frequency, the $\mathbf{y}_t = (1 - L)\mathbf{x}_t$ denotes the daily differences of $\mathbf{x}_t$. The multivariate discrete time real-valued stochastic process $\mathbf{y}_t$ can be decomposed into two parts, the predictable component, $E_{t-1}(\mathbf{y}_t) \equiv \boldsymbol{\mu}_t$, and the unpredictable component (or innovation process), $\boldsymbol{\varepsilon}_t \equiv \mathbf{y}_t - \boldsymbol{\mu}_t$.

$E_{t-1}(.)$ corresponds to the conditional mean given the information set $I_{t-1}$ available at time $t-1$. By $V_{t-1}(\mathbf{y}_t) \equiv \mathbf{H}_t$ we define the conditional covariance matrix of the innovation process. In a general form the underline framework can be presented as:

$$\begin{aligned}\mathbf{y}_t &= \boldsymbol{\mu}_t + \boldsymbol{\varepsilon}_t \\ \boldsymbol{\varepsilon}_t \mid I_{t-1} &\sim N(0, \mathbf{H}_t) \\ \boldsymbol{\varepsilon}_t &= \mathbf{H}_t^{1/2}\mathbf{z}_t \\ \mathbf{z}_t &\sim N(\mathbf{0}, \mathbf{I}) \\ \mathbf{H}_t &= g(\mathbf{H}_{t-1}, \mathbf{H}_{t-2}, \ldots, \boldsymbol{\varepsilon}_{t-1}, \boldsymbol{\varepsilon}_{t-2}, \ldots),\end{aligned} \tag{1}$$

where $N(.,.)$ is multivariate normal density function, $g(.)$ is a function of the lagged values of conditional covariance matrix and the innovation process, and $\mathbf{z}_t$ is an i.i.d. vector process such that $E(\mathbf{z}_t) = \mathbf{0}$ and $E(\mathbf{z}_t\mathbf{z}_t') = \mathbf{I}$.

In order to capture the autocorrelation that characterizes $\mathbf{y}_t$, the conditional mean is formulated as an AR($k$) model: $\boldsymbol{\mu}_t \equiv \mathbf{c}_0 \circ \left(\mathbf{i} - \sum_{i=1}^{k}\mathbf{c}_i\right) + \sum_{i=1}^{k}\mathbf{c}_i \circ \mathbf{y}_{t-i}$, where the symbol $\circ$ denotes the Hadamard product, $\mathbf{i}$ is a vector of ones, and $\mathbf{c}_0$, $\mathbf{c}_i$ are matrices with parameters to be estimated. The conditional covariance matrix is defined according to Bollerslev's *et al.* (1988) Diag-VECH($p,q$) framework, which has been modified in order to capture the asymmetric relationship between the unpredictable component of conditional mean and the conditional covariance[2]. Thus, by incorporating the Glosten's *et al.* (1993) asymmetric GARCH model, we define the Diag-aVECH($p,q$) framework:

$$vech(\mathbf{H}_t) = vech(\tilde{\mathbf{A}}_\mathbf{0}) + \sum_{i=1}^{p} vech(\tilde{\mathbf{A}}_i) \circ vech(\boldsymbol{\varepsilon}_{t-i}\boldsymbol{\varepsilon}_{t-i}') + vech(\tilde{\boldsymbol{\Gamma}}_i) \circ vech(\tilde{\boldsymbol{\varepsilon}}_{t-i}\tilde{\boldsymbol{\varepsilon}}_{t-i}') + \sum_{i=1}^{q} vech(\tilde{\mathbf{B}}_i) \circ vech(\mathbf{H}_{t-i}) \tag{2}$$

where $\tilde{\mathbf{A}}_\mathbf{0}$, $\tilde{\mathbf{A}}_i$, $\tilde{\boldsymbol{\Gamma}}_i$ and $\tilde{\mathbf{B}}_i$ are matrices with parameters to be estimated. The Diag-aVECH($p,q$) specification is preferable compared to models whose success depends on their ability to estimate extremely large time varying covariance matrices; i.e. Engle's (2002) Dynamic Conditional Correlation (DCC) model. Moreover, the Diag-aVECH is guaranteed to be positive definite and involves the estimation of less number of parameters than other multivariate ARCH models; i.e. Engle and Kroner's (1995) BEKK model, Engle's *et al.* (1986) VECH model, etc.

The asymmetric Diag-VECH model is estimated assuming that the non-diagonal elements of $\tilde{\mathbf{A}}_\mathbf{0}, \tilde{\mathbf{A}}_1$, $\tilde{\boldsymbol{\Gamma}}_1$ and $\tilde{\mathbf{B}}_1$ are time varying. Such a specification has the flexibility to estimate time-varying covariances. Otherwise, in case of constant non-diagonal elements of $\tilde{\mathbf{A}}_\mathbf{0}, \tilde{\mathbf{A}}_1$, $\tilde{\boldsymbol{\Gamma}}_1$ and $\tilde{\mathbf{B}}_1$, a time-varying correlation due to the time-varying standard deviations would lead to an increase(decrease) in correlations in less(more) volatile periods. For details

[2] The information criteria strongly suggest the estimation of the asymmetric Diag-VECH model.

about multivariate ARCH models, the interested reader is referred to Xekalaki and Degiannakis (2010, chapter 11).

*3.2.2. Value-at-Risk measure*

Having provided a dynamic multivariate model, we can proceed to the estimation of the one-step ahead α% confidence interval. The confidence intervals' estimation is based on the notion of the Value-at-Risk (VaR) measure; the most widely used risk measure in financial literature.

Given a confidence level $a \in (0,1)$, the VaR measure is given by the smallest number $l$ (in the real numbers set $\Re$) such that the probability that the loss $L$ exceeds $l$ is at most $1-a$:

$$VaR_t^{(a)} = \inf\{l \in \Re : P(L > l) \leq 1-a\}. \tag{3}$$

Based on our dynamic framework in eq.(1) the time $t$ VaR given the information available at previous time $t$-1 and a 95% lower confidence interval can be presented as:

$$VaR_{i,t|t-1}^{(95\%)} = \mu_{i,t|t-1} + f_{5\%}(z_t;0,1)\sqrt{\sigma_{i,t|t-1}^2}, \tag{4}$$

where $f_{5\%}(z_t;0,1)$ denotes the lower 5% percentile of the standard normal distribution, $\mu_{i,t|t-1}$ is the conditional mean estimate ($i^{th}$ element of vector $\boldsymbol{\mu}_t$) and $\sigma_{i,t|t-1}^2$ is the conditional variance estimate.

Our purpose is to estimate a 95% upper confidence interval (using the above definition). Thus, we replace in eq.(4) the $f_{5\%}(z_t;0,1)$ with $f_{95\%}(z_t;0,1)$. In this context, $CI_{i,t|t-1}^{(95\%)}$ interprets the 95% upper level confidence interval for the next day's air pollution level, i.e. the maximum value of the air pollution on a daily basis and at a 95% confidence level:

$$CI_{i,t|t-1}^{(95\%)} = (x_{i,t-1} + \mu_{i,t|t-1}) + f_{95\%}(z_t;0,1)\sqrt{\sigma_{i,t|t-1}^2}. \tag{5}$$

### 3.3 Framework layer II: Monitoring technique

After modelling the time series attitude of CO pollution in the city, the next step is to establish the process monitoring procedure, using techniques belonging to the toolbox of SPM (Montgomery, 2007).

In the following paragraphs, we propose and exhibit appropriate monitoring techniques. The proposed time series model has the ability of one-step ahead forecasting of the 95% upper level confidence interval for the next day levels of the air pollution variables under surveillance. In order to define the two monitoring techniques, we use the 95% upper bound defined in the previous sub-section ($CI_{i,t|t-1}^{(95\%)}$).

Before we proceed to the establishment of the monitoring procedure, there are several issues that have to be pointed out. One issue is that we have multivariate data. Another issue is that there is a correlation among the components of the vector with the CO measurements. In light of this, we are going to make use of suitable multivariate SPM methods (Wierda, 1994, Bersimis *et al.*, 2007) in order to define an appropriate procedure. Another issue is that we would like to monitor the multivariate CO levels in a way that it can be directly associated to the health safety limits. Specifically, the evolution of CO would be of no interest, in case that the CO level was independent of human health. Thus, the control procedure must have the ability of signalling alarms in case that the next day predicted CO levels are close or beyond the limits given by WHO.

WHO gives the following time-weighted average exposures for CO levels: (a) 100 $mg/m^3$ (87 ppm) for 15 min, (b) 60 $mg/m^3$ (52 ppm) for 30 min, (c) 30 $mg/m^3$ (26 ppm) for 1 h, (d) 10 $mg/m^3$ (9 ppm) for 8 h, (e) for indoor air quality 7 $mg/m^3$ (6 ppm) for 24 h. These values have been determined by WHO in such a way that a carboxyhaemoglobin (COHb) level of 2.5% is not exceeded, even when a regular subject engages in light or moderate exercise.

A first solution would be to monitor the consecutive $CI_{i,t|t-1}^{(95\%)}$ against an appropriate limit provided by WHO. However, this would not assess the spatiotemporal dynamics of the CO values. Thus, we propose two different techniques that take into account the spatiotemporal dynamics of the CO.

The first monitoring technique will assess a possible global change of the multivariate time series while the second one will aim to the component of the time series with the largest change. Both the monitoring techniques will be based on control charting appropriately statistics against appropriate control limits.

*3.3.1. Monitoring the time series for a global change*

The challenge of aiming towards a global change is to introduce an appropriate statistic that will take into account all the components of the time series as well as their correlation. Additionally, the statistic has to be compared against a constant limit, as it is usual in the statistical process control literature. For this reason, we introduce the following statistic:

$$T_{1,t} = \sum_{i=1}^{n} a_{1i} v_{t,i} + \sum_{i=1}^{n} a_{1i} v_{t-1,i} + \ldots + \sum_{i=1}^{n} a_{1i} v_{t-l,i} = \sum_{j=1}^{l} \sum_{i=1}^{n} a_{1i} v_{t-j,i}\,, \quad t = 1,2,\ldots, \tag{6}$$

which corresponds to first principal component after applying Dynamic Principal Component Analysis (DPCA) on the vector $\mathbf{v}_t$ containing the values $CI_{i,t|t-1}^{(95\%)}$, for $i = 1,2,\ldots,n$ (assuming that we study $n$ variables) and the associated $l$ time lagged values of $\mathbf{v}_t$. The application of DPCA on the $CI_{i,t|t-1}^{(95\%)}$ establishes an index of global change. High values of $T_{1,t}$ corresponds to global high values of $CI_{i,t|t-1}^{(95\%)}$, i.e. next day a% bound are high (even extremes).

DPCA is an extension of PCA method that takes into account the serial correlation, by augmenting each observation vector with the previous $l$ observation vectors. Chen and Liu

(2002) introduced the use of DPCA in industrial multivariate monitoring. The $T_{1,t}$ is plotted against a limit ($CL_1$) that can be obtained using a training data set during Phase I and the corresponding empirical distribution. Phase I in the statistical process monitoring literature corresponds to a retrospective analysis, which is applied to assess if the process is in control since the first sample was collected. Once this is accomplished, the control chart is used to define what is meant by statistical in-control. Then Phase II follows, where the control charts are employed to verify if the process remains in control in the future.

The $T_{1,t}$ is used to monitor the pollution time series for a global change since it incorporates information from all the variables analyzed.

*3.3.2. Identifying a local change*

In case we restrict our interest in identifying the most extreme changed element of the time series under study, we propose the use of the statistic:

$$T_{2,t} = \max_i \left( \frac{CI^{(95\%)}_{i,t|t-1}}{\sqrt{h_{ii}}} \right), t = 1,2,\dots,\ i = 1,2,\dots,n, \tag{7}$$

where $\sqrt{h_{ii}}$ is the $i^{th}$ diagonal element of matrix $\mathbf{H}_t$. The $T_{2,t}$ gives the highest next day forecast bound standardized with its dispersion. High values of $T_{2,t}$ corresponds to abnormally high movements of a measurement. As we exhibit later the above monitoring statistics can be used effectively for monitoring air-pollution data. The $T_{2,t}$ is plotted against a limit ($CL_2$) that can be obtained using a training data set during Phase I and the corresponding empirical distribution.

## 4. Monitoring CO Levels

In this section, after describing the data at hand, we apply the framework presented in the previous section for monitoring CO levels.

### 4.1. CO data

The data at hand are the daily CO measurements of Athens, Greece. The $(6 \times 1)$ vector $\mathbf{x}_t = (x_{1,t} \quad \dots \quad x_{6,t})'$ contains CO measurements from six different places of Athens (at time $t$=1,2,…). These measurements are acquired using the Athens' air quality network, which consists of sixteen stations recording air pollution data every 15 minutes. The daily CO level is computed as the average of the intra-day observations. The notion of averaging the intra-day CO levels relies on the attempt to model the average exposure to air pollution. CO monitoring is very important, especially in a daily basis, which can be seen considering that the US EPA National Ambient Air Quality Standards (NAAQS) has adopted for CO a standard of 35 ppm as a 1-h average and just 9 ppm as an 8-h average. Generally, continuous CO exposure to levels less than 10 mg/m3 should not cause carboxy-hemoglobin (COHb) levels more than 2% in normal non-smokers. This is because high CO concentrations can cause acute CO intoxication since CO is combined with the hemoglobin of human blood to produce carboxy-hemoglobin (COHb), and therefore disrupts the transfer of oxygen to human tissues causing hypoxia.

The six stations that are used in this study are Athinas area (city center), Geoponiki area (city center), Marousi (North suburb), Nea Smyrni (Southeast suburb), Patision area (city center), and Peristeri (Southwest suburb). These six stations represent the central area of Athens, which is the most heavily polluted area. The dataset is available for 2922 days.

**4.2 Modelling daily CO levels**

In the sequel, the AR( $k$ )Diag-aVECH( $p,q$ ) model framework is formulated to provide daily dynamic estimates of the variance and correlation for the 6 air pollution variables; for $\mathbf{x}_t$ denoting the $(6\times1)$ vector with the 6 variables on a daily frequency. The lag orders of both conditional mean, $k$ , and conditional variance, $p,q$ , have been investigated according to Akaike's (1973) and Schwarz's (1978) Bayesian information criteria.

The predictable component is defined as a 4th order autoregressive, or AR(4), model in order to capture the autocorrelation structure in $\mathbf{y}_t$ (the correlograms of all the variables highly indicate the existence of short memory autocorrelation). The lag orders of the Diag-aVECH( $p,q$ ) framework are defined, by the information criteria, to $p=q=1$. Thus, the proposed model is the six-dimensional multivariate AR(4)-Diag-aVECH(1,1) model:

$$\begin{gathered}\mathbf{y}_t=\mathbf{c}_0\circ(\mathbf{i}-\mathbf{c}_1-\mathbf{c}_2-\mathbf{c}_3)+\mathbf{c}_1\circ\mathbf{y}_{t-1}+\mathbf{c}_2\circ\mathbf{y}_{t-2}+\mathbf{c}_3\circ\mathbf{y}_{t-3}+\mathbf{c}_4\circ\mathbf{y}_{t-4}+\boldsymbol{\varepsilon}_t\\ \boldsymbol{\varepsilon}_t=\mathbf{H}_t^{1/2}\mathbf{z}_t\\ \mathbf{z}_t\sim N(\mathbf{z}_t;\mathbf{0},\mathbf{I})\\ vech(\mathbf{H}_t)=vech(\tilde{\mathbf{A}}_0)+vech(\tilde{\mathbf{A}}_1)\circ vech(\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}'_{t-1})+vech(\tilde{\boldsymbol{\Gamma}}_1)\circ vech(\tilde{\boldsymbol{\varepsilon}}_{t-1}\tilde{\boldsymbol{\varepsilon}}'_{t-1})+vech(\tilde{\mathbf{B}}_1)\circ vech(\mathbf{H}_{t-1}).\end{gathered}\tag{8}$$

Each $i^{th}$ diagonal element of $\mathbf{H}_t$ is estimated as:

$$\sigma_{i,t}^2=a_{i,i}+\tilde{a}_{i,i}\varepsilon_{i,t-1}^2+\gamma_{i,i}\varepsilon_{i,t-1}^2d_{i,t-1}+\tilde{b}_{i,i}\sigma_{i,t-1}^2,\tag{9}$$

whereas, each $(i,j)^{th}$ non-diagonal element is computed as:

$$\sigma_{i,j,t}=a_{i,j}+\tilde{a}_{i,j}\varepsilon_{i,t-1}\varepsilon_{j,t-1}+\gamma_{i,j}\varepsilon_{i,t-1}d_{i,t-1}\varepsilon_{j,t-1}d_{j,t-1}+\tilde{b}_{i,j}\sigma_{i,j,t-1},\tag{10}$$

where $d_{i,t-1}$ denotes the indicator function, i.e. $d_{i,t-1}=1$ if $\varepsilon_{i,t-1}<0$, and $d_{i,t-1}=0$ otherwise. The diagonal elements of $\mathbf{H}_t$ express the estimates of air pollution variables' conditional variance. Having estimated the elements of the time-varying covariance matrix, consequently, the time-varying correlations between $i^{th}$ and $j^{th}$ variables can be estimated as:

$$\rho_{i,j,t}=\frac{\sigma_{i,j,t}}{\sqrt{\sigma_{i,t}^2\sigma_{j,t}^2}}=\frac{a_{i,j}+\tilde{a}_{i,j}\varepsilon_{i,t-1}\varepsilon_{j,t-1}+\gamma_{i,j}\varepsilon_{i,t-1}d_{i,t-1}\varepsilon_{j,t-1}d_{j,t-1}+\tilde{b}_{i,j}\sigma_{i,j,t-1}^2}{\sqrt{\left(a_{i,i}+\tilde{a}_{i,i}\varepsilon_{i,t-1}^2+\gamma_{i,i}\varepsilon_{i,t-1}^2d_{i,t-1}+\tilde{b}_{i,i}\sigma_{i,t-1}^2\right)\left(a_{j,j}+\tilde{a}_{j,j}\varepsilon_{j,t-1}^2+\gamma_{j,j}\varepsilon_{j,t-1}^2d_{j,t-1}+\tilde{b}_{j,j}\sigma_{j,t-1}^2\right)}}.\tag{11}$$

The proposed model specification has been tested for residuals' serial correlation with Lütkepohl's (2007) multivariate Q-statistic and for presence of ARCH effects in the residuals with Tse's (2002) test. The model demands the estimation of 30 parameters for the conditional mean vector, as well as, the estimation of 84 parameters, or $2(n+1)n$, for $n=6$, for the computation of the conditional variance-covariance matrix[3].

**4.3. Estimating time-varying 95% upper bound confidence intervals and correlations**

Based on the generic dynamic framework in eq.(1) the day's $t$ conditional mean given the information available at previous day $t-1$ is estimated as an AR(4) process (from 1st row of eq.(8)):

$$\mu_{i,t|t-1} = c_0\left(1-\sum_{i=1}^{4}c_i\right)+\sum_{i=1}^{4}c_i y_{t-i}\,. \tag{12}$$

The conditional variance is estimated from eq.(9) as:

$$\sigma^2_{i,t|t-1} = a_{i,i} + \left(\widetilde{a}_{i,i} + \gamma_{i,i} d_{i,t-1}\right)\varepsilon^2_{i,t-1|t-1} + \widetilde{b}_{i,i}\sigma^2_{i,t-1|t-1}\,. \tag{13}$$

Moreover, we construct a confidence bound for the $\{\mathbf{x}_t\}$ vector. Hence, we adapt eq.(5) in order to quantify the 95% upper bound of the next day's air pollution, $CI^{(95\%)}_{i,t|t-1}$. The values of the 6 air pollution variables, on a daily basis, against the 95% upper level confidence interval, estimated by the AR(4)-Diag-aVECH(1,1) model are given in Figure 1. From Figure 1 it is noticeable that there are periods of high volatility, which are followed by periods of low volatility. Financial literature notes this effect as volatility clustering. In the case of air pollution the time series clustering expresses the seasonality. Mandelbrot (1963) was the first who noticed the volatility clustering in stock market data, noting down that "Large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes". As Engle (1982) first noted, the volatility clustering effect has been successively captured by ARCH modelling.

[Please Insert Figure 1 About Here]

Table 1 provides the 95% failure rate, i.e. the percentage of upper level confidence interval's violations, and the Kupiec's (1995) test. A violation occurs if the estimated 95% upper bound confidence interval is less than the actual value of the air pollution; or $x_{i,t} > CI^{(95\%)}_{i,t|t-1}$. The percentage of violations is computed as $N=\sum_{t=1}^{T}\widetilde{I}_t$, for $\widetilde{I}_t = 1$ if $x_{i,t} > CI^{(95\%)}_{i,t|t-1}$ and $\widetilde{I}_t = 0$ if $x_{i,t} \le CI^{(95\%)}_{i,t|t-1}$. Kupiec's (1995) test examines the null hypothesis that the observed violation rate, $N/T$, is statistically equal to the expected violation rate, $\rho$ =5%. The likelihood ratio statistic is chi-squared distributed with one degree of freedom. The likelihood ratio (LR) statistic equals to:

[3] Due to the large number of coefficients, their estimates are not reported, but they are available upon request.

$$LR = 2\log\left(\left(1-\frac{N}{T}\right)^{T-N}\left(\frac{N}{T}\right)^{N}\right) - 2\log\left((1-\rho)^{T-N}\rho^{N}\right) \qquad (14)$$

The LR test indicates that for all the air pollution variables the AR(4)-Diag-aVECH(1,1) model estimates accurately the 95% upper bound confidence interval. The p-values do not reject the null hypothesis that the observed violation rate is statistically equal to the expected violation rate.

[Please Insert Table 1 About Here]

Figure 2 presents the time-varying correlations between the 6 air pollution variables having estimated according to eq.(11). From Figure 2, it is noticeable that the values of air pollution are not always highly correlated among the six places of Athens. In the majority of the time the cross-correlation of air pollution among the 6 areas of Athens is highly positive. However, there are cases that the correlation approaches zero levels.

[Please Insert Figure 2 About Here]

**4.4. Monitoring procedure**

In this sub-section, we apply the monitoring procedures proposed above. The application will be done in two phases. We will use the data acquired during the first two years as a Phase I sample in order to estimate the control limits and then we apply this limits in the future VaR values.

*4.4.1. Monitoring the time series for a global change*

The $T_{1,t}$, $t$=1,2,…, is calculated by applying DPCA on the vector $\mathbf{v}_t$ containing the $CI^{(95\%)}_{i,t|t-1}$ values, for $i$=1,2,…,6, and the associated time lagged values of $\mathbf{v}_t$. The parameter $l$ was set equal to 4 after appropriate experimentation with criterion the robustness of the monitoring procedure (keeping the false alarm rate in a pre-specified level). Since the parameter $l$ was found to be equal to 4, it coincides with the estimated AR model (which is supported by the literature, as DPCA extracts autoregressive based components).

Using the statistic $T_{1,t}$, $t$=1,2,…, someone can argue that due to the summing nature of DPCA the approximate distribution of $T_{1,t}$ is normal. The values of $T_{1,t}$ for the first 500 observations are presented in Figure 3a. However, this assumption is not validated in practice.

Nevertheless, the two modes appearing in the distribution give us evidence that probably the distribution of $T_{1,t}$ is a mixture of two distributions with different parameters. In the literature review section, we referred to the fact that during the winter (cold periods) the values of CO are larger than the corresponding CO values during summer (Viras *et al.*, 1996). This fact motivated us to analyze the $T_{1,t}$ separately for winter and summer (appropriately assigning the other two seasons to either summer or winter using as a criterion the 10 year mean temperature). The distribution of $T_{1,t}$ for the first 500 observations (summer and winter are separated) is presented in the Figure 3b as we may see the distribution of $T_{1,t}$ after summer and winter are separated can be approximated by a normal distribution with appropriate parameters. Thus, someone can choose to calculate appropriate limits for $T_{1,t}$ either using a normal approximation or the empirical distribution estimated by a Phase I sample. However, the quality of approximation of the proposed statistic due to the nature of our data depends on a plethora of exogenous variables; i.e. temperature, humidity, air speed, etc. Henceforth, we propose the use of the empirical distribution.

In Figure 3c, we present the control chart for the phase I data set. The control limits were set to be equal with the upper 2% percentile of the empirical distributions of $T_{1,t}$ (winter [pointed out with 1] and summer [pointed out with 2] are separated). We adapt different control limits for each season because of the different distributional properties of $T_{1,t}$ in these two seasons.

[Please Insert Figure 3 About Here]

The $T_{1,t}$ is used to monitor the pollution time series for a global change since it incorporates information from all the variables analyzed. In Phase II, when the $T_{1,t}$ exceeds the limit a signal is alarmed indicating a global high concentration (i.e. the DPCA1 which represents a weighted mean of the values recorded by the six stations appropriately adjusted by the values of the same variables in a time window equal to 4, indicate high values in all variables measured). This way of thinking is enhanced by the fact that extreme values of $T_{1,t}$ are associated by global extremes of the $CI_{i,t|t-1}^{(95\%)}$, i=1,2,3,4,5,6 (high values of $CI_{i,t|t-1}^{(95\%)}$ warning that the CO values of next will be, with high probability, large or even extreme). By observing both Figures 3c and 3d we may conclude that the $T_{1,t}$ is plotted beyond the limits only rarely, which coincides with the previous studies that state that CO levels in broader Athens area are only rarely exceed safety limits.

*4.4.2. Identifying a local change*

Using the statistic $T_{2,t}$ and following the same way of thinking for $T_{2,t}$ we may identify an extremely local change since $T_{2,t}$ identifies the most extreme variable (CO measurement of one out of the 6 different areas of Athens). The $T_{2,t}$ is plotted against a limit ($CL_2$) which is acquired using the same spirit as in the case of $T_{1,t}$ during Phase I and the corresponding empirical distribution (see Figure 4). If $T_{2,t}$ exceeds the limit a signal is alarmed indicating a specific variable/area that shows extreme variation. By observing both Figures 4a and 4b one may conclude that the $T_{2,t}$ is plotted beyond the limits only rarely.

[Please Insert Figure 4 About Here]

## 5. Conclusion

The paper establishes and presents a SPM framework for monitoring the effects of CO air-pollution variable, over a network of stations. This framework combines the use of value-at-risk modelling and control charting, for detecting extreme pollution events. The application to Athens data reveals that the proposed methodology can provide accurate and robust results without excess false alarms. In particular, it is able to identify whether temporal exceeding come from a specific station (location) or it is attributed to a specific pollutant.

The operation steps of the proposed framework are briefly the following: The next day's estimated CO levels are estimated. Afterwards, using these forecasts, the next day's CO levels are being monitored and, in case that the actual CO levels are near or beyond a threshold level or limit, a signal alarms. Specifically, based on these forecasts, the complete area under examination (in our case the Athens area) is being monitored. At the same time, using the same forecasts and even if the actual CO level is not near or beyond a threshold level or limit in the examined area, particular regions of this area are checked. According to our knowledge, this is the first time that such a modelling framework is applied in environmental application.

To sum up, the proposed model consists of both an air-pollution forecasting layer borrowed from the area of finance as well as it constitutes a monitoring technique that automatically signals alarms all integrated in a complete framework.

Regarding its usefulness, this is evidently significant, since early signals of abnormal increases of CO levels can be used for public protection. It has been recognized that carbon monoxide pollution constitutes a remarkable threat for the public health mainly over the densely populated cities. According to the US Environmental Protection Agency (US EPA) AQI -Air Quality Index, "people with cardiovascular disease, such as coronary artery disease, are most at risk. They may experience chest pain and other cardiovascular symptoms if they are exposed to carbon monoxide, particularly while exercising. People with marginal or compromised cardiovascular and respiratory systems (for example, individuals with congestive heart failure, cerebrovascular disease, anemia, or chronic obstructive lung disease), and possibly young infants and fetuses, also may be at greater risk from carbon monoxide pollution. In healthy individuals, exposure to higher levels of carbon monoxide can affect mental alertness and vision". Therefore, it is quite obvious that carbon monoxide monitoring and modelling are very important issues public health protection and warning.

The proposed monitoring method of air pollution can be utilized in real-time, as it can be applied for subsequent points in time (i.e. daily observations) and requires rational computing time (i.e. less than a minute in an I7 Pentium). Moreover, it is flexible as it could be easily adapted depending on the demands of the monitoring, as well as it could be standardized and, consequently, incorporated in relevant atmospheric pollution monitoring devices and software. The program codes for estimating the proposed multivariate framework, the upper confidence interval, the Kupiec test, etc. are available to the readers upon request.

## References


**Akaike H. 1973.** Information Theory and an extension of the maximum likelihood principle. In Petrov BN and Csaki F (eds.) *Proceedings of the Second International Symposium on Information Theory*: 267-281.

**Anderson MJ, Thompson AA. 2004.** Multivariate control charts for ecological and environmental monitoring. *Ecological Applications* **14**(6): 1921-1935.

**Aron RH, Aron I-M. 1978.** Statistical forecasting models: I. Carbon monoxide concentrations in the Los Angeles basin. *Journal of the Air Pollution Control Association* **28**: 681-688.

**Asatar GI, Nair PR. 2010.** Spatial distribution of near-surface CO over Bay of Bengal during winter: role of transport. *Journal of Atmospheric and Solar-Terrestrial Physics* **72**: 1241–1250.

**Badr O, Probert SD. 1994.** Sources of atmospheric carbon monoxide. *Applied Energy* **49**: 145-195.

**Bel, G., Bolancé, C., Guillén, M., & Rosell, J. 2015.** The environmental effects of changing speed limits: A quantile regression approach. *Transportation Research Part D: Transport and Environment* **36**: 76-85.

**Bersimis S, Psarakis S, Panaretos J. 2007.** Multivariate statistical process control charts: an overview. *Quality and Reliability Engineering International* **23**: 517-543.

**Bollerslev T, Engle RF, Wooldridge JM. 1988.** A capital asset pricing model with time-varying covariances. *Journal of Political Economy* **96**: 116-131.

**Bogo H, Gómez DR, Reich SL, Negri RM, San Román E. 2001.** Traffic pollution in a downtown site of Buenos Aires City. *Atmospheric Environment* **35**: 1717 -1727.

**Bowman AW, Giannitrapani M, Scott EM. 2009.** Spatiotemporal smoothing and sulphur dioxide trends over Europe. *Working Paper*, 1-30.

**Chaloulakou A, Mavroidis I, Duci A. 2003a.** Indoor and outdoor carbon monoxide concentration relationships at different microenvironments in the Athens area. *Chemosphere* **52**: 1007–1019.

**Chaloulakou A, Kassomenos P, Spyrellis N, Demokritou P, Koutrakis P. 2003b.** Measurements of $PM_{10}$ and $PM_{2.5}$ particle concentrations in Athens, Greece. *Atmospheric Environment* **37**: 649–660.

**Chen JH, Liu KC. 2002.** On-line batch process monitoring using dynamic PCA and dynamic PLS models. *Chemical Engineering Science* **57**: 63–75.

**Chen R, Pan G, Zhang Y, Xu Q, Zeng G, Xu X, Chen B, Kan H. 2011.** Ambient carbon monoxide and daily mortality in three Chinese cities: The China Air Pollution and Health Effects Study (CAPES). *Science of The Total Environment* **409**(23): 4923–4928.

**Choi S-D, Chang Y-S. 2006.** Carbon monoxide monitoring in Northeast Asia using MOPITT: Effects of biomass burning and regional pollution in April 2000. *Atmospheric Environment* **40**: 686–697.

**Christoffersen, P. 2003.** Elements of Financial Risk Management. Academic Press, New York.

**Corbett CJ, Pan J-N. 2002.** Evaluating environmental performance using statistical process control techniques. *European Journal of Operational Research* **139**: 68–83.

**Degiannakis, S., Duffy, D., & Filis, G. 2014.** Business Cycle Synchronization in EU: A Time-Varying Approach. *Scottish Journal of Political Economy* **61**(4): 348-370.

**Diapouli E, Chaloulakou A, Spyrellis N. 2008.** Indoor and outdoor PM concentrations at a residential environment, in the Athens area. *Global NEST Journal* **10**(2): 201–208.

**Díaz-Robles LA, Ortega JC, Fu JS, Reed GD, Chow JC, Watson JG, Moncada-Herrera JA. 2008.** A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: The case of Temuco, Chile. *Atmospheric Environment* **42**: 8331–8340.

**Donnelly A, Misstear B, Broderick B. 2015.** Real time air quality forecasting using integrated parametric and non-parametric regression techniques. *Atmospheric Environment* **103**: 53–65.

**Duan, J. 1995.** The Garch Option Pricing Model, *Mathematical Finance* **5**(1): 31-32.

**Engle RF. 1982.** Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica* **50**: 987–1008.

**Engle, R.F., and Kroner, K.F. 1995.** Multivariate Simultaneous Generalized ARCH. *Econometric Theory*, **11**: 122-150.

**Engle RF, Manganelli S. 2004.** CAViaR: Conditional autoregressive value at risk by regression quantiles. *Journal of Business and Economic Statistics* **22**(4): 367–381.

**Engle, R.F., Granger, C.W.J., and Kraft, D. 1986.** Combining Competing Forecasts of Inflation Using a Bivariate ARCH Model. *Journal of Economic Dynamics and Control*, **8**: 151-165.

**Fasso A. 1998.** One-sided multivariate testing and environmental monitoring, *Austrian Journal of Statistics* **27**: 17–38.

**Flouris AD. 2006.** Modelling atmospheric pollution during the games of the XXVIII Olympiad: Effects on elite competitors. *International Journal of Sports Medicine* **27**: 137–142.

**Georgoulis LB, Hänninen O, Samoli E, Katsouyanni K, Künzli N, Polanska L, Bruinen de Bruin Y, Alm S, Jantunen M. 2002.** Personal carbon monoxide exposure in five European cities and its determinants. *Atmospheric Environment* **36**: 963–974.

**Glosten, L., Jagannathan, R. and Runkle, D. 1993.** On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *Journal of Finance*, **48**: 1779–1801.

**Gokhale S, Khare M. 2005.** A hybrid model for predicting carbon monoxide from vehicular exhausts in urban environments. *Atmospheric Environment* **39**: 4025–4040.

**Gray, S.F. 1996.** Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics* **42**: 27-62.

**Grivas G, Cheristanidis S, Chaloulakou A. 2012.** Elemental and organic carbon in the urban environment of Athens. Seasonal and diurnal variations and estimates of secondary organic carbon. *Science of the Total Environment* **414**: 535–545.

**Han X, Naeher LP. 2006.** A review of traffic-related air pollution exposure assessment studies in the developing world. *Environment International* **32**: 106–120.

**Ibarra-Berastegi G, Elias A, Barona A, Saenz J, Ezcurra A, Diaz de Argandoña J. 2008.** From diagnosis to prognosis for forecasting air pollution using neural networks: Air pollution monitoring in Bilbao. *Environmental Modelling & Software* **23**: 622–637.

**Kukkonen J, Partanen L, Karppinen A, Ruuskanen J, Junninen H, Kolehmainen M, Niska H, Dorling S, Chatterton T, Foxall R, Cawley G. 2003.** Extensive evaluation of neural network models for the prediction of $NO_2$ and $PM_{10}$ concentrations, compared with a deterministic modelling system and measurements in central Helsinki. *Atmospheric Environment* **37**: 4539–4550.

**Kumar U, Jain VK. 2010.** Time series models (Grey-Markov, Grey Model with rolling mechanism and singular spectrum analysis) to forecast energy consumption in India. *Energy* **35**: 1709–1716.

**Kur A, Gulbagci B, Karaca F, Alagha O. 2008.** An online air pollution forecasting system using neural networks. *Environment International* **34**: 592–598.

**Kupiec P. 1995.** Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives* **3**(2): 73-84.

**Lütkepohl H. 2007.** *New Introduction to Multiple Time Series Analysis*. Springer Verlag: New York.

**Luria M, Tanner RL, Valente RJ, Bairai ST, Koracin D, Gertler AW. 2005.** Local and transported pollution over San Diego, California. *Atmospheric Environment* **39**: 6765–6776.

**Mandelbrot B. 1963.** The variation of certain speculative prices. *Journal of Business* **36**: 394–419.

**Manly BFJ, Mackenzie D. 2000.** A cumulative sum type of method for environmental monitoring. *Environmetrics* **11**(2): 151–166.

**Mavroidis I, Gavriil I, Chaloulakou A. 2007.** Statistical modelling of CO and $NO_2$ concentrations in the Athens area. Evaluation of emission abatement policies. *Environmental Science and Pollution Research* **14**(2): 130–136.

**Min J-Y, Paek D, Cho S-I, Min K-B. 2009.** Exposure to environmental carbon monoxide may have a greater negative effect on cardiac autonomic function in people with metabolic syndrome. *Science of the Total Environment* **407**: 4807–4811.

**Mirasgedis S, Hontou V, Georgopoulou E, Sarafidis Y, Gakis N, Lalas DP, Loukatos A, Gargoulas N, Mentzis A, Economidis D, Triantafilopoulos T, Korizi K, Mavrotas G. 2008.** Environmental damage costs from airborne pollution of industrial activities in the greater Athens, Greece area and the resulting benefits from the introduction of BAT. *Environmental Impact Assessment Review* **28**: 39–56.

**Montgomery DC. 2007.** *Introduction to Statistical Quality Control*. Wiley: New York.

**Montoya T, Gurian PL, Velázquez-Angulo G, Corella-Barud V, Rojo A, Gram, JP. 2008.** Carbon monoxide exposure in households in Ciudad Juárez, México. *International Journal of Hygiene and Environmental Health* **211**: 40–49.

**Morrison LW. 2008.** The use of control charts to interpret environmental monitoring data. *Natural Areas Journal* **28**(1): 66–73

**Niska H, Hiltunen T, Karppinen A, Ruuskanen J, Kolehmainen M. 2004.** Evolving the neural network model for forecasting air pollution time series. *Engineering Applications of Artificial Intelligence* **17**; 159–167.

**Pan J-N, Chen S-T. 2008.** Monitoring long-memory air quality data using ARFIMA model. *Environmetrics* **19**: 209–219.

**Peng L, Zhao C, Lin Y, Zheng X, Tie X, Chan L-Y. 2007.** Analysis of carbon monoxide budget in North China. *Chemosphere* **66**: 1383–1389.

**Pettersson M. 1998.** Monitoring a freshwater fish population: Statistical surveillance of biodiversity. *Environmetrics* **9**: 139–150.

**Pisoni E, Farina M, Carnevale C, Piroddi L. 2009.** Forecasting peak air pollution levels using NARX models. *Engineering Applications of Artificial Intelligence* **22**: 593–602.

**Potoglou D, Kanaroglou PS. 2005.** Carbon monoxide emissions from passenger vehicles: predictive mapping with an application to Hamilton, Canada. *Transportation Research Part D* **10**: 97–109.

**Prybutok VR, Yi J, Mitchell D. 2000.** Comparison of neural network models with ARIMA and regression models for prediction of Houston's daily maximum ozone concentrations. *European Journal of Operational Research* **122**: 31–40.

**Raub JA, Mathieu-Nolf M, Hampson NB, Thom SR. 2000.** Carbon monoxide poisoning - a public health perspective. *Toxicology* **145**: 1–14.

**Sahu SK, Mardia KV. 2005.** A Bayesian kriged Kalman model for short-term forecasting of air pollution levels. *Applied Statistics* **54**(1): 223–244.

**Schwarz G. 1978.** Estimating the Dimension of a Model. *Annals of Statistics* **6**: 461–464.

**Tao Y, Zhong L, Huang X, Lu S-E, Li Y, Dai L, Zhang Y, Zhu T, Huang W. 2011.** Acute mortality effects of carbon monoxide in the Pearl River Delta of China. *Science of the Total Environment* **410**: 34–40.

**Tse YK. 2002.** Residual–based diagnostics for conditional heteroscedasticity models. *Econometrics Journal* **5**: 358–373.

**US EPA. 2000.** *Air Quality Criteria for Carbon Monoxide*. United States Environmental Protection Agency: Washington, DC.

**US EPA. 2009.** *AQI -Air Quality Index*. United States Environmental Protection Agency: Washington, DC.

**Vaughana WJ, Russella CS. 2012.** Monitoring point sources of pollution: answers and more questions from statistical quality control. *The American Statistician* **37**(4b): 476–487

**Vellopoulou AV, Ashore MR. 1998.** Personal exposures to carbon monoxide in the city of Athens: I. Commuters' exposures. *Environment International* **24**(7): 713–720.

**Viras LG, Paliatsos AG, Fotopoulos AG. 1996.** Nine-year trend of air pollution by CO in Athens, Greece. *Environmental Monitoring and Assessment* **40**: 203–214.

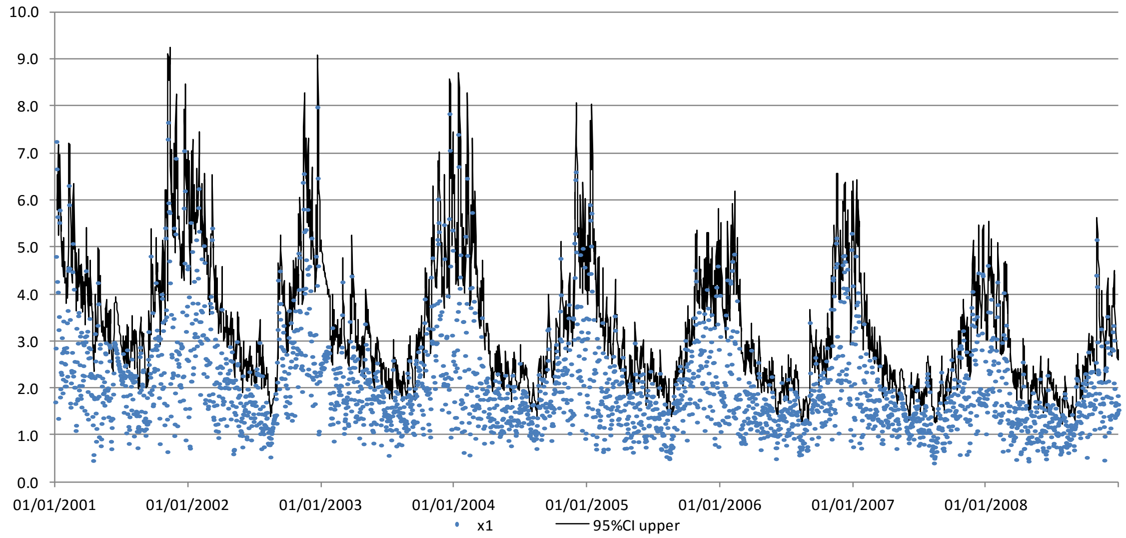
**Xekalaki E, Degiannakis S. 2010.** *ARCH Models for Financial Applications*. John Wiley and Sons: New York.

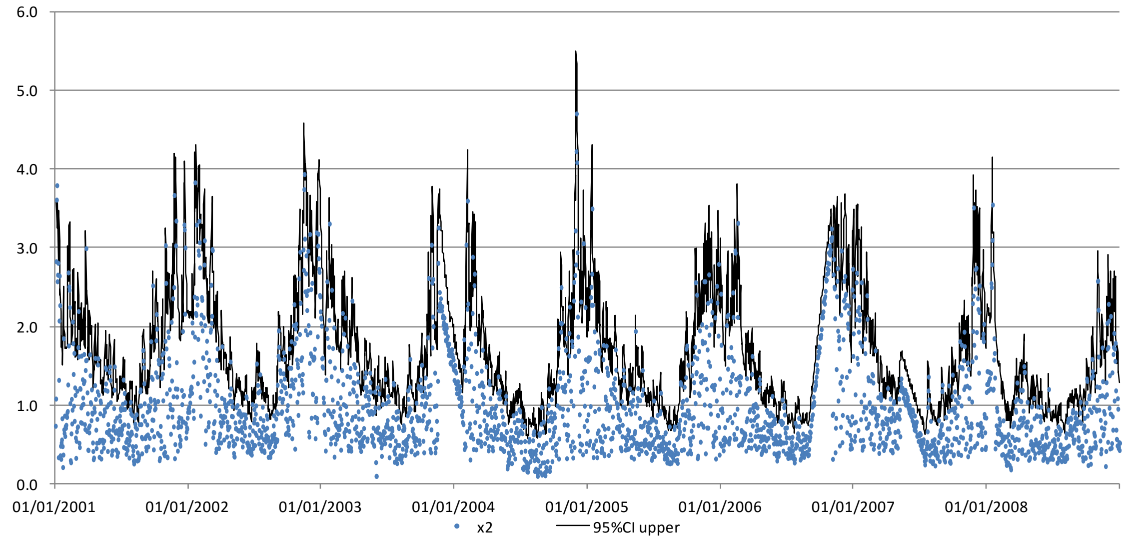
**Wierda SJ. 1994.** Multivariate statistical process control - recent results and directions for future research. *Statistica Neerlandica* **48**: 147–168.

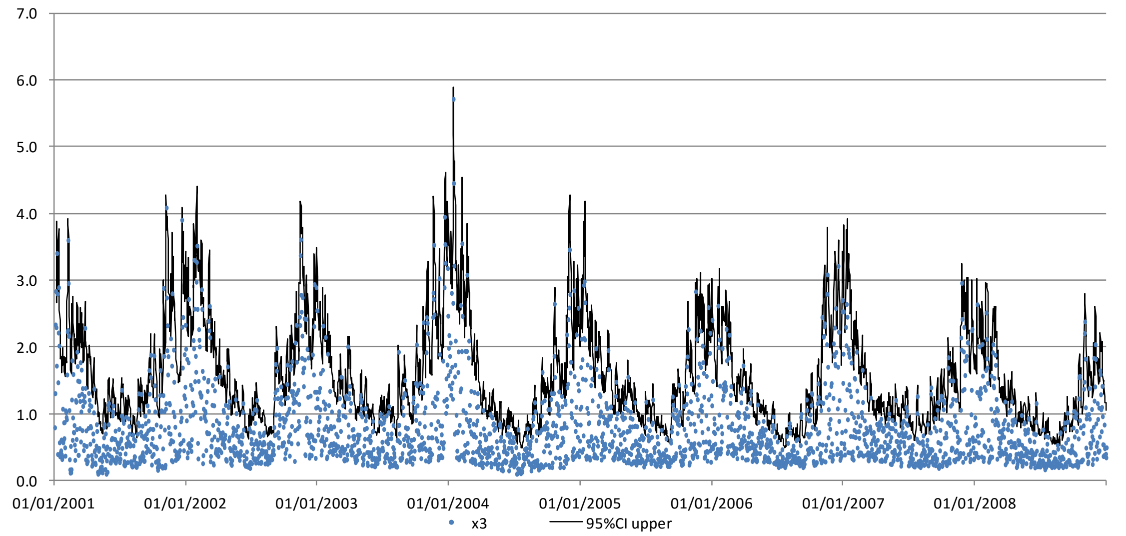
**Yoo CK, Villez K, Hulle SWH, Vanrolleghem PA. 2008.** Enhanced process monitoring for wastewater treatment systems. *Environmetrics* **19**: 602–617.

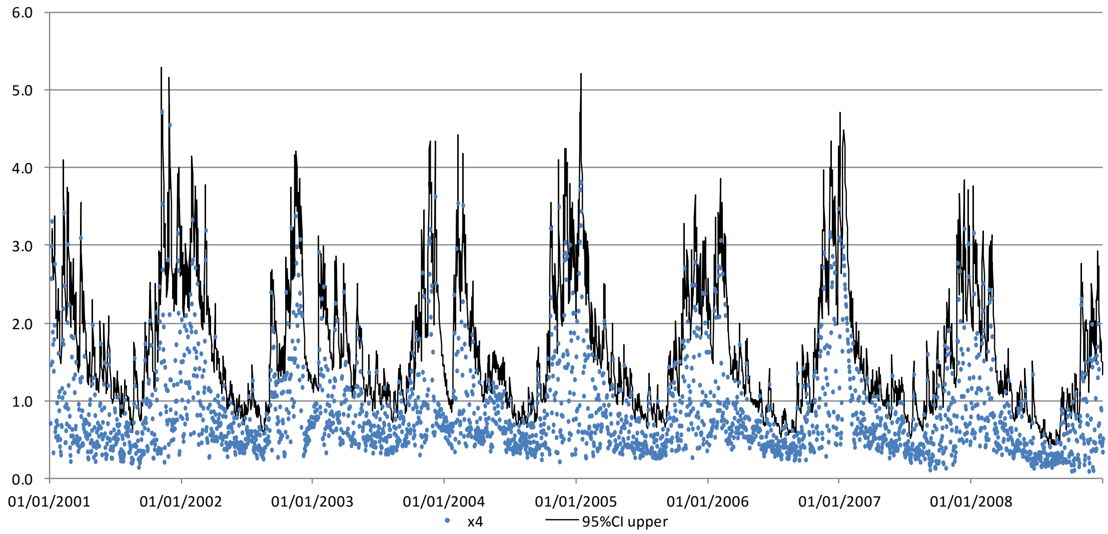
**Figures & Tables**

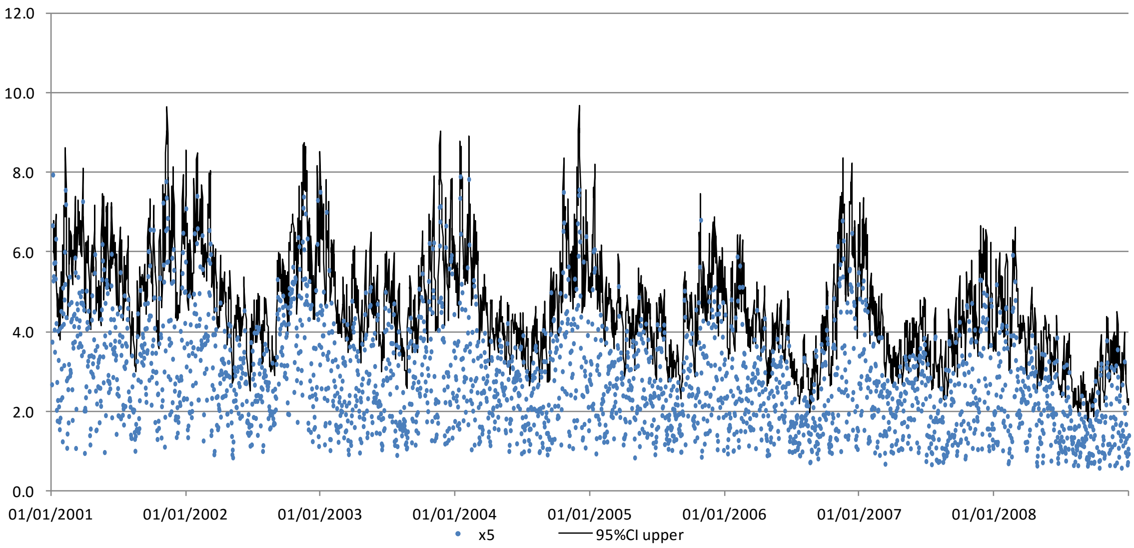
**Figure 1**. The values of the 6 air pollution variables, on a daily basis, against the 95% upper level confidence interval, estimated by the AR(4)-Diag-aVECH(1,1) model.
($x_1$: Athinas Station, $x_2$: Geoponiki Station, $x_3$: Marousi Station, $x_4$: Nea Smyrni Station, $x_5$: Patision Station, $x_6$: Peristeri Station).




6.0
5.0
4.0
3.0
2.0
1.0
0.0
01/01/2001
01/01/2002
01/01/2003
01/01/2004
01/01/2005
01/01/2006
01/01/2007
01/01/2008
x4
95%CI upper


12.0
10.0
8.0
6.0
4.0
2.0
0.0
01/01/2001
01/01/2002
01/01/2003
01/01/2004
01/01/2005
01/01/2006
01/01/2007
01/01/2008
x5
95%CI upper

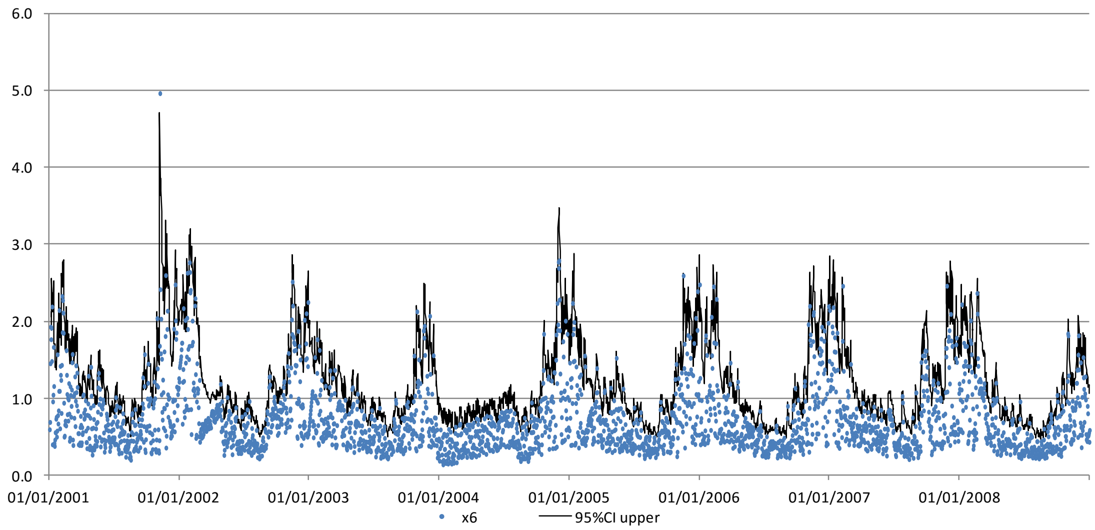

6.0
5.0
4.0
3.0
2.0
1.0
0.0
01/01/2001
01/01/2002
01/01/2003
01/01/2004
01/01/2005
01/01/2006
01/01/2007
01/01/2008
x6
95%CI upper

**Figure 2.** The daily time-varying correlations between the 6 air pollution variables estimated by the AR(4)-Diag-aVECH(1,1) model. Each figure plots the line graphs of time-varying correlation between variables $x_i$ and $x_j$, for $i$=1,…,6, $j$=2,…,6, and $j$>$i$. i.e. CORR_01_02 denotes the dymamic correlation bewteen $x_1$ and $x_2$
($x_1$: Athinas Station, $x_2$: Geoponiki Station, $x_3$: Marousi Station, $x_4$: Nea Smyrni Station, $x_5$: Patision Station, $x_6$: Peristeri Station).

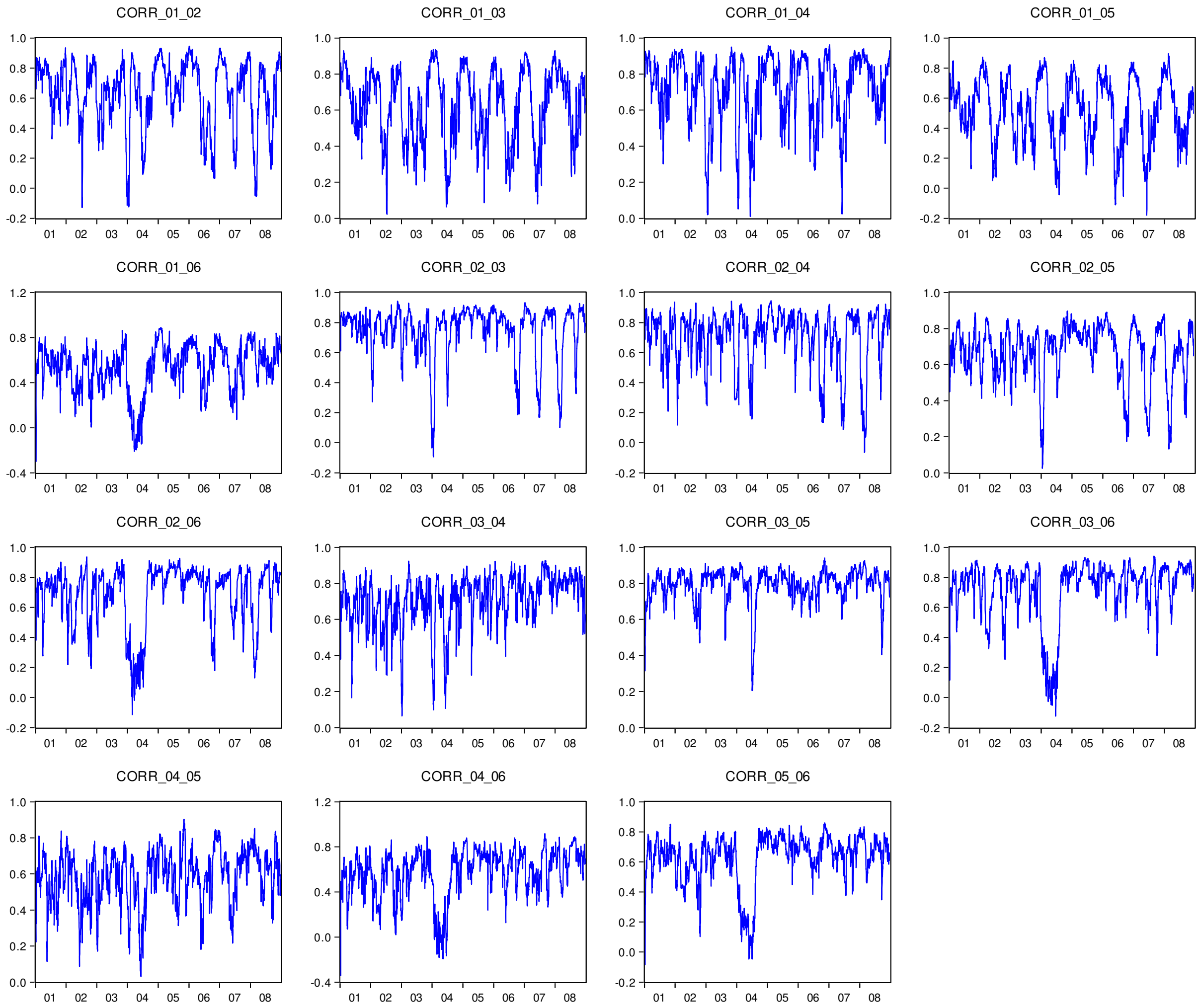


**Figure 3.** The distribution of $T_{1,t}$ and the phase I and phase II control charts.

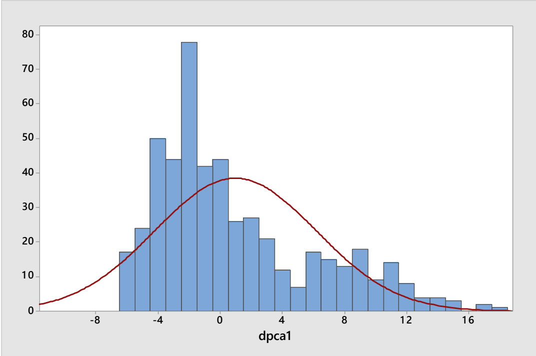



Figure 3a. The $T_{1,t}$ (DPCA1) distribution

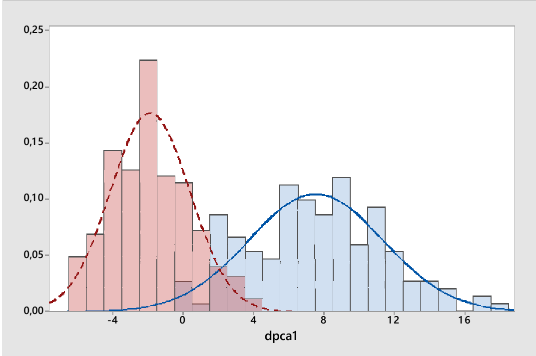



Figure 3b: The $T_{1,t}$ (DPCA1) distribution separately for Winter and Summer days

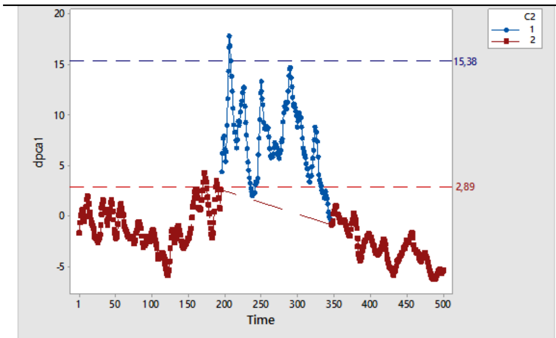



Figure 3c. Phase I control chart for $T_{1,t}$

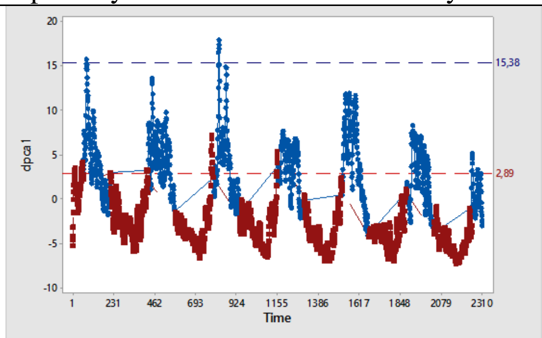



Figure 3d. Phase II control chart for $T_{1,t}$

**Figure 4.** The phase I (left panel) and phase II (right panel) control charts for $T_{2,t}$ .

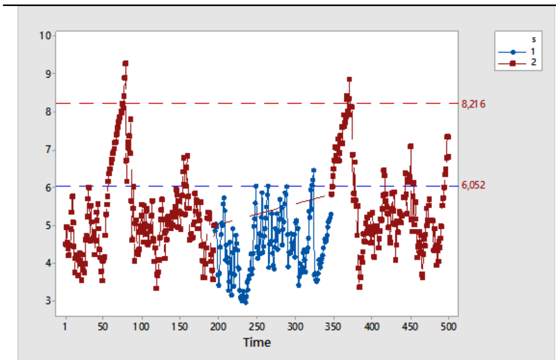



Figure 4a. Phase I control chart for $T_{2,t}$

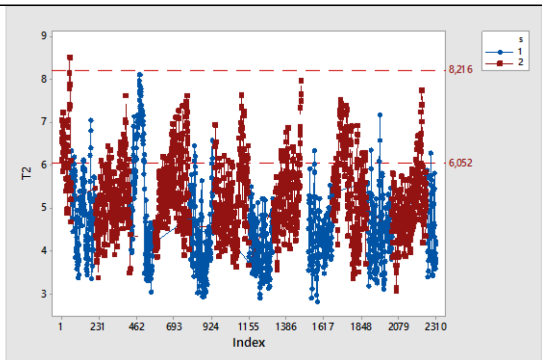



Figure 4b. Phase II control chart for $T_{2,t}$

**Table 1.** AR(4)-Diag-aVECH(1,1) model. 95% confidence interval's violations and the p-value of the LR statistic.

| Air pollution variable | % of violations | p-value of LR statistic |
|---|---|---|
| $x_{1,t}$ | 4.25% | 0.116 |
| $x_{2,t}$ | 4.38% | 0.117 |
| $x_{3,t}$ | 5.17% | 0.679 |
| $x_{4,t}$ | 4.87% | 0.726 |
| $x_{5,t}$ | 4.35% | 0.097 |
| $x_{6,t}$ | 5.00% | 0.990 |