

## A Threshold Autoregressive Asymmetric Stochastic Volatility Strategy to Alert of Violations of the Air Quality Standards

Montero Lorenzo, J. M.<sup>1\*</sup>, Garcia-Centeno, M. C.<sup>2</sup> and Fernandez-Aviles, G.<sup>1</sup>

<sup>1</sup> Facultad de Ciencias Jurídicas y Sociales de Toledo, University of Castilla-La Mancha  
Cobertizo San Pedro Mártir s/n, 45071, Toledo, Spain

<sup>2</sup> María Carmen García-Centeno, Facultad de Ciências Econômicas y Empresariales, Julián Romea  
23, 28003, Madrid, Spain

Received 12 July 2010;

Revised 17 Sep. 2010;

Accepted 27 Sep. 2010

**ABSTRACT:** Air quality is a topic of crucial importance, because air pollution is one of the most important pollution problems in the world. In particular, predicting or detecting a future extreme air pollution episode or predicting the violation of an air quality standard, is of crucial interest in the field of pollution control. There have been a variety of attempts to reach this purpose both from the perspective of the extreme value theory and the time series analysis, but as far as we know there is none successful strategy to alert of violations of the standards. This is why in this article we propose a new strategy, a threshold autoregressive asymmetric stochastic volatility strategy to alert of an immediate violation of the particulate matter quality standards, which take into account the different answer of the volatility to a positive or negative, but equal in magnitude, relative variation of the level of the pollutant in the previous period. Particulate matter is one of the still uncontrolled pollutants in big cities. This novel approach has been applied in Madrid City (Spain), the third-most populous municipality in the European Union, and it is able to predict a great percentage of violations of the standard.

**Key words:** Pollution, Particulate matter, Air quality violation, Volatility, environment, TA-ARSV model, Functional

### INTRODUCTION

Air quality is a topic of crucial importance, because air pollution is one of the most important pollution problems in the world. Many health problems (e.g., respiratory and cardiovascular) can be caused or worsened by exposure to air pollution on a day-to-day basis. The level of severity of these effects varies from advancing the day of death to less serious morbidity such as increased use of inhalers by asthmatics (Ayres, 2002). Epidemiological studies in the USA, continental Europe and the UK suggest that approximately 1% extra deaths may be brought forward by every 10  $\mu\text{g}/\text{m}^3$  increase in airborne particulate matter with a mass median diameter less than 10  $\mu\text{m}$  ( $\text{PM}_{10}$ ); it is estimated that particles contribute to around 8,100 deaths per year in urban areas of Great Britain (Ayres, 2002). Therefore, it is not surprising that the World Health Organization (WHO) has ranked the urban air pollution the 13th contributor to global deaths in its 2002 World Health Report.

In particular, predicting or detecting a future extreme air pollution episode or alert of an immediate violation of an air quality standard, is of particular interest in the field of pollution control. This is the reason why these two research fields are subjects of environmental concern.

The most widely used procedure to predicting or detecting a future extreme air pollution episode are extreme value models. One of the pioneers applications of the theory to air pollution is by Roberts [1979a and 1979b], who presents an application to  $\text{SO}_2$  and  $\text{NO}_2$  data in the Long Beach area, CA, USA. Surman *et al.* (1987) studied the ozone and applied the theory in Brisbane, Queensland, Australia. Smith (1989) proposed some extensions of the extreme values theory based on the point-processes view of high-level exceedances, and they were illustrated with ozone data collected in Houston, Texas. Sharma *et al.* (1999) provided predictions of the expected number of violations of the National Air Quality Standards of India

\*Corresponding author E-mail: jose.mlorenzo@uclm.es

for carbon monoxide concentrations from an urban road intersection. Ercelebi and Kirmanli (1999) predicted methane concentration in an underground mine environment. Lu and Fang (2003) applied the lognormal, Weibull and type V Pearson distributions to fit the parent distribution of the  $PM_{10}$  at five air monitoring stations in Taiwan from 1995 to 1999. Kan and Chen (2004) dealt with extreme value distributions for fitting the daily average concentration data of  $PM_{10}$ ,  $SO_2$ , and  $NO_2$  data in Shanghai. Hurairah *et al.* (2005) applied a new extreme value model to carbon monoxide data in Malaysia. Sfetos *et al.* (2006) modeled daily  $PM_{10}$  concentration values from an industrial area in Macedonia. Achcar *et al.* (2008a) considered the problem of estimating the number of times an air quality standard is exceeded in a given period of time; the theoretical development was applied to the measurements provided by the monitoring stations of Mexico City. Erceleby and Toros (2009) recently applied the extreme values theory to data of hourly  $SO_2$  and  $NO_2$  obtained from two permanent stations in Istanbul.

Linear and non-linear regression models have also been widely used for predicting future extreme air pollution episodes, overall in the ozone case (Robeson and Steyn, 1990, Hubbard and Cobourn, 1998, and Chaloulakou *et al.*, 1999, are some interesting works in the past). Classification and regression trees models (Burrows *et al.*, 1995) and neural networks (Baxt and White, 1995, van Aalst and de Leeuw, 1997, Comrie, 1997, and Gardner and Dorling, 2000, are good references in the past. Sharma *et al.* (2003), Wang and Lu (2006), Rost *et al.* (2009), and Barai *et al.* (2009) are recent recommended references) are other of the statistical procedures used to tackle this topic. However, they seem not to be a prominent research line nowadays. Canonical analysis and other techniques related with linear models have also been sporadically used with a certain success.

On the contrary, the more promising approach to the prediction, detection or alerting of a future extreme air pollution episode is time series analysis and, specifically, the use of volatility models. Time-varying volatility has been naturally relevant, topical, useful and interesting for modeling and forecasting financial data. The explosive growth of applications of econometrics to finance is due primarily to the increased availability of financial data, increased computer power and, of course, the greater interest in the performance of financial markets in economic discussions. But financial markets are not the only area where volatility plays an important role. Volatility, and over all extreme volatility, is relevant in other areas such as volcanic activity, the occurrence of earthquakes and tsunami, the evolution of weather patterns (such as temperature, wind, rainfall, motion of waves, and solar activity) and

environmental factors, such as air, water and soil pollution, among others. Fortunately, the two last decades have witnessed the increased availability of data about pollution measurements, mainly in big cities. This is why quantitative research is paying more and more attention to volatility of pollution data (without volatility, many temporal and spatial environmental variables would simply be constants) and to the modeling of such a volatility. Therefore, volatility also matters in the environmental field, and hence needs to be specified, estimated, tested and forecasted.

The different statistical models used to model air pollution data include time series modeling of the daily or weekly average pollution data (see for example, Loomis *et al.*, 1996; or Achcar *et al.*, 2008b). In this way, the use of stochastic volatility models (SV) has many advantages to analyze time series since they assume two processes to model the series: a process to model the observations and a process to model the latent volatilities (see for example, Ghysels *et al.*, 1996; Kim *et al.*, 1998; Meyer and Yu, 2000). This family of models have been extensively used to analyze financial time series (see for example, Danielsson, 1994; Yu, 2002), as a powerful alternative for the usual existing ARCH type models introduced in the seminal paper on Autoregressive Conditional Heteroskedasticity (ARCH) models by the 2003 joint Nobel Laureate in Economic Sciences, Engle (1982), which was subsequently generalized (GARCH modeling) in Bollerslev (1986), among others.

Another interesting approach frequently used to analyze SV models is Bayesian inference approach using Markov Chain Monte Carlo (MCMC) methods (see for example, Gelfand and Smith, 1990; Smith and Roberts, 1993). This approach allows overcoming great difficulties using standard classical inference approach, as high dimensionality, likelihood function with no closed form and high computational cost. Recently, Gyarmati-Szabó *et al.* (2008) have introduced the use of bivariate stochastic volatility models applied to air pollution data. In particular, they develop the Bayesian analysis using Markov Chain Monte Carlo (MCMC) methods to simulate samples for the joint posterior distribution of interest. As far as we have checked, relative variations in high frequency data in pollution are no significantly different from their financial analogous: constant and statistically null mean, and non constant variance with high volatility periods that alternates with others of low volatility (that is to say, there are evidence of volatility clusters).

But neither the above mentioned GARCH model nor the ARSV models are able to capture (or to appropriately capture) the stylized facts of financial series that are shared by the pollution series we have exam-

ined: persistence in volatility, Taylor effect, leverage effect, etc. This is the reason why it has emerged a wide range of derivations of GARCH and ARSV models that try to capture the above-mentioned stylized facts. This article focuses in a extremely important stylized fact in the environment field—the asymmetric answer of pollution volatility to shocks of equal relative magnitude but of different sign— and proposes an original model, the Threshold Asymmetric Autoregressive Stochastic Volatility Model (TA-ARSV model), to analyze that stylized fact. Thresholds were included in stochastic volatility models by So *et al.* (2002), and they have been recently developed by García and Mínguez (2009). As far as we know, it is the first time that a TA-ARSV model is used in the environment field.

In cases where the level of a particular pollutant is close to the standard, TA-ARSV strategies could be a guide for the authorities with competencies in pollution control that would avoid taking unnecessary and costly measures for citizens and companies through distinguishing between false and true alarms of violation of the air quality standards. It also would help them to avoid violations of the standards by taking the appropriate measures days before to the potential violation.

In particular, in this article a TA-ARSV strategy is used to help the authorities to prevent violations in particular matter (specifically  $PM_{10}$ ) standards. Particulate matter (PM) is the term used for a mixture of solid particles and liquid droplets suspended in the air. These particles originate from a variety of sources, such as power plants, industrial processes, and diesel trucks, and they are formed in the atmosphere by transformation of gaseous emissions. Their chemical and physical compositions depend on location, time of year, and weather.

Particulate matter is one of the six criteria pollutants, and the most important in terms of adverse effects on human health. This makes it especially dangerous and this is the reason that many epidemiological studies of PM health effects have been completed. Especially particles that are smaller than 10 microns ( $PM_{10}$ ), are likely to cause adverse health effects including increasing morbidity and mortality in susceptible individuals. In particular, the lifetime of  $PM_{10}$  is from minutes to hours, and its travel distance varies from less than 1km to 10 km. This is why governments have made great effort to maintain this pollutant under control. But, despite the significant improvements made over the last three decades,  $PM_{10}$  continues to exert a public health impact.

This is precisely the case of Madrid, the study site in this article. Whereas in the last years  $SO_2$  and CO seems to be under control, and measures to reduce the level of other pollutants have been certainly suc-

cessful,  $PM_{10}$  continues to be one of the air pollution problems that most worry the Madrid Municipality. In accordance to the current legislation, levels of  $PM_{10}$  are not satisfactory in Madrid City, although it is true that  $PM_{10}$  levels in Madrid have an important anthropogenic component: Saharan winds.

## MATERIALS & METHODS

The data used in this paper have been provided by the Atmosphere Pollution Monitoring System of Madrid municipality, Spain, the third-most populous municipality in the European Union (Fig. 1). They have been hourly measured at the 25 fixed operative monitoring stations since January 2000 until December 2008, and normalized to the temperature of 293 K and to the pressure of 101.3 kPa. Subsequently, we have computed daily means, to be able to compare the volume of the pollutant with the daily standard of Madrid City. Fig. 2 shows the locations of the air quality monitoring stations.

As can be seen in Fig. 2, most monitoring stations are located in the urban centre and relatively few in the peripheral sites. Note the reasonable coverage of the domain under study by the monitoring stations since most of Madrid population is concentrated in the urban centre.

The data base had some gaps due to breakdowns in the monitoring stations, the change of location of some of them, etc. These gaps ranged from one day to two months and it has been used an ordinary functional kriging (OFK) strategy for filling them, because it is considered the most appropriate strategy for predicting the values in long periods of time in a spatial framework. Other alternatives are universal kriging and cokriging, but the “leave one out” procedure shown that these non-functional alternatives provided greater mean square errors than the functional one.

OFK is an adaptation of Giraldo (2009) of the functional analysis recently developed by Ramsay and Silverman (2005) that tackles the problem of spatial prediction of functional data. In OFK strategy, the predicted curve is a linear combination of the observed curves in other locations (in our case monitoring stations) where the coefficients are real numbers.

The statistical statements are as follows: Let us consider a functional random process  $\chi_s : s \in D \subseteq \mathbb{R}^d$ , usually  $d=2$ , such that  $\chi_s$  is a functional variable for any  $s \in D$ . Let  $s_1, s_2, \dots, s_n$  be arbitrary locations in  $D$  (in our case the sites where the monitoring stations are located), and assume that we can observe a realization of the functional random process  $\chi_s$  at these sites,

$$\chi_{s_1}, \chi_{s_2}, \dots, \chi_{s_n}.$$

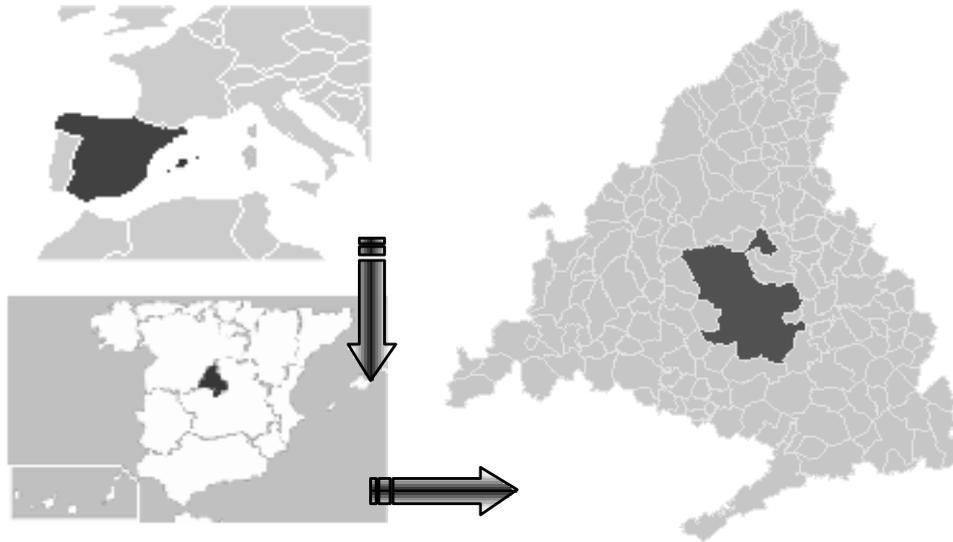


Fig. 1. Location of Madrid City

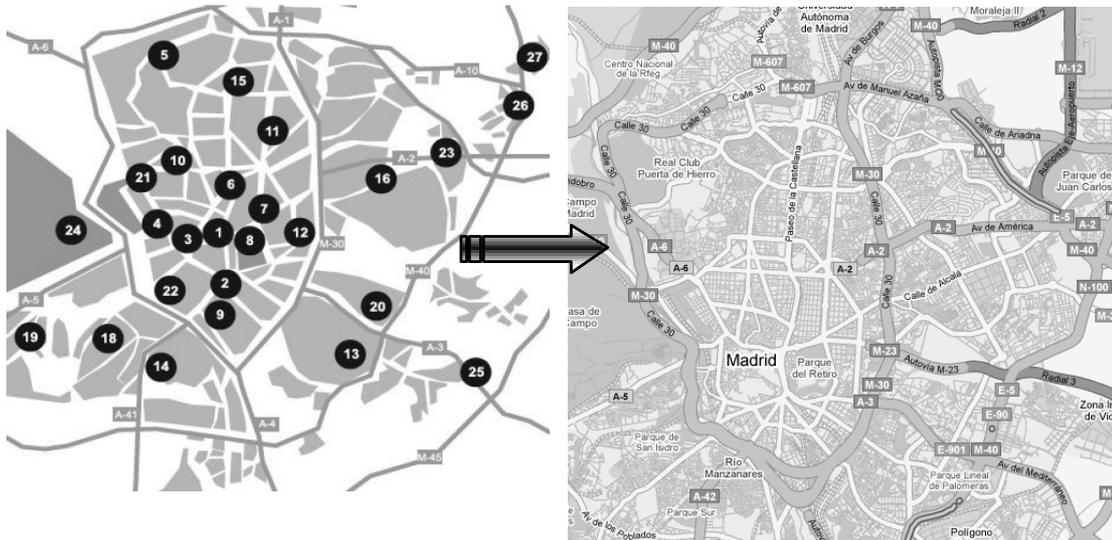


Fig. 2. Madrid City. Location of the air quality monitoring stations

The predictor we propose has the same expression as the ordinary kriging one (Cressie, 1993) but with curves instead of variables:

$$\chi_{s_0} = \sum_{i=1}^n \lambda_i \chi_{s_i} \quad (1)$$

That is to say, this approach treats the whole curve as a single entity, and the weights in the predictor give more influence to the curves of locations closer to the prediction point,  $s_0$ , than other separated. To find the best linear unbiased predictor (BLUP) we extend the criterion given in Myers (1982) to the functional case, assuming second-order stationarity and isotropy in the random process (this is the case in the light of the data base), and the  $n$  weights in the kriging predictor of  $\chi_{s_0}$  are given by the solution of the following optimization problem:

$$\min_{\lambda_1, \lambda_2, \dots, \lambda_n} \int_T V(\hat{\chi}_{s_0}(t) - \chi_{s_0}(t)) dt \quad \text{s.t.} \quad \sum_{i=1}^n \lambda_i = 1$$

(unbiasedness constraint)

Observe that the unbiasedness constraint and Fubini Theorem imply that

$$\min_{\lambda_1, \lambda_2, \dots, \lambda_n} \int_T V(\hat{\chi}_{s_0}(t) - \chi_{s_0}(t)) dt =$$

$$E \left[ \int_T (\hat{\chi}_{s_0}(t) - \chi_{s_0}(t))^2 dt \right]$$

On the other hand, the integral in the above equation can be written as:

$$\int_T V(\hat{\chi}_{s_0}(t) - \chi_{s_0}(t)) dt = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \int_T C_{ij}(t) dt + \int_T \sigma^2(t) dt - 2 \sum_{i=1}^n \lambda_i \int_T C_{i0}(t) dt$$

where  $C_{ij}(t)$  is the value of the spatial covariance function for the observed locations  $s_i$  and  $s_j$ ,  $C_{i0}(t)$  is the analogous for the observed location and the unobserved site, and  $\sigma^2(t)$  is the variance of the random process.

As a consequence, the objective function can be expressed as:

$$\int_T V(\hat{\chi}_{s_0}(t) - \chi_{s_0}(t)) dt = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \int_T C_{ij}(t) dt + \int_T \sigma^2(t) dt - 2 \sum_{i=1}^n \lambda_i \int_T C_{i0}(t) dt - 2 \mu \left( \sum_{i=1}^n \lambda_i - 1 \right)$$

The result of the optimization process, in matrix notation, is the following one:

$$\begin{pmatrix} \int_T C_{11}(t) dt & \dots & \int_T C_{1n}(t) dt & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \int_T C_{i0}(t) dt & \dots & \int_T C_{i0}(t) dt & 1 \\ \int_T C_{i0}(t) dt & \int_T C_{i0}(t) dt & \dots & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_i \\ \mu \end{pmatrix} = \begin{pmatrix} \int_T C_{10}(t) dt \\ \vdots \\ \int_T C_{i0}(t) dt \\ 1 \end{pmatrix} \quad (2)$$

The prediction variance can be easily obtained from the first  $n$  equations of the above system of equations:

$$\sigma^2 = \int_T \sigma^2(t) dt - \sum_{i=1}^n \lambda_i \int_T C_{i0}(t) dt - \mu \quad (3)$$

Once the data base gaps have been filled, data have been daily averaged or each hour. Table 1 displays the range, mean, and standard deviation of daily averaged  $PM_{10}$  values in the period under study.

**Table 1.  $PM_{10}$ : Main descriptive statistics (whole monitoring stations net), 2000-2008**

|            | Min.  | Max.  | Mean  | Standard deviation | Skewness Coeff. |
|------------|-------|-------|-------|--------------------|-----------------|
| $PM^{(a)}$ | 13.44 | 49.08 | 32.16 | 8.75               | 1.302           |

$\mu/m^3$ ; Source: Own elaboration.

After presenting the kriged strategy to complete the series of  $PM_{10}$  in the period under study, in the sequel

we proceed to develop the TA-ARSV model we propose as a useful tool for alerting of violations of the  $PM_{10}$  standard.

As it can be shown, the relative variation of the magnitude of pollutants is characterized by both a changing conditional variance and a mean statistically equal to zero. This is the reason why it is possible to model its behavior as follows:

$$y_t = \sigma_t \varepsilon_t \quad \varepsilon_t \sim iid(0,1), t = 1, \dots, T \quad (4)$$

where  $\sigma_t$ , the volatility, has a dynamic structure and it is assumed to be generated by a stationary process, so that its value at period  $t$  depends on the previous information ( $\Omega_{t-1}$ ).  $\varepsilon_t$  is a random disturbance (white noise), independent of  $\sigma_t$ , which is distributed with zero mean, unit variance and finite fourth order moments.

As said before, GARCH and ARSV models have been traditionally used to model the changing conditional variance over time. But they are not able to explain the possible asymmetric answer of volatility, and this is the reason why in this paper we propose as an alternative the Threshold Autoregressive Asymmetric Stochastic Volatility (TA-ARSV) model to explain pollutant dynamics.

TA-ARSV model includes two new parameters in the volatility equation of ARSV model,  $\phi_{11}$  and  $\phi_{12}$ , to capture the asymmetric behavior of volatility. They measure the effect of a positive and a negative relative variation of the same magnitude in the instant  $t-1$ , respectively, on the volatility of the relative variation of the magnitude of a particular pollutant in instant  $t$ . Additionally, it is necessary to define the two following indicator matrices:

$$I_{1t} = \begin{cases} 1 & \forall t \text{ in case that the pollutant variation is positive or zero} \\ 0 & \text{otherwise} \end{cases}$$

$$I_{2t} = \begin{cases} 1 & \forall t \text{ in case that a pollutant variation is negative} \\ 0 & \text{otherwise} \end{cases}$$

Specifically, the TA-ARSV model we propose is the TA-ARSV(1), because the estimated ACF and PACF computed for several air quality data bases have shown that it is only necessary one lag to explain the dynamics of volatility. TA-ARSV(1) is defined by the following equations:

• Mean equation:

$$y_t = (\sigma_* \exp(0.5h_t)) \varepsilon_t \quad \varepsilon_t \square N(0,1) \quad (5)$$

In particular, the steps of the algorithm are:

a) To obtain an initial Gaussian model for an initial vector of the parameters of the model. The initial val-

- Volatility equation:

$$h_t = (\phi_{11}I_{1t} + \phi_{12}I_{2t})h_{t-1} + \eta_t \quad (6)$$

$$|\phi_{11}| < 1; |\phi_{12}| < 1 \quad \eta_t \sim N(0, \sigma_\eta^2)$$

where,  $y_t$  is the variation in the relative magnitude of the pollutant;  $\sigma_*$  is a positive scale factor,  $\varepsilon_t$  is a random disturbance (white noise) in the mean equation that follows a zero mean and unit variance Gaussian distribution,  $\sigma_t$  is the volatility and it is modeled as an exponential function to guarantee its positivity, and  $h_t$  denotes the log-volatility. Equation (6) indicates that the log-volatility follows an AR(1) process in each regime;  $\eta_t$  is a white noise process and it follows a Gaussian distribution with zero mean and variance  $\sigma_\eta^2$ ; The distributions of  $\varepsilon_t$  and  $\eta_t$  are independent. It is also assumed that the disturbances of mean and volatility equation are independent. The values of  $\phi_{11}$  and  $\phi_{12}$  are less than one in absolute terms to assure that both regimes are stationary in covariance.

It can be noticed that in the TA-ARSV(1) model, in equation (5), volatility is defined as an exponential function, which implies that the model is non linear. However, as shown in Sandmann and Koopman, 1998, a simple logarithmic transformation turns it into a linear model. The transformed model is:

- Mean equation:

$$Y_t = \log(y_t^2) = \log(\sigma_*^2) + h_t + \xi_t \quad (7)$$

- Volatility equation:

$$h_t = (\phi_{11}I_{1t} + \phi_{12}I_{2t})h_{t-1} + \eta_t \quad (8)$$

$$|\phi_{11}| < 1; |\phi_{12}| < 1 \quad \eta_t \sim N(0, \sigma_\eta^2)$$

Equation (7) shows that the logarithm of the squared relative variation of the pollutant magnitude is obtained as the sum of a constant and two independent stochastic processes. These processes are the volatility ( $h_t$ ), which is a stationary linear process, and the random disturbance ( $\xi_t$ ), which follows Chi-square distribution with one degree of freedom. Note that this equation is the well-known measurement equation in terms of a state-space modeling.

Equation (8), the transition equation in a state-space strategy, represents the dynamics of the volatility along the time. Its behavior in the instant  $t$  depends on:

- Whether there has been a positive or negative relative variation of the magnitude in  $t-1$ .
- The magnitude of volatility in the previous period.
- And finally, the innovation, which is assumed to

follow a  $N(0, \sigma_\eta^2)$  distribution and to be uncorrelated with the mean equation disturbances.

Therefore, the TA-ARSV(1) strategy can be rewritten in a state space form as follows:

$$\begin{pmatrix} h_{t+1} \\ Y_t \end{pmatrix} = \delta_t + \Phi h_t + u_t$$

where

$$u_t \sim (0, \Omega_T); \quad \delta_t = \begin{pmatrix} 0 \\ \ln \sigma_*^2 \end{pmatrix}; \quad \Phi = \begin{pmatrix} \phi_{11}I_{1t} + \phi_{12}I_{2t} \\ 1 \end{pmatrix};$$

$$\Omega_T = \begin{pmatrix} \sigma_\eta^2 & 0 \\ 0 & \frac{\pi^2}{2} \end{pmatrix}$$

Provided that TA-ARSV(1) is a non Gaussian model, the estimation of the parametric

vector  $(\phi_{11}, \phi_{12}, \sigma_*, \sigma_\eta^2)$  requires that the likelihood function is evaluated by using the Monte Carlo method due that it approximates non Gaussian models by importance sampling.

The values of the parameters are estimated from the available information.

- To compute the Gaussian likelihood function for the initial model by using the Kalman filter.
- Repeat steps a) and b) until the desired level of convergence is achieved.

Finally, the parametric vector that maximizes the simulated likelihood function is obtained by using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method, a well-known method to solve unconstrained nonlinear optimization problems. The leverage effect is checked by testing the null hypothesis:  $H_0: \phi_{11} = \phi_{12}$  (both regimes have equal coefficients) against the alternative:  $H_1: \phi_{11} \neq \phi_{12}$ . Since the TA-ARSV(1) model encompasses the ARSV(1), this test can be viewed as testing the ARSV(1) model against the TA-ARSV(1) strategy. The fact that both the null and the alternative hypothesis refer to two nested models allows for the implementation of a likelihood ratio test, the test statistic being  $\lambda = -2(\ln L^R - \ln L)$ , which follows a chi-squared distribution with one degree of freedom. In case that the null hypothesis is not rejected, then there is not evidence of an asymmetric answer of volatility. In this case the ARSV(1) model is the right model. Alternatively, the rejection of the null hypothesis suggests a different effect of positive and negative shocks on the dynamics of the volatility.

## RESULTS & DISCUSSION

From the observation of the estimates of parameters  $\phi_{11}$  and  $\phi_{12}$ , which capture the asymmetric behavior of volatility in  $PM_{10}$  daily series in every one of the monitoring stations that have been operating in Madrid City during the period under study (Fig. 3), it can be deduced that there is not a general behavior across the monitoring net during the period 2000-2008. That is to say, in some of the monitoring stations  $\phi_{11}$  exceeds  $\phi_{12}$ , and in others it occurs just the opposite.

According to Section 2, those coefficients measure the effect of a positive and a negative relative variation of the same magnitude of  $PM_{10}$  in the instant  $t-1$ , respectively, on the volatility of the relative variation of the magnitude of the pollutant in instant  $t$ . It can be also noticed that the estimates of both parameters are positive, which indicates a positive relation between the volatility in two consecutive instants of time.

In particular,  $\phi_{11}$  exceed  $\phi_{12}$  in the South part of the city, which is characterized by its high density of industrial parks and an intensive build-up. The main sources of  $PM_{10}$  in this area are local re-suspension (the aridity of this part of the city give room for a cluster of factors that jointly favors this phenomenon); a poor vegetal coverage jointly with the fact that most of cultivation next to the South part of the city are dry-farmed crops, which maintain almost all year the soil exposed to the wind action; a low level of precipitations, that also helps to maintain the soil dry and subject to erosion; an intense convective dynamics induced by a high insolation over unprotected soils in spring and summer; and construction and demolition, a crucial factor in Madrid City, due to the intensive construction in the area during the two last decades. On the contrary,  $\phi_{11}$  is lesser than  $\phi_{12}$  in the city centre (extensive in pedestrian areas) and the North part of the city, where a handful of institutions are located, the income level is certainly high, build-up is dispersed and green areas characterizes the zone. However, the level of  $PM_{10}$  in the city centre, despite to be a pedestrian area, is extremely affected by the worse traffic that characterizes the surrounding areas. There is a particular source that affects the entire city: the long distance transport of mineral powder coming from Sahara or Shael. On the other hand, it is not negligible the effect of punctual accidents such as urban fires near to the monitoring stations or even forest fires in areas relatively remote from them.

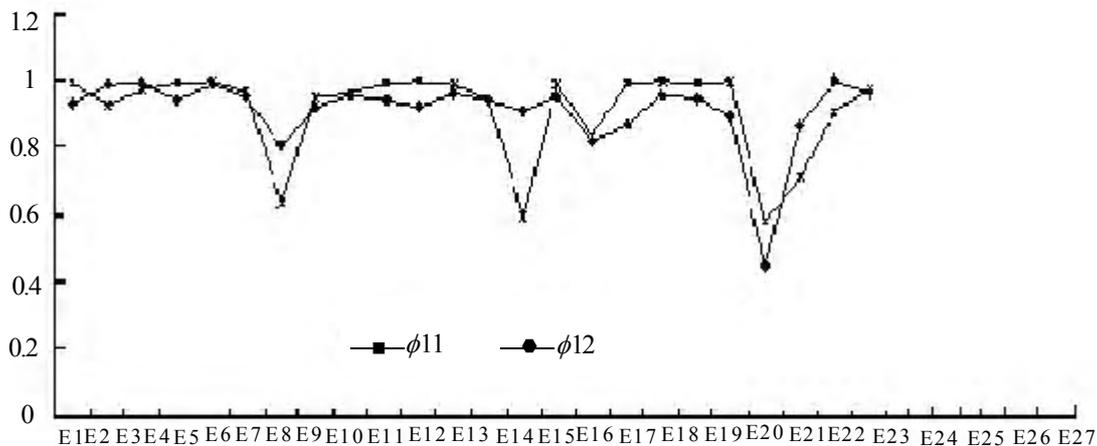
The fact that the estimated value of  $\phi_{12}$  does not exceed the estimated value of  $\phi_{11}$  (in the South part of the city) implies that the effect on volatility is greater in case that the relative variation of the level of  $PM_{10}$  increases than in case that it diminishes. Thus, neglecting the effect that other variables not considered in the analysis can have on the behavior of the level of  $PM_{10}$ , in case of a level close to the standard and a relative increment with respect to the previous period, authorities must be alert because there is a high probability that the standard will be exceeded in the next period. Additionally, the estimate of  $\phi_{11}$  is always very close to the unity (there is no cases with  $\phi_{11}$  equal to the unity, that is to say the processes are stationary), which implies that high volatility in  $t-1$  will lead to high volatility in  $t$ , and it reinforces the above stated thesis. It occurs just the opposite in the city centre and the North part of the city: the estimates of  $\phi_{11}$  are greater than the estimates of  $\phi_{12}$ , and they are certainly close to the unity. This implies that an increment in the magnitude of  $PM_{10}$  in period  $t-1$ , in a tessitura of high volatility, is not so worrying, in terms

of a violation of the standard in period  $t$ , as it is in the South part of the city. However, environmental authorities must be worried with small decreases of the level of the pollutant when both it is close to the standard and the volatility is significant, because in this case there is a high risk of violating the standard. Putting the focus of the analysis in the number of violations of the  $PM_{10}$  daily standard in Madrid City (the number of days that at least a monitoring station has violated the standard is 897 in the period under study), from the data it can be deduced that in 2008, the last year of the study, this standard was violated between 35 (the maximum number of times that the daily standard can be exceeded in a year) and 62 times in monitoring stations 1, 6, and 9; between 30 and 35 times in stations 5, 10, 22; and 25, and between 20 and 30 in stations 7, 8, 14, 19, and 20. In the remainder stations the number of violations did not exceed 20. There was found none monitoring station that does not violate the standard during 2008.

Regarding to the number of violations, in Madrid City the standard for human health protection (in 24 hours average terms) was set in October, 2002, in  $50\mu\text{g}/\text{m}^3$  (with a tolerance range or  $15\mu\text{g}/\text{m}^3$  that was reduced every year by  $5\mu\text{g}/\text{m}^3$ ) that could not be exceeded more than 35 days. In January, 2005 the new standard was maintained in  $50\mu\text{g}/\text{m}^3$ , but without tolerance range, and the maximum number of violations a year was fixed at 7. However, this new standard has come into effect in January, 2010. Additionally this new normative has established very close daily and annual legal limits, which requires not only low levels of  $PM_{10}$  but also a low variation of them along the year. This is not precisely the case of Madrid City (see Fig. 4) and, of course, it reinforces the thesis that volatility and its asymmetric answer is extremely important in case of a level of pollution close to the legal standard.

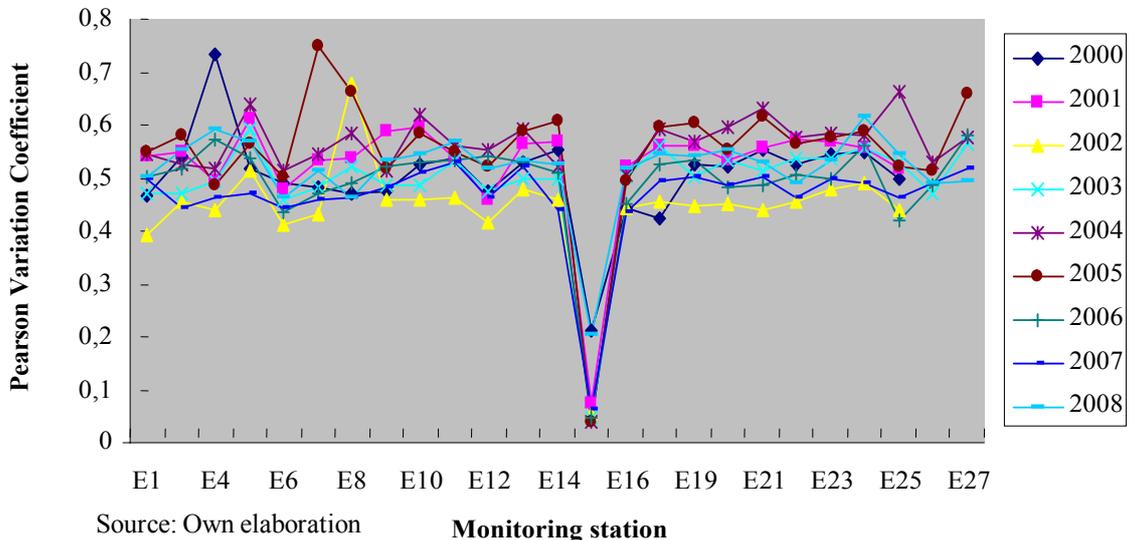
Note that monitoring stations 1 and 6 are in the city centre, whereas the station 9, the most conflictive, is in the South part of the city. Stations 5 and 10 are also in the centre of the city, whereas stations 22 and 25 are located in the South. Hence, it seems not to be a spatial relation in the number of violations, but there is in the magnitude relation of the parameters that estimate the asymmetric answer of volatility. In fact, monitoring stations 1, 6, 9, 5, 10 are located in areas with a high traffic density and a large volume of development of infrastructures and edification works in this decade. It is true that the area where station 1 is located is a pedestrian area, but it is also true that lifetime of  $PM_{10}$  is from minutes to hours, and its travel distance varies from less than 1 km to 10 km.; and the area that surrounds the zone where monitoring station 1 is sited is the area with worst traffic in the city.

Of course, a good system of alerts could be a crucial tool to avoid violations of the standard in the areas affected by traffic emissions and both edification and infrastructures works, without unnecessarily



Source: Own elaboration

Fig. 3. Estimated values of  $\phi_{11}$  and  $\phi_{12}$  at the operative monitoring stations. PM<sub>10</sub>, Madrid City



Source: Own elaboration

Fig. 4. PM<sub>10</sub> Pearson Variation Coefficient by monitoring station and year

stopping the activity in the construction sector. Thus, the core question is: How many violations of the standard could be alerted by taking into account the results obtained from the TA-ARSV model? This is a core question since the excessive number of violations of the PM<sub>10</sub> standard is not only a problem of Madrid City, but also of a handful of big cities. Mediterranean European cities in general are particularly problematic. In Table 2, it can be observed that a TA-ARSV strategy could have alerted at least 75% of the violations in 14 out of a total of 22 operative monitoring stations (stations 3, 5, 6, 10, 11, 12, 13, 14, 18, 19, 22, 23, 24 and 25). The number of violations that the TA-ARSV would have been able to predict in stations 5, 7, 8, 16 and 20 ranges from 50% to 75%. In monitoring stations 1, 9 (the most problematic ones) and 21, the violations previously detected by the proposed model does not reach the 50%, but in spite of that the prediction of 35% or more of the violations in the period 2000-2008 would have been of great help for Madrid environmental authorities. In the light of these results, it is clear that TA-ARSV strategies can be considered an extremely useful tool for environmental protection.

**CONCLUSION**

Given that there is a generalized consensus that air quality control is a topic of crucial importance, any help to prevent a violation of an air quality standard is of particular interest for the authorities with responsibility in the environmental field. Particulate matter is especially dangerous. In particular, particles that are smaller than 10 microns (PM<sub>10</sub>) are likely to cause adverse health effects including increasing morbidity and mortality in susceptible individuals. But, despite the significant successes in this field in the last three decades, PM<sub>10</sub> continues to exert a public health impact. This is precisely the case of Madrid, the study site in this article, where PM<sub>10</sub> continues to be one of the air pollution problems that most worry the Madrid Municipality. In accordance to the current legislation, levels of PM<sub>10</sub> are not satisfactory in Madrid City, although it is true that PM<sub>10</sub> levels in Madrid have an important anthropogenic component: Saharan winds.

Of course, in case that the level of PM<sub>10</sub> is close to the standard, authorities can implement a variety of measures that, in most cases, drastically affect both

**Table 2. Percentage of violations that could have been alerted by a TA-ARSV strategy (2000-2008) in the different monitoring stations (E)**

|            |            |            |            |            |            |            |            |            |            |            |
|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| <b>E1</b>  | <b>E3</b>  | <b>E4</b>  | <b>E5</b>  | <b>E6</b>  | <b>E7</b>  | <b>E8</b>  | <b>E9</b>  | <b>E10</b> | <b>E11</b> | <b>E12</b> |
| 46%        | 84%        | 61%        | 85%        | 89%        | 54%        | 70%        | 36%        | 80%        | 85%        | 86%        |
| <b>E13</b> | <b>E14</b> | <b>E16</b> | <b>E18</b> | <b>E19</b> | <b>E20</b> | <b>E21</b> | <b>E22</b> | <b>E23</b> | <b>E24</b> | <b>E25</b> |
| 82%        | 78%        | 63%        | 85%        | 87%        | 56%        | 35%        | 93%        | 84%        | 79%        | 80%        |

Source: Own elaboration

the economic activity and the daily life of citizens. To avoid such kind of disturbances, an efficient alert system must take into account volatility, because it can discriminate between situations close to the standard level followed by a violation and others followed by a decrease in the level of the pollutant. In this article we have proposed a new Threshold Autoregressive Asymmetric Stochastic Volatility model to explain PM<sub>10</sub> dynamics, and to serve as a system of alert of violations of the legal standard. This strategy includes two new parameters in the volatility equation of ARSV model ( $\phi_{11}$  and  $\phi_{12}$ ) to capture the asymmetric behavior of volatility, due to the asymmetric PM<sub>10</sub> answer to volatility. And this extremely important stylized fact must be taken into account. From the observation of the estimates of parameters  $\phi_{11}$  and  $\phi_{12}$  in PM10 daily series in every one of the monitoring stations that have been operating in Madrid City during the period under study, it can be deduced that there is not a general behavior for the monitoring net. It can be also noticed that the estimates of both parameters are positive, which indicates a positive relation between the volatility in two consecutive instants of time.

In particular, the estimated value of  $\phi_{12}$  does not exceed the estimated value of  $\phi_{11}$  (they capture the asymmetric behavior of volatility in instant  $t$  in case of a positive /negative PM<sub>10</sub> relative variation of the same magnitude in  $t-1$ , respectively), in the South part of the city, which implies that the effect on volatility is greater in case that the relative variation of the level of PM<sub>10</sub> increases than in case that it diminishes. Thus, in case of a level close to the standard that means a relative increment with respect to the previous period, authorities must be alert because there is a high probability that the standard will be exceeded in the next period. In the city centre and the North part of the city it occurs just the opposite: the estimates of  $\phi_{11}$  are greater than the estimates of  $\phi_{12}$ , and are certainly close to the unity. This implies that an increment in the magnitude of PM<sub>10</sub> in period  $t-1$ , in a tessitura of high volatility, is not so worrying, in terms of a violation of the standard in period  $t$ , as it is in the South part of the city. However, environmental authorities must be worried with small decreases of the level of the pollutant when it is close to the standard and the volatility is significant, because in this case there is a high risk of violating the standard. Regarding as the proposed strategy as an alert system of violation of the legal standard, the TA-ARSV strategy we propose would have alerted at least the 75% of the violations in 14 out of a total of 22

operative monitoring stations, between 50% and 75% in 5 stations and between 35% and 50% in the remainder 3 stations. Therefore, it can be concluded that TA-ARSV strategies can be considered as an extremely useful tool for environmental protection.

**ACKNOWLEDGEMENT**

We thank Emilio Porcu and Moreno Bevilaqua for valuable research assistance and helpful comments. We thank Dorota Witkowska, José Mondéjar and Manuel Vargas for helpful discussions on an earlier draft of the paper. We are grateful to the Madrid Municipality for providing most of the data.

**REFERENCES**

Achar, J. A., Fernández-Bremauntz, A. A., Rodrigues, E. R. and Tzintzun, G. (2008a). Estimating the number of ozone peaks in Mexico City using a non-homogeneous Poisson model. *Environmetrics*, **19** (5), 469-485.

Achar, J. A., Zozolotto, H. C. and Rodrigues, E. R. (2008b). Bivariate stochastic volatility models applied to Mexico City ozone pollution data. (In: F. Columbus (Ed.), *Air quality in the 21st century* (pp. 241-267). (New York: Nova Science Publishers Inc).

Ayres, J. G. (2002). Chronic effects of air pollution. *Occupational and Environmental Medicine*, **59**, 147-148.

Barai, S.V., Gupta, A.K. and Kodali, J. (2009). Air Quality Forecaster: Moving Window Based Neuro Models. In: (E. Avineri (Ed.) *Applications of soft computing* pp. 137-146), ASF 52, (Berlin: Springer-Verlag).

Baxt, W. and White, H. (1995). Bootstrapping confidence intervals for clinical input variable effects in a network trained to identify the presence of acute myocardial infarction. *Neural Comput*, **7**, 624-638.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heterocedasticity. *J. Econ.*, **31**, 307 -327.

Burrows, W. R., Benjamin, M., Beauchamp, S., Lord, E. R., McCollor, D. and Thomson, B. (1995). CART decision tree statistical analysis and prediction of summer season maximum surface ozone for the Vancouver, Montreal, and Atlantic Regions of Canada. *Journal of Applied Meteorology*, **34**, 1848-1862.

Chaloulakou, A., Assimacopoulos, D. and Lekkas, T. (1999). Forecasting daily maximum O<sub>3</sub> concentrations in the Athens Basin. *Environmental Monitoring and Assessment*, **56**, 97-112.

Comrie, A. (1997). Comparing neural networks and regression models for ozone forecasting. *J. Air Waste Manag. Assoc.*, **47**, 653-663.

- Cressie, N. (1993). *Statistics for Spatial Data*. Wiley Series in Probability and Mathematical Statistics, (New York: John Wiley and Sons Inc).
- Danielsson, J. (1994). Stochastic volatility in asset prices: estimation with simulated maximum likelihood. *J. Econ.*, **64**, 375-400.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, **50**, 987-1007.
- Ercelbebi, S. G. and Kirmanli, C. (1999). Statistical prediction of high methane concentration in underground coal mines. *Inst. Mining Metall. Eng. Trans.*, **108**, 132- 135.
- Ercelbebi, S. G. and Toros, H. (2009): Extreme Value Analysis of Istanbul Air Pollution Data. *Clean*, **37** (2), 122-131.
- García Centeno, M. C. and Mínguez Salido, R. (2009). Estimation of Asymmetric Stochastic Volatility Models for Stock-Exchange Index Returns. *International Advances Economic Research*, **15**, 71-87.
- Gardner, M. and Dorling, S. (2000). Statistical surface ozone models: an improved methodology to account for non-linear behaviour. *Atmos Environ.*, **34**, 21-34.
- Gelfand, A. E. and Smith, A. F. M. (1990): Sampling-based approaches to calculating marginal densities. *J. Am. Stat. Assoc.*, **85**, 398-409.
- Ghysels, E., Harvey, A. C. and Renault, E. (1996): Stochastic volatility. (In: C. R. Rao and A. G. S. Madalla (Eds.), *Statistical models in finance* (pp.119-191). Amsterdam: Statistics, Textbook and Monograph, 123).
- Giraldo, R. (2009). *Geostatistical Analysis of Functional Data*. Dissertation, Universitat Politècnica de Catalunya.
- Gyarmati-Szabó, J., Chen, H. and Bogachev, L. V. (2008). Multivariate Extreme Value Modelling of Air Pollution Concentrations and Analysis of Traffic and Meteorological Impact. *Rev. Bras. Biom.*, **26** (4), 67-81.
- Hubbard, M. C. and Cobourn, W. G. (1998). Development of a Regression Model to Forecast Ground Level Ozone in Louisville, Kentucky. *Atmospheric Environment*, **32**, 2637-2647.
- Hurairah, A., Ibrahim, N. A., Daud, I. B. and Haron, K. (2005): An application of a new extreme value distribution to air pollution data. *Manage. Environ. Qual.: Int. J.*, **16** (1), 17-25.
- Kan, H. D. and Chen, B. H. (2004). Statistical distributions of ambient air pollutants in Shanghai, China. *Biomed. Environ. Sci.*, **17** (3), 366-372.
- Kim, S., Shepard, N. and Chib, S. (1998). Stochastic volatility: likelihood inference and comparison with ARCH models. *Rev. Econ. Stud.*, **65**, 361-393.
- Loomis, D. P., Borja-Arbutó, V. H., Bangdinwalla, S. I. and Shy, C. M. (1996). Ozone exposure and daily mortality in Mexico City: a time series analysis, (Boston: Health Effects Institute Research).
- Lu, H. C. and Fang, G. C. (2003). Predicting the exceedences of a critical PM10 concentration, a case study in Taiwan. *Atmos. Environ.*, **37**, 3491-3499.
- Meyer, R. and Yu, J. (2000). BUGS for a Bayesian analysis of stochastic volatility models. *Econ. J.*, **3**, 198-215.
- Myers, D.E. (1983): Estimation of linear combinations and cokriging. *Math Geol.*, **15** (5), 633-637.
- Ramsay, J. O. and Silverman, B. W. (2005). *Functional data analysis*, second edition. (New York: Springer-Verlag).
- Roberts, E. M. (1979a). Review of statistics extreme values with applications to air quality data. Part I. Review. *J. Air Pollut. Control Assoc.*, **29**, 632-637.
- Roberts, E. M. (1979b). Review of statistics extreme values with applications to air quality data. Part II. Applications. *J. Air Pollut. Control Assoc.*, **29**, 733-740.
- Robeson, S.M. and Steyn, D.G. (1990). Evaluation and comparison of statistical forecast models for daily maximum ozone concentrations. *Atmos. Environ., Part B*, **24** (2), 303-312.
- Rost, J., Holst, T., Sahn, E., Klingner, M., Anke, K., Ahrens, D. and Mayer, H. (2009). Variability of PM<sub>10</sub> concentrations dependent on meteorological conditions. *Int. Journal of Environment and Pollution*, **36**, 3-18.
- Sandmann G. and Koopman, S. J. (1998). Estimation of stochastic volatility models via Monte Carlo maximum likelihood. *Journal of Econometrics*, **2**, 271-301.
- Sfetsos, A., Zoras, S., Bartzis, J. G., and Triantafyllou, A. G. (2006). Extreme value modeling of daily PM10 concentrations in an industrial area, *Fresenius Environ. Bull.*, **15**, 841-845.
- Sharma, P., Khare, M. and Chakrabarti, S. P. (1999). Application of extreme value theory for predicting violations of air quality standards for an urban road intersection. *Transport. Res., Part D*, **4**, 201-216.
- Sharma, S., Barai, S. V. and Dikshit, A. K. (2003). Studies of air quality predictors based on neural networks. *Int. Journal of Environment and Pollution*, **19** (5), 442-453.
- Smith, R. L. (1989). Extreme values analysis of environmental time series: an application to trend detection in ground-level ozone. *Stat. Sci.*, **4**, 367-393.
- Smith, A. F. M. and Roberts, G. O. (1993). Bayesian Computation via the Gibbs Samples and related Markov Chain Monte Carlo Methods (with discussion). *J. R. Stat. Soc., Ser. B, Stat. Meth.*, **55**, 3-23.
- So, M. K. P., Li, W. K., and Lam, K. (2002). A threshold stochastic volatility model. *Journal of Forecasting*, **21**, 473-500.
- Surman, P. G., Bodero, J. and Simpson, R. W. (1987). The prediction of the numbers of violations of standards and the frequency of air pollution episodes using extreme value theory. *Atmos. Environ.*, **21**, 1843-1848.
- WHO, (2002), *Reducing Risks, Promoting Healthy Life*. The World Health Organization (WHO), Geneva, 2002.
- van Aalst, R. and de Leeuw, F. (1997). National ozone forecasting systems and international data exchange in north-west Europe, *EEA Tech. Report*, 9, 50.
- Wang, D. and Lu, W-Z (2006). Interval estimation of urban ozone level and selection of influential factors by employing automatic relevance determination model. *Chemosphere*, **62**, 1600-1611.
- Yu, J. (2002). Forecasting volatility in the New Zealand stock market. *Appl. Financ. Econ.*, **12**, 193-202.