International Journal of Pure and Applied Mathematics

Volume 53 No. 1 2009, 1-20

ESTIMATING THE NUMBER OF OZONE PEAKS IN MEXICO CITY USING A NON-HOMOGENEOUS POISSON MODEL AND A METROPOLIS-HASTINGS ALGORITHM

Jorge A. Achcar[1], Gisela Ortiz-Rodríguez[2 §], Eliane R. Rodrigues[3]

[1]Faculdade de Medicina de Ribeirão Preto
Universidade de São Paulo – USP
Av. Bandeirantes, 3900
14049-900 – Ribeirão Preto – SP, BRAZIL
e-mail: achcar@fmrp.usp.br

[2,3]Instituto de Matemáticas
Universidad Nacional Autónoma de México – UNAM
Area de la Investigación Científica
Circuito Exterior, Ciudad Universitaria
México, D.F., 04510, MÉXICO
[2]e-mail: giselaortiz37@hotmail.com
[3]e-mail: eliane@math.unam.mx

Abstract: In this paper we consider the problem of estimating the probability of having an air quality standard exceeded a certain number of times in a time interval of interest. A non-homogeneous Poisson model is used to study this problem. The rate function at which the Poisson events occur is given by $\lambda(t) > 0$, $t \geq 0$ which depends on some parameters to be estimated. These parameters are estimated using a Bayesian formulation based on a Metropolis-Hastings algorithm. A comparison of the performance of this algorithm with the performance of the *WinBugs* software is also given.

AMS Subject Classification: 60J20, 60G55, 62F15, 62M99, 92F05
Key Words: Metropolis-Hastings algorithm, Bayesian inference, non-homogeneous Poisson model

Received: March 7, 2009

[§]Correspondence author

1. Introduction

One problem common to large cities throughout the world is air pollution. Among the many pollutants affecting the inhabitants of those cities we have ozone. When ozone concentration stays above a certain threshold for a given period of time, a very sensitive part of the population (elderly, newborn) may experience serious health problems (see for example [9], [10], [11], [22], [27], [32], [42]). Therefore, being able to predict when such exceedances may occur is a very important issue.

The legal threshold used to declare emergency situations may vary from country to country. The US Environmental Protection Agency (US-EPA) has established that to avoid adverse health effects, the fourth highest daily maximum 8-hour average ozone concentration measured at each monitor within an area over each year must not exceed 0.075 parts per million (0.075ppm) (see [13]). The official environmental law for ozone in Mexico is that an individual should not be exposed to a concentration of 0.11ppm or above for a period of one hour or more (see [31]). In Mexico City the threshold used to declare an emergency situation is 0.22ppm.

Among the several methodologies that have been used in the study of ozone air pollution we have extreme values theory, time series analysis, multivariate analysis, neural networks and Markov chains among other. Among the many works using one or another of the methodologies above we may quote, [6], [7], [8], [14], [16], [17], [19], [20], [23], [24], [25], [26], [33], [36], [37], [38] and [43].

When the aim is to estimate the number of times that a given environmental standard is violated, [21], [34], [38] use Poisson processes to model this problem. However, in all cases the processes used are time homogeneous. It is a well known fact that time homogeneity is not a property of ozone measurements. Aiming to overcome the time homogeneity hypothesis, [3], [4], [5] use non-homogeneous Poisson processes. However, one shortcoming of those works is the use of the Gibbs sampling algorithm internally implemented in the software *WinBugs* (see [39]). Since the Gibbs sampling algorithm not always converge, the algorithm implemented in *WinBugs* may present convergence problems. In the present work we keep the assumption of a non-homogeneous Poisson model. However, the parameters of the intensity function of the Poisson process are estimated by using a Metropolis-Hasting type algorithm (see [18], [28]). The advantage of using a Metropolis-Hastings algorithm is that convergence is always guaranteed.

This paper is organized as follows. In Section 2 the basic assumptions of the Poisson model are presented. Section 3 describes the Bayesian formulation considered. In Section 4 a Metropolis-Hastings type algorithm is proposed to estimate the parameters of the intensity function of the Poisson model. An application to the case of ozone measurements in Mexico City is given in Section 5. In Section 6 some comments about the results are given. Finally, in the Appendix we give the *MATLAB* code of the Metropolis-Hastings algorithm presented in this work.

In here the notation $X \sim F$ is used to indicate that the random variable X has distribution function F.

2. A Non-Homogeneous Poisson Model

There are several works that uses non-homogeneous Poisson model to study problems varying from reliability theory to discovery of new marine species (see for example [1], [2], [35], [40], and [41] among others). In here the non-homogeneous Poisson model is used to estimate the probability that an air pollution standard is exceeded a given number of times in a time interval of interest.

The problem of interest here is described as follows. Let $M_t \geq 0$, $t \geq 0$ be the number of times an environmental standard of a given pollutant is violated in the time interval $[0, t)$. In order to describe the behaviour of M_t in the case of ozone, [21] propose the use of a time homogeneous Poisson model with rate $\lambda > 0$. The rate λ was calculated using the environmental standard that dictates that the ozone standard should not be violated on average more than once in three years. However, in the case of Mexico City this rule is not true (see for instance [4]). Therefore, we assume that the number of times that the ozone standard is violated follows a non-homogeneous Poisson process with some rate function that depends on time.

Consider a function $\lambda(t) > 0$, $t \geq 0$ and assume that at time t the random variable M_t has a Poisson distribution with rate function $\lambda(t)$ and mean function $m(t)$ given by

$$m(t) = \int_0^t \lambda(s)\, ds, \qquad t \geq 0.$$

Hence, we have, for $k = 0, 1, 2, \ldots$ and $t, s \geq 0$, that

$$P(M_{t+s} - M_t = k) = \frac{[m(t+s) - m(t)]^k}{k!} \exp\left(-[m(t+s) - m(t)]\right). \qquad (1)$$

Note that, when the behaviour of the function $\lambda(t)$ has been understood we have that the behaviour of $M = \{M_t : t \geq 0\}$ may be predicted and explained. Hence, the problem of studying M is reduced to studying $\lambda(t)$, $t \geq 0$.

There are several forms that $\lambda(t)$, $t \geq 0$, may assume. It is possible to observe from [4] that, in general, the number of ozone exceedances has been decreasing from January 1, 1998 until December 31, 2004. Therefore, it is interesting to have a rate function $\lambda(t)$, $t \geq 0$ that presents a decreasing behaviour as the time passes. One suggestion (see [29] and also [35]) is to take $\lambda(t)$, $t \geq 0$, of the exponentiated-Weibull form given by

$$\lambda(t) = \frac{\alpha\,\beta\,\left[1 - e^{-(t/\sigma)^{\alpha}}\right]^{\beta-1}\, e^{-(t/\sigma)^{\alpha}}\,(t/\sigma)^{\alpha-1}}{\sigma\left[1 - \left(1 - e^{-(t/\sigma)^{\alpha}}\right)^{\beta}\right]}, \qquad t \geq 0, \tag{2}$$

where α, β and σ are parameters of $\lambda(t)$, $t \geq 0$.

Other forms of $\lambda(t)$ may also be considered. They are given by the Weibull (when $\beta = 1$ in (2)), Musa-Okumoto (see [30]) and the generalised Goel-Okumoto (see [15]) models. In here we consider only the exponentiated-Weibull because we want to compare the results obtained by the Metropolis-Hastings algorithm presented here with the results obtained when using the *WinBugs* software given in [4]. Therefore, in the present case the vector of parameters to be estimated is $\theta = (\alpha, \beta, \sigma) \in \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}_+$.

Remarks. 1. Note that the function (2) will have a monotone decreasing behaviour if both α and $(\alpha\,\beta)$ are smaller or equal to one.

2. Whenever convenient, we assume that the parameters α, β and σ are independent random variables. The case where α and β are not independent will be made explicit in the formulation of the model.

The problem here is reduced to estimating the vector of parameters θ that produces the behaviour of $\lambda(t)$, $t \geq 0$, that is the most adequate to fit the situation presented by the ozone measurements of the Mexico City monitoring network. In order to do so, a Bayesian approach will be used.

3. A Bayesian Formulation of the Model

In this section a Bayesian formulation of the model is presented in which the likelihood function follows a Poisson model with rate function that is time dependent. The aim is to estimate the parameters of this rate function.

The formulation considered here may be described as follows. Let $K > 0$

be a fixed natural number. Assume that there are K days in which an ozone environmental standard has been violated. Let $d_1, d_2, \ldots, d_K$ indicate the days at which these violations occurred. Denote by $\mathbf{D} = \{d_1, d_2, \ldots, d_K\}$ the set of observed data. There is a natural relationship among posterior and prior distributions and the likelihood function of the model given by

$$P(\theta \,|\, \mathbf{D}) \propto L(\mathbf{D} \,|\, \theta)\, P(\theta)\,, \tag{3}$$

where $P(\theta \,|\, \mathbf{D})$ is the posterior distribution of θ given the data $\mathbf{D}$, $P(\theta)$ is the prior distributions of θ and $L(\mathbf{D} \,|\, \theta)$ is the likelihood function. The components of (3) are given as follows.

1. (The Likelihood Function) Since a non-homogeneous Poisson model has been assumed for the problem, we have that the likelihood function will take the form

$$L(\mathbf{D} \,|\, \theta) = \left[\prod_{i=1}^{K} \lambda(d_i)\right] \exp\left(-m(d_K)\right).$$

The rate function $\lambda(t)$ given by (2) has mean function given by $m(t) = -\log\left[1 - F(t)\right]$ where $F(t) = (1 - \exp\left[-\left(t/\sigma\right)^{\alpha}\right])^{\beta}$. Therefore,

$$L(\mathbf{D} \,|\, \theta) \propto \frac{(\alpha\, \beta)^K \left[\prod_{i=1}^{K} d_i^{\alpha-1}\, e^{-(d_i/\sigma)^{\alpha}} \left(1 - e^{-(d_i/\sigma)^{\alpha}}\right)^{\beta-1}\right]}{\sigma^{\alpha K} \left[\prod_{i=1}^{K-1} \left(1 - \left[1 - e^{-(d_i/\sigma)^{\alpha}}\right]^{\beta}\right)\right]} \tag{4}$$

(see for example, [12], [25] and [35]).

2. (The Prior Distribution) The following prior distributions of the parameters are considered.

Due to the restrictions imposed on the behaviour of $\lambda(t)$, we need to have the random quantities α and $(\alpha\, \beta)$ in the interval $[0, 1]$ and $\sigma \in \mathbb{R}_+$. Hence, we take the following: $\alpha \sim \mathrm{Beta}(a_1, b_1)$; $\sigma \sim \mathrm{Gamma}(a_2, b_2)$ (in here, $\mathrm{Beta}(a, b)$ and $\mathrm{Gamma}(c, d)$ are the Beta and Gamma distributions with means $a/(a+b)$ and c/d, respectively, and variances $ab/[(a+b)^2(a+b+1)]$ and c/d^2, respectively). Assuming a dependence between α and β (since we need $\alpha\beta < 1$ given that $\alpha < 1$) we take $\beta \sim \mathrm{Uniform}(0, 1/\alpha)$. Then, the joint posterior density in this case takes the form:

$$\begin{aligned} P(\theta \,|\, \mathbf{D})) &= P(\alpha, \beta, \sigma \,|\, \mathbf{D}) \\ &\propto \alpha^{a_1+K}\, \beta^K\, \sigma^{a_2 - K\alpha - 1} e^{-\sigma\, b_2}\, (1-\alpha)^{b_1-1} \\ &\quad \frac{\prod_{i=1}^{K} \left[d_i^{\alpha-1}\, e^{-(d_i/\sigma)^{\alpha}} \left(1 - e^{-(d_i/\sigma)^{\alpha}}\right)^{\beta-1}\right]}{\prod_{i=1}^{K-1} \left(1 - \left[1 - e^{-(d_i/\sigma)^{\alpha}}\right]^{\beta}\right)}. \end{aligned} \tag{5}$$

The parameters $a_i,\ b_i,\ i = 1,2$ are know hyperparameters that will be specified later. The sampling of the values of θ will be performed using the Markov chain Monte Carlo algorithm given in the next section.

4. A Metropolis-Hastings Sampling Algorithm

The aim in this section is to present a Metropolis-Hastings type algorithm (see [18] and [28]) to obtain a sample from $P(\theta \,|\, \mathbf{D})$ given by (5). After a sample is obtained the law of large numbers may be used to estimate the behaviour of the posterior distribution.

In order to achieve the aim here, we are going to construct an ergodic Markov chain $X = \{X_n : n = 0, 1, \ldots\}$ whose stationary distribution is (5). Therefore, if at time n the actual state of the chain is $X_n = \theta$, a candidate θ' to be the next state of the chain is chosen as follows. Generate θ using the corresponding prior distributions (other distributions may also be used). The change from θ to θ' will occur with probability

$$q(\theta,\theta') = \min\left\{1, \frac{P(\theta' \,|\, \mathbf{D})}{P(\theta \,|\, \mathbf{D})}\,\frac{p(\theta',\theta)}{p(\theta,\theta')}\right\},$$

where $p(\theta,\theta')$ is the proposal distribution used to generate the values θ' (in the present case is the prior distribution of θ). Then, for $P(\theta \,|\, \mathbf{D})$ given by (5) we have that

$$\begin{aligned} q(\theta,\theta') \;=\; \min\Bigg\{ 1, & \left(\prod_{i=1}^{K} \frac{d_i^{\alpha'-1}\, e^{-(d_i/\sigma')^{\alpha'}} \left(1 - e^{-(d_i/\sigma')^{\alpha'}}\right)^{\beta'-1}}{d_i^{\alpha-1}\, e^{-(d_i/\sigma)^{\alpha}} \left(1 - e^{-(d_i/\sigma)^{\alpha}}\right)^{\beta-1}} \right) \\ & \left(\prod_{i=1}^{K} \frac{1 - \left[1 - e^{-(d_i/\sigma)^{\alpha}}\right]^{\beta}}{1 - \left[1 - e^{-(d_i/\sigma')^{\alpha'}}\right]^{\beta'}} \right) \frac{r(\theta',\theta)}{r(\theta,\theta')} \Bigg\}, \end{aligned}$$

where

$$\frac{r(\theta',\theta)}{r(\theta,\theta')} = \left(\frac{\alpha'\beta'\sigma'^{-\alpha'}}{\alpha\beta\sigma^{-\alpha}}\right)^{K}.$$

The aim now is to obtain a sample from the posterior distribution using the information provided by the data of the monitoring network of the Metropolitan Area of Mexico City. Once the sample is obtained inference about the behaviour of the rate function $\lambda(t)$, $t \geq 0$, can be performed.

5. An Application to the Ozone Measurements in Mexico City

In this section we apply the results described earlier in this paper to the case of ozone measurements of the Metropolitan Area of Mexico City monitoring network. The Metropolitan Area is divided into five regions or sections corresponding to the Northeast (NE), Northwest (NW), Center (CE), Southeast (SE) and Southwest (SW) and the ozone monitoring stations are placed throughout the city (see [4] and [6]). The data used in the analysis (obtained from www.sma.df.gob.mx/simat/) corresponds to seven years (from January 1, 1998 to December 31, 2004) of the daily maximum measurements of each region. The measurements are obtained minute by minute and the averaged hourly result is reported at each station. The daily maximum measurement for a given region is the maximum over all the maximum averaged values recorded hourly during a 24-hour period by each station placed in the region.

During the period considered here, the Mexican ozone standard of 0.11ppm was violated on 2063 days. Nevertheless, the daily peaks were double the Mexican standard on 237 days. In here we are going to use the threshold 0.17ppm. One of the reasons for choosing this threshold is that it is an intermediate value between 0.11ppm and 0.22ppm. Additionally, we would like to study the behaviour of M_t, $t \geq 0$, if 0.17ppm instead of 0.22ppm is used to declare emergency situations in Mexico City. Note that from January 1, 1998 until December 31, 2004, the threshold 0.17ppm was surpassed 980 days. Analysis were performed for each region separately and only considering the threshold 0.17ppm. Other values could also be considered.

The hyperparameters of the prior distributions were taken as follows. The parameters a_1 and b_1 were obtained by imposing the condition that the mean μ_{Beta} and the variance σ^2_{Beta} of the Beta distribution, were the estimated mean and variance of the parameter α given by [4]. Similar procedure was followed to obtain a_2 and b_2 of the Gamma prior distribution of the parameter σ. Table 1 gives the hyperparameters for the prior distributions of the parameters α and σ. Once the value of α was simulated, it was used to obtain the parameter β.

In order to initialise the algorithm and to perform analysis of its convergence and of the correlation among the generated values, three different values were used for each parameter. They were, in general, taken as the mean and the points near the extremes of the 95% credible interval obtained in previous studies. The values used to generate the sample considered to estimate the parameters in the present case were $\alpha_0 = 0.5$, $\beta_0 = 1$ and $\sigma_0 = 0.2$.

Even though convergence of the algorithm varied according to region and

	NE	NW	CE	SE	SW
a_1	0.4748	0.7266	0.9523	0.7619	1.3502
b_1	0.3582	0.3912	0.4081	0.3925	0.4264
a_2	1.6649	4.3719	5.5225	4.5918	10.4379
b_2	1.3874	19.0083	11.75	10.2041	24.8521

Table 1: Values of the hyperparameters for the prior distributions of the parameters α and σ for regions NE, NW, CE, SE and SW

parameter to be estimated, we have decided to use the same value of burn-in period as well as the same lag between two consecutive values taken to be part of the sample used to estimate the parameters of the model. We also use the same sample size. Therefore, estimates were performed using a sample of size 1000 selected after a burn-in period of 100000 steps and taking every 300-th generated value.

Table 2 presents the values of the estimated parameters. These values are similar to the ones obtained when using the *WinBugs* software (see [4]).

		Mean	SD	95% Credible Interval
NE	α	0.92912	0.082359	(0.92402, 0.93423)
	β	0.71094	0.27162	(0.6941, 0.72777)
	σ	0.98634	0.33729	(0.96543, 1.0072)
NW	α	0.55648	0.063033	(0.55257, 0.56038)
	β	1.2332	0.45217	(1.2051, 1.2612)
	σ	0.35806	0.14572	(0.34902, 0.36709)
CE	α	0.67963	0.062324	(0.67576, 0.68349)
	β	1.0221	0.34843	(1.0005, 1.0437)
	σ	0.69765	0.24648	(0.68237, 0.71292)
SE	α	0.68236	0.069786	(0.67803, 0.68668)
	β	1.0144	0.35447	(0.99246, 1.0364)
	σ	0.70566	0.26826	(0.68903, 0.72229)
SW	α	0.55208	0.061781	(0.54825, 0.55591)
	β	1.2403	0.47493	(1.2108, 1.2697)
	σ	0.52209	0.14876	(0.51287, 0.53131)

Table 2: Posterior mean, standard deviation (indicated by SD) and 95% credible interval of the parameters α, β and σ for regions NE, NW, CE, SE and SW

In order to illustrate how the results obtained here may be used to perform predictions, consider the case of region SW. We have decided to ilustrate using this region because it is the region with more problems related to ozone pollution. Hence, consider the period of time January 1-25, 2005 and we want to know what is the probability of having 5, 10, 15 and 20 days or less where the threshold 0.17ppm is surpassed. In this case we have that these probabilities are respectively, 0.16, 0.81, 1.0 and 1.0. During the first 25 days of 2005 there were two days in which the threshold 0.17ppm was surpassed. If we consider the probability of having five days or less with measurements above 0.17ppm, we have that the Metropolis-Hasting algorithm presented here underestimate the probability of this event when compared to the result given by the *WinBugs* software (which was 0.52).

6. Discussion

If we compare the estimated values for α, β and σ with the ones estimated using *WinBugs* (see [4]) we may see that they are very similar. Analysing the behaviour of $\lambda(t)$, $t \geq 0$, as function of α, β and σ we have that $\lambda(t)$ is a decreasing function as a function of β and σ and is an increasing function as a function of α. Comparing the parameters estimated by the *WinBugs* software and the Metropolis-Hastings algorithm presented here, we have that the values of $m(t)$ for each $t \geq 0$ are larger when using the *WinBugs* software and when considering regions NW, CE, SE and SW. When taking into account region NE we have that the value of $m(t)$ is larger when using the Metropolis-Hastings algorithm.

Comparing the fit of the observed mean and the estimated mean obtained by the *WinBugs* software (see [4]) and taking into account the comparison made above between $m(t)$ obtained by the *WinBugs* software and the Metropolis-Hastings algorithm, we have that graphically, the estimates given by the *WinBugs* produce better fit.

Looking at the values of the parameters estimated using the sample generated by the Metropolis-Hastings algorithm presented here, we have that the values do not differ much from the ones obtained by using the *WinBugs* software.

As pointed out before, the fit of the curve of the estimated mean $m(t)$ using the Metropolis-Hastings algorithm is not as good as the fit obtained when using the *WinBugs* software. Nevertheless, the Metropolis-Hastings algorithm has the advantage of always converging to the right distribution. Since *WinBugs* partly relies upon the Gibbs sampling algorithm, convergence may not always occur unless a very informative prior distribution is used. Convergence of the Gibbs sampling depends on the construction of an ergodic Markov chain. The Metropolis-Hastings always produces an ergodic chain. Additionally, the software *WinBugs* is presented as a black box, hence one is not sure how the internal implementation of the algorithm is written. In the present case we have an explicit formulation for the algorithm. Additionally, we present the *MATLAB* code of the algorithm. Alternatively, one may write one's own programme and have one's own code.

Acknowledgements

The authors wish to thank Guadalupe Tzintzun from the National Institute of Ecology of the Ministry of Environment of Mexico for providing the data from the Mexico City monitoring network and also for providing information about EPA-USA and Mexico environmental standards. E.R.R thanks the Department of Statistics at the University of Oxford, where part of this work was developed, for all the support received during her stay at the Department. G.O.R. thanks CONACyT-México for MSc scholarship. J.A.A. was partially funded by CNPq-Brazil grant number 300235/2005-4. E.R.R was partially funded by DGAPA-UNAM, Mexico, grant number 968SFA/2007.

References

[1] J.A. Achcar, Bayesian analysis for software reliability data, *Advances in Reliability, Handbook of Statistics*, **20** (2001) 733-748.

[2] J.A. Achcar, D.K. Dey, M. Niverthy, A Bayesian approach using nonhomogeneous Poisson process for software reliability models, In: *Frontiers in Reliability* (Ed-s: S.K. Basu, S. Mukhopadhyay), Series on Quality, Reliability and Engeneering Statistics, **4**, Calcutta University, India (1998).

[3] J.A. Achcar, E.R. Rodrigues, G. Tzintzun, Using non-homogeneous Poisson models with multiple change-points to estimate the number of ozone exceedances in Mexico City, *Environmetrics* (2008), Submitted.

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

[5] J.A. Achcar, E.R. Rodrigues, C.D. Paulino, P. Soares, Non-homogeneous Poisson processes with a change-point: An application to ozone exceedances in Mexico City, *Ecological and Environmental Statistics* (2008), Submitted.

[6] L.J. Álvarez, A.A. Fernández-Bremauntz, E.R. Rodrigues, G. Tzintzun, Maximum a posteriori estimation of the daily ozone peaks in Mexico City, *Journal of Agricultural, Biological, and Environmental Statistics*, **10** (2005), 276-290.

[7] L.J. Álvarez, E.R. Rodrigues, A trans-dimensional MCMC algorithm to estimate the order of a Markov chain: an application to ozone peaks in Mexico City, *International Journal of Pure and Applied Mathematics*, **48** (2008), 315-331.

[8] J. Austin, H. Tran, A characterization of the weekday-weekend behavior of ambient ozone concentrations in California, In: *Air Pollution*, **VII**, 645-661, WIT Press, UK (1999).

[9] M.L. Bell, A. McDermott, S.L. Zeger, J.M. Samet, F. Dominici, Ozone and short-term mortality in 95 US urban communities, 1987-2000, *Journal of the American Medical Society*, **292** (2004), 2372-2378.

[10] M.L. Bell, R. Peng, F. Dominici, The exposure-response curve for ozone and risk of mortality and the adequacy of current ozone regulations, *Environmental Health Perspectives*, **114** (2005), 532-536.

[11] M.L. Bell, R. Goldberg, C. Rogrefe, P.L. Kinney, K. Knowlton, B. Lynn, J. Rosenthal, C. Rosenzweig, J.A. Patz, Climate change, ambient ozone, and health in 50 US cities, *Climate Change*, **82** (2007), 61-76.

[12] D.R. Cox, P.A. Lewis, *Statistical Analysis of Series Events*, Methuen, UK (1996).

[13] EPA (US Environmental Protection Agency), *National Ambient Air Quality Standards*, http://www.epa.gov/air/criteria.html (2008).

[14] J.B. Flaum, S.T. Rao, I.G. Zurbenko, Moderating influence of meteorological conditions on ambient ozone concentrations, *J. Air and Waste Management Assoc.*, **46** (1996), 33-46.

[15] A.L. Goel, K. Okumoto, An analysis of recurrent software failures on a real-time control system, In: *Proceedings of ACM Conference*, Washington-DC, USA (1978), 496-500.

[16] R. Guardani, C.A.O. Nascimento, M.L.G. Guardani, M.H.R.B. Martins, J. Romano, Study of atmospheric ozone formation by means of a neural network based model, *J. Air and Waste Management Assoc.*, **49** (1999), 316-323.

[17] R. Guardani, J.L. Aguiar, C.A.O. Nascimento, C.I.V. Lacava, Y. Yanagi, Ground-level ozone mapping in large urban areas using multivariate analysis: application to the São Paulo Metropolitan Area, *J. Air and Waste Management Assoc.*, **53** (2003), 553-559.

[18] W.K. Hastings, Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, **57** (1970), 97-109.

[19] J. Horowitz, Extreme values from a nonstationary stochastic process: an application to air quality analysis, *Technometrics*, **22** (1980), 469-482.

[20] G. Huerta, B. Sansó, Time-varying models for extreme values, *Technical Report* **2005-4**, Department of Applied Mathematics and Statistics, University of California, USA (2005).

[21] J.S. Javits, Statistical interdependencies in the ozone national ambient air quality standard, *J. Air Poll. Control Assoc.*, **30** (1980), 58-59.

[22] K. Itô, S. de León, M. Lippman, Associations between ozone and daily mortality: a review and additional analysis, *Epidemiology*, **16** (2005), 446-457.

[23] M. Lanfredi, M. Macchiato, Searching for low dimensionality in air pollution time series, *Europhysics Lett.*, **40** (1997), 589-594.

[24] L.C. Larsen, R.A. Bradley, G.L. Honcoop, A new method of characterizing the variability of air quality-related indicators, In: *Air and Waste Management Association's International Specialty Conference of Tropospheric Ozone and the Environment*, California, USA (1990).

[25] J.F. Lawless, *Statistical Models and Methods for Lifetime Data*, John Wiley and Sons, USA (1982).

[26] M.R. Leadbetter, On a basis for "peak over threshold" modeling, *Statistics and Probability Letters*, **12** (1991), 357-362.

[27] D.P. Loomis, V.H. Borja-Arbuto, S.I. Bangdiwala, C.M. Shy, Ozone exposure and daily mortality in Mexico City: a time series analysis, *Health Effects Institute Research Report*, **75** (1996), 1-46.

[28] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, E. Teller, Equations of state calculations by fast computing machine, *J. Chem. Phys.*, **21** (1953), 1087-1091.

[29] G.S. Muldholkar, D.K. Srivastava, H. Friemer, The exponentiated-Weibull family: a reanalysis of the bus-motor failure data, *Technometrics*, **37** (1995), 436-445.

[30] J.D. Musa, K. Okumoto, A logarithmic Poisson execution time model for software reliability measurement, In: *Proceedings of Seventh International Conference on Software Engineering*, Orlando, USA (1984), 230-238.

[31] NOM, Modificación a la Norma Oficial Mexicana NOM-020-SSA1-1993, *Diario Oficial de la Federación*, 30 de Octubre de 2002 (2002).

[32] M.R. O'Neill, D. Loomis, V.H. Borja-Aburto, Ozone, area social conditions and mortality in Mexico City, *Environmental Research*, **94** (2004), 234-242.

[33] J.-N. Pan, S.-T. Chen, Monitoring long-memory air quality data using ARFIMA model, *Environmetrics*, **19** (2007), 209-219.

[34] A.E. Raftery, Are ozone exceedance rate decreasing?, Comment of the paper "Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone" by R.L. Smith, *Statistical Sciences*, **4** (1989), 378-381.

[35] J.E. Ramírez-Cid, J.A. Achcar, Bayesian inference for nonhomogeneous Poisson processes in software reliability models assuming nonmonotonic intensity functions, *Computational Statistics and Data Analysis*, **32** (1999), 147-159.

[36] E.M. Roberts, Review of statistics extreme values with applications to air quality data. Part I. Review, *Journal of the Air Pollution Control Association*, **29** (1979), 632-637.

[37] E.M. Roberts, Review of statistics extreme values with applications to air quality data. Part II. Applications, *Journal of the Air Pollution Control Association*, **29** (1979), 733-740.

[38] R.L. Smith, Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone, *Statistical Sciences*, **4** (1989), 367-393.

[39] D.J. Spiegelhalter, A. Thomas, N.G. Best, W.R. Gilks, *WinBugs: Bayesian Inference Using Gibbs Sampling*, MRC Biostatistics Unit, Cambridge, UK (1999).

[40] K. Storani, J.A. Achcar, Métodos Bayesianos em modelos de confiabilidade de software usando procesos de Poisson não homogêneos, *Revista Brasileira de Estatística*, **59** (1998), 59-80.

[41] S.P. Wilson, M.J. Costello, Predicting future discoveries of European marine species using non-homogeneous renewal processes, *Journal of the Royal Statistical Society Series C*, **54** (2005), 425-442.

[42] R. Wilson, S.D. Colone, J.D. Spengler, D.G. Wilson, *Health Effects if Fossil Fuel Burning: Assessment and Mitigation*, Ballenger, USA (1980).

[43] A. Zolghadri, D. Henry, Minmax statistical models for air pollution time series. Application to ozone time series data measured in Bordeaux, *Environmental Monitoring and Assessment*, **98** (2004), 275-294.

Appendix

In this section were present the computer code elaborated in *MATLAB* of the Metropolis-Hastings algorithm presented here and that was used to estimate the values of the parameters of the exponentiated-Weibull rate function.

```
% Metropolis-Hastings.
% Con  a~Beta(a1,b1), s~Gamma(a2,b2), b~U(0,1/a).
clc;
clear;
t1 = clock;
%---------- Reading the data set --------------------
fecha = '9804';
region = 'se';
d = leer_datos(fecha,region);
nivel = 2;
switch lower(num2str(nivel))
    case '1'
        nivelozono = 0.11;
    case '2'
        nivelozono = 0.17;
    case '3'
        nivelozono = 0.22;
end
```

```
m = length(d);
%---- Number of Markov chains considered and sample size --
ncad = 1;
N = 400000;
percal = 100000;
%---------- Parameters for the prior distributions---------
switch region
    case 'ne'
        mediart_a = 0.57;
        mediart_s = 1.2;
        var_a = (0.05)^2;
        var_s = (0.93)^2;
    case 'nw'
        mediart_a = 0.65;
        mediart_s = 0.23;
        var_a = (0.03)^2;
        var_s = (0.11)^2;
    case 'ce'
        mediart_a = 0.7;
        mediart_s = 0.47;
        var_a = (0.03)^2;
        var_s = (0.2)^2;
    case 'sw'
        mediart_a = 0.76;
        mediart_s = 0.42;
        var_a = (0.03)^2;
        var_s = (0.13)^2;
    case 'se'
        mediart_a = 0.66;
        mediart_s = 0.45;
        var_a = (0.04)^2;
        var_s = (0.21)^2;
end
[a1,b1] = param_beta(mediart_a,sqrt(var_a))
[a2,b2] = param_gama(mediart_s,sqrt(var_s))
%------- Seeds ------------------
for contador = 1:ncad;  % (***)
    a_0 = 0.50;
    s_0 = 0.20;
    b_0 = 1;
    I(a_0,b_0,s_0,d,nivel)   % This should be different of zero
    II(a_0,b_0,s_0,d,nivel)  % This should be different of zero
    %------- State of the Markov chain for the first step-----------
    a_ = betarnd(a1,b1);
            while (a_ == 0)
                a_ = betarnd(a1,b1);
            end
    s_ = gamrnd(a2,1/b2);
```

```
b_ = unifrnd(0,1/a_);
while (I(a_,b_,s_,d,nivel) == 0) | (II(a_,b_,s_,d,nivel) == 0)
    a_ = betarnd(a1,b1);
        while (a_ == 0)
            a_ = betarnd(a1,b1);
        end
    s_ = gamrnd(a2,1/b2);
    b_ = unifrnd(0,1/a_);
end
Q = [((a_*b_)/(a_0*b_0))^m] * [(s_^a_)/(s_0^a_0)]^(-m);
III = [ I(a_,b_,s_,d,nivel) / I(a_0,b_0,s_0,d,nivel) ] *
      [ II(a_0,b_0,s_0,d,nivel) / II(a_,b_,s_,d,nivel) ];
q = min(1,Q*III);
u = unifrnd(0,1);
if u <= q
    a(1,1) = a_;
    s(1,1) = s_;
    b(1,1) = b_;
else
    a(1,1) = a_0;
    s(1,1) = s_0;
    b(1,1) = b_0;
end
%------- State of the Markov chain for steps greater or equal to 2
for n = 2:N
    a_ = betarnd(a1,b1);
        while (a_ == 0)
            a_ = betarnd(a1,b1);
        end
    s_ = gamrnd(a2,1/b2);
    b_ = unifrnd(0,1/a_);
    while (I(a_,b_,s_,d,nivel) == 0) | (II(a_,b_,s_,d,nivel) == 0)
        a_ = betarnd(a1,b1);
            while (a_ == 0)
                a_ = betarnd(a1,b1);
            end
        s_ = gamrnd(a2,1/b2);
        b_ = unifrnd(0,1/a_);
    end
    Q = [((a_*b_)/(a(1,n-1)*b(1,n-1)))^m] *
         [(s_^a_)/(s(1,n-1)^a(1,n-1))]^(-m);
    III = [ I(a_,b_,s_,d,nivel) / I(a(1,n-1),b(1,n-1),
           s(1,n-1),d,nivel) ] * [ II(a(1,n-1),b(1,n-1),s(1,n-1),
                                d,nivel) / II(a_,b_,s_,d,nivel) ];
    q = min(1,Q*III);
    u = unifrnd(0,1);
    if u <= q
        a(1,n) = a_;
```

```
        s(1,n) = s_;
        b(1,n) = b_;
    else
        a(1,n) = a(1,n-1);
        s(1,n) = s(1,n-1);
        b(1,n) = b(1,n-1);
    end
end
dlmwrite(['C:results-alpha.dat'], a)
dlmwrite(['C:results-beta.dat'], b)
dlmwrite(['C:results-sigma.dat]', s)
%--------- Burn-in period (ergodic mean) ---------
S_a = prom_erg(a);
S_s = prom_erg(s);
S_b = prom_erg(b);
dlmwrite(['C:ergodic-alpha.dat'], S_a)
dlmwrite(['C:ergodic-sigma.dat'], S_s)
dlmwrite(['C:ergodic-beta.dat'], S_b)
hold off
fig = figure(1+9*(contador-1));
plot(1:N,S_a,'.')
xlabel(['Iteration'])
ylabel(['Ergodic mean'])
saveas(fig,['C:ergodic-alpha.eps'])
hold off
fig = figure(2+9*(contador-1));
plot(1:N,S_s,'.')
xlabel(['Iteration'])
ylabel(['Ergodic mean'])
saveas(fig,['C:ergodic-sigma.eps'])
hold off
fig = figure(3+9*(contador-1));
plot(1:N,S_b,'.')
xlabel(['Iteration'])
ylabel(['Ergodic mean'])
saveas(fig,['C:ergodic-beta.eps'])
%--  Eliminate the values generated during the burn-in period --
for t = 1:N-percal
    a_obs(1,t) = a(1,percal+t);
    s_obs(1,t) = s(1,percal+t);
    b_obs(1,t) = b(1,percal+t);
end
dlmwrite(['C:generated-alpha.dat'], a_obs)
dlmwrite(['C:generated-sigma.dat'], s_obs)
dlmwrite(['C:generated-beta.dat'], b_obs)
%--------- Correlation values -----
kdep = 600;
media_a_obs = mean(a_obs);
```

```
media_s_obs = mean(s_obs);
media_b_obs = mean(b_obs);
dim_a_obs = length(a_obs);
dim_s_obs = length(s_obs);
dim_b_obs = length(b_obs);
for j = 1:dim_a_obs
    pk2_a(1,j) = (a_obs(1,j)-media_a_obs)^2;
end
for j = 1:dim_s_obs
    pk2_s(1,j) = (s_obs(1,j)-media_s_obs)^2;
end
for j = 1:dim_b_obs
    pk2_b(1,j) = (b_obs(1,j)-media_b_obs)^2;
end
for k = 1:kdep
    for j = 1:dim_a_obs-k
        pk1(1,j) = (a_obs(1,j)-media_a_obs) *
                       (a_obs(1,j+k)-media_a_obs);
    end
    pk_a_obs(1,k) = sum(pk1)/sum(pk2_a);
    for j = 1:dim_s_obs-k
        pk1(1,j) = (s_obs(1,j)-media_s_obs) *
                       (s_obs(1,j+k)-media_s_obs);
    end
    pk_s_obs(1,k) = sum(pk1)/sum(pk2_s);
    for j = 1:dim_b_obs-k
        pk1(1,j) = (b_obs(1,j)-media_b_obs) *
                       (b_obs(1,j+k)-media_b_obs);
    end
    pk_b_obs(1,k) = sum(pk1)/sum(pk2_b);
end
dlmwrite(['C:corr-alpha.dat'], pk_a_obs)
dlmwrite(['C:corr-sigma.dat'], pk_s_obs)
dlmwrite(['C:corr-beta.dat'], pk_b_obs)
hold off
fig = figure(4+9*(contador-1));
plot(0:kdep,zeros(1,kdep+1))
hold on
plot(1:kdep,pk_a_obs,'.')
ylim([-1 1])
xlabel(['Lags'])
ylabel(['Auto-correlation'])
saveas(fig,['C:corr-alpha.eps'])
hold off
fig = figure(5+9*(contador-1));
plot(0:kdep,zeros(1,kdep+1))
hold on
plot(1:kdep,pk_s_obs,'.')
```

```
    ylim([-1 1])
    xlabel(['Lags'])
    ylabel(['Auto-correlation'])
    saveas(fig,['C:corr-sigma.eps'])
    hold off
    fig = figure(6+9*(contador-1));
    plot(0:kdep,zeros(1,kdep+1))
    hold on
    plot(1:kdep,pk_b_obs,'.')
    ylim([-1 1])
    xlabel(['Lags'])
    ylabel(['Auto-correlation'])
    saveas(fig,['C:corr-beta.eps'])
end  % end for (***)
%%%% Estimation of the density of the parameters
kind = 300;
tmuestra = floor((N-percal-1)/kind) + 1;
for t = 1:tmuestra
    a_ind(1,t) = a_obs(1, 1 + (t-1)*kind);
    s_ind(1,t) = s_obs(1, 1 + (t-1)*kind);
    b_ind(1,t) = b_obs(1, 1 + (t-1)*kind);
end
dlmwrite(['C:sample-alpha.dat'], a_ind)
dlmwrite(['C:sample-sigma.dat'], s_ind)
dlmwrite(['C:sample-beta.dat'], b_ind)
hold off
fig = figure(7+9*(contador-1));
hist(a_ind,20)
xlabel(['Alpha'])
ylabel(['Frequency'])
saveas(fig,['C:freq-alpha.eps'])
hold off
fig = figure(8+9*(contador-1));
hist(s_ind,20)
xlabel(['Sigma'])
ylabel(['Frequency'])
saveas(fig,['C:freq-sigma.eps'])
hold off
fig = figure(9+9*(contador-1));
hist(b_ind,20)
xlabel(['Beta'])
ylabel(['Frequency'])
saveas(fig,['C:freq-beta.eps'])
% -------- Mean, standard deviation and 95% credible interval
m_a = mean(a_ind);
m_s = mean(s_ind);
m_b = mean(b_ind);
dst_a = std(a_ind);
```

```
dst_s = std(s_ind);
dst_b = std(b_ind);
z = 1.96;               % cuantil 1-.05/2 de N(0,1)
linf_a = m_a-z*[dst_a/sqrt(length(a_ind))];
lsup_a = m_a+z*[dst_a/sqrt(length(a_ind))];
linf_s = m_s-z*[dst_s/sqrt(length(s_ind))];
lsup_s = m_s+z*[dst_s/sqrt(length(s_ind))];
linf_b = m_b-z*[dst_b/sqrt(length(b_ind))];
lsup_b = m_b+z*[dst_b/sqrt(length(b_ind))];
% -------- Saves results is results.txt -------------------
estimaciones = [m_a dst_a linf_a lsup_a alfa_a beta_a; m_s dst_s
                linf_s lsup_s alfa_s beta_s; m_b dst_b linf_b lsup_b
                alfa_b beta_b];
dlmwrite(['C:results.txt'], 'newline', 'pc')
% -------- Save variables in variables.txt -------------------
variables = [mediart_a; mediart_s; var_a; var_s; a1; b1; a2; b2; a_0;
             s_0; b_0; N; percal; kind; tmuestra];
dlmwrite(['C:variables.txt,'newline', 'pc')
t2 = clock;
tiempo_min = etime(t2,t1)/60
tiempo_hrs = etime(t2,t1)/3600
dlmwrite(['C:time_inf.txt'], tiempo_hrs)
fprintf('\n ****** F I N ****** \n')
beep;
```