

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

185,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



# A Gibbs Sampling Algorithm to Estimate the Occurrence of Ozone Exceedances in Mexico City

Eliane R. Rodrigues<sup>1</sup>, Jorge A. Achcar<sup>2</sup> and Julián Jara-Ettinger<sup>3</sup>

<sup>1</sup>*Instituto de Matemáticas, Universidad Nacional Autónoma de México*

<sup>2</sup>*Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo*

<sup>3</sup>*Facultad de Ciencias Físico-Matemáticas, Universidad Michoacana*

<sup>1,3</sup>*Mexico*

<sup>2</sup>*Brazil*

## 1. Introduction

Inhabitants of many cities around the world suffer the effects of high levels of ozone air pollution. When the ozone concentration stays above the threshold of 0.11 parts per million (0.11ppm) for a period of one hour or more, a very sensitive part of the population (e.g. elderly and newborn) in that environment may experience serious health deterioration (Bell et al., 2004, 2005, 2007; Gauderman et al., 2004; Loomis et al., 1996; O'Neill et al., 2004; WHO, 2006). Therefore, being able to predict when exceedances of the 0.11ppm threshold (or any threshold above this value) may occur is a very important issue.

Depending on the country, different thresholds may be considered to declare environmental alerts. The US Environmental Protection Agency (US-EPA) has established as its standard that the three-year average of the fourth highest daily maximum 8-hour average ozone concentration measured at each monitor within an area over each year must not exceed 0.075ppm (see EPA, 2008). In the case of Mexico, the standard is 0.11ppm and an individual should not be exposed, on average, for a period of an hour or more (NOM, 2002). In Mexico City when the ozone concentration surpasses the threshold 0.2ppm an emergency alert is issued and measures are taken to prevent population exposure to the pollutant (see <http://www.sma.df.gob.mx>).

It is possible to find in the literature a large number of works focussing on the study of the behaviour of air pollutants in general. Among the ones related to ozone air pollution we may quote, Horowitz (1980), Roberts (1979a, 1979b), Smith (1989) considering extreme value theory; Flaum et al. (1996), Gouveia & Fletcher (2000), Kumar et al. (2010), Lanfredi & Macchiato (1997) and Pan & Chen (2008) using time series analysis; Álvarez et al. (2005) and Austin & Tran (1999) considering Markov chain models; Achcar et al. (2008b, 2009a, 2009b) using stochastic volatility models; and Huerta & Sansó (2005) with an analysis of the behaviour the maximum measurements of ozone with an application to the data from Mexico City. Davis (2008) and Zavala et al. (2009) present studies that analyse the impact on air quality of the driving restriction imposed in Mexico City.

When the aim is to estimate the number of times that a given environmental standard is surpassed, Raftery (1989) and Smith (1989) use time homogeneous Poisson processes to model

this event. Aiming to overcome the time homogeneity hypothesis, Achcar et al. (2008a, 2010, 2011) consider non-homogeneous Poisson processes. However, one shortcoming of those works is that they estimate the parameters involved in the model using the Gibbs sampling algorithm internally implemented in the software WinBugs. One particular problem that occurred when using the data set from the monitoring network of Mexico City was that the software was very slow and sometimes, convergence of the algorithm could not be achieved unless very informative prior distributions were used. In the present work we keep the assumption of a non-homogeneous Poisson process. However, the estimation of the parameters of the intensity function of this Poisson process is made using a combination of Gibbs sampling and Metropolis-Hasting type algorithms implemented using the R software. The algorithm produces very fast running time (with rare exceptions) even with non informative prior distributions. We are going to apply the theoretical part of this work to the ozone data provided by the monitoring network of the Metropolitan Area of Mexico City. The Metropolitan Area is divided into five regions or sections corresponding to the Northeast (NE), Northwest (NW), Centre (CE), Southeast (SE) and Southwest (SW) and the ozone monitoring stations are placed throughout the city (see Álvarez et al., 2005; Achcar et al., 2008a and <http://www.sma.df.gob.mx>). We are going to analyse the data for each region separately. Even though the threshold to emit an environmental alert in Mexico City is 0.2ppm, we are going to consider the threshold 0.17ppm in order to see how the frequency of environmental alerts is affected if 0.17ppm were the new threshold for declaring that type emergency.

This work is organised as follows. In Section 2, the basic assumptions of the Poisson models are presented. Section 3 describes the Bayesian formulation used. In Section 4, the Gibbs sampling and the Metropolis-Hastings algorithms used to estimate the parameters of the intensity functions of the Poisson models are described as well as the methodology used to select the best model fitting the data considered. An application to the case of ozone measurements in Mexico City is given in Section 5. Finally, in Section 6 some comments about the results obtained are presented. The codes of the R programmes used to estimate the parameters of the models are given in an Appendix before the list of references.

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

## 2. Some non-homogeneous Poisson models

There are several works that use non-homogeneous Poisson model to study problems in a wide variety of research areas (see for example Achcar, 2001; Achcar et al., 1998; Ramírez-Cid & Achcar, 1999; Wilson & Costello, 2005, among others). As in Achcar et al. (2008a, 2010, 2011), in here, non-homogeneous Poisson processes are used as models to estimate the probability that an ozone measurement surpasses a given threshold a certain number of times in a time interval of interest. However, in here we focus mainly in the behaviour of the mean of the Poisson process. The description of the models considered here are given as follows.

Let  $N_t \geq 0$  record the number of times that a given threshold for a given pollutant is surpassed in the time interval  $[0, t)$ ,  $t \geq 0$ . We assume that  $N = \{N_t : t \geq 0\}$  follows a Poisson process which is non-homogeneous in time and with rate function  $\lambda(t) > 0$ ,  $t \geq 0$ . Therefore, at time  $t$ , the random variable  $N_t$  has Poisson distribution with rate function  $\lambda(t)$  and mean function

$m(t) = \int_0^t \lambda(s) ds$ , i.e., for  $k = 0, 1, \dots$  and  $s, t \geq 0$ ,

$$P(N_{t+s} - N_t = k) = \frac{[m(t+s) - m(t)]^k}{k!} \exp(-[m(t+s) - m(t)]). \quad (1)$$

The function  $\lambda(t)$ ,  $t \geq 0$ , will depend on some parameters that need to be estimated. Once those parameters are estimated, the resulting form for  $\lambda(t)$  may be substituted in (1) and the corresponding probabilities may be calculated. Additionally, since the rate function is related to the rate at which a surpassing occurs, we are able to obtain information on that as well.

There are several forms that  $\lambda(t)$ ,  $t \geq 0$ , may assume. Based on the information provided by previous works, we are going to consider three forms for the rate function. They are given by the Weibull (W), Musa-Okumoto (MO) (Musa & Okumoto, 1984), and a generalised form of the Goel-Okumoto (GGO) (Goel & Okumoto, 1978) models, i.e.,

$$\lambda^{(W)}(t) = (\alpha/\sigma) (t/\sigma)^{\alpha-1}, \quad (2)$$

$$\lambda^{(MO)}(t) = \beta/(t + \alpha), \quad (3)$$

$$\lambda^{(GGO)}(t) = \alpha\beta\gamma t^{\gamma-1} e^{-\beta t^\gamma}, \quad (4)$$

respectively. Hence, the vectors of parameters to be estimated are  $\theta = (\alpha, \sigma), (\alpha, \beta) \in \mathbb{R}_+^2$ , in (2) and (3), respectively, and is  $\theta = (\alpha, \beta, \gamma) \in \mathbb{R}_+^3$  in (4).

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

*Remark.* Note that, even though the rate functions (2), (3) and (4) have been considered in past works, the interest here is to analyse the fitting for more recent data. That is justified by the fact that the year 2000 was the point in time where the last of a series of important measures to control pollution emission were taken by the environmental authorities in Mexico City. Previous preliminary analysis shows that the behaviour of the daily maximum ozone measurements have changed.

### 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 describing this rate function. Therefore, assume that there are  $K \geq 1$  days in which a given threshold is surpassed by the daily maximum ozone measurement. Let  $d_1, d_2, \dots, d_K$  be those days and denote by  $\mathbf{D} = \{d_1, d_2, \dots, d_K\}$  the set of observed data. There is a natural relationship among posterior and prior distributions and the likelihood function of the model which is given by

$$P(\theta | \mathbf{D}) \propto L(\mathbf{D} | \theta) P(\theta) \quad (5)$$

where  $P(\theta | \mathbf{D})$  and  $P(\theta)$  are the posterior and the prior distributions of the vector of parameters  $\theta$ , respectively, and  $L(\mathbf{D} | \theta)$  is the likelihood function of the model. The components of (5) are given as follows.

1. **The likelihood function.** Since a non-homogeneous Poisson model has been assumed, we have that the likelihood function will take the form (see for instance Cox & Lewis, 1966; Lawless, 1986)

$$L(\mathbf{D} | \theta) = \left[ \prod_{i=1}^K \lambda(d_i) \right] \exp(-m(d_K)), \quad (6)$$

where  $\theta$  is the vector of parameters. The likelihood function takes different forms depending on the expression for the rate function  $\lambda(t)$ ,  $t \geq 0$ . Hence, taking that into account we have the following.

In the case of the Weibull function given by (2), we have that  $m^{(W)}(t) = (t/\sigma)^\alpha$ , and therefore,

$$L^{(W)}(\mathbf{D} | \alpha, \sigma) \propto \left( \frac{\alpha}{\sigma^\alpha} \right)^K \left( \prod_{i=1}^K d_i^{\alpha-1} \right) \exp[-(d_K/\sigma)^\alpha]. \quad (7)$$

When considering the Musa-Okumoto function given by (3), we have that  $m^{(MO)}(t) = \beta \log(1 + t/\alpha)$ , and then,

$$L^{(MO)}(\mathbf{D} | \alpha, \beta) \propto \beta^K \left( \prod_{i=1}^K \frac{1}{d_i + \alpha} \right) \exp[-\beta \log(1 + d_K/\alpha)]. \quad (8)$$

In the case of the generalised Goel-Okumoto function given by (4), we have that the mean is  $m^{(GGO)}(t) = \alpha [1 - \exp(-\beta t^\gamma)]$ , and therefore,

$$L^{(GGO)}(\mathbf{D} | \alpha, \beta, \gamma) \propto (\alpha \beta \gamma)^K \left( \prod_{i=1}^K d_i^{\gamma-1} \right) \exp \left( -\beta \sum_{i=1}^K d_i^\gamma \right) \exp \left[ -\alpha \left( 1 - e^{-\beta d_K^\gamma} \right) \right]. \quad (9)$$

2. **The prior distributions.** Assuming prior independence among the parameters of each model considered we have the following prior distributions. In the case of the Weibull intensity (2), in a first instance we consider for all regions and parameters, uniform prior distributions defined on appropriate intervals. In a second instance, we take  $\text{Beta}(a_1, b_1)$  and  $\text{Gamma}(a_2, b_2)$  prior distributions for  $\alpha$  and  $\sigma$ , respectively. (In here, we are denoting by  $\text{Beta}(a, b)$  and  $\text{Gamma}(c, d)$  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.) When considering the intensity function (3) we take uniform prior distributions, with appropriate hyperparameters, for all parameters and regions. In the case of the intensity function (4) we assume that  $\alpha$  and  $\beta$  have uniform prior distributions (with appropriate hyperparameters) for all regions. In the case of the parameter  $\gamma$  we have  $\gamma \sim \text{Gamma}(a_3, b_3)$  for all regions. The hyperparameters  $a_i$  and  $b_i$ , for  $i = 1, 2, 3$  are considered to be known and will be specified later. They will be such that either we have non-informative prior distributions or using prior information of experts we have informative prior distributions for the parameters of each model.

3. **The posterior distributions.** The posterior distributions for the models considered here have the following forms. In the case of the Weibull intensity (2), we have that

$$P(\alpha, \sigma | \mathbf{D}) \propto \alpha^{a_1+K-1} (1-\alpha)^{b_1-1} \sigma^{a_2-\alpha K-1} e^{-b_2\sigma} \left[ \prod_{i=1}^K d_i^{\alpha-1} \right] \exp[-(d_K/\sigma)^\alpha]. \quad (10)$$

*Remark.* In the case of uniform prior distributions for  $\alpha$  and  $\sigma$ , we have that  $P(\alpha, \sigma | \mathbf{D}) \propto (\alpha/\sigma^\alpha)^K \left[ \prod_{i=1}^K d_i^{\alpha-1} \right] \exp[-(d_K/\sigma)^\alpha]$ .

When considering  $\lambda(t)$  given by (3), we have that,

$$P(\alpha, \beta | \mathbf{D}) \propto \beta^K \left( \prod_{i=1}^K \frac{1}{d_i + \alpha} \right) \exp \left( -\beta \log \left[ 1 + \frac{d_K}{\alpha} \right] \right). \quad (11)$$

In the case of  $\lambda(t)$  given by (4) we have that

$$P(\alpha, \beta, \gamma | \mathbf{D}) \propto (\beta \alpha)^K \gamma^{K+a_3-1} e^{-\gamma b_3} \exp \left[ -\alpha \left( 1 - e^{\beta d_K^\gamma} \right) \right] \exp \left[ -\beta \left( \sum_{i=1}^K d_i^\gamma \right) \right] \left[ \prod_{i=1}^K d_i^{\gamma-1} \right]. \quad (12)$$

#### 4. Conditional marginal posterior distributions, a Gibbs sampling and a Metropolis-Hasting algorithm

In order to simulate values from the joint posterior distribution in each model, we use a combination of Gibbs and Metropolis-Hastings type algorithm (see Metropolis et al., 1953; Hastings, 1970; Gelfand & Smith, 1990). The estimation of the parameters in each model will be made through a sample obtained from their complete marginal conditional posterior distribution using a Metropolis-Hasting algorithm. The values obtained are used in the next step of the Gibbs sampling. The complete marginal conditional posterior distributions in each model are given as follows.

1. In the case of the Weibull intensity (2) and complete posterior distribution given by (10), we have that

$$P(\alpha | \sigma, \mathbf{D}) \propto \psi_1(\alpha, \sigma) \alpha^{a_1-1} (1-\alpha)^{b_1-1} \\ P(\sigma | \alpha, \mathbf{D}) \propto \psi_2(\alpha, \sigma) \sigma^{a_2-1} e^{-b_2\sigma},$$

where

$$\psi_1(\alpha, \sigma) = \exp \left[ K \log(\alpha) - K \alpha \log(\sigma) + (\alpha - 1) \sum_{i=1}^K \log(d_i) - \left( \frac{d_K}{\sigma} \right)^\alpha \right], \\ \psi_2(\alpha, \sigma) = \exp \left[ -K \alpha \log(\sigma) - \left( \frac{d_K}{\sigma} \right)^\alpha \right].$$



*Remark.* In the case of uniform prior distribution for  $\alpha$  and  $\sigma$  we have that  $P(\alpha | \sigma, \mathbf{D}) \propto \psi_1(\alpha, \sigma)$  and  $P(\sigma | \alpha, \mathbf{D}) \propto \psi_2(\alpha, \sigma)$ .

2. When considering the intensity function given by (3) we have that

$$P(\alpha | \beta, \mathbf{D}) \propto \psi_3(\alpha, \beta),$$

$$P(\beta | \alpha, \mathbf{D}) \propto \psi_4(\alpha, \beta),$$

where

$$\psi_3(\alpha, \beta) = \exp \left[ - \sum_{i=1}^K \log(d_i + \alpha) - \beta \log \left( 1 + \frac{d_K}{\alpha} \right) \right],$$

$$\psi_4(\alpha, \beta) = \exp \left[ K \log(\beta) - \beta \log \left( 1 + \frac{d_K}{\alpha} \right) \right],$$

i.e., conditioned on  $\alpha$  and  $\mathbf{D}$ ,  $\beta \sim \text{Gamma}(K, \log(1 + d_K/\alpha))$

3. In the case of the intensity function (4) we have

$$P(\alpha | \beta, \gamma, \mathbf{D}) \propto \psi_5(\alpha, \beta, \gamma),$$

$$P(\beta | \alpha, \gamma, \mathbf{D}) \propto \psi_6(\alpha, \beta, \gamma),$$

$$P(\gamma | \alpha, \beta, \mathbf{D}) \propto \psi_7(\alpha, \beta, \gamma) \gamma^{a_3-1} e^{-b_3 \gamma},$$

where

$$\psi_5(\alpha, \beta, \gamma) = \exp \left[ K \log(\alpha) - \alpha \left( 1 - e^{-\beta d_K^\gamma} \right) \right],$$

$$\psi_6(\alpha, \beta, \gamma) = \exp \left[ K \log(\beta) + \alpha e^{-\beta d_K^\gamma} - \beta \sum_{i=1}^K d_i^\gamma \right],$$

$$\psi_7(\alpha, \beta, \gamma) = \exp \left[ K \log(\gamma) + \alpha e^{-\beta d_K^\gamma} - \beta \sum_{i=1}^K d_i^\gamma + (\gamma - 1) \sum_{i=1}^K \log(d_i) \right].$$

Note that from the forms of most of the marginal conditional distributions obtained here it is not straightforward to obtain sampled values of the parameters directly from them. Hence, we are going to obtain a sample by using a Metropolis-Hastings algorithm.

We sample the proposed values for the parameters using their respective prior distributions. Therefore, if  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_n)$  is the vector of parameters in the present step of the algorithm, then the acceptance probability in the Metropolis-Hastings algorithm is given by

$$q(\theta_j, \theta'_j) = \min \left\{ 1, \frac{\psi_l(\theta'_j, \boldsymbol{\theta}_{(-j)})}{\psi_l(\theta_j, \boldsymbol{\theta}_{(-j)})} \right\}, \quad j = 1, 2, \dots, n, \quad (13)$$

where  $\boldsymbol{\theta}' = (\theta'_1, \theta'_2, \dots, \theta'_n)$  is the proposed vector where each coordinate was sampled using its prior distribution,  $\boldsymbol{\theta}_{(-j)}$  is the vector  $\boldsymbol{\theta}$  without its  $j$ th coordinate and  $\psi_l$  is the appropriate  $\psi$  function that appears in the expression for the marginal conditional distributions.

In order to select the model that best explains the behaviour of the data considered here, we use the plots of the Monte Carlo estimates for  $m(t)$  based on the simulated samples versus  $t$  and the accumulated number of occurrences up to  $t$  versus time. If the curves are similar,

we have an indication of a good fit of the proposed model to the data. We also consider a Bayesian discrimination method. Hence, from Raftery (1996) we have that the marginal likelihood function of the whole data set  $\mathbf{D}$  for Model  $l, l = 1, 2, \dots, J$  is given by

$$V_l = \int L(\mathbf{D} | \theta^{[l]}) P(\theta^{[l]}) d\theta^{[l]} \quad (14)$$

where  $\theta^{[l]}$  is the vector of parameters for Model  $l$  and  $P(\theta^{[l]})$  is the joint prior distribution for  $\theta^{[l]}$ . The Bayes factor criterion prefers Model  $i$  to Model  $j$  if  $V_j/V_i < 1$ . A Monte Carlo estimate for the marginal likelihood  $V_l$  is given by

$$\hat{V}_l = \frac{1}{M} \sum_{i=1}^M L(\mathbf{D} | \theta^{[l,i]}) \quad (15)$$

where  $M$  is the size of the simulated Gibbs sample and  $\theta^{[l,i]}, i = 1, 2, \dots, M$  is the sample obtained when considering the Model  $l, l = 1, 2, \dots, J$ .

## 5. An application to the ozone measurements in Mexico City

In this section we apply the results described in earlier sections to the case of daily maximum ozone measurements from the monitoring network of the Metropolitan Area of Mexico City. The data used in the analysis (obtained from <http://www.sma.df.gob.mx/simat/>) corresponds to seven years (from 01 January 2003 to 31 December 2009) of the daily maximum measurements in 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. The seven-year average measurements in regions NE, NW, CE, SE and SW are 0.085, 0.097, 0.098, 0.101 and 0.112, respectively, with standard deviation 0.028, 0.036, 0.036, 0.033 and 0.039. We also have that the threshold 0.17ppm was surpassed 13, 78, 53, 40 and 143 days in regions NE, NW, CE, SE and SW, respectively.

The computational details of the implementation of the computer programme of the algorithm used to estimate the parameters are given as follows. When considering the Weibull rate function, we have used three different initial values for the algorithm generating the values of  $\alpha$  and  $\sigma$ . Those initial values were based on the results given by Achcar et al. (2008a) for each specific region. Hence, we have taken one value near the left extreme of the 95% credible interval for the mean, another near the right extreme and the mean itself. In the case of the rate functions (3) and (4) similar procedure was considered.

Regarding the hyperparameters of the prior distributions when considering the Weibull rate function we have the following. In the case of the first stage of the sampling procedure (i.e., uniform prior distributions for  $\alpha$  and  $\sigma$ ), we have that in all regions the parameter  $\alpha$  has a  $U(0,1)$  prior distribution. The parameter  $\sigma$  has a  $U(0,10)$  prior distribution when using data from region CE and a  $U(0,100)$  prior distribution for the remaining regions. In the case of non uniform prior distributions, the value of the hyperparameters may vary from region to region. In the case of the parameter  $\alpha$  we have that the Beta distribution has hyperparameters  $a_1 = 3/25$  and  $b_1 = 2/25$  in the case of regions NW, SE and SW, and is  $Beta(1/8, 1/8)$  and  $Beta(7/200, 3/200)$  in the case of regions CE and NE, respectively. When considering the



parameter  $\sigma$  we have that the prior distributions are Gamma(3249/640, 57/640), Gamma(3/2, 1/2), Gamma(9/4, 3/4), Gamma(1.69, 0.13) and Gamma(49/20, 7/20) for regions NE, NW, CE, SE and SW, respectively.

*Remark.* Observe that taking either  $\alpha \sim U(0,1)$  or  $\alpha$  with a Beta distribution, we have that  $\lambda^{(W)}(t)$  is a decreasing intensity function.

In almost all cases and regions we consider a burn-in period of 30000 steps. The exceptions are the cases of Weibull rate function with  $\alpha$  having a Beta prior distribution and regions CE, NE and SW, and the case of GGO rate function and region SW where we have a burn-in period of 130000 steps. In all cases a sample of size 10500 was used to estimate the parameters of the models. The burn-in period was determined by trace plot analysis as well as by using the Gelman-Rubin test (see Gelman & Rubin, 1992).

*Remark.* In the case of GGO rate function and region SW, the sampling procedure were as follows, for each iteration producing a value of  $\beta$  we had 20 iterations for  $\alpha$  and  $\sigma$ . Hence, we have 20 iterations for  $\alpha$ , and the last generated value is used to generate a  $\beta$ . This generated  $\beta$  is used to generate  $\sigma$  and again we have 20 iterations for  $\sigma$  and the 20th generated value is used to generate  $\alpha$ , and so on. This artifice was necessary to obtain convergence of the Gibbs and Metropolis-Hastings algorithms.

The estimated mean, standard deviation (indicated by SD) and the 95% credible interval in the case of Weibull rate function are given in Table 1.

Weibull		Mean		SD		95% credible interval	
		uni	no-uni	uni	no-uni	uni	no-uni
NE	$\alpha$	0.7	0.891	0.11	0.14	(0.476, 0.917)	(0.575, 1)
	$\sigma$	3.47	91.05	2.37	33.404	(0.7635, 9.652)	(28.909, 152.84)
NW	$\alpha$	0.6	0.62	0.063	0.054	(0.543, 0.792)	(0.526, 0.726)
	$\sigma$	3.47	2.48	2.37	1.524	(0.763, 9.561)	(0.642, 5.884)
CE	$\alpha$	0.58	0.56	0.063	0.052	(0.462, 0.699)	(0.471, 0.674)
	$\sigma$	3.075	2.49	2.136	1.502	(0.432, 8.492)	(0.524, 6.228)
SW	$\alpha$	0.62	0.61	0.013	0.029	(0.591, 0.642)	(0.569, 0.675)
	$\sigma$	0.785	0.84	0.087	0.31	(0.626, 0.962)	(0.387, 1.541)
SE	$\alpha$	0.69	0.65	0.093	0.075	(0.525, 0.884)	(0.517, 0.811)
	$\sigma$	13.83	10.07	10	6.373	(2.129, 39.865)	(1.984, 25.831)

Table 1. Posterior mean, standard deviation (indicated by SD) and 95% credible interval of the parameters  $\alpha$  and  $\sigma$  when the Weibull rate function is considered.

In the case of the MOP model, we have that  $\alpha$  and  $\beta$  will have U(0,700) and U(1,2500) prior distributions, respectively, in all regions.

*Remark.* Note that in the case of the parameters  $\beta$  its marginal conditional posterior distribution depends on the values of  $K$  and  $d_K$  which may vary from region to region. Hence, we have that  $K = 13, 78, 53, 40$  and  $143$ , and  $d_K = 1927, 2258, 2283, 2282$  and  $2431$  for regions NE, NW, CE, SE and SW, respectively. Table 2 gives the estimated quantities of interest.

When the GGO rate function is taken into account, we have for all regions that the parameters  $\beta$  and  $\gamma$  have U(0.000001,0.001) and Gamma(0.5,1.5) prior distributions, respectively. In the case of the parameter  $\alpha$  we have that for regions NE, NW, CE and SE, its prior distribution is U(300,2000). In the case of region SW we have  $\alpha$  with U(0,3000) prior distribution. A summary of the estimated quantities of interest is given in Table 3.

The value of the Bayes factor for each model and region are given in Table 4.

We may see from Table 4 that the best model according to the Bayes factor criterion in each region is the one using the Weibull rate function. In the case of regions NE and CE the model

MO		Mean	SD	95% credible interval
NE	$\alpha$	199.178	356.8562	(0, 1115)
	$\beta$	2.981576	5.195376	(0, 16.09559)
NW	$\alpha$	96.21955	167.9046	(0, 536.9866)
	$\beta$	10.90067	17.844407	(0, 49.76161)
CE	$\alpha$	209.7355	102.5246	(73.91646, 476.3019)
	$\beta$	21.49208	4.840937	(13.94708, 32.71157)
SW	$\alpha$	507.8983	109.1856	(292.9967, 687.4573)
	$\beta$	81.2965	10.885889	(60.82191, 102.7176)
SE	$\alpha$	379.7262	148.8562	(134.278, 672.9357)
	$\beta$	20.83115	4.98808	(12.37612, 31.46646)

Table 2. Posterior mean, standard deviation (indicated by SD) and 95% credible interval of the parameters  $\alpha$  and  $\beta$  when the MO rate function is considered.

GGO		Mean	SD	95% credible interval
NE	$\alpha$	1024	483.15	(330.65, 1938.55)
	$\beta$	3.946E-04	0.0003	(2.33E-5, 0.000947)
	$\gamma$	0.522	0.126	(0.321, 0.8165)
NW	$\alpha$	1200	454.73	(389.173, 1957.981)
	$\beta$	0.00064	0.00022	(0.000212, 0.0009803)
	$\gamma$	0.624	0.0624	(0.5173, 0.7563)
CE	$\alpha$	1309	424.0873	(486.984, 1964.205)
	$\beta$	0.00067	0.00020	(0.000265, 0.000985)
	$\gamma$	0.5451	0.0573	(0.449, 0.671)
SW	$\alpha$	710	450.32	(0.6246, 0.8995)
	$\beta$	0.000724	0.000193	(0.000179, 0.000999)
	$\gamma$	0.77	0.0758	(0.583, 0.9612)
SE	$\alpha$	1059	480.94	(332.122, 1942.537)
	$\beta$	0.00052	0.0025	(0.000118, 0.000969)
	$\gamma$	0.592	0.076	(0.4511, 0.7464)

Table 3. Posterior mean, standard deviation (indicated by SD) and 95% credible interval of the parameters  $\alpha$ ,  $\beta$  and  $\gamma$  when the GGO rate function is considered.

	Weibull		MO	GGO
	uni	no-uni		
NE	3.499E-16	2.0712E-19	3.564126E-39	9.08166E-35
NW	1.684E-44	6.8451E-35	1.996936E-146	7.3953E-144
CE	8.2569E-26	5.6684E-29	2.673235E-108	1.52702E-104
SE	2.6252E-29	1.00755E-28	4.4002195E-89	4.7143E-86
SW	2.7104E-69	7.1844E-57	4.037815E-250	7.8386E-233

Table 4. Bayes factor for each model and region considered.

assuming uniform prior distribution for  $\alpha$  and  $\sigma$  is the one with larger Bayes factor. In the case of the other regions the version where  $\alpha$  has a Beta prior distribution is the one with larger Bayes factor.

Another way of verifying the fitting of a model to the data, is to perform graphical analysis. Hence, in Figure 1 we have the plots of the observed and estimated means for each of the

regions and rate functions considered here. The plain solid line represents the accumulated observed mean. The lines with ■, ◆, ▲ and ▼, correspond to the accumulated estimate means for models GGO, MO, Weibull where  $\alpha$  has a uniform prior and Weibull where  $\alpha$  has a Beta prior, respectively

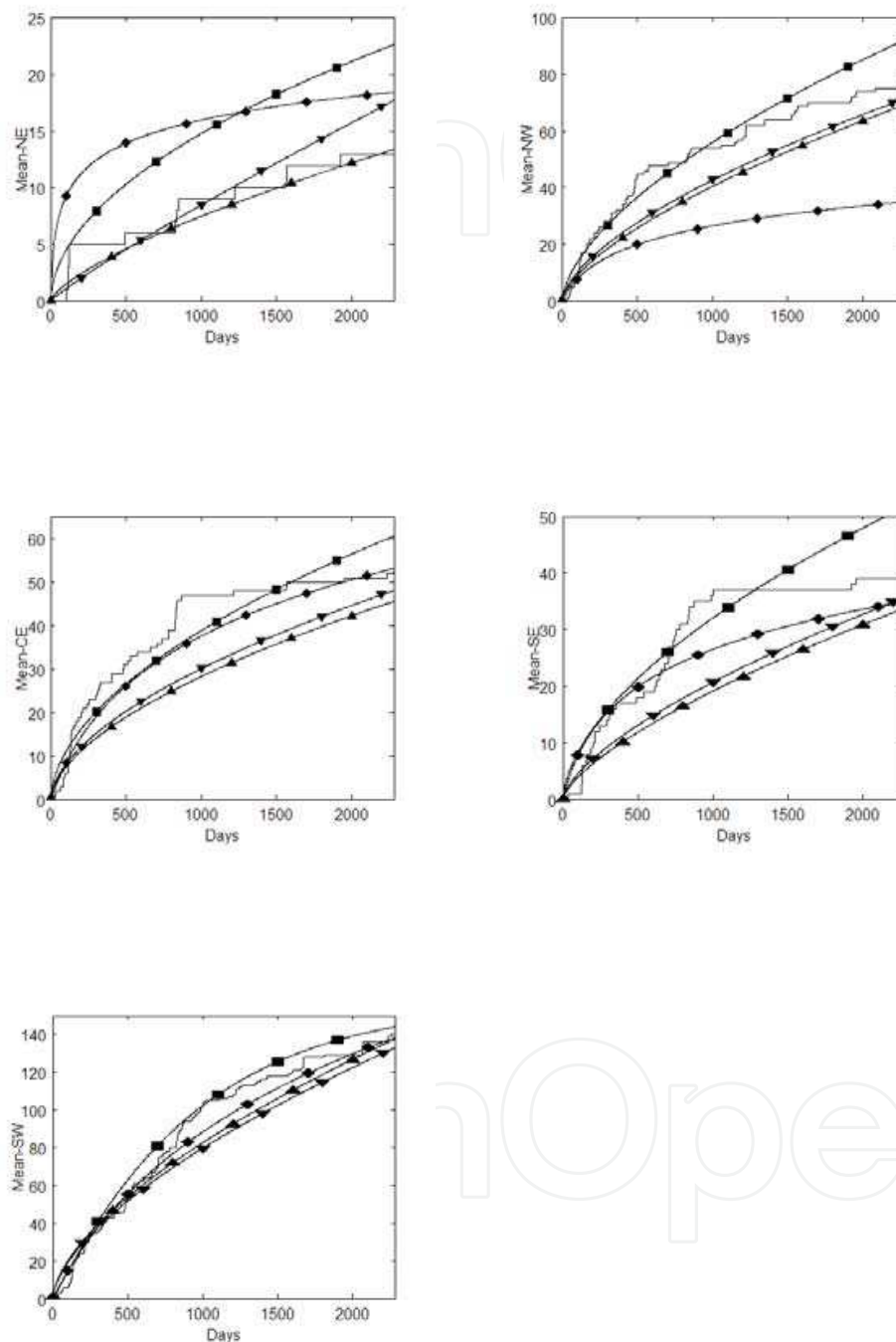


Fig. 1. Accumulated observed and estimated means for all models and all regions. Solid plain line represents the accumulated observed mean. Lines with ■, ◆, ▲ and ▼, correspond to the accumulated estimate means for models GGO, MO, Weibull where  $\alpha$  has a uniform prior and Weibull where  $\alpha$  has a Beta prior, respectively.

It is possible to see from Figure 1 that the model that best fit the data varies according to the region. We have that for region NE the estimated mean obtained by using the Weibull

rate function with  $\alpha$  uniformly distributed is the one that best fit the observed mean. In the case of region NW, we have that the GGO model is the most adequate to represent the behaviour of the observed mean. However, we may notice that towards the end of the observational period the estimated Weibull mean function (either formulation) also provides a good approximation. When we look at the observed mean in region CE, we have that the GGO model fits quite well the observed mean until approximately the year 2007 and then it starts to drift away from it. However, the estimated mean using the MO model gives a good approximation from the year 2007 on. The case of region SE is similar to that of region CE. However, the fitting of the estimated mean function using the GGO model ends earlier (around 2006) and the fitting using the MO model is not as good as in the case of region CE. When we consider region SW, we have that the estimated mean function using any of the models provides a very good approximation for the observed mean. The best approximation is the one given by the GGO model though.

## 6. Conclusion

In this work we have considered a Gibbs sampling algorithm to estimate the parameters of a rate function of a non-homogeneous Poisson process. The methodology was applied to some models used to study the behaviour of ozone measurements from the monitoring network of Mexico City. Two criteria were used to select the model that best adjusts to the observed data. One of them was graphical analysis and the other was the Bayes factor.

Looking at Table 3 we have that the Weibull rate function is the one producing the largest Bayes factor and therefore, under this criterion, that is the model that should be considered. However, when looking at Figure 1, we have that depending on the region of the city considered, we might have a better adjustment of the estimated and observed mean number of exceedances if other rate function were considered.

We have that the Weibull rate function has the best graphical adjustment only when region NE is considered. In the other regions we have that the best graphical adjustment either is when using the GGO rate function or when using the MO (see Figure 1, regions NW, CE and SW). There are situations where the GGO rate function adjusts well until a given point in time and afterwards some of the other rate functions considered are the ones that give the best graphical fitting (see Figure 1, regions CE and SE).

Looking at Figure 1, we have that perhaps in the case of regions NW, CE and SE one could consider the GGO model with the presence of change-points. Another option is to consider two different parametric forms for the rate function of the Poisson process. One of them explaining the behaviour of the data up to a point in time and another explaining the behaviour afterwards.

When comparing the results obtained here with previous works, we have the following. Taking into account the study performed using measurements from 01 January 1998 to 31 December 2004, in all regions the selected model (using the Bayes information criterion and the deviance information criterion) is when the Weibull rate function is used (see Achcar et al., 2008a). In terms of graphical adjustment, in most of the cases, results were similar to the ones obtained here using the 2003-2009 data. However, when considering regions CE and SE, the fitting of the Weibull rate function is better when using the 1998-2004 data. The values of the parameters  $\alpha$  and  $\sigma$  of the Weibull rate function are very different with the exception of regions NW and SW, where the parameters  $\alpha$  are very similar, but the parameter  $\sigma$  presents significant differences.

In Achcar et al. (2010), we have that the GGO and MO rate function were used when considering measurements also taken from 01 January 1998 to 31 December 2004. In that case the regional splitting of the city was not taken into account, but the ozone measurements were the overall measurements for the Metropolitan Area of Mexico City. In that work, the presence of a change-point was assumed and we have that the model selected using the deviance information criterion to represent the behaviour of the ozone measurements was the GGO with the presence of a change-point. In terms of graphical fitting we have that the GGO model with a change-point also provides the best fit. When considering the 2003-2009 data and the regional division of the city, we have that not always the GGO model is the one providing the best graphical fitting. Additionally, when using other criteria for selecting the models, the GGO may not be the selected model either, as we can see from Figure 1. In the cases where the GGO rate function is selected as providing the best graphical fitting, we have that, when compared to the results obtained using the 1998-2004 data, the values of the parameters  $\alpha$  and  $\beta$  are similar except in the case of region SW. However, the parameter  $\gamma$  presents very different behaviour (see Achcar et al., 2010).

As we can see from Figure 1, sometimes the MO rate function may be used to explain the behaviour of the data in the last years of the period 2003-2009. This rate function was also considered for the overall measurements during the period 1998-2004. The graphical fitting was much worse than the fitting obtained when using the 2003-2009 data.

We would like to recall that around the year 2000 was implemented the last of the major environmental measures taken by the environmental authorities in Mexico. We also have that from that year on, the daily maximum ozone concentration presents a decreasing tendency. The change in behaviour is captured by the models presented here, where different graphical fittings and values of parameters are detected when considering the 1998-2004 and 2003-2009 data sets.

An advantage of using the algorithm and consequently the programme presented here is that it may be modified by the user to adapt to each particular case studied. Each step of the algorithm and programme can be monitored and therefore, if there is any problem with the model considered, one can detect precisely where the problem is. Even though in some cases the burn-in period of the algorithm is larger than in previous cases, the overall running time is smaller than when considering some other software packages as the ones used by Achcar et al. (2008a, 2010, 2011).

## 7. Acknowledgements

The authors thank Guadalupe Tzintzun of the Instituto Nacional de Ecología of the Ministry of Environment of Mexico for providing the data from the monitoring network. This work is funded by the PAPIIT research project IN104110-3 of the DGAPA-UNAM, Mexico.

## 8. Appendix

In this Appendix we provide the R code of the Gibbs and Metropolis-Hastings algorithms for each rate function considered in this work.

### 1. Weibull rate function

In here we have the code for the case where the Weibull rate function is considered.

```
w<-function(zona,cant_iteraciones,alfa_inicial,
```



```

        sigma_inicial,a1,b1,a2,b2){
# creates the Markov chain and store the initial state
alfa<-vector("numeric",cant_iteraciones)
sigma<-vector("numeric",cant_iteraciones)
alfa[1]<-alfa_inicial
sigma[1]<-sigma_inicial
# upload the vector with the observed data
datos_observados<-scan(file = zona, what = double(0), nmax = -1,
  sep = " ", quote = "", dec = ".",
  skip = 0, nlines = 0, na.strings = "NA",
  flush = FALSE, fill = FALSE, strip.white = FALSE,
  quiet = FALSE, blank.lines.skip = TRUE, multi.line = TRUE,
  comment.char = "", allowEscapes = TRUE)
for(i in 2:cant_iteraciones){
  # obtain the proposed value from the Beta and Gamma prior
  # distributions
  alfa_propuesta<-rbeta(1,a1,b1)
  sigma_propuesta<-rgamma(1,a2,b2)
  # leave sigma fixed and check if the proposed
  # value for alpha is accepted
  aceptacion<-min(1,aceptar_alfa(sigma[i-1],alfa[i-1],
    alfa_propuesta,datos_observados))
  a<-runif(1)
  if(a<aceptacion)
    alfa[i]=alfa_propuesta # accepted
  else
    alfa[i]=alfa[i-1] #not accepted
  # leave alpha fixed and check if the proposed value
  # for sigma is accepted
  aceptacion<-min(1,aceptar_sigma(tail(datos_observados,1),
    length(datos_observados),alfa[i],
    sigma[i-1],sigma_propuesta))
  a<-runif(1)
  if(a<aceptacion)
    sigma[i]=sigma_propuesta #accepted
  else
    sigma[i]=sigma[i-1] # not accepted
  }
}
aceptar_alfa<-function(sigma_actual,alfa_actual,
  alfa_propuesta,datos_observados){
n<-length(datos_observados)
dn<-tail(datos_observados,1)
logsum<-0
for(k in 1:n){
  logsum<-logsum+log(datos_observados[k])
}
}

```



```

term<-n*log(alfa_propuesta/alfa_actual)+n*log(sigma_actual)
      *(alfa_actual-alfa_propuesta)+logsum
      *(alfa_propuesta-alfa_actual)
      +(dn/sigma_actual)^(alfa_actual)
      -(dn/sigma_actual)^(alfa_propuesta)
exp(term)
}
aceptar_sigma<-function(dn,n,alfa_actual,sigma_actual,
      sigma_propuesta){
term<-n*alfa_actual*log(sigma_actual/sigma_propuesta)
      +(dn/sigma_actual)^(alfa_actual)
      -(dn/sigma_propuesta)^(alfa_actual)
exp(term)
}

```

## 2. GGO rate function

In this subsection of the Appendix we have the code for the Gibbs and Metropolis-Hastings algorithms when we take the GGO rate function.

```

ggo<-function(zona,cant_iteraciones,alfa_inicial,
      beta_inicial,gamma_inicial,a5,b5,
      a6,b6,a7,b7){
# create Markov chain and store initial values
alfa<-vector("numeric",cant_iteraciones)
beta<-vector("numeric",cant_iteraciones)
gamma<-vector("numeric",cant_iteraciones)
alfa[1]<-alfa_inicial
beta[1]<-beta_inicial
gamma[1]<-gamma_inicial
# upload the vector with the observed data
datos_observados<-scan(file = zona, what = double(0), nmax = -1,
      sep = " ", quote = "", dec = ".",
      skip = 0, nlines = 0, na.strings = "NA",
      flush = FALSE, fill = FALSE, strip.white = FALSE,
      quiet = FALSE, blank.lines.skip = TRUE, multi.line = TRUE,
      comment.char = "", allowEscapes = TRUE)
logsum<-0
for(i in 1:length(datos_observados)){
  logsum<-logsum+log(datos_observados[i])
}
for(i in 2:cant_iteraciones){
  # obtain the proposed value from the uniform and Gamma prior
  # distributions
  alfa_propuesto<-runif(1,a5,b5)
  beta_propuesto<-runif(1,a6,b6)
  gamma_propuesto<-rgamma(1,a7,b7)

```

```

# leave the parameters beta and gamma fixed and check if the
# proposed value for alpha is accepted
aceptacion<-min(1,aceptar_alfa(beta[i-1],alfa[i-1],
                             alfa_propuesto,datos_observados,gamma[i-1]))
a<-runif(1)
if(a<aceptacion)
  alfa[i]=alfa_propuesto # accepted
else
  alfa[i]=alfa[i-1] # not accepted
# leave the parameters alpha and gamma fixed and check if the
# proposed value for beta is accepted
aceptacion<-min(1,aceptar_beta(tail(datos_observados,1),
                             length(datos_observados),alfa[i],
                             beta[i-1],beta_propuesto,
                             datos_observados,gamma[i-1]))
a<-runif(1)
if(a<aceptacion)
  beta[i]=beta_propuesto # accepted
else
  beta[i]=beta[i-1] # not accepted
# leave the parameters alpha and beta fixed and check if the proposed
# values for gamma is accepted
aceptacion<-min(1,aceptar_gamma(datos_observados,logsum,
                              gamma_propuesto,gamma[i-1],beta[i-1],alfa[i-1]))
a<-runif(1)
if(a<aceptacion)
  gamma[i]=gamma_propuesto # accepted
else
  gamma[i]=gamma[i-1] # not accepted
}
}
aceptar_alfa<-function(beta_actual,alfa_actual,
                      alfa_propuesto,datos_observados,
                      gamma_actual){
  n<-length(datos_observados)
  dn<-tail(datos_observados,1)
  res<-(alfa_propuesto/alfa_actual)^n
  *exp((alfa_actual-alfa_propuesto)
  *(1-exp(-beta_actual*dn^gamma_actual)))
}
aceptar_beta<-function(dn,n,alfa_actual,beta_actual,
                      beta_propuesto,datos_observados,
                      gamma_actual){
  disum<-0
  for(i in 1:n){
    disum<-disum+datos_observados[i]^gamma_actual
  }
}

```

```

term<-disum*(beta_actual-beta_propuesto)
      +alfa_actual*(exp(-beta_propuesto*dn^gamma_actual)
      -exp(-beta_actual*dn^gamma_actual))
res<-((beta_propuesto/beta_actual)^n)*exp(term)
res
}
aceptar_gamma<-function(datos_observados,logsum,
      gamma_propuesto,gamma_actual,
      beta_actual,alfa_actual){
n<-length(datos_observados)
dn<-tail(datos_observados,1)
expsum<-0
for(i in 1:n){
  expsum<-expsum+(datos_observados[i]^gamma_actual
    -datos_observados[i]^gamma_propuesto)
}
term<-logsum*(gamma_propuesto-gamma_actual)
      +beta_actual*expsum+alfa_actual
      *(exp(-beta_actual*dn^gamma_propuesto)
      -exp(-beta_actual*dn^gamma_actual))
result<-((gamma_propuesto/gamma_actual)^n)*exp(term)
result
}

```

### 3. MO rate function

In here we present the R code of the Gibbs and Metropolis-Hastings algorithms when the MO rate function is used.

```

mo<-function(zona,cant_iteraciones,alfa_inicial,
      beta_inicial,a3,b3,a4,b4){
  # creates the Markov chain and store the initial state
  alfa<-vector("numeric",cant_iteraciones)
  beta<-vector("numeric",cant_iteraciones)
  alfa[1]<-alfa_inicial
  beta[1]<-beta_inicial
  # upload the vector with the observed data
  datos_observados<-scan(file = zona, what = double(0), nmax = -1,
    sep = " ", quote = "", dec = ".",
    skip = 0, nlines = 0, na.strings = "NA",
    flush = FALSE, fill = FALSE, strip.white = FALSE,
    quiet = FALSE, blank.lines.skip = TRUE, multi.line = TRUE,
    comment.char = "", allowEscapes = TRUE)
  n<-length(datos_observados)
  dn<-tail(datos_observados,1)
  for(i in 2:cant_iteraciones){
    # obtain a proposed value for alpha from its Gamma prior distribution
    # and check is the value is accepted

```

```

    alfa_propuesta<-rgamma(1,a3,b3)
    acepcation<-aceptar_alfa(beta[i-1],alfa[i-1],alfa_propuesta,
                             datos_observados)

    a<-runif(1)
    if(a<aceptacion)
        alfa[i]=alfa_propuesta # accepted
    else
        alfa[i]=alfa[i-1] # not accepted
        # obtain a random sample for the parameter beta using its Gamma
        # posterior distribution
        beta[i]<-rgamma(1,n+1,log(1+(dn/alfa[i])))
    }
}

aceptar_alfa<-function(current_beta,alfa_actual,alfa_propuesta,
                       datos_observados){
    # obtain the cardinality of the observed data sets and obtain
    # the last value of the set
    n<-length(datos_observados)
    dn<-tail(datos_observados,1)
    # calculate the sum of log(di)
    logsum<-0
    for(i in 1:n){
        logsum<-logsum+log((datos_observados[i]+alfa_actual)
                           /(datos_observados[i]+alfa_propuesta))
    }
    term<-logsum+current_beta*log((alfa_actual*(alfa_propuesta+dn))
                                   /(alfa_propuesta*(alfa_actual+dn)))
    exp(term)
}

```

## 9. References

- Achcar, J.A. (2001). Bayesian analysis for software reliability data, *Advances in Reliability. Handbook of Statistics* Vol. 20: 733-748.
- Achcar, J.A., Dey, D.K. & Niverthy, M. (1998). A Bayesian approach using nonhomogeneous Poisson process for software reliability models, in: S.K. Basu & S. Mukhopadhyay (eds.), *Frontiers in Reliability. Series on Quality, Reliability and Engeneering Statistics* Vol. 4, Calcutta University, India, pp. 1-18.
- Achcar, J.A., Fernández-Bremauntz, A.A., Rodrigues, E.R. & Tzintzun, G. (2008a). Estimating the number of ozone peaks in Mexico City using a non-homogeneous Poisson model, *Environmetrics* Vol. 19: 469-485. (<http://www.interscience.wiley.com>, DOI: 10.1002/env.890)
- Achcar, J.A., Zozolotto, H.C. & Rodrigues, E.R. (2008b). Bivariate volatility models applied to air pollution data, *Rev. Bras. Biom.* Vol. 26: 67-81.
- Achcar, J.A., Zozolotto, H.C. & Rodrigues, E.R. (2009a). Bivariate stochastic volatility models applied to Mexico City ozone pollution data, in G.C. Romano & A.G. Conti (eds.), *Air Quality in the 21st Century*, Nova Publishers, New York, pp. 241-267.

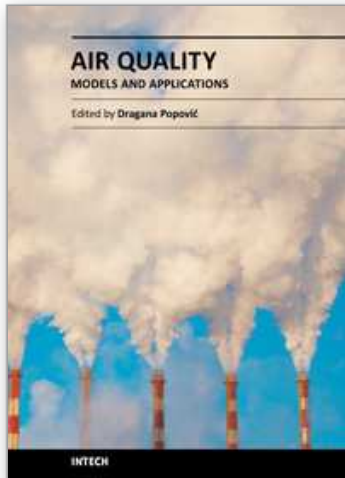
- Achcar, J.A., Rodrigues, E.R. & Tzintzun, G. (2009b). Using stochastic volatility models to analyse weekly ozone averages in Mexico City, *Environmental and Ecological Statistics*, (DOI:10.1007/s10651-010-0132-1).
- Achcar, J.A., Rodrigues, E.R., Paulino, C.D. & Soares, P. (2010). Non-homogeneous Poisson processes with a change-point: an application to ozone exceedances in Mexico City, *Environmental and Ecological Statistics* Vol. 17: 521-541. (DOI: 10.1007/s10651-009-0114-3)
- Achcar, J.A., Rodrigues, E.R. & Tzintzun, G. (2011). Using non-homogeneous Poisson models with multiple change-points to estimate the number of ozone exceedances in Mexico City, *Environmetrics* Vol. 22: 1-12. (<http://www.interscience.wiley.com>, DOI: 10.1002/env.1029)
- Álvarez, L.J., Fernández-Bremauntz, A.A., Rodrigues, E.R. & Tzintzun, G. (2005). Maximum a posteriori estimation of the daily ozone peaks in Mexico City, *Journal of Agricultural, Biological, and Environmental Statistics* Vol. 10: 276-290.
- Austin, J. & Tran, H. (1999). A characterization of the weekday-weekend behavior of ambient ozone concentrations in California, in *Air Pollution VII*, WIT Press, Ashurst Lodge, Ashurst, Southampton, UK, pp. 645-661.
- Bell, M.L., McDermott, A., Zeger, S.L., Samet, J.M. & Dominici, F. (2004). Ozone and short-term mortality in 95 US urban communities, 1987-2000, *Journal of the American Medical Society* Vol. 292: 2372-2378.
- Bell, M.L., Peng, R. & Dominici, F. (2005). The exposure-response curve for ozone and risk of mortality and the adequacy of current ozone regulations, *Environmental Health Perspectives* Vol. 114: 532-536.
- Bell, M.L., Goldberg, R., Rogrefe, C., Kinney, P.L., Knowlton, K., Lynn, B., Rosenthal, J., Rosenzwei, C. & Patz, J.A. (2007). Climate change, ambient ozone, and health in 50 US cities, *Climate Change* Vol. 82: 61-76.
- Cox, D.R. & Lewis, P.A. (1966). *Statistical analysis of series events*, Methuen, UK.
- Davis, L.W. (2008). The effect of driving restrictions on air quality in Mexico City, *Journal of Political Economy* Vol. 116: 39-81.
- EPA (US Environmental Protection Agency) (2008). *National Ambient Air Quality Standards*. Available at <http://www.epa.gov/air/criteria.html>.
- Flaum, J.B., Rao, S.T. & Zurbenko, I.G. (1996). Moderating influence of meteorological conditions on ambient ozone concentrations, *Journal of the Air and Waste Management Association* Vol. 46: 33-46.
- Gauderman, W.J., Avol, E., Gililand, F., Vora, H., Thomas, D., Berhane, K., McConnel, R., Kuenzli, N., Lurman, F., Rappaport, E., Margolis, H., Bates, D. & Peter, J. (2004). The effects of air pollution on lung development from 10 to 18 years of age, *The New England Journal of Medicine* 351: 1057-1067.
- Gelfand, A.E. & Smith, A.F.M. (1990). Sampling-based approaches to calculating marginal densities, *Journal of the American Statistical Association* Vol. 85: 398-409.
- Gelman, A. & Rubin, D.B. (1992). Inference from iterative simulation using multiple sequences, *Statistical Sciences* Vol. 7: 457-511.
- Goel, A.L. & Okumoto, K. (1978). An analysis of recurrent software failures on a real-time control system, *Proceedings of ACM Conference*, Washington-DC, USA, pp. 496-500.
- Gouveia, N. & Fletcher, T. (2000). Time series analysis of air pollution and mortality: effects by cause, age and socio-economics status, *Journal of Epidemiology and Community Health* Vol. 54: 750-755.



- Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications, *Biometrika* Vol. 57: 97-109.
- Horowitz, J. (1980). Extreme values from a nonstationary stochastic process: an application to air quality analysis, *Technometrics* Vol. 22: 469-482.
- Huerta, G. & Sansó, B. (2005). Time-varying models for extreme values, *Technical Report 2005-4*, Department of Applied Mathematics and Statistics. University of California, USA.
- Kumar, U., Prakash, A. & Jain, V.K. (2010). A multivariate time series approach to study the interdependence among  $O_3$ ,  $NO_x$  and VOCs in ambient urban atmosphere. *Environmental Modelling and Assessment*, (DOI: 10.1007/s10666-008-9167-1).
- Lanfredi, M. & Macchiato, M. (1997). Searching for low dimensionality in air pollution time series, *Europhysics Letters* Vol. 40: 589-594.
- Lawless, J. F. (1982). *Statistical models and methods for lifetime data*, John Wiley & Sons, USA.
- Loomis, D., Borja-Arbutó, V.H., Bangdiwala, S.I. & Shy, C.M. (1996). Ozone exposure and daily mortality in Mexico City: a time series analysis, *Health Effects Institute Research Report* Vol. 75: 1-46.
- Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. & Teller, E. (1953). Equations of state calculations by fast computing machine, *J. Chem. Phys.* Vol. 21: 1087-1091.
- Musa, J.D. & Okumoto, K. (1984). A logarithmic Poisson execution time model for software reliability measurement, *Proceedings of Seventh International Conference on Software Engineering*, Orlando, USA, pp. 230-238.
- NOM (2002). Modificación a la Norma Oficial Mexicana NOM-020-SSA1-1993, *Diario Oficial de la Federación*, 30 October 2002, Mexico. (In Spanish.)
- O'Neill, M.R., Loomis, D. & Borja-Aburto, V.H. (2004). Ozone, area social conditions and mortality in Mexico City, *Environmental Research* Vol. 94: 234-242.
- Pan, J.-N. & Chen, S.-T. (2008). Monitoring long-memory air quality data using ARFIMA model, *Environmetrics* Vol. 19: 209-219.
- Raftery, A.E. (1989). 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* Vol. 4: 378-381.
- Raftery, A.E. (1996). Hypothesis testing and model selection, in W. Gilks, S. Richardson & D.J. Spiegelhalter (eds.), *Markov chain Monte Carlo in practice*, Chapman and Hall, UK, pp. 163-187.
- Ramírez-Cid, J.E. & Achcar, J.A. (1999). Bayesian inference for nonhomogeneous Poisson processes in software reliability models assuming non monotonic intensity functions, *Computational Statistics and Data Analysis* Vol. 32: 147-159.
- Roberts, E.M. (1979a). Review of statistics of extreme values with applications to air quality data. Part I. Review, *Journal of the Air Pollution Control Association* Vol. 29: 632-637.
- Roberts, E.M. (1979b). Review of statistics of extreme values with applications to air quality data. Part II. Applications, *Journal of the Air Pollution Control Association* Vol. 29: 733-740.
- Smith, R.L. (1989). Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone, *Statistical Sciences* Vol. 4: 367-393.
- Wilson, S.P. & Costello, M.J. (2005). Predicting future discoveries of European marine species using non-homogeneous renewal processes, *Journal of the Royal Statistical Society Series C* Vol. 54: 425-442.



- WHO (World Health Organization) (2006). *Air Quality Guidelines - 2005. Particulate matter, ozone, nitrogen dioxide and sulfur dioxide*, World Health Organization Regional Office for Europe, EU.
- Zavala, M., Herndon, S.C., Wood, E.C., Onasch, T.B., Knighton, W.B., Marr, L.C., Kolb, C.E. & Molina, L.T. (2009). Evaluation of mobile emissions contributions to Mexico City's emissions inventory using on road and cross-road emission measurements and ambient data, *Atmos. Chem. Phys.* Vol. 9: 6305-6317.



## **Air Quality-Models and Applications**

Edited by Prof. Dragana Popovic

ISBN 978-953-307-307-1

Hard cover, 364 pages

**Publisher** InTech

**Published online** 09, June, 2011

**Published in print edition** June, 2011

Air pollution has been a major transboundary problem and a matter of global concern for decades. High concentrations of different air pollutants are particularly harmful to large cities residents, where numerous anthropogenic activities strongly influence the quality of air. Although there are many books on the subject, the one in front of you will hopefully fulfill some of the gaps in the area of air quality monitoring and modeling, and be of help to graduate students, professionals and researchers. The book is divided in five sections, dealing with mathematical models and computing techniques used in air pollution monitoring and forecasting; air pollution models and application; measuring methodologies in air pollution monitoring and control; experimental data on urban air pollution in China, Egypt, Northeastern U.S, Brazil and Romania; and finally, the health effects due to exposure to benzene, and on the influence of air pollutants on the acute respiratory diseases in children in Mexico.

### **How to reference**

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Eliane R. Rodrigues, Jorge A. Achcar and Julian Jara-Ettinger (2011). A Gibbs Sampling Algorithm to Estimate the Occurrence of Ozone Exceedances in Mexico City, *Air Quality-Models and Applications*, Prof. Dragana Popovic (Ed.), ISBN: 978-953-307-307-1, InTech, Available from: <http://www.intechopen.com/books/air-quality-models-and-applications/a-gibbs-sampling-algorithm-to-estimate-the-occurrence-of-ozone-exceedances-in-mexico-city1>

**INTECH**  
open science | open minds

### **InTech Europe**

University Campus STeP Ri  
Slavka Krautzeka 83/A  
51000 Rijeka, Croatia  
Phone: +385 (51) 770 447  
Fax: +385 (51) 686 166  
[www.intechopen.com](http://www.intechopen.com)

### **InTech China**

Unit 405, Office Block, Hotel Equatorial Shanghai  
No.65, Yan An Road (West), Shanghai, 200040, China  
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元  
Phone: +86-21-62489820  
Fax: +86-21-62489821

© 2011 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](https://creativecommons.org/licenses/by-nc-sa/3.0/), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen