Brazilian Journal of Probability and Statistics
2011, Vol. 25, No. 2, 183–204
DOI: 10.1214/10-BJPS116

Modelling interoccurrence times between ozone peaks in Mexico City in the presence of multiple change points

Jorge Alberto Achcar[a], Eliane R. Rodrigues[b] and Guadalupe Tzintzun[c]

[a]*Universidade de São Paulo*
[b]*Universidad Nacional Autónoma de México*
[c]*Secretaría de Medio Ambiente y Recusos Naturales*

Abstract. In this article we consider the problem of analysing the interoccurrence times between ozone peaks. These interoccurrence times are assumed to have an exponential distribution with some rate $\lambda > 0$ (which may have different values for different interoccurrence times). We consider four parametric forms for λ. These parametric forms depend on some parameters that will be estimated by using Bayesian inference through Markov Chain Monte Carlo (MCMC) methods. In particular, we use a Gibbs sampling algorithm internally implemented in the software WinBugs. We also present an analysis to detect the possible presence of change points. This is performed using the 95% credible interval of the difference between two consecutive means. Results are applied to the maximum daily ozone measurements provided by the monitoring network of Mexico City. An analysis in terms of the number of possible change points present in the model in terms of different years and seasons of the year is also presented.

1 Introduction

It is a well-known fact that individuals exposed for a long period of time to high levels of ozone may experience serious health problems (see, e.g., Bell et al., 2004; Bell et al., 2005; Gauderman et al., 2004; Itô et al., 2005; Loomis et al., 1996; O'Neill et al., 2004; and for a review of the subject see, e.g., ARB, 2005; Itô et al., 2005; Seinfeld, 2004 and references therein). Hence, due to its impact on the population health, to understand the behaviour of pollutants in general and in particular of ozone is a very important issue.

There are several works in the literature analysing the problem of predicting pollution emergency episodes. Among them we may quote Álvarez et al. (2005), Austin and Tran (1999), Flaum et al. (1996), Guardani et al. (1999, 2003), Horowitz (1980), Huerta and Sansó (2005), Lanfredi and Macchiato (1997), Larsen et al. (1990), Leadbetter (1991), Pan and Chen (2008), Roberts (1979a, 1979b), Smith (1989), Zolghadri and Henry (2004). See Piegorsch et al. (1998)

Key words and phrases. Poisson models, multiple change points, Bayesian inference, ozone peaks, interoccurrence times.

Received September 2009; accepted January 2010.

for a review of some statistical methodologies commonly used in the study of environmental problems and see also Seinfeld (2004) and Itô et al. (2005). When the aim is to estimate the number of times that a given environmental standard is violated, we could use Poisson processes (see, e.g., Achcar et al., 2008, 2009a, 2009b, 2009c; Javits, 1980; Raftery, 1989). In this paper, instead of modelling the number of times that a given environmental standard is violated, we consider the problem of modelling the interoccurrence times between two ozone peaks.

The modelling of interoccurrence times between two consecutive violations of an environmental standard is not very common. However, the approach is very popular in the topic of software reliability (see, e.g., Jelinski and Moranda, 1972). In the case of environmental problems, the study of interoccurrence times instead of the number of occurrences of violation, has the advantage of avoiding the problem of convergence of the Gibbs sampling algorithm within the Bayesian framework in the presence of multiple change points (see Achcar et al., 2009a).

Remark. Even though the Mexican standard for ozone is 0.11 parts per million (0.11 ppm) (see NOM, 2002) and 0.22 ppm is the threshold used in Mexico City for declaring an emergency situation, we are going to consider the value 0.17 ppm as the threshold to indicate that an ozone peak has occurred. The main reason being that the threshold 0.11 ppm is surpassed quite frequently and the threshold 0.22 ppm is rarely surpassed. Hence, 0.17 ppm is an intermediate value between these two values. An additional reason is that the environmental authorities in Mexico City have been thinking of reducing the value of the threshold used to declare emergency situations. The value 0.17 ppm could be an alternative and the reduction would not be so drastic as to 0.11 ppm in which case the city might have to stop every day. Any other value could be used, specially if the interest is either in obtaining information about the behaviour of the pollutant when a certain threshold is considered or in seeing how this behaviour could be affected by government decisions.

We would like to call attention to the fact that several measures have been taken by the environmental authorities in order to improve the air quality in large cities in Mexico and in particular Mexico City. In addition to an inspection programme for vehicles, in 1990 a driving restriction for cars was introduced. There were also a restriction on 20% of public transportation vehicles that were allowed to circulate during working days. In 1997, 1999 and 2000 further measures were taken. Additionally, in 2001, restrictions were imposed on diesel vehicles. There were also a renewal of the taxis and bus fleet and 300 of the most polluting industries were regulated.

Therefore, it would be interesting to know if these regulations have produced any changes in the behaviour of ozone levels in Mexico City. In here we are going to analyse this under the point of view of mean waiting times between two consecutive violation of an ozone environmental standard. The aim is to provide a way of knowing the possible changes in terms of the lengths of these mean waiting times. We would like to know if it is possible to detect changes around the period when environmental measures were introduced. We would also like to know how many possible changes in the behaviour of the interoccurrence times (i.e., possible presence of change points) there exist in terms of different years and seasons of the year.

This paper is organised as follows. In Section 2 the parametric models used to analyse the problem considered here are presented. A Bayesian formulation of the problem is given in Section 3. In Section 4 the method for selecting the best model to explain the behaviour of the data is presented as well as a way of detecting the possible presence of change points. An application to ozone data provided by the monitoring network of Mexico City is given in Section 5. In Section 6 we present a discussion of the results. Finally, in the Appendix we give the WinBugs codes used to obtain the samples for estimating the parameters of the models.

2 Description of the statistical models

In this section the notation used throughout this work as well as the models considered to study the problem are presented. We start by setting some of the notation. Let $N \geq 1$ be an integer which will represent the number of interoccurrence times between two ozone peaks during the entire observational period. Denote by T_i, $i = 1, 2, \dots, N$, these interviolation times and consider that they are independent. We assume that T_i has an exponential distribution with parameter $\lambda_i > 0$, that is, for $i = 1, 2, \dots, N$, T_i has density

$$f(t_i|\lambda_i) = \lambda_i e^{-\lambda_i t_i}, \qquad t_i \geq 0. \tag{2.1}$$

(Therefore, T_i has mean $1/\lambda_i$ and variance $1/\lambda_i^2$, $i = 1, 2, \dots, N$.) In here we are going to consider four parametric forms for the parameter λ_i, $i = 1, 2, \dots, N$.

Model I. In this model we consider the following form for λ_i, $i = 1, 2, \dots, N$. Set $t_0 = 0$ and let W_i be a latent variable (nonobservable) associated to each interviolation time, $i = 1, 2, \dots, N$. Define λ_i, $i = 1, 2, \dots, N$, by

$$\lambda_i = \alpha(M - i + 1) + W_i + \phi t_{i-1}, \tag{2.2}$$

where $M \geq N$, $0 < \phi < 1$ and $\alpha \geq 0$ are unknown quantities to be estimated. We assume that W_i, $i = 1, 2, \dots, N$, are random quantities with a Gamma distribution Gamma(a, b). (In here, Gamma(a, b) is the Gamma distribution with mean a/b and variance a/b^2.)

Remarks. 1. The term ϕt_{i-1} in (2.2) gives an autoregressive contribution of the previous time in the rate of the ozone exceedances. Observe that if W_i and ϕ are both zero, the term $\alpha(M - i + 1)$ in (2.2) gives a decreasing rate λ_i, that is, the mean times between the ozone exceedances are increasing as time progresses.

2. It is important to point out that the model $\lambda_i = \alpha(M - i + 1)$ has been explored in software reliability theory (Jelinski and Moranda, 1972; Moranda, 1975), where M is an unknown parameter denoting the number of bugs in a software during a debugging period, where each encountered bug is eliminated by correction of the software. In the environmental applications, M could be tentatively interpreted as an overall number (unknown) of ozone violation since the beginning of the observational period.

Model II. In this version of the model we consider the same parametric form for λ_i given in Model I. However, instead of the latent variables W_i, $i = 1, 2, \ldots, N$, having a Gamma distribution, they will have an uniform distribution in an interval $[0, K]$, that is, W_i has distribution $\mathrm{U}(0, K)$, $i = 1, 2, \ldots, N$.

Remark. The choice of an uniform distribution on the interval $(0, K)$ for the latent variable W_i was considered in order to have a better convergence of the Gibbs sampling algorithm implemented in the software WinBugs. The value will be selected in order to reflect the lack of knowledge of the behaviour of the latent variable.

Model III. This model is a particular case of Model I. In here the dependence on the latent variables W_i, $i = 1, 2, \ldots, N$, is eliminated. Hence, for $i = 1, 2, \ldots, N$, the rate λ_i is given by

$$\lambda_i = \alpha(M - i + 1) + \phi t_{i-1}. \tag{2.3}$$

Model IV. This model also is a particular case of Model I. Although we consider the presence of the latent variables W_i, $i = 1, 2, \ldots, N$, the term ϕt_{i-1} is not taken into account. Hence, for $i = 1, 2, \ldots, N$, the rate λ_i has the following form:

$$\lambda_i = \alpha(M - i + 1) + W_i. \tag{2.4}$$

In here, the latent variables W_i, $i = 1, 2, \ldots, N$, are assumed to have an uniform distribution in the interval $[0, K]$ as in Model II with a possibly different value of K.

Remark. The parameters a and b of the Gamma distribution and the parameter K of the uniform distribution, are considered to be known and will be specified later.

The set of observed data is indicated by $\mathbf{D} = \{N; T_1, T_2, \ldots, T_N\}$ and the parameters to be estimated are the ones related to the parametric forms of λ_i, $i = 1, 2, \ldots, N$. Hence, the vector of parameters are $\boldsymbol{\theta}_I = \boldsymbol{\theta}_{II} = (\alpha, M, \phi)$ when either Model I or II is used, $\boldsymbol{\theta}_{III} = (\alpha, \phi)$ when Model III is considered, and is $\boldsymbol{\theta}_{IV} = (\alpha, M)$ when Model IV is taken into account.

3 A Bayesian formulation of the models

In this section we describe the Bayesian formulation of the problem considered here. Even though each model will be analysed separately, the same prior distributions for the parameters are used in all models and for all regions and we assume prior independence among them. We also assume that α, M and ϕ have a $\mathrm{U}(0, a_1)$, $\mathrm{U}(N, b_1)$ and a Beta $\mathrm{B}(e_1, e_2)$ prior distributions, respectively. (In here, $\mathrm{Beta}(a, b)$ denotes a Beta distribution with mean $a/(a+b)$ and variance $ab/[(a+b)(a+b+1)]$.) The hyperparameters a_1, b_1, e_1 and e_2 are considered to be known and will be specified later.

Remark. The choice of the hyperparameters for the prior distributions will be made in a way to have approximately noninformative prior distributions (see, e.g., Bernardo and Smith, 1994).

Since we are assuming that the interviolation times, have density given by (2.1) we have that the general form of the likelihood function is given by

$$L(D|\boldsymbol{\theta}) = \prod_{i=1}^{N} \lambda_i e^{-\lambda_i t_i}, \tag{3.1}$$

where $\boldsymbol{\theta} = \boldsymbol{\theta}_I, \boldsymbol{\theta}_{II}, \boldsymbol{\theta}_{III}, \boldsymbol{\theta}_{IV}$ is the vector of parameters associated to each chosen parametric form of λ_i, $i = 1, 2, \dots, N$.

Model I. When Model I is considered, we have that the likelihood function has the following form:

$$\begin{aligned} L(\mathbf{D}|\boldsymbol{\theta}_I) = A(\boldsymbol{\theta}_I) \exp\Bigg(-\alpha M \sum_{i=1}^{N} t_i + \alpha \sum_{i=1}^{N} i t_i - \alpha \sum_{i=1}^{N} t_i \\ - \sum_{i=1}^{N} W_i t_i - \phi \sum_{i=1}^{N} t_i t_{i-1} \Bigg), \end{aligned} \tag{3.2}$$

where

$$A(\boldsymbol{\theta}_I) = \prod_{i=1}^{N} [\alpha(M - i + 1) + W_i + \phi t_{i-1}]. \tag{3.3}$$

Hence, the joint posterior distribution of $\boldsymbol{\theta}_I$ and $\mathbf{W} = (W_1, W_2, \dots, W_N)$ is given by

$$P(\boldsymbol{\theta}_I, \mathbf{W}|\mathbf{D}) \propto \phi^{e_1-1}(1-\phi)^{e_2-1} \left(\prod_{i=1}^{N} W_i^{a-1} e^{-bW_i} \right) A(\boldsymbol{\theta}_I) \tag{3.4}$$

$$\times \exp\Bigg(-\alpha M \sum_{i=1}^{N} t_i + \alpha \sum_{i=1}^{N} i t_i - \alpha \sum_{i=1}^{N} t_i - \sum_{i=1}^{N} W_i t_i - \phi \sum_{i=1}^{N} t_i t_{i-1}\Bigg),$$

where $\phi \in (0, 1)$, $0 \le \alpha \le a_1$ and $N \le M \le b_1$.

Model II. If we consider Model II, then the likelihood function of the model is (3.2) but now using $\boldsymbol{\theta}_{II}$ instead of $\boldsymbol{\theta}_I$. Hence, the joint posterior distribution of $\boldsymbol{\theta}_{II}$ and $\mathbf{W}$ is given by

$$P(\boldsymbol{\theta}_{II}, \mathbf{W}|\mathbf{D}) \propto \phi^{e_1-1}(1-\phi)^{e_2-1} A(\boldsymbol{\theta}_{II}) \times \exp\Bigg(-\alpha M \sum_{i=1}^{N} t_i + \alpha \sum_{i=1}^{N} i t_i - \alpha \sum_{i=1}^{N} t_i - \sum_{i=1}^{N} W_i t_i - \phi \sum_{i=1}^{N} t_i t_{i-1}\Bigg), \tag{3.5}$$

where $\phi \in (0, 1)$, $0 \le \alpha \le a_1$ and $N \le M \le b_1$, with $A(\boldsymbol{\theta}_{II})$ given by (3.3) using $\boldsymbol{\theta}_{II}$ instead of $\boldsymbol{\theta}_I$.

Model III. The likelihood function when Model III is used is obtained by setting $W_i = 0$, $i = 1, 2, \dots, N$, in (3.2) and in (3.3) and using $\boldsymbol{\theta}_{III}$ instead of $\boldsymbol{\theta}_I$. In order to obtain the expression for the posterior distribution $P(\boldsymbol{\theta}_{III}|\mathbf{D})$, we just exclude from (3.5) the terms involving W_i, $i = 1, 2, \dots, N$, and use $\boldsymbol{\theta}_{III}$ instead of $\boldsymbol{\theta}_{II}$.

Model IV. When Model IV is considered, then the likelihood function is obtained by setting $\phi = 0$ in (3.2) and in (3.3) and using $\boldsymbol{\theta}_{IV}$ instead of $\boldsymbol{\theta}_I$. The joint posterior distribution of $\boldsymbol{\theta}_{IV}$ and $\mathbf{W}$ given the data is obtained from (3.5) by excluding the terms involving ϕ and by using $\boldsymbol{\theta}_{IV}$ instead of $\boldsymbol{\theta}_{II}$.

Posterior summaries of interest are obtained from simulated samples from the respective joint posterior distributions by using standard Markov Chain Monte Carlo (MCMC) methods as the Gibbs sampling algorithm (see, e.g., Gelfand and Smith, 1990) or the Metropolis–Hastings algorithm (see, e.g., Smith and Roberts, 1993). A great simplification in the generation of the samples is given by using the WinBugs software (Spiegelhalter et al., 1999) where we only need to specify the distribution of the data and the prior distributions of the parameters of the model.

4 Model selection

The selection of the best model to explain the behaviour of the ozone data from the monitoring network of Mexico City is made by using standard existing model discrimination methods. In here, we use the sum of the absolute values for the differences between the Bayesian estimates of $\mathrm{E}(T_i|\mathbf{D})$, based on a sample generated by the Gibbs sampling algorithm, and the observed interoccurrence times t_i, $i = 1, 2, \ldots, N$, which is given by

$$c(l) = \sum_{i=1}^{N} \left|\hat{T}_i^{(l)} - t_i\right|, \tag{4.1}$$

where $\hat{T}_i^{(l)}$ is the Monte Carlo estimate of the posterior mean $\mathrm{E}(T_i|\mathbf{D})$ when Model l is used, $l = I, II, III, IV$.

Considering the model that best fit the data, we can obtain accurate inference results for the rates λ_i, $i = 1, 2, \ldots, N$. We may also construct credible intervals for the differences between the means of two consecutive interoccurrence times, $\Delta_{(i)} = 1/\lambda_i - 1/\lambda_{i-1}$, $i = 2, 3, \ldots, N$, to detect multiple change points. Observe that if zero is not included in a specified Bayesian credible intervals for $\Delta_{(i)}$, there is an indication that the means $1/\lambda_i$ and $1/\lambda_{i-1}$ are different, and therefore, there is an indication of the presence of a change point. These inference results are of great practical interest in the control and analysis of pollution data. That is so because if a change is detected by a model then one may try to find out what may have caused the change (e.g., government measures or an environmental factor).

5 Analysis of ozone peaks in Mexico City

In this section we apply the models described in Section 2 to analyse the data provided by the monitoring network of Mexico City. Since the behaviour of ozone varies from one region of the city to another, the environmental authorities of Mexico City have decided to split the Metropolitan Area 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., 2008). Hence, when the environmental threshold of interest is surpassed in one or more regions, an emergency is declared and measures are taken to bring the ozone level down only in those regions. In this paper, we have considered the same spatial division used by the environmental authorities of Mexico City.

Remark. We would like to call attention to the fact that this spatial division may not be the best, but it is the one currently used by the environmental authorities. There is a study where a different spatial division is being considered and tested. However, that is an ongoing research.

We are also considering measurements used by the environmental authorities to declare an emergency. The data used is described as follows. The primary data used in the analysis corresponds to 18 years (from 1 January 1990 to 31 December

2007) of the daily maximum ozone 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 that region. This maximum is taken independently of the number of active monitoring station in each region. There are 3, 4, 4, 4 and 5 fixed monitoring stations in regions NE, NW, CE, SE and SW, respectively (see, e.g., Álvarez et al., 2005 and Achcar et al., 2008). The 18-year average measurements in those regions are 0.98, 0.131, 0.135, 0.153 and 0.128, respectively, with standard deviation 0.041, 0.057, 0.056, 0.062 and 0.048.

Remark. Note that measurements in different monitoring stations in a given region may not be similar. However, for the purpose of reporting the air quality only the maximum among the measurements in the region is considered and those measurements are the ones used in the present study. Nevertheless, for other specific questions, an analysis station by station may be performed.

The data that actually is used in the analysis corresponds to the length of time between ozone peaks, that is, the length of time between days in which the maximum ozone measurement surpassed the threshold 0.17 ppm. The threshold 0.17 ppm was surpassed in 316, 1589, 1691, 1201 and 2394 days in regions NE, NW, CE, SE and SW, respectively. Hence, we have that N has a different value for each region.

In Figure 1, we have the plots of the ordered interviolation times (i.e., the time between two consecutive ozone peaks) for all regions during the period of time considered here.

Observing Figure 1, we may notice that for all regions the interoccurrence times have increased in the last years of the observational period. We also observe the possible presence of multiple change points for each region. The existence of these change points is justified by the alternation between periods of large and small interoccurrence times as is observed in Figure 1.

In region NE, we observe the presence of larger interoccurrence times interchanged with smaller interviolation times since the beginning of the observational period not only at the end of it. Taking into account the other regions, we observe a more standard pattern with smaller interoccurrence times during the period ranging from 1 January 1990 until approximately the years 2001–2002. That fact shows that before approximately the year 2002, the occurrence of ozone peaks were more frequent than after that year.

The sample used to estimate the parameters of the models were obtained after a burn-in period of 1000 steps to eliminate the dependence on the initial values in the Gibbs sampling algorithm. A final Gibbs sample of size 1000 was obtained by taking every 10th generated value in order to have approximately uncorrelated samples.

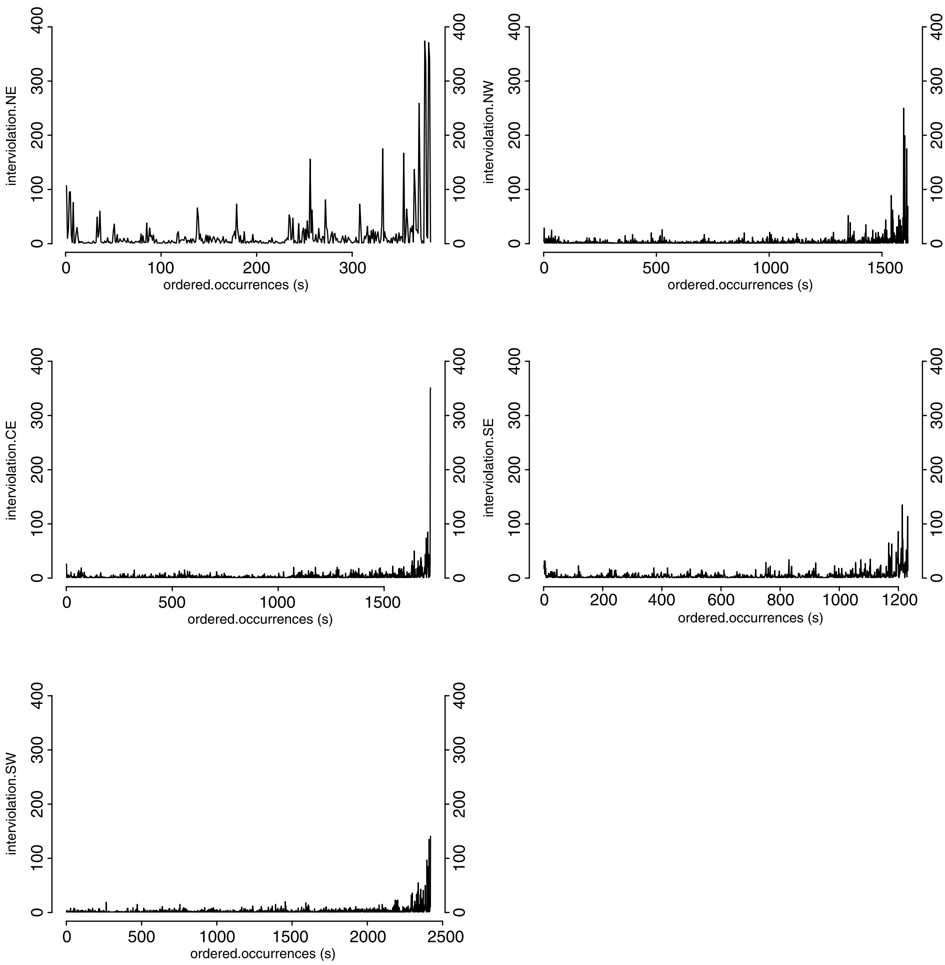

Figure 1 *Observed interviolation ozone times versus ordered occurrences for regions NE, NW, CE, SE and SW.*

In all models and regions, we assume that the hyperparameters of the prior distributions have the following values, $a_1 = 1$, $b_1 = 10{,}000$ and $e_1 = e_2 = 1$. We also assume for all cases that $a = b = 1$ and that $K = 100$.

In Table 1 we present the values of $c(l)$, $l = I, II, III, IV$, for all regions.

Observing Table 1 we may notice that for all regions the model with smallest absolute value for the differences of estimated and observed interviolation times is Model IV. Hence, this is an indication that Model IV is the one that best explain the behaviour of the data provided by the monitoring network of the Metropolitan Area of Mexico City. Hence, we are going to report only the results given by that model.

Table 1 *Values of $c(l)$, $l = I, II, III$, for regions NE, NW, CE, SE and SW*

	$c(I)$	$c(II)$	$c(III)$	$c(IV)$
NE	5868.02	980.32	689.14	587.23
NW	4862.44	1837.89	477.16	371.68
CE	4434.36	1747.25	382.21	296.88
SE	4845.51	1552.89	461.91	248.61
SW	4087.23	1951.09	355.73	292.29

Table 2 *Posterior mean and standard deviation (in parentheses) of the parameters M and α for Model IV for regions NE, NW, CE, SE and SW*

	M	α
NE	2186 (2243)	0.00000268 (0.000004193)
NW	3703 (2182)	0.00000212 (0.00000266)
CE	4055 (2139)	0.00000209 (0.000002506)
SE	3597 (2347)	0.00000234 (0.00000312)
SW	5675 (2049)	0.00000224 (0.000002442)

Table 2 presents the posterior mean and standard deviation (in parentheses) of the quantities of interest when Model IV is used.

In Figure 2 we have the plots of the estimated means $1/\lambda_i$, $i = 1, 2, \ldots, N$, of the interviolation times considering Model IV. Comparing Figures 1 and 2 we may see that Model IV fits well the observed interoccurrence times for all regions.

Since Model IV is the one that best fits the data recording the interviolation times for all regions of Mexico City, we also consider only this model to obtain inference for the possibly existing change points. In Figure 3, we have the plots for the Monte Carlo estimates of the posterior means for the differences $\Delta_{(i)} = 1/\lambda_i - 1/\lambda_{i-1}$, $i = 2, 3, \ldots, N$, for the five regions of Mexico City.

Observing Figure 3, we note that towards the end of the data set the difference between two consecutive mean waiting times tends to be larger. It is worthwhile to call attention to the fact that there is an oscillation between positive and negative differences. Positive differences implicate that larger mean waiting times are preceded by smaller ones. Negative differences implicate that smaller waiting times are preceded by larger ones. This exchange in length are reflected in the plots of Figure 3. Even though we have differences between two consecutive mean times in the beginning of the measurements, they tend to be more homogeneous. We also observe the presence of multiple change points for all regions (reflected by the fact that periods of time when smaller differences occur are placed in between periods of time when large ones occur, we may see that more clearly in the plots for region NE).

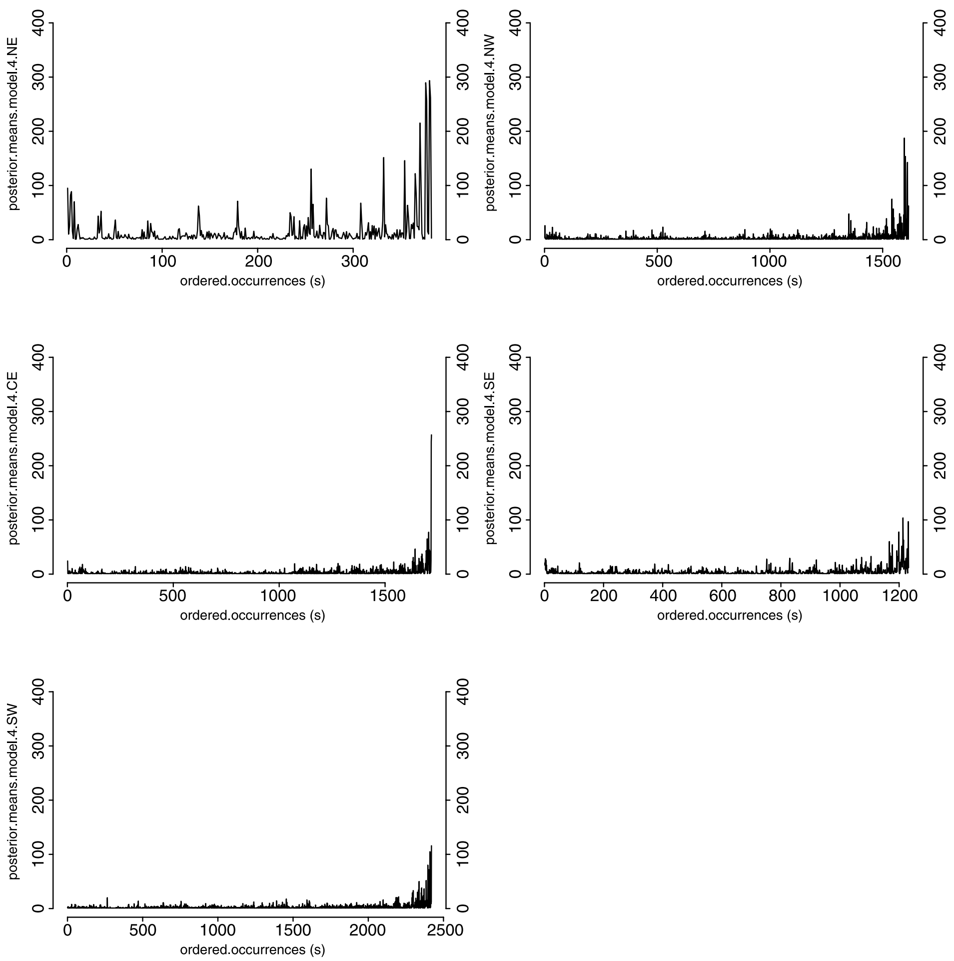

Figure 2 *Estimated mean interviolation times versus ordered occurrences for regions NE, NW, CE, SE and SW.*

Table 3 presents the 95% credible intervals for the difference between two consecutive means for region SW that do not have zero in them. We have decided to present the table only for this region as an illustration and because this is the one with more severe ozone problem. In Table 3, the "order j" means the index of the jth ozone peak.

Note from Table 3 that there are 69 possible change points for region SW. Also note that the first 265 interoccurrence times were ruled by the same exponential distribution. After that possible change point, we have that about 200 interoccurrence times were ruled by an exponential distribution with a different rate λ, and so on.

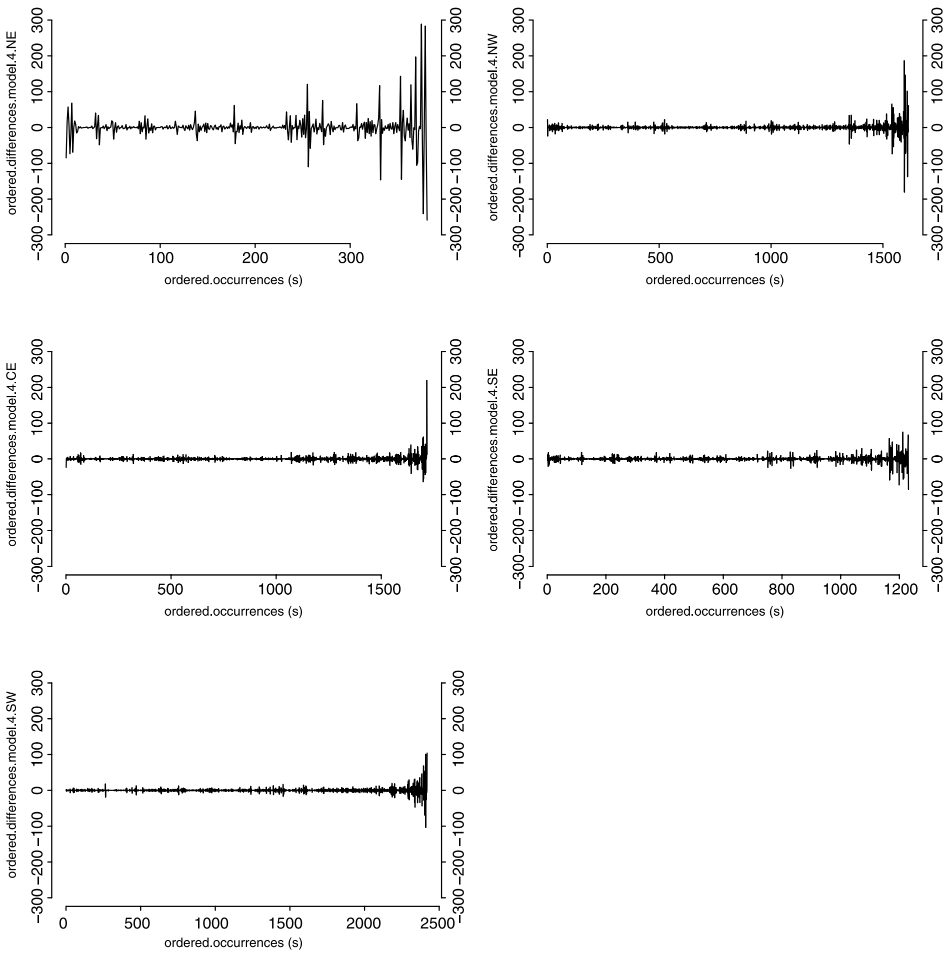

Figure 3 *Estimated posterior means of the differences of means versus ordered occurrences for regions NE, NW, CE, SE and SW.*

In Table 4 we have the number of possible change points (negative and positive) for each region and year during the observational period.

Observing Table 4 we have that there are 66, 87, 68, 83 and 69 possible change points for regions NE, NW, CE, SE and SW, respectively. It is possible to observe that many of the change points occur in the period prior to 2001 for all regions. We also have that for regions NW and SW the number of possible change points stay more or less with the same behaviour until 2004 and 2007, respectively. We have a more homogeneous distribution of the number of possible change points in the time interval from 1996 until 2000. Also, from Table 4 we may notice that from 2003 there is a stabilisation in the behaviour of the ozone in regions NE, CE and SE and hence a decrease in the number of possible change points. This behaviour

Table 3 *95% credible intervals for the ordered differences of means not including zero for region SW*

i	Order	95% cr. interv.	Day/Month/Year	*i*	Order	95% cr. interv.	Day/Month/Year
1	266	(−75.940; −2.365)	09/04/1991 (Tue)	36	2185	(3.096; 79.880)	31/10/2001 (Wed)
2	443	(0.1046; 32.640)	22/11/1991 (Fri)	37	2186	(−76.480; −1.760)	23/11/2001 (Fri)
3	444	(−32.140; −0.04672)	01/12/1991 (Sun)	38	2199	(−48.970; −1.196)	14/02/2002 (Thu)
4	471	(0.7814; 59.670)	21/01/1992 (Tue)	39	2203	(−90.970; −3.004)	14/03/2002 (Thu)
5	472	(−57.350; −1.035)	05/02/1992 (Wed)	40	2207	(−41.440; −0.3738)	31/03/2002 (Sun)
6	639	(−43.300; −0.009635)	09/11/1992 (Mon)	41	2212	(0.2686; 34.480)	05/04/2002 (Fri)
7	755	(1.646; 55.200)	24/06/1993 (Thu)	42	2237	(0.2994; 38.270)	22/05/2002 (Wed)
8	756	(−54.540; −1.190)	09/07/1993 (Fri)	43	2244	(0.2374; 48.760)	14/06/2002 (Fri)
9	779	(−32.570; −0.3690)	25/08/1993 (Wed)	44	2261	(0.1596; 49.500)	30/07/2002 (Tue)
10	1166	(0.02847; 35.350)	05/07/1995 (Wed)	45	2271	(0.5203; 59.390)	17/09/2002 (Tue)
11	1239	(0.2316; 59.130)	15/12/1995 (Fri)	46	2272	(−59.170; −0.3415)	29/09/2002 (Sun)
12	1240	(−55.180; −0.4211)	27/12/1995 (Wed)	47	2332	(0.4096; 48.280)	26/11/2003 (Wed)
13	1295	(0.4864; 38.470)	21/03/1996 (Thu)	48	2336	(−138.500; −5.922)	08/01/2004 (Thu)
14	1296	(−38.100; −0.7308)	01/04/1996 (Mon)	49	2337	(1.823; 81.480)	09/01/2004 (Fri)
15	1343	(0.5332; 41.310)	21/06/1996 (Fri)	50	2339	(−198.900; −6.237)	20/04/2004 (Tue)
16	1369	(−46.860; −0.815)	27/08/1996 (Tue)	51	2343	(0.0382; 45.960)	28/04/2004 (Wed)
17	1390	(1.097; 50.710)	27/09/1996 (Fri)	52	2344	(−45.550; −0.5998)	10/05/2004 (Mon)
18	1429	(0.4857; 47.240)	04/01/1997 (Sat)	53	2349	(−57.39; −1.151)	10/06/2004 (Thu)
19	1430	(−45.720; −0.5675)	15/01/1997 (Wed)	54	2356	(−141.900; −0.6993)	15/10/2004 (Fri)
20	1455	(2.067; 68.070)	12/03/1997 (Wed)	55	2358	(1.705; 97.07)	20/10/2004 (Wed)
21	1456	(−63.960; −2.163)	01/04/1997 (Tue)	56	2366	(3.124; 109.600)	06/12/2004 (Mon)
22	1460	(−43.180; −0.8046)	28/04/1997 (Mon)	57	2370	(−84.090; −2.979)	19/02/2005 (Sat)
23	1592	(1.887; 59.000)	16/12/1997 (Tue)	58	2371	(6.046; 144.100)	20/02/2005 (Sun)
24	1597	(−39.680; −0.3755)	15/01/1998 (Thu)	59	2372	(−141.900; −3.368)	02/04/2005 (Sun)
25	1598	(0.4082; 36.830)	25/01/1998 (Sun)	60	2400	(−49.510; −0.5335)	04/05/2006 (Thu)
26	1611	(0.8759; 41.460)	12/03/1998 (Thu)	61	2405	(−269.500; −13.060)	08/09/2006 (Fri)
27	1612	(−41.450; −0.6997)	24/03/1998 (Tue)	62	2409	(−149; −0.04385)	26/11/2006 (Sun)
28	1728	(0.2519; 39.790)	17/10/1998 (Sat)	63	2410	(18.600; 362.600)	30/11/2006 (Thu)
29	1729	(−39.690; −0.2366)	27/10/1998 (Tue)	64	2411	(−352; −22.060)	14/04/2007 (Sat)
30	1848	(0.2779; 38.510)	09/06/1999 (Wed)	65	2413	(0.3958; 106.500)	18/04/2007 (Wed)
31	1923	(0.3626; 49.900)	17/12/1999 (Fri)	66	2414	(−107.300; −1.864)	17/05/2007 (Thu)
32	2078	(0.7915; 51.440)	20/12/2000 (Wed)	67	2415	(0.07568; 75.320)	19/05/2007 (Sat)
33	2100	(−82.160; −1.207)	03/03/2001 (Sat)	68	2417	(−43.350; 24.660)	23/06/2007 (Sat)
34	2169	(0.2854; 40.050)	19/08/2001 (Sun)	69	2420	(4.383; 382.600)	27/07/2007 (Fri)
35	2181	(−48.340; −1.372)	21/10/2001 (Sun)	–	–	–	–

is not observed in regions NW and SW. We would like to point out that the last change point that appears in regions NE, NW and SE were points representing a negative change. The last ones in region CE and SW were points representing a positive change.

Remark. Another way of visualising the behaviour of the number of positive and negative change points is through a plot of the number of change points versus years. This representation is given in Figure 4. In that figure we have that the information regarding regions NE, NW, CE, SE and SW is given by the plots represented by the thick grey, thin black, dashed, thin grey and thick black lines.

Table 4 *Number of possible change points for each region and each year during the observational period*

	NE		NW		CE		SE		SW	
	Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg
1990	1	1	3	2	3	3	2	3	–	–
1991	3	3	1	–	1	–	4	1	1	2
1992	4	5	1	1	2	1	4	5	1	2
1993	2	3	4	4	2	2	1	1	1	2
1994	3	3	1	–	1	1	3	4	–	–
1995	1	2	2	3	1	–	2	2	2	1
1996	3	2	4	3	3	2	2	2	3	2
1997	2	3	1	2	5	4	3	3	3	3
1998	4	3	2	1	6	3	6	3	3	3
1999	1	2	3	2	2	1	4	2	2	–
2000	5	1	4	1	–	2	2	6	1	–
2001	2	2	5	8	4	5	5	4	2	3
2002	1	–	5	4	4	1	1	2	5	4
2003	1	–	5	3	1	1	1	1	1	–
2004	–	–	5	4	1	2	1	1	4	5
2005	1	1	1	1	2	2	1	1	1	2
2006	–	–	1	–	–	–	–	–	1	3
2007	1	–	–	–	–	–	–	–	3	3
Total	35	31	48	39	38	30	42	41	34	35

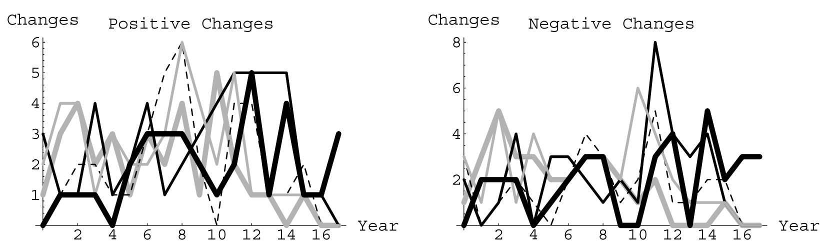

Figure 4 *Number of positive and negative change points versus years for regions NE, NW, CE, SE and SW.*

The numbers in the “Year” axis mean the first, second and so on years where measurements were taken. In the “Changes” axis we have the number of change points in for each year.

In Table 5 we have the distribution according to the seasons of the year of the possible change points for each region and separated by type of change that they might represent.

Table 5 *Number of possible change points for each region and each season of the year during the observational period*

	NE		NW		CE		SE		SW	
	Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg
Winter	15	11	8	9	12	7	14	14	10	8
Spring	12	59	15	8	10	8	9	6	8	14
Summer	8	4	14	14	9	7	11	11	8	5
Autum	0	57	11	8	7	8	8	10	8	8

Observing Table 5 we have that most of the change points related to region NE occur during winter and spring. The number of change points representing positive and negative changes do not differ much for that region. If we consider region NW we may see that most of the change points occur during spring and summer (cases indicating positive changes). If we consider the change points indicating a decrease of the mean of the waiting time between two peaks, we have that the largest number of such change points occur during the summer. When considering the remaining regions we have that in region CE the largest number of change points indicating a positive change occur during winter and spring. The number of change points indicating negative changes are more or less equally distributed among the different seasons. In region SE the change points (either indicating positive or negative changes) occur mostly during winter and summer. Regarding region SW, we have that the largest number of positive and negative changes occur during winter and spring, respectively.

6 Discussion and remarks

Looking at Figure 2, we may notice that for all regions of Mexico City, there was a consistent increase in the interviolation times. We may observe from Figure 3 that the difference between consecutive peaks also increases towards the end of the observational period. Note that even though there are negative changes they are always followed by a period of positive changes. In particular, in region CE we may notice that towards the end of the dataset the periods representing a increase in the mean waiting time, have in general, larger length than the periods representing a negative change.

If we look at Table 4 (or Figure 4), we may see that for region SE there are more positive change points in the years 1991, 1992, 1998, 1999 and 2001. We may also observe a decreasing number of change points from 2002 to 2005 and in particular there are no change points from the year 2006. Taking into account region CE, we may see from Table 4 that there is an increase in the number of change points representing a positive change in the years 1997, 1998, 2001 and

2002. In region NE similar behaviour occurs in the years 1998 and 2000. In region NW, the increase of positive changes occurs in the years 1993, 1996, 2001 and from 2001 to 2004. Similar behaviour for region SW occurs in the year 2000, however, from 2004 to 2006 there was an increase in the number of change points indicating a negative change.

Recall that in the years 1990, 1997, 1999, 2000 and 2001 important environmental measures were implemented by the Mexican authorities. Observing Figure 3 and Table 4 (or Figure 4) we have that there is an indication that some measures might have helped to increase the mean waiting time between two consecutive ozone peaks. As an example of this fact, take for instance regions NE and SE. We may notice that there was an increase in the number of change points indicating an increase in the waiting times in the year 1991, however, in the following year an increase of change points indicating a decrease in this mean waiting time occurred. Hence, we have that some improvement in the ozone air quality occurred right after the first environmental measure was implemented. In the years of 1997 and 1998, there was also an increase in the mean number of change points representing a positive change. Nevertheless, that was not enough to keep a consistent increase in the mean waiting time between ozone peaks. After the implementation of another environmental measure in 1999, in region SE there was a very large increase in the number of change points representing a negative change. However, it seems that altogether, the set of environmental measures taken by the environmental authorities throughout the years have produced a positive effect in those regions. These are reflected by the decrease in the number of possible change points after the year 2002.

Also, note that the largest increase in the number of change points indicating an increase in the mean waiting time occur in region NW after 2002. We also have that in general after 2001 (except for regions NW and SW) there is a decrease in the number of change points. In fact, for region SW, we have that the number of change points keeps more or less the same behaviour throughout the observational period. One reason for that could be that the ozone pollution in region SW is very severe. Also note that for that region the number of change points indicating a decreasing in the mean waiting time between ozone peaks has increased while the number of those indicating an increase in the mean time has decreased. Hence, even though several environmental measures were taken, it seems that they were not enough to cause a substantial change in the behaviour or the pollutant in that region. Note that for region NW, even though the number of change points are large after the year 2002, we would like to point out that they are mostly change points indicating a positive change, that is, an increase in the mean waiting time between ozone peaks. Hence, it seems that in that case, we have that a large number of changes indicating an increase in the mean waiting time between ozone peaks have been produced.

Looking at the results presented in here, we may conclude that, in general, the environmental policies that have been implemented by the Mexican authorities have had a positive effect in almost all region in Mexico City. The exception being region SW. However, we would like to recall that this region is the one that receives and keeps (because of its geographic location and wind direction in the city from NE to SW) a large amount of ozone precursors and also ozone itself. Therefore, it is possible that in order to have a more substantial improvement in the ozone air quality in region SW we should have more drastic measures in order to have a huge decrease of ozone concentration and ozone precursors in region NE and CE.

Appendix: WinBugs codes

In this section we present the WinBugs codes used in each model.

Model I.

model{

$t[1] \sim \text{dexp}(\text{lambda}[1])$

$\text{lambda}[1] \leftarrow \text{alpha} * M + w[1]$

$w[1] \sim \text{dgamma}(1, 1)$

$\text{theta}[1] \leftarrow 1/\text{lambda}[1]$

for $(i \text{ in } 2 : n)\{$

$t[i] \sim \text{dexp}(\text{lambda}[i])$

$\text{lambda}[i] \leftarrow \text{alpha} * (M - i + 1) + w[i] + \text{phi} * t[i - 1]$

$\text{theta}[i] \leftarrow 1/\text{lambda}[i]$

$w[i] \sim \text{dgamma}(1, 1)$

}

$\text{phi} \sim \text{dbeta}(a, b)$

$\text{alpha} \sim \text{dunif}(c, d)$

$M \sim \text{dunif}(e, f)$

}

Model II.

model{

$t[1] \sim \text{dexp}(\text{lambda}[1])$

$\text{lambda}[1] \leftarrow \text{alpha} * M + w[1]$

$w[1] \sim \text{dunif}(0, 100)$

$\text{theta}[1] \leftarrow 1/\text{lambda}[1]$

$\text{for } (i \text{ in } 2:n)\{$

$t[i] \sim \text{dexp}(\text{lambda}[i])$

$\text{lambda}[i] \leftarrow \text{alpha} * (M - i + 1) + w[i] + \text{phi} * t[i-1]$

$\text{theta}[i] \leftarrow 1/\text{lambda}[i]$

$w[i] \sim \text{dunif}(0, 100)$

$\}$

$\text{phi} \sim \text{dbeta}(a, b)$

$\text{alpha} \sim \text{dunif}(c, d)$

$M \sim \text{dunif}(e, f)$

$\}$

Model III.

model{

$t[1] \sim \text{dexp}(\text{lambda}[1])$

$\text{lambda}[1] \leftarrow \text{alpha} * M$

$\text{theta}[1] \leftarrow 1/\text{lambda}[1]$

$\text{for } (i \text{ in } 2:n)\{$

$t[i] \sim \text{dexp}(\text{lambda}[i])$

$\text{lambda}[i] \leftarrow \text{alpha} * (M - i + 1) + \text{phi} * t[i-1]$

$\text{theta}[i] \leftarrow 1/\text{lambda}[i]$

$\}$

$\text{phi} \sim \text{dbeta}(a, b)$

$\text{alpha} \sim \text{dunif}(c, d)$

$M \sim \text{dunif}(e, f)$

$\}$

Model IV.

model{

$t[1] \sim \text{dexp}(\text{lambda}[1])$

$\text{lambda}[1] \leftarrow \text{alpha} * M + w[1]$

$w[1] \sim \text{dunif}(0, 100)$

$\text{theta}[1] \leftarrow 1/\text{lambda}[1]$

for $(i \text{ in } 2:n)\{$

$t[i] \sim \text{dexp}(\text{lambda}[i])$

$\text{lambda}[i] \leftarrow \text{alpha} * (M - i + 1) + w[i]$

$\text{theta}[i] \leftarrow 1/\text{lambda}[i]$

$w[i] \sim \text{dunif}(0, 100)$

}

$\text{alpha} \sim \text{dunif}(a, b)$

$M \sim \text{dunif}(c, d)$

}

Multiple change points.

model{

$t[1] \sim \text{dexp}(\text{lambda}[1])$

$\text{lambda}[1] \leftarrow \text{alpha} * M + w[1]$

$w[1] \sim \text{dunif}(0, 100)$

$\text{theta}[1] \leftarrow 1/\text{lambda}[1]$

for $(i \text{ in } 2:n)\{$

$t[i] \sim \text{dexp}(\text{lambda}[i])$

$\text{lambda}[i] \leftarrow \text{alpha} * (M - i + 1) + w[i]$

$\text{theta}[i] \leftarrow 1/\text{lambda}[i]$

$w[i] \sim \text{dunif}(0, 100)$

$\text{diff}[i] \leftarrow \text{theta}[i] - \text{theta}[i-1]$

}

$\text{alpha} \sim \text{dunif}(a, b)$

$M \sim \text{dunif}(c, d)$

}

Acknowledgments

The authors would like to thank the two referees for all their comments, suggestions and interesting questions. J. A. Achcar was partially funded by a Conselho Nacional de Pesquisa—Brazil, grant number 300235/2005-4. E. R. Rodrigues 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. E. R. Rodrigues was partially funded by Dirección General de Apoyo al Personal Académico of the Universidad Nacional Autónoma de México grant number 968SFA/2007.

References

Achcar, J. A., Fernández-Bremauntz, A. A., Rodrigues, E. R. and Tzintzun, G. (2008). Estimating the number of ozone peaks in Mexico City using a non-homogeneous Poisson model. *Environmetrics* **19**, 469–485. MR2523909

Achcar, J. A., Rodrigues, E. R. and Tzintzun, G. (2009a). Using non-homogeneous Poisson models with multiple change-points to estimate the number of ozone exceedances in Mexico City. *Environmetrics*. Published on line in September 2009, DOI:10.1002/env.1029.

Achcar, J. A., Rodrigues, E. R., Paulino, C. A. and Soares, P. (2009b). Non-homogeneous Poisson process with a change-point: An application to ozone exceedances in Mexico City. *Environmental and Ecological Statistics*. Published on line in May 2009, DOI:10.1007/s10651-009-0114-3.

Achcar, J. A., Ortíz-Rodríguez, G. and Rodrigues, E. R. (2009c). Estimating the number of ozone peaks in Mexico City using a non-homogeneous Poisson model and a Metropolis–Hastings algorithm. *International Journal of Pure and Applied Mathematics* **53**, 1–20.

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

ARB (Air Resource Board) (2005). Review of the California ambient air quality standard for ozone. Staff report. California Environmental Protection Agency, USA.

Austin, J. and Tran, H. (1999). A characterization of the weekday–weekend behavior of ambient ozone concentrations in California. In *Air Pollution VI* 645–661. Ashurst Lodge, UK: WIT Press.

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

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

Bernardo, J.-M. and Smith, A. F. M. (1994). *Bayesian Theory*. New York: Wiley. MR1274699

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

Gauderman, W. J., Avol, E., Gililand, F., Vora, H., Thomas, D., Berhane, K., McConnel, R., Kuenzli, N., Lurmman, F., Rappaport, E., Margolis, H., Bates, D. and 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. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85**, 398–409. MR1141740

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

Guardani, R., Aguiar, J. L., Nascimento, C. A. O., Lacava, C. I. V. and Yanagi, Y. (2003). 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**, 553–559.

Horowitz, J. (1980). Extreme values from a nonstationary stochastic process: An application to air quality analysis. *Technometrics* **22**, 469–482. MR0596792

Huerta, G. and Sansó, B. (2005). Time-varying models for extreme values. Technical Report 2005–4. Department of Applied Mathematics and Statistics, University of California, Santa Cruz.

Itô, K., de León, S. and Lippman, M. (2005). Associations between ozone and daily mortality: A review and additional analysis. *Epidemiology* **16**, 446–457.

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

Jelinski, Z. and Moranda, P. B. (1972). Software reliability research. In *Statistical Computer Performance Evaluation* (Freiberger, W., ed.) 465–497. New York: Academic Press.

Lanfredi, M. and Macchiato, M. (1997). Searching for low dimensionality in air pollution time series. *Europhysics Lett.* **40**, 589–594.

Larsen, L. C., Bradley, R. A. and Honcoop, G. L. (1990). 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*. Los Angeles.

Leadbetter, M. R. (1991). On a basis for "peak over threshold" modeling. *Statistics and Probability Letters* **12**, 357–362. MR1131065

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

Moranda, P. B. (1975). Prediction of software reliability and its applications. In *Proceedings of the Annual Reliability and maintainability Symposium* 327–332. Washington, DC.

NOM (2002). Modificación a la Norma Oficial Mexicana NOM-020-SSA1–1993. Diario Oficial de la Federación. 30 de Octubre de 2002.

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

Pan, J.-N. and Chen, S.-T. (2008). Monitoring long-memory air quality data using ARFIMA model. *Environmetrics* **19**, 209–219. MR2420186

Piegorsch, W. W., Smith, E. P., Edwards, D. and Smith, L. (1998). Statistical advances in environmental sciences. *Statistical Sciences* **13**, 186–208.

Raftery, A. E. (1989). Are ozone exceedance rate decreasing?, Comment on 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**, 378–381. MR1041763

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

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

Seinfeld, J. H. (2004). Air pollution: A half century of progress. *American Institute of Chemical Engineers Journal* **50**, 1098–1108.

Smith, R. L. (1989). Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone. *Statistical Sciences* **4**, 367–377. MR1041763

Smith, A. F. M. and Roberts, G. O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society Series B* **55**, 3–23. MR1210421

Spiegelhalter, D. J., Thomas, A. and Best, N. G. (1999). *WinBugs: Bayesian inference using Gibbs sampling*. Cambridge, UK: MRC Biostatistics Unit.

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

J. A. Achcar
Departamento de Medicina Social
Faculdade de Medicina de Ribeirão Preto
Universidade de São Paulo
14049-900—Ribeirão Preto—SP
Brazil
E-mail: achcar@fmrp.usp.br

E. R. Rodrigues
Instituto de Matemáticas
Universidad Nacional Autónoma de México
México

G. Tzintzun
Instituto Nacional de Ecología
Secretaría de Medio Ambiente
y Recusos Naturales
México