Journal of Environmental Statistics

August 2013, Volume 5, Issue 3. http://www.jenvstat.org

Non-homogeneous Poisson process in the presence of one or more change-points: an application to air pollution data

Lorena Vicini
Federal University of Santa Maria

Luiz Koodi Hotta
University of Campinas

Jorge Alberto Achcar
Medical School of Ribeirão Preto

Abstract

We consider the problem of modeling the number of times that an air quality standard is exceeded in a certain period of time. We assume that the number of times the threshold is exceeded takes place according to a non-homogeneous Poisson process (NHPP) with the mean function modeled by the generalized gamma distribution. We consider models with and without change-points. When the presence of change-points is assumed, we have none , one, two or three change-points, depending on the data set. We use the Bayesian approach, where the posterior summaries of interest are obtained using standard Markov Chain Monte Carlo (MCMC) methods. We also discuss the use of different prior distributions for the parameters of the models, with an analysis of the convergence of the Gibbs sampling algorithm and sensitivity for the choice of different priors. To illustrate the proposed method we consider simulated data and a pollution data set from of a region of Mexico City.

Keywords: Bayesian analysis, Markov Chain Monte Carlo methods and simulation, multiple change-points, ozone air pollution.

1. Introduction

One problem that has affected many regions around the world is air pollution. In some places, such as in big cities and industrial regions, there is a higher concentration of pollution. However, due to wind, the air pollution can spread to other regions .

Air pollution has become a public health problem, since an increase in pollution can cause serious public health problems, such as diseases related to respiratory and cardiovascular systems; these have been highlighted in many health studies (see, for example, Braga et al., 2002 ; Gouveia et al., 2006). Air pollution is characterized by the presence of toxic gases and liquid or solid particles in the air. An important example of a pollutant is ozone, because when its concentration remains above a threshold level for a certain period of time, individuals exposed to it can suffer serious health problems (see, for example, Air Resource Board (ARB), 2005).

In this paper we analyze a series of data for ozone (O_3), which is a gas composed of three oxygen atoms and formed by chemical reactions between nitrogen oxides (NOx) and volatile organic compounds ($VOC's$) in the presence of sunlight. Ozone has the same chemical structure miles above the earth or at ground level and it can be "good" or "bad" depending on its location in the atmosphere. In the lower atmosphere, the tropospheric ozone is considered "bad".

As pollution levels have increased at an alarming rate in recent years, exceeding the limits for acceptable standard of air quality on certain days, studies related to the problem are gaining prominence around the world. As a direct result, new models and statistical methods have been developed to analyze air pollution data. Considering as the main interest the estimation of the number of times that a given environmental standard is violated, Javits (1980) assumes Bernoulli and Poisson models; Raftery (1989) uses a mixture of homogeneous Poisson models. Since time homogeneity is usually not verified in applications to air pollution data some authors assume non-homogeneous Poisson processes. It is important to point out that even assuming non-homogeneous Poisson models, usually we could have the presence in the model of one or more change-points (see, for example, Achcar et al., 2010; Achcar et al., 2011).

In relation to pollution by ozone gas, the literature contains several studies (see, for example, Wilson et al., 1980; Loomis et al., 1996; Galizia and Kinney, 1999; Bell et al., 2004; Gauderman et al., 2004; Álvarez et al.; 2005, ARB, 2005; Bell et al., 2005; Bell et al., 2007 ; and Achcar et al., 2010).

Other studies are also related to diseases caused by an increase in the level of air pollution (see, for example, Martins et al., 2002; Farhat et al., 2005).

In this paper, we consider how to model the data by a non-homogeneous Poisson process (NHPP) with the rate function modeled by the generalized gamma distribution. The model is used to analyze the daily data set collected by the monitoring network of the Metropolitan Area of Mexico City. The set contains 18 years of daily average ozone measurements in the period from 1 January 1990 to 31 December 2008. The Metropolitan Area of Mexico City is divided into five regions, corresponding to the Center (CE), Northwest (NW), Northeast (NE), Southeast (SE) and Southwest (SW).

This paper considers the modeling of the data in the presence or not of one or more change-points. We get posterior summaries of interest using standard MCMC methods, in special,the Gibbs sampling algorithm or the Metropolis-Hastings algorithm. Also it is discussed the sensitivity of the choice of different prior distributions for the parameters of the model and their effect in the convergence of the simulation algorithm. The studies are illustrated with simulated data and a pollution real data set.

This article is organized as follows: in Section 2, the model is presented; Section 3 presents the Bayesian formulation, first without taking into account change-points, which are incorporated later; examples considering simulated data set are given in Section 4; in Section 5, the proposed models are applied to a data set collected by the monitoring network of the Metropolitan Area of Mexico City; Section 6 concludes with final remarks and discussions of the results.

2. Description of the model

The NHPP model has been used to model various phenomena. For example, times from remission for patients with leukemia (Matthews and Farewell, 1982), intervals between coal-mining disasters (Raftery and Akman, 1986, and Yang and Kuo, 2001), time of failures in repairable systems (Ruggeri and Sivaganesan, 2005), arrival times of calls to a call center (Weinberg et al., 2007). Other examples can be found in Leemis (1991). The problem can be described as follows. Let $T > 0$ be a real number and $M = \{M(t) : t \in (0, T]\}$, where the random variable $M(t)$ represents the cumulative number of events in the time interval $[0, t)$ for $t \geq 0$. In the NHPP model the random variable $M(t)$ has a Poisson distribution with mean $m(t)$. One can also characterize the distribution by the intensity function $\lambda(t) = \frac{dm(t)}{dt}$. If $\lambda(t)$ is a constant, so that $m(t)$ is linear, then $M(t)$ is called a homogeneous Poisson process; otherwise the process is called a non-homogeneous Poisson process.

Different choices for the function $m(t)$ are considered in the literature, especially in software reliability modeling (see, for example, Achcar et al., 1998). Goel and Okumoto (1979) stated that the expected number of software failures for time t is given by the mean value function $m(t)$, which is non-decreasing and bounded above. Specifically, they considered the mean and intensity functions given, respectively by,

$$m_1(t) = \theta \left(1 - e^{-\beta t}\right), \quad \text{and} \tag{1}$$

$$\lambda_1(t) = \theta\beta e^{-\beta t}, \tag{2}$$

where, in our case, θ represents the expected maximum number of days in which the air quality standard is violated by a particular pollutant and β is considered to be the rate at which events occur. Goel (1983) generalized model (1) proposing the intensity function

$$\lambda_2(t) = \theta\beta\alpha t^{\alpha-1} e^{-\beta t^{\alpha}}, \tag{3}$$

with mean function

$$m_2(t) = \theta \left(1 - e^{-\beta t^{\alpha}}\right). \tag{4}$$

Note that (1) and (4) can be written as special cases of the general form where the mean value function is given as

$$m(t) = \theta F(t), \tag{5}$$

where $F(t)$ is a distribution function. On the other hand, for any distribution function $F(t)$ we have a valid model.

A widely used distribution function, given its high flexibility, is the generalized gamma distribution. In this case the mean function is

$$m_{GG}(t) = \theta I_k(\beta t^{\alpha}), \tag{6}$$

where $I_k(s)$ is the integral of the gamma function given by

$$I_k(s) = \frac{1}{\Gamma(k)} \int_0^s x^{k-1} e^{-x} dx. \tag{7}$$

From (7), we obtain the intensity function given by

$$\lambda_{GG}(s) = m'_{GG}(t) = \frac{1}{\Gamma(k)} \theta \beta^k \alpha t^{\alpha k - 1} e^{-\beta t^\alpha}. \tag{8}$$

This model is called generalized gamma. When k is an integer we can write $F(t)$ as

$$F(t) = 1 - e^{-\beta t^\alpha} \sum_{j=o}^{k-1} \frac{(\beta t^\alpha)^j}{j!}. \tag{9}$$

The three models used in this article are described as follows:
a) Model I: Here all parameters are unknown, $\lambda_{GG}(t)$ is given by (8).
b) Model II: With $k = 1$, $\lambda_{GG}(t)$ is given by $\lambda_2(t)$ (3).
c) Model III: With $k = 1$ and $\alpha = 1$, $\lambda_{GG}(t)$ reduces to $\lambda_1(t)$ (2).

In the Bayesian analysis of these models, we may have some difficulties in obtaining the Bayesian inferences using MCMC simulation methods. These difficulties are mainly related to the convergence of the MCMC chains. For this reason we will study the effect of different prior distributions on the performance of sample simulation algorithms of the posterior distribution of interest.

The main idea of this paper is to model the number of times that an air quality standard is exceeded in a period of time, under a Bayesian approach. To do this, we explore the generalized intensity functions that can best fit the data set. In particular, we fit the generalized intensity function with and without the presence of change-points. We also study several prior distributions for the parameters because of the convergence problem mentioned above, especially when all the parameters are estimated simultaneously. For this reason we perform sensitivity analysis in relation to the choice of prior distributions. The biggest concern is related to the parameter k, a parameter of the gamma distribution.

The convergence of the MCMC algorithm was analyzed by graphical methods and by the Gelman-Rubin statistic (Gelman and Rubin, 1992). The Gelman-Rubin statistic relies on parallel chains to test whether they all converge to the same posterior distribution. Brooks and Gelman (1998) suggested the introduction of a correction factor. This statistic is evaluated using the coda package in R. This statistic is larger or equal to 1.0. The closest this statistic is to 1.0, more evidence we have that the chain is near convergence. The limit value of 1.2 is sometimes used as a guideline for “approximate convergence” (Gelman, 1996).

3. Bayesian inference

Denote by $D_T = \{n; t_1, t_2, ..., t_n; T\}$, the data set, where n is the number of events observed such that $0 \leq t_1 < t_2 < ... < t_n \leq T$, and where t_i are the times of the events observed during the period of time $(0, T]$.

We consider that the parameters θ, β, α and k are unknown and will be estimated. In the Bayesian framework, for each parameter, we must select prior distributions which describe the uncertainty about them.

In this article we consider the presence or absence of change-points to NHPP. We have two different forms for the likelihood function of the model, one for each formulation. First, we define the notation and expressions for the case without change-points, and then we do the same for the case when there are one or more change-points.

3.1. Models without change-points

For Model I, the likelihood function for the vector $\Theta = (\theta, \beta, \alpha, k)$, considering T as the truncation time of the truncated model (see, for example, Cox and Lewis, 1966), is given by

$$L(\Theta \mid D_T) = \left(\prod_{i=1}^{n} \lambda(t_i)\right) \exp(-m(T)). \tag{10}$$

In some cases it is advisable to enter a latent variable, as this may serve as a computational aid. We consider the introduction of the latent variable N' which has Poisson distribution with parameter $\theta[1 - F(T \mid \beta)]$ (see, for example, Achcar et al., 1998).

Considering the generalized gamma distribution model given in (6), the likelihood function for the vector of parameters $\Theta = (\theta, \beta, \alpha, k)$ is expressed as

$$L(\Theta \mid D_T) = \left\{\prod_{i=1}^{n} \theta \frac{\beta^k}{\Gamma(k)} \alpha t_i^{\alpha k-1} e^{-\beta t_i^{\alpha}}\right\} \exp\left\{-\int_o^T \theta \frac{\beta^k}{\Gamma(k)} \alpha u^{\alpha k-1} e^{-\beta u^{\alpha}} du\right\},$$

such that

$$L(\Theta \mid D_T) = \frac{\theta^n \alpha^n \beta^{kn}}{\{\Gamma(k)\}^n} \left\{\prod_{i=1}^{n} t_i^{\alpha k-1}\right\} \exp\left\{-\beta \sum_{i=1}^{n} t_i^{\alpha} - \theta I_k(\beta T^{\alpha})\right\}. \tag{11}$$

The prior distributions and the conditional posterior distributions of the MCMC algorithm are presented in the appendix.

3.2. Models with change-points

The ozone pollution often changes during the time interval $(0, T]$, due to some type of intervention or change. In special, some political decisions by public authorities could implies in a decreasing or an increasing in ozone gas emission and, therefore, a similar effect can occur in the daily measurements of ozone.

In this case we can have J change-points, which we denote by τ_j, $j = 1, ..., J$, and in each interval we use the generalized gamma model, the NHPP model, presented in Section 2.

We assume here that all the parameters are unknown: $(\theta_i, \alpha_i, \beta_i,\ k_i)$ and τ_j, where $i = 1, ..., J+1$ and $j = 1, ..., J$. They must be estimated, and the change-points can occur at any time $\tau_0, \tau_1, \cdots, \tau_i$, where $\tau_0 = 0$.

In this case, we have that the rate function of the NHPP process is of the form (Achcar et al., 2010)

$$\lambda(t \mid \Theta) = \begin{cases} \lambda(t \mid \Theta_1), & \text{se } 0 \leq t < \tau_1 \\ \lambda(t \mid \Theta_j), & \text{se } \tau_{j-1} \leq t < \tau_j, j = 2, 3, ..., J, \\ \lambda(t \mid \Theta_{J+1}), & \text{se } \tau_J \leq t \leq T, \end{cases} \tag{12}$$

where the intensity functions $\lambda(t \mid \Theta_j), j = 1, 2, ..., J+1$ are related to equation (12) and $\Theta_j = (\theta_j, \beta_j, \alpha_j, k_j)$, $j = 1, 2, ..., J+1$ are the parameters associated with the NHPP process in each interval limited by the change-points.

For the generalized gamma distribution the rate function is of the form

$$\lambda(t \mid \Theta) = \frac{1}{\Gamma(k_j)} \theta_j \beta_j^{k_j} \alpha_j t^{\alpha_j k_j - 1} e^{-\beta_j t^{\alpha_j}}.$$

The mean function $m(t \mid \Theta_j), j = 1, 2, ..., J+1$, as given in Achcar et al. (2010) is of the form

$$m(t \mid \Theta) = \begin{cases} m(t \mid \Theta_1), & \text{se } 0 \leq t < \tau_1 \\ m(\tau_1 \mid \Theta_1) + m(t \mid \Theta_2) - m(\tau_1 \mid \Theta_2), & \text{se } \tau_1 \leq t < \tau_2, \\ m(t \mid \Theta_{j+1}) - m(\tau_j \mid \Theta_{j+1}) & \\ + \sum_{i=2}^{j} [m(\tau_i \mid \Theta_i) - m(\tau_{i-1} \mid \Theta_i)] + m(\tau_1 \mid \Theta_1), & \text{se } \tau_j \leq t < \tau_{j+1}, \\ j = 2, 3, ..., J, & \end{cases} \tag{13}$$

where $\tau_{J+1} = T$. For the generalized gamma distribution the mean function is given by

$$m(t \mid \Theta) = \theta_j I_{k_j}(\beta_j t^{\alpha_j}).$$

Let $\boldsymbol{w} = (\Theta_1, \Theta_2, ..., \Theta_{J+1}; \tau_1, \tau_2, ..., \tau_J)$ is the vector of all parameters. Having assumed an NHPP, the likelihood function is given by

$$\begin{aligned} L(w \mid D_T) \quad \propto \quad & \prod_{i=1}^{N_{\tau_1}} \lambda(t_i \mid w_1) e^{-m(\tau_1 \mid w_1)} \left[\prod_{j=2}^{J} \left(\prod_{i=N_{\tau_{j-1}}+1}^{N_{\tau_j}} \lambda(t_i \mid w_j) e^{-[m(\tau_j \mid w_j) - m(\tau_{j-1} \mid w_j)]} \right) \right] \\ & \times \prod_{i=N_{\tau_J}+1}^{N_T} \lambda(t_i \mid w_{J+1}) e^{-[m(T \mid w_{J+1}) - m(\tau_J \mid w_{J+1})]}. \end{aligned} \tag{14}$$

The inclusion of change-points in the restricted Models II and III is carried out in the same way as in the inclusion in Model I.

4. Application to simulated data

We consider models without change-points and with one and two change-points. These models will initially be tested with simulated data sets and later they will be used to model ozone pollution data from Mexico City based on when a threshold in the levels of ozone concentration is exceeded for a certain period of time. In both cases we will discuss the convergence of the chains and the adjustment of the function $m(t)$. In most cases the number of simulated data is not large because we want to test whether the method works in this situation. Thus, we do not expect to find small variances for the posterior distributions.

4.1. Simulated data without change-point

We first simulated 300 observations with $\theta = 300$, $\beta = 0.02$, $\alpha = 1$ and $k = 1$, i.e. Model III without a change-point. Models I and II are also correct because they incorporate this model, but they are over-specified.

We obtained a summary of estimates of the posterior distributions of Model I based on $900,000$ MCMC replications after a burn-in of size $50,000$ and jumps of 500 samples (A jump of 500 means that every 500th sample is chosen from the simulated Gibbs samples to have approximately uncorrelated samples used to get Monte Carlo estimates for the random quantities of interest.) For the restricted Models II and III, summaries of the estimates of the posterior distributions were obtained using $100,000$ iterations, a burn-in of size $10,000$ and jumps of 10 samples.

As the analysis of the adjustment of the models without change-point was carried out in Vicini et al. (2012), we only present the fit of the three models and the summaries of the estimated posterior distributions. Vicini et al. (2012) analyzed the sensitivity of the estimates of all the parameters, and they found difficulties in obtaining convergence of the chains for the parameter k_i of the Model, when assuming the improper prior used in the first set of distributions. They tested various specifications of prior distributions and different values for hyperparameters and suggested the truncated exponential distribution, which is used here.

Figure 1 shows the adjustment of Models I, II and III in the absence of a change-point. This figure presents the graphs of the theoretical function $m(t)$, its empirical (or nonparametric) estimation, and its estimation by Models I, II and III without a change-point. We observe a good fit of all assumed models to the data set. We observe that the estimated curves are superimposed on each other and on the empirical curve. The curves are evaluated at the parameter values given by the median of the posterior distribution, since these estimates provide better results than using the mean values. From now on we will always use the posterior medians in place of the posterior means.

Table 1 presents the estimates of the mean, median and standard deviation of the parameters of Models I, II and III without change-points. The 95% credible intervals (CI) were obtained by taking the 2.5% and 97.5% percentiles of the simulated Gibbs sample. We observe that the parameter values were generally in their respective 95% CI. In this paper, we use the 2.5% and 97.5% percentiles for all the CI.

Table 1: Summary of the posterior distributions of the parameters for Models I, II and III, with no change-points. Real model:$\theta = 300$, $\beta = 0.02$, $\alpha = 1$ and $k = 1$. Gibbs sample of size 300. Prior distributions for the parameters θ, β and α are presented in Section (3). The prior distribution for the parameter k in Model I is the exponential distribution with mean parameter equal to 0.95 and truncated at 3.

Model	parameter	mean	median	S.D.	(CI)2.5%	(CI)97.5%
I	θ	310.00	309.47	17.86	282.30	340.10
	α	1.234	1.219	0.174	0.976	1.543
	β	0.007687	0.005503	0.007301	0.000941	0.022115
	k	0.765	0.751	0.161	0.528	1.057
II	θ	310.28	309.89	17.63	282.27	339.70
	α	1.002	1.002	0.049	0.922	1.085
	β	0.019282	0.018933	0.004088	0.013193	0.026475
III	θ	310.03	309.82	17.77	281.62	339.49
	β	0.019115	0.019087	0.001214	0.017144	0.021113

We observe in Table 1 that the Monte Carlo estimates for the posterior mean and posterior median of the parameters of the subsequent Models I, II and III were satisfactory, because their values are very close to the true values of the parameters. We can also check that all CI contain the true value of the parameters. We observed good convergence of the simulation

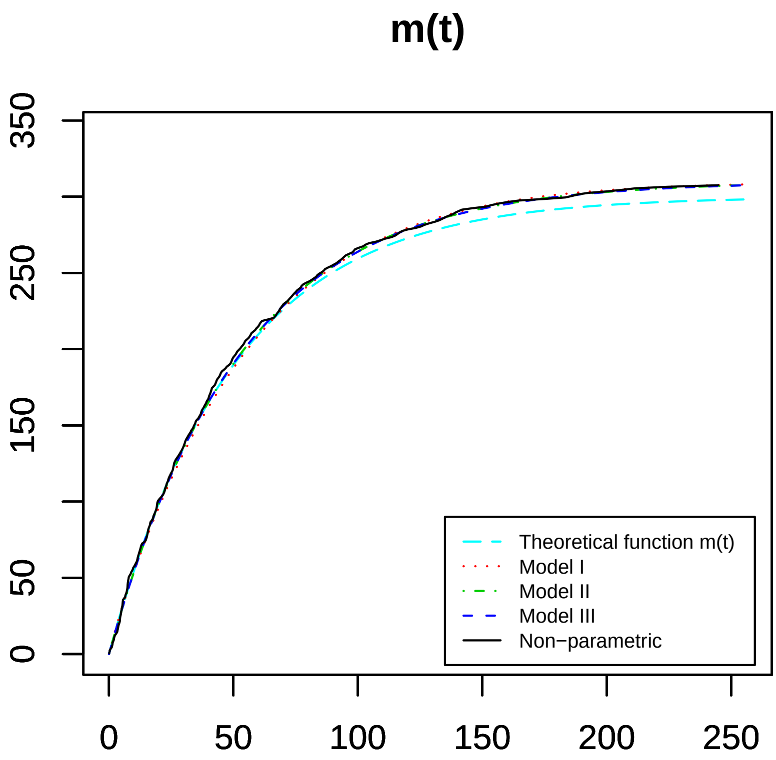

Figure 1: Simulation from the model without change-point with parameters $(k, \alpha) = (1, 1)$. Sample size equal to 300. Graphs of the theoretical function $m(t)$, the empirical estimation and its estimates by models without change-point: a) Model I, b) Model II c) Model III.

algorithm, as observed in the trace plots of the simulated samples. The largest value of the Gelman-Rubin statistic is equal to 1.01 also indicating convergence of the MCMC chains. The results indicate that the information that $\alpha = k = 1$ is not relevant to the estimation of the parameter θ, as the estimate of the posterior distribution remains almost the same whether or not the constraint is taken into account.

4.2. Simulated data with two change-points

In this case, we simulated 150 observations with parameter values given by $\theta_1 = 190$, $\theta_2 = 130$, $\theta_3 = 365$, $\beta_1 = 0.009$, $\beta_2 = 0.007$, $\beta_3 = 0.003$, $\alpha_1 = 1$, $\alpha_2 = 1$, $\alpha_3 = 1$, $\tau_1 = 50$, $\tau_2 = 100$, $k_1 = 1$, $k_2 = 1$ and $k_3 = 1$, with change-points $\tau_1 = 50$ and $\tau_2 = 100$, with 50 observations before τ_1, 50 observations between τ_1 and τ_2 and 50 after τ_2. The prior distributions for the parameters k_1 and k_2 in Model I are the second set of the proposed distributions presented in Section 3, where we have the exponential prior distribution with mean parameter equal to 0.99 and truncated at 6, for both parameters.

Here all three models, Models I, II and III, are correct, but Model II incorporates the information that $k_1 = k_2 = 1$ and Model III incorporates the information that $\alpha_1 = \alpha_2 = \alpha_3 = k_1 = k_2 = k_3 = 1$, while in Model I these parameters are estimated.

The posterior summaries of Model I were obtained using $100,000$ iterations, a burn-in of size $30,000$ and jumps of size 50 in the simulation algorithm. For the posterior summaries of Models II and III we used $80,000$ iterations, a burn-in of size $20,000$ and jumps of size 50.

We also performed a sensitivity analysis with respect to the specifications of the a priori distributions. Here we present only the best obtained parameter estimates for these models, which are given using the following specifications of prior distributions for the parameters: $\theta_1 \sim Gamma(0.001, 0.001)$; $\theta_2 \sim Gamma(0.001, 0.001)$; $\theta_3 \sim Gamma(0.001, 0.001)$; $\beta_1 \sim Gamma(1, 100)$; $\beta_2 \sim Gamma(1, 100)$; $\beta_3 \sim Gamma(0.01 \times 100, 100)$; $\alpha_1 \sim Gamma(1, 1)$; $\alpha_2 \sim Gamma(1, 1)$; $\alpha_3 \sim Gamma(1, 1)$; $\tau_1 \sim Uniform(40, 60)$ and $\tau_2 \sim Uniform(90, 110)$.

We only present the graphical results of Model I, where all the parameters are free.

Figure 2 shows the graphs of the chains for the parameters of Model I with two change-points. Figure 3 shows the posterior and prior distributions and the true value of each parameter for Model I with two change-points.

The analyses of Figures 2 and 3 give an indication that the chains are converging to their true values, as can be seen from the horizontal trace in Figure 2 and the vertical line in Figure 3.

Figure 4 shows the fit of Models I, II and III, with the presence of two change-points. This figure presents the graphs of the theoretical function $m(t)$, its empirical estimation, and its estimation by Models I, II and III. All models with two change-points give a good fit for the simulated data. This result can be seen from the curves that overlap with each other.

Table 2 presents the estimates for the parameters of Models I, II and III with two change-points. Analyzing Table 2 we can observe that the estimates of the mean and median of Models I, II and III with two change-points were satisfactory, because their values are very close to the true values of the parameters. All the credibility intervals 95% contain the true value of the parameters. There was no problem in the convergence of the chains of the parameters, as shown by the chains and also by the good fit of the curves. The largest value of the Gelman-Rubin statistic is equal to 1.15, obtained for the parameter θ_2, also indicating convergence of the MCMC chains. Note that the three models are correct, but Models II and III incorporate more information, and Model III incorporates more information than Model II. We also see, by comparing the results of Model III with the other models, that adding the information that $k_1 = k_2 = k_3$ changes very little the posterior distribution of the parameter β_1, indicating that information on the parameters $k's$ is not relevant to the parameter β_1. Simultaneously analyzing Tables 1 and 2, we can see that the range of the credibility intervals of the parameters are larger in Table 2, because the sample size is smaller in this case.

5. Application to ozone data

Air pollution is one of the main problems facing large cities. There are many pollutants that cause problems to the population, but what most affects large cities is ozone gas. In some cities, like Mexico City, the authorities have been concerned with the high levels of pollution that it presents. Related to this problem, the environmental authorities have implemented measures aimed at reducing the level of pollutants. These measures are extremely important, because when pollution levels reach a certain threshold of concentration for a given period of

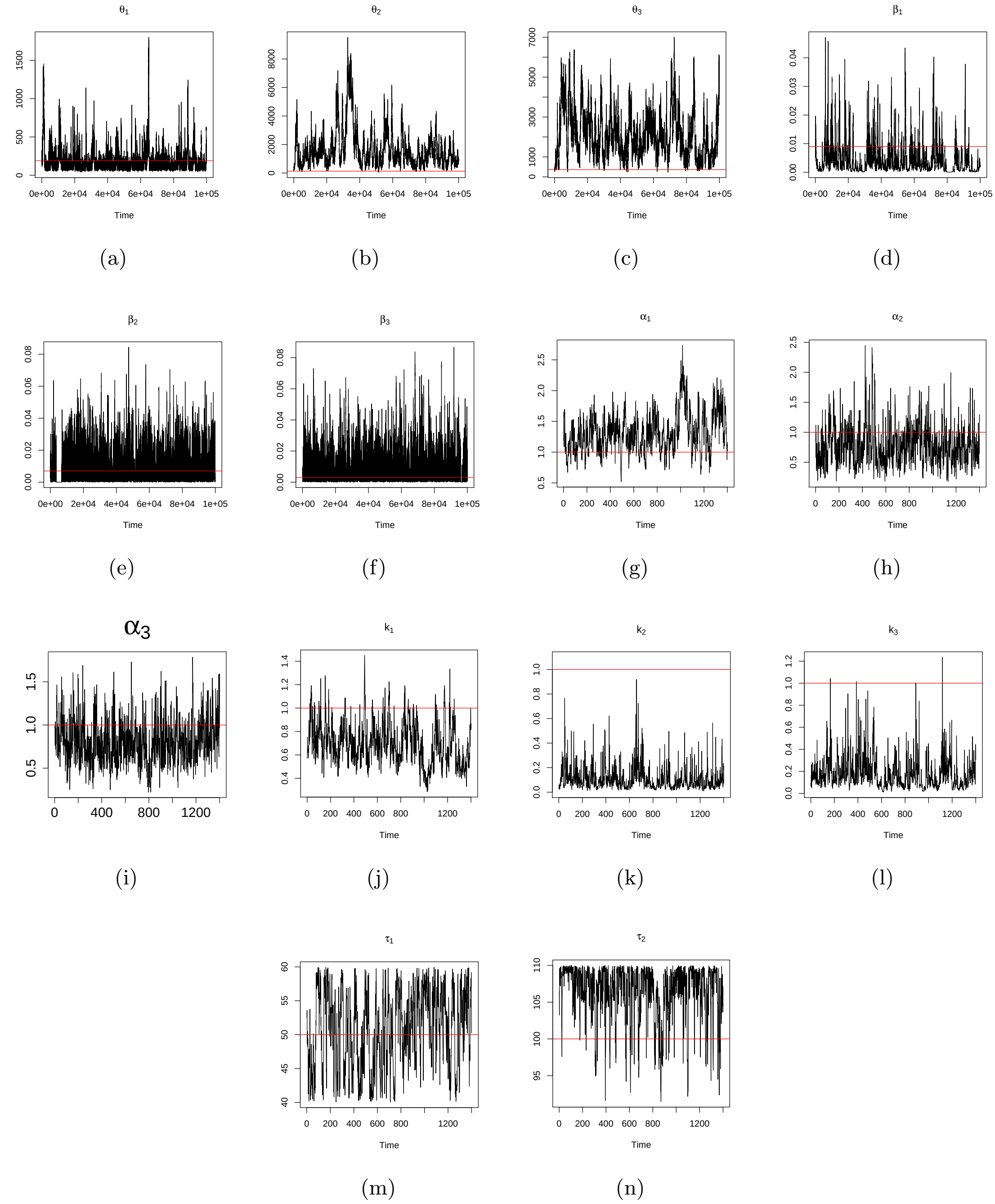

Figure 2: Simulation of the chains of Model I with two change-points. The horizontal traces are the true values of each parameter.

time, individuals exposed to the pollutant can suffer serious health problems (see, for example, Wilson et al., 1980; Loomis et al., 1996; Galizia and Kinney, 1999; Bell et al., 2004; Bell et al., 2005; Bell et al., 2007; Gauderman et al., 2004; and ARB, 2005). In this section, we apply the proposed models to the measurements of ozone in the Metropolitan Area of Mexico City. The Mexican ozone standard is 0.11 parts per million (0.11 ppm) and the threshold used in

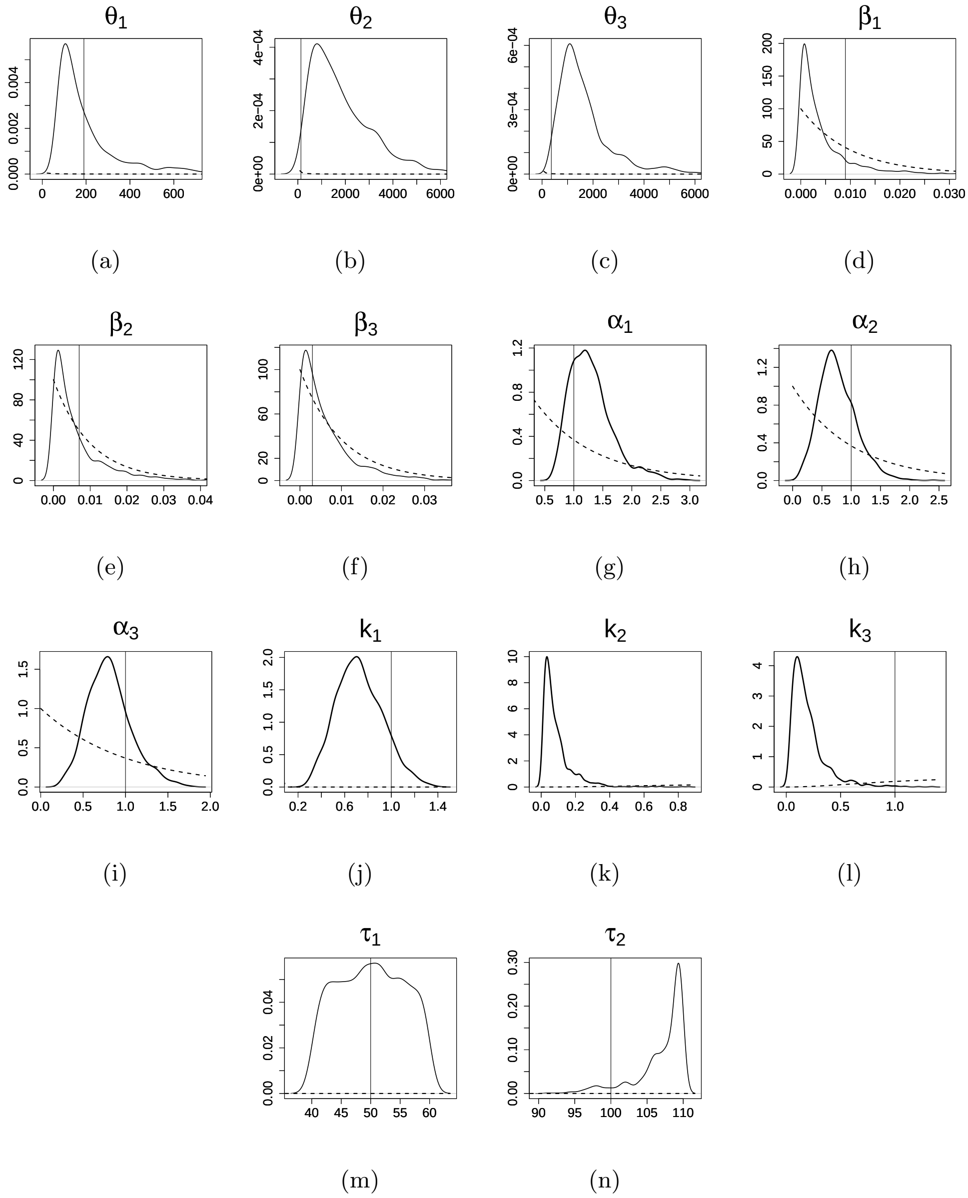

Figure 3: Simulation of the posterior distribution of the parameters of Model I with two change-points. The posterior distributions are shown with a solid line and the prior distributions are shown with the dashed line. The true values are indicated by vertical lines.

Mexico City to declare an emergency is 0.22 ppm (see, for example, Achcar et al., 2010). Here we consider a threshold equal to 0.20 ppm. This value is used because it is between the other two. We applied the proposed models to fit the data corresponding to the maximum daily average measurements of ozone gas, based on data measured in the Northeast region (NE) of Mexico City with a sample of 981 observations, which correspond to times when a certain threshold established for the air quality standard is violated during the period of time T which we considered. These data are collected from www.sma.df.gob.mx/simat/, which account for about 18 years of observations (1 January 1990 to 31 December 2008).

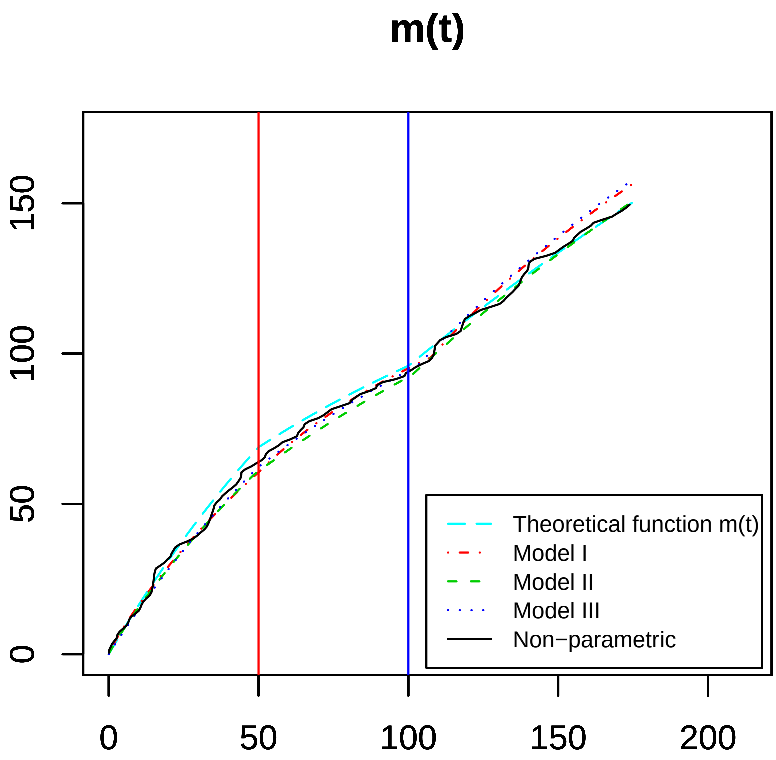

Figure 4: Model with two change-points. Sample size equal to 150. Graphs of the theoretical function $m(t)$, the empirical estimation and its estimates by models with two change-points: a) Model I, b) Model II c) Model III. The vertical lines indicate the change-points.

The posterior summaries of interest for Model I with no change-points were obtained using $700,000$ iterations, a burn-in of size $15,000$ and jumps of size 100. The specification of the prior distributions of the parameters for this model were: $\theta \sim Gamma(0.001, 0.001)$; $\beta \sim Gamma(10, 100)$; $\alpha \sim Gamma(0.01, 0.01)$. The assumed prior distribution for the parameter k is the exponential distribution with the mean parameter equal to 0.95 and truncated at 3.

For Model I with one change-point, the posterior estimations are obtained using $400,000$ iterations, a burn-in of size $200,000$ and jumps of size 100. The specifications of the prior distributions of the parameters for this model were: $\theta_1 \sim Gamma(0.001, 0.001)$; $\theta_2 \sim Gamma(0.001, 0.001)$; $\beta_1 \sim Gamma(0.001, 0.001)$; $\beta_2 \sim Gamma(0.001, 0.001)$; $\alpha_1 \sim Gamma(1, 1)$; $\alpha_2 \sim Gamma(1, 1)$ and $\tau_1 \sim Uniform(2900, 2950)$. The assumed prior distributions for the parameters k_1 and k_2 are the exponential distribution with mean parameter 0.99 and truncated at 6 for both parameters.

For Model I with two change-points the posterior estimations were obtained using $400,000$ iterations, a burn-in of $300,000$ and jumps of 200. The specifications of the prior distributions of the parameters for this model were $\theta_1 \sim Gamma(0.001, 0.001)$; $\theta_2 \sim Gamma(0.001, 0.001)$;

Table 2: Summary of estimates of the posterior distributions for Models I, II and III with two change-points. Real model: $\theta_1 = 190$, $\theta_2 = 130$, $\theta_3 = 365$, $\beta_1 = 0.009$, $\beta_2 = 0.007$, $\beta_3 = 0.003$, $\alpha_1 = 1$, $\alpha_2 = 1$, $\alpha_3 = 1$, $\tau_1 = 50$, $\tau_2 = 100$, $k_1 = 1$, $k_2 = 1$ and $k_3 = 1$. Sample size equal to 150. Prior distributions for the parameters θ_1, θ_2, β_1, β_2, α_1, α_2 and τ, are presented in Section 3. The prior distributions for the parameters k_1 and k_2 in Model I are the exponential distribution with the mean parameter equal to 0.99 and truncated at 6, for both parameters.

Model	parameter	mean	median	S.D.	2.5%	97.5%
I	θ_1	244.23	153.65	246.14	70.73	952.61
	θ_2	1587.49	1322.18	1080.72	321.76	4459.74
	θ_3	2121.27	1804.27	1353.79	552.48	5761.59
	β_1	0.00405	0.00221	0.00518	0.00013	0.01768
	β_2	0.00611	0.00368	0.00736	0.00006	0.02716
	β_3	0.00606	0.00366	0.0069	0.00007	0.0255
	τ_1	50.38	50.55	5.50	40.83	59.26
	τ_2	106.00	107.34	4.07	96.44	109.92
	α_1	1.263	1.232	0.335	0.716	2.022
	α_2	0.786	0.741	0.337	0.260	1.570
	α_3	0.797	0.768	0.284	0.336	1.512
	k_1	0.737	0.707	0.198	0.417	1.168
	k_2	0.093	0.063	0.096	0.013	0.339
	k_3	0.151	0.118	0.115	0.030	0.450
II	θ_1	271.77	209.24	205.09	79.88	856.36
	θ_2	456.94	211.64	556.71	78.73	2240.62
	θ_3	650.06	424.86	538.86	192.49	2201.55
	β_1	0.01021	0.00902	0.00578	0.00252	0.02462
	β_2	0.01179	0.00911	0.01005	0.00041	0.03658
	β_3	0.01031	0.00759	0.00963	0.00046	0.0368
	τ_1	49.20	48.45	5.53	40.55	59.20
	τ_2	101.05	101.12	4.73	92.18	108.27
	α_1	0.947	0.932	0.139	0.708	1.255
	α_2	0.867	0.859	0.297	0.320	1.452
	α_3	0.867	0.845	0.259	0.431	1.439
III	θ_1	221.80	171.96	153.60	85.79	700.27
	θ_2	184.82	143.24	130.11	89.92	598.77
	θ_3	397.48	331.83	226.51	231.97	1179.89
	β_1	0.01001	0.009	0.00567	0.00202	0.02358
	β_2	0.0097	0.0082	0.0069	0.00108	0.02642
	β_3	0.00638	0.00561	0.00412	0.00074	0.01603
	τ_1	50.00	49.85	5.53	40.79	59.13
	τ_2	102.16	102.68	4.55	92.58	108.37

$\theta_3 \sim Gamma(0.001, 0.001)$; $\beta_1 \sim Gamma(6, 1000)$; $\beta_2 \sim Gamma(1, 1000)$; $\beta_3 \sim Gamma(1, 100)$; $\alpha_1 \sim Gamma(1, 1)$; $\alpha_2 \sim Gamma(1, 1)$; $\alpha_3 \sim Gamma(1, 1)$; $\tau_1 \sim Uniform(2900, 2950)$ and $\tau_2 \sim Uniform(4340, 4360)$. The prior distributions for the parameters k_1, k_2 and k_3 are the exponential distribution with mean parameter 0.99 and truncated at 6 for all three parameters.

Figure 5 presents the empirical and estimated theoretical mean function m(t) assuming the models with no change-points and with one or two change-points. We can observe that the graphs of the function $m(t)$ estimated by the models without any change-points and with one and two change-points fit very well the data set. This result can be observed in the adjusted curves, which overlap the empirical curve.

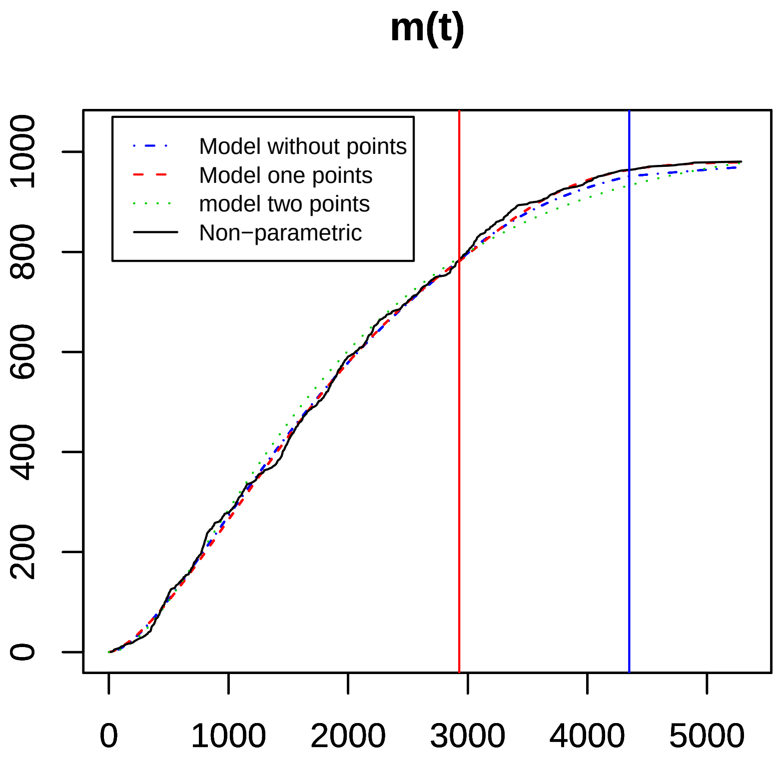

Figure 5: Ozone pollution from NE region of Mexico City. Graphs of the estimates of the mean function $m(t)$, empirical, and estimates by the models without any change-points and with one and two change-points. The vertical lines show the estimated change-points.

Table 3 presents the posterior summaries for the models without any change-point and with one and two change-points for Model I. Prior distributions for the parameters θ_i, β_i, α_i and τ_j, $i = 1, 23$, are presented in the appendix. The prior distributions for the parameters k_i, are those of the second proposed set of distributions, in which we have the prior exponential distribution with mean parameter 0.95 truncated at 3 for the model without change-point and with mean parameter 0.99 truncated at 6 for the models with one and two change-points.

We also use the Deviance Information Criterion (DIC) proposed by Spiegelhalter et al. (2002) to compare the models. This criterion is widely used in Bayesian model selection when samples of posterior distributions of parameters are obtained by MCMC simulation. The DIC selects the more complex models with two change-points, with DIC equal to -2413.32. The DIC for the models with no change-points and with one change-points are, respectively, 4863.14 and -2217.76.

Table 3: Summary of estimates of the posterior distributions for the models without any change-point and with one and two change-points for Model I. Prior distributions for the parameters θ_i, β_i, α_i and τ_j, are presented in Section 3. The assumed prior distributions for the parameters k_i are the exponential distribution with mean parameter equal to 0.95 and truncated at 3 for the model without change-points and mean equal to 0.99 and truncated at 6 for the models with one and two change-points.

Model	parameter	mean	median	S.D.	2.5%	97.5%
without	θ	1035.04	1034.45	35.20	979.42	1093.69
change-points	β	0.0105081	0.0104119	0.0022471	0.0068893	0.0144018
	α	0.744	0.743	0.028	0.700	0.793
	k	2.898	2.925	0.093	2.710	2.995
one	θ_1	994.25	986.88	62.87	903.48	1105.67
change-point	θ_2	348.85	344.48	44.87	282.13	426.69
	β_1	0.0000100	0.0000097	0.0000025	0.0000067	0.0000146
	β_2	0.0000052	0.0000028	0.0000069	0.0000003	0.0000175
	τ	2923.45	2922.35	13.51	2902.91	2946.19
	α_1	1.501	1.502	0.038	1.437	1.561
	α_2	1.810	1.809	0.139	1.591	2.086
	k_1	1.000	0.999	0.045	0.929	1.075
	k_2	5.982	5.987	0.019	5.945	5.999
two	θ_1	1175.47	1166.73	107.61	991.77	1403.03
change-points	θ_2	3416.58	3215.40	1380.97	1331.79	6816.79
	θ_3	2354.08	1951.01	1448.58	601.19	5735.09
	β_1	0.00504	0.00471	0.00221	0.00167	0.0105
	β_2	0.00048	0.00038	0.00042	0.00007	0.00154
	β_3	0.0059	0.00381	0.00613	0.00015	0.023
	τ_1	2924.12	2923.43	13.40	2902.10	2946.76
	τ_2	4350.87	4350.88	4.99	4341.21	4359.29
	α_1	0.796	0.793	0.051	0.709	0.914
	α_2	1.110	1.104	0.111	0.890	1.314
	α_3	0.527	0.518	0.152	0.273	0.853
	k_1	2.298	2.290	0.213	1.917	2.730
	k_2	1.044	0.936	0.571	0.320	2.600
	k_3	0.186	0.139	0.168	0.026	0.597

6. Final Remarks

Considering the application with real data presented in Section 5, associated to air pollution in Mexico City, we observe that the intensity function of a NHPP could increase or decrease depending on the intervention measures adopted by the environmental authorities. In this case the use of NHPP processes with change-points to model the data could be an appropriate way to analyze the data set.

The use of a Bayesian approach based on MCMC methods for this kind of model could be very useful in the analysis of air pollution data. We also observed, from the obtained results of this paper, that sensitivity tests to different choices of prior distributions showed robust inferences, which guarantees the usefulness of the proposed methodology to analyze air pollution data.

The Bayesian sensitivity analysis was based on the effect on the posterior summaries using different choices of the hyperparameters for the prior distributions and also in the convergence of the MCMC algorithm. To illustrate the proposed methodology we also considered simulated data presented in Section 4.

The analysis was performed in two stages. Initially we worked with models without the presence of change-points; these models are of fundamental importance, because the information obtained from them serves to support more complex models which use change-points. The use of the information obtained in previous stages of the Bayesian analysis is important to achieve convergence of the MCMC chains.

The application of the generalized gamma model using Bayesian inference for NHPP with change-points proved to be an excellent tool to analyze the air pollution caused by ozone gas. It was observed that when applying more complex models, with the inclusion of change-points, the improvement in the fit tended to compensate for the increase in the number of parameters. This is also confirmed by the graphs of the MCMC chains of the parameters, the fitting of the curves and the DIC discrimination criterion.

Usually, it is observed that for air pollution data, the change-points occur after a period during which environmental action was taken by the government to reduce pollution levels. This is observed in the example of ozone air pollution in Mexico City presented in Section 5, where from 1990 to 1998 there was a reduction in the use of highly polluting vehicles and, around this period, the model detected a change-point. This was also observed in Achcar et al. (2010). It is noteworthy that, from 1999 to 2002, a series of measures was taken by the environmental authorities in Mexico, with the goal of reducing ozone levels; these measures included manufacturers being encouraged to produce cars with modern clean technology. The year of 2001 was a key date here. Matching this, around 2002 another turning point was detected.

Meteorological variables (temperature, wind speed, etc.) have a significant impact on daily ozone levels. This is beyond the goals of this paper and it is left as a goal of a future study.

Acknowledgments

This work was partially supported by grants from CNPq and FAPESP. We thank Guadalupe Tzintzun for providing the data set and Epifisma Lab (Unicamp). We thank two anonymous reviewers for the valuable comments and suggestions.

Appendix: Priors and MCMC sampling scheme

In the appendix it is presented the prior distributions and the conditional posterior distributions used in the MCMC sampling scheme.

Prior distributions for the model without change point

This work used two sets of a priori distributions for Model I; the first set (1) of a priori distributions are the distributions suggested by Achcar et al. (1998)

$$
\begin{array}{rl}
(i) & N' \sim \text{ Poisson } \{\theta[1 - I_k(\beta T^{\alpha})]\}\,; \\
(ii) & \theta \sim \text{ Gamma } (a, b);\ \text{known a and b}\ ; \\
(iii) & \beta \sim \text{ Gamma } (c, d);\ \text{known c and d}\ ;
\end{array}
$$

$$(iv) \quad \alpha \sim \pi_1(\alpha) \text{ where } \pi_1(\alpha) \propto \frac{1}{\alpha} \text{ for } \alpha(\alpha > 0);$$
$$(v) \quad k \sim \pi_2(k) \text{ where } \pi_2(k) \propto \frac{1}{k} \text{ for } k(k > 0).$$

In the second set (2), the prior distributions for α and k are modified as

$$(iv) \quad \alpha \sim \text{ Gamma } (e, f);\ \text{known e and f };$$
$$(v) \quad k \sim \text{ truncated exponential distribution } (a_n, g);\ \text{known } a_n \text{and g.}$$

$P(\lambda)$ denotes the Poisson distribution with parameter λ, a_n, g are the hyperparameters of the truncated exponential distribution, and a, b, c, d, e and f are known hyperparameters of the gamma distributions where $Gamma(a, b)$ denotes a gamma distribution with expected value $\frac{a}{b}$ and variance $\frac{a}{b^2}$. The choice of the first set of priors for the parameters of the model introduced by Achcar et al (1998) was based in the domain of the parametric space or using very non-informative priors (gamma priors for theta and β and improper priors for α and κ). In the second choice of the priors, the parameters e, f, g and a_n were chosen in order to have very non-informative priors (gamma priors for α, and for κ a truncates exponential such that the distribution is near uniform).

We assume independence of the prior distributions of parameters. The values of the hyperparameters are given in the applications. In the restricted Models II and III we use the same prior distributions for the free parameters.

Posterior distributions for the model without change point

The inference will be conducted based on information supplied by the posterior distribution of the parameters. Assuming independence of the prior distributions, the likelihood function for the Poisson processes, both for the homogeneous case and for the non-homogeneous case, are given as

$$L(\Theta \mid D_T) = \left\{ \prod_{i=1}^{n} \lambda(t_i) \exp\left(- \int_{t_{i-1}}^{t_i} \lambda(x)dx \right) \right\} \exp\left\{ - \int_{t_n}^{T} \lambda(u)du \right\},$$

that is,

$$L(\Theta \mid D_T) = \left\{ \prod_{i=1}^{n} \lambda(t_i) \right\} \exp\left\{ - \int_{0}^{T} \lambda(u)du \right\}.$$

Thus, we have

$$L(\Theta \mid D_T) = \left\{ \prod_{i=1}^{n} \lambda(t_i) \right\} \exp\left\{ -m(T) \right\}. \tag{15}$$

Initially, we assume the prior distribution for k proposed by Achcar et al.(1998). The posterior distribution is given by

$$\begin{aligned} P(N', \alpha, \beta, \theta / D_T) \quad \propto \quad & \frac{\theta^{N'+n+a-1} \alpha^n \beta^{kn+c-1}}{N'! \left\{ \Gamma(k) \right\}^n} \left\{ \prod_{i=1}^{n} t_i^{\alpha k - 1} \right\} \left\{ 1 - I_k(\beta T^\alpha) \right\}^{N'} \\ & \times e^{-(b+1)\theta - (d + \sum_{i=1}^{n} t_i^\alpha)\beta} \pi_1(\alpha) \pi_2(k). \end{aligned} \tag{16}$$

Since the posterior distribution in (16) has no closed form, we resort to MCMC methods to simulate samples of the joint posterior distribution. The algorithm used to obtain posterior distribution samples in (16) is given as follows. For the subset of parameters whose full conditional posterior distributions is known, we sample directly from it, using the Gibbs Sampling algorithm and for: and for the subset of parameters in which the conditional densities are not known, the samples are simulated using the steps of the Metropolis-Hastings algorithm (see Metropolis et al., 1953; Hastings, 1970). Recommended references for the review of MCMC methods are given in Casella and Berger (2001). The required full conditional posterior distributions needed for the MCMC algorithms are given by,
$(i)\ N' \mid \theta, \alpha, \beta, k, D_T \sim Poisson[\theta(1 - I_k(\beta T^{\alpha})];$
$(ii)\ \theta \mid N', \alpha, \beta, k, D_T \sim Gamma[a + n + N', b + 1];$
$(iii)\ \beta \mid N', \alpha, \theta, k, D_T \propto \beta^{kn+c-1} e^{-\beta[\sum_{i=1}^{n} x_i^{\alpha} + d]} \{1 - I_k(\beta T^{\alpha})\}^{N'};$
$(iv)\ \alpha \mid N', \theta, \beta, k, D_T \propto \alpha^{n} \left\{\prod_{i=1}^{n} t_i^{\alpha k-1}\right\} e^{-\beta \sum_{i=1}^{n} t_i^{\alpha}} \{1 - I_k(\beta T^{\alpha})\}^{N'} \pi_1(\alpha);$
$(v)\ k \mid \alpha, N', \theta, \beta, D_T \propto \frac{\beta^{kn}}{\{\Gamma(k)\}^n} \left\{\prod_{i=1}^{n} t_i^{\alpha k-1}\right\} \{1 - I_k(\beta T^{\alpha})\}^{N'} \pi_2(k).$
The parameter θs have closed form distributions, so their posterior distributions are obtained through the Gibbs sampling method. For the other parameters we use the Metropolis-Hastings algorithm.

When we adopt the prior specification of the second set, proposed in Model I, we have that the posterior distribution is given by

$$\begin{aligned} P(N', \alpha, \beta, \theta / D_T) \quad \propto \quad & \frac{\theta^{N'+n+a-1} \alpha^{n+e-1} \beta^{kn+c-1}}{N'! \{\Gamma(k)\}^n} \left\{\prod_{i=1}^{n} t_i^{\alpha k-1}\right\} \{1 - I_k(\beta T^{\alpha})\}^{N'} \\ & \times e^{-(b+1)\theta - (d + \sum_{i=1}^{n} t_i^{\alpha})\beta - f\alpha - a_n k} I_{\{0,g\}}(k), \end{aligned} \tag{17}$$

and the only changes in the full conditional posterior distributions occur in (iv) and (v), which are replaced by
$(iv)\ \alpha \mid N', \theta, \beta, k, D_T \propto \alpha^{n+e-1} \left\{\prod_{i=1}^{n} t_i^{\alpha k-1}\right\} e^{-\beta \sum_{i=1}^{n} t_i^{\alpha} - f\alpha} \{1 - I_k(\beta T^{\alpha})\}^{N'};$
$(v)\ k \mid \alpha, N', \theta, \beta, D_T \propto \frac{\beta^{kn}}{\{\Gamma(k)\}^n} \left\{\prod_{i=1}^{n} t_i^{\alpha k-1}\right\} \{1 - I_k(\beta T^{\alpha})\}^{N'} e^{-a_n k} I_{\{0,g\}}(k).$

Prior and posterior distributions for models with change-points

The prior distributions for the parameters are given as follows:
$(i)\ \tau_j \sim Uniform(f_j, g_j), j = 1, 2;$
$(ii)\ N_i' \sim Poisson[\theta_i(1 - I_{k_i}(\beta_i t_i^{\alpha_i}))], i = 1, 2, 3;$
$(iii)\ \theta_i \sim Gamma(a_i, b_i),\ i = 1, 2, 3,\ a_i, b_i$ known;
$(iv)\ \beta_i \sim Gamma(c_i, d_i),\ i = 1, 2, 3,\ c_i, d_i$ known;
$(v)\ \alpha_i \sim Gamma(e_i, h_i),\ i = 1, 2, 3,\ e_i, h_i$ known;
$(vi)\ k_i \sim$ truncated exponential (m_i, u), $i = 1, 2, 3$ and m_i, u known.

We also assume independence among the prior distributions.

Let N_i' be a latent variable that has Poisson distribution with parameters $P[\theta_i(1 - I_{k_i}(\beta_i \tau_j^{\alpha_i}))]$, and let $\mathbf{\Phi} = (\boldsymbol{w}, N_i')$. The inference is performed using information supplied by the posterior distribution of the parameters. Assuming independence among the prior distributions, in the presence of change-points the posterior and the prior distributions and the likelihood function are related as follows,

$$P(\mathbf{\Phi} \mid D_T) \propto L(D_T \mid \mathbf{\Phi})P(\mathbf{\Phi}), \tag{18}$$

where $P(\mathbf{\Phi} \mid D_T)$ is the joint posterior distribution of $\mathbf{\Phi}$ conditional to data D_T; $P(\mathbf{\Phi})$ corresponds to the prior distributions of parameters and $L(D_T \mid \mathbf{\Phi})$ is the likelihood function. The full conditional posterior distributions needed for the Gibbs sampling algorithm are given by,

- $P(N_i') \sim P[\theta_i(1 - I_{k_i}(\beta_i \tau_j^{\alpha_i}))]$;
- $P(\theta_i) \sim Gamma(N_{\tau_j} + N_i' + a_i - 1, 1 + b_i)$;
- $P(\beta_i) \propto \beta_i^{K_i N_{\tau_j} + c_i - 1} e^{-\beta_i(\sum_{i=1}^{N_{\tau_j}} t_i^{\alpha_i} + d_i)}[1 - I_{k_i}(\beta_i \tau_j^{\alpha_i})]^{N_i'}$;
- $P(\alpha_i) \propto \alpha_i^{N_{\tau_j} + e_i - 1} \left(\prod_{i=1}^{N_{\tau_j}} t_i^{\alpha_i k_i - 1}\right) e^{-\beta_i \sum_{i=1}^{N_{\tau_j}} t_i^{\alpha_i} - h_i \alpha_i}[1 - I_{k_i}(\beta_i \tau_j^{\alpha_i})]^{N_i'}$;
- $P(k_i) \propto \left(\prod_{i=1}^{N_{\tau_j}} \frac{1}{\Gamma(k_i)}\right) \beta_i^{k_i N_{\tau_j}} \left(\prod_{i=1}^{N_{\tau_j}} t_i^{\alpha_i k_i - 1}\right) e^{-m_i k_i} I_{(0,u)} k_i [1 - I_{k_i}(\beta_i \tau_j^{\alpha_i})]^{N_i'}$;
- $P(\tau_j) \propto \left(\prod_{i=1}^{N_{\tau_j}} \frac{1}{\Gamma(k_i)}\right) \theta_i^{N_{\tau_j}} \beta_i^{k_i N_{\tau_j}} \alpha_i^{N_{\tau_j}} \left(\prod_{i=1}^{N_{\tau_j}} t_i^{\alpha_i k_i - 1}\right) e^{-\beta_i \sum_{i=1}^{N_{\tau_j}} t_i^{\alpha_i}}$
$\times \left(\prod_{i=N_{\tau_j}+1}^{N_{\tau_j}} \frac{1}{\Gamma(k_i)}\right) \theta_i^{N_{\tau_j} - N_{\tau_{j-1}}} \beta_i^{k_i N_{\tau_j} - N_{\tau_{j-1}}} \alpha_i^{N_{\tau_j} - N_{\tau_{j-1}}} \left(\prod_{i=N_{\tau_j}+1}^{N_{\tau_j}} t_i^{\alpha_i k_i - 1}\right)$
$\times e^{-\beta_i \sum_{i=N_{\tau_j}+1}^{N_{\tau_j}} t_i^{\alpha_i}} [1 - I_{k_i}(\beta_i \tau_j^{\alpha_i})]\}^{N_i'} [1 - I_{k_i}(\beta_i \tau_j^{\alpha_i}) + I_{k_i}(\beta_i \tau_{j-1}^{\alpha_i})]^{N_i'}$.

The parameters θ_i have distributions in closed form, so their posterior distributions are obtained using Gibbs Sampling, and for the other parameters α_i, β_i, k_i and τ_j the samples are obtained using the Metropolis-Hastings algorithm.

References

Achcar, J. A.; Dey, K.D. and Niverthi, M. – A Bayesian approach using nonhomogeneous Poisson process for software reliability models, In: A.S.Basu; S.K.Basu; S.Mukhopadhyay, (Org.). *Frontiers in Reliability.* ed.1 River Edge, NJ: World Scientific, v. 4, p. 1–18, 1998.

Achcar, J. A.; Rodrigues, E.R.; Paulino, C. D. and Soares, P. – Non-homogeneous Poisson models with a change-point: an application to ozone peaks in Mexico City, *Environmental and Ecological Statistics*, v. 17, p. 521–541, 2010.

Achcar, J. A.; Rodrigues, O.G. and Rodrigues, E.R. – The behaviour of a Metropolis-Hastings algorithm under different prior distributions: an application to ozone measurements in Mexico City, 3^{rd} WSEAS International Conference on Urban Planning and Transportation (UPT'10), Corfu Island, Greece, p. 160–165, 2010.

Achcar, J. A.; Rodrigues, and Tzintzun, G. – Using non-homogeneous Poisson models with multiple change-points to estimate the number of ozone exceedances in Mexico City, *Environmetrics*, v. 22, p. 1–12, 2011.

Air Resource Board (ARB) – Review of air quality standard for ozone in California, Environmental Protection Agency, Staff Report, California, USA, 2005.

Álvarez, L.J.; Fernández-Bremauntz, A.A.; Rodrigues, E.R. and Tzintzun, G. – Maximum a posteriori estimation of the daily ozone peaks in Mexico City, *Journal of Agricultural, Biological and Environmental Statistics*, v. 10, n. 3, p. 276–290, 2005.

Bell, M.L.; Mcdermontt, A.; Zeger, S.L.; Samet, J.M. and Dominici, F. – Ozone and short-term mortality in 95 US urban communities, *Journal of the American Medical Society*, v. 292, p. 2372–2378, 2004.

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

Bell, M.L.; Goldberg, R.; Rogrefe, C.; Kinney, P.L.; Knowlton, K.; Lynn, B.; Rosenthal, J.; Rosenzweweig, C. and Patz, J.A. – Climate change, ambient ozone, and health in 50 US cities, *Climate Change*, v. 82, p. 61–76, 2007.

Braga, A.L.F.; Zanobetti, A. and Schwartz, J. – The effect of weather on respiratory and cardiovascular deaths in 12 U.S. Cities, *Enviromental Health Perspectives*, v. 110, p. 859–863, 2002.

Brooks, Stephen P. and Andrew Gelman. – General methods for monitoring convergence of iterative simulations, *Journal of Computational and Graphical Statistics*, v. 7, p. 434–455, 1998.

Casella, G. and Berger, R. L. – *Statistical Inference*, Belmont, CA: Duxbury Press, second Edition, 2001.

Cox, D.R. and Lewis, P.A. – *Statistical Analysis of Series of Events*, London: Methuen Press, 1966.

Farhat, S.C.L.; Paulo, R.L.P.; Shimoda, L.M.; Conceição, G.M.S.; Lin, C.A.; Braga, A.L.F.; Warth, M.P.N. and Saldiva, P.H.N. – Effect of air pollution on pediatric respiratory emergency room visits and hospital admissions, *Brazilian Journal of Medical and Biological Research*, v. 38, p. 227–235, 2005.

Galizia, A. and Kinney, P.L. – Long-term residence in areas of high ozone: associations with respiratory health in a nationwide sample of nonsmoking young adults, *Environmental Health Perspectives*, v. 107, p. 675–679, 1999.

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.; Peter, J. – The effects of air pollution on lung development from 10 to 18 years of age, *New England Journal of Medicine*, p. 1057–1067, 2004.

Gelman, A. and Rubin, D. R. A single series from the Gibbs sampler provides a false sense of security. Em *Bayesian Statistics* 4 (eds. J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith), Oxford: Oxford University Press, p. 625–631, 1992.

Gelman, A. Inference and monitoring convergence. Em *Markov Chain Monte Carlo in Practice* (eds. W. R. Gilks, S. Richardson and D. J. Spiegelhalter), London: Chapman and Hall, p. 131–143, 1996.

Goel, A.L. and Okumoto, K. – An Analysis Of Recurrent Software Errors In A Real-Time Control System, In Proceedings of ACM Conference, Washington, D.C., USA, p. 496–501, 1979.

Goel, A.L. – A guidebook for software reliability assessment, Technical Report, RADC-TR-83-176, 1983.

Gouveia, N.; Freitas, C.U.; Martins, L.C. and Marcilio, I.O. – Hospitalizações por causas respiratórias associadas à contaminação atmosférica no município de São Paulo, Brasil, *Cadernos Saúde Pública*, v. 22, p. 2669–2677, 2006.

Hastings, W. K. – Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, v. 57, p. 97–109, 1970.

Javits J.S. - Statistical interdependencies in the ozone national ambient air quality standard, *Journal of the Air Pollution Control Association*, v. 30, p. 58–59, 1980.

Leemis, L. M. – Nonparametric estimation of the cumulative intensity function for a nonhomogeneous Poisson process, *Management Science*, v. 37, p. 886–900, 1991.

Loomis, D.P; Borja-Arbuto, V.H; Bangdiwalla, S.I. and Shy, C.M. – Ozone exposure and daily mortality in Mexico City: a time series analysis, Health Effects Institute Research Report, v. 75, p. 1–46, 1996.

Martins, L.C.; Latorre, M.R.D.O.; Cardoso, M.R.A.; Gonçalves, F.L.T.; Saldiva, P.H.N. and Braga, A.L.F. – Poluição atmosférica e atendimentos por pneumonia e gripe em São Paulo, Brasil, *Revista de Saúde Pública*, v. 36, p. 88–94, 2002.

Matthews D.E, Farewell VT. - On testing for a constant hazard against a change-point alternative, *Biometrics*, v. 38, p. 463–468, 1982.

Metropolis, N.; Rosenbluth, A.W.; Rosenbluth, M.N.; Teller, A.H.; Teller, E. –Equations of state calculations by fast computing machines, *Journal of Chemical Physics*, v. 21, p. 1087–1092, 1953.

Raftery A.E. and Akman, V.E. - Bayesian analysis of a Poisson process with a change-point, *Biometrika*, v. 73, p. 85–89, 1986.

Raftery A.E. - 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*, v. 4, p. 378–381, 1989.

Ruggeri F., Sivaganesan S. - On modelling change-points in nonhomogeneous Poisson processes, *Statistical Inference in Stochastic Models*, v. 8, p. 311–329, 2005.

Spiegelhalter, D.J.; Best, N.G.; Carlin, B.P. and Van der Linde, A. – Bayesian measures of model complexity and fit (with discussion), *Journal of the Royal Statistical Society*, B(64), p. 583–639, 2002.

Vicini, L.; Hotta, L. K. and Achcar, J. A. – Non-homogeneous Poisson processes applied to count data: a Bayesian approach considering different prior distributions, *Journal of Environmental Protection*, v. 3. p. 1336-1345, 2012.

Weinberg, J.; Brown, L. D.; Stroud, J. R. - Bayesian forecasting of an inhomogeneous Poisson process with applications to call center data, *Journal of the American Statistical Association*, v. 102, p. 1185-1198, 2007.

Wilson, R., Colome, S.D., Spengler, J.D. and Wilson, D.G. – *Health Effects in Fossil Fuel Burning: Assessment and Mitigation*, Cambridge, USA: Ballinger, 1980.

Yang T.E and Kuo L. - Bayesian binary segmentation procedure for a Poisson process with multiple change-points, *Journal of Computational and Graphical Statistics*, v. 10, p. 772–785, 2001.

Affiliation:

Lorena Vicini
Department of Statistics, UFSM
Avenida Roraima, 1000, CCNE
Santa Maria, RS, Brazil, 97105-900
E-mail: vicini22@gmail.com

Luiz Koodi Hotta
Department of Statistics, UNICAMP
Rua Sérgio Buarque de Holanda, 651
Campinas, SP, Brazil 13083-859
E-mail: hotta@ime.unicamp.br

Jorge Alberto Achcar
Department of Social Medicine
Medical School of Ribeirão Preto
Ribeirão Preto, SP, Brazil, 14049-900
E-mail: achcar@fmrp.usp.br

Journal of Environmental Statistics http://www.jenvstat.org
Volume 5, Issue 3 Submitted: 2012-07-21
August 2013 Accepted: 2013-06-27