

Chapter 3

Poisson Models and Their Application to Ozone Data

3.1 Introduction

In this chapter we address the question of estimating the probability that a given environmental standard is surpassed by a pollutant's concentration a certain number of times within a time interval of interest. In order to do so, we assume a Poisson model for the problem.

Poisson processes are a special case of continuous-time Markov chains and also of counting processes (see for instance [47, 71], Grimmett and Stirzaker (1982)). Such processes are ideal for modeling problems involving counting of occurrences of events. Thus for N_t recording the number of times that an event of interest occurred in the time interval $[0, t]$, $t \geq 0$, let $\mathcal{N} = \{N_t : t \geq 0\}$ be a Poisson process. If $\mathcal{N}$ is a time homogeneous process, then we denote its rate by $\lambda > 0$. In the case of a non-homogeneous Poisson process we denote by $\lambda(t) > 0$ and $m(t) = \int_0^t \lambda(s) ds$ its rate and mean functions, respectively, $t \geq 0$. Note that the time homogeneous case is obtained from the non-homogeneous one by taking $\lambda(t) = \lambda$ and therefore $m(t) = \lambda t$.

Among the properties presented by Poisson processes we have, for $s, t \geq 0$ and $k = 1, 2, \dots$, that

$$P(N_{t+s} - N_t = k) = \frac{[m(t+s) - m(t)]^k}{k!} \exp(-[m(t+s) - m(t)]), \quad (3.1)$$

i.e., the number of events of interest that occur in a time interval $[t, t+s]$ has a Poisson distribution with parameter $[m(t+s) - m(t)]$.

Remarks. 1. In the case of the time homogenous Poisson model, (3.1) becomes

$$P(N_{t+s} - N_t = k) = P(N_s = k) = \frac{(\lambda s)^k}{k!} e^{-\lambda s}, \quad s \geq 0, \quad k = 1, 2, \dots \quad (3.2)$$

2. Note that if we have a non-homogeneous Poisson process, then the time between two occurrences of the event of interest is such that they form an independent sequence and if X_i is the random variable recording the time between the $(i-1)$ th and the i th occurrences of the event, then for t_{i-1} the time of the $(i-1)$ th occurrence of the event of interest, we have that

$$\begin{aligned} P(X_i > s) &= P(N_{t_{i-1}+s} - N_{t_{i-1}} = 0) = \exp(-[m(t_{i-1} + s) - m(t_{i-1})]) \\ &= \exp\left(-\int_{t_{i-1}}^{t_{i-1}+s} \lambda(u) du\right). \end{aligned}$$

Hence, the density of X_i is

$$f_{X_i}(s) = \frac{d}{ds} [1 - P(X_i > s)] = \lambda(t_{i-1} + s) \exp(-[m(t_{i-1} + s) - m(t_{i-1})]).$$

In the case of a time homogeneous Poisson process with rate λ we have that $f_{X_i}(s) = \lambda e^{-\lambda s}$, $s \geq 0$, i.e., X_i has an exponential distribution with parameter λ .

When considering a homogeneous Poisson model, note that the rate of the process is the parameter that needs to be estimated. The rate may be a function of some variables, but it must not depend on time. In the case of a non-homogeneous Poisson process the rate function may be written as a function of some parameters as well as the time. The rate functions we consider in the non-homogeneous case are well-known functions borrowed from the area of reliability theory and survival analysis. Given the nature of the Poisson process, these borrowed functions are very suitable to represent the rate of a non-homogeneous Poisson process.

In the present work the random variables N_t will count the number of times that an environmental threshold is surpassed by a pollutant's concentration in the time interval $[0, t)$, $t \geq 0$.

3.2 Homogeneous Poisson Models

When time homogeneous Poisson models are considered, what we are interested in is estimating the rate $\lambda > 0$. Once we have done that, we may use (3.2) to calculate the probability that a given environmental threshold is surpassed a number of times within a time interval of interest.

Time homogeneous Poisson processes have been used to model several types of problems mainly those of determining the count of occurrences of events. Javits [45] considers that type of process to study the problem of counting the number of surpassings of an environmental threshold by the ozone concentration. The threshold used was 0.12 ppm. The rate λ was set in such a way that the mean number of surpassings of the ozone standard of 0.12 ppm in one year was on average 1 and that of three years was on average equal to 3. Hence, during a non-leap

year, the exceedance rate per day is $1/365$. Therefore, the Poisson process used by Javits [45] to estimate the probability of ozone exceedances in a time interval of interest has rate $1/365$ (see also [61] for a review).

Javits [45] also considers the case where two and three exceedances occur on two and three consecutive days, respectively. Again, in order to estimate the probability of the number of those 2 and 3 days in a row of threshold surpassing, the ozone environmental standard was imposed to obtain the rate of the Poisson process. Take for instance the number of two consecutive days of threshold exceedances in a one-year period. In the case of a non-leap year, there would be about $182 \approx 365/2$ possible pairs (see [61]). Therefore, the rate of the Poisson process counting the number of those days in a year is about $0.00275 \approx 0.5/182$ (since one exceedance is expected in 1 year, 0.5 exceeding pairs of days are expected in that same period).

The estimated values of λ can then be inserted into (3.2) and the appropriate probabilities may be calculated. In [61] are presented computer programs, written in Basic, used to calculate the Poisson probability and distribution.

In the case of Mexico City, the environmental standard limiting exposure to a concentration of 0.11 ppm or above for a period of 1 h or more on average once a year is rarely obeyed. Hence, using that threshold and regulation is not helpful. Nevertheless, a Bayesian point of view may be considered in order to estimate the rate λ of the homogeneous Poisson process. In that direction, Pérez-Muñoz [62] considers λ an unknown quantity and assigns to it a Gamma prior distribution with hyperparameters α and β that are considered to be known and are given by the researcher. Hence, the parameter to be estimated is $\theta = \lambda$. The observed data $\mathbf{D}$ are composed of the number of ozone exceedances of the threshold of interest that occurred in the period of 1 year. Using (1.1) and the fact that a Poisson model is used to count the number of ozone exceedances, we have from [62] that

$$\begin{aligned} P(\theta = \lambda | \mathbf{D} = k) &\propto \frac{e^{-\lambda} \lambda^k}{k!} \frac{\beta^\alpha \lambda^{\alpha-1} e^{-\beta\lambda}}{\Gamma(\alpha)} \\ &\propto \frac{\beta^\alpha \lambda^{\alpha+k-1} e^{-(\beta+1)\lambda}}{\Gamma(\alpha+k)}. \end{aligned}$$

Therefore, the posterior distribution of θ is proportional to a Gamma distribution with parameters $(\alpha + k)$ and $(\beta + 1)$. (This occurs because the family of Gamma densities forms a conjugate family with the likelihood function of Poisson models—see for instance [23,34].) Note that the unit of time is one year; hence, the likelihood function of the model is given by (3.2) setting $s = 1$.

The hyperparameters of the prior distribution of θ are obtained by taking advantage of the fact that the mean and variance of a $\text{Gamma}(\alpha, \beta)$ density are given, respectively, by $\mu = \alpha/\beta$ and $\sigma^2 = \alpha/\beta^2$. Hence, we can calculate the empirical mean and variance of the number of surpassings per year of the threshold of interest during the observed years. For instance, if we have T' years of measurements, then for n_i the number of exceedances of the threshold of interest in the i th year ($i = 1, 2, \dots, T'$), the empirical mean is $\mu = (1/T') \sum_{i=1}^{T'} n_i$.

A similar calculation is given for the empirical variance (see for instance [65]). Then, using the relation between μ and σ^2 and α and β we solve the system of equations to obtain the hyperparameters α and β of the Gamma prior distribution.

In [62] data from the Metropolitan Area of Mexico City were used. They were the overall daily maximum ozone measurements (MAMC) from 1 January 1997 to 31 December 2003. Inference was performed year by year and the thresholds used were 0.11, 0.17, 0.23, and 0.278. The hyperparameters were calculated using the empirical mean and variance obtained in the period 1997–1999 (inclusive) and 2000–2003 (inclusive). Hence, two sets of hyperparameters were used for each threshold considered. The reason for splitting the observational period to calculate the hyperparameters of the prior distributions is that in 1999 the last of a series of environmental measures aiming to reduce the level of ozone in Mexico City was implemented. Hence, changes in the behavior of that pollutant may be observed. (Note that splitting implies that a change-point is forced on the model at that period of time.)

The values used as an estimate for λ was the mode of the posterior distribution (defined for $\alpha > 1$) given by $(\alpha - 1)/\beta$ (see for instance [65]). Whenever the mode was not available the mean of the posterior distribution was used to estimate the rate λ .

As an example of the results obtained by Pérez-Muñoz [62], take the threshold 0.17. There were 208, 209, 183, 172, 135, 105, and 55 days on which this threshold was surpassed in the years 1997, 1998, 1999, 2000, 2001, 2002, and 2003, respectively. The empirical means of the number of surpassings during the periods 1997–1999 and 2000–2003 were, respectively, 200 and 217, and the respective variances were 217 and 2,445.5. The hyperparameters of the prior distribution of λ were thus $\alpha = 184.33$ and $\beta = 0.92$ in the case of years during the period 1997–1999 and $\alpha = 2.22$ and $\beta = 0.21$ in the case of years during the period 2000–2003. Hence, during the years 1997, 1998, 1999, 2000, 2001, 2002, and 2003 the modes of the posterior distribution of $\theta = \lambda$ were 203.64, 204.16, 190.63, 168.53, 133.21, 104.58, and 56.86, respectively. (Note that the value of the mode decreases from year to year.) Therefore, the environmental measures implemented in Mexico City since 1994 to decrease the number of environmental alerts are reflected in the results obtained by the model.

As an example of the calculations that may be performed, if we substitute the value of λ for the year 2000 into (3.2), then we have that the probability of having more than 150 and more than 180 days in the year where the threshold 0.17 is surpassed are approximately 0.8423 and 0.1777, respectively. Hence, the probability of having between 150 and 180 days of surpassings of the threshold 0.17 in the period of 1 year is $P(150 \leq N_t \leq 180) = P(N_t \leq 180) - P(N_t \leq 150) = [1 - P(N_t > 180)] - [1 - P(N_t > 150)] = 0.8233 - 0.1577 = 0.6653$. (Note that the time scale is 1 year, so $s = 1$ in (3.2).)

The drawback here is that in order to use the homogeneous Poisson model, we need to assume homogeneity of the sequence of measurements, which is not always a valid assumption—hence, the splitting into years of the complete

observational period. In order to overcome the shortcoming of time homogeneity, non-homogeneous Poisson models may be considered. These are the subject of study of the next section.

Another way to overcome the problem is to use a middle-term formulation. Hence, we describe the model considered in [8]. In that work the discretized data, assuming observational periods according to approximately the length of time corresponding to the seasons of the year, are taken into account. Then, piecewise homogeneous Poisson processes with different intensity functions for each segment are considered. Note that even though time homogeneous Poisson models are used for each season, some time inhomogeneity is captured when for different seasons a rate function with possibly different parameters may be considered.

Remark. The difference between this approach and the one presented in [62] is that here, in spite of letting the rate of the Poisson model be an unknown quantity, it is not estimated directly. Instead, some parametric forms are considered for λ , and once the parameters are estimated, the rate can be estimated as well. The general formulation of the mathematical model considered in the present case may be described as follows.

Let $N \geq 1$ be a fixed integer number and $a_1, a_2, \dots, a_N$ be positive real numbers. Set $a_0 = 0$. The numbers a_i , $i = 1, 2, \dots, N$, will represent the extremes of the subintervals of the time interval $[0, T]$ where the ozone measurements were taken. Let $I_j = [a_{j-1}, a_j]$, $j = 1, 2, \dots, N$, indicate the subintervals into which $[0, T]$ is split. Let X_j record the number of times the ozone concentration surpasses a given threshold in the time interval I_j , $j = 1, 2, \dots, N$. Assume that the behavior of X_j is explained by a Poisson model with parameter $\lambda_j > 0$, i.e., for $x_j = 0, 1, 2, \dots$,

$$P(X_j = x_j) = \frac{\lambda_j^{x_j} e^{-\lambda_j}}{x_j!}, \quad j = 1, 2, \dots, N. \quad (3.3)$$

The quantities λ_j , $j = 1, 2, \dots, N$, need to be estimated.

Remark. Note that, by the way the distribution of X_j is defined, we have that in the j th time subinterval, the time between two successive exceedances of the threshold of interest has an exponential distribution with mean $1/\lambda_j$, $j = 1, 2, \dots, N$.

Three different parametric forms are considered for the rate λ_j , $j = 1, 2, \dots, N$. The models described are analyzed separately for each region of Mexico City. The variations considered here are given as follows [8]:

Model I. Assume that for $j = 1, 2, \dots, N$, we have

$$\lambda_j = \lambda_0 \kappa^{j-1}, \quad (3.4)$$

where $\lambda_0 > 0$ and $\kappa \in (0, 1)$ are quantities to be estimated.

Model II. This model is a modification of Model I. Here, a non-observable latent variable W_j , $j = 1, 2, \dots, N$, is introduced into (3.4). Thus, for $j = 1, 2, \dots, N$ we assume that λ_j has the form

$$\lambda_j = \lambda_0 W_j \kappa^{j-1}, \quad (3.5)$$

where W_j is a random quantity with a $\text{Gamma}(\phi^{-1}, \phi^{-1})$ prior distribution. The parameter ϕ^{-1} is considered to be an unknown quantity that will be estimated.

Model III. The third model considered here also is a modification of Model I. In the present formulation latent variables W_j , $j = 1, 2, \dots, N$, are also taken into account but now with a different distribution from the one considered in Model II. These variables will also affect the mean of the Poisson model in a different way. Hence, for $j = 1, 2, \dots, N$, we take λ_j given by

$$\lambda_j = \lambda_0 \kappa^{j-1} e^{W_j}, \quad (3.6)$$

where W_j has a $N(0, \sigma_w^2)$ prior distribution. The parameter σ_w^2 also needs to be estimated.

Remarks. 1. Observe that λ_j defined by (3.4) describes a strictly decreasing mean number of ozone exceedances throughout the period of time considered here. This assumption is made to accommodate the expected results associated with the governmental measures aimed at improving the air quality in Mexico City that have been taken during the past 20 years. The smaller the value of λ_j , the larger the mean waiting time between the occurrences of two consecutive ozone exceedances of a given threshold of interest.

2. It is important to point out that the use of models such as Model I has been explored by software researchers (e.g., [55]) to predict the number of bugs in a debugging software test, where each encountered bug is eliminated by correction of the software.
3. Note that when we consider W_j with a distribution $\text{Gamma}(\phi^{-1}, \phi^{-1})$, we have that given ϕ^{-1} , the expected value and variance of W_j are $E(W_j) = 1$ and $\text{Var}(W_j) = \phi$, respectively. Since X_j has a Poisson distribution, we observe that $E(X_j | W_j) = \text{Var}(X_j | W_j) = \lambda_0 W_j \kappa^{j-1}$. Thus, given ϕ^{-1} , we have that $E(X_j) = \lambda_0 \kappa^{j-1}$ and $\text{Var}(X_j) = \lambda_0^2 \kappa^{2(j-1)} \phi + \lambda_0 \kappa^{j-1}$. Assuming that ϕ^{-1} is known, it is possible from the expression for $\text{Var}(X_j)$ to see that, when compared to the value given by Model I, we have an extra Poisson variability given by $\lambda_0^2 \kappa^{2(j-1)} \phi$.
4. Note that by hypothesis, we have that given σ_w^2 , the mean and variance of W_j are given by $E(W_j) = 0$ and $\text{Var}(W_j) = \sigma_w^2$, respectively. Also, observe that $E(X_j | W_j) = \text{Var}(X_j | W_j) = \lambda_0 \kappa^{j-1} e^{W_j}$.
5. Since given σ_w^2 , the quantity W_j has a normal distribution $N(0, \sigma_w^2)$, we have that e^{W_j} has a log-normal distribution with mean $E(e^{W_j}) = e^{\frac{\sigma_w^2}{2}}$ and $\text{Var}(e^{W_j}) = (e^{\sigma_w^2} - 1)e^{\sigma_w^2}$. Thus, if σ_w^2 is known, we have that $E(X_j) = \lambda_0 \kappa^{j-1} e^{\frac{\sigma_w^2}{2}}$ and that

$\text{Var}(X_j) = \lambda_0^2 e^{2(j-1)} (e^{\sigma_w^2} - 1) e^{\sigma_w^2} + \lambda_0 \kappa^{j-1} e^{\sigma_w^2/2}$. We see that for a known value of σ_w^2 , when compared to the value given by Model I, in Model III we also have an extra Poisson variability for $E(X_j)$ which in this case is given by $e^{\sigma_w^2/2}$.

The observed data now are the number of surpassings of a threshold of interest in each time subinterval. Therefore, let n_j be the number of ozone exceedances of a threshold of interest in the time interval I_j , $j = 1, 2, \dots, N$. Then, the observed data set is $\mathbf{D} = \{n_1, n_2, \dots, n_N\}$. The vector of parameters to be estimated in each model is $\theta_I = (\lambda_0, \kappa)$, $\theta_{II} = (\theta_I, \phi)$, and $\theta_{III} = (\theta_I, \sigma_w^2)$ in Models I, II, and III, respectively. Since we are using a Poisson model for the data, the likelihood function is given by

$$L(\mathbf{D} | \theta) \propto \prod_{j=1}^N \lambda_j^{x_j} e^{-\lambda_j}, \quad (3.7)$$

where $\theta = \theta_I, \theta_{II}, \theta_{III}$ depending on the specific model considered. In all cases we consider prior independence of the parameters. The hyperparameters of the prior distributions will be considered known and will be specified when the application of the models is given.

When Model I is taken into account, we assume that λ_0 and κ have $\text{Gamma}(a, b)$ and $\text{Beta}(c, d)$ prior distributions, respectively. Hence, from (1.1) and (3.7), the posterior distribution of θ_I given $\mathbf{D}$ is (see for instance [18])

$$P(\theta_I | \mathbf{D}) \propto \lambda_0^{a+v_1-1} \kappa^{c+v_2-1} (1-\kappa)^{d-1} \exp \left[-\lambda_0 \left(b + \sum_{j=1}^N \kappa^{j-1} \right) \right], \quad (3.8)$$

where $v_1 = \sum_{j=1}^N n_j$ and $v_2 = \sum_{j=1}^N (j-1)n_j$.

In the case of Model II, we consider the same prior distribution for θ_I as in Model I with possibly different values for its hyperparameters. We assume that ϕ will also have as prior distribution a $\text{Gamma}(a', b')$ for all regions with the exception of region SW. In that case ϕ will have a uniform prior distribution. (The change in the prior distribution for this region is due to the fact that there were problems in the convergence of the MCMC algorithm internally implemented in the software WinBugs.)

Therefore, when ϕ has a Gamma prior distribution, the joint posterior distribution of $\theta_{II} = (\theta_I, \phi)$ and $\mathbf{W} = (W_1, W_2, \dots, W_N)$ is given by [8]

$$\begin{aligned} P(\theta_{II}, \mathbf{W} | \mathbf{D}) &\propto \lambda_0^{a+v_1-1} \kappa^{c+v_2-1} (1-\kappa)^{d-1} \phi^{a'-1} e^{-b'\phi} \\ &\times \left(\prod_{j=1}^N W_j^{n_j+\phi-1} \right) \exp \left(-\phi^{-1} \sum_{j=1}^N W_j \right) \\ &\times \exp \left[-\lambda_0 \left(b + \sum_{j=1}^N W_j \kappa^{j-1} \right) \right], \end{aligned} \quad (3.9)$$

where v_1 and v_2 are as in Model I. In the case where ϕ has uniform prior distribution, the only modification made in (3.9) is that the terms $\phi^{a'-1}$ and $e^{-b'\phi}$ do not appear.

When Model III is considered, the parameter θ_I will have the same prior distribution as the one considered in Model I, with possibly different values for its hyperparameters. The random quantity σ_w^2 is assumed to have a Gamma(a'' , b'') prior distribution. Hence, the joint posterior distribution of $\theta_{III} = (\theta_I, \sigma_w^2)$ and $\mathbf{W}$ is given by [8]

$$\begin{aligned} P(\theta_{III}, \mathbf{W} | \mathbf{D}) &\propto \lambda_0^{a+v_1-1} \kappa^{c+v_2-1} (1-\kappa)^{d-1} (\sigma_w^2)^{a''-1} e^{-b''\sigma_w^2} \\ &\times \exp\left(\sum_{j=1}^N W_j n_j\right) \exp\left[-\frac{1}{2\sigma_w^2} \sum_{j=1}^N W_j^2\right] \\ &\times \exp\left[-\lambda_0 \left(b + \sum_{j=1}^N \kappa^{j-1} e^{W_j}\right)\right], \end{aligned} \quad (3.10)$$

where v_1 and v_2 are as in Models I and II.

Posterior summaries of interest are obtained from simulated samples from the joint posterior distribution using the MCMC algorithm internally implemented in the software WinBugs. The model selection is performed using the DIC and the SDM criteria (1.5) as well as visual inspection of the corresponding plots of the quantities of interest (observed and estimated).

The sum of absolute values of the differences between the Monte Carlo estimates for the posterior means $E(\lambda_j | \mathbf{D})$ and the observed mean numbers m_j of ozone exceedances is

$$\text{SDM}(l) = \sum_{j=1}^N |\hat{\lambda}_j^{(l)} - m_j|, \quad (3.11)$$

where $\hat{\lambda}_j^{(l)}$ is the Monte Carlo estimate of $E(\lambda_j | \mathbf{D})$ when using Model l , $l = \text{I, II, III}$ in the j th time subinterval.

Assuming the model that best fits the data, we can obtain accurate inference results for the mean number of exceedances in each interval I_j , $j = 1, 2, \dots, N$. We also can construct Bayesian credible intervals for the differences $\Delta_{(j)} = \lambda_{j+1} - \lambda_j$, $j = 1, 2, \dots, N-1$, between the estimated means of the models that best fit the data in consecutive subintervals to detect multiple change-points. Observe that if zero is not included in a specified Bayesian credible interval for $\Delta_{(j)}$, then that is an indication that the means λ_{j+1} and λ_j are different, and therefore, that is an indication of the possible presence of a change-point.

In order to illustrate the application of Models I, II, and III, we consider the daily maximum ozone data from 1 January 1990 to 31 December 2007. The 18-year average measurements in regions NE, NW, CE, SE, and SW are 0.13, 0.1, 0.136, 0.128, and 0.153, respectively, with respective standard deviations given by

Table 3.1 Values of DIC and $\text{SDM}(i)$, $i = \text{I, II, III}$ for regions NE, NW, CE, SE, and SW when time is split into subintervals corresponding roughly to the seasons of the year

	Model I		Model II		Model III	
	DIC	SDM(I)	DIC	SDM(II)	DIC	SDM(III)
NE	543.263	287.4	295.784	47.6723	308.883	46.2616
NW	844.046	721.351	449.049	84.839	457.584	78.949
CE	851.493	719.457	447.476	84.626	453.959	76.774
SE	721.992	530.496	421.062	77.589	430.045	74.863
SW	786.894	768.095	502.408	112.467	506.366	101.916

0.058, 0.041, 0.056, 0.048, and 0.062. We also have that the Mexican environmental standard for ozone of 0.11 was surpassed on 4,030, 2,863, 4,563, 4,492, and 5,137 days in regions NE, NW, CE, SE, and SW, respectively. The double of the Mexican standard was exceeded on 577, 43, 520, 264, and 1,028 days in the same regions, respectively. Analyses were performed for each region separately. Note that the threshold 0.17 ppm was surpassed on 1,763, 470, 1,892, 1,415, and 2,638 days in regions NE, NW, CE, SE, and SW, respectively.

Here the threshold 0.17 is also used to indicate the presence or absence of an ozone exceedance. The total number of measurements taken is 6,574 and we have $N = 73$ time subintervals. They were formed as follows: the first subinterval corresponds to the months of January and February 1990, the last interval corresponds to the month of December 2007, and the remaining months are split into subintervals 3 months long. Hence, the second subinterval correspond to the period ranging from 1 March 1990 to 31 May 1990 (corresponding to spring of 1990), the third subinterval corresponds to the period ranging from 1 June 1990 to 30 September 1990 (corresponding to summer of 1990), and so on. Therefore, we obtain 73 subintervals, one with length 2 months, one with length 1 month and 71 with length 3 months each.

In all cases, we have used the software WinBugs to simulate samples from the joint posterior distribution of interest, considering a burn-in period of size 5,000. A final Gibbs sample of size 1,000 was obtained taking every 10th generated value in order to have an approximately uncorrelated sample.

In all three models, the hyperparameters used were $a = b = 0.01$ and $c = d = 1$ for the prior distributions of λ_0 and κ , respectively. We also assume that $a' = b' = 0.01$ for the prior distribution of ϕ in Model II, and $a'' = b'' = 1$ for the prior distribution of σ_w^2 in Model III. In the case of region SW and Model II we have ϕ with a uniform prior distribution in $(0, 10)$.

In Table 3.1 [8] we present the values of DIC and $\text{SDM}(l)$, $l = \text{I, II, III}$ for all regions and models.

Looking at Table 3.1 it is possible to see that when the DIC is used to select the best model, the one that best fits the data provided by all regions is Model II (smallest DIC), followed by Model III. However, note that with the exception of region NE, the difference between the DIC for Model II and the DIC for Model III is smaller than 10. Hence, from [21] there is no strong evidence to say that Model

Table 3.2 Posterior mean and 95 % credible intervals of the parameters ϕ , κ , λ_0 , and σ_w^2 for Models II and III and regions NE, NW, CE, SE, and SW

		Mean		95 % credible interval	
		Model II	Model III	Model II	Model III
NE	ϕ	0.9742	—	(0.5998, 1.479)	—
	κ	0.9542	0.9535	(0.9405, 0.9687)	(0.9395, 0.9678)
	λ_0	19.73	12.9	(11.37, 32.16)	(7.296, 20.48)
	σ_w^2	—	0.9788	—	(0.5621, 1.602)
NW	ϕ	0.4058	—	(0.2566, 0.6227)	—
	κ	0.9588	0.957	(0.9505, 0.9671)	(0.9496, 0.9648)
	λ_0	71.4	63.05	(51.48, 97.96)	(45.17, 82.89)
	σ_w^2	—	2.184	—	(1.391, 3.316)
CE	ϕ	0.4132	—	(0.2577, 0.6483)	—
	κ	0.9548	0.9526	(0.9456, 0.9631)	(0.9445, 0.9604)
	λ_0	83.58	73.36	(60.86, 113.8)	(51.54, 97.8)
	σ_w^2	—	2.092	—	(1.246, 3.254)
SE	ϕ	0.4135	—	(0.2539, 0.6389)	—
	κ	0.9553	0.954	(0.9465, 0.9636)	(0.9444, 0.9629)
	λ_0	60.71	51.81	(43.36, 84.25)	(36.61, 71.39)
	σ_w^2	—	2.117	—	(1.255, 3.333)
SW	ϕ	0.2353	—	(0.1491, 0.365)	—
	κ	0.9677	0.9654	(0.9614, 0.9739)	(0.9593, 0.9793)
	λ_0	86.5	83.89	(67.03, 108.4)	(66.15, 104.5)
	σ_w^2	—	3.93	—	(2.431, 5.965)

The symbol “—” is used to indicate that a particular parameter was not part of that specific model

II is the best for regions NW, CE, SE, and SW when using only the DIC to make the decision. If $\text{SDM}(I)$, $I = \text{I, II, III}$ is used to select the most suitable model to explain the behavior of the ozone data of Mexico City, then in all regions the best model is given by Model III, which produces the smallest values of $\text{SDM}(I)$, $I = \text{I, II, III}$. Therefore, we report estimates for the quantities of interest considering both Models II and III for all regions.

In Table 3.2 [8] we have the posterior estimates of the mean as well as the 95 % credible intervals for the parameters of Models II and III and all regions.

Note that both Models II and III produce very similar estimated values for the parameter κ and that they are very close to 1 for all regions. Regarding the other parameter shared by both models, namely the parameter λ_0 , the estimates produced by both Models II and III are such that in most of the cases (except in the case of region CE) they differ by a value smaller than 10. Hence, the difference that might exist between the estimated mean using Model II and the one using Model III is governed mainly by the behavior of the variables W_i , $I = 1, 2, \dots, N$.

In Fig. 3.1 [8] we have the plots of the observed (continuous line) and estimated posterior means for all regions when Models I, II, and III (dashed, dotted, and dash-dotted lines, respectively) are used.

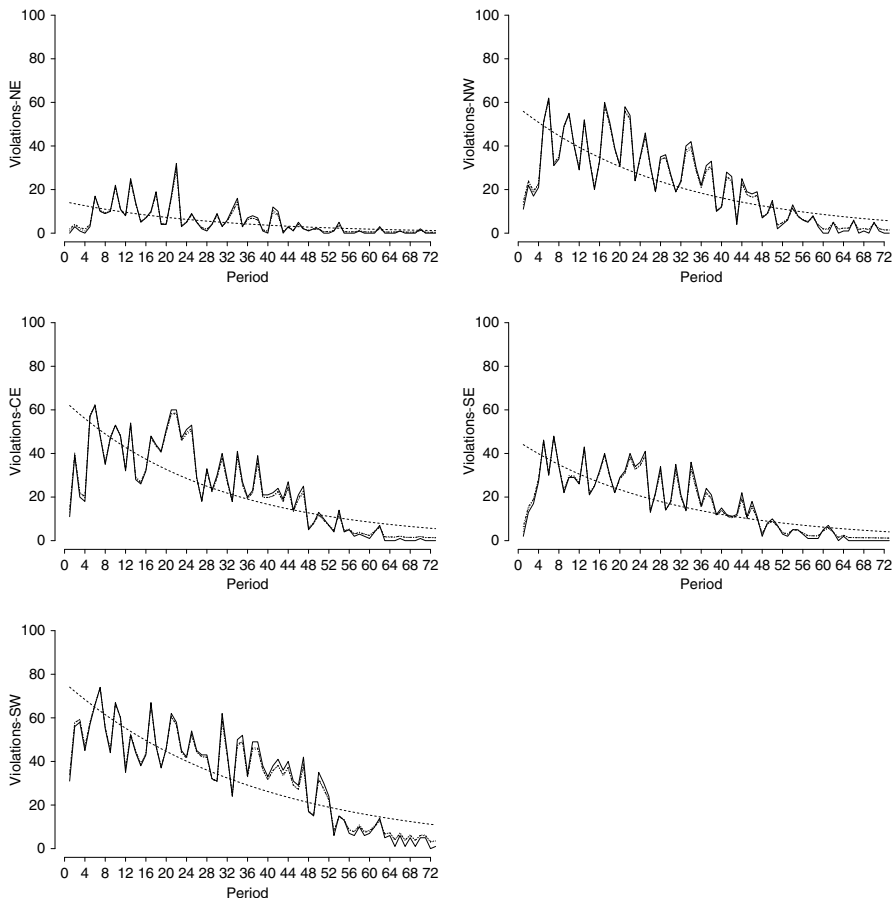


Fig. 3.1 Accumulated observed and estimated means for each rate function for each interval using the threshold 0.17 for regions NE, NW, CE, SE, and SW

By observing Fig. 3.1, it is possible to see an excellent fit of the estimated mean plots using either Model II or III and the observed mean (the lines are practically indistinguishable).

Using the information from Table 3.1 and Fig. 3.1, we have the confirmation that either Model II or Model III is an excellent model to provide information about the behavior of the data provided by the Mexico City monitoring network.

Remark. Note that even though the estimated value of λ_0 is larger in Model II, in the end the behavior of the means λ_i , $i = 1, 2, \dots, N$, is very similar when using parameters estimated by either Model II or Model III. That could be partially explained by the fact that the influence of W_i in λ_i , $i = 1, 2, \dots, N$, is linear in Model II but is exponential in Model III. Hence, the exponential term in Model III might compensate for the larger values of λ_0 in Model II. That difference seems not to have any effect when region CE is considered, though.

Table 3.3 Mean and 95 % credible interval for the ordered differences of the means not including zero for region CE for both Models II and III

$\Delta_{(j)} = \lambda_{j+1} - \lambda_j$		Mean	95 % credible interval	Calendar time
$\Delta_{(1)}$	Model II	27.15	(13.25; 41.6)	Jan, 1990 to May, 1990
	Model III	25.23	(10.38; 39.69)	Jan, 1990 to May, 1990
$\Delta_{(2)}$	Model II	-18.38	(-34.19; -2.591)	Mar, 1990 to Aug, 1990
	Model III	-17.64	(-31.27; -2.792)	Mar, 1990 to Aug, 1990
$\Delta_{(4)}$	Model II	37.41	(19.68; 55.01)	Sep, 1990 to Feb, 1991
	Model III	34.31	(21.07; 55.98)	Sep, 1990 to Feb, 1991
$\Delta_{(12)}$	Model II	20.65	(3.611; 38.02)	Sep, 1992 to Feb, 1993
	Model III	20.54	(3.838; 38.3)	Sep, 1992 to Feb, 1993
$\Delta_{(13)}$	Model II	-24.35	(-40.85; -7.299)	Dec, 1992 to May, 1993
	Model III	-24.51	(-41.46; -8.242)	Dec, 1992 to May, 1993
$\Delta_{(25)}$	Model II	-21.75	(-38.9; -4.677)	Dec, 1995 to May, 1996
	Model III	-22.96	(-39.52; -7.073)	Dec, 1992 to May, 1996
$\Delta_{(27)}$	Model III	13.97	(1.259; 27.06)	Jun, 1996 to Nov, 1996
$\Delta_{(33)}$	Model II	20.44	(6.012; 34.99)	Dec, 1997 to May, 1998
	Model III	21.07	(6.878; 35.77)	Dec, 1997 to May, 1998
$\Delta_{(37)}$	Model II	13.43	(0.2182; 27.07)	Dec, 1998 to May, 1999
	Model III	14.58	(0.05444; 29.1)	Dec, 1998 to May, 1999
$\Delta_{(38)}$	Model II	-15.71	(-29.19; -2.762)	Mar, 1999 to Aug, 1999
	Model III	-16.63	(-30.79; -3.035)	Mar, 1999 to Aug, 1999
$\Delta_{(44)}$	Model II	-10.82	(-22.12; -0.1185)	Sep, 2000 to Feb, 2001
	Model III	-11.5	(-23.13; -0.3275)	Sep, 2000 to Feb, 2001
$\Delta_{(47)}$	Model II	-15.84	(-26.05; -6.854)	Jun, 2001 to Nov, 2001
	Model III	-17.13	(-27.56; -7.701)	Jun, 2001 to Nov, 2001
$\Delta_{(53)}$	Model II	7.362	(0.8528; 15.38)	Dec, 2002 to May, 2003
	Model III	7.702	(1.431; 15.13)	Dec, 2002 to May, 2003
$\Delta_{(54)}$	Model II	-7.325	(-14.4; -1.081)	Mar, 2003 to Aug, 2003
	Model III	-7.965	(-15.381; -1.299)	Mar, 2003 to Aug, 2003
$\Delta_{(62)}$	Model II	-4.687	(-9.72; -0.5274)	Mar, 2005 to Aug, 2005
	Model III	-4.391	(-9.511; -0.3662)	Mar, 2005 to Aug, 2005

As we have seen earlier, either Models II or III may be considered as the most suitable to explain the data gathered from the monitoring network of Mexico City. Using both models for all regions, we have performed inference for the change-points that might be present. We describe in detail only the results for regions CE and SW. The main reason for choosing those regions is that region SW is the one with major ozone problems and region CE may be considered a region that presents an ozone behavior that could be a typical behavior for ozone in the remaining regions. Note that, for simplicity, we are performing the analysis in terms of the mean λ of the Poisson model. However, inference for the mean $1/\lambda$ of the inter-exceedance times may be obtained directly from those results.

First of all, consider region CE. In Table 3.3 [8] we have the posterior mean for the ordered differences between the posterior estimated means λ_{i+1} and λ_i whose 95 % credible interval does not contain zero, $i = 1, 2, \dots, N - 1$, i.e., those

Table 3.4 Mean and 95 % credible interval for the ordered differences of the means not including zero for region SW for both Models II and III

$\Delta_{(j)} = \lambda_{j+1} - \lambda_j$		Mean	95 % credible interval	Calendar time
$\Delta_{(1)}$	Model II	24.16	(6.044; 42.1)	Jan, 1990 to May, 1990
	Model III	23.39	(5.585; 40.49)	Jan, 1990 to May, 1990
$\Delta_{(9)}$	Model II	20.61	(1.153; 40.64)	Dec, 1991 to May, 1992
	Model III	21.48	(1.848; 42.01)	Dec, 1991 to May, 1992
$\Delta_{(11)}$	Model II	-23.81	(-41.81; -4.9)	Jun, 1992 to Nov, 1992
	Model III	-23.02	(-42.07; -3.375)	Jun, 1992 to Nov, 1992
$\Delta_{(16)}$	Model II	22	(2.844; 41.57)	Sep, 1993 to Feb, 1994
	Model III	22.55	(3.134; 42.81)	Sep, 1993 to Feb, 1994
$\Delta_{(30)}$	Model II	27.12	(10.98; 44.59)	Mar, 1997 to Aug, 1997
	Model III	28.71	(10.38; 47.33)	Mar, 1997 to Aug, 1997
$\Delta_{(32)}$	Model II	-17.83	(-32.06; -3.235)	Sep, 1997 to Feb, 1998
	Model III	-18.17	(-33.54; -2.54)	Sep, 1997 to Feb, 1998
$\Delta_{(33)}$	Model II	22.41	(5.906; 38.81)	Dec, 1997 to May, 1998
	Model III	23.09	(8.107; 40.19)	Dec, 1997 to May, 1998
$\Delta_{(47)}$	Model II	-20.34	(-34.13; -8.014)	Jun, 2001 to Nov, 2001
	Model III	-22.05	(-36.37; -8.335)	Jun, 2001 to Nov, 2001
$\Delta_{(48)}$	Model II	16.1	(3.373; 29.19)	Dec, 2001 to May, 2002
	Model III	16.66	(4.05; 30.1)	Dec, 2001 to May, 2002
$\Delta_{(52)}$	Model II	-14.19	(-24.57; -4.751)	Sep, 2002 to Feb, 2003
	Model III	-14.18	(-24.07; -4.911)	Sep, 2002 to Feb, 2003

subintervals where we have an indication of possible presence of a change-point. We also present the calendar time corresponding to the 95 % credible interval of difference of means that possibly contains the change-point.

Observing Table 3.3, we may notice that even though the mean of the Poisson model presents a steadily decreasing behavior from around March 2003, it is possible to see that around the years 1993, 1996, 1999, and also at the end of the year 2001 are the times when there are more meaningful changes in the behavior of the mean of the Poisson model in terms of decreasing behavior, i.e., an increasing behavior of the mean inter-exceedance times. We may notice that there are many possible change-points. Many of these involve days in spring and summer. This could be explained by the extremely dry winters and extremely wet summers that are common in Mexico City, as well as a sunny and warm spring.

In Table 3.4 [8] we have the estimates of the differences of means when data from region SW are taken into account.

Looking at Table 3.4, we observe a decrease in the mean of the Poisson model around June 1992, September 1997, June 2001, and September 2002. Note that the mean of the Poisson models presents a substantial decrease in the year 1992 and also in 2001 (around summer and autumn of 1992 and 2001). However, following a period of decrease between two consecutive means, an increase occurs. We also may notice that the increase is, in general, larger than the decrease.

Observing Tables 3.3 and 3.4, we see that the location of the possible changes of parameters of the models more or less repeat themselves in similar periods in

subsequent years. Table 3.3, for example, shows results for region CE. In both models the possible existing change-points that indicate a decrease in the means of the Poisson models (i.e., an increase in the mean inter-occurrence times) could be located in the following time intervals: from March to August of 1990, 1999, 2003, and 2005; from December 1992, 1995 to May 1993, 1996; from September 2000 to February 2001; and from June to November 2001. The change-points indicating an increase in the means of the Poisson models (and therefore a decrease in the mean inter-exceedance times) could be located in the following time intervals: from January to May 1990; from September 1990 and 1992 to February 1991 and 1993; and from December 1997, 1998, and 2002 to May 1998, 1999, and 2003.

Also from Table 3.3 we may notice that shortly after some of the measures proposed by the environmental authorities were implemented—see a description of some of them in the Introduction of this work—there was a decrease (an increase) in the mean of the Poisson process (mean of the inter-exceedance times). Some of the changes occurred during wintertime and others during spring and summer. Note that the times where an increase between two successive means occurs are approximately during autumn and winter. During the period ranging from 1990 to 1993 the increase in the mean was very large. However, from 1999 to 2001 the amount of decrease is very large when compared to the amount of increase during the end of 2002 and beginning of 2003. Note that for region CE, Model III detects the possible existence of an extra change-point (see $\Delta_{(27)}$) which is not detected when Model II is used. The models and the data suggest that the mean of the Poisson process has had a steady decrease since 2003. Both models also suggest the possible existence of fourteen change-points (with an extra point detected by Model III). A similar pattern can also be seen in the results for the remaining regions.

Looking at Table 3.4, we may observe that when compared to the results for CE, the SW region may have fewer change-points. In this region, the increase between two consecutive means is very large and is more or less of the same order as decreases. It is also possible to notice that half of the instances of decrease in the difference of consecutive means have occurred during summer and autumn and the other half have occurred during the wintertime. They also have occurred shortly after some of the environmental measures were implemented. An increase in two consecutive means occurred mostly during a period belonging either to autumn or to winter.

Observing these results, we could say that the models considered here capture the positive effect that some governmental decisions have had in reducing the level of ozone in Mexico City through reducing the level of ozone precursors released into the atmosphere. Note that even though the environmental measures may have had an effect on reducing the mean of the Poisson model and as a consequence an increase in the mean time between two consecutive exceedances of the threshold of 0.17 ppm, it is common knowledge that the seasons of the year also play an important role in the behavior of ozone air pollution. The models considered here also capture these differences.

In order to use the methodology described here to perform predictions about the occurrence of a surpassing of a given threshold, consider for instance the following.

Take the results for region SW and assume that we are using Model II. Also, consider the last part of the observed data. Looking at Table 3.4, we have that the last change-point occurs in the transition from the 52nd to the 53rd subinterval. Hence, we take $j = 53$. Therefore, from the results given in Table 3.4, the parameter λ is given by $\lambda_{53} = 86.5 \times W_{53} \times 0.9677^{53-1}$, where W_{53} is a random quantity with distribution $\text{Gamma}(4.25, 4.25)$, i.e., $\lambda_{53} = 15.7 \times W_{53}$. Now, we generate a sample of the random quantity W_{53} and from that we have a sample of the parameter λ_{53} . Let $\bar{\lambda}$ be the mean of this sample. Therefore, if no drastic change occurs in the actual environmental conditions, the probability of having $k \geq 0$ exceedances of the threshold 0.17 ppm in the next time subinterval is given by

$$P(X = k) = \frac{\bar{\lambda}^k e^{-\bar{\lambda}}}{k!}, \quad k = 0, 1, 2, \dots$$

If what we are interested in is estimating the probability of the length of the next inter-surpassing time, given that the present one is of length s , then we have the following. Let S_{pres} and S_{next} be the most recent and the next inter-exceedance interval, respectively. Then by the properties of the Poisson process (see for instance [71]), the probability of having S_{next} greater than t is

$$P(S_{\text{next}} > t | S_{\text{pres}} = s) = e^{-\bar{\lambda}t}, \quad t \geq 0.$$

In order to give a numerical example, we have generated a sample of size 1,000 of W_{53} , and from that we have obtained a sample of λ_{53} . The mean $\bar{\lambda}$ is 0.96032. Therefore, the probability of having three surpassings of the threshold 0.17 ppm in the next subinterval is

$$P(X = 3) = \frac{(0.96032)^3 e^{-0.96032}}{3!} \approx 0.06,$$

and the probability of having just one surpassing is 0.38919. Additionally, the next inter-exceedance interval will have a length greater than 10 with probability $e^{-0.96032 \times 10}$, which is approximately 0.00012, and will be greater than 3 with probability approximately equal to 0.067.

Note that the maximum number of times that the threshold 0.17 ppm was surpassed occurred during either spring or winter for all regions. We had 14, 10, 12, 6, and 9 exceedances during spring for regions NE, NW, CE, SE, and SW, respectively, and there were 5, 7, 4, 5, and 5 exceedances during winter. We also have that from the year 2002 the exceedances occurred only during spring for all regions except for region SE, where from the winter 2005–2006, the threshold 0.17 ppm was not surpassed in any of the seasons. Autumn is the season with no exceedances of the threshold in regions NE, NW, and SW. However, there were 2 and 3 exceedances during autumn in regions CE and SE, respectively. There were no exceedances during summer in region NE and there were 2, 1, 6, and 4 exceedances in regions CE, NW, SW, and SE, respectively. The results presented here reflect

these findings when they detect an increase in the mean waiting times between two consecutive exceedances in the later years and also the presence of change-points depending on the seasons of the year.

As was said earlier, another approach to take into account the time inhomogeneity of the data is to use non-homogeneous Poisson processes. This approach is presented in the following section.

3.3 Non-homogeneous Poisson Models

When using non-homogeneous Poisson models, we consider that the rate function of the process is time dependent, and we also consider some parametric forms for that function. In this section, therefore, we present different types of rate functions for the non-homogeneous Poisson process $\mathcal{N}$. These rate functions have been used mostly in the area of reliability theory. Here, they are used to study problems related to air pollution. We present the Bayesian model as well as some of the results obtained when applying the Poisson formulation to the case of Mexico City ozone measurements.

The rate functions considered here are the exponentiated-Weibull (EW) [57, 64], the Musa–Okumoto (MO) [58], the Goel–Okumoto (GO) [39], and a generalized Goel–Okumoto (GGO). These rate functions are given, for $t \geq 0$, as follows:

$$\begin{aligned}\lambda^{(\text{EW})}(t|\theta) &= \frac{\alpha\beta \left[1 - e^{-(t/\sigma)^\alpha}\right]^{\beta-1} e^{-(t/\sigma)^\alpha} (t/\sigma)^{\alpha-1}}{\sigma \left[1 - \left(1 - e^{-(t/\sigma)^\alpha}\right)^\beta\right]}, \\ \lambda^{(\text{GGO})}(t|\theta) &= \alpha\beta\gamma t^{\gamma-1} e^{-\beta t^\gamma}, \\ \lambda^{(\text{MO})}(t|\theta) &= \frac{\beta}{t + \alpha}, \\ \lambda^{(\text{GO})}(t|\theta) &= \alpha\beta e^{-\beta t},\end{aligned}\tag{3.12}$$

where $\alpha > 0$, $\beta > 0$, $\gamma > 0$, and $\sigma > 0$. The mean functions associated with these rate functions are given, respectively, by

$$\begin{aligned}m^{(\text{EW})}(t|\theta) &= -\log \left[1 - \left(1 - e^{-(t/\sigma)^\alpha}\right)^\beta\right], \\ m^{(\text{GGO})}(t|\theta) &= \alpha \left[1 - e^{-\beta t^\gamma}\right], \\ m^{(\text{MO})}(t|\theta) &= \beta \log(1 + t/\alpha), \\ m^{(\text{GO})}(t|\theta) &= \alpha \left[1 - e^{-\beta t}\right],\end{aligned}$$

for $t \geq 0$ and where θ represents the vector of parameters to be estimated. We have that θ is (α, β, σ) in the case of a exponentiated-Weibull rate function, is (α, β, γ) in the case of the generalized Goel–Okumoto rate function, and is (α, β) for the remaining functions.

- Remarks.* 1. Note that other rate functions may also be used. See, for instance, [11], where the Weibull-geometric [15] and the Beta-Weibull [26,32] rate functions are used.
2. When $\beta = 1$ in the exponentiated-Weibull, we have the so-called Weibull (W) rate function (sometimes also referred to as power law and indicated by PLP). In that case, $\lambda^{(W)}(t|\theta) = (\alpha/\sigma)(t/\sigma)^{\alpha-1}$, $m^{(W)}(t) = (t/\sigma)^\alpha$, and $\theta = (\alpha, \sigma)$. Also, note that the Goel–Okumoto rate function is obtained from the generalized Goel–Okumoto by setting $\gamma = 1$.
3. The exponentiated-Weibull rate function is such that (see for instance [2]) if $\alpha \geq 1$ and $\alpha\beta \geq 1$, then $\lambda^{(EW)}(t|\theta)$, $t \geq 0$, is an increasing function; if $\alpha \leq 1$ and $\alpha\beta \leq 1$, then $\lambda^{(EW)}(t|\theta)$, $t \geq 0$, is a decreasing function; if $\alpha > 1$ and $\alpha\beta < 1$, then $\lambda^{(EW)}(t|\theta)$, $t \geq 0$, has a bathtub form; if $\alpha < 1$ and $\alpha\beta > 1$, then $\lambda^{(EW)}(t|\theta)$, $t \geq 0$, is unimodal. Correspondingly, the Weibull rate function is such that for $\alpha < 1$, $\alpha = 1$, and $\alpha > 1$, the rate function $\lambda^{(W)}(t|\theta)$, $t \geq 0$, is a decreasing, constant, and increasing function, respectively.

Since a non-homogenous Poisson model is assumed, for $\mathbf{D} = \{d_1, d_2, \dots, d_K\}$ ($K > 0$, known), the days at which the environmental threshold of interest has been surpassed during the time interval $[0, T)$, ($T > 0$), we have from [27] and [49] that

$$L(\mathbf{D}|\theta) \propto \left[\prod_{i=1}^K \lambda(d_i|\theta) \right] e^{-m(T|\theta)}, \quad (3.13)$$

where $\lambda(t|\theta)$ and $m(t|\theta)$ are the rate and mean functions, respectively, of the Poisson process $\mathcal{N}$.

Remark. In [2] and [70], the expression for the likelihood function has the factor $\exp[-m(d_K|\theta)]$ instead of $\exp[-m(T|\theta)]$. This is so because the observation stopped at the K th surpassing (see [49]).

In order to illustrate the use of non-homogeneous Poisson process, take for instance the case of the exponentiated-Weibull rate function considered in [2]. Then, using (3.13) and $\lambda^{(EW)}(t|\theta)$ defined in (3.12), we have from [2] that

$$L^{(EW)}(\mathbf{D}|\alpha, \beta, \sigma) = \frac{(\alpha\beta)^K \left[\prod_{i=1}^K d_i^{\alpha-1} e^{-(d_i/\sigma)^\alpha} \left(1 - e^{-(d_i/\sigma)^\alpha} \right)^{\beta-1} \right]}{\sigma^{\alpha K} \left[\prod_{i=1}^{K-1} \left(1 - [1 - e^{-(d_i/\sigma)^\alpha}]^\beta \right) \right]}. \quad (3.14)$$

Using (1.1), (3.14), and assuming a Gamma(a_1, b_1) and a Gamma(a_2, b_2) prior distribution for α and σ , respectively, and taking, for instance, a uniform prior

distribution for β , we have that the posterior distribution of the parameter $\theta = (\alpha, \beta, \sigma)$ is

$$P^{(EW)}(\theta | \mathbf{D}) \propto \alpha^{a_1+K-1} \beta^K \sigma^{a_2-K\alpha-1} e^{-(\alpha b_1 + \sigma b_2)} \times \frac{\prod_{i=1}^K \left[d_i^{\alpha-1} e^{-(d_i/\sigma)^\alpha} \left(1 - e^{-(d_i/\sigma)^\alpha} \right)^{\beta-1} \right]}{\prod_{i=1}^{K-1} \left(1 - \left[1 - e^{-(d_i/\sigma)^\alpha} \right]^\beta \right)}. \quad (3.15)$$

(Note that we are assuming prior independence of the parameters of the model. Hence, $P(\theta) = P(\alpha, \beta, \sigma) = P(\alpha) P(\beta) P(\sigma)$.)

The complete marginal conditional posterior distributions of the parameters are [2]

$$\begin{aligned} P^{(EW)}(\beta | \alpha, \sigma, \mathbf{D}) &\propto \exp[K \log(\beta)] f_1(\alpha, \beta, \sigma), \\ P^{(EW)}(\alpha | \beta, \sigma, \mathbf{D}) &\propto \alpha^{a_1-1} e^{-\alpha b_1} f_2(\alpha, \beta, \sigma), \\ P^{(EW)}(\sigma | \beta, \alpha, \mathbf{D}) &\propto \sigma^{a_2-1} e^{-\sigma b_2} f_3(\alpha, \beta, \sigma), \end{aligned}$$

where

$$\begin{aligned} f_1(\alpha, \beta, \sigma) &= \exp \left\{ \sum_{i=1}^K \left[(\alpha - 1) \log(d_i) - \left(\frac{d_i}{\sigma} \right)^\alpha + (\beta - 1) \log \left(1 - e^{-(d_i/\sigma)^\alpha} \right) \right] \right. \\ &\quad \left. - \sum_{i=1}^K \log \left[1 - \left(1 - e^{-(d_i/\sigma)^\alpha} \right)^\beta \right] \right\}, \\ f_2(\alpha, \beta, \sigma) &= \exp[K \log(\alpha) - K\alpha \log(\sigma)] f_1(\alpha, \beta, \sigma), \\ f_3(\alpha, \beta, \sigma) &= \exp[-K\alpha \log(\sigma)] f_1(\alpha, \beta, \sigma). \end{aligned}$$

Due to the complexity of the expressions for the posterior distributions, it is very difficult to obtain information about the parameters directly from them. Hence, we use a sample drawn from the posterior distribution and then use the Law of Large Numbers to estimate the quantities such as mean, variance, and mode. Using that sample, we may also estimate the value of the posterior distribution at specific values. However, obtaining a sample directly from the posterior distribution is a task with a degree of difficulty similar to that of calculating the mean and variance directly from it. Therefore, one solution is the use of MCMC algorithms to draw a sample from each of the complete marginal conditional posterior distributions. The task here is facilitated by the use of the software WinBugs [52, 74]. In that case we only need to specify the likelihood function of the model and the prior distribution of the parameters.

Since very little information on the behavior of the parameters is available, an empirical Bayesian approach is considered. Hence, initially we assume that

Table 3.5 Number of days at which the surpassing of the respective threshold occurred in regions NE, NW, CE, SE, and SW and the overall measurements MAMC

Threshold \ Data Set	NE	NW	CE	SE	SW	MAMC
0.11	1,406	816	1,631	1,581	1,863	2,063
0.17	429	95	433	131	774	980
0.22	64	4	57	38	178	237

$\beta = 1$ and we take uniform prior distributions for α and σ . Then, in a second approach, based on the information provided by the complete marginal conditional posterior distribution of the parameters, a uniform distribution with appropriate hyperparameters is considered for β , and Gamma prior distributions with suitable hyperparameters are assigned to α and σ . Therefore, we have two versions of the model—one assuming $\beta = 1$ and another assuming β an unknown quantity that needs to be estimated.

Remark. Note that when $\beta = 1$ and the prior distributions of α and σ are uniform, then $P^{(EW)}(\beta | \alpha, \sigma, \mathbf{D}) \propto f_1(\alpha, \beta, \sigma)$, $P^{(EW)}(\alpha | \beta, \sigma, \mathbf{D}) \propto f_2(\alpha, \beta, \sigma)$, and $P^{(EW)}(\sigma | \beta, \alpha, \mathbf{D}) \propto f_3(\alpha, \beta, \sigma)$.

In order to illustrate the application of the model described here, consider the daily maximum ozone measurements from 1 January 1998 until 31 December 2004 (a total of $T = 2,557$ days) for each region and also the overall maximum measurements of the city (MAMC). The seven-year average measurements in regions NE, NW, CE, SE, and SW are 0.121, 0.098, 0.126, 0.122, and 0.143, respectively, with standard deviations 0.049, 0.035, 0.046, 0.042, and 0.052. For MAMC we have an average measurement of 0.153 with standard deviation of 0.05. The thresholds considered in [2] were $L = 0.11, 0.17, 0.22$. The number of days on which a surpassing of the thresholds occurred varied according to region and threshold. Those numbers are given in Table 3.5 and, for the remainder of this chapter, they represent the number K in the respective data set $\mathbf{D}$.

Remark. The analysis is performed for each set of data separately, hence the different values of K for different data sets.

Hence, in the first approach ($\beta = 1$) the uniform prior distributions of α and σ are defined on the intervals $(0, 2)$ and $(0, 100)$, respectively. In the second approach (β unknown) β has a uniform prior distribution defined on the interval $(0, 100)$. The parameters of the Gamma prior distributions of α and σ varied according to the region and threshold considered. Since the purpose is to illustrate the application of the model, we consider only the threshold $L = 0.17$. Therefore, in this case the prior distribution of α has the following hyperparameters. When considering regions NW, SE, and NE, the hyperparameters of the Gamma prior distribution are $a_1 = 0.06$, and for regions CE and MAMC we have $a_1 = 0.07$. For region SW we have $a_1 = 0.08$. In all cases we have $b_1 = 0.1$. When considering the case of the prior distribution of σ we have that $b_2 = 0.1$ in all cases. The hyperparameter a_2 was equal to 0.02, 0.12, 0.05, 0.04, 0.04, and 0.03 for data from regions NW, NE, CE, SW, SE and for the MAMC data, respectively.

Considering several lags (from length 0 to 50), the covariance between the values generated by the MCMC algorithm was calculated. Based on these calculations, we have decided to collect every thirtieth generated value to be part of the sample used to estimate the parameters of the model. The burn-in period was 3,000 steps in the first stage of the simulation ($\beta = 1$) and 2,000 steps in the second stage, for all sets of data considered. A sample of size 2,000 was taken for all regions and stages, except for the region NW in the second stage, where a sample of size 1,756 was used.

Since we have two versions of the model ($\beta = 1$ and β unknown), we have used the DIC and the modified BIC criteria (1.4) to select the version that best fits the data. In Table 3.6 [2], we have Monte Carlo estimates for DIC and modified BIC obtained from the simulated Gibbs sample for threshold 0.17 and for all sets of data.

Looking at Table 3.6, note that for all data sets, the selected model is the one assuming $\beta = 1$, i.e., the one considering the Weibull rate function. The estimated values of the posterior mean, standard deviation (indicated by SD), and the 95 % credible interval for each parameter are given in Table 3.7 [2].

Remark. When using the threshold $L = 0.22$ and data from regions NE and SE, the version considering β an unknown quantity is the one selected. In those cases the estimated values of β were 18.93 and 49.15, respectively. In all other cases and thresholds, the selected model was the one assuming $\beta = 1$ (see [2]).

As an example of the type of concrete results that it is possible to obtain from the analysis described in the present work, consider regions CE and SW. These two regions are of importance because of the wind direction in Mexico City (NE to SW) and also because of the fact that some of the ozone precursors are produced in region NE and therefore are carried by the wind in the direction of regions CE and SW. Additionally, as we may recall, region CE can be taken as representative of the other regions with the exception of region SW, which is an extreme situation due to its geographical position.

Consider the period of time from the 1st to the 25th January 2005. We want to know the probabilities of having 5, 10, 15, and 20 days or fewer where an ozone environmental threshold of interest is surpassed. Table 3.8 [2] presents the results obtained. The values are approximate, and whenever the value 1 appears, it is because the probability of having more than that specific number k of days with ozone above the given threshold is of order 10^{-4} or smaller. The calculation of these probabilities was made using the Poisson distribution (3.1) with rate function $\lambda^{(w)}(t | \theta)$, $t \geq 0$, where the parameters α and σ in $\lambda^{(w)}(t | \theta)$ are the mean values estimated using their complete marginal conditional distributions given in Table 3.7.

Remark. Recall that $P(N_{t+s} - N_t \leq k) = \sum_{l=1}^k P(N_{t+s} - N_t = l)$ and that $P(N_{t+s} - N_t > k) = 1 - P(N_{t+s} - N_t \leq k)$. In the present example we have $s = 25$ and $t = 2,557$ in (3.1).

In the case of the threshold 0.11, the estimated values of α for regions SW and CE were 0.93 and 0.91, respectively, and the respective estimated values of σ were 0.82 and 0.79. When considering the threshold 0.22, the estimated values of α for

Table 3.6 Estimated values for DIC and BIC for each model region and also for the MAMC data

	NE		NW		CE		SE		SW		MAMC	
	W	EW	W	EW	W	EW	W	EW	W	EW	W	EW
DIC	771.3	861.2	2,281.7	4,219.5	2,337	4,681.1	1,870.7	2,780.2	3,330.5	7,255.8	3,713.6	9,593.2
BIC	-780.1	-891	-2,292.9	-4,243.5	-2,348.2	-4,727.3	-1,881.4	-2,835.2	-3,342.5	-7,304.9	-3,726.2	-9,615.1

Table 3.7 Posterior mean, standard deviation (indicated by SD) and 95 % credible intervals of the parameters α and σ for the case of the Weibull rate function for all sets of data

		Mean	SD	95 % credible interval
NE	α	0.57	0.05	(0.47, 0.69)
	σ	1.2	0.93	(0.18, 3.7)
NW	α	0.65	0.03	(0.59, 0.71)
	σ	0.23	0.11	(0.08, 0.51)
CE	α	0.7	0.03	(0.63, 0.77)
	σ	0.47	0.2	(0.18, 0.95)
SW	α	0.76	0.03	(0.71, 0.81)
	σ	0.42	0.13	(0.22, 0.74)
SE	α	0.66	0.04	(0.59, 0.73)
	σ	0.45	0.21	(0.14, 0.98)
MAMC	α	0.75	0.03	(0.7, 0.8)
	σ	0.27	0.08	(0.14, 0.47)

Table 3.8 Poisson probabilities of having a number of days equal to or smaller than $k = 5, 10, 15, 20$ in which exceedances of an ozone environmental standard occurred during the first $s = 25$ days of the year 2005 for regions SW and CE

		$s = 25$	$k \leq 5$	$k \leq 10$	$k \leq 15$	$k \leq 20$
SW	11 ppm		0.001	0.073	0.453	0.86
	17 ppm		0.52	0.97	1.0	1.0
	22 ppm		1.0	1.0	1.0	1.0
CE	11 ppm		0.006	0.182	0.678	0.955
	17 ppm		0.93	1.0	1.0	1.0
	22 ppm		1.0	1.0	1.0	1.0

The value 1.0 appears whenever the corresponding probability of having more than k days with the ozone measurement larger than the specific threshold is of order 10^{-4} or smaller

regions SW and CE were 0.64 and 0.65, respectively, and the estimated values of σ were 0.93 and 6.72. (Recall that for both regions and all thresholds the model that best fits the data is the one with known $\beta = 1$.)

The results obtained reflect what happens in reality in Mexico City. In regions CE and SW in the first 25 days of the year 2005, the threshold $L = 0.11$ was surpassed on 20 and 14 days in region SW and CE, respectively. Note that in region SW, if we consider the threshold $L = 0.11$, then there is an approximately 0.8 probability of having 10–20 days on which the environmental standard is exceeded in the first 25 days of the year 2005. Additionally, the probability of having five or fewer days of ozone environmental standard surpassings is only 0.001. Therefore, it is more probable that there will be a larger number of days than a smaller number in which the threshold 0.11 is surpassed. Similar results are also obtained when considering region CE and the first 25 days of the 2005. One explanation for the results given above relies on the fact that the wind direction goes from NE to SW and the production of ozone precursors lies in regions NE and CE.

If the threshold $L = 0.22$ is considered, we have that in the first 25 days of the year 2005 there occurred 1 and 0 days above that threshold for regions SW and CE, respectively. It is possible to see that the probability of having 5–10 days in the first 25 days of the year 2005 surpass the threshold 0.22 is zero. However, there is a large probability (namely one) of having fewer than 5 days on which the double of the environmental standard is exceeded. Similar results are valid for region SW and $L = 0.22$.

Remark. Recall that the value of α for regions SW and CE and threshold 0.11 is very close to one. Hence, a time homogeneous Poisson model could be considered in those cases, and expression (3.2) could be used to estimate the probabilities of the number of surpassings. In those cases the rate λ of the Poisson process (which in the present case is given by $1/\sigma$) would be approximately 1.22 and 1.27 for regions SW and CE, respectively.

It is important to point out that the fit of the non-homogenous Poisson model when an exponentiated-Weibull intensity function (which in the present case corresponds to the Weibull rate function) and the threshold 0.17 are used is a reasonable one. This can be observed in the plots of the accumulated observed and estimated means against time given in Fig. 3.2 [2]. Looking at Fig. 3.2, we may observe a reasonable fit in the beginning of the observational period. However, in the cases of region SW and MAMC data, from about half of the observational period, the fit is not so good, and perhaps we could obtain better results by considering more informative prior distributions for the parameters or by considering other parametric forms for the rate function of the non-homogeneous Poisson model. That option is explored next with a different set of measurements.

We would like to call attention to the fact that the fit of a Poisson model assuming a given rate function may also vary according to the subset of data considered. Next we present an example where some rate functions other than the Weibull are also used as well as a different set of data. Therefore, consider the case studied in [70]. In that work the use of the Weibull, Musa–Okumoto, and the generalized Goel–Okumoto rate functions is shown. The data considered were the daily maximum measurements taken from 1 January 2003 to 31 December 2009 for each region of the Metropolitan Area of Mexico City. The averaged values of the measurements during that period were 0.085, 0.097, 0.098, 0.101, and 0.112 for regions NE, NW, CE, SE, and SW, respectively. The corresponding standard deviations were 0.028, 0.036, 0.036, 0.033, and 0.039. The threshold used was 0.17, and it was surpassed on 13, 78, 53, 40, and 143 days, respectively, in regions NE, NW, CE, SE, and SW (see [44, 70]).

The estimation of the parameters was done using a mixture of Gibbs and Metropolis–Hastings algorithms. The Gibbs sampling and Metropolis–Hastings algorithms were programmed in R. The code is freely available in [70] for the cases of the Musa–Okumoto, generalized Goel–Okumoto, and Weibull rate functions, and it is also given in the Appendix.

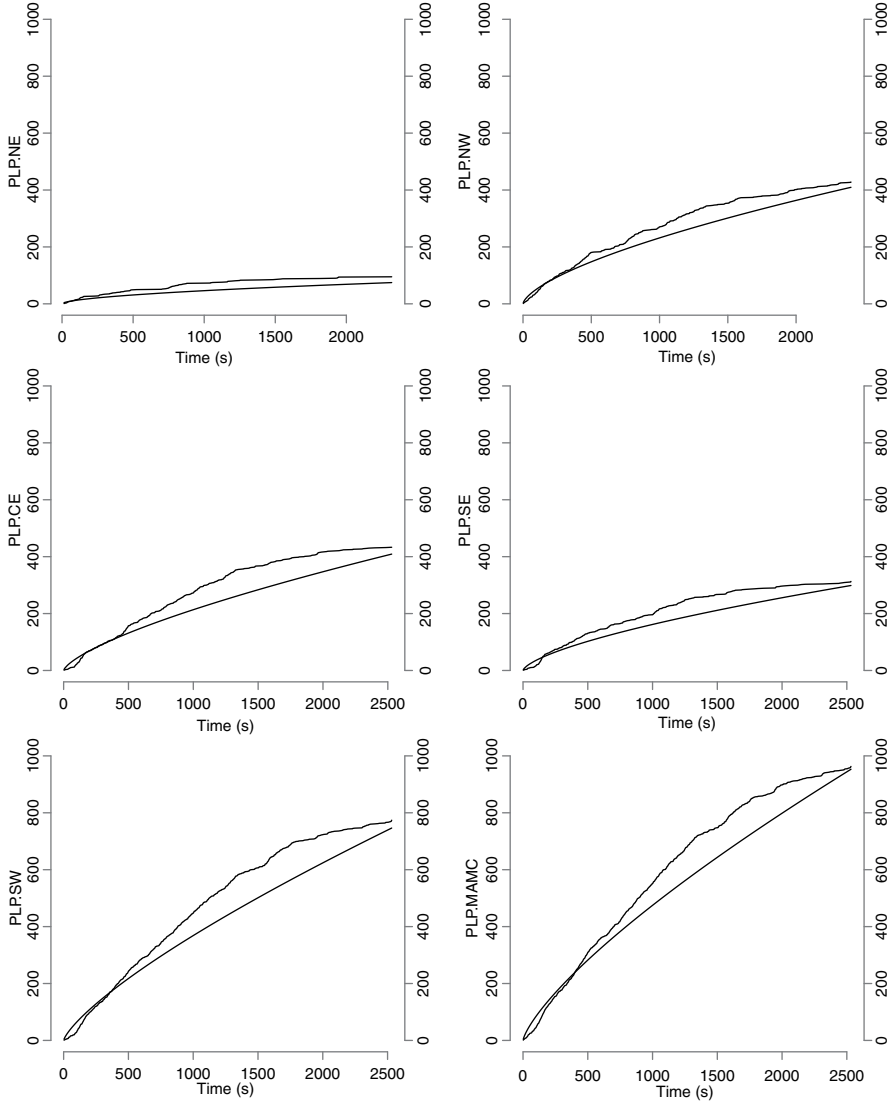


Fig. 3.2 Accumulated observed and estimated means for the Weibull rate function (indicated by PLP in the plots) and the threshold 0.17 for regions NE, NW, CE, SE, and SW and for the MAMC data

Using (1.1) and (3.13), we have that the posterior distributions of θ when considering the Weibull, Musa–Okumoto, and generalized Goel–Okumoto rate functions are given, respectively, by

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

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

and

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

where $P(\alpha, \sigma)$, $P(\alpha, \beta)$, and $P(\alpha, \beta, \gamma)$ are the prior distributions of the corresponding vector of parameters θ in each model (see [70]).

Note that when considering the Weibull, Musa–Okumoto, and generalized Goel–Okumoto rate functions we also have very complex forms for the posterior distributions involved in the analysis. This is the reason for the use of the mixture of the Gibbs sampling and Metropolis–Hastings algorithms. Here we also consider a priori independence of the parameters.

As an example, consider from [70] the Weibull rate function. In that case the complete marginal conditional posterior distributions of the parameters (used in the Gibbs sampling part of the algorithm) are given by

$$P^{(\text{W})}(\alpha | \sigma, \mathbf{D}) \propto \psi_1(\alpha, \sigma) P(\alpha), \\ P^{(\text{W})}(\sigma | \alpha, \mathbf{D}) \propto \psi_2(\alpha, \sigma) P(\sigma),$$

where

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

and

$$\psi_2(\alpha, \sigma) = \exp \left[-K \alpha \log(\sigma) - \left(\frac{d_K}{\sigma} \right)^\alpha \right].$$

The complete marginal conditional posterior distributions are used in the Metropolis–Hastings algorithm in the following way. Consider the case of the parameter α (the case for the parameter σ is similar). First, a proposed value α' is sampled using the prior distribution of α . (The prior distribution is going to play the role of the transition matrix Q presented in the description of the Metropolis–Hastings algorithm given in the Introduction of this work.) Second, calculate $P(\cdot | \sigma, \mathbf{D})$ for the proposed and current values of the parameter α . Finally, calculate (1.2) for the

corresponding values. If the new value of α' is accepted, then use α' in the next step of the Gibbs sampling; otherwise, use the value of α . Let α'' be the accepted value. Proceed with the Gibbs sampling, now for the parameter σ . Begin by generating a value σ' using the prior distribution of σ . Calculate $P(\cdot | \alpha'', \mathbf{D})$ for the proposed and current values of σ and the ratio (1.2). Use the accepted value in the next step of the Gibbs sampling (i.e., repeat the procedure for α , but now with the accepted σ to evaluate the complete marginal conditional distribution of α). Proceed with the Metropolis–Hastings and the Gibbs sampling steps until the convergence of the algorithm is attained. In the case of the other rate functions the procedure is similar.

More formally, we have the following. We sample the proposed values for the parameters using their respective prior distributions and, if θ_j is the parameter in the present step of the Gibbs algorithm, then the acceptance probability in the Metropolis–Hastings algorithm is given by

$$q(\theta_j, \theta'_j) = \min \left\{ 1, \frac{\psi_l(\theta'_j, \theta_{(-j)})}{\psi_l(\theta_j, \theta_{(-j)})} \right\}, \quad j = 1, 2, \dots, n, \quad (3.16)$$

where θ'_j is the proposed parameter sampled using its prior distribution, $\theta_{(-j)}$ is the vector of parameters without its j th coordinate, and ψ_l is the appropriate ψ function that appears in the expression for the complete marginal conditional distributions. (Note that $\theta_{(-j)}$ will have some coordinates that have been updated and some that have not.)

In the case of the rate functions Musa–Okumoto and generalized Goel–Okumoto, the complete marginal conditional posterior distributions are given, respectively, by (see [70] and [44])

$$\begin{aligned} P^{(\text{MO})}(\alpha | \beta, \mathbf{D}) &\propto \psi_3(\alpha, \beta) P(\alpha), \\ P^{(\text{MO})}(\beta | \alpha, \mathbf{D}) &\propto \psi_4(\alpha, \beta) P(\beta), \end{aligned}$$

where

$$\begin{aligned} \psi_3(\alpha, \beta) &= \exp \left[- \sum_{i=1}^K \log(d_i + \alpha) - \beta \log \left(1 + \frac{d_K}{\alpha} \right) \right], \\ \psi_4(\alpha, \beta) &= \exp \left[K \log(\beta) - \beta \log \left(1 + \frac{d_K}{\alpha} \right) \right] \end{aligned}$$

(note that conditioned on α and $\mathbf{D}$, β has as its complete marginal conditional posterior distribution a $\text{Gamma}(K, \log(1 + d_K/\alpha))$ distribution), and

$$\begin{aligned} P^{(\text{GGO})}(\alpha | \beta, \gamma, \mathbf{D}) &\propto \psi_5(\alpha, \beta, \gamma) P(\alpha), \\ P^{(\text{GGO})}(\beta | \alpha, \gamma, \mathbf{D}) &\propto \psi_6(\alpha, \beta, \gamma) P(\beta), \\ P^{(\text{GGO})}(\gamma | \alpha, \beta, \mathbf{D}) &\propto \psi_7(\alpha, \beta, \gamma) \gamma^{a_3-1} e^{-b_3 \gamma}, \end{aligned}$$

Table 3.9 Bayes factor for each model and region considered

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

where

$$\begin{aligned}\psi_5(\alpha, \beta, \gamma) &= \exp \left[K \log(\alpha) - \alpha \left(1 - e^{-\beta d_k^\gamma} \right) \right], \\ \psi_6(\alpha, \beta, \gamma) &= \exp \left[K \log(\beta) + \alpha e^{-\beta d_k^\gamma} - \beta \sum_{i=1}^K d_i^\gamma \right], \\ \psi_7(\alpha, \beta, \gamma) &= \exp \left[K \log(\gamma) + \alpha e^{-\beta d_k^\gamma} - \beta \sum_{i=1}^K d_i^\gamma + (\gamma - 1) \sum_{i=1}^K \log(d_i) \right].\end{aligned}$$

Remarks. 1. In the case of the Weibull rate function, Jara-Ettinger [44] and Rodrigues et al. [70] use two possible distributions as prior distributions for α . In one case a uniform distribution on the interval $(0, 1)$ is used and in another a $\text{Beta}(a, b)$ distribution with appropriate hyperparameters is considered. In the case of the parameter σ either a $\text{Gamma}(c, d)$ or a uniform prior distribution with suitable hyperparameters is considered.

2. Note that since the prior distributions assigned to α are defined on $(0, 1)$, the rate function will have a decreasing behavior. This is justified by the way the data behave (see Fig. 1.2).

Consider now the Weibull rate function. In that case the hyperparameters a and b of the Beta prior distribution were $a = 3/25$ and $b = 2/25$ for regions NW, SE, and SW; $a = b = 1/8$ for region CE; and $a = 7/200$ and $b = 3/200$ for region NE. The uniform distribution for σ is defined on the interval $(0, 10)$ in the case of region CE and $(0, 100)$ for the remaining regions. The hyperparameters of the Gamma prior distribution of σ were $c = 3,249/640$ and $d = 57/640$ in the case of region NE, $c = 3/2$ and $d = 1/2$ for region NW, $c = 9/4$ and $d = 3/4$ for region CE, $c = 1.69$ and $d = 0.13$ for region SE, and $c = 49/20$ and $d = 7/20$ for region SW.

The selection of the model that best fits the data was made using the Bayes factor using expression (1.3). The values obtained are given in Table 3.9 [44, 70].

By looking at Table 3.9, it is possible to see that the selected model was the one that considers the Weibull rate function. The estimation of the parameters was performed using a sample of size 10,500. The burn-in period varied according to region and rate function and was monitored using the Gelman–Rubin test [36].

Table 3.10 Posterior mean, standard deviation (indicated by SD) and 95 % credible intervals of the parameters α and σ when the Weibull rate function is considered

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

In Table 3.10, [44, 70] we have the estimated quantities of interest obtained when using the selected model.

In Fig. 3.3 [44, 70] we have the fitting of the several accumulated estimated means to the accumulated observed means for all the rate functions taken into account.

Observing Fig. 3.3, we may notice that even though, in all cases, the selected model using the Bayes factor was the one assuming the Weibull rate functions, when considering region CE, for instance, the Weibull model does not provide as good a fit as the ones presented by the GGO and MO models. However, when considering region SW, we may see that all rate functions, including the Weibull, provide a good fit, with the best one given by the GGO model. It is worthwhile to call attention to the fact that the fit of the estimated and observed accumulated means varies not only according to region and rate function, but also according to the threshold and subset of the data set that is considered (compare with the Weibull model using a different data set described earlier in this section).

Remark. Another option for generating values to estimate the parameters of the models is to use a Metropolis–Hastings algorithm without the inclusion of the Gibbs sampling step. That path was followed in [60]. In that work the exponentiated-Weibull rate function is considered and the complete posterior distribution (3.15) is used in the ratio (1.2). The proposed values for the parameters are generated using their respective prior distributions. Hence, in that case the ratio (1.2) becomes

$$\alpha(\theta, \theta') = \min \left\{ 1, \frac{L^{(EW)}(\mathbf{D} | \theta')}{L^{(EW)}(\mathbf{D} | \theta)} \right\},$$

where the parameters are considered independent a priori. A case where the prior density of β is dependent of α , is also taken into account. Since the data used in the analysis (same data set as in [2]) presented a decreasing behavior in the number of surpassings as the time passed, a prior distribution for α and β such that $\alpha\beta \leq 1$

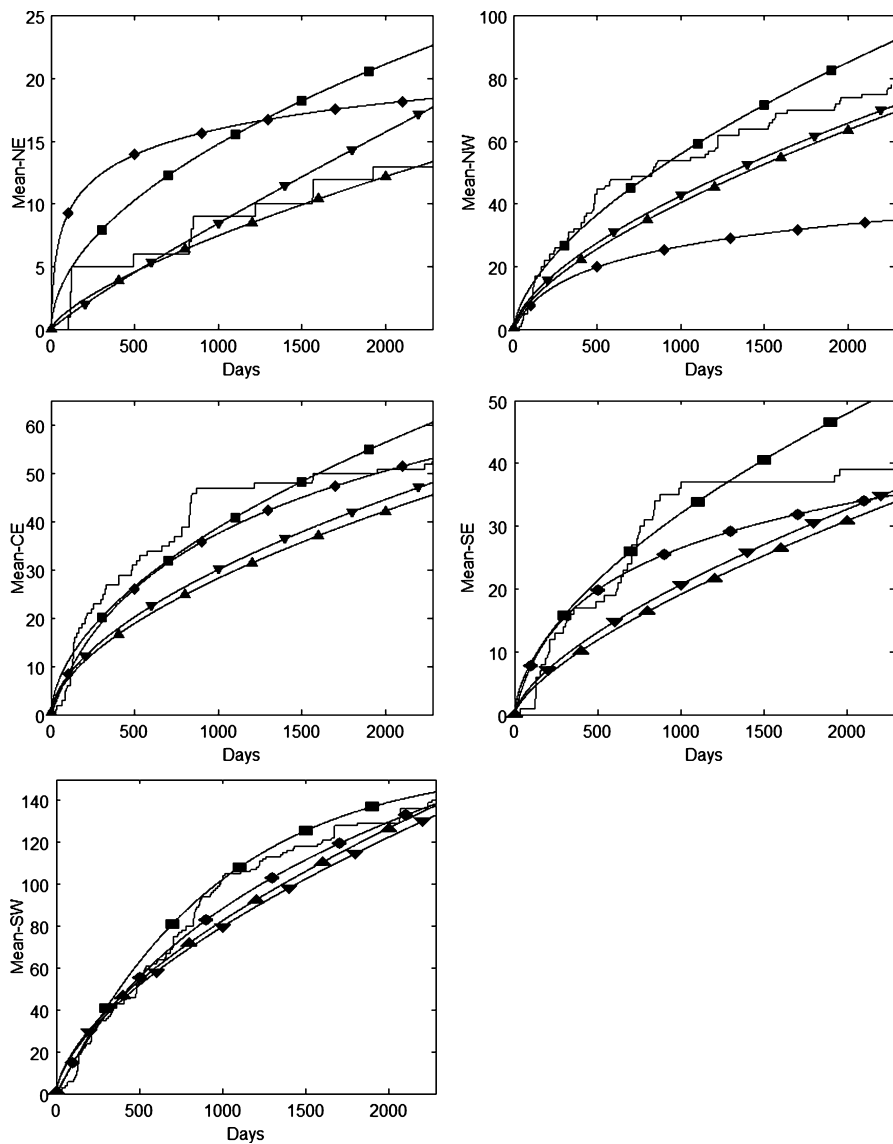


Fig. 3.3 Accumulated observed and estimated means for all regions. The *solid plain line* represents the accumulated observed mean. Lines with *filled square*, *filled diamond*, *filled triangle*, and *filled inverted triangle* correspond to the accumulated estimated mean for models GGO, MO, Weibull where α has a uniform prior, and Weibull where α has a Beta prior, respectively

with $\alpha < 1$ was considered. (Note that condition implied that $\lambda^{(EW)}(t | \theta)$, $t \geq 0$, has a decreasing behavior.) Hence, the prior distribution of θ was $P(\theta) = P(\alpha, \beta, \sigma) = P(\beta | \alpha) P(\alpha) P(\sigma)$, where $P(\alpha)$ is the uniform distribution in $(0, 1)$ and $P(\beta | \alpha)$ is

the $U(0, 1/\alpha)$. Several cases for the prior distributions were also considered for that version of the model. The complete analysis is given in [6]. The MATLAB code for the MCMC algorithm is also given in [6] and in [60].

In order to deal with the poor adjustment between estimated and observed accumulated means, besides using other forms of rate function, we may also allow the presence of change-points in the models. That is considered in the next section.

3.4 Models with the Presence of Change-Points

Looking at Figs. 3.2 and 3.3, we may see that perhaps, in some cases, such as when considering either region NE or region SE in Fig. 3.3 and region SW and MAMC data in Fig. 3.2, it would be more adequate to consider a model where one or more change-points are allowed. Following in that direction, denote by $\tau_1, \tau_2, \dots, \tau_J$ the possible change-points and let $\tau = \{\tau_1, \tau_2, \dots, \tau_J\}$. Since we are dealing with non-homogeneous Poisson processes, the rate function is of the form (see [1, 3, 5, 77])

$$\lambda(t|\theta) = \begin{cases} \lambda(t|\theta_1), & 0 \leq t < \tau_1, \\ \lambda(t|\theta_j), & \tau_{j-1} \leq t < \tau_j, \quad j = 2, 3, \dots, J, \\ \lambda(t|\theta_{J+1}), & \tau_J \leq t \leq T, \end{cases} \quad (3.17)$$

where $\lambda(t|\theta_j)$ and θ_j are, respectively, the rate function and the vector of parameters of the non-homogeneous Poisson model in-between change-points, $j = 1, 2, \dots, J+1$. Set $\theta = (\theta_1, \theta_2, \dots, \theta_{J+1})$. Equivalently, if $m(t|\theta_j)$, $j = 1, 2, \dots, J+1$, are the corresponding mean value functions, we have that

$$m(t|\theta) = \begin{cases} m(t|\theta_1), & 0 \leq t < \tau_1, \\ m(\tau_1|\theta_1) + m(t|\theta_2) - m(\tau_1|\theta_2), & \tau_1 \leq t < \tau_2, \\ m(t|\theta_{j+1}) - m(\tau_j|\theta_{j+1}) \\ \quad + \sum_{i=2}^j [m(\tau_i|\theta_i) - m(\tau_{i-1}|\theta_i)] \\ \quad + m(\tau_1|\theta_1), & \tau_j \leq t < \tau_{j+1}, \quad j = 2, 3, \dots, J, \end{cases} \quad (3.18)$$

where we take $\tau_{J+1} = T$.

- Remarks.* 1. When $J = 0$ the mean and rate functions of the Poisson process become the expressions used in the case where no change-points are allowed.
2. Note that when only one change-point, τ' , is present, we have that (3.17) and (3.18) simplify to [3]

$$\lambda(t|\theta) = \begin{cases} \lambda(t|\theta_1), & 0 \leq t \leq \tau', \\ \lambda(t|\theta_2), & t > \tau', \end{cases}$$

and

$$m(t|\theta) = \begin{cases} m(t|\theta_1), & 0 \leq t \leq \tau, \\ m(\tau|\theta_1) + m(t|\theta_2) - m(\tau|\theta_2), & t > \tau, \end{cases}$$

respectively.

Now, the vector of parameters of the model in the presence of the change-points is $\phi = (\theta, \tau)$, where θ is the vector of parameters associated to the rate function and τ is the vector of possible change-points. Hence, using (1.1), the joint posterior distribution $P(\phi|\mathbf{D})$ of ϕ is

$$P(\phi|\mathbf{D}) \propto L(\mathbf{D}|\theta, \tau) P(\theta, \tau), \quad (3.19)$$

where $L(\mathbf{D}|\theta, \tau)$ is the corresponding likelihood function and $P(\theta, \tau)$ is the joint prior distribution of θ and τ .

In the presence of multiple change-points, the likelihood function has the following explicit form (see [5]):

$$\begin{aligned} L(\mathbf{D}|\phi) &\propto \prod_{i=1}^{N_{\tau_1}} \lambda(d_i|\theta_1) e^{-m(\tau_1|\theta_1)} \\ &\times \left[\prod_{j=2}^J \left(\prod_{i=N_{\tau_{j-1}}+1}^{N_{\tau_j}} \lambda(d_i|\theta_j) e^{-[m(\tau_j|\theta_j) - m(\tau_{j-1}|\theta_j)]} \right) \right] \\ &\times \prod_{i=N_{\tau_J}+1}^{N_T} \lambda(d_i|\theta_{J+1}) e^{-[m(T|\theta_{J+1}) - m(\tau_J|\theta_{J+1})]} \end{aligned} \quad (3.20)$$

(see also [1] and [77]).

Remarks. 1. Note that when no change-point is present, (3.20) becomes (3.13).
2. When only one change-point is present we have that [3]

$$\begin{aligned} L(\theta, \tau|\mathbf{D}_T) &= \left(\prod_{i=1}^{N_{\tau}} \lambda(t_i|\theta_1) \right) \exp[-m(\tau|\theta_1)] \\ &\times \left(\prod_{i=N_{\tau}+1}^{N_T} \lambda(t_i|\theta_2) \right) \exp[-m(T|\theta_2) + m(\tau|\theta_2)]. \end{aligned}$$

In order to illustrate the application of this type of model we are going to consider the case analyzed by Achcar et al. [5]. Hence, we take as possible rate functions the Weibull and the Goel–Okumoto functions. Due to the lack of more precise information on the behavior of the random quantities α , β , and τ when related

to the present data, we have considered a four-stage approach using an empirical Bayes-type analysis.

1. In the first stage, we assume that there are no change-points present in the model. We take α with a uniform prior distribution, with appropriate hyperparameters, in both W and GO models. The parameter σ in the W model has a uniform prior distribution, and the parameter β in the GO model has a Gamma prior distribution.
2. In the second stage, after analyzing the behavior of the accumulated estimated mean, whenever necessary, we allow the presence of a change-point τ_1 . Using the information provided by the case where no change-points are allowed, more informative prior distributions for the random quantities α , β , and σ are used. Hence, we have the following choices. The parameters α_i will have uniform and Gamma prior distributions in the case of the W and GO models, respectively, $i = 1, 2$. The parameters σ_i and β_i , $i = 1, 2$, will have Gamma prior distributions with appropriate hyperparameters. Finally, since no precise information about the change-point is provided in the first stage of the procedure, we assume that τ_1 has a uniform prior distribution defined in an appropriate range.

Remark. One way of choosing appropriately the values at which the prior distribution of the change-point τ will be defined is to look at the plot of the accumulated observed mean and see where there is an indication that its behavior has changed, and then center the support of the prior distribution of τ in the neighborhood of that value allowing for a large variance.

3. In the third stage, in the cases where needed, we use the additional information provided by the model with only one change-point to select the prior distributions for the parameters in the formulation where the presence of two change-points is allowed. Therefore, we have taken a uniform prior distribution for α and σ and for β a Gamma prior distribution. The change-points τ_1 and τ_2 both have appropriate (large variance) normal prior distributions in both W and GO models.
4. Finally, whenever necessary, we include the possibility of the presence of a third change-point. In that case, when the W model is assumed, α_i , σ_i , $i = 1, 2, 3, 4$, will have uniform prior distributions. When the GO model is considered, α_i , β_i , $i = 1, 2, 3, 4$, will have Gamma prior distributions. In both models τ_i , $i = 1, 2, 3$, will have normal prior distributions with appropriate hyperparameters.

The hyperparameters of the prior distributions are assumed to be known and will be specified later. Here we are also assuming prior independence of the parameters.

Remark. We stop at three possible change-points because in the case of the present data that was enough. However, more steps can be added. Another option is to consider the number J of possible change-points an unknown quantity that also needs to be estimated. In that case we may use, for instance, a trans-dimensional MCMC algorithm to jointly estimate J , τ , and θ .

The selection of the model that best fits the data is made by using the DIC. The models are also discriminated by comparing the plots of the accumulated and estimated mean value functions versus time of occurrence.

The sampling of the values of θ and τ is made by using the Gibbs sampling and Metropolis–Hastings algorithm that are internally implemented in the WinBugs software.

The data used in the analysis correspond to 16 years (from 1 January 1990 to 31 December 2005, a total of $T = 5,844$ days) of the daily maximum ozone measurements in each region of Mexico City, as well as the overall daily maximum measurements (MAMC). The 16-year average measurements in regions NE, NW, CE, SE, and SW are 0.136, 0.1, 0.141, 0.132, and 0.158, respectively, with standard deviations 0.058, 0.042, 0.055, 0.048, and 0.062. In the case of the MAMC data we have an average measurement of 0.174 with standard deviation of 0.062.

We are also going to consider the threshold 0.17, which was surpassed on 1,575, 364, 1,689, 2,372, and 1,201 days in regions NE, NW, CE, SW, and SE, respectively. We are going to report only those results obtained when the data from MAMC and region SE are used. The results for the MAMC data are reported because they represent the overall measurements for the Metropolitan Area. Those related to the SE data are reported because they illustrate the case where we need to consider the presence of three change-points.

Convergence of the Gibbs sampling algorithm was monitored by the usual time series plots for the simulated samples and also using the Gelman–Rubin test. Inference was performed using a sample of size 1,000, taken every tenth generated value. When no or one change-point was allowed, the burn-in period was 1,000 steps. In the case where two or three change-points were present, the burn-in period was 5,000 steps. The prior distribution of the parameters involved in each model and for each set of data considered, as well as the results obtained in each case, are given as follows.

Model without the presence of a change-point. When considering both data sets when the W model is considered, we take a $U(0, 1)$ and a $U(0, 100)$ prior distribution for the parameters α and σ , respectively. In the case of the GO model, we take a $U(1, 10,000)$ and a $\text{Gamma}(0.01, 0.01)$ prior distribution for the parameters α and β , respectively. These prior distributions were selected in order to have them approximately non-informative.

The posterior estimates for the parameters, together with the standard deviation and the 95 % credible interval obtained using the model with no change-points, are given in Table 3.11 (see [5]).

In Fig. 3.4 (see [5]), we present the plots of the accumulated observed and estimated mean value functions versus time for the W and the GO models without the presence of change-points for the MAMC and SE data.

Observing Fig. 3.4, we may notice that for the data from region SE the fit produced by the GO model with no change-points is not very good. In the case of the MAMC the fit is not so bad. In the case of the W model, the fit of the accumulated observed and estimated means gets worse in the case of region SE, but improves in the case of the MAMC. Therefore, we go further into the analysis and consider for both models and sets of data the case where a change-point is allowed.

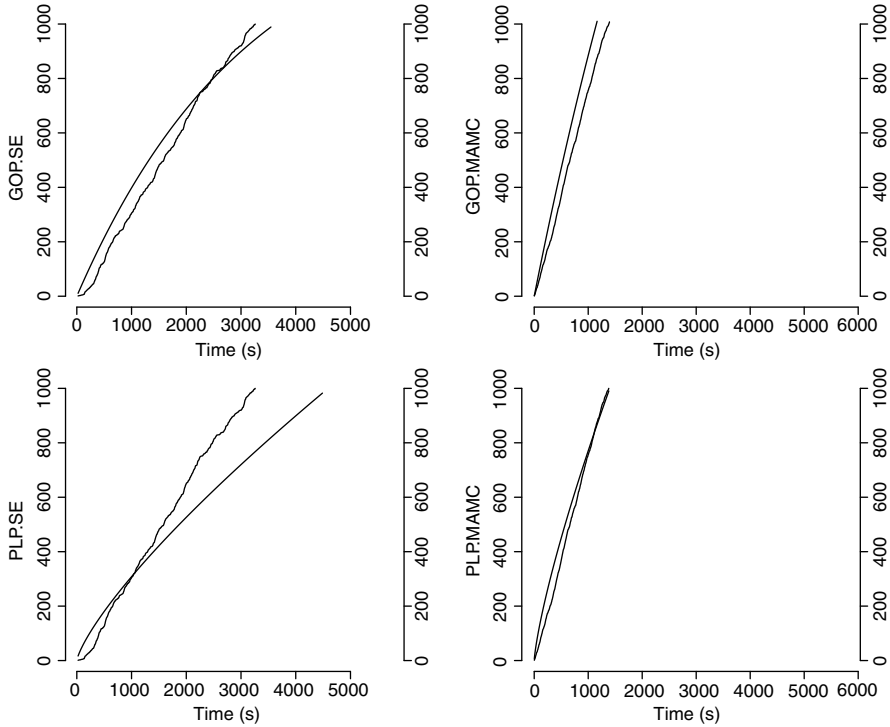


Fig. 3.4 Accumulated number of ozone exceedances and estimated means versus time for region SE and for the MAMC data when the GO (indicated by GOP) and the W (indicated by PLP) models without the presence of change-points are considered. (*Rough lines* represent the observed means and *smooth lines* the estimated means)

Model with the presence of one change-point. In the case of the W model and for both data sets, we assume that α_1 and α_2 have a $U(0, 1)$ prior distribution. We also assume that σ_1 and σ_2 have a $\text{Gamma}(77, 100)$ prior distribution. In the case of the GO model and both data sets, we have that β_1 and β_2 have a $\text{Gamma}(0.01, 1)$ prior distribution. The prior distributions of α_1 and α_2 vary according to the data set. The hyperparameters (a, b) of the Gamma distributions are $(150,000, 100)$ for the data from region SE, and for the MAMC data the hyperparameters are $(40,000, 10)$. In region SE and the MAMC data, and in both W and GO models, the change-point τ_1 has a $U(1, 5, 844)$ prior distribution. (Note that we take the interval at which the prior distribution of the change-point is defined as the entire observational period. In this way we allow the algorithm to place the change-point wherever it feels appropriate.)

Table 3.11 presents the estimates of interest in the case of W and GO models with the presence of one change-point.

Table 3.11 Posterior mean, standard deviation (indicated by SD) and 95 % credible intervals of the parameters of the models W and GO for MAMC and SE data

		Mean		SD		95 % credible interval	
		W	GO	W	GO	W	GO
MAMC							
No change-points	α	0.7758	3,998	0.01333	108.9	(0.7477, 0.8017)	(3,794, 4,217)
	β	0.1901	0.00025	0.03356	0.000014	(0.127, 0.2634)	(0.000226, 0.00027)
One change-point	α_1	0.9187	4,014	0.01051	21.13	(0.8986, 0.9395)	(3,972, 4,055)
	α_2	0.7984	4,000	0.01048	20.08	(0.7786, 0.8189)	(3,960, 4,039)
	β_1	0.7629	0.00024	0.07529	0.0000076	(0.6312, 0.9227)	(0.000226, 0.000256)
	β_2	0.7628	0.000599	0.08654	0.0000815	(0.6132, 0.9501)	(0.00049, 0.00076)
SE region	τ_1	4,261	5,170	10.27	427.3	(4,245, 4,280)	(4,678, 5,634)
No change-points	α	0.7726	1,465	0.02176	51.99	(0.7324, 0.8133)	(1,362, 1,574)
	β	0.6011	0.000317	1.582	0.000018	(0.3525, 0.9343)	(0.00028, 0.00035)
One change-point	α_1	0.3957	1,501	0.06727	3,793	(0.2615, 0.5254)	(1,494, 1,509)
	α_2	0.772	1,500	0.01025	3,899	(0.7519, 0.7928)	(1,493, 1,508)
	β_1	0.7771	0.00028	0.09046	0.000013	(0.6148, 0.971)	(0.00026, 0.00031)
	β_2	0.5631	0.000025	0.06861	0.0000049	(0.4359, 0.7132)	(0.000017, 0.000036)
	τ_1	130.8	4,564	5.775	65.1	(109.5, 135.9)	(4,266, 4,651)
	α_1	0.8328	1,500	0.1302	40.22	(0.5176, 0.9932)	(1,491, 1,576)
Two change-points	α_2	0.7995	1,575	0.03306	34.29	(0.7337, 0.867)	(1,506, 1,648)
	α_3	0.8337	1,500	0.1261	37.94	(0.5738, 0.9956)	(1,421, 1,574)
	β_1	25.68	0.0000309	15.93	0.0000125	(5.376, 67.53)	(0.0000111, 0.0000608)
	β_2	0.6082	0.000298	0.2312	0.0000162	(0.2574, 1.202)	(0.000267, 0.000331)
	β_3	9.981	0.0000254	8.058	0.000049	(0.167, 0.00071625, 77)	(0.0000173, 0.0000359)
	τ_1	132	132.2	4.208	3,375	(121.5, 136)	(124.8, 136.3)
	τ_2	4236	4563	24.67	47.1	(4,176, 4,266)	(4,502, 4,628)

(continued)

Table 3.11 (continued)

	Mean		SD		95 % credible interval			
	W	GO	W	GO	W	GO	W	GO
Three change-points	α_1	0.8213	1.499	0.1402	37.68	(0.5036, 0.9945)	(1.430, 1.576)	
	α_2	0.9137	1.583	0.039	35.37	(0.8354, 0.9864)	(1.517, 1.652)	
	α_3	0.891	1.503	0.09553	38.33	(0.64, 0.9966)	(1.431, 1.580)	
	α_4	0.901	1.498	0.08349	38.89	(0.6882, 0.9968)	(1.423, 1.580)	
	β_1	26.78	0.000099	17.37	0.000016	(4.76, 76.5)	(0.000069, 0.000133)	
	β_2	1.696	0.00032	0.5567	0.0000185	(0.7701, 2.941)	(0.00029, 0.00036)	
	β_3	3.8991	0.000269	2.274	0.000082	(0.1149, 7.688)	(0.000135, 0.000431)	
	β_4	18.58	0.00065	9.281	0.000044	(1.63, 35.78)	(0.000568, 0.00074)	
	τ_1	129.5	296.5	7.714	16.8	(108.1, 135.9)	(244.9, 312)	
	τ_2	3.414	3.505	29.28	95.08	(3.348, 3.484)	(3.336, 3.696)	
	τ_3	4.566	4524	20.23	96.82	(4.528, 4.619)	(4.242, 4.625)	

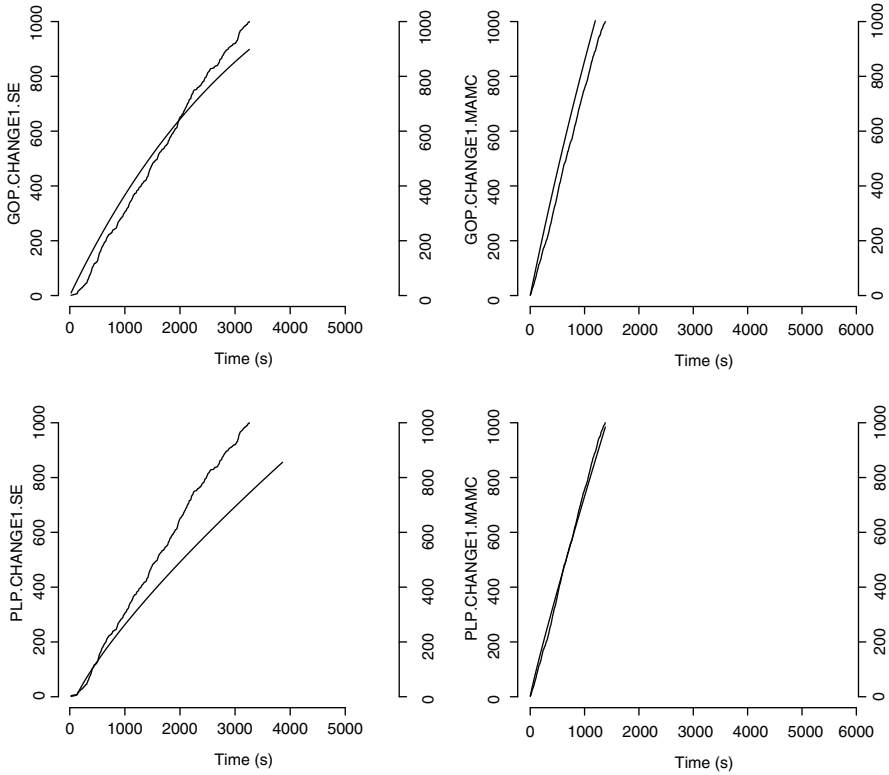


Fig. 3.5 Accumulated number of ozone exceedances and estimated mean value function versus time for region SE and for the MAMC data when the W (indicated by PLP) and the GO (indicated by GOP) models with one change-point are considered. (*Rough lines* represent the observed means and *smooth lines* the estimated means)

In Fig. 3.5 (see [5]) we have the accumulated observed and estimated mean value functions versus time for the GO (indicated by GOP) and W (indicated by PLP) models in the presence of one change-point for the MAMC and SE data.

When looking at Figs. 3.4 and 3.5, we may observe that the W model without a change-point produces a good fit for the MAMC data, but the W model with one change-point gives an almost perfect fit. This may be an indication that the W model with one change-point is the one providing the best fit for the MAMC data. This is confirmed when we use the DIC to select the model. We have that the W model with a change-point has a DIC value of 9,456.6 whereas the W model without a change-point and the GO model without and with a change-point have DIC values 9,867.06, 9,571.85, and 9,484.96, respectively. In the case of the data from region SE, it seems advisable to allow the presence of more than one change-point. Hence, for the SE region, we have decided to go further and to consider both the W and GO models with the presence of two change-points.

Remark. Looking at Table 3.11, we see that the estimated change-points for the W and the GO models in the case of region SE are approximately 130 and 4,565. In the case of the MAMC data they are, respectively, 4,261 and 5,170. Hence, they are placed either in the very beginning of the observational period or at the very end.

If in the W model the mode of the marginal posterior distributions was used to estimate the change-point, then for the MAMC data there are two points with similar frequencies. They are approximately 4,250 and 4,265, with the largest frequency for 4,265, which is very close to the value 4,261 estimated using the mean. In the case of the GO model, the possible candidate using the mode for τ_1 would be 5,620, which is very different from the one using the mean and given in Table 3.11. We would like to point out that, except for the GO model, the values estimated using the mode are very similar to those given in Table 3.11, which were obtained using the mean of the generated values.

Remark. Recall that the mode of the posterior distributions of the parameters may also be obtained by using the MCMC generated values and the Law of Large Numbers.

As noted before, in the case of region SE it seems that further analysis is needed. Therefore, from now on we are going to continue the analysis taking into account only the data from region SE.

Model in the presence of two change-points. The prior distributions assumed for the parameters are as follows. In the case of the W we assume that α_i and σ_i have $U(0, 1)$ and $U(0, 100)$ prior distributions, respectively, $i = 1, 2, 3$. In the case of GO model, the parameters α_i and β_i have $\text{Gamma}(1, 500, 1)$ and $\text{Gamma}(0.1, 1)$ prior distributions, respectively, $i = 1, 2, 3$. In both the W and the GO models, the change-points τ_1 and τ_2 have $N(130, 1,000)$ and $N(4,500, 10,000)$ prior distributions, respectively.

In Table 3.11 we have a summary of the posterior estimates of the parameters together with the standard deviation and the 95 % credible intervals. Observe that the choice of the hyperparameters of the normal prior distributions of the change-points is made by inspection of the plot for the accumulated observed mean value (see for instance Fig. 3.4).

Figure 3.6 [5] shows the plots of the accumulated observed and estimated means versus time for the W and GO models in the presence of two change-points when region SE is considered.

We may observe from Fig. 3.6 that perhaps further change-points should be allowed. Hence, we consider a model in the presence of three change-points.

Model in the presence of three change-points. When considering region SE we assume that for the W model, α_i and σ_i have $U(0, 1)$ and $U(0, 100)$ prior distributions, respectively, $i = 1, 2, 3, 4$. When the GO model is considered, we assume that α_i and β_i have $\text{Gamma}(1, 500, 1)$ and $\text{Gamma}(0.01, 1)$ prior distributions, respectively, $i = 1, 2, 3, 4$. In both models the change-points τ_1 , τ_2 , and τ_3 have $N(130, 10,000)$, $N(3,500, 10,000)$, and $N(4,500, 10,000)$ prior distributions, respectively.

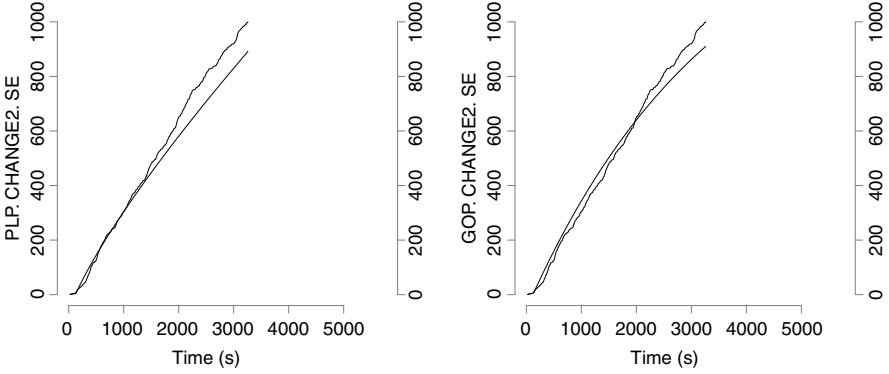


Fig. 3.6 Accumulated number of ozone exceedances and estimated means versus time for region SE when the GO (indicated by GOP) and the W (indicated by PLP) models when two change-points are considered. (*Rough lines* represent observed means and *smooth lines* represent estimated means)

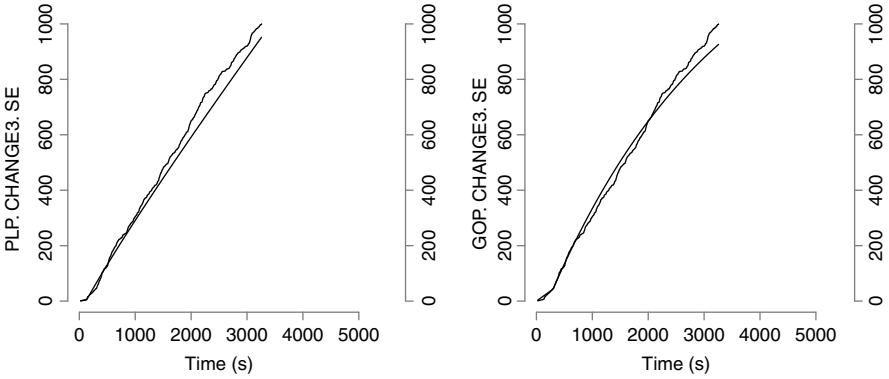


Fig. 3.7 Accumulated number of ozone exceedances and estimated means versus time for region SE when the GO (indicated by GOP) and the W (indicated by PLP) models with three change-points are considered. (*Rough lines* represent observed means and *smooth lines* represent estimated means)

Table 3.11 presents the estimated values of the posterior means of α_i , σ_i , $i = 1, 2, 3, 4$, and τ_i , $i = 1, 2, 3$, as well as the standard deviation and the 95 % credible intervals.

In Fig. 3.7 [5] we have the plots of the accumulated observed and estimated means versus time for the W and GO models when three change-points are present and region SE is considered.

It is possible to see from Fig. 3.7 that there is a very good fit when both models are used. In particular the W model is the one that best fits the data. This is confirmed

Table 3.12 Value of DIC for models W and GO when none, one, two, and three change-points are allowed for the SE data

No change-points		One change-point		Two change-points		Three change-points	
W	GO	W	GO	W	GO	W	GO
6,209.6	5,977.49	6,122.17	5,910.07	5,820.84	5,828.7	5,788.58	5,816.99

when the DIC is used to select the model. In Table 3.12 [5] we have the DIC values for both the W and GO models where no, one, two, and three change-points are allowed and data from region SE is used.

If the mode of the distribution is used to estimate the change-points, the candidate for τ_1 when the W model is considered is 135. In the case of τ_2 the possible value is 3,400. A possible candidate for τ_3 is approximately 4,560. Those values are very close to the ones obtained using the mean which are given in Table 3.11.

Remark. Note that the presence of change-points may be assumed not only for the Poisson models but also for the other models considered in this work. That possibility is discussed further in the next chapter.

Applications of Discrete-time Markov Chains and
Poisson Processes to Air Pollution Modeling and
Studies

Rodrigues, E.R.; Achcar, J.A.

2013, X, 107 p. 12 illus., Softcover

ISBN: 978-1-4614-4644-6