This is a pre-peer-review, pre-copyedit version of an article
published in the journal "Applied Acoustics" (Elsevier).
The final authenticated version is available online at http://dx.doi.org/10.1016/j.apacoust.2014.12.002.

An analysis of airport noise data using a non-homogeneous Poisson model with a change-point

Claudio Guarnaccia[a,1], **Joseph Quartieri**[a], **Carmine Tepedino**[a], **Eliane R. Rodrigues**[b]

[a] *Department of Industrial Engineering, University of Salerno, Via Giovanni Paolo II 132, I-84084 Fisciano, Italy.* E-mails: cguarnaccia@unisa.it, quartieri@unisa,it, ctepedino@unisa.it,

[b] *Instituto de Matemáticas, Universidad Nacional Autónoma de México, Area de la Investigación Científica, Circuito Exterior, Ciudad Universitaria, Mexico DF 04510, Mexico.* E-mail: eliane@math.unam.mx

ABSTRACT

In this work a non-homogeneous Poisson model is considered to study the behaviour of airport noise levels. The model is used to count the number of times that the noise level exceeds a given threshold in a time interval of interest. The rate function of the Poisson process is assumed to be of a Weibull type. Two threshold values are considered. One of them make it evident the change of behaviour of the data. Hence, it requires the use of a model with a change-point. The models considered here depend on some parameters that need to be estimated. After the estimation of the parameters are made, a way of obtaining the probability that the noise threshold is exceeded a certain number of times in a given time interval is presented. Results related to the mean number of exceedances of the threshold are also provided. Those results can be very useful in the study of the

[1] Corresponding author. Department of Industrial Engineering, University of Salerno, Via Giovanni Paolo II 132, I-84084 Fisciano, Italy. Tel.: ++ (39) (89) 969 356. E-mail: cguarnaccia@unisa.it

population's exposure to noise produced by air traffic. The model is also useful to predict, given the current behaviour of the data, the probability of occurrence of high levels of noise in a near future. An application to the airport noise data from Nice International Airport is given.

Keywords: Non-homogeneous Poisson models, Bayesian inference, airport noise, Markov chain Monte Carlo algorithm

1 Introduction

Community noise pollution may cause serious health problems to individuals either living or spending long hours in an environment affect by it (see, for instance, [1, 2, 3, 4, 5]). Among the many hazardous effects of noise pollution we have hearing impairment, sleeping disturbance ([2]) and adverse cardiovascular effects. The latter can occur after long-term exposure to air and/or road traffic noise ([5]). Therefore, it is very important to be able to predict the behaviour of these types of community noise.

There are several ways of measuring sound levels. In the case of continuous noise, such as road traffic noise, the so-called energy average equivalent noise of the A-weighted sound over a period of time T, indicated by $L_{Aeq,T}$, should be used (see [5]). However, when there are particular events such as noise from aircraft or train transit, then the maximum noise level, indicated by L_{Amax}, should also be used together with the $L_{Aeq,T}$ (see [5]). Those types of measures are given as follows ([5]). The $L_{Aeq,T}$ is given by

$$L_{Aeq,T} = 10 \log \left[\frac{1}{T} \int_0^T \frac{p_A^2(t)}{p_0^2} \, dt \right] \tag{1}$$

where $p_A^2(t)$ and p_0^2 represent the square of the A-weighted pressure at time t and the square of the reference pressure, respectively. The L_{Amax} is the maximum value that the A-weighted sound pressure level reaches during a given measurement period ([6]). In the present work, data are taken on a time basis of 0.5 of a second and they have been averaged using (1) with T corresponding to one hour.

The aim in the present work is to provide a way of analysing the behaviour of air traffic noise. There are many questions that might be asked and, depending on their nature, different methodologies may be used. For instance, in [7] several methodologies to assess road traffic noise are reported from a statistical as well as a dynamical point of view. In [8, 9], time series analysis is used to study road traffic noise level. In these works, the aim was to describe the present and past behaviour of the observed data and with that predict how they will behave in the future.

Herein, we are interested in knowing the probability that a noise level exceeds a given threshold in a time interval of interest. Additionally, we are also interested in estimating the mean number of times that a threshold is exceeded in a given time interval. In order to answer these questions, a non-homogeneous Poisson model will be used. Even though, this approach has already been used before ([10]) to study road traffic noise, in the present case we have the additional difficulty that the data considered here change behaviour during the observational period. Hence, in some cases, we have to allow the presence of the so-called change-point. That change implies that some events might have occurred affecting the measurements' behaviour. Hence, we might have lower (higher) levels of noise before a change-point and after that point in time we might have higher (lower) levels. Therefore, once the change-point is identified, it would be possible to point out the cause of the change in behaviour. If we have a good change, then it would be desirable to try to induce the cause in other situations as to repeat the good effects. If the change is a bad one, then actions should be taken in order to revert the bad effect and prevent it from happening in another scenario.

In [10] a non-homogeneous Poisson model with no change-points was applied to noise proceeding mostly from cars. In that case, even though a non-homogeneous model was considered, an approximately homogenous behaviour was detected. Here, we will see that depending on the threshold considered, we might have either a homogeneous or a non-homogeneous behaviour. Additionally, in some cases it will also be necessary to use a model with the presence of a change-point.

Models with change-points have been used in several areas. In the case of environmental applications we have for instance [11, 12] where the models are applied to air pollution data, and [13] with applications to prediction of marine species abundance. A couple of other related studies are [14] with applications to the epidemiological area and [15] applying the models to problems in genetics, medicine and finance.

Since some parameters are present in the models, in order to obtain the results sought, they need to be estimated. That will be performed using the classic statistics point of view ([16, 17, 18]) as well as the Bayesian one ([19]). In the classical sense, estimation will be made using the maximum likelihood approach. In the Bayesian sense, that will be performed through a Markov chain Monte Carlo (MCMC) algorithm ([20]).

This work is presented as follows. In Section 2 the models with and without change-points are described. Section 3 gives the statistical setting. In Section 4 we apply the models to the data from the Nice-Côte d'Azur International Airport, France. Section 5 presents a discussion of the results obtained as well as some comments on future work. An Appendix with the Matlab code of the function that is used in the maximum likelihood procedure is presented before the list of references.

2 Description of the mathematical model

Non-homogeneous Poisson models ([21, 22]) have been used to study problems in several areas of application ([10, 22, 23, 24], among others). The model considered here may be described as follows.

Let $S \geq 1$ and $L > 0$ be natural and real numbers representing the amount of observed data and the threshold of interest, respectively. We also consider a natural number $K \geq 0$ indicating the number of times in $[0, S]$ that the threshold L is surpassed by an observed measurement. Indicate by $N = \{N_t : t \geq 0\}$ the stochastic process ([21]) such that N_t registers the number of times that the noise level has exceeded the threshold L in the time interval $[0, t)$, $t \geq 0$. Assume that N is a non-homogeneous Poisson model with rate and mean functions $\lambda(t) > 0$ and $m(t) = \int_0^t \lambda(s)\, ds$, respectively, $t \geq 0$; i.e., for $t, s \geq 0$, and $k = 0, 1, 2, \ldots$,

$$P(N_{t+s} - N_t = k) \quad = \quad \frac{[m(t+s) - m(t)]^k}{k!} \exp\left(-[m(t+s) - m(t)]\right). \qquad (2)$$

Let $h_1, h_2, \ldots, h_K$ indicate the times (hours in the present case) at which the threshold L was exceeded in the time interval $[0, S]$. Let $\mathbf{D} = \{h_1, h_2, \ldots, h_K\}$ denote the set of those times. The set $\mathbf{D}$ is the set of observed data.

In the present case we assume that $\lambda(t)$, $t \geq 0$ is of the Weibull type; i.e., $\lambda(t) = (\alpha/\sigma)\,(t/\sigma)^{\alpha-1}$, giving the mean function $m(t) = (t/\sigma)^{\alpha}$, $t \geq 0$, $\alpha, \sigma \in (0, \infty)$ (see, for instance, [25]).

Two versions of the model will be considered. In one version a model without change-points is used. Whenever necessary, a second version with the presence of a change-point $\tau \in (0, \infty)$ is considered. When the presence of a change-point is assumed, we have that the rate and mean functions need to be defined for times before and after the change-point.

Hence, the rate function should be of the following form,

$$\lambda(t) = \begin{cases} \lambda_1(t), & t \leq \tau \\ \lambda_2(t), & t > \tau, \end{cases}$$

where $\lambda_i(t) = (\alpha_i/\sigma_i)\,(t/\sigma_i)^{\alpha_i - 1}$, $t \geq 0$, $\alpha_i, \sigma_i \in (0, \infty)$, $i = 1, 2$, and therefore, the mean function is (see, for instance, [11])

$$m(t) = \begin{cases} m_1(t), & t \leq \tau \\ m_1(\tau) + m_2(t) - m_2(\tau), & t > \tau, \end{cases}$$

where $m_i(t) = (t/\sigma_i)^{\alpha_i}$, $i = 1, 2$.

Therefore, when no change-points are allowed, the vector of parameters to be estimated is $\boldsymbol{\theta} = (\alpha, \sigma) \in \mathbb{R}_+^2$. If the presence of a change-point is considered, then the vector of parameters is $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \tau) \in \mathbb{R}_+^5$, where $\boldsymbol{\theta}_1 = (\alpha_1, \sigma_1)$ and $\boldsymbol{\theta}_2 = (\alpha_2, \sigma_2)$ are the vectors of parameters of the rate function when $0 \leq t \leq \tau$ and $\tau < t \leq T$, respectively.

3 Statistical inference

In order to estimate the parameters present in the models, we will use both the classic and the Bayesian statistics points of view. In the case of the classical approach, we will use the maximum likelihood method. Using the Matlab software, in the case without the presence of change-points, the estimation of the value that maximises the likelihood function of the model is relatively easy. However, in some other cases (for instance, when we allow the presence of a change-point) that may not be true. Therefore, we also use the Bayesian point of view. Those two approaches have been used extensively in many areas of applications such as air pollution data, as in [11, 12, 24, 26]; genetics, as in [27, 28];

image recovery ([29]); neurosciences, as in [30]; crystallography, see [31]; evolutionary biology, as in [32]; and phylogenetic analysis ([33]), among others.

When using the Bayesian approach, estimation of the parameters is performed through a sample obtained from the respective posterior distribution. The posterior distribution of a parameter $\boldsymbol{\theta}$, denoted by $P(\boldsymbol{\theta} \,|\, \mathbf{D})$, is such that $P(\boldsymbol{\theta} \,|\, \mathbf{D}) \propto L(\mathbf{D} \,|\, \boldsymbol{\theta})\, P(\boldsymbol{\theta})$, where $P(\boldsymbol{\theta})$ is the prior distributions of $\boldsymbol{\theta}$, $L(\mathbf{D} \,|\, \boldsymbol{\theta})$ is the likelihood function of the model, and $\mathbf{D}$ the set of observed data.

Since a Poisson model is assumed, we have that the likelihood function in the case of no change-points, is given by (see, for example, [16, 17])

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

where $\lambda(t)$ and $m(t)$ are the rate and mean functions, respectively, $t \geq 0$. When the presence of a change-point is allowed we have (see, for instance, [11, 12, 24, 34]), that the likelihood function takes the form,

$$L(\mathbf{D} \,|\, \boldsymbol{\theta}) \;\; \propto \;\; \left(\prod_{i=1}^{N_\tau} \lambda_1(h_i)\right) e^{-m_1(\tau)} \left(\prod_{i=N_\tau+1}^{N_S} \lambda_2(h_i)\right) e^{-[m_2(S)-m_2(\tau)]},$$

where $\lambda_j(t)$ and $m_j(t)$ are the corresponding rate and mean functions, $j = 1, 2$, $t \geq 0$.

When using the Bayesian point of view, we will assume a prior independence of the parameters. We also assume that the hyperparameters of the prior distributions are known and they will be specified when applying the model to the data.

3.1 Model Selection

The selection of the model that best explains the behaviour of the data will be made using a graphical comparison, the sum of the absolute differences (SAD) between the observed and estimated values of the accumulated means as well as using the coefficient of variation (CV) of a sample. The sum of the absolute differences is defined as $\mathrm{SAD} = \sum_i |p(i) - \hat{p}(i)|$, where $\hat{p}(i)$ is the estimated value of the observed $p(i)$. Hence, in our case we have that $p(\cdot) = m(\cdot)$ and $\hat{p}(\cdot) = \hat{m}(\cdot)$, where $\hat{m}(\cdot)$ is the mean function when we use the respective estimated parameters, and where $m(\cdot)$ is the observed mean. Therefore, in the present case we have that

$$\mathrm{SAD} = \sum_{i=1}^{K} |m(h_i) - \hat{m}(h_i)| \,.$$

The coefficient of variation is defined by $CV = \sigma''/\mu$, where here, σ'' is the standard deviation of the sample $\{|m(h_i) - \hat{m}(h_i)| : 1 = 1, 2, \ldots, K\}$, and μ is its mean.

4 Application to the Nice airport data

Nice International Airport is located in Côte d'Azur, South of France. The location is sought by a large number of individuals that want to spend their holidays in the region. The data considered here correspond to measurements taken from 4:48pm of 21 December 2000 to 9:48pm of 01 January 2001; i.e., the period around Christmas and New Year's Eve holidays. That gives a total of 249 observed values. The data may be obtained from the dBTrait software sample data base (used with permission). Figure 1 shows the plots of the hourly measurements during the period considered.

The minimum and maximum measurements are 29.4 decibels (29.4dB) and 74.7dB, respectively. The data considered have an average value of 55.1dB with a standard deviation of 11.2dB. Two threshold values are taken into account. In a first instance, we take $L = 55$dB. Later on, we use $L = 60$dB. During the observational period there were 157 and 98 hours in which the thresholds 55dB and 60dB were exceeded, respectively.

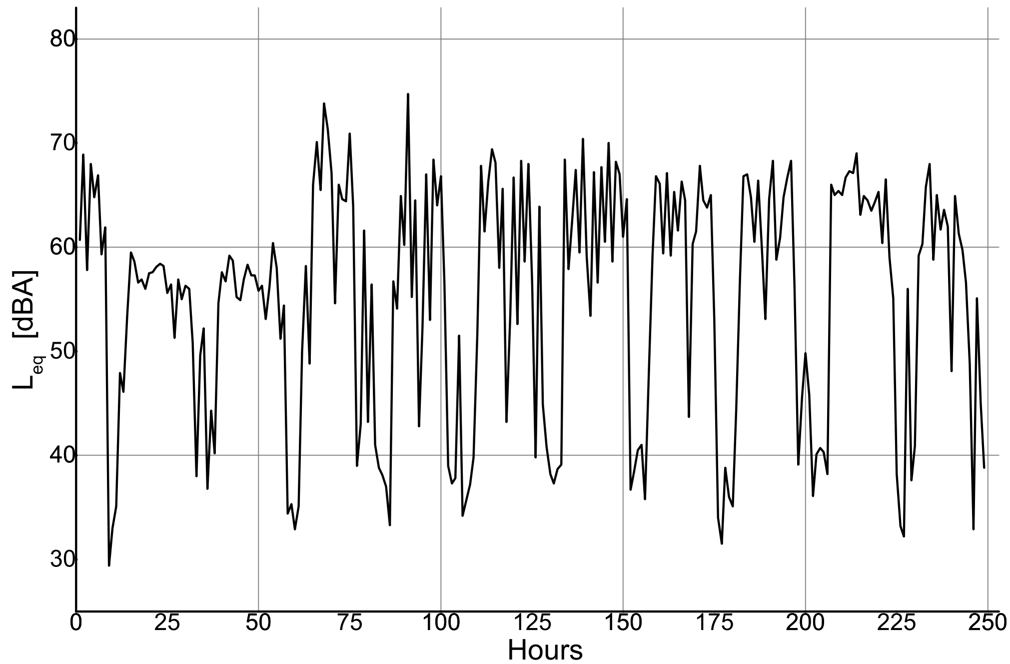


Figure 1: Measurements at the Nice-Côte d'Azur International Airport, South of France.

4.1 Maximum likelihood estimation

In order to obtain the maximum likelihood estimator of the vector of parameters of the non-homogeneous Poisson model, the software Matlab is used. In the Appendix we give the code with the definition of the function to be optimised using the Matlab optimisation tool. Due to the complexity of the maximisation procedure in the presence of change-points, we use the maximum likelihood method only in the version of the model with no change-points.

The search regions for the maximum value $\hat{\boldsymbol{\theta}} = (\hat{\alpha}, \hat{\sigma})$ are the intervals [0.1, 2] for the parameter α and [0.1, 100] for σ, for both thresholds. The search started at points 0.5 and 30, for α and σ, respectively. When the threshold 55dB is used, thirteen iterations are needed to obtain the values of α and σ that maximise the likelihood function. They are $\hat{\alpha} = 0.999$ and $\hat{\sigma} = 1.574$. In the case of the threshold 60dB, we have that after twelve iterations those values are $\hat{\alpha} = 1.312$ and $\hat{\sigma} = 7.554$.

Note that in the case of the threshold 55dB, the process N may be considered a homogeneous Poisson process ($\alpha \approx 1$). Nevertheless, a non-homogeneous Poisson with an increasing rate function ($\alpha > 1$) is detected when the threshold is 60dB.

4.2 Bayesian estimation

In order to estimate the parameters using the Bayesian point of view, samples from their respective posterior distributions are used. Since the posterior distributions involved do not have a simple form, a Markov chain Monte Carlo algorithm is used. The one considered here is the Gibbs sampling algorithm internally implemented in the OpenBugs software (see, for instance, www.openbugs.net/w and [35]). The code used is a slight variation of the one found in [11]. The version in R is given in [36].

4.2.1 No change-points

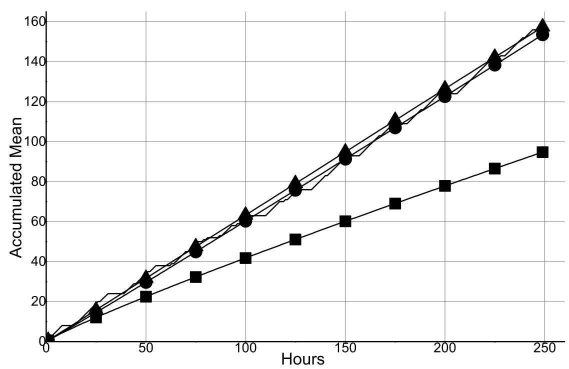
In a first instance, and for both threshold values, uniform prior distributions are assumed for the parameters α and σ. Hence, for all threshold values, we take a uniform distribution U(0.1, 2) for α and a U(0.1, 100) for σ. When the threshold 55dB is considered and more informative prior distributions are used, we take a Normal distribution N(1.032, 157.468) as prior distribution for α and a Gamma prior distribution Gamma(6.9054, 3.5358) for σ. When taking into account the threshold 60dB the respective prior distributions were N(1.351, 57,13) and Gamma(8.9584, 1.0367). (Here, Gamma(a, b) is the Gamma distribution with mean a/b and variance a/b^2.) In all cases estimation of the parameters was made with a sample of size 22500 obtained from five chains after a burn-in period of 5000 steps, taking every 10th generated value. In Table 1 we present the estimated quantities of interest.

		Mean		SD		95% Credible Interval	
		55dB	60dB	55dB	60dB	55dB	60dB
α	(uni)	0.897	1.35	0.277	0.131	(0.355, 1.187)	(1.108, 1.617)
	(non-uni)	1.024	1.34	0.6471	2.842	(0.9323, 1.116)	(3.942, 14.91)
σ	(uni)	1.555	8.64	1.032	0.077	(2.12E-04, 1.617)	(1.192, 1.493)
	(non-uni)	1.825	8.24	0.4178	1.647	(1.096, 2.728)	(5.307, 11.71)

Table 1: Posterior means, standard deviations (indicated by SD), and the 95% credible intervals for the estimated parameters α and σ when the Weibull rate function is used and for both threshold considered. We use the notation "uni" and "non-uni" to indicate the cases of uniform and non-uniform prior distributions, respectively.

Looking at Table 1, we may see that when the threshold 55dB is considered and the non informative prior distributions are used, the Poisson model is such that the rate function presents a decreasing behaviour ($\alpha < 1$). However, when we use more informative prior distributions, the Poisson model is approximately a homogeneous process (since α is close to one). However, strictly speaking, the rate function is an increasing one, since the value of α is larger than one, even though it is only slightly larger. When the threshold 60dB is used, we have in both cases (non informative and informative prior distributions), an increasing rate function. That means that as the time passes, exceedances of the threshold, become more frequent.

In order to check the adequacy of the models considered here, Figure 2 shows the plots of the observed and estimated accumulated means in the case of the thresholds 55db (top) and 60dB (bottom).



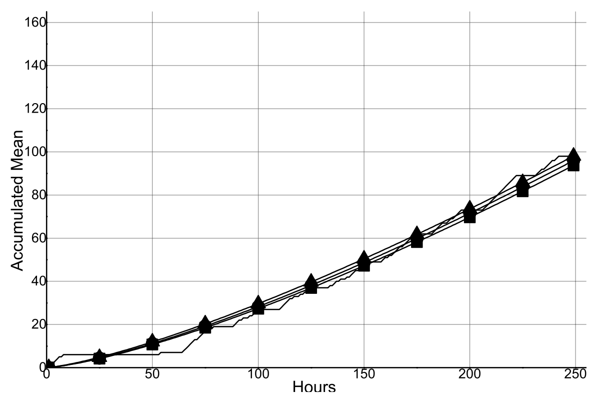


Figure 2: Observed (plain solid line) accumulated mean of the non-homogeneous Poisson model, and the corresponding estimated accumulated means using the maximum likelihood function (▲), using the Bayesian inference with non-informative (■), and informative (●) prior distributions when no change-points are allowed. Plots on the top of figure correspond to the case where the threshold 55dB is considered, and the plots on the bottom correspond to the case where we use the threshold 60dB.

Looking at Figure 2, we may see that in the case of 55dB, even though the maximum likelihood estimator produces an estimated accumulated mean function with behaviour similar to the observed one, we have that the Bayesian model with informative prior distributions also explains well the behaviour of the observed accumulated mean. However, the values provided by the Bayesian method using non-informative prior distributions give an estimated accumulated mean that is far away from the observed one. When the threshold 60dB is considered, it is possible to see that all methods provide estimated values producing an estimated accumulated mean that explains well the behaviour of the observed one.

4.2.2 Presence of a change-point

It is possible to see, by looking at Figure 2, that when the threshold is 60dB, perhaps a model with the presence of a change-point should be considered. Hence, in this section we deal with that case. Based on the information obtained when using the model with no change-points, we have chosen as prior distributions for the parameters, a Gamma(64, 80) and a N(1.351, 57.13) for α_1 and α_2, respectively; and a Gamma(64,16) and Gamma(8.9584, 1.0367) for σ_1 and σ_2, respectively. The change-point τ will have as prior distribution a N(65, 0.03). A sample of size 22500, obtained from five chains after a burn-in period of 5000 steps and taking every 10th generated value, was used to estimate the parameters of interest. Their values are given in Table 2.

Looking at Table 2, we have that in the time interval $(0, \tau)$, the rate function has a decreasing behaviour ($\alpha_1 < 1$). However, after τ, we have that a change occurs not only in the value of the parameter α, but also on the behaviour of the rate function. We can see that from that point in time on, the rate function becomes an increasing function ($\alpha_2 > 1$). Hence, the model specifies that from about the 62nd measurement (approximate value of the change-point), exceedances are more frequent.

In Figure 3, we have the observed and estimated accumulated means. Looking at Figure 3, we may see that there is an almost perfect fit of the estimated accumulated

	Mean	SD	95% Credible Interval
α_1	0.722	0.073	(0.581, 0.867)
α_2	1.338	0.087	(1.172, 1.512)
σ_1	4.02	0.488	(3.121, 5.045)
σ_2	7.718	1.858	(4.483, 11.74)
τ	62.44	2.554	(55.8, 66.11)

Table 2: Posterior means, standard deviations (indicated by SD), and 95% credible intervals of the parameters α_1, α_2, σ_1, σ_2 and τ when the threshold 60dB is used.

mean to the observed one. Therefore, based on Figures 2 and 3, we have that the best model to explain the data is the non-homogeneous model allowing the presence of a change-point.

4.3 Other methods for model selection

In addition to the graphical comparison, we have obtained the SAD and CV values in all cases considered here. Their values as well as the values of μ and σ" used to calculate CV are given in Table 3.

If we consider the SAD as the method to select the best model, then looking at Table 3, we may see that the chosen one is the version whose estimated parameters are obtained by the Bayes estimation procedure with an informative prior distribution for both threshold values. When we use the CV to select the best estimating method we have that in the case of the threshold 55dB, the chosen model is the one whose parameters were estimated using the Bayesian method with informative prior distributions. However, in the case of the threshold 60dB, the chosen model is the one whose parameters are estimated using

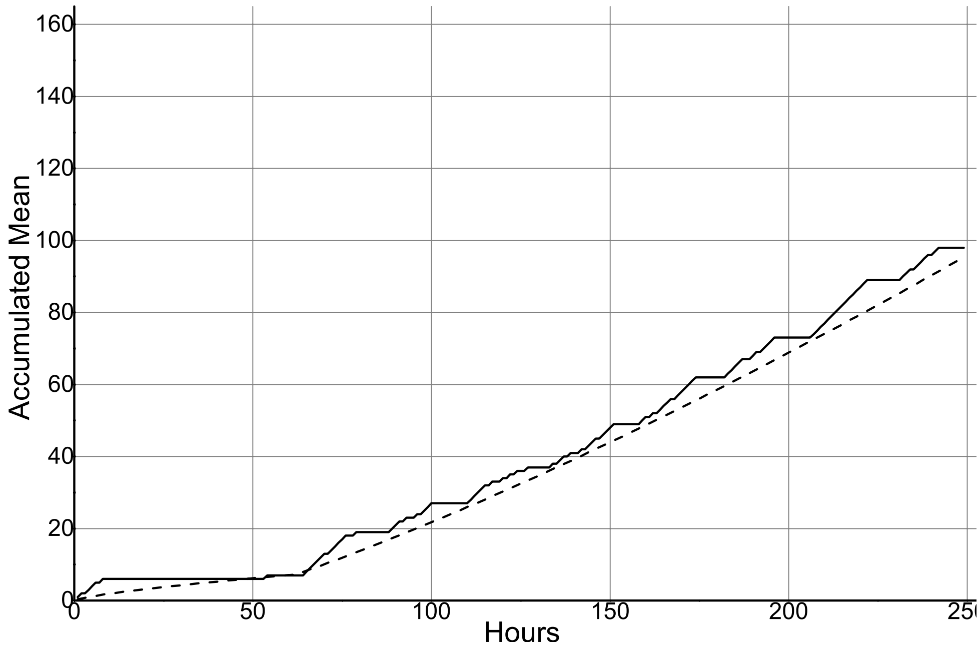


Figure 3: Observed (solid line) and estimated (dashed line) accumulated means of the non-homogeneous Poisson process when the presence of a change-point is allowed.

the Bayesian framework and allowing the presence of a change-point.

Note that small values of CV indicate that, on average, the difference between estimated and observed mean functions is concentrated around its mean. In the case of the threshold 60dB, we have that the model using the Bayesian method with informative prior distribution and also the one allowing the presence of a change-point, produce similar values of CV with a slight advantage of the change-point formulation. Looking at the values of μ and σ'' in those same cases, we have that they are also similar with a slight advantage to the version without the change-point. However, given the order of the mean and the standard deviation, we may say that the version adjusting best to the observed values is the one with a change-point. That can be corroborated by looking at Figures 2

		MaxLike	BayesNonInfo	BayesInfo	BayesChPt
55dB	SAD	670.07	6762.99	550.95	–
	CV = μ/σ''	0.719	0.65	0.63	–
	μ	2.69	27.16	2.21	–
	σ''	1.934	17.67	1.4	–
60dB	SAD	835.87	768.89	739.35	797.36
	CV = μ/σ''	0.604	0.67	0.599	0.565
	μ	3.36	3.09	2.97	3.2
	σ''	2.03	2.07	1.78	1.81

Table 3: Values of the sum of the absolute values of the differences (SAD), of the coefficient of variation (CV), of the mean μ, and of the standard deviation σ'' for all thresholds and versions of the model considered here. We use "MaxLike", "BayesNonInfo", "BayesInfo", "BayesChPt" to indicate the cases where we use the maximum likelihood method, the Bayesian method with non-informative and informative prior distributions, and the Bayesian method with the presence of a change-point, respectively. The symbol "-" is used to indicate that the corresponding version of the model was not used.

and 3. When we consider the threshold 55dB, it is clear the advantage of the model with parameter estimated using the Bayesian method with informative prior distribution.

5 Discussion

In this work, we have used a non-homogeneous Poisson model to study the behaviour of noise level data. The main question to be answered was related to the average number of times that a given threshold was exceeded by a sound measurement. The model was applied to the noise level data obtained at Nice-Côte d'Azur International Airport, located in the South of France.

Two threshold values were considered. They are, $L = 55$dB and $L = 60$dB. When taking into account the threshold 60dB, the presence of a change-point was detected and therefore also allowed in the model.

We have that the graphical comparison, between the observed and estimate accumulated means, says that in the case of $L = 55$dB, the model that best fits the data is the Bayesian model assuming informative prior distributions for the parameters. In the case of $L = 60$dB, the best fit is given by the model with a possible change-point and using informative prior distributions.

The results show that when the threshold 55dB is taken into account, the process N behaves as an approximately homogeneous Poisson process with rate $\lambda(t) = \lambda \approx (1/\sigma) = (1/1.825) \approx 0.55$. Therefore, if we want to calculate the probability that during a time interval $[t, t+s)$, $t, s, \geq 0$, we have k ($k = 0, 1, \ldots$) hours in which the noise level exceeds 55dB, we just have to use the following simplified version of (2)

$$P(N_s = k) = \frac{(\lambda\, s)^k}{k!}\, e^{-\lambda\, s}, \quad s \geq 0.$$

Therefore, if we want to know the probability of having four exceedances of the threshold 55dB in a time interval of 10 hours, then we calculate

$$P(N_{10} = 4) = \frac{(0.55 \times 10)^4}{4!}\, e^{-0.55 \times 10} \approx 0.16.$$

When the threshold considered is 60dB and we take into account the model with a change-point, we have a non homogeneous model with an increasing rate function after the 62nd measurement. If we want to know if in the next 24 hours following the end of the observational period, we have 10 measurements with levels exceeding 60dB, we just need to use (2) with the appropriate values of s and t and the corresponding estimated values of α and σ. Note that in this case we have $t = 249$. We also have (see Table 3) that the estimated values of α and σ are $\hat{\alpha} = 1.338$ and $\hat{\sigma} = 7.718$. Hence, we have that $(1/\hat{\sigma}) \approx 0.13$. Therefore, substituting the corresponding values in (2) we have that

$$P(N_{249+s} - N_{249} = k) = \frac{[(0.13 \times (249+s))^{1.338} - (0.13 \times 249)^{1.338}]^k}{k!} \exp\left(-\left[0.13 \times (249+s))^{1.338} - (0.13 \times 249)^{1.338}\right]\right). \quad (3)$$

Hence, taking $s = 24$, the probability sought is $P(N_{273} - N_{249} = 10) \approx 0.071$. Whereas, if we take $k = 4$ then we have that $P(N_{273} - N_{249} = 4) \approx 0.0016$. Suppose that we want to know the probability that four exceedances of the threshold 60dB will occur in the next 10 hours after the observational period has ended. Then, when we need to take $s = 10$ and $k = 4$ in (3). Hence, we have that $P(N_{259} - N_{249} = 4) \approx 0.148$. In the same manner we may calculate the probability for other values of k and s.

Note that the value of the change-point $\tau \approx 62$ corresponds to the time around 5:48am of the 24 December 2000. Since on the morning of Christmas Day the air traffic is bound to increase, the averaged noise levels tend to be higher around that time. This leads to a larger number of exceedances of the threshold of interest. Hence, the model with a change-point reflects well the behaviour of the data. The other versions of the model do not capture this change in behaviour, though.

In an attempt to improve the results obtained here, we have decided to use a hybrid method. Using the change-point estimated via the Bayesian point of view, we have applied the maximum likelihood estimation method separately for each subset of the observed data; i.e., those prior and those located after the change-point. The estimated values of α_1, α_2, σ_1, and σ_2, produce an even better model approximation to the observed data.

Those estimated values are, respectively, 0.4, 1.02, 0.5, and 2.21, and the SAD and CV values are, respectively, 398.68 and 0.74, with $\mu = 1.6$ and $\sigma'' = 11.8$.

Using the information provided by the maximum likelihood estimated values of α_1, α_2, σ_1, and σ_2, we consider very informative prior distributions for those parameters and apply the Bayesian estimation method again. The prior distributions of the change-point τ is as before. In the case of the other parameters, we consider the following prior distributions: Gamma(16, 40), Gamma(26, 25, Gamma(25, 50), and Gamma(122, 55), for α_1, α_2, σ_1, and σ_2, respectively. The estimated value of τ was again about 62. In the case of the other parameters, we have that the estimated values of α_1, α_2, σ_1, and σ_2 are 0.39, 1.02, 0.5, and 2.23, respectively. Those estimated values produce a estimated mean function that gives SAD and CV values of, respectively, 411.16 and 0.78 with $\mu = 1.65$ and $\sigma'' = 1.29$. Therefore, comparing these results with those of Table 3, we have that the hybrid estimation of the parameters, in the case of the presence of a change-point, would be the one with the smallest SAD and CV. Note that, in this case, we have a non-homogeneous (with a decreasing rate function) model for the first part of the data (before the change-point), and a homogeneous Poisson model for the second part of the data (after the change-point). The drawback of this method, though, is that the interval where the prior distributions have larger values is an extremely narrow one. Figure 4 shows the graphical fit of the estimated accumulated mean to the observed one.

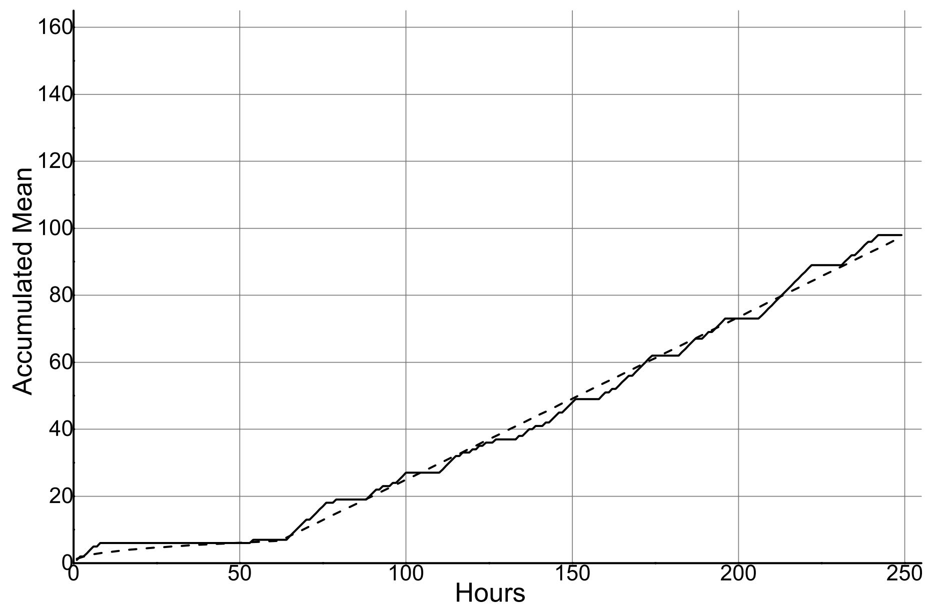


Figure 4: Observed (solid line) and estimated (dashed line) accumulated means of the non-homogeneous Poisson model with a change-point using very informative prior distributions obtained when the hybrid method is considered.

Acknowledgements

The authors thank 01dB-Metravib for allowing us to use the data set considered here. ERR was partially funded by the project PAPIIT-IN10271-3 of the Dirección General de Apoyo al Personal Académico of the Universidad Nacional Autónoma de México, Mexico. CG received partial support from FARB-2009. ERR thanks all at the Department of Industrial Engineering of the University of Salerno, Italy, for their hospitality and support.

Appendix

In the section we present the Matlab code for the likelihood function to be maximised when we consider the non-homogeneous Poisson model without the presence of change-points. We show only the case where the threshold is 55dB. The variable S is the number of hours in which the measurements were taken, K is the number of hours, among the S, that have values exceeding 55dB, and M is $\sum_{i=1}^{K} \log(h_i)$, with h_i, $i = 1, 2, \ldots, K$, the hours where the threshold was exceeded.

Matlab Code

```
function f = cinq(x)
S = 249;
K = 157;
M = 709;
f  = -((K*log(x(1)))-(K*x(1)*log(x(2)))+((x(1)-1)*M) -((S/x(2))^x(1)));
```

References

[1] Guarnaccia C, Mastorakis NE, Quartieri J. Noise Sources Analysis in a Wood Manufacturing Company, International Journal of Mechanics 2013; 7:37-44.

[2] De Kluizenaar Y, Janssen SA, van Lenthe FL, Miedema HME, Mackenbach JP. Long-term road traffic noise exposure associated with an increase morning tiredness. J. Acoust. Soc. Am. 2009; 126:626-633.

[3] De Kluizenaar Y, Salomons EM, Janssen SA, van Lenthe FJ, Vos H, Zhou H, Miedema HME, Mackenbach JP. Urban road traffic noise and annoyance: the effect of a quiet façade. J. Acoust. Soc. Am. 2011; 130:1936-1942.

[4] Ouis D. Annoyance from road traffic noise: a review. Journal of Environmental Psychology 2001; 21:101-120.

[5] WHO. World Health Organization. In: Berlund B, Lindvall T, and D.H. Schewela DH, editors. Guidelines for Community Noise. Geneva: World Health Organization; 1999, p. v-xviii, 19-20.

[6] ENSTAT. NV1 - Definition of noise and vibration terms. Downloaded from www2.luton.gov.uk/NapierPark/Environmental_Statement/Appendices/Noise_and_Vibration/NV1.pdf, on 14 May 2014.

[7] Guarnaccia C. Advanced Tools for Traffic Noise Modelling and Prediction, WSEAS Transactions on Systems 2013; 12:121-130

[8] Guarnaccia C, Quartieri J, Mastorakis NE, Tepedino C. Acoustic Noise Levels Predictive Model Based on Time Series Analysis. In: Latest Trends in Circuits, Systems, Signal Processing and Automatic Control. Proceedings of the 2nd International Conference on Acoustics, Speech and Audio Processing (ASAP '14), Salerno, Italy, June 3-5, 2014.

[9] Guarnaccia C, Quartieri J, Rodrigues ER, Tepedino C. Time Series Model Application to Multiple Seasonality Acoustical Noise Levels Data Set. In: Latest Trends in Circuits, Systems, Signal Processing and Automatic Control. Proceedings of the 2nd Int. Conf. on Acoustics, Speech and Audio Processing (ASAP '14), Salerno, Italy, June 3-5, 2014.

[10] Guarnaccia C, Quartieri J, Barrios JM, Rodrigues ER. Modelling environmental noise exceedances using non-homogeneous Poisson processes. Revised version submitted to the Journal of the Acoustical Society of America; 2013.

[11] Achcar JA, Rodrigues ER, Paulino CD, Soares P. Non-homogeneous Poisson processes with a change-point: an application to ozone exceedances in Mexico City'. Environmental and Ecological Statistics 2010; 17:521-541. doi:10.1007/s10651-009-0114-3

[12] Achcar JA, Rodrigues ER,Tzintzun G. Using non-homogeneous Poisson models with multiple change-points to estimate the number of ozone exceedances in Mexico City. Environmetrics 2011; 22:1-12. doi:10.1002/env.1029

[13] Wilson SP, Costello MJ. Predicting future discoveries of European marine species using non-homogeneous renewal processes. Journal of the Royal Statistical Society Series C 2005; 54:425-442.

[14] Kass-Houter TA, Xu Z, McMurray P, Park S, Buckeridge DL, Brownstein JS, Finelli L, Groseciose SL. Application of change-point analysis to daily influenza-like ilness emergency department visits. Journal of the Am. Med. Inform. Assoc. 2012; 19:1075-1081. doi: 1136/amiajnl-2011-000793

[15] Chen J, Gupta AK. Parametric change-point analysis with applications to genetics, medicine, and finance. 2nd ed. New York: Springer; 2012.

[16] Cox DR, Lewis PA. Statistical analysis of series of events. London: Methuen; 1966.

[17] Lawless JF. Statistical models and methods for lifetime data. New York: John Wiley and Sons; 1982.

[18] Rice JA. Mathematical statistics with data analysis. New York: Wadsworth and Brooks; 1988.

[19] Carlin BP, Louis TA. Bayes and empirical Bayes methods for data analysis. 2nd ed. New York: Chapman and Hall/CRC; 2000.

[20] Robert CR, Casella G. Monte Carlo statistical methods. New York: Springer; 1999.

[21] Grimmett GR, Stirzaker DR. Probability and random processes. London: Clarendon Press; 1982.

[22] Ross SM. Stochastic Processes. 2nd ed. New York: John Wiley and Sons; 1996.

[23] Achcar JA, Dey DK, Niverthy M. A Bayesian approach using nonhomogeneous Poisson process for software reliability models. In: Basu SK, Mukhopadhyay S, editors. Frontiers in Reliability, Series on Quality, Reliability and Engineering Statistics 4, Calcutta University, India, 1998, p. 1-18. doi: 10.1142/9789812816580_0001

[24] Achcar JA., Fernández-Bremauntz AA, Rodrigues ER, Tzintzun, G. Estimating the number of ozone peaks in Mexico City using a non-homogeneous Poisson model. Environmetrics 2008; 19:469-485. doi: 10.1002/env.890

[25] Ramírez-Cid JE, Achcar JA. Bayesian inference for nonhomogeneous Poisson processes in software reliability models assuming nonmonotonic intensity functions. Computational Statistics and Data Analysis 1999; 32:147-159.

[26] Álvarez LJ, Fernández-Bremauntz AA, Rodrigues ER, Tzintzun, G. Maximum a posteriori estimation of the daily ozone peaks in Mexico City. Journal of Agricultural, Biological, and Environmental Statistics 2005; 10:276-290. doi:10.1198/108571105X59017

[27] Boys RJ, Henderson DA. On determining the order of Markov dependence of an observed process governed by a hidden Markov model. Special Issue of Scientific Programming 2002; 10:241-251.

[28] Churchill GA, Lazareva B. Bayesian restoration of a hidden Markov chain with applications to DNA sequencing. Journal of Computational Biology 1999; 6:261–277.

[29] Geman D. Random fields and inverse problems in imaging. Lecture Notes in Mathematics 1990; 1427:113-193.

[30] Freston KJ, Glaser DE, Henson RNA, Kiebel S, Phillips C, Ashburner J. Classical and Bayesian inference in neuroimaging: applications, NeuroImage 2002; 16:484-512. doi: 10.1006/ning.2002.1091

[31] Murshudov GN, Vagin AA, Dodson EJ. Refinement of molecular structure by maximum likelihood method. Acta Cryst. 1997; D53:240-255.

[32] Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA 5: molecular evolutionary genetics. Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution 2011; 28:2731-2739. doi: 10.1093/molbev/msr121

[33] Yang Z. PAML 4: Phylogenetic analysis by maximum likelihood. Molecular Biology and Evolution 2007; 24:1586-1591. doi: 10.1009/molbev/msr088

[34] Yang TE, Kuo L. Bayesian binary segmentation procedure for a Poisson process with multiple change-points. Journal of Computational and Graphical Statistics 2001; 10:772-785. doi:10.1198/106186001317243449

[35] Lunn D, Spiegelhalter D, Thomas A, Best N. The BUGS project: Evolution, critique and future directions (with discussion). Statistics in Medicine 2010; 28:3049-3082.

[36] Rodrigues ER, Achcar JA, Jara-Ettinger J. A Gibbs sampling algorithm to estimate the occurrence of ozone exceedances in Mexico City. In: Popovic D, editor. Air Quality: Models and Applications. InTech Open Access Publishers: Croatia; 2011, p. 131-150.