Elsevier Editorial System(tm) for Applied Acoustics
Manuscript Draft

Manuscript Number: APAC-D-16-00088

Title: A time series analysis and a non-homogeneous Poisson model with multiple change-points applied to acoustic data

Article Type: Research Paper

Section/Category: Europe and Rest of the World

Keywords: Markov chain Monte Carlo algorithms; Statistical inference; Community noise; Non-homogeneous Poisson models; Time-series models

Corresponding Author: Dr. Eliane R. Rodrigues, PhD

Corresponding Author's Institution: Universidad Nacional Autonoma de Mexico

First Author: Claudio Guarnaccia, PhD

Order of Authors: Claudio Guarnaccia, PhD; Joseph Quartieri, PhD; Carmine Tepedino, MSc; Eliane R. Rodrigues, PhD

Abstract: High levels of the so-called community noise may produce hazardous effect on the health of a population exposed to them for large periods of time. Hence, studying the behaviour of those noise measurements is very important. In this work we analyse that in terms of the probability of exceeding a given threshold level a certain number of times in a time interval of interest. Since the datasets considered contain missing measurements, we use a time series model to estimate the missing values and complete the datasets. Once the data is complete, we use a non-homogeneous Poisson model with multiple change-points to estimate the probability of interest. Estimation of the parameters of the models are made using the usual time series methodology as well as the Bayesian point of view via Markov chain Monte Carlo algorithms. The models are applied to data obtained from two measuring sites in Messina, Italy.

Dear Editor,

Please, find the manuscript entitled "A time series analysis and a non-homogeneous Poisson model with multiple change-points applied to acoustic data" co authored with C. Guarnaccia, J. Quartieri, and C Tepedino, that we are submitting for possible publication in the journal Applied Acoustics.

Looking forward to hearing from you I send you my best wishes.

Your sincerely,

Dr. Eliane R. Rodrigues

A time series analysis and a non-homogeneous Poisson model with multiple change-points applied to acoustic data

Claudio Guarnaccia, **Joseph Quartieri** and **Carmine Tepedino**

Department of Industrial Engineering

University of Salerno, Italy

E-mails: cguarnaccia@unisa.it, quartieri@unisa.it, ctepedino@unisa.it

Eliane R. Rodrigues[1]

Instituto de Matemáticas

Universidad Nacional Autónoma de México, Mexico

E-mail: eliane@math.unam.mx

[1]Corresponding author.

Abstract

High levels of the so-called community noise may produce hazardous effect on the health of a population exposed to them for large periods of time. Hence, studying the behaviour of those noise measurements is very important. In this work we analyse that in terms of the probability of exceeding a given threshold level a certain number of times in a time interval of interest. Since the datasets considered contain missing measurements, we use a time series model to estimate the missing values and complete the datasets. Once the data is complete, we use a non-homogeneous Poisson model with multiple change-points to estimate the probability of interest. Estimation of the parameters of the models are made using the usual time series methodology as well as the Bayesian point of view via Markov chain Monte Carlo algorithms. The models are applied to data obtained from two measuring sites in Messina, Italy.

Keywords: Markov chain Monte Carlo algorithms; Statistical inference; Community noise; Non-homogeneous Poisson models; Time-series models

1 Introduction

Individuals spending time in an environment with high levels of the so-called community noise or environmental noise pollution may suffer a deterioration in their health. Among the many adverse effects caused by high levels of noise are hearing impairment, sleeping disturbance ([1]), and cardiovascular problems. Therefore, it is a very important issue to be able to understand the behaviour of this type of pollution. Once that behaviour is understood, the corresponding environmental authorities may implement preventive/palliative measures in a way that either the population is able to avoid a hazardous situation or the authorities are able to bring the levels down.

There are several ways of measuring sound levels. To give an approximation to the frequency response of our hearing system, the most common procedure used for environmental noise is the so-called A-weighting (see for instance [2]). That gives low weights to low frequencies and higher weights to middle and high frequencies. When we have continuous noise such as road traffic noise (which is the type of noise considered here), a suggested measure ([2]) is the energy average equivalent level of the A-weighted sound pressure over a period of time R, which is indicated by $L_{Aeq,R}$ and defined by

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

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.

Note that sound pressure levels for 24 hours can be between 75dBA and 80dBA alongside roads and other noisy areas. Therefore, since the majority of human beings live in urban and suburban areas, that part of the population is largely affected by noise proceeding from road traffic. Hence, the importance of studying the behaviour of that type of data.

One of the aims in the present work is to estimate the probability that a given population is exposed to a noise level that exceeds a threshold a certain number of times in a given time interval. Two types of questions are of interest here. One of them is related to the ability of predicting future behaviour of the data in terms of exceeding a given noise threshold. The other is related to the behaviour of the actual measurements. In the latter type of question also resides the interest in comparing how the data change from one period of time to another. This change may be captured by the so-called change-points which will be considered in the analysis.

The datasets analysed here present many missing data. In order to solve this problem we will use time series analysis to estimate the missing values. Once the dataset is complete (i.e., with observed and estimated measurements), then a non-homogeneous Poisson model allowing the presence of multiple change-points is used to estimate the number of exceedances of a given threshold. In addition to the time series method, the non-homogeneous Poisson model allows the prediction of the possible behaviour of future measurements.

Both methodologies considered here (time series and Poisson process) have been used in several areas of application. When considering environmental problems, we have, for instance, that non-homogeneous Poisson models are applied to the areas of air pollution (see for instance [3, 4, 5]) and in species abundance ([6]). When the problem is related to community noise, we have [7]) where the non-homogeneous Poisson model is applied to two datasets collected in two locations in the city of Messina, Italy, and [8] where a non-homogeneous Poisson model with one change-point is applied to data from an airport in the South of France. In the case of times series applications to air pollution problems we have for instance [5] and [9]. In [10, 11] two time series models were used to analyse a subset of one of the datasets considered here. In these works, a multiplicative time series was used as well as a mixed one where two seasonal effects could be detected. In the present work we use the model given in [10] to analyse the behaviour of the data and to fill the gaps related to the missing values.

The daily observational data at a measuring site, are represented by a 16-hour energy average sound level, $L_{Aeq,16h}$, for the day period (corresponding to 6am to 10pm), and an 8-hour energy average sound level, $L_{Aeq,8h}$, for the night period (corresponding to 10pm to 6am). The measuring sites considered here are the Viale Boccetta and Via La Farina located in the city of Messina, Italy.

Remarks. 1. Even though in the present work we also use the Messina data, the entire dataset is used and not only subsets of the measurements as in [7, 10, 11].

2. Note that the methods considered here could be used in conjunction with traffic noise models to predict the behaviour of noise levels when changes are made in a given environment. Using the traffic noise models it would be possible to observe how the noise levels would change if, for instance, traffic is reduced in busy roads next to a residential area. Taking into account that information we could apply the methodology considered here to estimate the number of times that a noise level would be surpassed if traffic is restricted. Additionally, we could predict future behaviour of the noise measurements under the new restriction. Therefore, the behaviour of the noise levels could be theoretically studied before the noise reducing measures are implemented in a given community.

This paper is organised as follows. In Section 2 the mathematical models are presented. In Section 3 the methods used to estimate the parameters of the models are given as well as criteria for selecting the best model to represent the behaviour of the datasets. Section 4 gives an application to the data from Viale Boccetta and Via La Farina sites in Messina, a city located in Sicily, Italy. Finally, in Section 5, we present a discussion of the results obtained.

2 Description of the mathematical models

A two-step approach will be used in order to analyse the problem considered here. The first step consists of using a time series model to reconstruct the missing data. The second step consists of using the reconstructed dataset, formed by the actual measurements and the ones imputed using the time series model, to obtain the days in which exceedances of a noise threshold of interest occurred. Once these days are obtained a non-homogeneous Poisson model is used to estimate the probability of having a given number of exceedances in a time interval of interest. The time series and the non-homogeneous Poisson models are described as follows.

2.1 The time series model

Time series is a stochastic process, i.e., a sequence of random variables recording the outcome of a random experiment ([12, 13, 14, 15]). The present study deals with the case where the random variables registers the daily (day and night periods) noise levels at a given site of interest.

The time series considered here is described mainly by three components: the trend component which explain the long time direction of the series, the seasonal component which accounts for cyclical changes, and the random noise component, also called residual, to account for other random fluctuations.

Let $\mathbf{X} = \{X_t : t \geq 0\}$ indicate the time series of interest. Denote by $\mathbf{T} = \{T_t : t \geq 0\}$ the trend component of the series, $\mathbf{S} = \{S_t : t \geq 0\}$ the seasonal component, and $\mathbf{E} = \{E_t : t \geq 0\}$ the random noise component.

A mixed times series is used to describe the behaviour of the data. Therefore, we consider a multiplicative form in the trend and seasonal components and an additive random component, i.e.,

$$X_t = T_t \times S_t + E_t, \quad t \geq 0. \tag{1}$$

A trend of type $T_t = \sum_{k=0}^{n} b_k \, t^k$ is taken. In some datasets taking $n = 1$ will be enough. However, in some cases higher values of n will be adopted. The seasonal and random components are given as in [10, 11]. In particular, the seasonal effect S_t at a given period t, is obtained by the ratio between the actual (measured/estimated) data X_t and the moving average value M_t,

$$S_t = \frac{X_t}{M_t}.$$

The moving average M_t is calculated with a span of length k. The value of k is the value that maximises the autocorrelation function of the series $\mathbf{X}$.

Once the seasonal effect S_t is calculated for every period, k seasonal coefficients, one for each period of the chosen span, are evaluated averaging on all the homologous periods, according to the following formula,

$$\overline{S}_i = \frac{\sum_{l=1}^{m_i-1} S_{(i+l)k}}{m_i},\ i = 1, 2, \ldots, k,$$

where m_i is the number of homologous ith periods in the overall time range of the dataset. (In our case, we will have a span of length seven and each period will correspond to a day of the week, i.e., we have one for Monday, one for Tuesday, and so on.)

As for the random component, we estimate the error of the model in the calibration dataset as follows,

$$\hat{E}_t = X_t - F_t,$$

where F_t is the so-called point forecast as given in [10, 11] by,

$$F_t = T_t \times S_t, \quad t \geq 0. \tag{2}$$

The random variables E_t are expected to be independent and identically distributed. Thus, $\hat{E}_t$ is expected to be normally distributed. Therefore, its mean coincides with its mode. Thus, the mean error, indicated by m_ϵ, can be added to the forecast, in order to draw the final model prediction Y_t, i.e.,

$$Y_t = T_t \times \overline{S}_i + m_\epsilon,$$

where the value of $\overline{S}_i$ used is the one corresponding to the cycle starting on day t of the observational period, i.e., if t corresponds to a Monday, then the $\overline{S}_i$ corresponds to the estimated value for that cycle component.

Remark. Note that for independent and identically distributed errors, the mean of the distribution is expected to be zero but as we add m_ϵ to the forecast, it is possible to balance the possible presence of distortions in the model.

In order to impute the missing values in the dataset, we consider the point forecast F_t, given by (2), evaluated on the missing periods. A comparison of this imputation method, with a standard regression method is reported in [16]. Once the missing values have been imputed, taking the whole series, the exceedance days are obtained.

2.2 The non-homogeneous Poisson process model

Let $\hat{\mathbf{X}} = \{\hat{X}_t \,:\, t \geq 0\}$ indicate the sequence of measurements formed by both the actual measured community noise levels and the imputed ones using the time series model. In order to estimate the probability of having the noise level above a given threshold a certain number of times, a non-homogeneous Poisson model is used.

Poisson processes ([17, 18]) are a particular case of continuous-time Markov chains ([12, 17]) and they are usually used to count occurrences of events (see[18]). Since in the present work we are interested in counting the number of times that a given environmental noise threshold is surpassed, Poisson processes are a suitable choice.

In order to set the model, consider the following notation. Let $N_t \geq 0$ be the number of times that a given community noise threshold is surpassed in the time interval $[0, t)$, $t \geq 0$. Assume that $\mathbf{N} = \{N_t \,:\, t \geq 0\}$ evolves according to a non-homogeneous Poisson process with rate and mean functions given by $\lambda(t) > 0$ and $m(t) = \int_0^t \lambda(s)\, ds$, $t \geq 0$, respectively ([18]). Hence, we have that, for $k = 0, 1, 2, \ldots$,

$$P(N_{t+s} - N_t = k) \;=\; \frac{[m(t+s) - m(t)]^k}{k!} \exp\left(-[m(t+s) - m(t)]\right). \tag{3}$$

Take $\lambda(t)$, $t \geq 0$ of the Weibull type, i.e., $\lambda(t) = (\alpha/\sigma)\,(t/\sigma)^{\alpha-1}$, where $\alpha > 0$ and $\sigma > 0$ are parameters that need to be estimated. When $\lambda(\cdot)$ is of the Weibull form, the mean function associated to it is $m(t) = (t/\sigma)^\alpha$, $t \geq 0$ (see for instance [19]).

Remark. If $\alpha < 1$ ($\alpha > 1$), then the rate function $\lambda(\cdot)$ is a decreasing (increasing) function of t. If $\alpha = 1$, then the rate function is a constant function of t. An increasing rate function $\lambda(\cdot)$ means that exceedances become more frequent events as the time passes. A decreasing one indicates that exceedances become rarer events as the time passes. If $\lambda(\cdot)$ is constant, then no changes occur in the behaviour of the time between two consecutive exceedances.

3 Estimation of the parameters of the models

There are several ways in which the parameters of a model may be estimated. When estimating the parameters involved in the time series model, a simple spreadsheet suffices. In the case of the parameters of the non-homogeneous Poisson model, we use the Bayesian point of view ([20, 21, 22]). Within the Bayesian framework, we assign prior distributions to the parameters to describe our uncertainty about them. In this way, they become random quantities.

3.1 The time series model

When using a spreadsheet, we just need to specify the expression for the trend and also the lag of the moving average in the case of the seasonal component. These expressions are given as follows. In the case of the trend, the coefficients of the function are obtained by means of linear regression methods. As for the lag, the choice is made by maximising the autocorrelation function. All the other parameters of the model (seasonal coefficients and mean of the error) are evaluated according to the formulas presented in subsection 2.1 and the detailed description is reported in subsection 4.1.

3.2 Non–homogeneous Poisson model

In the estimation of the parameters of the non-homogeneous Poisson model under the Bayesian point of view, we take advantage of the natural relationship involving the posterior and the prior distributions and the likelihood function of the model. Hence, we have ([22]), that $P(\boldsymbol{\theta} \,|\, \mathbf{D}) \propto L(\mathbf{D} \,|\, \boldsymbol{\theta})\, P(\boldsymbol{\theta})$ where $P(\boldsymbol{\theta} \,|\, \mathbf{D})$ is the posterior distribution of $\boldsymbol{\theta}$ given the data $\mathbf{D}$, $P(\boldsymbol{\theta})$ is the prior distribution of the parameter $\boldsymbol{\theta}$, and $L(\mathbf{D} \,|\, \boldsymbol{\theta})$ is the likelihood function of the model. Those components will be specified as follows and when applying the model to the data.

Let $V > 0$ and $K > 0$ be fixed real and natural numbers representing, respectively, the total number of observed days and the number of days in which a chosen environmental noise threshold has been surpassed in the time interval $[0, V)$. Let $d_1, d_2, \ldots, d_K$ indicate those days. The set $\mathbf{D} = \{d_1, d_2, \ldots, d_K\}$ will denote, from now on, the set of observed data.

By hypothesis we have a non-homogeneous Poisson model for the problem. Therefore, when no change-points are allowed, the likelihood function is of the following form ([23, 24])

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

where $\lambda(t)$ and $m(t)$ are the rate and mean functions, respectively, with $\boldsymbol{\theta}$ the vector of parameters that need to be estimated. Therefore, with the form considered for the rate function we have that, in the case of no change-points, $\boldsymbol{\theta} = (\alpha, \sigma)$ and

$$L(\mathbf{D} \,|\, \alpha, \sigma) \propto \left(\frac{\alpha}{\sigma^{\alpha}} \right)^{K} \left(\prod_{i=1}^{K} d_i^{\alpha-1} \right) \exp\left[-(V/\sigma)^{\alpha}\right], \tag{4}$$

(see for example [3, 4]).

In some cases it is necessary to consider the presence of change-points. Hence, if $I \geq 0$ change-points are present, let $\tau_1, \tau_2, \ldots, \tau_I$ indicate them. Therefore, we have that the rate function $\lambda(\cdot)$ has the following form,

$$\lambda(t) = \begin{cases} \lambda_1(t), & 0 \le t < \tau_1 \\ \lambda_i(t), & \tau_{i-1} \le t < \tau_i, \qquad i = 2, 3, \dots, I \\ \lambda_{I+1}(t), & \tau_I \le t \le V, \end{cases} \tag{5}$$

where $\lambda_i(t) = (\alpha_i/\sigma_i)\,(t/\sigma_i)^{\alpha_i - 1}$, with $\boldsymbol{\theta}_i = (\alpha_i, \sigma_i)$, $i = 1, 2, \dots, I+1$, the parameters of the non-homogeneous Poisson model between change-points. The mean associated to this rate function is (see for instance [4])

$$m(t \,|\, \boldsymbol{\theta}) = \begin{cases} m_1(t), & 0 \le t < \tau_1, \\ m_1(\tau_1) + m_2(t) - m_2(\tau_1), & \tau_1 \le t < \tau_2 \\ m_{j+1}(t) - m_{j+1}(\tau_j) + & \\ \quad \sum_{i=2}^{j} [m_i(\tau_i) - m_i(\tau_{i-1})] + m_1(\tau_1), & \tau_j \le t \le V, \qquad j = 2, 3, \dots, I, \end{cases} \tag{6}$$

where $m_i(\cdot)$, $i = 1, 2, \dots, I+1$ are the mean functions of the non-homogeneous Poisson process between change-points. In the case of multiple change-points, we take $\boldsymbol{\phi} = (\boldsymbol{\theta}, \boldsymbol{\tau})$, where $\boldsymbol{\tau} = (\tau_1, \tau_2, \dots, \tau_I)$, as the vector of parameters to be estimated. We use $\boldsymbol{\phi}_i$ to denote $\boldsymbol{\phi}$ when $\boldsymbol{\theta} = \boldsymbol{\theta}_i$, $i = 1, 2, \dots, I+1$. Therefore, the likelihood function is of the form (see for instance [3, 4, 25])

$$\begin{aligned} L(\mathbf{D} \,|\, \boldsymbol{\phi}) \;\; \propto \;\; & \left[\prod_{i=1}^{N_{\tau_1}} \lambda_1(d_i) \right] e^{-m_1(\tau_1)} \\ & \left[\prod_{j=2}^{I} \left(\prod_{i=N_{\tau_{j-1}}+1}^{N_{\tau_j}} \lambda_j(d_i) e^{-[m_j(\tau_j) - m_j(\tau_{j-1})]} \right) \right] \\ & \left[\prod_{i=N_{\tau_I}+1}^{K} \lambda_{I+1}(d_i) \right] e^{-[m_{I+1}(V) - m_{I+1}(\tau_I)]}, \end{aligned} \tag{7}$$

where N_{τ_i} represents the number of exceedance days before the change-point τ_i, $i = 1, 2, \dots, I$.

We also assume prior independence of the parameters of the Poisson model. Hence, we have that, $P(\boldsymbol{\theta}) = P(\alpha)\, P(\sigma)$ and, in the case of one change-point, $P(\boldsymbol{\phi}) = P(\boldsymbol{\theta}, \tau) =$

$P(\boldsymbol{\theta} \,|\, \tau)\, P(\boldsymbol{\tau}) = P(\alpha, \sigma \,|\, \tau)\, P(\tau) = P(\alpha \,|\, \tau)\, P(\sigma \,|\, \tau)\; P(\tau)$. The case of multiple change-points follows in a similar way. The prior distributions will be taken, in most of the cases, as uniform distributions defined on appropriate range. However, gamma distributions may also be used.

Remark. Note that when we have uniform prior distributions, then $P(\boldsymbol{\phi} \,|\, \mathbf{D}) \propto L(\mathbf{D} \,|\, \boldsymbol{\phi})$ and/or $P(\boldsymbol{\theta} \,|\, \mathbf{D}) \propto L(\mathbf{D} \,|\, \boldsymbol{\theta})$.

The sampling of the values of $\boldsymbol{\theta}$ and/or $\boldsymbol{\phi}$ will be made using a Gibbs sampling algorithm ([22]) internally implemented in the software OpenBugs (see www.openbugs.net /w, [26, 27]).

3.3 Model selection

Since several versions of the non-homogeneous Poisson model will be used, we need some criteria to select the best model fitting the data. Two criteria will be used. One of them is the graphical criterion where we compare the fit of the estimated and observed accumulated means associated to a given non-homogeneous Poison model. The other criterion is the so-called deviance information criterion (DIC). The smaller the value of DIC the better the model. This criterion may be described as follows. The deviance is defined by $\mathrm{Dev}(\boldsymbol{\theta}) = -2\,\log[L(\mathbf{D} \,|\, \boldsymbol{\theta})] + c$, where $\boldsymbol{\theta}$ is the vector of parameters of the model, $\mathbf{D}$ is the observed data, and c is a constant that is not needed when comparing the models. The DIC ([28]) is given by $\mathrm{DIC} = \mathrm{Dev}(\hat{\boldsymbol{\theta}}) + 2\, n_D$, where $\mathrm{Dev}(\hat{\boldsymbol{\theta}})$ is the deviance evaluated at the posterior mean $\hat{\boldsymbol{\theta}}$ and $n_D = \mathrm{E}[\mathrm{Dev}(\boldsymbol{\theta})] - \mathrm{Dev}(\hat{\boldsymbol{\theta}})$ is the effective number of parameters of the model.

4 An application to the Messina data

In this section we apply the models, described earlier, to the community noise data (obtained from http://mobilitamessina.it/index.php/monitoraggio-ambientale) from the

Viale Boccetta and Via La Farina measuring sites in the city of Messina, Italy. The total time intervals considered are from 11 May 2007 until 10 January 2011 in the case of the Viale Boccetta site, and from 22 April 2008 to 09 November 2010 in the La Farina. Measurements were split into "Day" (corresponding to 6am – 10pm) and "Night" (corresponding to 10pm – 6am) periods.

The observational period has 1341 days in the case of the Viale Boccetta site, and has 932 in the La Farina. Of those days, 214 and 216 had missing measurements in the Viale Boccetta Day and Night periods, respectively. In the case of La Farina Day and Night periods, these numbers were 177 and 179, respectively.

Remark. As in previous works (see for instance [7, 10, 11]) we will use the notation BD and BN to indicate that measurements are from the Viale Boccetta site obtained during the "Day" and "Night" periods, respectively. Similarly we use LFD and LFN to represent the data from the La Farina site.

4.1 Time series analysis

Using the maximisation of the autocorrelation function of the series $\mathbf{X}$ a lag $k = 7$ was detected. This value accommodates the weekly periodicity. The reconstruction of the missing data was performed as follows. In the BD dataset, we have that the first 321 measurements were present. This first group of data is used to calibrate a model that can be used to impute the following 26 missing measurements, as in [10] and [16].

Using the first 321 measurements, a moving average smoothing of the series has been applied. Let $i = 1, 2, \ldots, 7$ correspond to the periods related to Friday, Saturday, Sunday, ..., Thursday, respectively. The estimated seasonal coefficients corresponding to $\overline{S}_i$, $i = 1, 2, \ldots, 7$ are, respectively, 1.01, 1.00, 0.99, 1.00, 1.00, 1.00, and 1.00. We take $n = 1$ in the trend, and the estimated coefficients are $b_0 = 72.80$ and $b_1 = 0.0017$. Finally, the missing values from the 322nd to the 347th days, were estimated using the forecast formula (2).

The distribution of the forecast error $\hat{E}_t$, is characterized by a mean value of -0.009 and a standard deviation of 0.44. After the missing values have been estimated, the dataset is complete for the first 544 days. In order to impute the following missing measurements, another moving average smoothing with lag 7 was applied to this dataset (made of measurements and imputed data). The new estimated coefficients of $\overline{S}_i$ are, respectively, 1.00, 1.00, 0.99, 1.00, 1.00, 1.00, and 1.00. The value $n = 1$ was also used in the trend and the estimated coefficients are $b_0 = 72.94$ and $b_1 = 0.00073$. With the new seasonal coefficients and the new trend line, the missing data have been imputed according to the forecast formula (2).

The resulting mean error of the forecast is -0.027 and the standard deviation is 0.44. With the latter reconstruction, a complete dataset of size 1280 has been obtained. Again, using these values another smoothing is performed using a moving average of lag 7. The estimated coefficients of $\overline{S}_i$, $i = 1, 2, \ldots, 7$ are, respectively, 1.00, 1.00, 0.98, 1.00, 1.00, 1.00, and 1.00. After that, a trend with $n = 1$ has been considered with the estimated coefficients being $b_0 = 73.49$ and $b_1 = -0.0016$. The remaining missing values have been imputed using the usual forecast formula. A mean forecast error of 0.005 and a standard deviation of 0.88 were detected.

Similar procedure is applied to the remaining datasets. In all cases, with the exception of the LFN dataset, a trend with $n = 1$ was needed. In this dataset, a trend with $n = 4$ had to be considered. The estimated parameters of this trend are $b_0 = 69.671$, $b_1 = 0.298$, $b_2 = 0.00011$, $b_3 = -1.35E-07$ and $b_4 = 5.47E-11$, and are the same for all reconstruction of missing data in LFN dataset. Once all datasets have been completed, we proceed to apply the non-homogeneous Poisson model.

Figure 1 shows the plots of the time series composed by the actual measurements and the imputed data (when measurements were missing) using the estimated trend and seasonality in a moving average setting with a span of length $k = 7$.

In Table 1 we have the mean, standard deviation (indicated by SD) as well as the

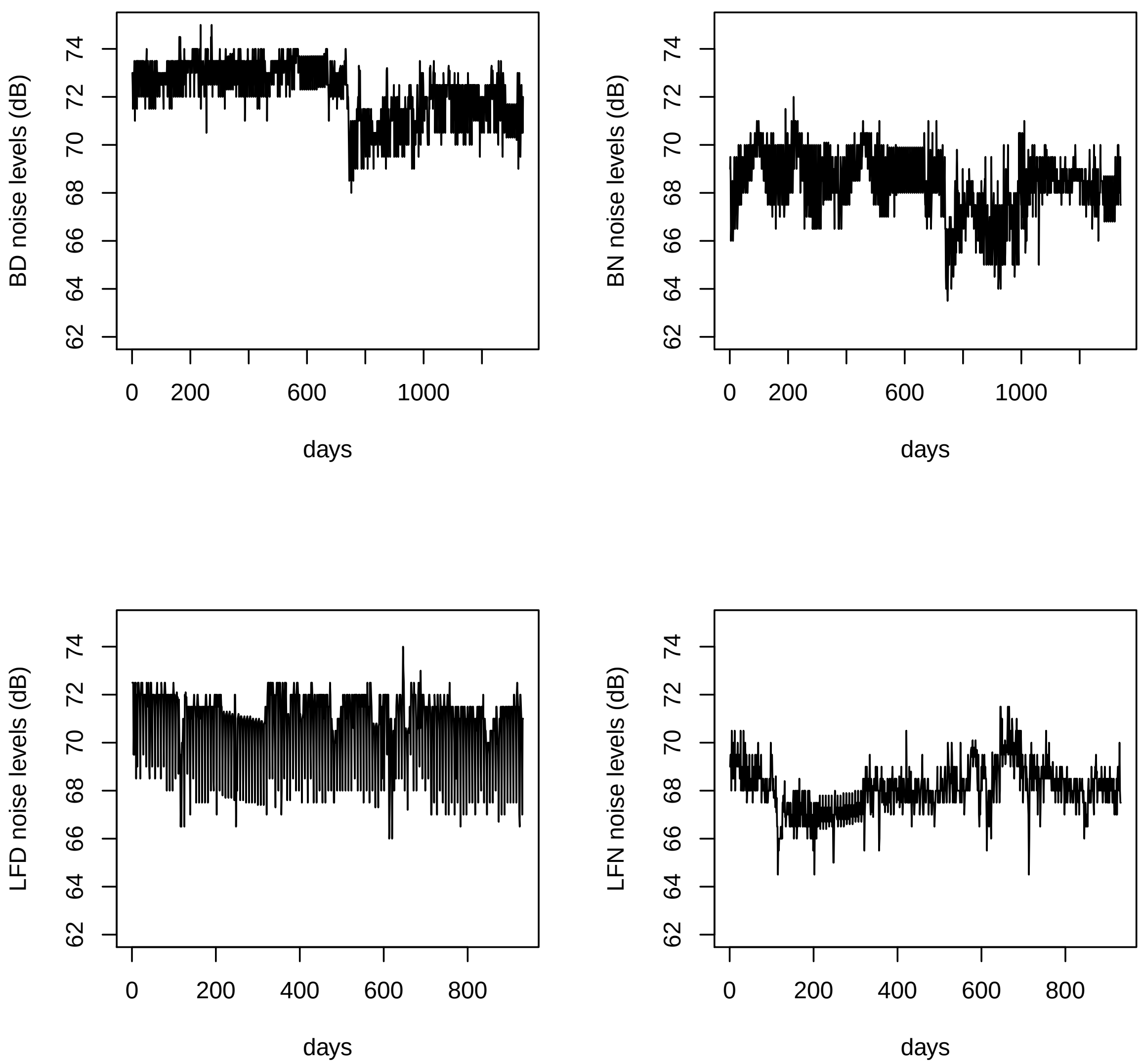


Figure 1: Complete dataset, obtained by merging the actual and the estimated community noise level for both measuring sites and “Day” and “Night” periods..

minimum (Min) and the maximum (Max) measurements in each complete dataset.

	Mean	SD	Min	Max
BD	72.41	1.19	68.0	75.0
BN	68.24	1.31	63.5	72.0
LFD	70.72	1.51	66.0	74.0
LFN	67.6	1.45	62.4	71.5

Table 1: Mean, standard deviation (indicated by SD), maximum and minimum measurements for all datasets after the missing values are estimated.

4.2 Bayesian estimation and the Poisson models

Even though the recommended interval to which the environmental threshold for noise levels in countries in the European Community is 50-55dBA for outdoor noise ([29, 30]), due to the high levels of the measurements used in the present work, we are taking the threshold values 72dBA for the "Day" and 68dBA for the "Night" periods. The use of an artificial, higher thresholds (mentioned above) was made only for the purpose of illustrating the application of the models considered here.

The number of days in which the threshold 72dBA was exceeded in the "Day" period in the case of the Viale Boccetta and La Farina datasets were 971 and 242, respectively. In the case of the Night period the threshold 68dBA was surpassed in 900 and 569 days in those same sites.

Several cases are considered for each dataset. We start by assuming that no change-points are present and then we include them as necessary.

- **No change-points are present**

  In this case the vector of parameters to be estimated is $\boldsymbol{\theta} = (\alpha, \sigma)$ and the prior distributions for the parameters α and σ are the uniform distributions U(0, 3) and U(0, 60), respectively. In the case of the BD dataset a sample of size 20000 was obtained from five chains after a burn-in period of 30000 using a sampling gap of 10. In the remaining datasets, the sample size was 25000 and the burn-in period was 20000. The sampling gap was the same as in the BD dataset. Table 2 gives the means, standard deviations (indicated by SD), the 95% credible intervals of the parameters of the model as well as the values of the DIC when different datasets are considered.

		Mean	SD	95% Credible Interval	DIC
BD	α	0.834	0.027	(0.781, 0.887)	2517
	σ	0.363	0.099	(0.203, 0.58)	
BN	α	0.924	0.03	(0.866, 0.985)	2514
	σ	0.872	0.214	(0.539, 1.362)	
LFD	α	0.734	0.044	(0.6496, 0.825)	1107
	σ	0.575	0.267	(0.198, 1.228)	
LFN	α	1.082	0.045	(0.997, 1.171)	1703
	σ	2.712	0.661	(1.602, 4.179)	

Table 2: Bayesian estimates of the parameters of the non-homogeneous Poisson model for all datasets when no change-points are allowed.

Figure 2 shows the plots of the observed and estimated accumulated means when all datasets are considered and no change-points are allowed

It is possible to see by looking at Figure 2 that even though in some cases such as BN and LFD, the fit is good, we may need to allow the presence of change-points.

- **Presence of one change-point**

In this case the vector of parameters to be estimated is $\boldsymbol{\phi} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \tau)$, where $\boldsymbol{\theta}_i = (\alpha_i, \sigma_i)$, $i = 1, 2$. The uniform prior distributions varied from dataset to dataset. Table 3 gives those distributions.

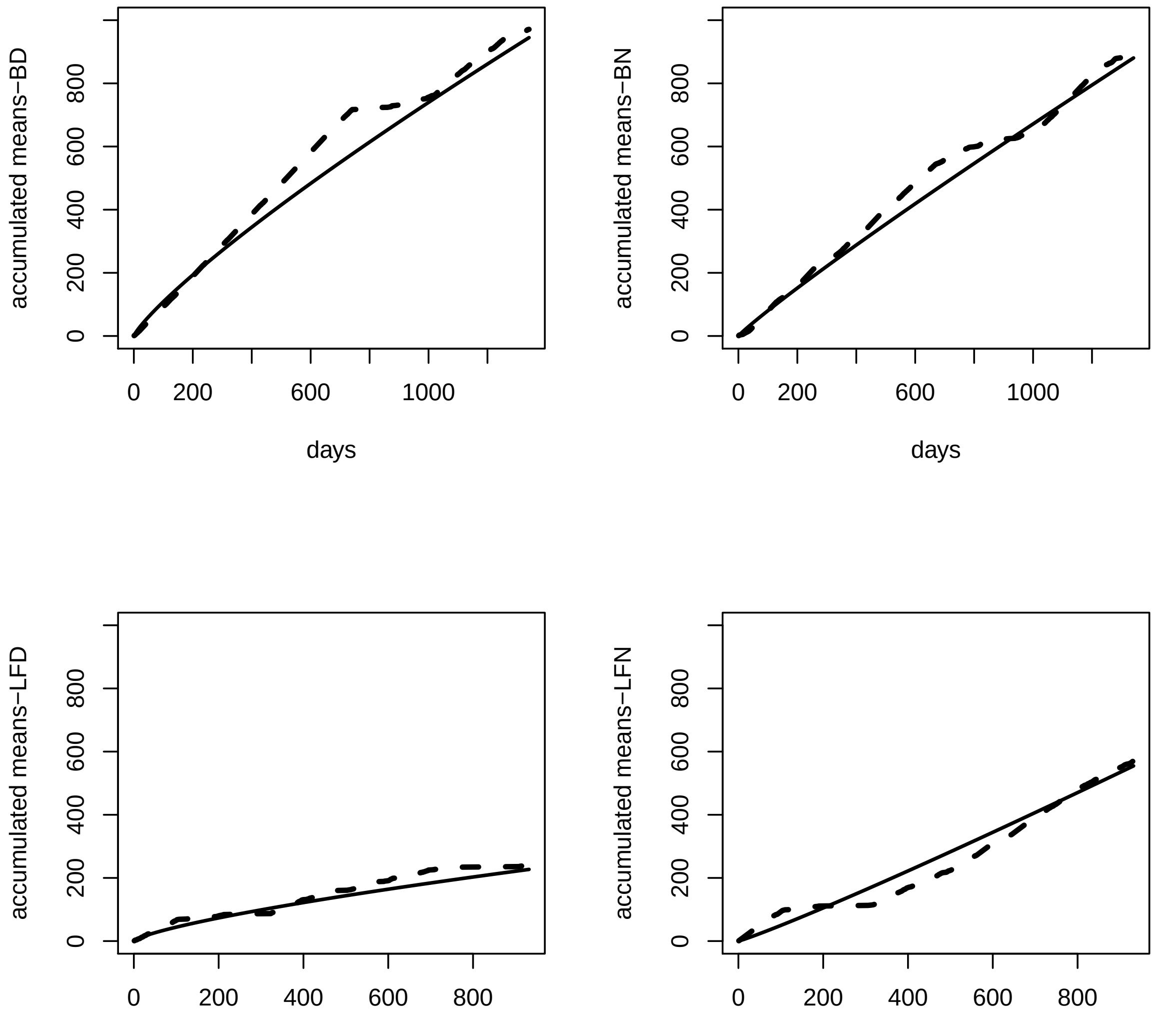


Figure 2: Observed (dashed line) and estimated (continuous line) accumulated means when all datasets are considered and no change-points are allowed.

In the case of the BD dataset, a sample of size 20000 was obtained from five chains after a burn-in period of 30000 steps using a sampling gap of length 10. When the LFN dataset is used, the sample size was 15000 and the burn-in period was of

	BD	BN	LFD	LFN
α_1	U(0.8, 1.5)	U(0.8, 1.5)	U(0.5, 3)	U(0.8, 2)
α_2	U(2.5, 5)	U(2, 3)	U(0.5, 3)	U(1.5, 2.5)
σ_1	U(0.1, 3)	U(1, 7)	U(1, 20)	U(0.8, 10)
σ_2	U(150, 400)	U(90, 150)	U(1, 20)	U(2, 60)
τ	U(720, 800)	U(720, 740)	U(90, 110)	U(95, 120)

Table 3: Prior distributions of the parameters when all datasets are considered and one change-point is allowed.

40000 iterations. In the case of the BN and LFD datasets the sample size was 25000 and the sample was collected after a burn-in period of 20000 steps. The number of chains and the sampling gap was as in the BD dataset.

Table 4 presents the estimated quantities of interest as well as the 95% credible intervals and the value of the DIC when each dataset is considered.

In Figure 3 we have the plots of observed and estimated accumulated means for all dataset when one change-point is allowed.

Note that even though the fit is almost perfect. There are some indication that a second change-point might exist. Hence, we consider that case as well.

- **Presence of two change-points**

  When two change-points are allowed, we have that the vector of parameters is $\boldsymbol{\phi} = (\boldsymbol{\theta_1}, \boldsymbol{\theta}_2, \boldsymbol{\theta_3}, \boldsymbol{\tau})$, where $\boldsymbol{\theta_i} = (\alpha_i, \sigma_i)$ $i = 1, 2, 3$, and $\boldsymbol{\tau} = (\tau_1, \tau_2)$. The uniform prior distributions also varied according to the dataset. Table 5 gives the prior distributions in each case.

  Estimation of the parameters was made using a sample of size 15000 in the case of the BD and LFN datasets. They were collected from five chains after a burn-in

		Mean	SD	95% Credible Interval	DIC
BD	α_1	1.032	0.039	(0.957, 1.111)	2449
	α_2	3.91	0.371	(3.177, 4.556)	
	σ_1	1.297	0.314	(0.767, 2.007)	
	σ_2	315.7	46.66	(222.9, 391.7)	
	τ	741	1.844	(736.5, 744.6)	
BN	α_1	1.058	0.040	(0.983,1.138)	2451
	α_2	2.6	0.106	(2.33, 2.737)	
	σ_1	1.807	0.4154	(1.11, 2.726)	
	σ_2	134.4	13.05	(137.8, 149.5)	
	τ	732	4.553	(722.5, 739.3)	
LFD	α_1	1.126	0.1186	(0.929, 1.382)	146500
	α_2	1.075	0.118	(0.846, 1.302)	
	σ_1	2.582	1.03	(1.12, 5.02)	
	σ_2	7.699	4.052	(1.723, 17.24)	
	τ	103	3.019	(95.04, 108.8)	
LFN	α_1	1.05	0.08	(0.93, 1.235)	1882
	α_2	2.039	0.1007	(1.834, 2.22)	
	σ_1	1.43	0.486	(0.825, 2.687)	
	σ_2	45.45	6.794	(32.07, 58.05)	
	τ	107	2.871	(102.4, 112.3)	

Table 4: Bayesian estimates of the parameters of the Poisson model for all datasets when one change-point is allowed.

period of 50000 and 40000, respectively. In the case of BN and LFD the sample size was 10000 and the burn-in period was 100000 iterations. The sampling gap was the same in all cases and it was equal to 10. Tables 6 and 7 shows the estimated

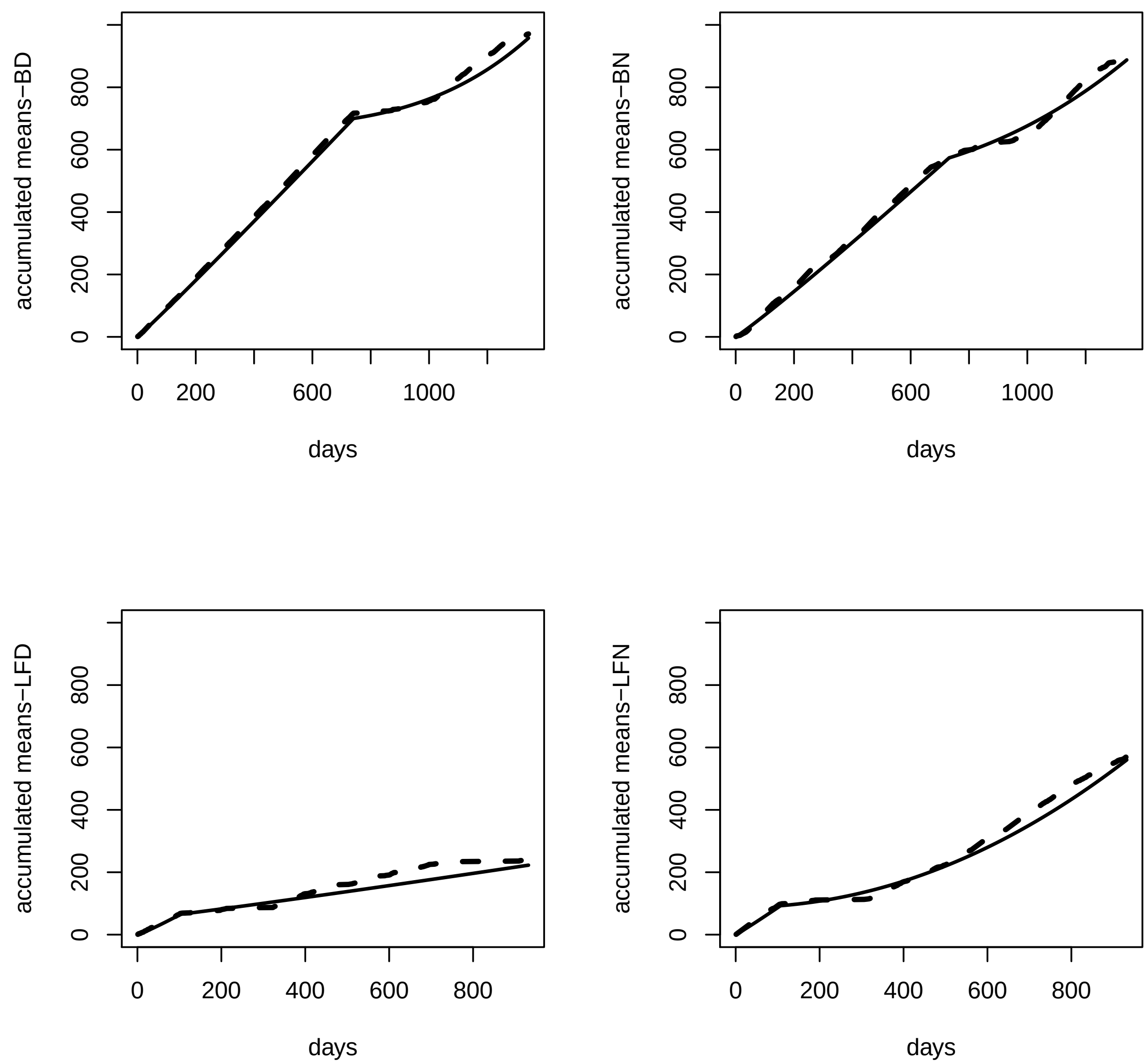


Figure 3: Observed (dashed line) and estimated (continuous line) accumulated means when all datasets are considered and one change-point is allowed.

quantities of interest as well as the values of DIC in all cases.

Figure 4 shows the plots of the estimated and observed accumulated means for all datasets when two change-points are present.

	BD	BN	LFD	LFN
α_1	U(0.8, 1.5)	U(0.8, 1.8)	U(0.1, 2)	U(0.8, 1.2)
α_2	U(0.5, 1.5)	U(0.5, 0.8)	U(0.5, 1.5)	U(0.5, 1.8)
α_3	U(0.6, 1.9)	U(1.1, 1.8)	U(0.1, 1)	U(0.9, 1.8)
σ_1	U(0.1, 3)	U(0.1, 5)	U(0.1, 10)	U(0.5, 2)
σ_2	U(0.1, 20)	U(0.1, 40)	U(0.1, 40)	U(0.1, 50)
σ_3	U(0.1, 20)	U(0.9, 40)	U(0.1, 10)	U(0.1, 20)
τ_1	U(700, 750)	U(730, 750)	U(100, 130)	U(100, 130)
τ_2	U(950, 1050)	U(950, 1000)	U(300, 400)	U(300, 3400)

Table 5: Prior distributions of the parameters when all datasets are considered and two change-points are allowed.

Note that even though in the case of the BD and LFD datasets the value of the DIC is smaller, we may notice that the fit of the estimated accumulated means to the observed ones are worse for measurements towards the end of the observational period when compared to the case of one change-point. The overall fit in the case of one change-point is better than when two change-points are allowed. However, in the beginning of the observational period the fit is improved when two change-points are present. Note that in the case of LFN the smallest DIC is when no change-points are allowed. However, looking at Figures 2, 3, and 4, we may see that the best graphical fit is when only one change-point is present. In the case of LFD we may see that even with two change-points the fit has not improved substantially given the best fit in the case of only one-change point. Therefore, we have decided to consider the case of three change-points for the LFD dataset and see if any improvement is achieved.

- **LFD with three change-points**

		Mean	SD	95% Credible Interval	DIC
BD	α_1	1.031	0.037	(0.962, 1.111)	2385
	α_2	1.095	0.114	(0.823, 1.251)	
	α_3	1.229	0.173	(0.901, 1.503)	
	σ_1	1.291	0.278	(0.795, 1.973)	
	σ_2	12.54	5.271	(2.051, 19.72)	
	σ_3	7.867	5.492	(0.689, 18.96)	
	τ_1	741	1.74	(736.6, 744.2)	
	τ_2	994	13.34	(953.3, 1020)	
BN	α_1	1.05	0.0425	(0.966,1.135)	2761
	α_2	1.318	0.174	(0.912, 1.551)	
	α_3	1.424	0.191	(1.123, 1.772)	
	σ_1	1.729	0.018	(1.005, 2.66)	
	σ_2	23.41	10.79	(2.779, 39.25)	
	σ_3	14.9	3.161	(12.2, 35.73)	
	τ_1	735	3.737	(730.3, 741.8)	
	τ_2	986	5.152	(974, 993.6)	

Table 6: Bayesian estimates of the parameters of the Poisson model for the Viale Boccetta datasets when two change-points are allowed.

When the LFD dataset is considered and three change-points are allowed, we have that the vector of parameters to be estimated is $\boldsymbol{\phi} = (\boldsymbol{\theta_1}, \boldsymbol{\theta}_2, \boldsymbol{\theta_3}, \boldsymbol{\theta_4}, \boldsymbol{\tau})$, where $\boldsymbol{\theta_i} = (\alpha_i, \sigma_i)$ $i = 1, 2, 3, 4$, and $\boldsymbol{\tau} = (\tau_1, \tau_2, \tau_3)$. The forms and hyperparameters of the prior distributions vary according to the parameter. In the case of α_1, α_2 , α_3 , α_4, σ_1, τ_1, τ_2, and τ_3, the prior distributions are, respectively, the uniform distributions U(0.1, 2), U(0.1, 1.5), U(0.8, 1.5), U(0.5, 1.5), U(0.1, 8), U(90, 110), U(300, 400), and U(700, 800). When we consider the parameters σ_2, σ_3, and σ_4, Gamma(a, b)

		Mean	SD	95% Credible Interval	DIC
LFD	α_1	1.095	0.1274	(0.8638, 1.353)	1042
	α_2	1.037	0.227	(0.5624, 1.428)	
	α_3	0.695	0.059	(0.6264, 0.844)	
	σ_1	2.35	1.028	(0.788, 4.635)	
	σ_2	16.55	10.391	(0.543, 37.07)	
	σ_3	0.337	0.337	(0.104, 1.301)	
	τ_1	104	2.762	(100.7, 110.2)	
	τ_2	319	4.045	(308.3, 323.6)	
LFN	α_1	1.003	0.075	(0.8646, 1.137)	1807
	α_2	1.088	0.265	(0.605, 1.597)	
	α_3	1.333	0.1467	(1.006, 1.592)	
	σ_1	1.163	0.3799	(0.545, 1.907)	
	σ_2	20.49	13.62	(1.169, 47.68)	
	σ_3	8.454	4.25	(1.415, 17.68)	
	τ_1	108	2.96	(102.8, 112.9)	
	τ_2	315	5.087	(303.8, 323.6)	

Table 7: Bayesian estimates of the parameters of the Poisson model for the La Farina datasets when two change-points are allowed.

prior distributions are considered. (Here, we consider a gamma distribution whose mean and variance are, respectively, a/b and a/b^2.). Therefore, σ_2, σ_3, and σ_4 have as their prior distributions Gamma(42, 3), Gamma(16, 4), and Gamma(1.44, 0.24), respectively.

Estimation of the parameters was made using a sample of size 10000 collected from five chains after a burn-in period of length 50000 using a sampling gap of 10 iterations. The means, standard deviations (indicated by SD), the 95% credible intervals

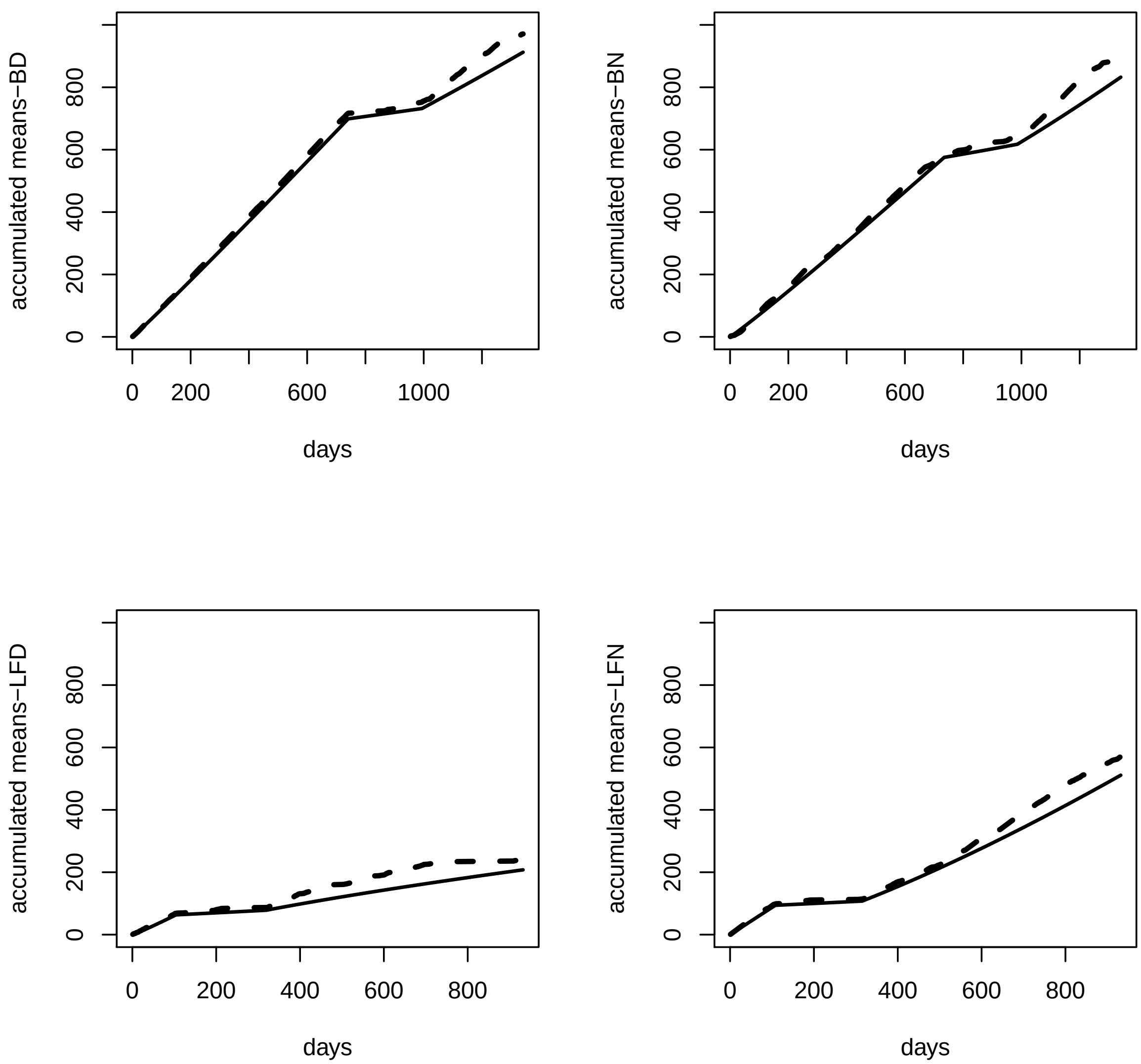


Figure 4: Observed (dashed line) and estimated (continuous line) accumulated means when all datasets are considered and two change-points are allowed.

of the quantities of interest as well as the value of the DIC are given in Table 8.

In Figure 5 we have the estimated and observed accumulated means in the case of the LFD dataset when three change-points are allowed.

		Mean	SD	95% Credible Interval	DIC
LFD	α_1	1.11	0.12	(0.89, 1.37)	1006
	α_2	1.03	0.08	(0.87, 1.89)	
	α_3	1.04	0.05	(0.94, 1.14)	
	α_4	0.83	0.12	(0.6, 1.08)	
	σ_1	2.49	1.05	(0.88, 4.95)	
	σ_2	13.92	2.17	(10.02, 18.45)	
	σ_3	3.73	0.98	(2.03, 5.85)	
	σ_4	7.1	5.48	(0.61, 21.06)	
	τ_1	104	2.52	(100.5, 109.1)	
	τ_2	319	3.96	(308.9, 324.7)	
	τ_3	726	18.42	(700.6, 761.2)	

Table 8: Bayesian estimates of the parameters of the Poisson model for the La Farina Day dataset when three change-points are allowed.

Looking at Figure 5 we may see that the fit is good even thought the estimated accumulated mean underestimate the observed one.

4.3 Model selection

If we use the DIC to decide which model fits best the observed behaviour of the data, we have, by looking at Tables 2, 4, 6, 7, 8, that the selected model is the Poisson with no change-points in the case of the LFN dataset, Poisson with one change-point in the case of BN dataset, Poisson with two change-points in the case of BD dataset, and Poisson with three change-points in the case of LFD dataset. However, looking at Figures 2, 3, 4, and 5, we may see that in the best overall fit is provided by the Poisson model with one change-point in all cases. Therefore, this is the case we are going to consider to illustrate the applications of the model to the estimation of the probability that a population is

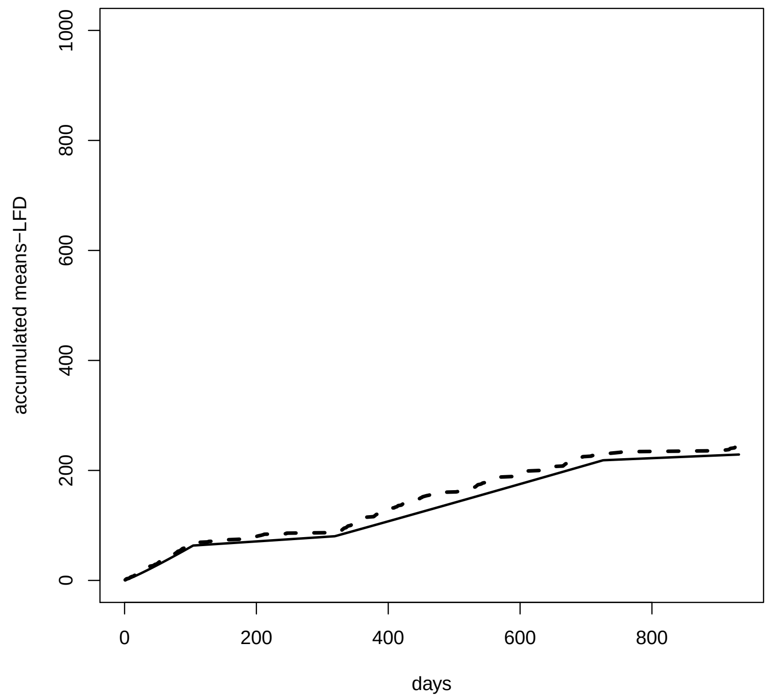


Figure 5: Observed (dashed line) and estimated (continuous line) accumulated means when the LFD dataset is considered and three change-points are allowed.

exposed to noise levels above a given threshold a certain number of times in a time interval of interest.

4.4 Calculating probabilities

In this subsection we will provide a way of calculating the probability of some events of interest. In all cases we use the LFN dataset. Considering the graphical criterion as the one used to select the best model fitting the data, we have that the chosen model is the non-homogeneous Poisson model with one change-point. Hence, that is the model we take.

- **Estimating probabilities of exceedances in a future time**

  Assume that we want to calculate the probability that during the night a population will be exposed to noise levels above 68dBA five times in the next 30 days after the observational period is over. Hence, we want to calculate the probability that during the time interval [932, 962] the threshold 68dBA is exceeded five times. Since the time interval belongs to the time segment after the change-point, we have that the parameters of the mean function of the Poisson process are $\alpha_2 = 2.039$ and $\sigma_2 = 45.45$ (see Table 4). Hence, from (3) the probability of interest is

$$\begin{aligned}
P(N_{932+30} - N_{932} = 5) &= \frac{\left[\left(\frac{932+30}{45.45}\right)^{2.039} - \left(\frac{932}{45.45}\right)^{2.039}\right]^5}{5!} \\
&\times \exp\left(-\left[\left(\frac{932+30}{45.45}\right)^{2.039} - \left(\frac{932}{45.45}\right)^{2.039}\right]\right) \\
&\approx 5.09E - 09.
\end{aligned}$$

  Another question that may be asked is related to the probability that in those same 30 days we have between five and eight exceedances of the threshold 68dBA. In this case the probability is

$$\begin{aligned}
P(5 \leq N_{932+30} - N_{932} \leq 8) &= P(N_{962} - N_{932} \leq 8) - P(N_{962} - N_{932} < 5) \\
&= \sum_{k=5}^{8} P(N_{962} - N_{932} = k) \\
&= \sum_{k=5}^{8} \left\{ \frac{\left[\left(\frac{962}{45.45}\right)^{2.039} - \left(\frac{932}{45.45}\right)^{2.039}\right]^k}{k!} \right. \\
&\times \left. \exp\left(-\left[\left(\frac{962}{45.45}\right)^{2.039} - \left(\frac{932}{45.45}\right)^{2.039}\right]\right)\right\} \\
&\approx 6.297E - 07.
\end{aligned}$$

- **Comparing probabilities of events before and after the change-point**

Take now the time intervals [50, 70] and [120, 140]. Suppose we want to know the probability of having five exceedances of the threshold 68dBA in each of them. Note that the change-point is $\tau = 107$. Thus, we are comparing the probabilities of an event in time intervals of equal lengths, but one of them is before the change-point and the other is after. Hence, we want to know the values of $P(N_{70} - N_{50} = 5)$ and $P(N_{140} - N_{120} = 5)$. Before the change-point we have that the estimated parameters of the Poisson model are $\alpha_1 = 1.05$ and $\sigma_1 = 1.43$ (see Table 4), then

$$\begin{aligned} P(N_{70} - N_{50} = 5) &= \frac{\left[\left(\frac{70}{1.43}\right)^{1.05} - \left(\frac{50}{1.43}\right)^{1.05}\right]^5}{5!} \\ &\times \exp\left(-\left[\left(\frac{70}{1.43}\right)^{1.05} - \left(\frac{50}{1.43}\right)^{1.05}\right]\right) \\ &\approx 2.098E-04, \end{aligned}$$

and in the case where the time interval is located after the change-point, we have

$$\begin{aligned} P(N_{140} - N_{120} = 5) &= \frac{\left[\left(\frac{140}{45.45}\right)^{2.039} - \left(\frac{120}{45.45}\right)^{2.039}\right]^5}{5!} \\ &\times \exp\left(-\left[\left(\frac{140}{45.45}\right)^{2.039} - \left(\frac{120}{45.45}\right)^{2.039}\right]\right) \\ &\approx 7.86E-07. \end{aligned}$$

- **Estimating the probability of exceedances in a time interval containing a change-point**

  Take for instance the time interval [90, 120] and assume that we want to know the probability of having three exceedances in this interval. Note that the estimated change-point is $\tau = 107$ which belongs to the time interval in consideration. The change-point marks the point in time where the process changes behaviour. Hence, we have to split [90, 120] into two parts, one before the change-point and another after the change-point. Recall that we have $\lambda(t) = \lambda_1(t)$, $t < \tau$ and $\lambda(t) = \lambda_2(t)$, $t \geq \tau$. Therefore, the time interval is split into $V_1 = [90, 107)$ and $V_2 = [107, 120]$. In the first time interval the parameters of the Poisson rate function are $\alpha_1 = 1.05$ and $\sigma_1 = 1.43$. In the second, those parameters change to $\alpha_2 = 2.039$ and $\sigma_2 = 45.45$. There are several ways in which those exceedances may occur. We may have no exceedances in V_1 and all of them occurring in V_2, or we may have one exceedance occurring in V_1 and two occurring in V_2, or we may have two exceedances occurring in V_1 and one occurring in V_2, or or we may have three exceedance occurring in V_1 and none in V_2. Hence, the probability sought is

$$\begin{aligned}
P(N_{120} - N_{90} = 3) &= \sum_{k=0}^{3} \left[P(N_{107} - N_{90} = k)\, P(N_{120} - N_{107} = 3 - k) \right] \\
&= \sum_{k=0}^{3} \Bigg[\Bigg\{ \frac{\left[\left(\frac{107}{1.43}\right)^{1.05} - \left(\frac{90}{1.43}\right)^{1.05} \right]^k}{k!} \\
&\qquad \exp\left(- \left[\left(\frac{107}{1.43}\right)^{1.05} - \left(\frac{90}{1.43}\right)^{1.05} \right] \right) \Bigg\} \\
&\times \Bigg\{ \frac{\left[\left(\frac{120}{45.45}\right)^{2.039} - \left(\frac{107}{45.45}\right)^{2.039} \right]^{3-k}}{(3-k)!} \\
&\qquad \exp\left(- \left[\left(\frac{120}{45.45}\right)^{2.039} - \left(\frac{107}{45.45}\right)^{2.039} \right] \right) \Bigg\} \Bigg] \\
&\approx 2.0E - 07 \times 0.13 + 3.0E - 06 \times 0.25 \\
&+ 2.38E - 05 \times 0.33 + 1.22E - 04 \times 0.22 \\
&= 3.585E - 05.
\end{aligned}$$

Additionally, note that the probability of having three or less exceedances in the time interval [107, 120] is

$$
\begin{aligned}
P(N_{120} - N_{107} \leq 3) &= \sum_{k=0}^{3} P(N_{120} - N_{107} = k) \\
&= \sum_{k=0}^{3} \left\{ \frac{\left[\left(\frac{120}{45.45}\right)^{2.039} - \left(\frac{107}{45.45}\right)^{2.039}\right]^k}{k!} \right. \\
&\qquad \left. \exp\left(-\left[\left(\frac{120}{45.45}\right)^{2.039} - \left(\frac{107}{45.45}\right)^{2.039}\right]\right)\right\} \\
&\approx 0.22 + 0.33 + 0.25 + 0.13 \approx 0.93.
\end{aligned}
$$

Hence, the probability of having more than three exceedances in that interval is

$$
P(N_{120} - N_{107} > 3) = 1 - \sum_{k=0}^{3} P(N_{120} - N_{107} = k) \approx 1 - 0.93 = 0.07
$$

Other variations of the questions considered here may be posed and they may be answered in the same fashion.

5 Discussion

In this work we have considered two modelling stages to study the behaviour of community noise data. In the first stage, subsets of missing data were estimated using a time series model. The model considered is multiplicative involving the trend and the seasonal components and additive in the error component. Using the estimated components of the time series model, missing data were estimated through the forecast values. In the second stage, after the missing data are imputed, a non-homogeneous Poisson model with Weibull rate function is used in the complete dataset to estimate the probability that a population is exposed to community noise level above a certain threshold a given number of times in a time interval of interest.

The components of the time series were estimated using standard time series methodologies. In the case of the non-homogeneous Poisson model, a Bayesian point of view is followed. Due to the complexity of the distribution functions involved in the model, parameters were estimated using a Markov chain Monte Carlo (MCMC) algorithms. The algorithm used was the Gibbs sampling internally implemented in the software OpenBugs (http://www.openbugs.info/w). In this case we simply have to specify the likelihood function of the model and the prior distributions of the parameters involved. Programmes used to estimate the parameters are a straightforward modification of the ones given in [7] and [31].

Looking at Tables 2, 4, 6, 7, and 8, we may see that the smallest value of the DIC corresponds to the Poisson model with two change-points in the case of BD, one change-point in the case of BN dataset, three change-points in the LFD case, and no change-points in the case of LFN datase. However, we may see from Figures 2, 3, 4, and 5, that for all datasets, with the exception of the LFD, the best graphical fit corresponds to the non-homogeneous Poisson model with one change-point. In the case of the LFD dataset we have that the model with three change-points provide a very good fit. However, we may also notice that the observed accumulated mean is approximated by a curve composed by four segments which are approximately straight lines. Similar approximation is observed in the case of two change-points. When one change-point is considered we may see that a good fit is provided and the approximating curves are not straight lines. Hence, it is clear that the larger the number of change-points the better the fit, but the computational time increases accordingly. Thus, we must optimise the adequacy of the fit of the estimated curve to the observed one and the time spent in the estimation of the parameters of interest. Therefore, when comparing the computational time and the fit of the curves it seems to be enough to take only one change-point in the LFD dataset.

Using the graphical criterion for selecting the best model to explain the behaviour of the data, we have obtained the probabilities that a population is exposed to noise levels that exceed a given threshold a certain number of time in time intervals of interest. The dataset considered to illustrate this calculation was the LFN dataset. We may see that when taking two time intervals of the same length but with one located before and and the other after the change-point, there are substantial changes in the probability of exposure. It would be interesting to investigate the causes of this changes and, in particular, what have caused the presence of the change-point.

The change-point in the case of the LFN, corresponds to a day in the beginning of August 2008. The change that occurred was a decrease in the rate function at which noise exceedances occurred. If we plot the rate functions with the estimated parameters before the estimated change-point, we may see that the plot is always larger for all values of $t \geq 0$, than when we use the estimated parameters after the change-point. That may be observed by looking at Figure 6 top plots.

The decrease in the rate function (in the case of one change-point) could be caused by a decrease in road traffic due to the holiday period. The other possible change-point corresponds to a day in the beginning of March 2009. When we consider the plots of the rate functions using the estimated parameters in the case of two change-points, the plots using parameters before the first change-point produce a figure that is larger for all $t \geq 0$ than the curve using the estimated parameters after that change-point and before the second change-point (see plots at the bottom of Figure 6). It is also possible to see that using the estimated parameters after the second change-point, the rate function lies above the curve with the estimated parameters between the first and second change-points, but below the corresponding curve using the estimated parameters before the first change-point. That means that after the second change-point, the rate at which exceedances occur is larger than that between the first and second change-points. However, this rate is smaller than at times prior to the first change-point. Therefore, even though a deterioration in the noise levels has occurred around March 2009, that is not as bad as in the beginning of the observational period.

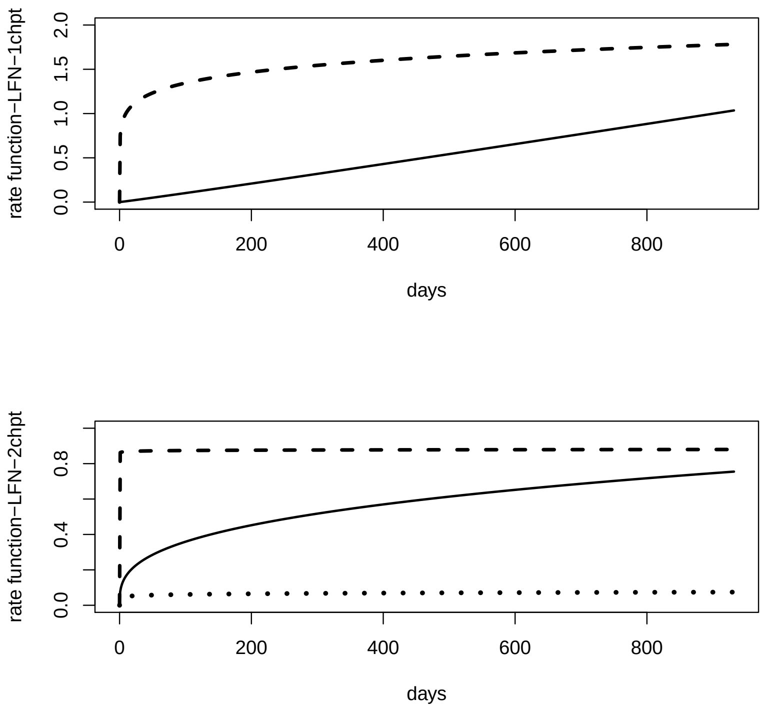


Figure 6: Estimated rate functions in the case of LFD dataset when different scenarious are considered. Plots at the top of the figure are the case where only one change-point is considered. Dashed line represents the rate function when the estimated parameters are the ones before the change-point and the continuous line is the case of parameters after the change-point. Bottom plots represent the case with two change-points. In that case the dashed lines indicate the rate function with the estimated parameters before the first change-point, the dotted line is the rate function with estimated parameters between the first and the second change-points and the continuous line represents the rate function with the estimated parameters after the second change-point

Acknowledgements

The authors thank Department of Urban Mobility of the city of Messina–Italy, for making public the noise data. This work was partially funded by the research project PAPIIT-

IN102416 of the Dirección General de Apoyo al Personal Académico of the Universidad Nacional Autónoma de México, Mexico, and by the research grant number 3651, Legge 5/02, Regione Campania, Italy. ERR thanks all at the Department of Industrial Engineering of the University of Salerno, Italy, for their hospitality and support.

References

[1] De Kluizenaar Y, Janssen SA, van Lenthe FL, Miedema HME, Mackenbach JP. Long-term road traffic noise exposure associated with an increase morning tiredness. Journal of the Acoustic Society of America 2009; 126:626-633.

[2] WHO). World Health Organization. Guidelines for Community Noise. Berlund B, Lindvall T, Schewela DH, editors. Geneva: World Health Organization; 1999.

[3] 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. http://dx.doi.org/10.1002/env.890.

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

[5] Gouveia N, Fletcher T. Time Series Analysis of Air Pollution and Mortality: Effects by Cause, Age and Socio-Economics Status. Journal of Epidemiology and Community Health 2000; 54:750-755.

[6] 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.

[7] Guarnaccia C, Quartieri J, Barrios JM, Rodrigues ER. Modeling environmental noise exceedances using non-homogeneous Poisson processes. Journal of the Acoustic Society of America 2014; 136:1631-1639. http://dx.doi.org/10.1171/1.4895662.

[8] Guarnaccia C, Quartieri J, Tepedino C, Rodrigues ER. An analysis of airport noise data using a non-homogeneous Poisson model with a change-point. Applied Acoustics 2015; 91:33-39. http://dx.doi.org/10.1016/j.apacoust.2014.12.002

[9] Guarnaccia C, Cerón-Bretón TC, Quartieri J, Tepedino C, Cerón-Bretón, RM. An application of time series analysis for forecasting and control of carbon monoxide concentrations. International Journal of Mathematical Models and Methods in Applied Sciences 2014; 8:505-515.

[10] Guarnaccia C, Quartieri J, Rodrigues ER, Tepedino C. Acoustical Noise Analysis and Prediction by means of Multiple Seasonality Time Series Model. International Journal of Mathematical Models and Methods in Applied Sciences 2014; 8:384-393.

[11] Guarnaccia C, Quartieri J, Mastorakis NE, Tepedino C. Development and Application of a Time Series Predictive Model to Acoustical Noise Levels. WSEAS Transactions on Systems 2014; 13:745-756.

[12] Karlin S, Taylor HM. A first course in Stochastic Processes. 2nd ed. USA: Academic Press; 1975.

[13] Box GEP, Jenkins G, Reinsel GC. Time Series Analysis: Forecasting and Control. 4th ed. New Jersey: John Wiley and Sons; 2008.

[14] Weber R. Time Series Lecture Notes. University of Cambridge. Department of Pure Mathematics and Mathematical Statistics; 2013. (www.statslab.cam.ac.uk/ rrw1/timeseries/t.pdf)

[15] STAT-510. Applied Times Series. Penn State Online Courses. (http://onlinecourses.sciences.psu.edu/stat510)

[16] Guarnaccia C, Quartieri J, Tepedino C, Petrovic CL. A Comparison of Imputation Techniques in Acoustic Level Datasets. International Journal of Mechanics 2015; 9: 272-278.

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

[18] Ross SM. Stochastic Processes. 2nd ed. USA: John Wiley and Sons; 1996.

[19] 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.

[20] Smith AFM, Roberts GO. Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods (with discussion). Journal of the Royal Statistical Society Series B 1993; 55:3-23

[21] Robert CP, Casella G. Monte Carlo statistical methods. New York, USA: Springer; 1999.

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

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

[24] Lawless JF. Statistical Models and Methods for Lifetime Data. USA: John Wiley and Sons; 1982.

[25] 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. http:// dx.doi.org/10.1198/106186001317243449.

[26] Gelfand AE, Smith AFM. Sampling-based approaches to calculating marginal densities. Journal of the American Statistical Association 1990; 85:398–409

[27] Lunn D, Spiegelhalter D, Thomas A, Best N. The BUGS project: Evolution, critique and future directions (with discussion). Statistics in Medicine 2009; 28:3049-3082.

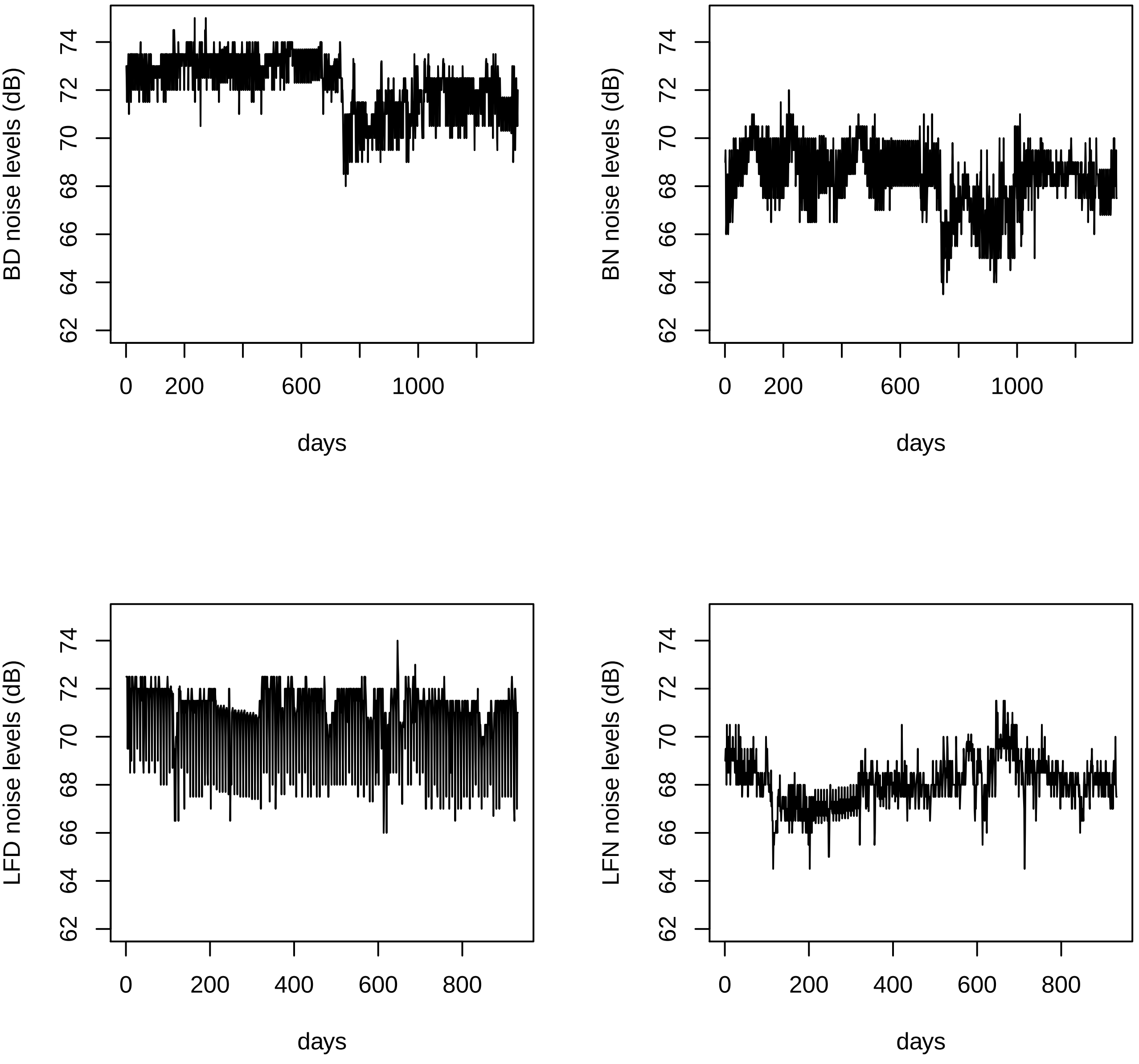
[28] Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society Series B 2002; 64: 583-639.

[29] WG-AEN. European Working Group for Assessment of Exposure to Noise (WG-AEN). *Good practice guide for strategic noise mapping and the production of associated data on noise exposure*. Version 2. Europe; 2006.

[30] DEPC. Directive 2002/49/EC of the European Parliament and of the Council of 25 June 2002. Official Journal of the European Communities. Europe 2002:L189/12-L189/25.

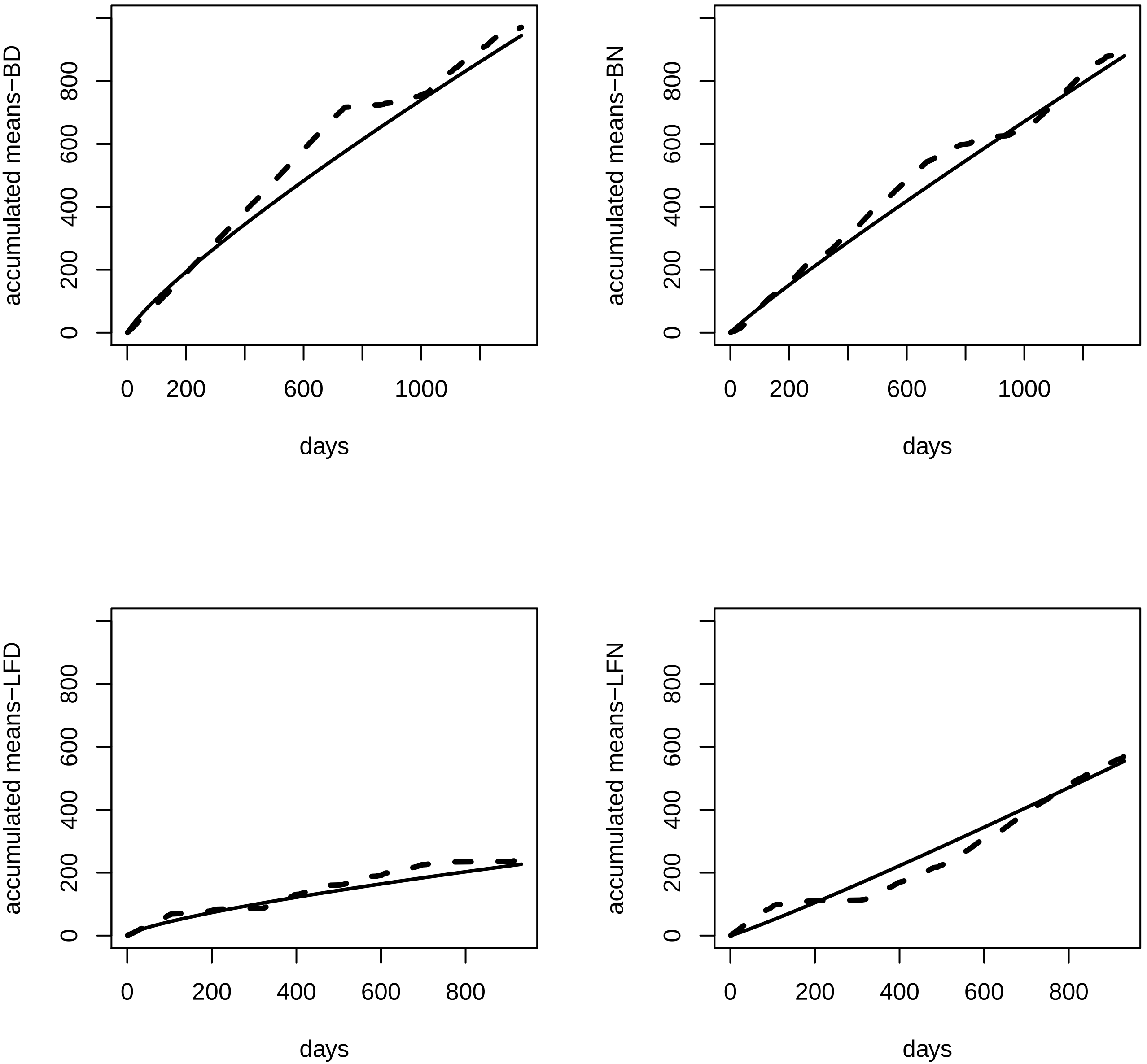
[31] Achcar JA, Loibel S, Andrade MG. Interfailure data with constant hazard function in the presence of change-points. REVSTAT - Statistical Journal 2007; 5:209-226.

Figure



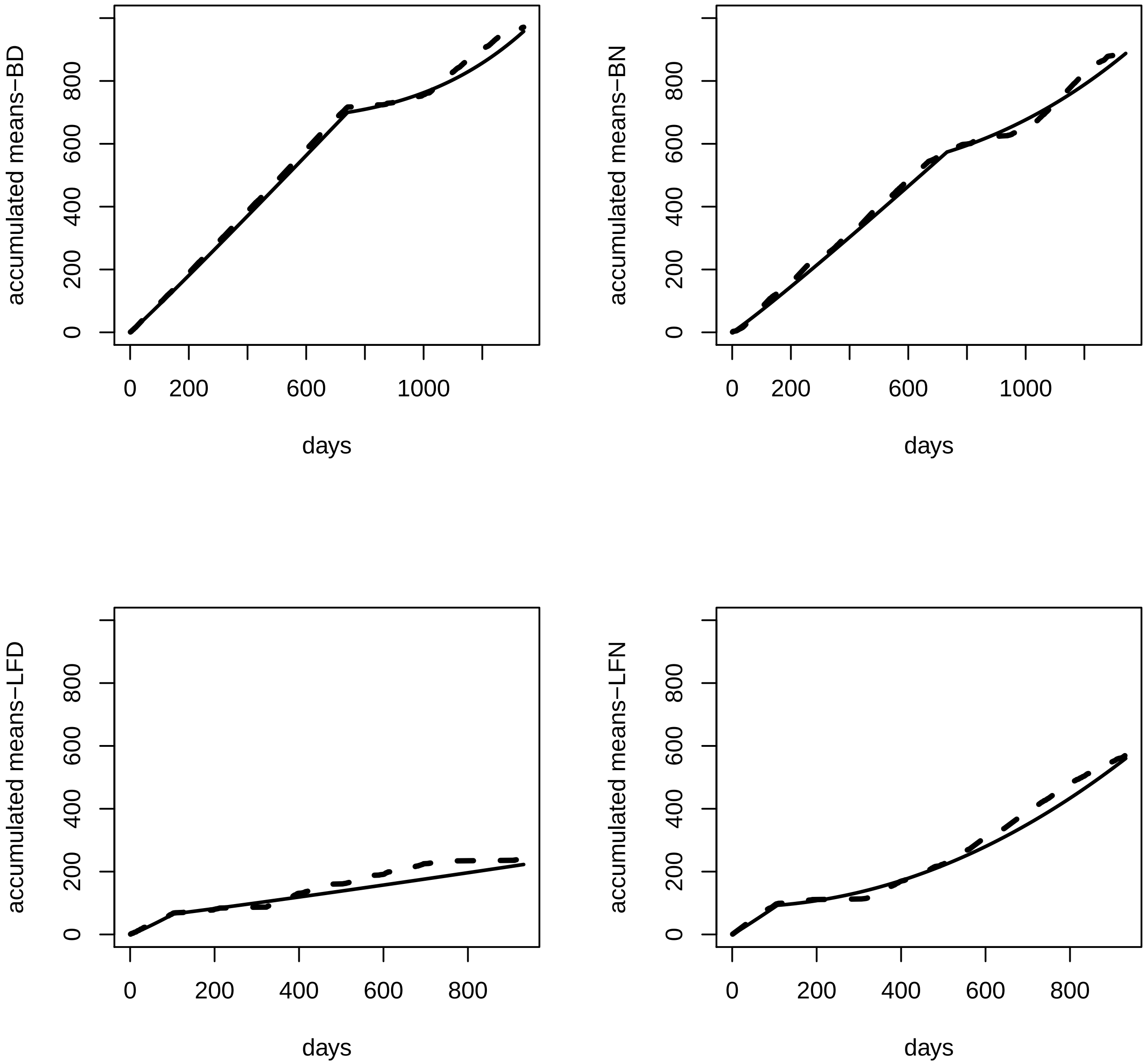
BD noise levels (dB)
BN noise levels (dB)
LFD noise levels (dB)
LFN noise levels (dB)
62
64
66
68
70
72
74
0
200
400
600
800
1000
days

Figure

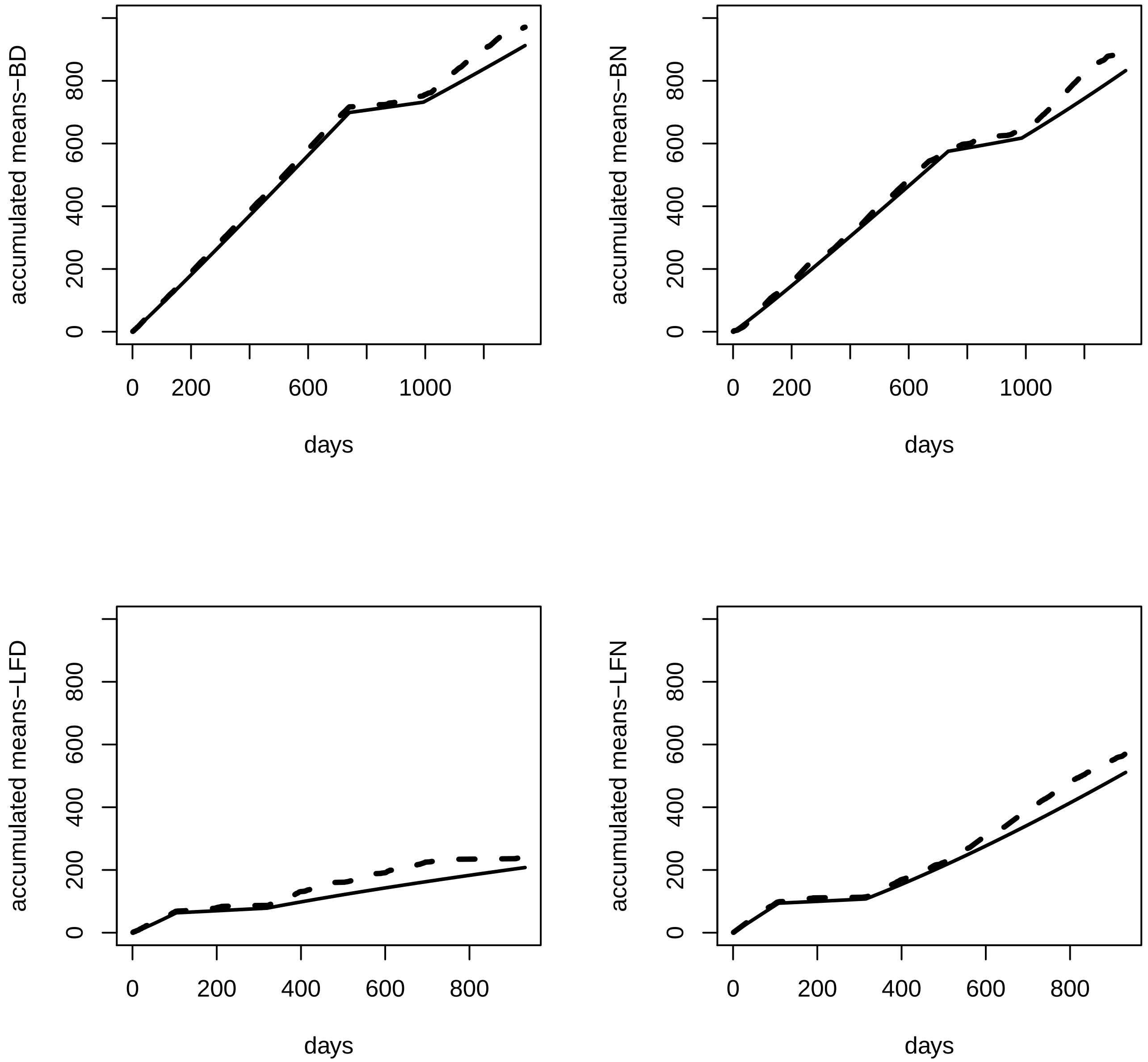


accumulated means−BD
0
200
400
600
800
0
200
600
1000
days
accumulated means−BN
0
200
400
600
800
0
200
600
1000
days
accumulated means−LFD
0
200
400
600
800
0
200
400
600
800
days
accumulated means−LFN
0
200
400
600
800
0
200
400
600
800
days

Figure

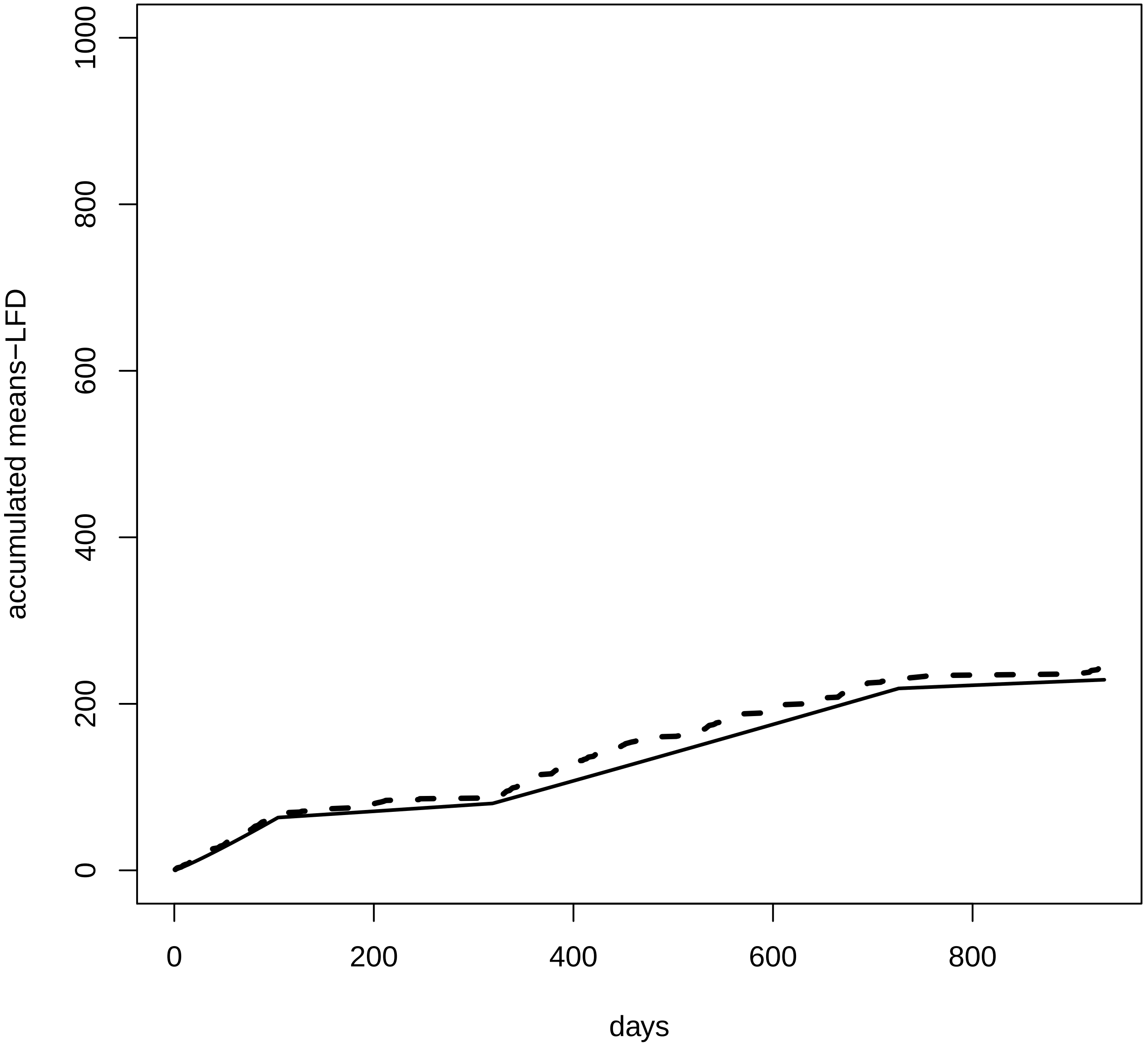

accumulated means−BD
accumulated means−BN
accumulated means−LFD
accumulated means−LFN
days
0
200
400
600
800
1000

Figure

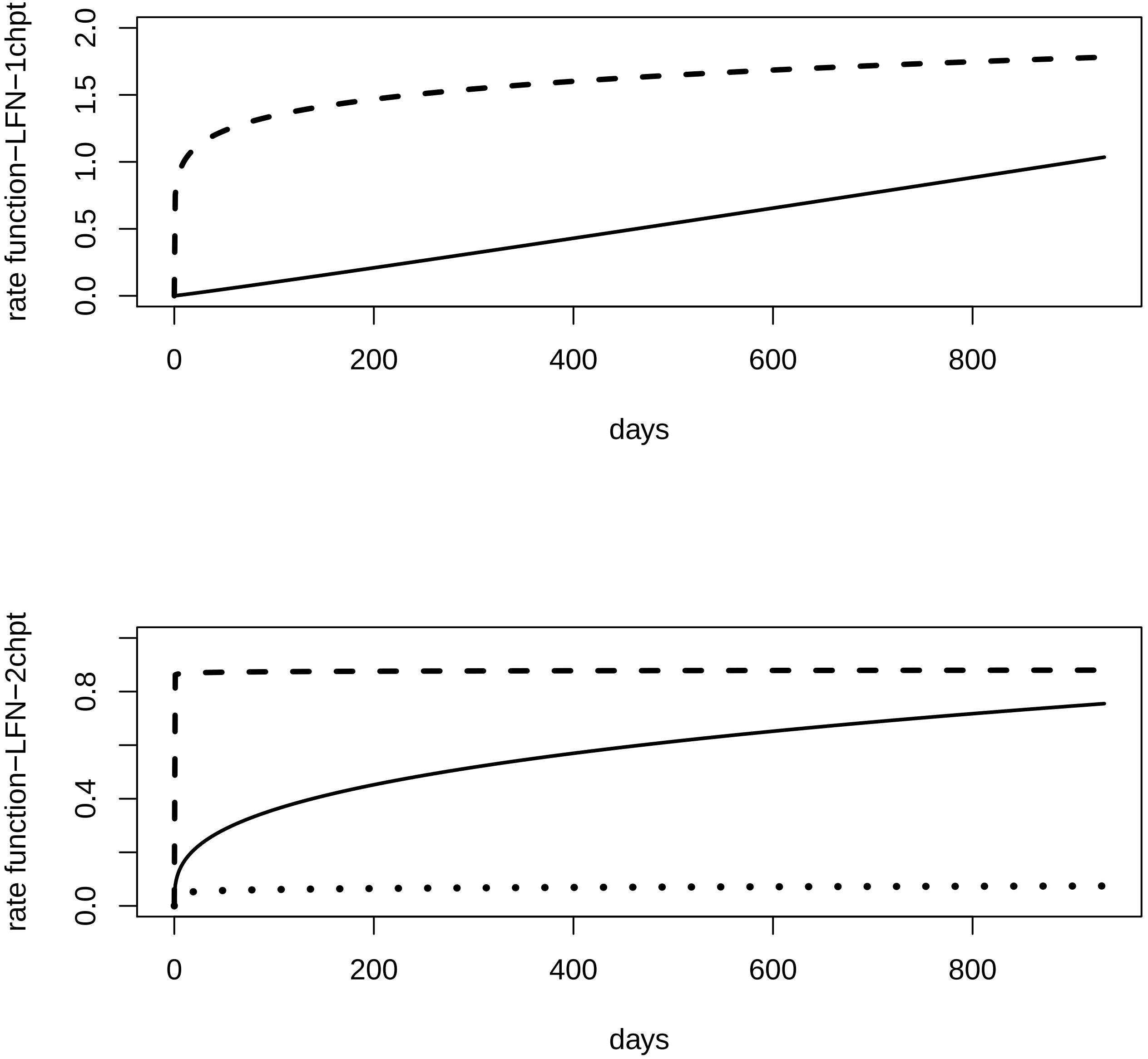


accumulated means−BD
accumulated means−BN
accumulated means−LFD
accumulated means−LFN
days
0
200
400
600
800
1000

Figure



Figure



rate function−LFN−1chpt
0.0
0.5
1.0
1.5
2.0
0
200
400
600
800
days
rate function−LFN−2chpt
0.0
0.4
0.8
0
200
400
600
800
days