Applied Acoustics 114 (2016) 203–212

Contents lists available at ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust

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

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

[a] Department of Industrial Engineering, University of Salerno, Italy
[b] Instituto de Matemáticas, Universidad Nacional Autónoma de México, Mexico

ARTICLE INFO

Article history:
Received 9 February 2016
Received in revised form 20 June 2016
Accepted 22 July 2016

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

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, the study of 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.

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] and the one used here, is the energy average equivalent level of the A-weighted sound pressure over a given period of time. Note that sound pressure levels for 24 h can be between 75 dBA and 80 dBA 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.

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. When the behaviour of the measurements is considered, several lines may be followed. One may, for instance, be interested 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, in particular, in the non-homogeneous Poisson model. Another line to be followed is to be able to estimate missing data. Regarding that, one may consider a time series approach.

Therefore, 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. In order to do that, a non-homogeneous Poisson model is considered. Even though the methodology was used in the past [7] to subsets of the datasets considered here, the measurements taken into account were those belonging to the time frames where no missing data were present. That gives only partial information since the overall behaviour of the data is ignored and only information provided by the measurements in the specific time frame is

* Corresponding author.
E-mail addresses: cguarnaccia@unisa.it (C. Guarnaccia), quartieri@unisa.it (J. Quartieri), ctepedino@unisa.it (C. Tepedino), eliane@math.unam.mx (E.R. Rodrigues).

http://dx.doi.org/10.1016/j.apacoust.2016.07.022

taken into account. To solve this problem a time series model will be used to estimate the missing values. Once the dataset is complete (i.e., with observed and estimated measurements), then a non-homogeneous Poisson model is used to estimate the number of exceedances of a given threshold.

Both methodologies considered here (time series and Poisson process) have been used in several areas of application. In the case of environmental problems, we have, for instance, that non-homogeneous Poisson models are applied to the areas of air pollution (see for instance [3–5]) 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 [5,9]. 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 produced by the missing values.

The daily observational data at a measuring site, are represented by a 16-h energy average sound level, indicated by $L_{Aeq,16h}$, for the day period (corresponding to 6 am to 10 pm), and an 8-h energy average sound level, denoted by $L_{Aeq,8h}$, for the night period (corresponding to 10 pm to 6 am). 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 non-homogeneous Poisson models 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. The traffic model could incorporate changes that could be made in order to reduce traffic noise, for instance, traffic is reduced in busy roads next to a residential area. After that, simulations could be performed and noise levels under the new situation would be produced. Taking into account that information we could apply the non-homogeneous Poisson model to the simulated noise levels. By doing that we may estimate the number of times that a population would be exposed to noise levels above a certain threshold in a time interval of interest if traffic is restricted. Note that the information produced by the existing traffic noise models would be related to the values of the measurements under the modification considered and the non-homogeneous Poisson model would provide information on the probability that the population, in the new scenario, would be exposed to levels above a given threshold a certain number of times in a time interval of interest. 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 a criterion 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. In Section 5, we present a discussion of the results obtained. Finally, in Section 6 we conclude.

2. Description of the mathematical models

A two-step approach will be taken 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 [12–15]. The present study deals with the case where the random variables register 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 explains 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 $\mathbf{X} = \{X_t : t \geqslant 0\}$ indicate the time series of interest. Denote by $\mathbf{T} = \{T_t : t \geqslant 0\}$ the trend component of the series, $\mathbf{S} = \{S_t : t \geqslant 0\}$ the seasonal component, and $\mathbf{E} = \{E_t : t \geqslant 0\}$ 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.,

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

A trend of type $T_t = \sum_{k=0}^{n} b_k t^k$ is taken. In some datasets taking $n = 1$ will suffice. 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, and is given by, $S_t = X_t/M_t$. 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,

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

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]; i.e.,

$$F_t = T_t \times S_t, \quad t \geqslant 0. \tag{3}$$