MULTIPLE CHANGE-POINT DETECTION USING A GENETIC ALGORITHM AND A BAYESIAN MINIMUM DESCRIPTION LENGTH FOR TIME EXCEEDANCES

Biviana Marcela Suárez Sierra  
Computing and analytics area. School of Applied Sciences and Engineering.  
EAFIT University  
Medellín, Colombia  
bmsuarezs@eafit.edu.co

Arrigo Coen  
Department of Mathematics, Faculty of Sciences.  
National Autonomous University of México  
CDMX, México, México  
coen@ciencias.unam.mx

Carlos Alberto Taimal  
Computing and analytics area. School of Applied Sciences and Engineering.  
EAFIT University  
Medellín, Colombia  
cataimaly@eafit.edu.co

ABSTRACT

The change-point detection problem has been widely studied in time series and signal processing literature. The current methods can be resumed in the search for the appropriate partitions of a whole time series such that the problem can be approached as one of optimization; nevertheless, an exact optimization approach could result computationally expensive and approximate ones discard potential scenarios for change-points configurations in a non-rigorous manner. Thus, a framework it is presented to detect change-points in a univariate time series using a decision criterion based on the Minimum Description Length (MDL), modified such that a Bayesian analysis is included. To search for the points of change, the times where mean value deviations occur (exceedances) are analyzed and then it is evaluated which of these could constitute a change-point through a genetic algorithm using as a fitness function the previously described MDL. The effectiveness of the method it is assessed through a simulation study and on the other hand, it is analyzed its practical validity in a real dataset for the presence of Particulate Matter of less than 2.5 microns (µm) \( PM_{2.5} \) in Bogotá, Colombia for the 2018-2020 period under different settings to understand the algorithm convergence. It is found that this definition for the objective function tends to find better results for both the number of change-points and their location in the series for most of cases reducing the error in comparison to other available methods in the literature.

Keywords  Multiple Change-point Detection · Genetic Algorithms · Minimum Description Length · Non-homogeneous Poisson Processes · Maximum A Posteriori Estimation

arXiv:2210.14807v1  [stat.AP]  26 Oct 2022
1 Introduction

The change-point analysis problem has been widely covered in the time series and signal processing literature [1, 2]. A change in a time series or signal is defined as a deviation from the normal behaviour in reference to a parameter or feature of the generating process of interest, and, consequently, a change-point is defined as the time where such deviation occurs. Some of these shifts include but are not limited to changes in the mean, variance or standard deviation, location and scale parameters and in general, any parameter of the assumed generating process to mention a few.

Now as there are multiple scenarios that determine a change and its respective time, also there are multiple methodologies available to finding them. Two exhaustive surveys are found in the works of [3, 2] to which the reader is referred to for a detailed review on the subject. From the second work it is noteworthy the typology used to describe change-point detection algorithms that are expressed as an optimization problem composed of an objective function, a search method, and a penalty on the analyzed parameters. Thus, the aim of a change-point detection algorithm is to segment a time series according to an optimality criterion which generally corresponds to a value that maximizes a penalized likelihood function.

On the other hand, to find the total of change-points and their location a method to search the solution space is used. Nevertheless, as the problem’s complexity increases when the time series granularity does it as well, the use of exact methods implies a high computation cost. For example, a method used by [4] based on dynamic programming had a time complexity of \(O(T^2)\) with \(T\) the length of the series. Likewise, considering \(J\) as the total of change-points, there are \(\binom{T}{J}\) possible settings to evaluate and, if it was decided to compute the objective function value for all of them, i.e., for \(J = 0, 1, 2, \ldots, T\), there would be \(2^T\) scenarios to analyze [5].

Thus, and based on the work of [6, 5, 7, 8] is presented a method which uses a genetic algorithm as a search method modifying the objective function such that a Bayesian analysis is included and the Maximum A Posteriori estimation (MAP) function is used instead of the Maximum Likelihood one. On the other hand, we did not considered the underlying distribution for the observed data but the times where deviations from the mean or exceedances did occur, methodology adapted from Survival Analysis overcoming the need of a prior parametric definition for the generating process.

The rest of this paper is organized as follows: section (2) presents the inferential analysis corresponding to the existence of exceedances from the mean of the process, modeled using the rate and intensity functions of a non-homogeneous Poisson process for the times; in the subsections of this, (2.1) and (2.2) is presented the proposal for detecting such exceedances, using a genetic algorithm to search for these points and using as an objective function a modified version of the Minimum Description Length (MDL) principle. Next, in section (3), the derivation for the objective function are exposed for each one of the considered intensity functions such that in section (4) the performance of the algorithm is assessed through simulated and real data. Finally in section (5) the conclusions in regards of the found results are presented.

2 Inferential Analysis of the Univariate Non-Homogeneous Poisson Process

Based on the construction of the likelihood from a non-homogeneous Poisson process discussed in [9] and [10], we have the following expression, from which we will start to construct the penalized MAP function. The developments that will be shown in the following sections will be based on the assumption that some of the functions of the expression correspond to the intensity function of the Non-homogeneous Poisson Process (NHPP).

\[
\begin{align*}
\lambda^{(W)}(t|\theta) &= (\alpha/\beta)(t/\beta)^{\alpha-1}, \quad \alpha, \beta > 0 \\
\lambda^{(MO)}(t|\theta) &= \frac{\beta}{t+\alpha}, \quad \alpha, \beta > 0 \\
\lambda^{(GO)}(t|\theta) &= \alpha\beta \exp(-\beta t), \quad \alpha, \beta > 0 \\
\lambda^{(GGO)}(t|\theta) &= \alpha\beta\gamma t^{-\gamma-1} \exp(-\beta t\gamma), \quad \alpha, \beta, \gamma > 0.
\end{align*}
\]

(1)

Each letter in the superscript of the left-hand term corresponds to the initial letter of the name of the distribution used as intensity function as follows, Weibull \((W)\) [11][12], Musa-Okumoto \((MO)\) [13], Goel-Okumoto \((GO)\) and a generalization of the Goel-Okumoto model \((GGO)\) [14], for which the cumulative average function, \(m(t|\theta)\) is defined respectively by:
A change point is defined as an instant where the structural pattern of a time series shifts. We assumed the presence of $J$ change-points, $\{\tau_1, \tau_2, \cdots, \tau_J\}$ such that there are variations on the model parameters in between segments $\tau_{j-1} < t < \tau_j$, $j = 0, 1, 2, \ldots, J + 1$, $j_0 = 1, j_{J+1} = T$. These changes can be attributed to environmental policies or legislations in a certain year, the suspension of some network station due to maintenance for a climate case, macroeconomic policies from an economic standpoint or the presence of an stimulus in a neuroscience context. Then, the intensity functions of the NHPP have the form,

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

(3)

where $\theta_j$ is the vector of parameters between the change points $\tau_{j-1}$ and $\tau_j$, for $j = 2, ..., J$, and $\theta_1$ and $\theta_{J+1}$ are the parameter vectors before and after the first and last change points, respectively. With $n$ observations, the functions for the means are (see, e.g., [15]),

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

(4)

where $\tau_{J+1} = T$. That is, because $m(t|\theta_1)$ represents the average number of exceedances of the standard, before the first change point, $m(\tau_1|\theta_1) + m(t|\theta_2) - m(\tau_1|\theta_1)$ is the average number of exceedances of the standard between the first change point $\tau_1$ and the second one $\tau_2$, given that the vector of parameters $\theta_2$ is known, and so on.

Be $D = d_1, d_2, \ldots, d_n$, where $d_k$ (as in the case without change points), is the time of occurrence of the $k$th event (the $k$th time the maximum level of the environmental standard is exceeded), with $k = 1, 2, \ldots, n$, the likelihood function is determined by the expression below where $N_{t_i}$ represents the number of exceedances before the change point $\tau_{i}$, with $i = 1, 2, \ldots, J$ (see [16])

$$L(D|\phi) \propto \prod_{i=1}^{N_{\tau_1}} \lambda(d_i|\theta_1) e^{-m(\tau_1|\theta_1)} \times \prod_{j=2}^{J} \left( \prod_{i=N_{\tau_{j-1}}+1}^{N_{\tau_j}} \lambda(d_i|\theta_j) e^{-[m(\tau_j|\theta_j) - m(\tau_{j-1}|\theta_j)]} \right) \times \prod_{i=N_{\tau_{J+1}}+1}^{n} \lambda(d_i|\theta_{J+1}) e^{-[m(\tau_{J+1}|\theta_{J+1}) - m(\tau_j|\theta_{J+1})]},$$

(5)

Using the expression (5), we infer the parameters $\phi = (\theta, \tau)$, with $\theta = (\theta_1, ..., \theta_J)$ and $\tau = (\tau_1, ..., \tau_J)$ using a Bayesian approach. This perspective consists of finding the relationship between the a priori distribution of the parameter $\theta$, on
whose intensity function $\lambda(t|\theta)$ is dependent and the a posteriori distribution of the same, after taking into consideration the observed information $D$. In [18], this method was applied to obtained results very close to the observed ones, hence the descriptive capacity of the model and the methodology used. In such work, the criteria used to select the model that best fits the data together with the graphic part was the MDL.

2.2 Detection of multiple change points using genetic algorithm

2.2.1 MDL framework

Since finding $J$ change-points implies finding out $J + 1$ regimes for the time series or fitting $J + 1$ models with different parameters, statistical criteria has been used for such purpose in the available literature. Some include the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC), Cross-Validation methods, and MDL-based methods. For problems involving regime shift detection, MDL methods usually provide superior empirical results. This superiority is probably due to the fact that both AIC and BIC apply the same penalty to all parameters, regardless of the nature of the parameter. On the other hand, MDL methods can adapt penalties to parameters depending on their nature be it continuous or discrete, bounded or not. In short, MDL defines the best fitting model as the one that enables the best compression of the data minimizing a penalized likelihood function. That is,

$$MDL = - \log_2(L_{opt}) + P.$$  

(6)

Here $\log_2(L_{opt})$ is the required amount of information needed to store the fitted model, term taken from information theory. More details on this can be found in [19]. $L_{opt}$ is obtained from replacing the maximum likelihood estimator in the likelihood function [5]. This will be explained in more detail in the next section.

Because of the above, it is possible to make the natural connection between the likelihood and the MDL objective function by means of the penalty $P$ (see [19]). The broad penalty methodology is summarized in three principles as stated by [5]. The first one is to penalize the real valued parameters by the number of observations. Say $k$ that are used to estimate it, then, the penalty will be $\frac{\log_2 k}{2}$. For this principle, it is important to take into consideration how the observations are arranged to calculate the parameter of interest because this arrangement will be reflected in the penalty.

The second principle involves the penalty of how many integer parameters, such as the number of change points $J$ and where they are located represented by $\tau_1, ..., \tau_J$ should be charged. This charging is calculated based on the value for each of them. For example, the quantity $J$, which is bounded by the total number of observations $T$ is charged an amount of $\frac{\log_2 T}{2}$. For each of the $\tau_j$ with $j = 1, ..., J$, we have that $\tau_j < \tau_{j+1}$, therefore the cost of its penalty will be $\frac{\log_2 \tau_i}{2}$ for $j = 2, ..., J$.

The last principle, mentioned in [5], is the additivity principle. It involves constructing $P$ based on the sum of all the partial penalties mentioned above. The more parameters the model has, the higher $P$ will be. However, if despite adding parameters, the expression $\log_2(L_{opt})$ does not grow larger than the penalty $P$ of the extra parameters, the simpler model will be preferred. For the purposes of this paper, the following will be used as the penalty function $P_\tau(\theta)$ for a fixed change point configuration,

$$P_\tau(\theta) = R \sum_{j=1}^{J+1} \frac{\ln(\tau_i - \tau_{i-1})}{2} + \ln(J) + \sum_{j=2}^{J} \ln(\tau_j),$$  

(7)

where $R = 2, 3$ depending on whether $\theta = (\alpha, \beta)$ or $\theta = (\alpha, \beta, \gamma)$, i.e. if $\theta$ has one or two parameters. The first summand of the right-hand term of expression represents that each of the real-valued parameters $(\alpha_j, \beta_j, \gamma_j)$ will be penalized by $\frac{\ln(\tau_i - \tau_{i-1})}{2}$ of the $j$-th regime to which they belong and since there are $J + 1$ regimes, the sum goes from 1 to $J + 1$. The second summand of the right-hand term is derived from the penalty of the number of points of change, and the last term comes from the sum of each of the penalties of each of the change points.

2.2.2 Genetic Algorithm Schema

As exposed in [5] the total possible cases to evaluate the MDL corresponds to $\binom{J}{J}$, where $T$ is the number of observations in the time series and $J$ is the number of change points. However, this number of parametric configurations
is a quantity that does not make a computationally efficient optimization algorithm that aims to choose the best of the parametric configurations that minimize the MDL. For this reason, we will use the genetic algorithm that, by natural selection criteria will establish the best of the parameters configurations that we will call chromosomes. Each chromosome will be labeled as \((J, \tau_1, \ldots , \tau_J)\), where the first component \(J\) stands for the number of change points, located respectively at times \(\tau_1, \ldots , \tau_J\), corresponding to the respective coordinates. The following is to establish how the genetic algorithm (GA) evaluates each of the chromosomes, while avoiding those with a low probability of being optimal.

Let us now see how a complete generation is produced from an initial one with a given size, although the size can also be a couple. For this purpose, suppose there are \(k\) individuals or chromosomes in the initial generation set at random. Each of the \(T\) observations in the time series is allowed to be a change-point, independent of all other change-points, with probability, for example as seen in [5], of 0.06. The number of change points for each chromosome in the initial generation has a binomial distribution with parameters \(T - 1\) and 0.06, respectively.

Two chromosomes are taken out of the initial generation, one mother and one father chromosome, to make a child of the next generation. This is done by a probabilistic combination of the parents. The principle of natural selection in this context will be performed by selecting a pair of chromosomes that best optimize the expression (6) since this couple is considered to have the highest probability of having offspring. Therefore, the chromosomes are arranged from the most likely to the least likely to have children, and each individual of the same generation is assigned a ranking, say \(S_j\), being the ranking of the \(j\)th individual, with \(S_j = 1, \ldots , k\). If \(S_j = k\) then \(j\) is the individual that best fits the objective function (6). If \(S_j = 1\) then \(j\) is the individual that least well fits the objective function (6).

Once this ranking has been made for each of the chromosomes of the same generation, we proceed to establish the probability of selection using the following expression that simulates the principle of natural selection of the parents that will generate the next generation.

\[
\frac{S_j}{\sum_{i=1}^{k} S_i}
\]  

(8)

The chromosome that has the highest probability of being selected from the \(k\) chromosomes, is chosen as the mother. Among the remaining \((k - 1)\) chromosomes, the father is chosen under the same selection criteria as the mother. Suppose that the mother chromosome has \(m\) change points, located in some \(\tau_1, \ldots , \tau_m\), i.e. with the parameter configuration \((m, \tau_1, \ldots , \tau_m)\). Similarly, suppose the father chromosome has the parametric configuration \((n, \delta_1, \ldots , \delta_n)\). A child of these parents can arise simply by joining the two chromosomes, i.e., the child chromosome will initially have the following configuration, \((m + n, \epsilon_1, \ldots , \epsilon_m + n)\), where the \(m + n\) change-points contain the mother’s \(m\) and the father’s \(n\) change-points.

After this, we remove the duplicated change-points from the child \((m + n, \epsilon_1, \ldots , \epsilon_m + n)\). From this last configuration, we keep all or some change-points. For this, in [5] use the dynamics of flipping a coin for each change-point in the child configuration. If heads comes up, the change-point is left, otherwise it is removed. That is, a binomial distribution will be used with probability parameter 1/2 and number of trials, the length of the configuration of the child minus duplicities. All this with the aim that the offspring will keep traits of the parents, without being exact replicas.

Each point of change in the child chromosome, can undergo a mutation; a schema taken from [5] is that one of the following three alternatives may happen. We start by generating, with some random mechanism, the numbers \(-1, 0\) and, 1 with respective probabilities 0.4, 0.3, 0.4. If \(-1\) comes out, the change point is subtracted by one unit; if 0 comes out, it stays at the time it is at, and if 1 comes out, the current change point is added by one unit. Again, duplicates are eliminated. With this last procedure, we have finished the construction of Child 1. Child 2 up to \(k\) are generated in the same way as the previous one. New parents are selected if chromosomes are duplicated in the same generation with the previous parents.

The process of generation is repeated as many times as generations are to be obtained. In fact, one of the criteria for establishing the completion of the genetic algorithm is to fix the number of generations \(r\). Another approach could be to reach the solution that minimizes the objective function. The objective function for we used was \(\ln P_r(\theta) - \ln f(D|\theta) - \ln f(\theta)\), i.e.
We are particularly interested in the likelihood function of expression (9) in order to establish what we have called the
L function, where
f corresponds, and
f. From (14), we start by establishing the general form of the a priori joint function
\( L\) of the parameter vector \( \theta \). Taking logarithm for (13) we have that,
\[ \ln f(\theta | D) \propto \ln L(D|\theta)) + \ln(f(\theta)) \] (14)
From (14), we start by establishing the general form of the a priori joint function \( f(\theta) = f(\alpha, \beta, \tau) \) in the first three cases, and \( f(\theta) = f(\alpha, \beta, \gamma, \tau) \) in the last function \( m \) of the term (2).
3.1 A priori distributions

If we take \( \alpha \sim \text{Gamma}(\phi_{11}, \phi_{12}) \), then,

\[
f(\alpha) = \frac{\phi_{11}^{\phi_{12}-1}e^{-\phi_{11}\alpha}}{\Gamma(\phi_{12})}
\]

After applying logarithm we obtain,

\[
\log f(\alpha) = \phi_{12} \log \phi_{11} - \log \Gamma(\phi_{12}) + (\phi_{12} - 1) \log \alpha - \phi_{11} \alpha
\]

(15)

Similarly for \( \beta \), if we take \( \beta \sim \text{Gamma}(\phi_{21}, \phi_{22}) \), then,

\[
\log f(\beta) \propto (\phi_{22} - 1) \log \beta - \phi_{21} \beta
\]

(16)

On the other hand, assuming every time in the series can be chosen as a change-point \( \tau_j \) with the same probability, thus \( \tau_j \sim \text{Uniform}(0, T) \), \( j = 1, 2, \ldots, J \).

Then we have,

\[
f(\tau_j) = \frac{1}{T}
\]

(17)

Taking logarithm we obtain,

\[
\log f(\tau_j) = -\log(T)
\]

(18)

Rebuilding the joint function for \( \theta = (\alpha, \beta, \tau_j) \), under the assumption of independence, we have,

\[
\log f(\alpha, \beta, \tau_j) \propto (\phi_{12} - 1) \log \alpha - \phi_{11} \alpha + (\phi_{22} - 1) \log \beta - \phi_{21} \beta - \log(T)
\]

(19)

In the three-parameter model for the intensity function, we have that, under the assumption of independence and that all the parameters have an a priori gamma distribution; considering also, the distribution for the change-points, thus we have,

\[
\log f(\alpha, \beta, \gamma, \tau_j) = -\alpha \phi_{11} + (\phi_{12} - 1) \log \alpha - \beta \phi_{21} + (\phi_{22} - 1) \log \beta - \gamma \phi_{31} + (\phi_{32} - 1) \log \gamma - \log(T).
\]

(20)

Up to this point, the second summand of the right-hand side of (14) has been obtained. Next, the first summand of the right-hand side of (14) will be derived, but this will be done depending on the intensity function of the non-homogeneous Poisson process that was established previously and whose cumulative mean functions are expressed in the four possibilities of (2).

3.1.1 Weibull intensity rate (W)

After taking the expressions for the intensity function \( \lambda^{(W)}(t|\theta) \) and the cumulative mean function \( m^{(W)}(t|\theta) \) using (2) and (1) respectively, and replacing these in (12) we have,
Change-point Detection Using a GA and a BMDL for Time Exceedances

\[
\log L(D|\phi) = \sum_{j=1}^{J+1} \left( m(\tau_{j-1}|\theta_j) - m(\tau_j|\theta_j) + \sum_{i=N_{\tau_{j-1}}+1}^{N_{\tau_j}} \log \lambda(d_i|\theta_j) \right)
\]

\[
= \sum_{j=1}^{J+1} \left( \frac{\tau_{j-1}}{\beta_j} \alpha_j - \left( \frac{\tau_j}{\beta_j} \right) \alpha_j + \sum_{i=N_{\tau_{j-1}}+1}^{N_{\tau_j}} \log \left( \frac{\alpha_j d_i}{\beta_j} \right)^{\alpha_j-1} \right)
\]

\[
= \sum_{j=1}^{J+1} \left( \frac{\tau_{j-1}}{\beta_j} \alpha_j - \left( \frac{\tau_j}{\beta_j} \right) \alpha_j + (N_{\tau_j} - N_{\tau_{j-1}}) \log(\alpha_j) - \alpha_j \log(\beta_j) \right)
\]

\[+ (\alpha_j - 1) \sum_{i=N_{\tau_{j-1}}+1}^{N_{\tau_j}} \log(d_i) \]  \hspace{1cm} (21)

Substituting the expressions (7), (19), (21) in the objective function of the expression (9) we have that,

\[
\ln P_{\tau}(\theta) - \ln f_{\tau}(D|\theta) - \ln f_{\tau}(\theta) = 2 \sum_{i=1}^{J+1} \ln(\tau_i - \tau_{i-1}) \, 2 + \ln(J) + \sum_{i=2}^{J} \ln(\tau_i) - \sum_{j=1}^{J+1} \left( \frac{\tau_{j-1}}{\beta_j} \alpha_j - \left( \frac{\tau_j}{\beta_j} \right) \alpha_j + (N_{\tau_j} - N_{\tau_{j-1}}) \log(\alpha_j) - \alpha_j \log(\beta_j) \right)
\]

\[+ (\alpha_j - 1) \sum_{i=N_{\tau_{j-1}}+1}^{N_{\tau_j}} \ln(d_i) \]

\[- \sum_{j=1}^{J+1} \left( (\phi_{12} - 1) \alpha_j - \phi_{11} \alpha_j + (\phi_{22} - 1) \beta_j - \phi_{21} \beta_j \right) J \ln(T) \]  \hspace{1cm} (22)

### 3.1.2 Musa-Okumoto (MO)

Likewise, taking the expressions for the intensity function \(\lambda^{(MO)}(t|\theta)\) and the cumulative mean function \(m^{(MO)}(t|\theta)\) from (2) and (1) respectively, and replacing these values in (12) we have,

\[
\log L(D|\phi) = \sum_{j=1}^{J+1} \left( m(\tau_{j-1}|\theta_j) - m(\tau_j|\theta_j) + \sum_{i=N_{\tau_{j-1}}+1}^{N_{\tau_j}} \log \lambda(d_i|\theta_j) \right)
\]

\[
= \sum_{j=1}^{J+1} \left( \beta_j \log \left( \frac{\alpha_j + \tau_{j-1}}{\alpha_j} \right) - \beta_j \log \left( \frac{\alpha_j + \tau_j}{\alpha_j} \right) + \sum_{i=N_{\tau_{j-1}}+1}^{N_{\tau_j}} \log \left( \frac{\beta_j}{\alpha_j + d_i} \right) \right)
\]

\[
= \sum_{j=1}^{J+1} \left( \beta_j \log \left( \frac{\alpha_j + \tau_{j-1}}{\alpha_j} \right) - \beta_j \log \left( \frac{\alpha_j + \tau_j}{\alpha_j} \right) + (N_{\tau_j} - N_{\tau_{j-1}}) \log(\beta_j) - \sum_{i=N_{\tau_{j-1}}+1}^{N_{\tau_j}} \log(\alpha_j + d_i) \right)
\]

\[= \sum_{j=1}^{J+1} \left( \beta_j \log \left( \frac{\alpha_j + \tau_{j-1}}{\alpha_j} \right) - \log(\alpha_j + \tau_j) + (N_{\tau_j} - N_{\tau_{j-1}}) \log(\beta_j) - \sum_{i=N_{\tau_{j-1}}+1}^{N_{\tau_j}} \log(\alpha_j + d_i) \right) \]  \hspace{1cm} (23)

Now, after replacing (7), (23), and (19) in (9) we have,
As for the previous cases, we take the expressions \(\lambda^{(GO)}(t|\theta)\) and \(m^{(GO)}(t|\theta)\) from (2) and (1) respectively, and replace this values in (12). Then we have,

\[
\log L(D|\phi) = \sum_{j=1}^{J+1} \left( m(\tau_{j-1} | \theta_j) - m(\tau_j | \theta_j) + \sum_{i=N_{\tau_j-1}+1}^{N_{\tau_j}} \log \lambda(d_i | \theta_j) \right)
= \sum_{j=1}^{J+1} \left( \alpha_j \left[ 1 - e^{-\beta_j \tau_j} \right] - \alpha_j \left[ 1 - e^{-\beta_j \tau_j} \right] + \sum_{i=N_{\tau_j-1}+1}^{N_{\tau_j}} \log (\alpha \beta e^{-\beta d_i}) \right)
= \sum_{j=1}^{J+1} \left( \alpha_j \left[ e^{-\beta_j \tau_j} - e^{-\beta_j \tau_j} \right] + (N_{\tau_j} - N_{\tau_j-1}) \log(\alpha \beta) - \beta \sum_{i=N_{\tau_j-1}+1}^{N_{\tau_j}} d_i \right)
\]

Replacing the expressions (7), (25), (19) in the objective function of the expression (9) we have that

\[
P_\tau(\theta) - \ln f_\tau(D|\theta) - \ln f_\tau(\theta) = 2 \sum_{i=1}^{J+1} \frac{\ln(\tau_i - \tau_{i-1})}{2} + \ln(J) + \sum_{i=2}^{J} \ln(\tau_i)
- \sum_{j=1}^{J+1} \left( \alpha_j \left[ e^{-\beta_j \tau_j} - e^{-\beta_j \tau_j} \right] + (N_{\tau_j} - N_{\tau_j-1}) \log(\alpha \beta) - \beta \sum_{i=N_{\tau_j-1}+1}^{N_{\tau_j}} d_i \right)
- \sum_{j=1}^{J+1} ((\phi_{12} - 1) \ln \alpha_j - \phi_{11} \alpha_j + (\phi_{22} - 1) \ln \beta_j - \phi_{21} \beta_j) + J \ln(T)
\]

3.1.3 Goel-Okumoto (GO)

As for the previous cases, we take the expressions \(\lambda^{(GO)}(t|\theta)\) and \(m^{(GO)}(t|\theta)\) from (2) and (1) respectively, and replace this values in (12). Then we have,

\[
\log L(D|\phi) = \sum_{j=1}^{J+1} \left( m(\tau_{j-1} | \theta_j) - m(\tau_j | \theta_j) + \sum_{i=N_{\tau_j-1}+1}^{N_{\tau_j}} \log \lambda(d_i | \theta_j) \right)
= \sum_{j=1}^{J+1} \left( \alpha_j \left[ 1 - e^{-\beta_j \tau_j} \right] - \alpha_j \left[ 1 - e^{-\beta_j \tau_j} \right] + \sum_{i=N_{\tau_j-1}+1}^{N_{\tau_j}} \log (\alpha \beta e^{-\beta d_i}) \right)
= \sum_{j=1}^{J+1} \left( \alpha_j \left[ e^{-\beta_j \tau_j} - e^{-\beta_j \tau_j} \right] + (N_{\tau_j} - N_{\tau_j-1}) \log(\alpha \beta) - \beta \sum_{i=N_{\tau_j-1}+1}^{N_{\tau_j}} d_i \right)
\]

3.1.4 Generalized Goel-Okumoto (GGO)

Finally and once again, we take \(\lambda^{(GGO)}(t|\theta)\) and \(m^{(GGO)}(t|\theta)\) from (2) and (1) respectively, and replace these values in (12) and then,
We considered the existence of one, two, three and four change-points and for each of these cases, such number of change-points were located in different ways which are resumed in the table (1).

For the synthetic data generation for data with an approximate log-normal distribution was used, such that 1096 observations were simulated, being this, the total of observations for the real data we considered in the following section.

Therefore, the expressions (22), (24), (26) and (28) are the objective functions that by minimizing the Bayesian MDL is obtained for each of the different functions \( \lambda(t|\theta) \) of the expression (1), respectively.

Each of the members of the same generation will have a Bayesian-MDL, of which the smallest is chosen. This is done for all the generations. At the end, we will have as many Bayesian-MDLs as generations, and the minimum corresponding to the solution sought in the problem of determining the points of change of the time series of intervals is chosen.

### 4 Results and discussion

#### 4.1 Simulation study

To analyze the performance of the algorithm for the selection of different number of change-points under different settings, we conducted a set of simulations. Thus, a similar scheme as the one exposed by [5] for the synthetic data generation for data with an approximate log-normal distribution was used, such that 1096 observations were simulated, being this, the total of observations for the real data we considered in the following section.

We considered the existence of one, two, three and four change-points and for each of these cases, such number of change-points were located in different ways which are resumed in the table (1).
Table 1: Different settings for change-points simulations

| Number of Change-points | Setting | Locations       |
|-------------------------|---------|-----------------|
| 1                       | 1       | $\tau_1 = 548$ |
| 1                       | 2       | $\tau_1 = 275$ |
| 1                       | 3       | $\tau_1 = 825$ |
| 2                       | 1       | $\tau_1 = 275, \tau_2 = 821$ |
| 2                       | 2       | $\tau_1 = 547, \tau_2 = 823$ |
| 2                       | 3       | $\tau_1 = 365, \tau_2 = 730$ |
| 3                       | 1       | $\tau_1 = 275, \tau_2 = 549, \tau_3 = 823$ |
| 3                       | 2       | $\tau_1 = 548, \tau_2 = 823, \tau_3 = 973$ |
| 3                       | 3       | $\tau_1 = 169, \tau_2 = 413, \tau_3 = 1027$ |
| 4                       | 1       | $\tau_1 = 220, \tau_2 = 439, \tau_3 = 658, \tau_4 = 877$ |
| 4                       | 2       | $\tau_1 = 23, \tau_2 = 319, \tau_3 = 345, \tau_4 = 434$ |
| 4                       | 3       | $\tau_1 = 182, \tau_2 = 507, \tau_3 = 533, \tau_4 = 904$ |

For each setting, 1000 simulations were conducted such that for every possible value of $J$ we obtained 3000 results. Then, these results are analyzed for every setting.

4.1.1 One Change-point

We started by locating a change-point in the simulated time series; thus, two sub-series with an approximate log-normal distribution with parameters $\sigma$ and $\mu$ were generated. The scale parameter, $\sigma$ remained constant with a value of $\sigma \approx 0.32$ for the two regimes, and in an increasing manner, the location parameter was varied, such that for the first regime, this took a value of 3.50 and for the second one, 4.00. This behaviour was maintained for the three settings for one change-point under analysis.

For the three settings, $\tau_1 = 548, \tau_1 = 275$ and $\tau_1 = 82$ the results are reported in the figure (1) and the tables (2), (3) and (4). First, it is noteworthy in the figure that in general, the algorithm overestimates the real value of $J$ such that the overall trend estimates between 2 and 3 change-points for the simulated series. Nevertheless, the number of values detected as change-points was not so far from the real one as can be seen in the three panels in figure (1) where $J = 1$ is represented by a black dashed vertical line.

Now, in the tables can be verified the exposed result for the real estimation of the number of change-points under the three settings for the 1000 simulations considered where the proportions of times that the algorithm detected $J = 1$ was at least 12.20% of the times and at most, 17.80% of the times, favoring principally values of $J = 2$ or greater and even 3. While on a first approach the algorithm seems not to present a proper behavior, we will see that as the number of change-points increase, the algorithm stabilizes and the trend for the number of change-points detected gets closer to the real one.
4.1.2 Two Change-points

In this case, \( J + 1 = 3 \) regimes were simulated for the time series; in a similar manner for the case of one change-points, a constant value for the scale parameter, \( \sigma \approx 0.32 \) was considered, while, the location parameter \( \mu \) was varied in an increasing manner in 0.5 units taking values of 3.50, 4.00 and 4.50 because we are interested in detecting exceedances from the mean; on the other hand, the locations of the change-points is the one reported in table (1) in rows four, five and six, thus, the cases considered are \((\tau_1, \tau_2) = (275, 821), (\tau_1, \tau_2) = (547, 823)\) and \((\tau_1, \tau_2) = (365, 730)\). The results for each one of these simulations are resumed in figure (2) and tables (5), (6) and (7).
Change-point Detection Using a GA and a BMDL for Time Exceedances

![Graph showing estimation of change-points for different settings](image)

**Figure 2:** Estimation of the number of Change-points for $J = 2$

| $J$ | Percent |
|-----|---------|
| 1   | 0.00    |
| 2   | 15.10   |
| 3   | 21.60   |
| 4   | 20.30   |
| 5+  | 43.00   |

Table 5: Proportions of estimated change-point numbers. The correct value of $J$ is two and the setting is the second one.

| $J$ | Percent |
|-----|---------|
| 1   | 0.30    |
| 2   | 16.30   |
| 3   | 19.70   |
| 4   | 21.80   |
| 5+  | 41.90   |

Table 6: Proportions of estimated change-point numbers. The correct value of $J$ is two and the setting is the second one.

| $J$ | Percent |
|-----|---------|
| 1   | 0.10    |
| 2   | 11.90   |
| 3   | 19.80   |
| 4   | 23.60   |
| 5+  | 44.60   |

Table 7: Proportions of estimated change-point numbers. The correct value of $J$ is two and the setting is the third one.

While, still it can be observed and overestimation for the number of change-points by the algorithm, it can be appreciated as well, that this starts to stabilize around the real value we are interested in estimate for the three settings of $J$ under analysis. The proportion of number of change-points estimated starts to get consider the estimation closer to the real values as in this case such that at least in 12.00% of the times two change-points were detected as the optimal number...
and 16.30% such estimation occurred. Two change-points was considered as the third most common case under the first setting and as the fourth most common one under the second and third settings which supposes an improvement in the behavior of the algorithm in contrast to the detection of one change-point.

4.1.3 Three Change-points

Now, three possible change-points were considered, such that four sub-series or regimes with a log-normal distribution were simulated; again, the value of the scale parameter was held constant as the previous cases, that is, $\sigma \approx 0.32$ and similarly, for $\mu$, the different values were varied in an increasing manner in 0.5 units, starting from 3.50; thus, the other values considered for this parameter were 4.00, 4.50 and 5.00.

On the other hand, the settings in the seventh, eighth and ninth rows in table (1) were considered for the location of the change-points, that is, $(\tau_1, \tau_2, \tau_3) = (275, 549, 823), (\tau_1, \tau_2, \tau_3) = (548, 823, 973)$ and $(\tau_1, \tau_2, \tau_3) = (169, 413, 1027)$ and the results were reported in figure (3) and tables (8), (9) and (10).

![Figure 3: Estimation of the number of Change-points for $J = 3$](image)

| $J$ | Percent |
|-----|---------|
| 1   | 0.30    |
| 2   | 16.50   |
| 3   | 20.60   |
| 4   | 19.40   |
| 5+  | 43.20   |

Table 8: Proportions of estimated change-point numbers. The correct value of $J$ is three and the setting is the first one.

| $J$ | Percent |
|-----|---------|
| 1   | 0.10    |
| 2   | 12.60   |
| 3   | 20.50   |
| 4   | 21.90   |
| 5+  | 44.90   |

Table 9: Proportions of estimated change-point numbers. The correct value of $J$ is three and the setting is the second one.
Change-point Detection Using a GA and a BMDL for Time Exceedances

| $J$ | Percent |
|-----|---------|
| 1   | 0.30    |
| 2   | 17.10   |
| 3   | 21.40   |
| 4   | 20.50   |
| 5+  | 40.70   |

Table 10: Proportions of estimated change-point numbers. The correct value of $J$ is three and the setting is the third one.

As previously mentioned, the algorithm’s precision for the detection of the number of change-points improves significantly as this value increases. Thus, for the three considered settings, in the first and third one, the number of change-points most commonly detected corresponded to three, being this value, the real one; then for the 1000 simulations under each setting, in the first one, 20.60% of the times the algorithm detected three change-points, in the second setting, 20.50% of the times and in the third setting, 21.40% of the times.

4.1.4 Four Change-points

Finally, we have the case of four change-points. We proceeded in the same manner as the three previous cases defining a constant value for the parameter $\sigma$ of the log-normal distribution, $\sigma \approx 0.32$ and considered five regimes such that the values of $\mu$ were again varied in an incremental manner in 0.5 units; thus, this parameter took the values 3.00, 3.50, 4.00, 4.50 and 5.00. The locations of the change-points are the ones in the last three rows of the table (1) and the results are reported in the figure (4) and the tables (11), (12) and (13).

| $J$ | Percent |
|-----|---------|
| 1   | 0.10    |
| 2   | 13.90   |
| 3   | 21.70   |
| 4   | 20.50   |
| 5   | 14.70   |
| 6+  | 29.10   |

Table 11: Proportions of estimated change-point numbers. The correct value of $J$ is four and the setting is the first one.
Change-point Detection Using a GA and a BMDL for Time Exceedances

|   |   |
|---|---|
| 1 | 0.10 |
| 2 | 16.60 |
| 3 | 22.90 |
| 4 | 19.90 |
| 5 | 14.40 |
| 6+ | 26.10 |

Table 12: Proportions of estimated change-point numbers. The correct value of $J$ is four and the setting is the second one.

|   |   |
|---|---|
| 1 | 0.20 |
| 2 | 18.20 |
| 3 | 20.20 |
| 4 | 18.90 |
| 5 | 15.30 |
| 6+ | 27.20 |

Table 13: Proportions of estimated change-point numbers. The correct value of $J$ is four and the setting is the third one.

As in the case of three change-points, in this one, the behavior of the algorithm in contrast to cases were one and two change-points were considered, improves in a significant manner. While, the estimations of the algorithms for the real value of $J$ usually returns three change-points, the difference of the proportions between the detection of four change-points and this one, is usually lower and ranges lower than 2% for the three settings. Thus, the detection of four change-points constituted the second most recurring case for the three settings such that in the first setting, four change-points were detected as the optimal number 20.50% of the times, in the second setting, four change-points were detected as the optimal number 19.90% of the times and in the third setting, 18.90%. On the other hand, the overall trend oscillates around $J = 4$ as can be appreciated in figure (4).

4.2 Real Data Analysis

Two computational experiments with $PM_{2.5}$ data for Bogotá, Colombia in the timespan January 1, 2018 - December 31, 2020 were conducted.

In the first one, different generations sizes were used each one with 200 individuals and the previously described algorithm was implemented comparing it with the one presented by [5] (frequentist MDL from now on) in order to assess if there existed an improvement after modifying the objective function. For each case and for each generation, the number of change-points, $J$ was stored, as well as their positions and using such positions, the parameters for the underlying generating process were estimated; for the last procedure, and as exposed by [20] and using the results obtained using the package fitdistrplus [21] from the R statistical software [22], the parameters were estimated for a log-normal distribution.

Now with the number of change-points, we compared the distribution for each number of generations versus the optimal true value, $J = 8$ in a graphical manner; likewise, we analyzed the behaviour for the objective function in relation to each generation of the genetic algorithm, again, for each one of the six mentioned cases. In addition, the fitness for the estimated parameters was assessed under a simulation study.

On the other side, the optimal estimation was conducted using the here proposed algorithm and some conclusions based on the nature of the problem were drawed.

The analyses were made using the statistical software R and the code can be shared on request by the reader.

4.2.1 Objective function evaluation

We analyzed different scenarios for the generations of the genetic algorithm. For each one of these, 200 individuals or chromosomes were considered and the procedure previously explained in section 4.2.2 adapted from [5] was applied to every one of them. The objective function Bayesian MDL makes reference to expression (22) which considers exceedances from the mean and the count of these as realizations from a non-homogeneous Poisson process and
the one named as frequentist MDL, makes reference to the procedure based on the Maximum Likelihood function for the underlying generating process of the data, as presented in [5], that in this case corresponds to a log-normal distribution with parameters \( \mu \) and \( \sigma \). For the Bayesian MDL it was needed to define as well, a threshold to account for the exceedances; considering the environmental regulations norms for Bogotá, Colombia, we fixed this threshold as 37 \( \mu g/m^3 \) which is roughly the arithmetic average, \( \bar{X} \), for the data.
To initialize the genetic algorithm population, again, we based again, on the procedure exposed in [5] who point that for environmental data collected from geolocated stations the change-points rate corresponds to around 6% of the observations.

Thus, an initial population of 200 possible change-points was generated, and the genetic algorithm was stared for each of the generation sizes considered, 50, 100, 150, 200, 300 and 500 such that in every case, the objective function was evaluated.

After storing each value for the Bayesian MDL and the frequentist one, we plotted them against each generation for every size considered; such plots can be seen in figure 5 with the black solid line and the red dashed one for each respective objective function.

It is noteworthy that, while the frequentist MDL decreases as the number of generations increases, this behavior is slow in comparison to the Bayesian MDL that after 10 generations approximately, for all sizes, decreases rapidly and reaches a steady behavior afterwards, such that it takes values that seem to show no major variation from each other in contrast to the previously mentioned case.

The latter can be clearly appreciated in figures 6 and 7 where for the frequentist case, it is present a wide spread with respect to the percentile 50 or median and likewise, there seems to exist the presence of a sizable number of values that could be considered as outliers. On the other hand, while the Bayesian MDL presents some outliers as well, these values can be associated with roughly the first ten generations in each scenario such that if we omit them, the associated distribution seems symmetric and highly clustered around the median.

Then, we can consider the scheme exposed here more stable, in contrast to the one presented in [5] when it comes to the values that takes the objective function allowing then for consistency in the replication of results and eventually, convergence as the algorithm iterates for a larger number of generations accordingly to the features the method adapts from natural selection processes.

4.2.2 Estimation of number of change-points

Using the same experimental scheme as the previous subsection, it was stored as well for each scenario, the number of change-points, \( J \) detected as optimal by the genetic algorithm for each objective function.

We denoted by \( \hat{J}_{\text{Freq}} \) the total of change-points estimated by the frequentist MDL and by \( \hat{J}_{\text{Bayes}} \), the estimation using the Bayesian MDL. Then, we plotted the frequency with that a time index for the 1096 considered and associated to the time series observed data was detected as a change-point, this for 50, 100, 150, 200, 300 and 500 generations.

Thus, for example, if the tenth observation constitutes a change-point, then, the count for this index increases, as well as the bar in the barplot in a vertical manner, this as many times as this index is detected as such for all generation sizes. On the other hand, the real number of change-points, \( J = 8 \) is represented in all panels as a positive vertical black dashed line for such index.

When we plotted the number of change-points detected by both objective functions for each generation sizes (figure 8), the algorithm using the frequentist MDL underestimates \( J \) such that the most common values that \( \hat{J}_{\text{Freq}} \) takes are in the set \( \{2, 3, 4\} \) with this last element, the mode for most of the cases. On the other hand, the Bayesian MDL represents an improvement in comparison to its predecessor, such that, if there are instances where the the genetic algorithm overestimates the number of change-points, the overall behavior shows that the values taken by \( \hat{J}_{\text{Bayes}} \) are clustered around the real one, that is, \( J = 8 \) regardless of the number of generations considered.

Besides that, it is worth to notice the consistency in the results of the genetic algorithm when the Bayesian MDL is used for all evaluated generation sizes, such that the values returned for \( J \) are close to the real optimum; then this estimation could be considered as one of lower bias in contrast to the one based on the Maximum Likelihood function for the data.
4.2.3 Parameter and Values Estimation Assessment

For this experiment we began by taking for each generation for each considered case, the estimation (piecewise and using the detected change-points) of the parameters $\mu$ and $\sigma$ as exposed in [5] when the data is approximately distributed log-normal. Thus, when 50 generations were considered, there were 50 estimations for $\mu$ and $\sigma$, 100 estimations for 100 generations and so on.

Because beforehand we did not have the density values for the data and as well, the true values of the parameters, to assess the fitting using both objective functions we proceeded in the following way; let’s suppose the first estimation for the parameters in any of the scenarios for the size of the generations, with such estimates and using the statistical software R, we simulated 1096 observations from a log-normal distribution. With these simulations, then we computed the Mean Squared Error (MSE),

$$MSE = \frac{1}{1096} \sum_{i=1}^{1096} (Y_t - \hat{Y}_t)^2,$$

$Y_t$, observed values and $\hat{Y}_t$, Simulated values.

This was repeated 1000 times. Then, we took the average for the MSE for the 1000 repetitions and stored. This was made for each one of the other generations for each size such that we obtained an approximated distribution for the MSE in all cases, which can be appreciated in figure (9).

Analyzing the behavior of the simulations, first, it is noteworthy the range associated with the MSE when using the estimates through the Bayesian MDL that for all cases, was found to be under the range for the estimates using the frequentist MDL. As the generation size increases, eventually can be observed that for both objective functions, there...
exists a lower dispersion for the MSE, nevertheless, the Bayesian MDL estimated MSE seems to have a slightly higher dispersion when we consider the generation sizes 200, 300 and 500.

Yet, for the mentioned cases, the presence of outliers seems to be lower than for the frequentist MDL.

Thus, the use of a genetic algorithm with the objective function here exposed does not only provides stability when it comes to the values it takes, but as well, it does estimates better the number of change-points, $J$, and as well, the parameters for the generating process underlying to the observed data.

### 4.2.4 Optimal values analysis

We implemented the previous algorithm for $PM_{2.5}$ data in Bogotá, Colombia for the time interval January 1, 2018 - December 31, 2020. The time series can be observed in (10) where different exceedances of the Colombian standard are shown in dotted line, that is, a threshold of 37 µg/m$^3$ was established.

The conditions considered for the present implementation included taking generations of size 50 with 50 offspring each. The distributions used to generate the first generation and the mutations were those mentioned in section (2.2.2). As initial settings for the values to be optimized in the first equation of expression (2), $\alpha$ and $\beta$ were taken to be 0.1 and 0.5, respectively. For this process we made use of the R statistical software `optim` function.
The a priori distributions used for $\alpha$ and $\beta$ were gammas with hyperparameters $\phi_{11} = 1, \phi_{12} = 2, \phi_{21} = 3$ and $\phi_{22} = 1.2$, as defined in section (3.1.1), given that the intensity rate in use is Weibull.

The optimal chromosome was found to be (8, 400, 408, 445, 488, 627, 654, 661, 798), such that there are $m = 8$ change-points; next the respective location for these points in the 1096 days of the observed timeline are given, it can be observed as well, that the first 4 change-points are very close. Also, the corresponding Bayesian-MDL value was 2065. The following table (14) shows the date corresponding to each change-point.

These surpasses occurred during the rainy season in Bogotá. If the dates in the table above are compared with the measurements in figure (10), the seasons with the most consistent peaks over time can be captured.

The graph in figure (11) shows that before day 400, i.e., Monday, February 4, the rate of exceedances of the 37 $\mu g/m^3$ threshold had been decreasing sharply, but after it, the highest emission rate for eight consecutive days was recorded. This high average rate is around 1.004, as shown in (Table 15), for the second regime. In the third regime, it decreased to an average rate of 0.9468; in the fourth, it jumps to 0.6748, and it is in the fifth regime that it achieves the lowest drop, 0.2090, before the intensity function rises again to levels above the 0.2 threshold overshoots per unit of time. This rate present in the fifth regime goes from May 3 to September 19, 2019. Thus, it is evident that the fifth regime occurred...
Change-point Detection Using a GA and a BMDL for Time Exceedances

Figure 9: Mean Squared Error for different generation sizes

| Change Point | Day of the week | Date               |
|--------------|-----------------|--------------------|
| 400          | Monday          | February 4, 2019   |
| 408          | Tuesday         | February 12, 2019  |
| 445          | Thursday        | March 21, 2019     |
| 488          | Friday          | May 3, 2019        |
| 627          | Thursday        | September 19, 2019 |
| 654          | Wednesday       | October 16, 2019   |
| 661          | Wednesday       | October 23, 2019   |
| 798          | Sunday          | March 8, 2020      |

Table 14: Points of change with their respective dates

before the COVID-19 pandemic lockdown was declared in Colombia. Therefore, this may be the result of a public policy aimed at reducing the emissions of $PM_{2.5}$.

As part of the 2010-2020 ten-year plan for air pollution control, the use of emission control systems in cargo transport vehicles and motorcycles, and as well as the integrated public transport system (SITP for its acronym in Spanish) policies were implemented. The later includes the replacement of old buses with internal combustion engine with electric or hybrid buses. In addition to the above, a few days before the fifth regime, resolution 383 (see RES19 (2019)) was issued, which declared a yellow alert for particulate matter exceedances. Considering the first regime represented in figure (11), the rapid deceleration in the emission of threshold exceedances can also be seen as a consequence of this resolution. Such is the case of restrictions on the use of transportation and the mobility sector, in addition to those
aimed at the operations of industries that use combustion processes associated mainly with the burning of biomass and the use of fossil or liquid fuels.

5 Conclusions and future work

The detection of change-points using a genetic algorithm and a Bayesian-MDL as selection criterion yielded results that agree with the public policies implemented in Bogotá, Colombia, both regarding the contamination alerts issued by
Change-point Detection Using a GA and a BMDL for Time Exceedances

| Regime | Min  | Mean | Max   |
|--------|------|------|-------|
| 1      | 0.1894 | 0.2379 | 0.6555 |
| 2      | 1.002  | 1.004 | 1.007 |
| 3      | 0.9363 | 0.9468 | 0.9577 |
| 4      | 0.6657 | 0.6748 | 0.6841 |
| 5      | 0.1991 | 0.2090 | 0.2201 |
| 6      | 0.4801 | 0.4832 | 0.4864 |
| 7      | 0.5765 | 0.5773 | 0.5781 |
| 8      | 0.4761 | 0.4896 | 0.5043 |
| 9      | 0.1744 | 0.1849 | 0.1971 |

Table 15: Minimum, Maximum and Mean for each Regime

the monitoring network, as well as the mobility restrictions due to the quarantine caused by the SARS-CoV-2 virus pandemic. It is seen that the algorithm gives good results both in regions with stations, as is the case of Mexico City, ([18], [17]), and in tropical areas where these periodic changes in temperature do not occur, showing the robustness of the technique. We hope that in the near future these methods will prove to be a useful tool for the government agencies in charge of measuring the effectiveness of the actions taken to reduce air pollution, and thus reduce its impacts both in the environment and the health of the inhabitants.

**Conflict of interest**

The authors declare that they have no conflict of interest.

**References**

[1] Michèle Basseville and Igor V. Nikiforov. *Detection of Abrupt Changes - Theory and Application*. Prentice Hall, Englewood Cliffs, NJ, 1993.

[2] Charles Truong, Laurent Oudre, and Nicolas Vayatis. Selective review of offline change point detection methods. *Signal Processing*, 167:107299, 2020.

[3] Samaneh Aminikhanghahi and Diane Cook. A survey of methods for time series change point detection. *Knowledge and Information Systems*, 51:339–367, 2017.

[4] Toby Hocking, Guillem Rigaill, and Guillaume Bourque. Peakseg: constrained optimal segmentation and supervised penalty learning for peak detection in count data. In Francis Bach and David Blei, editors, *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 324–332. PMLR, 2015.

[5] Shanghong Li, Robert Lund, Li, and Robert Lund. Multiple Changepoint Detection via Genetic Algorithms. *Journal of Climate*, 25(2):674–686, Jan 2012.

[6] J. Achcar, A. Fernandez-Bremauntz, E. Rodrigues, and G. Tzintzun. Estimating the number of ozone peaks in mexico city using a non-homogeneous poisson model. *Environmetrics*, 19:469–485, 2008.

[7] B. Suarez-Sierra, E. Rodrigues, and G. Tzintzun. Rate of ozone and pm10 exceedances: A case study comparing data from mexico city and bogota. *Communications in Statistics: Case Studies, Data Analysis and Applications*, 5(2):153–165, 2019.

[8] B. Suarez-Sierra, E. Rodrigues, and G. Tzintzun. An application of a non-homogeneous poisson model to study pm2.5 exceedances in mexico city and bogota. *Journal of Applied Statistics*, 49(9):2430–2445, 2022.

[9] J. F. Lawless. *Statistical Models and Methods for Lifetime Data, Second Edition*. John Wiley and Sons, Inc., Hoboken, NJ, 2 edition, 2002.

[10] D. R. Cox and P. A. W. Lewis. *The statistical analysis of series of events*. Methuen’s Monographs on Applied Probability and Statistics. Springer Dordrecht, 1 edition, 1966.

[11] G. S. Mudholkar, D. K. Srivastava, and M. Freimer. The exponentiated weibull family: A reanalysis of the bus-motor-failure data. *Technometrics*, 37(4):436–445, 1995.

[12] Cid Ramirez, Juan Esteban, and Jorge Alberto Achcar. Bayesian inference for nonhomogeneous poisson processes in software reliability models assuming nonmonotonic intensity functions. *Computational Statistics & Data Analysis*, 32(2):147–159, December 1999.
[13] J. D. Musa and K. Okumoto. A logarithmic poisson execution time model for software reliability measurement. In Proceedings of the 7th International Conference on Software Engineering, ICSE ’84, pages 230–238, Piscataway, NJ, USA, 1984. IEEE Press.

[14] Amrit L. Goel and Kazu Okumoto. An analysis of recurrent software errors in a real-time control system. In Proceedings of the 1978 Annual Conference, ACM ’78, pages 496–501, New York, NY, USA, 1978. ACM.

[15] E. R. Rodrigues and J. Achcar. Applications of Discrete-time Markov Chains and Poisson Processes to Air Pollution Modeling and Studies. Number 1 in SpringerBriefs in Mathematics. Springer-Verlag New York, 2013.

[16] Bo-Yi Yang, Shujun Fan, and E. Thiering. Ambient air pollution and diabetes: A systematic review and meta-analysis. Environmental Research, 180, 2020.

[17] J. Achcar, E. Rodrigues, and G. Tzintzun. Using non-homogeneous poisson models with multiple change-points to estimate the number of ozone exceedances in mexico city. Environmental and Ecological Statistics, 22, 2011.

[18] J.A. Achcar, E. Rodrigues, C. Paulino, and P. Soares. Non-homogeneous poisson processes with a change-point: an application to ozone exceedances in méxico city. Environmental and Ecological Statistics, 17:521–541, 2010.

[19] Richard A Davis, Thomas C. M Lee, and Gabriel A Rodriguez-Yam. Structural Break Estimation for Nonstationary Time Series Models. Journal of the American Statistical Association, 101(473):223–239, mar 2006.

[20] Brian Rumburg, Richard Alldredge, and Candis Claiborn. Statistical distributions of particulate matter and the error associated with sampling frequency. Atmospheric Environment, 35(16):2907–2920, 2001.

[21] Marie Laure Delignette-Muller and Christophe Dutang. fitdistrplus: An R package for fitting distributions. Journal of Statistical Software, 64(4):1–34, 2015.

[22] R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2022.