Chapter 17

A Non-Homogeneous Markov Chain Model to Study Ozone Exceedances in Mexico City

Eliane R. Rodrigues, Mario H. Tarumoto and Guadalupe Tzintzun

Additional information is available at the end of the chapter

1. Introduction

In many cities around the world, air pollution is among the many environmental problems that affect their population. Among the many known facts about the impact of pollution on human health, we have that for ozone concentration levels above 0.11 parts per million (0.11ppm), the susceptible part of the population (e.g., the elderly, ill, and newborn) staying in that environment for a long period of time, may experience serious health deterioration (see, for example, [1–10]). Therefore, to understand the behaviour of ozone and/or pollutants in general, is a very important issue.

It is possible to find in the literature a vast amount of works that try to answer some of the many issues arising in the study of pollutants' behaviour. Depending on the type of questions that one is trying to answer, different methodologies may be used. Among the many works concentrating on the study of ozone behaviour are, [11–13] using extreme value theory to study the behaviour of the maximum ozone measurements; [14] using time series analysis; [15] using volatility models to study the variability of the weekly average ozone measurements; [13, 16] using homogeneous Poisson processes and [17, 18] using non-homogeneous Poisson models to analyse the probability of having a certain number of ozone exceedances in a time interval of interest; [19] using compound Poisson models to study the occurrence of clusters of ozone exceedances as well as their mean duration time; and [20] using queueing model to study the occurrence of cluster of ozone exceedances as well as their size distribution.

In the environmental area, it is also possible to find works using Markov chains models. Some of them are, [21, 22] where non-homogeneous Markov models are used to study the occurrence of precipitation. We also have [23] where those types of models are used to study tornado activity. In the case of ozone modelling we have, for instance, the works of [24–26] using time homogeneous Markov chains. In those works the interest was in estimating the probability that the ozone measurement would be above (below) a given threshold, conditioned on where it lays in the present and in the past days.

In [24], the order of the Markov chain was estimated using auto-correlation function. Its transition matrix was estimated using the maximum likelihood method (see, for instance, [27, 28], among others). In [25], the order of the chain was also considered an unknown quantity that needed to be estimated. The Bayesian approach (see, for example, [29]) was used to estimate the order as well as the transition probabilities of the chain. In particular, the maximum *à posteriori* method was used. In [26], the estimation of the order of the chain is performed using the Bayesian approach using the so-called trans-dimensional Markov chain Monte Carlo algorithm ([30, 31]). The transition matrix of the chain was obtained through the maximum *à posteriori* method. However, the common denominator of those works is that the Markov chain model used was a time homogeneous one. Since ozone data are not, in general, time homogeneous, the data had to be split into time homogeneous segments and the analysis was made for each segment separately.

Here, the interest also resides in estimating, for instance, the probability that the ozone measurement will be above a given threshold some days into the future, given where it stands today and in the past few days. Although in the present work we also use Markov chain models and the Bayesian approach, the novelty here is that the time-homogeneous assumption is dropped. Here, we consider a non-homogeneous Markov chain model. We assume that the order of the chain as well as its transition probabilities are unknown and need to be estimated. The chosen method of estimation is also the maximum *à posteriori*.

This work is presented as follows. In Section 2 the non-homogeneous Markov chain model is given. Section 3 presents the Bayesian formulation of the model. An application to ozone measurements from Mexico City is given in Section 4. In Section 5 some comments about the methodology and results are made. In an Appendix, before the list of references, we present the code of the programme used to estimate the order and the transition probabilities of the Markov chain.

2. A non-homogeneous Markov chain model

The mathematical model considered here may be described as follows. Let $N > 0$ be a natural number representing the number of years in which measurements were taken. Let T_i, $i = 1, 2, \ldots, N$, be natural numbers representing the amount of observations in each year. Hence, we have that for a given year i, either $T_i = 366$ or $T_i = 365$, depending on whether or not we have a leap year, $i = 1, 2, \ldots, N$.

Let $Z_t^{(i)}$ be the ozone concentration on the tth day of the ith year, $t = 1, 2, \ldots, T_i$, $i = 1, 2, \ldots, N$. Following [23], we will set $T_i = T = 366$, $i = 1, 2, \ldots, N$, with the convention that for non leap year, we assign $Z_T^{(i)} = 0$.

Remark. Since, we are taking all years of the same length we will drop the index i from the notation.

Denote by $L > 0$ the environmental threshold we are interested in knowing if the ozone concentration has surpassed or not. Define $\mathbf{Y} = \{Y_t : t \geq 0\}$ by,

$$Y_t = \begin{cases} 0, \text{ if } Z_t < L \\ 1, \text{ if } Z_t \geq L. \end{cases} \tag{1}$$

Hence, Y_t indicates whether or not in the tth day the threshold L was exceeded.

As in [25], we assume that $\mathbf{Y}$ is ruled by a Markov chain of order $K \geq 0$. In contrast with that work, in the present case the Markov chain is a non-homogeneous one. Hence, denote by $X^{(K)} = \{X_t^{(K)} : t = 1, 2, \dots T\}$, the corresponding non-homogeneous Markov chain of order K. We assume that K has as state space a set $\mathcal{S} = \{0, 1, \dots, M\}$, for some fixed integer $M \geq 0$, such that, $M \leq T$ with probability one.

Note that, $X^{(K)}$ has as state space the set $S_1^{(K)} = \{(x_1, x_2, \dots, x_K) \in \{0,1\}^K\}$, with $S_1^{(0)} = S_1^{(1)}$. Also, note that (see [25]), if the set of observed value is $(y_1, y_2, \dots, y_T)$, then the transition probabilities of $X^{(K)}$ are such that

$$P(X_{t+1}^{(K)} = w \,|\, X_t^{(K)} = x_t = (y_{t+1}, y_{t+2}, \dots, y_{t+K})),$$

is different of zero if, and only if, $w = (y_{t+2}, y_{t+3}, \dots, y_{t+K+1}) \in S_1^{(K)}$, with $0 \leq t \leq T - K$. Therefore, w occurs, if and only if, the observation following $y_{t+1}, y_{t+2}, \dots, y_{t+K}$, is y_{t+K+1}. This enables us to work with a more treatable state space for $X^{(K)}$, and therefore, to have a better form for the transition matrix.

Hence, as in [25, 32], we consider the transformed state space $S_2^{(K)} = \{0, 1, \dots, 2^K - 1\}$, which is obtained from $S_1^{(K)}$ by using the transformation $f : S_1^{(K)} \to S_2^{(K)}$, given by, $f(w_1, w_2, \dots, w_K) = \sum_{l=0}^{K-1} w_{l+1} 2^l$. Let $(x_1, x_2, \dots, x_K) \leftrightarrow \overline{m}$ indicate that the state $(x_1, x_2, \dots, x_K) \in S_1^{(K)}$ corresponds to the state $\overline{m} \in S_2^{(K)}$. Hence, the transition probabilities of $X^{(K)}$ may be written as (see, for instance, [25]),

$$P_{\overline{m}j}^{(K)}(t) = P(Y_{t+K+1} = j \,|\, X_t^{(K)} = (y_{t+1}, y_{t+2}, \dots, y_{t+K}) \leftrightarrow \overline{m}), \tag{2}$$

where $\overline{m} \in S_2^{(K)}$, $j \in \{0, 1\}$, and $0 \leq t \leq T - K$.

Now, indicate by $Q_{\overline{m}}^{(K)}(t)$, $\overline{m} \in S_2^{(K)}$, $\overline{m} \in S_2^{(K)}$, the probability $P(X_t^{(K)} = \overline{m})$. Hence, when $t = 1$, we have that $Q_{\overline{m}}^{(K)}(1)$, $\overline{m} \in S_2^{(K)}$, is the initial distribution of $X^{(K)}$. When $K = 0$, we have that $P_{\overline{m}j}^{(0)}(t) = Q_j^{(0)}(t)$, $t = 1, 2, \dots, T$, $j = 0, 1$, $\overline{m} \in S_2^{(0)} = \{0, 1\}$.

Remarks. 1. When $K = 1$, we have that $P_{\overline{m}j}^{(0)}(t)$, $j = 0, 1$, $t = 1, 2, \dots, T - K$, are the usual one-step transition probabilities. When $K = 0$, the transition probabilities are just the probabilities $Q_{\overline{m}}^{(0)}(t)$, associated to each state $\overline{m} \in S_2^{(0)} = \{0, 1\}$, with $t = 1, 2, \dots, T$.

2. Unless otherwise stated, from now on, we are going to use the state space $S_2^{(K)}$ and the corresponding transition probabilities.

3. $\mathbf{Y}$ is going to represent our observed data.

In addition to estimating the order K of the Markov chain, we will also estimate its transition probabilities $P_{\overline{m}j}^{(K)}(t)$ as well as the probabilities $Q_{\overline{m}}^{(K)}(t)$, $j \in \{0, 1\}$, $\overline{m} \in S_2^{(K)}$, for each t. We

indicate by $P^{(K)}(t) = \left(P^{(K)}_{\overline{m}j}(t)\right)_{j\in\{0,1\},\overline{m}\in S_2^{(K)}}$, the transition matrix at time t. Note that, if $K = 0$, then $P^{(0)}_{\overline{m}j}(t) = Q^{(0)}_j(t)$, $j \in \{0,1\}$, $\overline{m} \in S_2^{(0)}$, $t = 1, 2, \ldots, T$.

3. A Bayesian estimation of the parameters of the model

There are many ways of estimating the order and the transition matrix of a non-homogeneous Markov chain. One way of estimating the order is via the auto-correlation function associated to the chain throughout the years. Another way is to use the Bayesian approach. When it comes to estimating the transition probabilities we have, for instance, the maximum likelihood method ([33]) and the empirical estimator ([34]) which are essentially the same. In the present work, we will use the Bayesian approach (see, for instance, [29, 35]) to estimate the order and the transition probabilities. In particular, we are going to adopt the maximum *à posteriori* approach. Inference will be performed using the information provided by the so-called posterior distribution of the parameters. The posterior distribution of a vector of parameters $\boldsymbol{\theta}$ given the observed data $\mathbf{D}$, indicated by $P(\boldsymbol{\theta} \mid \mathbf{D})$, is such that $P(\boldsymbol{\theta} \mid \mathbf{D}) \propto L(\mathbf{D} \mid \boldsymbol{\theta})\, P(\boldsymbol{\theta})$, where $L(\mathbf{D} \mid \boldsymbol{\theta})$ is the likelihood function of the model, and $P(\boldsymbol{\theta})$ is the prior distribution of the vector $\boldsymbol{\theta}$.

In the present case, we have that the vector of parameter is $\boldsymbol{\theta} = (K, Q^{(K)}(1), P^{(K)}(t), t = 1, 2, \ldots, T-K)$. If $K = 0$, the range of t is $\{1, 2, \ldots, T\}$. The vector $\boldsymbol{\theta}$ belongs to the following sample space

$$\Theta = \bigcup_{K=0}^{M} \left(\{K\} \times \Delta_2^{2^K} \times \Delta_2^{(T-K)\,2^K}\right)$$

where $\Delta_2 = \{(x_1, x_2) \in \mathbb{R}^2 : x_i \geq 0,\, i = 1, 2,\, x_1 + x_2 = 1\}$ is the one dimensional simplex. (Note that if we have $K = 0$, then the parametric space reduces to $\Theta = \Delta_2^T$.) In the present case we have $\mathbf{D} = \mathbf{Y}$

Let $(x_1, x_2, \ldots, x_K)$ be such that $Y_t = x_1$. Indicate by $n^{(K)}_{\overline{m}i}(t)$ the number of years in which the vector $(x_1, x_2, \ldots, x_K)$ corresponding to a state $\overline{m} \in S_2^{(K)}$ is followed by the observation i, $i = 0, 1$. Also define $n^{(0)}_{\overline{m}}(t)$, $\overline{m} \in S_2^{(0)} = \{0, 1\}$, as the number of years in which we have the observation $\overline{m}$ at time t, $t = 1, 2, \ldots, T$. Additionally, let $n^{(K)}_{\overline{m}}$ indicate the number of years in which the state corresponding to the initial K days is equivalent to the value $\overline{m} \in S_2^{(K)}$, $K \geq 0$. In the case of $K = 0$, we have $n^{(0)}_{\overline{m}} = n^{(1)}_{\overline{m}}$, and $\overline{m} \in S_2^{(0)} = S_2^{(1)} = \{0, 1\}$.

Therefore, since a Markovian model is assumed, the likelihood function is given by (see, for instance, [23, 33])

$$L(\mathbf{Y} \mid \boldsymbol{\theta}) \propto \left(\prod_{\overline{m}\in S_2^{(K)}} \left[Q^{(K)}_{\overline{m}}(1)\right]^{n^{(K)}_{\overline{m}}}\right) \left(\prod_{t=1}^{T-K} \left[P^{(K)}_{\overline{m}0}(t)\right]^{n^{(K)}_{\overline{m}0}(t)} \left[1 - P^{(K)}_{\overline{m}0}(t)\right]^{n^{(K)}_{\overline{m}1}(t)}\right). \tag{3}$$

Note that when $K = 0$, the expression (3) simplifies to

$$L(\mathbf{\Upsilon} \mid \boldsymbol{\theta}) \propto \prod_{t=1}^{T} \prod_{m=0}^{1} \left[Q_{\overline{m}}^{(0)}(t) \right]^{n_{\overline{m}}^{(0)}(t)},$$

where $Q_{\overline{m}}^{(0)}(t) = P(X_t^{(0)} = \overline{m}) = P(Y_t = \overline{m})$, $t = 1, 2, \ldots, T$, $\overline{m} \in S_2^{(0)} = \{0, 1\}$.

The prior distribution of the vector of parameters is given as follows. We assume a prior independence of $P^{(K)}(t)$ as functions of t. Also, since the forms of $P^{(K)}(t)$ and $Q^{(K)}(1)$ depend on the value of K, we have that for $\boldsymbol{\theta} = (K, Q^{(K)}(1), P^{(K)}(t), t = 1, 2, \ldots, T - K)$,

$$P(\boldsymbol{\theta}) = P(Q^{(K)}(1) \mid K) \left[\prod_{t=1}^{T-K} P\left(P^{(K)}(t) \mid K \right) \right] P(K),$$

where $P(Q^{(K)}(1) \mid K)$ and $P\left(P^{(K)}(t) \mid K \right)$ are the prior distributions of the initial distribution $Q^{(K)}(1)$ and of the transition matrix $P^{(K)}(t)$ given the order of the chain, respectively, and $P(K)$ the prior distribution of the order K.

Remark. When we have $K = 0$, the vector of parameters is $\boldsymbol{\theta}' = (Q_{\overline{m}}^{(0)}(t);\ t = 1, 2, \ldots, T)$, whose prior distribution is

$$P(\boldsymbol{\theta}') = \left[\prod_{t=1}^{T} P(Q^{(0)}(t)) \right],$$

where $P(Q^{(0)}(t))$ is the prior distribution of the probability vector $Q^{(0)}(t) = (Q_{\overline{m}}^{(0)}(t),\ \overline{m} = 0, 1)$, $t = 1, 2, \ldots, T$.

Given the nature of transition matrices, we are going to assume that rows are independent. We also assume that, given the order K of the chain, each row of the transition matrix $P^{(K)}(t)$ will have as prior distribution a Dirichlet distribution with appropriate hyperparameters. Therefore, given that $K = k$, row $(P_{\overline{m}0}^{(k)}(t), P_{\overline{m}1}^{(k)}(t))$ has as prior distribution a Dirichlet$(\alpha_{\overline{m}0}^{(K)}(t), \alpha_{\overline{m}1}^{(K)}(t))$, $t = 1, 2, \ldots, T$; i.e.,

$$P\left(P^{(K)}(t) \mid K \right) = \prod_{\overline{m} \in S_2^{(K)}} \left(\frac{\Gamma(\alpha_{\overline{m}0}^{(K)}(t) + \alpha_{\overline{m}1}^{(K)}(t))}{\Gamma(\alpha_{\overline{m}0}^{(K)}(t))\, \Gamma(\alpha_{\overline{m}1}^{(K)}(t))} \left\{ \left[P_{\overline{m}0}^{(K)}(t) \right]^{\alpha_{\overline{m}0}^{(K)}(t) - 1} \left[P_{\overline{m}1}^{(K)}(t) \right]^{\alpha_{\overline{m}1}^{(K)}(t) - 1} \right\} \right)$$

for t in the appropriate range. In the case of initial distribution $Q^{(K)}(1)$, we also have a Dirichlet prior distribution, but now with hyperparameters $(\alpha_{\overline{m}}^{(K)}; \overline{m} \in S_2^{(K)})$.Therefore,

$$P\left(Q^{(K)}(1) \mid K \right) = \frac{\Gamma(\sum_{\overline{m} \in S_2^{(K)}} \alpha_{\overline{m}}^{(K)})}{\prod_{\overline{m} \in S_2^{(K)}} \Gamma(\alpha_{\overline{m}}^{(K)})} \prod_{\overline{m} \in S_2^{(K)}} \left[Q_{\overline{m}}^{(K)}(1) \right]^{\alpha_{\overline{m}}^{(K)} - 1}.$$

If $K = 0$, then $Q^{(0)}(t)$ has as prior distribution a Dirichlet distribution with hyperparameters $(\alpha_{\overline{m}}^{(0)}(t); t = 1, 2, \ldots, T; \overline{m} \in S_2^{(0)} = \{0, 1\})$. Therefore, for any given t,

$$P\left(Q^{(0)}(t) \mid K = 0\right) = \frac{\Gamma(\sum_{\overline{m}=0}^{1} \alpha_{\overline{m}}^{(0)}(t))}{\prod_{\overline{m}=0}^{1} \Gamma(\alpha_{\overline{m}}^{(0)}(t))} \prod_{\overline{m}=0}^{1} \left[Q_{\overline{m}}^{(0)}(t)\right]^{\alpha_{\overline{m}}^{(0)}(t)-1}.$$

We assume that K has as prior distribution a truncated Poisson distribution defined on the set $\mathcal{S}$ with rate $\lambda > 0$; i.e.,

$$P(K) = \frac{\lambda^K}{K!} I_{\mathcal{S}}(K),$$

where $I_A(x) = 1$, if $x \in A$ and is zero otherwise.

Therefore, we have from [25, 32, 36], that the conditional posterior distribution of $P^{(K)}(t)$ given K, is

$$P\left(P^{(K)}(t) \mid K, \mathbf{Y}\right) \propto \prod_t \left\{ \prod_{\overline{m} \in S_2^{(K)}} \left(\left[P_{\overline{m}0}^{(K)}(t)\right]^{n_{\overline{m}0}^{(K)}(t)+\alpha_{\overline{m}0}^{(K)}(t)-1} \left[P_{\overline{m}1}^{(K)}(t)\right]^{n_{\overline{m}1}^{(K)}(t)+\alpha_{\overline{m}1}^{(K)}(t)-1} \right) \right\}.$$

Hence, $P\left(P^{(K)}(t) \mid K, \mathbf{Y}\right)$ is proportional to the product of Dirichlet distributions with hyperparameters $(n_{\overline{m}0}^{(K)}(t) + \alpha_{\overline{m}0}^{(K)}(t), n_{\overline{m}1}^{(K)}(t) + \alpha_{\overline{m}1}^{(K)}(t))$. The mode of each Dirichlet distribution is known and is given by (see [37]),

$$P_{\overline{m}i}^{(K)}(t) = \frac{n_{\overline{m}i}^{(K)}(t) + \alpha_{\overline{m}i}^{(K)}(t) - 1}{\sum_{j=0}^{1} \left[n_{\overline{m}j}^{(K)}(t) + \alpha_{\overline{m}j}^{(K)}(t) - 1\right]}, \quad i = 0, 1; \overline{m} \in S_2^{(K)}; K \in \mathcal{S}; t = 1, 2, \ldots, T - K. \quad (4)$$

Additionally, the posterior distribution of the initial distribution $Q^{(K)}(1)$ given K is

$$P(Q^{(K)}(1) \mid K, \mathbf{Y}) \propto \prod_{\overline{m} \in S_2^{(K)}} \left[Q_{\overline{m}}^{(K)}(1)\right]^{\alpha_{\overline{m}}^{(K)}+n_{\overline{m}}^{(K)}-1}.$$

Therefore, $P(Q^{(K)}(1) \mid K, \mathbf{Y})$ is proportional to a Dirichlet distribution with hyperparameters $(\alpha_{\overline{m}}^{(K)} + n_{\overline{m}}^{(K)}; \overline{m} \in S_2^{(K)})$. Hence, as in the case of the posterior distribution of $P^{(K)}(t)$, the mode of $P(Q^{(K)}(1) \mid K, \mathbf{Y})$ is,

$$Q_{\overline{m}}^{(K)}(1) = \frac{n_{\overline{m}}^{(K)} + \alpha_{\overline{m}}^{(K)} - 1}{\sum_{\overline{m}' \in S_2^{(K)}} \left[n_{\overline{m}'}^{(K)} + \alpha_{\overline{m}'}^{(K)} - 1\right]}, \quad \overline{m} \in S_2^{(K)}; K \in \mathcal{S}. \quad (5)$$

When $K = 0$, we have

$$P(Q^{(0)}(t) \mid K = 0, \mathbf{Y}) \propto \left(\prod_{\overline{m}=0}^{1} \left[Q_{\overline{m}}^{(0)}(t) \right]^{n_{\overline{m}}^{(0)}(t) + \alpha_{\overline{m}}^{(0)}(t) - 1} \right), \quad t = 1, 2, \ldots, T.$$

Therefore, for each $t = 1, 2, \ldots, T$,

$$Q_{\overline{m}}^{(0)}(t) = \frac{n_{\overline{m}}^{(0)}(t) + \alpha_{\overline{m}}^{(0)}(t) - 1}{\sum_{\overline{m}'=0}^{1} \left[n_{\overline{m}'}^{(0)}(t) + \alpha_{\overline{m}'}^{(0)}(t) - 1 \right]}, \quad \overline{m} = 0, 1. \tag{6}$$

Furthermore, we also have, from [25], that

$$L(\mathbf{Y} \mid K) \propto \frac{\Gamma\left(\sum_{\overline{m} \in S_2^{(K)}} \alpha_{\overline{m}}^{(K)} \right)}{\Gamma\left(\sum_{\overline{m} \in S_2^{(K)}} \left[\alpha_{\overline{m}}^{(K)} + n_{\overline{m}}^{(K)} \right] \right)} \left(\prod_{\overline{m} \in S_2^{(K)}} \frac{\Gamma\left(n_{\overline{m}}^{(K)} + \alpha_{\overline{m}}^{(K)} \right)}{\Gamma\left(\alpha_{\overline{m}}^{(K)} \right)} \right)$$
$$\prod_{t=1}^{T-K} \left\{ \prod_{\overline{m} \in S_2^{(K)}} \left(\frac{\Gamma[\alpha_{\overline{m}0}^{(K)}(t) + \alpha_{\overline{m}1}^{(K)}(t)]}{\Gamma(\sum_{j=0}^{1} [n_{\overline{m}j}^{(K)}(t) + \alpha_{\overline{m}j}^{(K)}(t)])} \prod_{j=0}^{1} \frac{\Gamma(n_{\overline{m}j}^{(K)}(t) + \alpha_{\overline{m}j}^{(K)}(t))}{\Gamma(\alpha_{\overline{m}j}^{(K)}(t))} \right) \right\},$$

with the appropriate adaptation for the case of $K = 0$. Hence, the posterior distribution of the order K is

$$P(K \mid \mathbf{Y}) = \frac{1}{c} L(\mathbf{Y} \mid K) \frac{\lambda^K}{K!} \tag{7}$$

where $c = \sum_{k \in \mathcal{S}} L(\mathbf{Y} \mid K = k) \left(\lambda^k / k! \right)$ is the normalising constant.

Therefore, in order to obtain the probability of interest, we just have to use (7) to estimate the value of K that maximises that posterior probability, and then use (4), (5), and/or (6) (depending on the case), in order to calculate the corresponding transition matrix and initial distribution, respectively.

The hyperparameters appearing in the prior distribution will be considered known and will be specified later.

4. Application to ozone data from the monitoring network of Mexico City

In this section we apply the model to the Mexico City's ozone measurements. The data used consist of twenty two years of the daily maximum ozone measurements (from 01 January 1990 to 31 December 2011) provided by the monitoring network of the Metropolitan Area of Mexico City. The Metropolitan Area is divided into five regions, namely, Northeast (NE), Northwest (NW), Centre (CE), Southeast (SE), and Southwest (SW). The monitoring stations are placed throughout the city. Measurements in each monitoring station are

obtained minute by minute and the averaged hourly result is reported at each station. The daily maximum measurement for a given region is the maximum over all the maximum averaged values recorded hourly during a 24-hour period by each station placed in the region. Since emergency alerts in Mexico City are declared regionally, we will analyse each region separately.

The Mexican ozone standard considers the threshold 0.11ppm (see [38]). Hence, we will take that value as one of our thresholds. Additionally, for comparison purpose, we will also take the threshold values 0.15ppm and 0.17ppm. One of the reasons for choosing these latter values is that we would like to know what would happen if the threshold for declaring emergency alerts in Mexico City was lowered to 0.17ppm. The reason for choosing the threshold 0.15ppm is because it is an intermediate value between the Mexican standard and 0.17ppm.

During the observational period considered here, we have that the mean of the daily observed measurements were 0.12, 0.098, 0.13, 0.12, and 0.14, in regions NE. NW, CE, SE, and SW, respectively, with corresponding standard deviations of 0.06, 0.04, 0.06, 0.05, and 0.06, for those same regions. The threshold 0.11ppm was either reached or exceeded in 4280, 3139, 4921, 4921, and 5711 days in regions NE, NW, CE, SE, and SW, respectively. In those same regions, the threshold 0.15ppm was reached or exceeded in 2460, 963, 2819, 2299, and 3594 days, and the numbers in the case of the threshold 0.17ppm are, 1769, 479, 1896, 1419, and 2660, respectively.

Even though it is a general belief that ozone measurements depend on the measurements of only a few days in the past, we are taking $M = 16$ when we consider the threshold values 0.15ppm and 0.17ppm. We have decided to do that because in previous works the order for homogeneous segments could have higher order. In the case of $L = 0.11$ppm, in some cases, larger values of M were needed. Hence, we also take $M = 16$, in the case of region NW, and we take $M = 18$ in the case of regions CE, NE, SE, and SW. In order to account also for the possibility of low order, we take $\lambda = 1$ in the prior distribution of K.

The hyperparameters of the Dirichlet prior distributions are assigned as in [25]. Therefore, the values of $\alpha_{\overline{m}i}^{(K)}(t)$, $\alpha_{\overline{m}}^{(0)}(t)$, and $\alpha_{\overline{m}}^{(K)}$ will belong to the set $\{3,4,5,6,7,8\}$. Hence, assign $\alpha_{\overline{m}i}^{(K)}(t) = 8$ for the coordinate corresponding to the $\max\{n_{\overline{m}0}^{(K)}(t), n_{\overline{m}1}^{(K)}(t)\}$. Depending on the difference $\max\{n_{\overline{m}0}^{(K)}(t), n_{\overline{m}1}^{(K)}(t)\} - \min\{n_{\overline{m}0}^{(K)}(t), n_{\overline{m}1}^{(K)}(t)\}$, an integer value in $\{3,4,5,6,7\}$ is assigned to the hyperparameter corresponding to $\min\{n_{\overline{m}0}^{(K)}(t), n_{\overline{m}1}^{(K)}(t)\}$. If we have $n_{\overline{m}i}^{(K)}(t) = 0$, then the value 3 is automatically assigned to the corresponding $\alpha_{\overline{m}i}^{(K)}(t)$. Similar procedure is applied in the cases of $\alpha_{\overline{m}}^{(0)}(t)$ and $\alpha_{\overline{m}}^{(K)}$.

Table 1 gives the values of $P(K \mid \mathbf{Y})$. Even though, $\mathcal{S}$ includes the values 0, 1, 2, and 3, since the posterior probabilities at those points are of order 10^{-8} and below, we have omitted those values of K. We use the symbol "-" to indicate that the specific value of K either was not considered in the corresponding region or the probability associated to it was small compared to the values shown.

Looking at Table 1 we may see that, if we consider the threshold $L = 0.11$ppm, then the selected order of the chain is K equal to 16 in the case of region NE, equal to 12 for region NW, and equal to 17 for regions CE and SE. When we consider region SW, the value of K is

	NE			NW			CE			SE			SW		
	0.11	0.15	0.17	0.11	0.15	0.17	0.11	0.15	0.17	0.11	0.15	0.17	0.11	0.15	0.17
$K=4$	–	–	–	–	–	$<10^{-7}$	–	–	–	–	–	–	–	–	–
$K=5$	–	–	–	–	0.33	1	–	–	–	–	–	0.641	–	–	–
$K=6$	–	–	0.02	–	0.67	$<10^{-16}$	–	–	–	–	–	0.359	–	–	–
$K=7$	–	–	0.93	–	–	–	–	–	0.33	–	–	–	–	–	–
$K=8$	–	0.007	0.05	–	–	–	–	–	0.67	–	0.09	–	–	–	0.024
$K=9$	–	0.173	–	–	–	–	–	–	–	–	0.9	–	–	–	0.635
$K=10$	–	0.67	–	–	–	–	–	0.05	–	–	0.01	–	–	–	0.383
$K=11$	–	0.15	–	0.47	–	–	–	0.68	–	–	–	–	–	–	0.003
$K=12$	–	–	–	0.53	–	–	–	0.26	–	–	–	–	–	0.008	–
$K=13$	–	–	–	–	–	–	–	–	–	–	–	–	–	0.095	–
$K=14$	–	–	–	–	–	–	–	–	–	–	–	–	–	0.792	–
$K=15$	–	–	–	–	–	–	–	–	–	–	–	0.012	–	0.105	–
$K=16$	0. 66	–	–	–	–	–	0.01	–	–	0.004	–	–	–	–	–
$K=17$	0.07	–	–	–	–	–	0.99	–	–	0.984	–	–	–	–	–
$K=18$	0.27	–	–	–	–	–	–	–	–	–	–	–	–	–	–

Table 1. Posterior distribution of the order of the chain for all regions and threshold considered. The symbol "-" is used to indicate that the specific value of K either was not considered in the corresponding region or the probability associated to it was small compared to the values shown.

either larger than or equal to 18 with probability one. If we take into account the threshold $L = 0.15$ppm, then, also by looking at Table 1, we have that the chosen orders are 10, 6, 11, 9, and 14, in the cases of regions NE, NW, CE, SE, and SW, respectively. When we consider the threshold $L = 0.17$ppm, then the selected orders are 7, 8, and 9, for regions NE, CE, and SW, respectively. In the cases of regions NW and SE, the estimated order is 5. Therefore, using this information and (4), the corresponding transition and initial probabilities may then be calculated.

As an example, consider the case of region CE and the threshold 0.17ppm. In that case, we have that the order of the chain is $K = 8$. Therefore, $S_2^{(K)} = \{0, 1, \ldots, 255\}$. In Table 2, we have the approximated estimated values of the initial distribution $Q_{\overline{m}}^{(K)}(1)$, and of the transition probabilities $P_{\overline{m}0}^{(K)}(t)$, $t = 1, 2$. (We have truncated the values and the total sum is approximately one.) We use the notation $\overline{m}' - \overline{m}''$, to indicate that for all values of $\overline{m}$ in $\{\overline{m}', \overline{m}' + 1, \ldots, \overline{m}''\}$, the estimated probabilities are equal to the values shown.

Looking at Table 2, it is possible to see that the highest initial probability is that associated to the state $\overline{0}$, i.e., the first eight days of the year form a string of zeros, meaning that the concentration levels are below 0.17ppm. Additionally, once you have the information that the ozone concentration levels on the first eight days are below 0.17ppm, then the highest transition probability is also associated to the transition to zeros, i.e., the two days following the eight initial days with concentration below 0.17ppm are more likely to present lower concentration levels as well.

In order to illustrate the type of information that may be obtained using the methodology considered here, take the case of the year 2012 and region CE. Suppose we want to calculate the probability that during the first nine days of January we have that the ozone concentration is below 0.17ppm from the first eight days, and it is above it on the ninth. Therefore, we want to know the probability that (0, 0, 0, 0, 0, 0, 0, 0) is followed by one. Hence, we want the probability of having the following sequence of zeros and ones: 0, 0, 0, 0, 0, 0, 0, 0, 1. Therefore,

$$P((0,0,0,0,0,0,0,0,1)) = P_{\overline{0}1}^{(8)}(1) \times Q_{\overline{0}}^{(8)}(1) = 0.238 \times 0.0327 \approx 0.008.$$

$\overline{m}$	$Q^{(8)}_{\overline{m}}(1)$	$\overline{m}$	$P^{(8)}_{\overline{m}0}(1)$	$P^{(8)}_{\overline{m}0}(2)$
$\overline{0}$	0.0327	$\overline{0}$	0.762	0.889
$\overline{1}-\overline{15}$	0.0036	$\overline{1}-\overline{11}$	0.5	0.5
$\overline{16}$	0.0073	$\overline{12}$	0.5	0.286
$\overline{17}-\overline{23}$	0.0036	$\overline{13}-\overline{62}$	0.5	0.5
$\overline{24}$	0.0073	$\overline{63}$	0.286	0.8
$\overline{25}-\overline{27}$	0.0036	$\overline{64}-\overline{127}$	0.5	0.5
$\overline{28}$	0.0073	$\overline{128}$	0.5	0.444
$\overline{29}-\overline{31}$	0.0036	$\overline{129}-\overline{158}$	0.5	0.5
$\overline{32}$	0.0073	$\overline{159}$	0.5	0.286
$\overline{33}-\overline{62}$	0.0036	$\overline{160}-\overline{247}$	0.5	0.5
$\overline{63}$	0.0073	$\overline{248}$	0.286	0.5
$\overline{63}-\overline{125}$	0.0036	$\overline{249}$	0.5	0.5
$\overline{126}$	0.0073	$\overline{250}$	0.286	0.5
$\overline{127}-\overline{191}$	0.0036	$\overline{251}$	0.5	0.6
$\overline{192}$	0.0073	$\overline{252}-\overline{253}$	0.5	0.286
$\overline{193}-\overline{241}$	0.0036	$\overline{254}-\overline{255}$	0.5	0.5
$\overline{242}$	0.0073	–	–	–
$\overline{243}$	0.0036	–	–	–
$\overline{244}$	0.0073	–	–	–
$\overline{245}-\overline{247}$	0.0036	–	–	–
$\overline{248}$	0.0073	–	–	–
$\overline{249}$	0.0036	–	–	–
$\overline{250}$	0.0073	–	–	–
$\overline{251}-\overline{255}$	0.0036	–	–	–

Table 2. Transition probabilities $P^{(8)}_{\overline{m}0}(1)$ and $P^{(8)}_{\overline{m}0}(2)$ as well as the initial probabilities $Q^{(8)}_{\overline{m}}(1)$, for all values of $\overline{m} \in S_2^{(K)}$ in the case of region CE and threshold 0.17ppm. The notation $\overline{m}' - \overline{m}''$ is used to indicate that for all values of $\overline{m}$ in $\{\overline{m}', \overline{m}'+1, \ldots, \overline{m}''\}$, the estimated probabilities are equal to the values shown.

In order to obtain the values of the probabilities of interest, recall that $P^{(K)}_{\overline{m}0}(t) = 1 - P^{(K)}_{\overline{m}1}(t)$, $\overline{m} \in S_2^{(K)}$, $t = 1, 2, \ldots, T-K$. Therefore, looking at Table 2, we have that $P^{(8)}_{\overline{01}}(1)$ is one minus the value on the column corresponding to $P^{(K)}_{\overline{m}0}(1)$ with $\overline{m} = \overline{0}$. Similar comment is valid in the case of $Q^{(8)}_{\overline{m}}(1)$, $\overline{m} \in S_2^{(K)}$.

Suppose now that we want to know the probability of having (0, 0, 0, 0, 0, 0, 0, 0) followed by (1, 1). Hence, we want to know what the probability that (0, 0, 0, 0, 0, 0, 0, 0) is followed by one and that (0, 0, 0, 0, 0, 0, 0, 1) is followed by one. Therefore, we need to calculate

$$P((0,0,0,0,0,0,0,0,1,1)) = P^{(8)}_{\overline{255}1}(2) \times P^{(8)}_{\overline{01}}(1) \times Q^{(8)}_{\overline{0}}(1) = 0.5 \times 0.238 \times 0.0327 \approx 0.0004.$$

Proceeding in this way we may calculate the probability of having any string of states at any time of the year.

If we compare to the actual measurements in the year 2012, then we have that in the first ten days, the sequence $\mathbf{Y}$, in the case of region CE, has the configuration 0, 0, 0, 0, 0, 0, 0, 0, 0, 0. In fact, the estimated probability of that sequence of zeros and ones is $0.5 \times 0.762 \times 0.0327 \approx 0.0125$ which is three times higher than the probability of having (0, 0, 0, 0, 0, 0, 0, 0) followed by (1, 1). If we consider also the year 2013, the results are similar. Hence, the methodology used here can produce estimated values that may describe well the behaviour of the data.

5. Conclusion

In this work we have considered a non-homogeneous Markov chain model to study the ozone's behaviour in Mexico City. The interest resides in estimating the probability that the ozone level will be above (below) a certain threshold given that it is either above or below it in the present and in the recent past. Due to the nature of the questions asked here, a natural way of trying to answer them is to use Markov chain models. However, due to the non-homogeneity of the data, a non-homogeneous version of the chain is used.

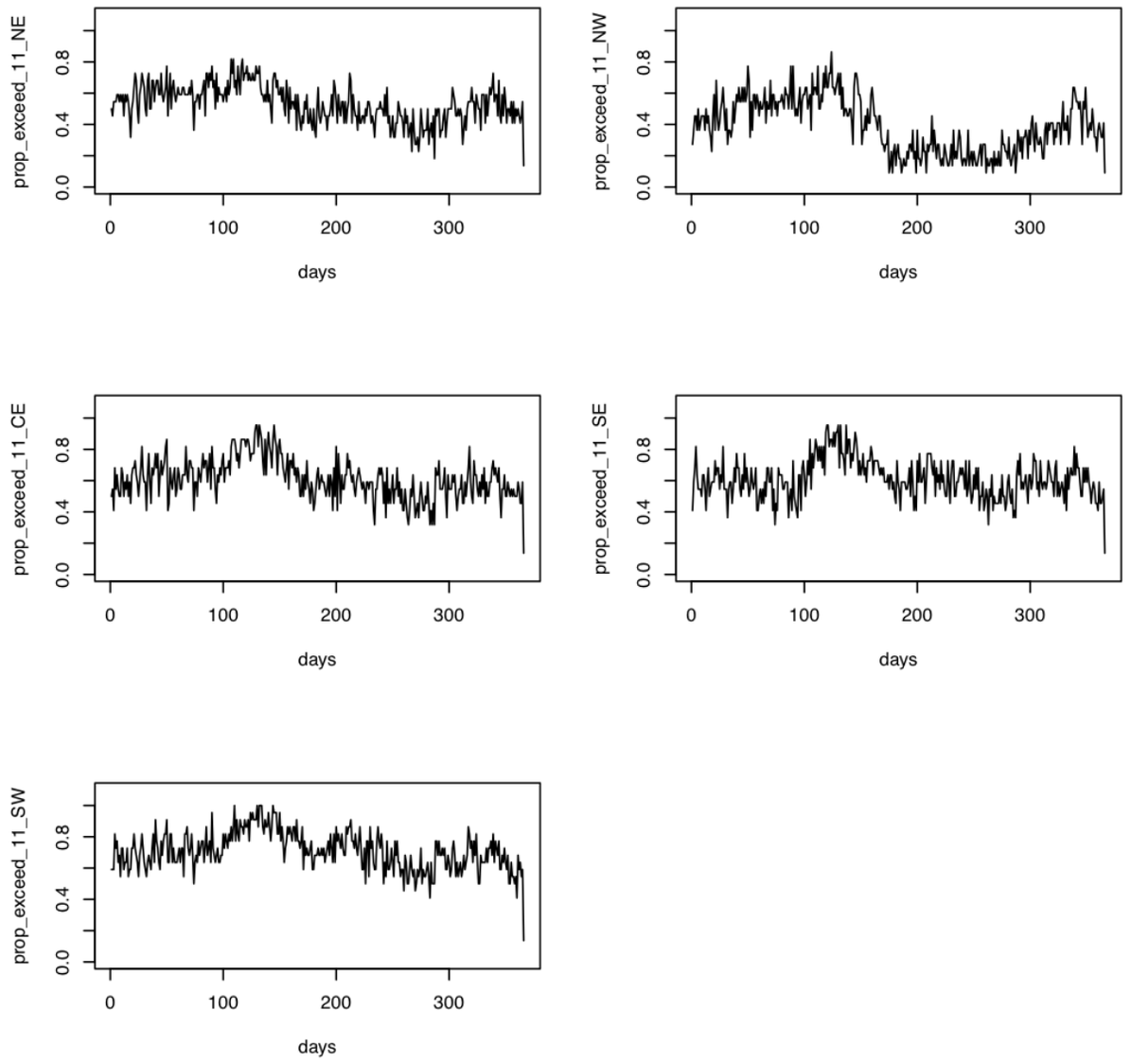

Figure 1. Proportion of years in which, for a given day, the threshold 0.11ppm was exceeded by the ozone concentration.

Using the Bayesian approach a maximum *à posteriori* estimation of the order of the matrix as well as its transition matrix and initial distribution was made. The results have shown that higher order should be considered for the chain. One explanation for that could be the

way the empirical probability of having an exceedance in a given day behaves. That can be seen in Figure 1 where, as an illustration, the proportion of exceedances of the threshold 0.11ppm is presented for each region. The values correspond to the proportion of years in which in a fixed day the threshold was exceeded. As we vary the days, we have the behaviour throughout the 366 days. In that figure we have in the horizontal axis the days and in the vertical axis we have the values of $prop(t) = (1/N)\sum_{i=1}^{N} Y_t^{(i)}$, which represents the proportion of years with an exceedance in the tth day, $t = 1, 2, \ldots, T$. The notation $Y_t^{(i)}$ is used to indicate the variable Y_t defined in (1) on the ith year.

The plots in Figure 1 reflect well the fact that in region SW, in most of the days of the year, there are exceedances of the threshold 0.11. We may also see the influence of the seasons of the year. The hill between days 100 and 200 appearing in every plot, corresponds to measurements taken between April and June. Higher values occur during the days corresponding to approximately mid April to mid May. Those months are in the middle of Spring. During this season it does rain much in Mexico City. Additionally, there is a lot of sunlight. Hence, the ozone concentration is bound to be high, and as a consequence, the proportion of years in which exceedances occur at that period is large. The values decrease when the raining season starts (around the beginning of June).

If we consider the threshold values 0.15ppm and 0.17ppm, the behaviour of the proportion of years where in a given day exceedances occurred is similar to the case of 0.11ppm. The difference is that the values of the proportions are smaller. It is possible to see that the proportion of exceedances may vary according to the seasons of the year, and that, within a given season, changes are, in general, not drastic. Therefore, it is possible that measurements from more than a few days may have an influence on the behaviour of future measurements, and with that, make the estimation method considered here to produce high values for the order of the chain.

Acknowledgements

The authors thank the Editor for sending comments that helped to improve the presentation of the results. The authors also thank Peter Guttorp for providing a copy of his works related to applications of non-homogeneous Markov chains. ERR and MHT were partially funded by the project PAPIIT-IN102713-3 of the Dirección General de Apoyo al Personal Académico de la Universidad Nacional Autónoma de México (DGAPA-UNAM), Mexico. ERR also received funds from a sabbatical year grant from DGAPA-UNAM. ERR is grateful to the Departments of Statistics of the Universidade Estadual Paulista "Júlio de Mesquita Filho" – Campus Presidente Prudente, Brazil, and of the University of Oxford, UK, where parts of this work were developed, for all the support and hospitality received during her stay at those departments.

Appendix

In this Appendix we present the code of the programme in R used to estimate the order of the non-homogeneous Markov chain as well as its transition probabilities. The code it not optimal and can be highly improved, but in its present form it provides elements for estimating the necessary quantities.

```
# ESTIMATING THE ORDER OF THE CHAIN
ozonio_sw=read.table('data.txt', header=T)
attach(ozonio_sw)
anos=NCOL(ozonio_sw)
dias=NROW(ozonio_sw)
# assigning the value of M
M = 15 # for instance
# initialisation of the matrices to store the values of the likelihood
# for each order and day
term_init_like <- matrix(0, M+1, 1)
term_general_like <- matrix(0, dias, M+1)
#######
# Case K = 0
#######
# counting the numbers of ones and zeros in each row
count_init_0 <- matrix(0, dias, 2)
for(i in 1:dias){
for(j in 1:anos){
if(ozonio_sw[i,j] == 0){count_init_0[i,1] = count_init_0[i,1] + 1} }
count_init_0[i,2] = anos - count_init_0[i,1]}
# assigning the values of alpha
alpha_init_0 <- matrix(0,dias,2)
for(i in 1:dias){
for(j in 1:2){if(count_init_0[i,j] == 0){alpha_init_0[i,j] = 3} }
if ((count_init_0[i,1] == min(count_init_0[i,])) & (count_init_0[i,1] != 0)){
if(alpha_init_0[i,1] == 0){alpha_init_0[i,1] = 4
if(alpha_init_0[i,2] == 0){alpha_init_0[i,2] = 8}} }
if ((count_init_0[i,1] == max(count_init_0[i,])) & (count_init_0[i,2] != 0)){
if(alpha_init_0[i,1] == 0){alpha_init_0[i,1] = 8
if(alpha_init_0[i,2] == 0){alpha_init_0[i,2] = 4}} }
if ((count_init_0[i,1] == min(count_init_0[i,])) & (count_init_0[i,1] == 0)){
if(alpha_init_0[i,2] == 0){alpha_init_0[i,2] = 8} }
if ((count_init_0[i,1] == max(count_init_0[i,])) & (count_init_0[i,2] == 0)){
if(alpha_init_0[i,1] == 0){alpha_init_0[i,1] = 8} }
}
# calculating the value of the likelihood L(Y | K = 0)
prod_1_k0 <- matrix(0,dias,1)
prod_2_k0 <- matrix(0,dias,1)
prod_k0 <- matrix(0,dias,1)
for(i in 1:dias){
for(j in 1:2){
prod_1_k0[i] = (gamma(count_init_0[i,j]+alpha_init_0[i,j])/gamma(alpha_init_0[i,j])) }
prod_2_k0[i] = (gamma(sum(alpha_init_0[i,]))/gamma(sum(count_init_0[i,] +
 alpha_init_0[i,])))
prod_k0 [i] = prod_1_k0[i]*prod_2_k0[i]}
for(i in 1:dias){
term_general_like[i, 1] = prod_k0[i]}
#######
# Caso K = 1
#######
# the initial states in this case is the first row of the case K=0
# assigning the values of count_init_0[1,i] to count_init_1[i]
count_init_1 <- matrix(0,1,2)
for(j in 1:2){count_init_1[j]= count_init_0[1,j]}
# assigning the same values of alpha_init_0[1,i] to alpha_init_1[i]
alpha_init_1 <- matrix(0,1,2)
for(j in 1:2){alpha_init_1[j]= alpha_init_0[1,j]}
##############
# counting the number of transitions 0 -> 0, 0 -> 1, 1 -> 0 and 1 -> 1
# count_0_k1[i,1] counts the number of transitions from zero to zero in row i,
```

```
# count_0_k1[i,2] counts the number of transitions from zero to one in row i,
#count_1_k1[i,1] counts the number of transitions from one to zero in row i
# count_1_k1[i,2] counts the number of transitions from ne to one in row i
##############
count_0_k1 <- matrix(0, dias-1, 2)
count_1_k1 <- matrix(0, dias-1, 2)
for(i in 1:(dias-1)){
for(j in 1:anos){
if((ozonio_sw[i, j] == 0) & (ozonio_sw[i+1,j] == 0))
{count_0_k1[i,1] = count_0_k1[i,1] + 1}
if((ozonio_sw[i, j] == 0) & (ozonio_sw[i+1,j] == 1))
{count_0_k1[i,2] = count_0_k1[i,2] + 1}
if((ozonio_sw[i, j] == 1) & (ozonio_sw[i+1,j] == 0))
{count_1_k1[i,1] = count_1_k1[i,1] + 1}
if((ozonio_sw[i, j] == 1) & (ozonio_sw[i+1,j] == 1))
{count_1_k1[i,2] = count_1_k1[i,2] + 1}}}
# assigning the values of the values of alpha
# alpha_0_k1 is associated with the transitions 0 -> 0 and 0 -> 1
# alpha_1_k1 is associated with the transitions 1 -> 0 and 1 -> 1
alpha_0_k1 <- matrix(0, dias-1, 2)
alpha_1_k1 <- matrix(0, dias-1, 2)
for(i in 1:(dias-1)){
for(j in 1:2){
if(count_0_k1[i,j] == 0){alpha_0_k1[i,j] = 3}
if(count_1_k1[i,j] == 0){alpha_1_k1[i,j] = 3} }
if((count_0_k1[i,1] == count_0_k1[i,2]) & (count_0_k1[i,1] != 0) &
(count_0_k1[i,1] < 5) & (alpha_0_k1[i, 1] == 0)){alpha_0_k1[i,1] = 5
if(alpha_0_k1[i,2] == 0){alpha_0_k1[i,2] = 5} }
if((count_1_k1[i,1] == count_1_k1[i,2]) & (count_1_k1[i,1] != 0) &
(count_1_k1[i,1] < 5) & (alpha_1_k1[i, 1] == 0)){alpha_1_k1[i,1] = 5
if(alpha_1_k1[i,2] == 0){alpha_1_k1[i,2] = 5} }
if((count_0_k1[i,1] == count_0_k1[i,2]) & (count_0_k1[i,1] != 0) &
(count_0_k1[i,1] >= 5) & (alpha_0_k1[i, 1] == 0)){alpha_0_k1[i,1] = 7
if(alpha_0_k1[i,2] == 0){alpha_0_k1[i,2] = 7} }
if((count_1_k1[i,1] == count_1_k1[i,2]) & (count_1_k1[i,1] != 0) &
(count_1_k1[i,1] >= 5) & (alpha_1_k1[i, 1] == 0)){
alpha_1_k1[i,1] = 7
if(alpha_1_k1[i,2] == 0){alpha_1_k1[i,2] = 7} }
if((count_0_k1[i,1] == min(count_0_k1[i,])) & (count_0_k1[i,1] != 0)){
if(alpha_0_k1[i,1] == 0){alpha_0_k1[i,1] = 4}
if(alpha_0_k1[i,2] == 0){alpha_0_k1[i,2] = 8} }
if((count_1_k1[i,1] == min(count_1_k1[i,])) & (count_1_k1[i,1] != 0)){
if(alpha_1_k1[i,1] == 0){alpha_1_k1[i,1] = 4}
if(alpha_1_k1[i,2] == 0){alpha_1_k1[i,2] = 8}}
if((count_0_k1[i,1] == min(count_0_k1[i,])) & (count_0_k1[i,1] == 0)){
if(alpha_0_k1[i,2] == 0){alpha_0_k1[i,2] = 8} }
if((count_1_k1[i,1] == min(count_1_k1[i,])) & (count_1_k1[i,1] == 0)){
if(alpha_1_k1[i,2] == 0){alpha_1_k1[i,2] = 8} }
if((count_0_k1[i,1] == max(count_0_k1[i,])) & (count_0_k1[i,2] != 0)){
if(alpha_0_k1[i,1] == 0){alpha_0_k1[i,1] = 8}
if(alpha_0_k1[i,2] == 0){alpha_0_k1[i,2] = 4} }
if((count_1_k1[i,1] == max(count_1_k1[i,])) & (count_1_k1[i,2] != 0)){
if(alpha_1_k1[i,1] == 0){alpha_1_k1[i,1] = 8}
if(alpha_1_k1[i,2] == 0){alpha_1_k1[i,2] = 4} }
if((count_0_k1[i,1] == max(count_0_k1[i,])) & (count_0_k1[i,2] == 0)){
if(alpha_0_k1[i,1] == 0){alpha_0_k1[i,1] = 8} }
if((count_1_k1[i,1] == max(count_1_k1[i,])) & (count_1_k1[i,2] == 0)){
if(alpha_1_k1[i,1] == 0){alpha_1_k1[i,1] = 8} }
}
# calculating the value of the likelihood L(Y | K = 1)
# term corresponding to the initial distribution
```

```
prod_1_k1 = 1
for(j in 1:2){
prod_1_k1 = prod_1_k1*((gamma(count_init_1[j]+alpha_init_1[j])/
gamma(alpha_init_1[j])))*((gamma(sum(alpha_init_1))/
gamma(sum(count_init_1+ alpha_init_1))))}
term_init_like[2] = prod_1_k1
# term corresponding to the rest of the days (rows)
prod_2_k1 <- matrix(0,(dias - 1),1)
prod_3_k1 <- matrix(0,(dias - 1),1)
prod_k1 <- matrix(0,(dias - 1),1)
for(i in 1:(dias - 1)){
for(j in 1:2){
prod_2_k1[i] = (gamma(count_0_k1[i,j] + alpha_0_k1[i,j])/gamma(alpha_0_k1[i,j]))*
(gamma(count_1_k1[i,j] + alpha_1_k1[i,j])/gamma(alpha_1_k1[i,j]))}
for(j in 1:2){
prod_3_k1[i] = (gamma(sum(alpha_0_k1[i,])+1)*gamma(sum(alpha_1_k1[i,])))/
(gamma(sum(count_0_k1[i,]+alpha_0_k1[i,])+sum(count_1_k1[i,]+alpha_1_k1[i,])+1))}
prod_k1[i] = prod_2_k1[i] * prod_3_k1[i]}
for(i in 1:(dias-1)){
term_general_like[i, 2] = prod_k1[i]}
####
# Case K >=2
#####
# Transforming vector of zeros and ones in an element of S_{2}^{(K)}
# counting the initial values of the chain (K >= 2) initialisation of the
# matrices to store the values of the likelihood for each order and day
##########
prod_init <- matrix(0, 1, M)
for(K in 2:M){
count_init_k <- matrix(0, 1, (2^K))
mbase_init <- matrix(0, 1, anos)
alpha_init_k <- matrix(0, 1, 2^K)
# transforming the k-dimensional initial vector into an integer number
for(j in 1:anos){
for(l in 0:(K-1)){
mbase_init[j] = mbase_init[j] + ozonio_sw[1+l,j]*2^l}
# counting the number of m in the initial state
for(n in 0:(2^K-1)){if(mbase_init[j] == n){
count_init_k[mbase_init[j]+1] = count_init_k[mbase_init[j]+1] + 1}}}
# assigning the respective values of alpha_init_k
for(n in 0:(2^K - 1)){
if(count_init_k[n+1] == 0){
alpha_init_k[n+1] = 3}
if((count_init_k[n+1] == min(count_init_k)) & (alpha_init_k[n+1] == 0)){
alpha_init_k[n+1] = 3}
if((count_init_k[n+1] == max(count_init_k)) & (alpha_init_k[n+1] == 0)){
alpha_init_k[n+1] = 8}
if((abs(count_init_k[n+1]-max(count_init_k)) > 5) & (alpha_init_k[n+1] == 0)){
alpha_init_k[n+1] = 4}
if((abs(count_init_k[n+1]-max(count_init_k)) <= 5) & (alpha_init_k[n+1] == 0))
{alpha_init_k[n+1] = 6}}
# calculation of the first term in the product in the likelihood L(Y | K)
prod_1_init = 1
prod_2_init = 1
prod_init_k = 0
for(n in 1:(2^K)){
prod_1_init = prod_1_init*(gamma(count_init_k[n]+alpha_init_k[n])
/gamma(alpha_init_k[n]))}
prod_2_init = 1
sumalphacount = sum(alpha_init_k+count_init_k)
sumalpha = sum(alpha_init_k)
```

```
limsup = sum(alpha_init_k+count_init_k) - sum(alpha_init_k)
for(k in 0:(limsup-1)){prod_2_init = prod_2_init*((sumalphacount-1)-k)^(-1)}
prod_init_k = prod_1_init*prod_2_init
term_init_like[K+1] = prod_init_k
# initialisation of the matrices of interest in the case of i not the initial state
mbase <- matrix(0, dias - K, anos) # that is overline{m}
count_Km <- matrix(0, 2^K, 2) # matrix counting the transitions from m
s <- matrix(0, (dias-K), 2^K)
# transforming the vectors of length K into a number in the base 2 for
# each day for all years
for(i in 1:(dias-K)){
for(j in 1:anos){
for(l in 0:(K-1)){
mbase[i,j] = mbase[i,j] + ozonio_sw[i+l,j]*2^l}
# storing the number of mbase in row i in the vector s[day, mbase]
for(n in 0:(2^K-1)){
if(mbase[i,j] == n){s[i,n+1] = s[i, n+1] + 1} }
} # closes the j loop
} # closes the i loop
# counting the number of transitions for each day (day is kept fixed while
# counting goes through years)
# count_Km[m,1] counts the number of transitions m -> 0 in the 22 years for fixed i
# count_Km[m,2] counts the number of transitions m -> 1 in the 22 years for fixed i
# n_m0[dias,m] counts the number of transitions m -> 0 in the 22 years for each i
# n_m1[dias,m] counts the number of transitions m -> 1 in the 22 years for each i
n_m0 <- matrix(0, (dias - K), 2^K) # matrix counting m -> 0
n_m1 <- matrix(0, (dias - K), 2^K) # matrix counting m -> 1
for(i in 1:(dias-K)){
for(j in 1:anos){
if(ozonio_sw[i+K,j] == 0){
for(n in 0:(2^K-1)){
if(mbase[i,j] == n){
count_Km[n+1,1] = count_Km[n+1,1]+1}}}
if(ozonio_sw[i+K,j] == 1){
for(n in 0:(2^K-1)){if(mbase[i,j] == n){
count_Km[n+1,2] = count_Km[n+1,2]+1}}}  }
for(m in 1:(2^K)){
n_m0[i,m] = count_Km[m,1]
n_m1[i,m] = count_Km[m,2] }
count_Km <- matrix(0, 2^K, 2)
} # closes de i loop
# assignation of the values of the corresponding values of
# alpha the hyperparameter of the Dirichlet prior distribution
# alpha_m0 is associated to the transitions m -> 0 for each day and each m
# alpha_m1 is associated to the transitions m -> 1 for each day and each m
alpha_m0 <- matrix(0, (dias - K), 2^K)
alpha_m1 <- matrix(0, (dias - K), 2^K)
for(i in 1:(dias-K)){
for(m in 1:2^K){
if(n_m0[i,m] == 0){alpha_m0[i,m] = 3}
if(n_m1[i,m] == 0){alpha_m1[i,m] = 3}
} #closes the m loop
for(m in 1:2^K){
if((n_m0[i,m] == min(n_m0[i,m], n_m1[i,m])) & (alpha_m0[i,m] == 0)){
alpha_m0[i,m] = 4
if((abs(n_m0[i,m] - n_m1[i,m]) >= 5) & (alpha_m1[i,m] == 0)){alpha_m1[i,m] = 7}
if((abs(n_m0[i,m] - n_m1[i,m]) < 5) & (alpha_m1[i,m] == 0)){alpha_m1[i,m] = 5}}
if((n_m0[i,m] == min(n_m0[i,m], n_m1[i,m])) & (alpha_m0[i,m] != 0)){
if((abs(n_m0[i,m] - n_m1[i,m]) >= 5) & (alpha_m1[i,m] == 0)){alpha_m1[i,m] = 7}
if((abs(n_m0[i,m] - n_m1[i,m]) < 5) & (alpha_m1[i,m] == 0)){alpha_m1[i,m] = 5} }
if((n_m0[i,m] == max(n_m0[i,m], n_m1[i,m])) & (alpha_m0[i,m] == 0))
```

```
{ alpha_m0[i,m] = 8
if((abs(n_m0[i,m] - n_m1[i,m]) >= 5) & (alpha_m1[i,m] == 0)){alpha_m1[i,m] = 4}
if((abs(n_m0[i,m] - n_m1[i,m]) < 5) & (alpha_m1[i,m] == 0)){alpha_m1[i,m] = 7} }
if((n_m0[i,m] == min(n_m0[i,m], n_m1[i,m])) & (alpha_m0[i,m] != 0)){
if((abs(n_m0[i,m] - n_m1[i,m]) >= 5) & (alpha_m1[i,m] == 0)){alpha_m1[i,m] = 4}
if((abs(n_m0[i,m] - n_m1[i,m]) < 5) & (alpha_m1[i,m] == 0)){alpha_m1[i,m] = 7} }
} #closes the second m loop
} #closes the i loop
# calculation of the likelihood L(Y | K)
prod_1_km <- matrix(0, (dias - K), 1)
prod_2_km <- matrix(0, (dias - K), 1)
prod_km <- matrix(0, (dias-K), 1)
for(i in 1:(dias-K)){
for(n in 1:2^K){
prod_1_km[i] = (gamma(n_m0[i,m]+alpha_m0[i,m])*gamma(n_m1[i,m]+alpha_m1[i,m]))/
(gamma(alpha_m0[i,m])*gamma(alpha_m1[i,m]))
prod_2_km[i] = gamma(alpha_m0[i,m] + alpha_m1[i,m])/gamma(alpha_m0[i,m] +
n_m0[i,m] + alpha_m1[i,m] + n_m1[i,m])
} #closed the m loop
prod_km[i] = prod_1_km[i]*prod_2_km[i]
} # closes the i loop
for(i in 1:(dias-K)){
term_general_like[i,K+1] = prod_km[i]}
} #close the K loop
write.csv(term_general_like, file = "results-file-1.csv")
write.csv(term_init_like, file = "results-file-2.txt")
#
# TRANSITION PROBABILITIES - CASE OF REGION CE, THRESHOLD 0.17
#
K = 8 # may change for other regions and thresholds
# calculation of the normalising constant in the case of the initial distribution
somainit = 0
for(m in 1:2^K){
somainit = somainit + (alpha_init_k[m] + count_init_k[m] - 1)}
# calculation of the initial distribution
prob_init <- matrix(0, 2^K, 1)
for(m in 1: 2^K){
prob_init[m] = (alpha_init_k[m] + count_init_k[m] - 1)/somainit}
write.csv(prob_init, file = "prob_init_chain.txt")
# limiting the number of days
dd = 2 # because I need p_{mj}(1) and p_{mj}(2)
# alpha_m1[i,m] where i is day and m is the value of the vector
# the m1 indicates that m is followed by 1
p_m0 <- matrix(0, dd, 2^K)
p_m1 <- matrix(0, dd, 2^K)
# calculation of the transition probabilities m -> 0 and m -> 1
for(m in 1:2^K){
for(i in 1:dd){
p_m0[i,m] = (alpha_m0[i,m] + n_m0[i,m] - 1)/((alpha_m1[i,m] + n_m1[i,m] - 1)
+(alpha_m0[i,m] + n_m0[i,m] - 1))
p_m1[i,m] = (alpha_m1[i,m] + n_m1[i,m] - 1)/((alpha_m1[i,m] + n_m1[i,m] - 1)
+(alpha_m0[i,m] + n_m0[i,m] - 1)) }}
trans_mat <- matrix(0, 2^K, dd)
for(m in 1:2^K){
for(i in 1:dd){
trans_mat[m,i] = p_m0[i,m]}}
write.csv(trans_mat, file = "trans_mat.txt")
```

Author details

Eliane R. Rodrigues[1*], Mario H. Tarumoto[2] and Guadalupe Tzintzun[3]

1 Instituto de Matemáticas, Universidad Nacional Autónoma de México, Mexico

2 Faculdade de Ciênicas e Tecnologia, Universidade Estadual Paulista Júlio de Mesquita Filho, Brazil

3 Instituto Nacional de Ecología y Cambio Climático, Secretaría de Medio Ambiente y Recursos Naturales, Mexico

*Address all correspondence to: eliane@math.unam.mx

References

[1] Bell ML, McDermontt A, Zeger SL, Samet JM, Dominici F. Ozone and short-term mortality in 95 US urban communities, 1987-2000. Journal of the American Medical Society 2004; 292: 2372-2378.

[2] Bell ML, Peng R, Dominici F. The exposure-response curve for ozone and risk of mortality and the adequacy of current ozone regulations. Environmental Health Perspectives 2005; 114: 532-536.

[3] Cifuentes L, Borja-Arbuto VH, Gouveia N, Thurston G, Davis DL. Assessing the health benefits of urban air pollution reduction associated with climate change mitigation (2000-2020): Santiago, São Paulo, Mexico City and New York City. Environmental Health Perspectives 2001; 109: 419-425.

[4] Dockery DW, Schwartz J, Spengler JD. Air pollution and daily mortality: association with particulates and acid aerosols. Environmental Research 1992; 59: 362-373.

[5] Galizia A, Kinney PL. Long-term residence in areas of high ozone: association with respiratory health in a nationwide sample of nonsmoking adults. Environmental Health 1999; 99: 675-679.

[6] Gauderman WJ, Avol E, Gililand F, Vora H, Thomas D, Berhane K, McConnel R, Kuenzli N, Lurmman F, Rappaport E, Margolis H, Bates D, Peter J. The effects of air pollution on lung development from 10 to 18 years of age. The New England Journal of Medicine 2004; 351: 1057-1067.

[7] 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.

[8] Loomis D, Borja-Arbuto VH, Bangdiwala SI, Shy CM. Ozone exposure and daily mortality in Mexico City: a time series analysis. Health Effects Institute Research Report 1996; 75: 1-46.

[9] Martins LC, de Oliveira Latorre MRD, Saldiva PHN, Braga ALF. Air pollution and emergency rooms visit due to chronic lower respiratory diseases in the elderly: an ecological time series study in São Paulo, Brazil. J. Occupational and Environmental Medicine 2002; 44: 622-627.

[10] WHO. Air Quality Guidelines-2005, Particulate Matter, Ozone, Ditrogen dioxide and Sulfur Dioxide. European Union: World Health Organization Regional Office for Europe; 2006.

[11] Horowitz J. Extreme values from a nonstationary stochastic process: an application to air quality analysis. Technometrics 1980; 22: 469-482.

[12] Smith RL. Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone. Statistical Sciences 1989; 4: 367-393.

[13] Raftery AE. Are ozone exceedance rate decreasing?. Comment of the paper *Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone by R. L. Smith*. Statistical Sciences 1989; 4: 378–381.

[14] Flaum JB, Rao ST, Zurbenko IG. Moderating Influence of Meteorological Conditions on Ambient Ozone Concentrations. Journal of the Air and Waste Management Association 1996; 46: 33-46.

[15] Achcar JA, Rodrigues ER, Tzintzun G. Using stochastic volatility models to analyse weekly ozone averages in Mexico City. Environmental and Ecological Statistics 2011; 18: 271-290.

[16] Javits JS. Statistical interdependencies in the ozone national ambient air quality standard. Journal of Air Pollution Control Association 1980; 30: 58-59.

[17] 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.

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

[19] Villaseñor-Alva JA, González-Estrada E. On modelling cluster maxima with applications to ozone data from Mexico City. Environmetrics 2010; 21: 528-540.

[20] Barrios JM, Rodrigues ER. A queueing model to study occurrence and duration of ozone exceedances in Mexico City. Journal of Applied Statistics 2014. dx.doi.org/101080/02664763.2014.939613

[21] Rajagopalan B, Upmanu L, Tarboton DG. Non-homogeneous Markov model for daily precipitation. Journal of Hydrology Engineering 1996; 1: 33-40.

[22] Hughes JP, Guttorp P, Charles SP. A non-homogeneous hidden Markov model for precipitation occurrence. Applied Statistics, Part 1 1999; 48: 16-30.

[23] Drton M, Marzban C, Guttorp P, Schafer JT. A Markov chain model for tornadic activity. American Meteorological Society 2003; 131: 2941-2953.

[24] Larsen LC, Bradley RA, Honcoop GL. A new method of characterizing the variability of air quality-related indicators. Air and Waste Management Association. International Specialty Conference, Tropospheric Ozone and the Environment. Pittsburgh, USA: California Air and Waste Management Series; 1990.

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

[26] Álvarez LJ, Rodrigues ER. A trans-dimensional MCMC algorithm to estimate the order of a Markov chain: an application to ozone peaks in Mexico City. International Journal of Pure and Applied Mathematics 2008; 48: 315-331.

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

[28] Rice JA. Mathematical statistics with data analysis. New York, USA: Wadsworth and Brook; 1988.

[29] Carlin BP, Louis TA. Bayes and Empirical Bayes Methods for Data Analysis. Second Edition. USA: Chapman and Hall/CRC; 2000.

[30] Green PJ. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika 1995; 82: 711-732.

[31] Carlin BP, Chib S. Bayesian model choice via Markov chain Monte Carlo methods. Journal of the Royal Statistical Society Series B 1995; 57: 473-484.

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

[33] Fleming TR, Harrington DP. Estimation for discrete time non-homogeneous Markov chains. Stochastic Processes and their Applications 1978; 7: 131-139.

[34] Aalem OO, Johansen S. An empirical transition matrix for non-homogeneous Markov chains based on censored observation. Scandinavian Journal of Statistics 1978; 5: 141-150.

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

[36] Fan T-H, Tsai C-A. A Bayesian method in determining the order of a finite Markov chain. Communications in Statistics – Theory and Methods 1999; 28: 1711-1730.

[37] Evans M, Swartz T. Approximating Integrals via Monte Carlo and Deterministic Methods. Oxford Statistical Series 20. Oxford, UK: Oxford University Press; 2000.

[38] NOM. Modificación a la Norma Oficial Mexicana NOM-020-SSA1-1993. Diario Oficial de la Federación. 30 Octubre 2002. Mexico: 2002. (In Spanish.)