Scientific Research Publishing

Journal of Environmental Protection, 2023, 14, 561-582
https://www.scirp.org/journal/jep
ISSN Online: 2152-2219
ISSN Print: 2152-2197

Air Quality Estimation Using Nonhomogeneous Markov Chains: A Case Study Comparing Two Rules Applied to Mexico City Data

Eliane R. Rodrigues[1]*, Juan A. Cruz-Juárez[2], Hortensia J. Reyes-Cervantes[2], Guadalupe Tzintzun[3]

[1]Instituto de Matemáticas, Universidad Nacional Autónoma de México, Mexico City, Mexico
[2]Facultad de Ciencias Fsico-Matemáticas, Benemérita Universidad Autónoma de Puebla, Puebla, Mexico
[3]Instituto Nacional de Ecologa y Cambio Climático, Secretara de Medio Ambiente y Recursos Naturales, Mexico City, Mexico
Email: *eliane@math.unam.mx

How to cite this paper: Rodrigues, E.R., Cruz-Juárez, J.A., Reyes-Cervantes, H.J. and Tzintzun, G. (2023) Air Quality Estimation Using Nonhomogeneous Markov Chains: A Case Study Comparing Two Rules Applied to Mexico City Data. *Journal of Environmental Protection*, **14**, 561-582.
https://doi.org/10.4236/jep.2023.147033

Received: June 2, 2023
Accepted: July 25, 2023
Published: July 28, 2023

Abstract

A nonhomogeneous Markov chain is applied to the study of the air quality classification in Mexico City when the so-called criterion pollutants are used. We consider the indices associated with air quality using two regulations where different ways of classification are taken into account. Parameters of the model are the initial and transition probabilities of the chain. They are estimated under the Bayesian point of view through samples generated directly from the corresponding posterior distributions. Using the estimated parameters, the probability of having an air quality index in a given hour of the day is obtained.

Keywords

Air Quality Index, Air Pollution, Mexico City, Nonhomogeneous Markov Chains, Bayesian Inference

1. Introduction

Air quality may be defined as the air characteristics in an environment when the levels of the so-called criterion pollutants are taken into account. These criterion pollutants are those considered as keys to establish how polluted an area is and they may vary from location to location. The classification of air quality in a given environment at a particular time is made, for instance, in terms of the criterion pollutants concentrations in sites in that environment at that time (see, for example, [1] [2]). In the case of Mexico City, the pollutants considered as criterion are ozone (O_3), sulphur dioxide (SO_2), carbon monoxide (CO), nitrogen dioxide (NO_2), particulate matter with diameter smaller that 10 microns (PM_{10}), and those with diameter smaller than 2.5 ($PM_{2.5}$). These pollutants are chosen because of their possible harmful impacts on human health, as well as the environment ([3] [4]). For instance, if we have high levels of ozone, the ill, newborn, and elderly may experience serious health deterioration (see, for example, [5] [6] [7] [8], among others). We also know that SO_2 and NO_2 when in contact with the right level of humidity in the atmosphere, may produce acid rain ([9] [10]). Additionally, exposure of pregnant women to CO, PM_{10}, and $PM_{2.5}$ may produce adverse effects on the newborn ([11] [12] [13] [14]), and exposure to PM_{10} and $PM_{2.5}$ may cause cardiovascular problems to the population in general and an increase in mortality of at-risk groups ([15] [16] [17] [18]).

The air quality in the metropolitan area of Mexico City follows the so-called "Metropolitan Index of Air Quality" (IMECA for its name in Spanish)—legislation NADF-009-AIRE-2017 (see [1]). The IMECA is a value without a unit of measure which is obtained from the pollutants' measurements through a linear by parts transformation. This transformation is given in [1]. In 2019 a new legislation—NOM-172-SEMARNAT-2019 ([2])—was introduced in Mexico as a country and also adopted in Mexico City. This new legislation considers what is called the "Air and Health Index". This index is obtained directly from the pollutants' measurements and more on that will be said when the model is applied to the data.

When there is a new legislation introduced, one question that may rise is how the existing and new rules compare when assigning the air quality at a given hour in a given region. For instance, if we take the case of Mexico City, one question that may arise is related to how strict NOM-172 is when compared to NADF-009. That is one of the questions studied here. Other interests are the estimation of the probability that a given air quality index occurs at an hour of interest in a given region; the probability of having a given sequence of air quality indices in consecutive hours in a given region; the probability of having a given air quality index in a given hour taking into account the observed indices some hours in the past, as well as the probability of having an air quality index few hours ahead given that you have the present hour index.

In order to analyze the questions posed in this work, we consider the sets of criterion pollutants measurements obtained in the year 2020 and see how the air quality indices assigned to them, as dictated by NADF-009 and NOM-172 rules, behave. In order to analyze the behavior under the different regulations we use a nonhomogeneous Markov chain model.

Even though, nonhomogeneous Markov models have already been used to study air pollution (such as exceedances of environmental thresholds), as well as other environmental problems (such as tornado activity and rain)—see, for instance, [19]-[25]—in the present work we use this type of model to study air quality indices. Air quality classification has also been studied by [26] [27] [28] [29]. However, these works use homogeneous Markov chains (with and without a spatial component) to analyze Malaysia's data.

Although we use a nonhomogeneous Markov chain model and the Bayesian point of view ([30]) to estimate the parameters present in the model, the novelty and difference of the present study, when compared to previous works, is that this approach is used to study the behavior of air quality indices associated with some pollutants' measurements obtained from the Mexico City's monitoring network. Additionally, the data used differ from the ones used previously since the subject of air quality indices is not tackled in any of the previous works using nonhomogeneous Markov chains. Another novelty here is that we are comparing two different regulations and analyszing how strict they are when contrasted with each other.

The aim of this work is to use a nonhomogeneous Markov chain model to study the sequence of air quality indices assigned to each hour in a day in Mexico City. Using this model we aim to compare the performance of the two rules applied in Mexico City analyzing how strict they are when contrasted with each other. This work is organized as follows. In Section 2, we present the mathematical and the Bayesian formulations of the model. Section 3 gives an application to the case of Mexico City air quality indices. In Section 4, a series of comments regarding the results, as well as more general comments are presented. Finally, in Section 5 we conclude. An appendix, placed after the list of references, gives some of the plots and results mentioned in the main text.

2. The Mathematical and Bayesian Models

Let $N>0$ and $T>0$ be, respectively, the number of days and the number of hours in a given day where we have air quality indices assigned. Assume there are $d\geq 1$ criterion pollutants whose concentrations are taken into account when obtaining the air quality indices. Let $S=\{1,2,\cdots,M\}$, for some known $M\geq 1$, be a set of integer numbers which are associated with the air quality indices that may be assigned to a pollutant, where smaller numbers are assigned to better air quality and larger to worse. (Air quality classification and the integer numbers associated with them will be used indistinctly).

Denote by $Z_{i,t}^{(j)}$ the index associated with the *i*th pollutant at the *t*th hour of the *j*th day; $i=1,2,\cdots,d$; $t=1,2,\cdots,T$; $j=1,2,\cdots,N$. Denote by $Z_t^{(j)}$ the air quality index at the *t*th hour of the *j*th day defined as

$Z_t^{(j)}=\max\left\{Z_{i,t}^{(j)},i=1,2,\cdots,d\right\}$; $t=1,2,\cdots,T$; $j=1,2,\cdots,N$. Let

$\mathbf{Z}^{(j)}=\left\{Z_t^{(j)}:t=1,2,\cdots,T\right\}$ be the process recording the air quality indices in the *j*th day, $j=1,2,\cdots,N$. As in [31] and [32], we assume that $\mathbf{Z}^{(j)}$ is ruled by a nonhomogeneous Markov chain. Denote this chain by $\mathbf{X}=\{X_t:t=1,2,\cdots,T\}$ and let $\mathcal{S}$ be its state space. Therefore, in the present case we will have a total of 366 realizations of the the chain $\mathbf{X}$. The corresponding transition probabilities are given by $P_{ij}(t)=P_{ij}(t,t+1)=P(X_{t+1}=j\mid X_t=i)$; $i,j\in S$; $1\leq t\leq T-1$. We define $Q_i(t)$, $i\in\mathcal{S}$, as the probability $P(X_t=i)$, $t=1,2,\cdots,T$. Hence, when $t=1$, we have $Q_i(1)$, $i\in\mathcal{S}$, the initial distribution of $\mathbf{X}$. Denote by

$P(t)$ the transition matrix whose components are $P_{ij}(t)$; $i, j \in \mathcal{S}$; *i.e.*, $P(t) = (P_{ij}(t))_{i,j\in S}$; $t = 1, 2, \cdots, T-1$. These initial and transition probabilities are parameters to be estimated.

Remark. Note that even though we do not have explicit time dependent formulas for the transition probabilities they do depend on time, since for different values of *t* we allow different values for the probability of a given transition.

As additional information provided by the model, we are able to obtain $P(X_t = i)$, $i \in \mathcal{S}$. The function $P(X_t = \cdot)$ reports the probability of having a given air quality index at time *t*. They are obtained by taking advantage of the Markov property and using a recursive form as follows ([22] [33]). For a given time *t*, we have, for $k \in \mathcal{S}$,

$$P(X_{t+1} = k) = \sum_{i\in S} P(X_{t+1} = k \mid X_t = i) P(X_t = i) = \sum_{i\in S} P_{ik}(t) P(X_t = i), \tag{1}$$

where $P(X_1 = i) = Q_i(1)$, $i \in \mathcal{S}$; $t = 1, 2, \cdots, T-1$.

Remark. Note that even though the recursive formula has similar form as that valid for homogeneous Markov chain, in the nonhomogeneous case the transition probabilities are dependent of *t*, *i.e.*, we have $P_{ij}(t)$ instead of P_{ij}, and the values, for the same transition, may vary for different values of *t*; $i, j \in \mathcal{S}$; $t = 1, 2, \cdots, T-1$. The principle is the same, but the values of the transition probabilities may differ for different values of *t*.

The initial and transition probabilities may be estimated, for instance, using the maximum likelihood method ([34]) and empirical estimators ([35]). In the present work, we use the Bayesian approach to estimate them. Inference is performed using information provided by the so-called posterior distribution of the parameters. The posterior distribution of a vector of parameters $\boldsymbol{\theta}$ of a model describing an observed data set *D*, denoted 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 $\boldsymbol{\theta}$.

In the present case, the vector of parameters is $\boldsymbol{\theta} = (Q(1), P(t), t = 1, 2, \cdots, T-1)$ which belongs to the sample space $\Theta = \left\{\Delta_{|\mathcal{S}|} \times \Delta_{|\mathcal{S}|}^{|\mathcal{S}|(T-1)}\right\}$, where $\Delta_l = \left\{(x_1, x_2, \cdots, x_l) \in \mathbb{R}^l : x_i \geq 0, i = 1, 2, \cdots, l; \sum_{l=1}^{l} x_i = 1\right\}$ is the ($l-1$)-dimensional simplex. We will use as our observation the values given by $\mathbf{Z} = \left(\mathbf{Z}^{(1)}, \cdots, \mathbf{Z}^{(N)}\right)$.

Since a nonhomogeneous Markov model is assumed, the likelihood function is given by (see, for instance, [25] [34] [36] [37])

$$L(\mathbf{Z} \mid \boldsymbol{\theta}) \propto \left(\prod_{i\in\mathcal{S}} \left[Q_i(1)\right]^{n_i}\right) \left\{\prod_{t=1}^{T-1} \left[\prod_{j\in\mathcal{S}} \left(\prod_{k\in\mathcal{S}} \left[P_{jk}(t)\right]^{n_{jk}(t)}\right)\right]\right\}, \tag{2}$$

where n_i is the number of days in which we have state *i* at time $t = 1$ and $n_{jk}(t)$ is the number of days in which a transition from a state *j* at time *t* to a state *k* at time $t+1$ has occurred; $i, j, k \in \mathcal{S}$; $t = 1, 2, \cdots, T-1$.

Another component to be established is the prior distribution of the vector of parameters. In order to do that, we assume a prior independence of $P(t)$ as functions of t, and also a prior independence between the initial and transition probabilities. Given the nature of transition matrices, we assume that rows are independent and that each row of $P(t)$ will have as prior distribution a Dirichlet distribution with appropriate hyperparameters. Therefore, row $(P_{i1}(t), P_{i2}(t), \cdots, P_{iM}(t))$ has as prior distribution a Dirichlet with hyperparameters $(\alpha_{i1}(t), \alpha_{i2}(t), \cdots, \alpha_{iM}(t))$, $\alpha_{ik}(t) > 0$; $i, k \in \mathcal{S}$; $t = 1, 2, \cdots, T-1$. The initial distribution $Q(1)$, will also have a Dirichlet prior distribution, but now with hyperparameters $(\alpha_i; i \in \mathcal{S})$, $\alpha_i > 0$, $i \in \mathcal{S}$. The hyperparameters of the prior distributions will be considered known and will be specified when the model is applied to the data. Hence, we have the following,

$$\begin{aligned} P(\boldsymbol{\theta}) &= P(Q(1))\left[\prod_{t=1}^{T-1} P(P(t))\right] \\ &\propto \left(\prod_{i\in\mathcal{S}} [Q_i(1)]^{\alpha_i - 1}\right)\left\{\prod_{t=1}^{T-1}\left[\prod_{i\in\mathcal{S}}\left(\prod_{k\in\mathcal{S}} [P_{ik}(t)]^{\alpha_{ik}(t)-1}\right)\right]\right\}. \end{aligned} \tag{3}$$

(Recall that the hyperparameters α_i, $\alpha_{jk}(t)$; $i, j, k \in \mathcal{S}$; $1 \le t \le T-1$, are given and hence, are known).

Therefore, because we have a likelihood function proportional to a product of multinomial distributions and the prior distribution of the vector of parameters also a product of Dirichlet distributions, the marginal conditional posterior distributions of the initial distribution and each row of the transition matrices are also Dirichlet distributions (see, for instance, [32]). The hyperparameters of these posterior distributions are, respectively, $(n_i + \alpha_i; i \in \mathcal{S})$ and $(n_{i1}(t) + \alpha_{i1}(t), n_{i2}(t) + \alpha_{i2}(t), \cdots, n_{iM}(t) + \alpha_{iM}(t))$, for the initial distribution and row i of the transition matrix at time t; $i \in \mathcal{S}$; $t = 1, 2, \cdots, T-1$. Therefore, we may generate samples of these probabilities directly from their posterior distributions and use the law of large numbers to obtain the parameters estimates without the need to perform an explicit Markov chain Monte Carlo algorithm.

Remark. The posterior distribution of each row of the transition matrices may be obtained using the expression for the joint posterior distribution which is the product of the expressions for the likelihood function and prior distribution given, respectively, by the formulas (2) and (3) displayed above, and integrating with respect to the remaining variables

3. Application to Air Quality Indices of Mexico City

Application will be made to measurements of the criterion pollutants collected at Mexico City's monitoring network during the year 2020 (http://www.aire.cdmx.gob.mx/default.php?opc=%27aKBh%27). Pollutants' concentrations are measured in parts per million (ppm) in the cases of O_3, SO_2, NO_2, and CO, and in micrograms per cubic meter (μg/m^3) when we consider PM_{10} and $PM_{2.5}$. The monitoring network comprises of several monitoring stations placed throughout the metropolitan area. Measurements in each monitoring station are obtained minute by minute and the averaged hourly results are reported at each station. The data primarily considered are the hourly averaged measurements of each of the criterion pollutants collected during the year 2020.

Even though the pollutants considered as criterion are the same in both regulations, the difference lies in how the air quality indices are assigned to each pollutant. These indices are based on the pollutants measurements and how they are taken into account. Hence, we have the following. In the case of NADF-009, the respective hourly reported averaged measurements are considered in the cases of O_3 and NO_2; the 24-hour moving averages are taken into account in the cases of SO_2, PM_{10}, and $PM_{2.5}$; and in the case of CO data, the 8-hour moving averages are used. If we switch to the NOM-172 rule, then there are two types of measurements for O_3, the hourly reported and the 8-hour moving averages. In the cases of PM_{10} and $PM_{2.5}$, weighted 12-hour moving averages are used. When SO_2, NO_2, and CO are taken into account, the data are as in NADF-009. Another difference between these two rules is that in the NADF-009 legislation only one set of subintervals is used in order to classify the air quality. This set is a result of the linear by parts transformation applied to the pollutants measurements in order to produce the IMECA values (see [1]). In the case of NOM-172, measurements are used directly in order to classify the air quality. Hence, there is a set of intervals for each pollutant (see [2]). Additionally, depending on the rule used, there are different classifications for the air quality. Therefore, in [1] six states for the air quality index were considered. They were "*Good*" (G), "*Regular*" (R), "*Bad*" (B), "*Very Bad*" (VB), "*Extremely Bad*" (EB), and "*Dangerous*" (D), if the calculated IMECA felt in the intervals [0, 50], (50, 100], (100, 150], (150, 200], (200, 300], and (300, 500], respectively. When the NOM-172 was implemented, the intervals depended on each particular pollutant and the air quality was classified as "*Good*" (G), "*Acceptable*" (A), "*Bad*" (B), "*Very Bad*" (VB), and "*Extremely Bad*" (EB) (see [2]). In both cases, the assigned index to a given region at a given hour of the day is the worst of the indices associated with each criterion pollutant at that particular hour when measurements from all stations in the region are taken into account. For instance, if the indices associated with O_3, SO_2, CO, NO_2, PM_{10}, and $PM_{2.5}$ at a given hour are, respectively, G, G, B, A, VB, and VB, then the air quality assigned to that hour is VB.

Therefore, in the present case we have $N = 366$ (since 2020 is a leap year) and $T = 24$, since we have twenty four hours in a day. Even though, in both regulations we have six criterion pollutants, we have $d = 6$ and $d = 7$ in the cases of NADF-009 and NOM-172, respectively. The value $d = 7$ in the NOM-172 rule is because there are two types of O_3 measurements used: the hourly averaged measurements and the 8-hour moving averages. Hence, using the classification provided by NADF-009 and NOM-172, air quality indices were assigned to each hour of the day. Depending on the rule considered we have either six or five possible states for the indices. If we adopt NADF-009 we have the following association: G, R, B, VB, EB, and D corresponding to 1, 2, 3, 4, 5, and 6, respectively. If we take the NOM-172 rule, then we associate G, A, B, VB, and EB with 1, 2, 3, 4, and 5, respectively. Hence, the state space of $\mathbf{X}$ will be either

$\mathcal{S}_{NADF} = \{1,2,3,4,5,6\}$ or $\mathcal{S}_{NOM} = \{1,2,3,4,5\}$ depending if we assume NADF-009 or NOM-172 rule, respectively.

Remark. Note that for each pair of states *ij* and each time *t*, we have 366 values to obtain the empirical transition probabilities, as well as the counting variables $n_{ik}(t)$; $i,k \in \mathcal{S}$; $t = 1,2,\cdots,T-1$.

3.1. Data Analysis

Since the metropolitan area of Mexico City is divided into five regions: northwest (NW), northeast (NE), center (CE), southwest (SW), and southeast (SE), we will assign to each region its own air quality indices and analysis will be performed for each region separately. During the observed time considered here, we have 8784 hours. In each of these hours the air quality index is produced by at least one of the criterion pollutants. In **Table 1** we have the number of times each pollutant was responsible by the air quality index in each region. Note that we may have more than one pollutant producing an air quality index since we might have, for instance, two or more pollutant with value "3" assigned to them with no larger value associated to any other pollutants. In this case, we have that these two pollutants are responsible for the air quality assigned to that particular region.

Looking at **Table 1**, we see that, in all regions, when NADF-009 is taken into account the pollutant with the largest number of times in which it was responsible for the air quality index is $PM_{2.5}$, followed by PM_{10} and ozone as the pollutants with the second and third largest numbers. If we consider the NOM-172 rule, then the pollutants with the largest, second, and third largest number of times producing the air quality indices vary. In the cases of region NE and CE,

Table 1. Number of times each pollutant dictated the air quality index in each region according to the rule used.

		NW	NE	CE	SE	SW
PM_{10}	NADF	6113	6611	5253	5105	3536
	NOM	6466	6012	5737	4958	4450
$PM_{2.5}$	NADF	7769	7073	8163	8187	8158
	NOM	3980	4323	5423	5433	5558
O_3	NADF	1864	1935	2280	2484	3129
	NOM_h	3305	3195	4264	4345	4723
	NOM_8h	3209	3357	4421	4882	5221
NO_2	NADF	877	969	1131	1118	1683
	NOM	2962	2522	3780	3518	3826
CO	NADF	874	966	1131	1117	1683
	NOM	2073	1719	2780	2642	2996
SO_2	NADF	1026	1092	1147	1117	1683
	NOM	3579	2797	3830	3165	3948

we have that the pollutants with the largest, second, and third largest numbers are, respectively, PM_{10}, $PM_{2.5}$, and O_3 with the 8-hour moving average. In the cases of regions NW, SE, and SW, the pollutants are, respectively, PM_{10}, $PM_{2.5}$, and SO_2; $PM_{2.5}$, PM_{10}, and O_3 with the 8-hour moving averages; and $PM_{2.5}$, O_3 with the 8-hour moving averages, and O_3 hourly averages. Therefore, we see that depending on the rule, we have different pollutants as the responsible for the air quality indices, with NOM-172 producing more heterogeneous results depending on the region analized.

Of all regions, SW and CE are considered critical regions because of their geographic positions and the prevailing wind direction (from NE to SW). Region NE has several industries and pollution produced there may be transported to the center and southwest regions; region CE has many vehicles circulating, hence in this region high levels of pollution are produced and it also receives part of that produced in region NE; and region SW receives the pollution produced in the first two regions in addition to that produced at the south end of the city. The SW region also has some mountains which trap the pollution in that area. Region SE has large areas of dry patches with scarce vegetation and hence, it is not surprising that it has high indices of particulate matter.

Focussing on regions CE and SW, looking at **Table 1** we see that under the legislation NADF-009 most of the times (about 92% of the hours) we have that the air quality indices in these regions had the contribution of $PM_{2.5}$ measurements. Under the legislation NOM-172, most of the times, about 65% and 63% of the hours, respectively, air quality indices were results of the contribution of PM_{10} measurements in region CE and $PM_{2.5}$ in the case of region SW. Note that, when comparing the two rules, there is a large difference between the percentages of hours in which particulate matter was a contributor to the air quality indices assigned to the regions, with NOM-172 giving the lower percentage. However, this rule captures the contribution pollutants, other than ozone and particulate matter, may have in the air quality indices assignation. For instance, we have region NW where SO_2 appears as having the third largest number of times in which it contributed to the assignation of the air quality index to that region.

3.2. Results

The assignation of the hyperparameters of the Dirichlet prior distributions, described in Section 2 when the model was presented, is made in a similar manner as in [22] [33]. Hence, we assign values to $\alpha_{ik}(t)$ and α_k; $i,k \in \mathcal{S}_{NADF}, \mathcal{S}_{NOM}$ as follows: $\alpha_{ik}(t) = \alpha_k = 1/16$, for $k = 1,2,3$, and $\alpha_{ik}(t) = \alpha_k = 1/64$, for $k = 4,5,6$ in the case of NADF-009, and $\alpha_{ik}(t) = \alpha_k = 1/16$, for $k = 1,2$, and $\alpha_{ik}(t) = \alpha_k = 1/64$, for $k = 3,4,5$ in the case of NOM-172, $i \in \mathcal{S}_{NADF}, \mathcal{S}_{NOM}$, for all regions.

Remark. Note that the hyperparameters of the Dirichlet prior distributions are chosen to reflect the occurrences of the events with which they are associated. For instance, events that have not been observed during the time interval taken into account, but are part of the space of possible events, will have small probabilities assigned to them whereas those that have been observed will have probabilities that absorb the contributions of both observed data and prior distributions.

Estimation of the initial and transition probabilities was performed using a sample of size 1000 generated directly from their corresponding posterior distributions. As an example of how the fit of the estimated to the observed values are, in Figures A1-A5 given in **Appendix A.1** we give the plots of the transition probabilities in the case of region SW when both rules are applied. Looking at Figures A1-A5 we see that the estimated values represent well the observed and that indeed the Markov chain is nonhomogeneous. The fit and non-homogeneous behavior in the remaining cases is also corroborated.

The estimated values of the initial and transition probabilities are then used to obtain the corresponding values of $P(X_t = i)$; $i \in \mathcal{S}$; $t = 1, 2, \cdots, 24$, using (1). Figure A6 and Figure A7, in **Appendix A.2**, show the plots of these estimated probabilities for all regions and both rules. Since, the values of $P(X_t = k)$; $k = 5, 6$, for NADF-009 are negligible, we have clumped them together and use the combined state "5" to compare to state 5 of NOM-172. Looking at Figure A6 and Figure A7 we see that depending on the regulation and region, sometimes NADF-009 rule provides larger probabilities to some states in some regions whereas in other times and regions this behavior is given by NOM-172. For instance, in the case of state 2 (corresponding to regular and acceptable air qualities in NADF-009 and NOM-172, respectively), in all regions the values of the probabilities using NADF-009 is higher than those given by NOM-172 at all hours of the day. The opposite happens when we consider states 4 and 5 (corresponding, respectively, to states VB and EB+D states in NADF-009 and VB and EB air quality indices in NOM-172). The rules given by NOM-172 favors state 3 (bad air quality under both rules) over NADF-009 if we consider the time period after around 8 am (one of the rush hours—when people go to work) in almost all regions with the exception of region SW where both rules give similar weights until around 3 pm which is another of the rush hours in Mexico City (lunch time around 2 pm). After 3 pm, NOM-172 gives higher probabilities to state 3 than NADF-009. Additionally, NOM-172 always gives a higher probability to good air quality (state 1) until around noon when we compare to the probabilities given by NADF-009. This is consistent with the fact that during the early hours of the day some activities that threatens to increase the levels of pollution have not started yet and those that have started have not yet produced enough quantity to be considered hazardous. Notice that under NADF-009, state 1 (good air quality) has always low probability of occurrence. Similar behavior may be observed when we consider states 4 and 5 (VB and EB air quality in NOM-172 and VB and EB+D states in NADF-009). However, the reason for the small values of the probabilities of their occurrences may be that the thresholds associated with those states are high.

4. Comments

In this study we have used a nonhomogeneous Markov chain model to investigate the behavior of air quality indices in Mexico City when different rules are applied. We consider both the NADF-009 and NOM-172 regulations. Using the rules specified by them, air quality indices were assigned to the criterion pollutants and the regional air quality indices were assigned to each region of Mexico City for every hour of the day during the year 2020. In the present case, results show that the behaviors of the estimated quantities of interest mimic well those of the observed (see plots given in the appendix) showing the suitability of the model and results presented in this work in the study of air quality indices.

We may also compare, as in [22] and [32], the probabilities of having the occurrence of a given air quality index in a given hour of the day when we have information of few hours earlier. For instance, consider region SW and NOM-172 and assume we want to know the probability that at 1 pm the air quality index will be 3, *i.e.*, B under the NOM-172 legislation, and that at 10 am, 11 am, and 12 pm we have the sequence of states 1, 1, and 2 (*i.e.*, G, G, and A air quality). Therefore, we want to know the probability of the following sequence of states, $X_{10}=1, X_{11}=1, X_{12}=2, X_{13}=3$. Hence, we have,

$$\begin{aligned}
&P\left(X_{10}=1, X_{11}=1, X_{12}=2, X_{13}=3\right) \\
&=P\left(X_{13}=3 \mid X_{12}=2\right) \times P\left(X_{12}=2 \mid X_{11}=1\right) \times P\left(X_{11}=1 \mid X_{10}=1\right) \times P\left(X_{10}=1\right) \\
&\approx 0.202 \times 0.483 \times 0.625 \times 0.491 \approx 0.03.
\end{aligned}$$

(The values of the transition probabilities used in the calculations are those estimated using the samples generated from their corresponding posterior distributions. The probability of a state at time 10 am, *i.e.*, $P\left(X_{10}=i\right)$, $i \in \mathcal{S}$, is obtained using the estimated transition and initial probabilities and the recursive formula (1).)

If we compare to the other probabilities, *i.e.*, of having at 1 pm either state 1, 2, 4, or 5 (G, A, VB, or EB), then we have

$$\begin{aligned}
&P\left(X_{10}=1, X_{11}=1, X_{12}=2, X_{13}=i\right) \\
&\approx \begin{cases}
0.013 \times 0.483 \times 0.625 \times 0.491 \approx 1.927\mathrm{E}-03, & i=1 \\
0.763 \times 0.483 \times 0.625 \times 0.491 \approx 0.113, & i=2 \\
0.005 \times 0.483 \times 0.625 \times 0.491 \approx 7.41\mathrm{E}-04, & i=4 \\
0.017 \times 0.483 \times 0.625 \times 0.491 \approx 2.52\mathrm{E}-03, & i=5
\end{cases}
\end{aligned}$$

Hence, this sequence of states gives a high probability of having an A (acceptable) air quality at 1 pm.

When we consider the NADF-009 legislation, the same sequences of states, *i.e.*, $X_{10}=1, X_{11}=1, X_{12}=2, X_{13}=i$ have probabilities approximately equal to 1.13E−04, 0.0145, 1.135E−03, 8.488E−07, 6.34E−07, and 1.369E−07 for $i=1,2,3,4,5$, and 6 respectively. Thus, under the NADF-009 legislation, we also have the occurrence of the sequence $X_{10}=1, X_{11}=1, X_{12}=2, X_{13}=2$ with the highest probability. However, under NADF-009 the probability of this string of states is small when compared to that given the case of NOM-172. Similar analysis may be performed for any sequence of states in a given day in a given region of interest.

Another question that may rise is related to the probability of having a given state in a time $t+s$ into the future given the state at time *t*. In order to do that, we just need to obtain the product of matrices $\prod_{l=t}^{t+s-1} P(l)$. Hence, take for instance the results associated with region SW. Suppose we have a state at time $t = 10\text{ am}$ and want to know the probability of a given state at time $t+s = 4\text{ pm}$. Hence, we need the values of the transition matrices at times 10, 11, 12, 13, 14, and 15, since $P(15)$ will give the transition from time 3 pm to 4 pm. The values of the matrices at different times, using both the NADF-009 and NOM-172 rules, are given in **Appendix A.3** and **Appendix A.4**, respectively.

Consider first the cases where NADF-009 is used. We may see that the highest probability is associated with the transition from state 3 to state 3, *i.e.*, if at 10 am we have bad air quality, then the highest probability is given to the event that at 4 pm we will still have bad air quality. However, if we look at the transitions at the intervals $(t = 10, t+s = 10+s)$, $s = 2,3,4,5$, we see that $P_{33}(t, t+s) = P(X_{t+s} = 3 \mid X_t = 3)$ decreases as *s* increases. Hence, even though we have a high probability of having bad air quality, it decreases as we move further away in time from the morning hour 10 am. We also note that a fast decrease occurs when we start with good air quality (state 1) and obtain the probability of having good air quality at time 4 pm. We see that the transition probability from 1 at 12 pm, which is approximately 0.83, drops to approximately 0.49 at 4 pm, *i.e.*, the probability of continuing to have a good air quality at 4 pm given that we have a good air quality at 12 pm drops almost 50%. On the other hand, we see a substantial increase in the probability of having state 3 (bad air quality) at 4 pm given that we have good air quality (state 1) at 10 am. Similar behavior occurs when we consider state 2 (regular air quality) at time 10 am. Proceeding in this way, we may analyze the other possible transitions.

If we take into account NOM-172, then the highest transition probability is of going from state 4 to state 3, *i.e.*, going from extremely bad to very bad air quality. This behavior is also observed when we take the intervals $(t = 10, t+s = 10+s)$, $s = 2,3,4$. Note that the extremely high transition probability obtained for the interval $(t = 10, t+s = 10+2)$ (corresponding to the transition from state 2 to 2, *i.e.*, acceptable to acceptable air quality) with value approximately 0.838, drops drastically to approximately 0.38 when we consider the interval $(t = 10, t+s = 10+6)$. We also see that the low value of the transition from state 2 (acceptable air quality) at time 10 am to state 3 (bad air quality) at time 12 pm (approximately 0.11) increases to approximately 0.41 if we consider the transition to state 3 at time 4 pm. Similar analysis may be performed for the other transitions.

Consider now a specific transition. Say, we have good air quality (*i.e.*, state 1) at time 10 am. The highest probability is given by the transition to state 2 (*i.e.*, acceptable air quality) followed closely by a transition to state 3 (bad air quality) when we consider the NOM-172. In the case of NADF-009, we have that the highest value is to a transition to state 1 (good air quality) followed, not so closely, by a transition to state 2 (regular air quality) at time 4 pm. Therefore, in this specific case we have NOM-172 giving a more pessimistic value to the air quality at 4 pm, giving high probability to a worse scenario. This is more consistent with what happens in reality, since region SW is the one with more serious air pollution problems followed by region CE.

Analyses similar as that made for region SW may be performed for the other regions. Looking at the results given by the model considered here, we see that they reflect the behavior of the data. Using them as a first approach to predict the behavior of the air quality in a given hour of the day can be very useful, since if there is a high probability of having an ill-suited air quality in a given hour, measures can be taken in order to avoid it, as well as to prevent population exposure to unsuitable air quality. Using the estimated probabilities and the method for simulating Markov chains, we may also simulate different scenarios for the sequence of states in a given day.

Remark. Recently, a working group has been assembled in order to revise the intervals considered in the "Air and Health Index". The new extremes would take into account the information provided by the limits of the intervals given in the regulations considered in 2019 in the case of SO_2.

5. Conclusion

Some conclusion points regarding the results obtained in this work are: estimated initial, transition and probabilities are at time *t* represent well the corresponding observed (empirical) probabilities; the air quality indices associated using the NOM-172 represent well the observed behavior of the pollutants considered in this analysis, however, it is worth mentioning that in some cases NADF-009 also gives a good representation; the air quality indices produced by the NOM-172 allow to detect more times the contribution of pollutants other than ozone and PM_{10} to the air quality classification in Mexico City; the model used in this study allows us to obtain the probability of having a given air quality in a given time of the day using information of the current and past air quality indices. We may also use the model to simulate strings of air quality states in given hours.

Acknowledgements

The authors thank an anonymous reviewer for the comments and suggestions that helped to improve the presentation of the results. This work is part of JACJ's Ph.D. Thesis developed at the Benemérita Universidad Autónoma de Puebla, Puebla, México. JACJ thanks the Ph.D. Scholarship received from CONACyT-Mexico. Part of this work was developed while ERR was in a sabbatical visit at the Department of Statistics of the Universidade Estadual Paulista "Júlio de Mesquita Filho" (UNESP)—Campus Presidente Prudente, Brazil, with a grant from the Dirección General de Apoyo al Personal Académico of the Universidad Nacional Autónoma de México, Mexico (PASPA-Aug./2022-Jan./2023). ERR is grateful to the Department of Statistics of UNESP, for support and hospitality during the development of this work.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] N.A.D.F. (2018) Norma Ambiental para el Distrito Federal NADF-009-AIRE-2017. Gaceta Oficial de la Ciudad de México. 14 de noviembre de 2018. Mexico. (In Spanish) http://www.aire.cdmx.gob.mx/descargas/monitoreo/normatividad/NADF-009-AIRE-2017.pdf

[2] N.O.M. (2019) Norma Oficial Mexicana NOM-172-SEMARNAT-2019. Diario Oficial de la Federación. 20 de noviembre de 2019. Mexico. (In Spanish) https://www.dof.gob.mx/nota_detalle.php?codigo=5579387fecha=20/11/2019#gsc.tab=0

[3] World Health Organization (2006) Air Quality Guidelines-2005, Particulate Matter, Ozone, Nitrogen Dioxide and Sulphur Dioxide. European Union, World Health Organization Regional Office for Europe, Geneva.

[4] Gauderman, W.J., Avol, E., Gililand, F., Vora, H., Thomas, D., Berhane, K., McConnel, R., Kuenzli, N., Lurmman, F., Rappaport, E., *et al*. (2004) The Effects of Air Pollution on Lung Development from 10 to 18 Years of Age. *The New England Journal of Medicine*, **351**, 1057-1067. https://doi.org/10.1056/NEJMoa040610

[5] Bell, M.L., McDermontt, A., Zeger, S.L., Samet, J.M. and Dominici, F. (2004) Ozone and Short-Term Mortality in 95 US Urban Communities, 1987-2000. *Journal of the American Medical Society*, **292**, 2372-2378. https://doi.org/10.1001/jama.292.19.2372

[6] Bell, M.L., Peng, R.D. and Dominici, F. (2006) The Exposure-Response Curve for Ozone and Risk of Mortality and the Adequacy of Current Ozone Regulations. *Environmental Health Perspectives*, **114**, 532-536. https://doi.org/10.1289/ehp.8816

[7] Lippmann, M. (1989) Health Effects of Ozone: A Critical Review. *Journal of Air Pollution Control Association*, **39**, 672-695. https://doi.org/10.1080/08940630.1989.10466554

[8] Loomis, D., Borja-Arbuto, V.H., Bangdiwala, S.I. and Shy, C.M. (1996) Ozone Exposure and Daily Mortality in Mexico City: A Time Series Analysis. Health Effects Institute Research Report, No. 75, 1-46.

[9] Mohnen, V.A. (1988) The Challenge of Acid Rain. *Scientific American*, **259**, 30-38. https://doi.org/10.1038/scientificamerican0888-30

[10] Likens, G. (2010) Acid Rain. In: Cutler, J., Ed., *Encyclopedia of Earth*, Environmental Protection Agency, USA. https://editors.eol.org/eoearth/wiki/Acid_rain

[11] Ritz, B. and Yu, F. (1999) The Effects of Ambient Carbon Monoxide on Low Birth Weight among Children Born in Southern California between 1989 and 1993. *Environmental Health Perspectives*, **107**, 17-25. https://doi.org/10.1289/ehp.9910717

[12] Ritz, B., Yu, F., Chapa, G. and Fruin, S. (2000) Effects of Air Pollution on Preterm

Birth among Children Born in Southern California between 1989 and 1993. *Epidemiology*, **11**, 502-511. https://doi.org/10.1097/00001648-200009000-00004

[13] Ritz, B., Wilhelm, M., Hoggart, K.J. and Ghosh, J.K.C. (2007) Ambient Air Pollution and Preterm Birth in the Environment and Pregnancy Outcomes Study at the University of California, Los Angeles. *American Journal of Epidemiology*, **166**, 1045-1052. https://doi.org/10.1093/aje/kwm181

[14] Salam, M.T., Millstein, J., Li, Y.-F., Lurmann, F.W., Margolis, H.G. and Gilliland, F.D. (2005) Birth Outcomes and Prenatal Exposure to Ozone, Carbon Monoxide, and Particulate Matter: Results from the Children's Health Study. *Environmental Health Perspectives*, **113**, 16-38. https://doi.org/10.1289/ehp.8111

[15] Itô, K. and Thurston, G.D. (1996) Daily PM_{10}/Mortality Associations: An Investigation of At-Risk Subpopulations. *Journal of Exposure Analysis and Environmental Epidemiology*, **6**, 79-95. https://doi.org/10.1097/00001648-199503000-00241

[16] Mauderly, J. (1997) Relevance of Particle-Induced Rat Lung Tumors for Assessing Lung Carcinogenic Hazard and Human Lung Cancer Risk. *Environmental Health Perspectives*, **105**, 1337-1346. https://doi.org/10.1289/ehp.97105s51337

[17] Pope III, C.A. and Dockery, D.W. (2006) Health Effects of Fine Particulate Air Pollution: Lines That Connect. *Journal of Air and Waste Management and Assessment*, **56**, 709-742. https://doi.org/10.1080/10473289.2006.10464485

[18] Thurston, G.D. (1996) A Critical Review of PM_{10} Mortality Time Series Studies. *Journal of Exposure Analysis and Environmental Epidemiology*, **6**, 3-21.

[19] Paroli, R., Pistollato, S., Rosa, M. and Spezia, L. (2005) Non-Homogeneous Markov Mixture of Periodic Autoregressions for the Analysis of Air Pollution in the Lagoon of Venice. *Proceedings of Applied Stochastic Models and Data Analysis*, Brest, 16-20 May 2005, 1124-1132.

[20] Lagona, F., Maruotti, A. and Picone, M. (2011) A Non-Homogeneous Hidden Markov Model for Analysis of Multi-Pollutant Exceedances Data. In: Dymarski, P., Ed., *Hidden Markov Models*: *Theory and Applications*, InTech, Croatia, 207-222. https://doi.org/10.5772/14749

[21] Rodrigues, E.R., Tarumoto, M.H. and Tzintzun, G. (2019) Application of a Non-Homogeneous Markov Chain with Seasonal Transition Probabilities to Ozone Data. *Journal of Applied Statistics*, **46**, 395-415. https://doi.org/10.1080/02664763.2018.1492527

[22] Gallegos-Herrada, M.A., Rodrigues, E.R., Tarumoto, M.H. and Tzintzun, G. (2023) A Multi-Dimensional Non-Homogeneous Markov Chain of Order K to Jointly Study Multi-Pollutant Exceedances. *Environmental and Ecological Statistics*, **30**, 157-187. https://doi.org/10.1007/s10651-023-00557-8

[23] Hughes, J.P., Guttorp, P. and Charles, S.P. (1999) A Non-Homogeneous Hidden Markov Model for Precipitation Occurrence. *Journal of the Royal Statistical Society Series C*: *Applied Statistics*, **48**, 16-30. https://doi.org/10.1111/1467-9876.00136

[24] Rajagopalan, B., Upmanu, L. and Tarboton, D.G. (1996) Non-Homogeneous Markov Model for Daily Precipitation. *Journal of Hydrology Engineering*, **1**, 33-40. https://doi.org/10.1061/(ASCE)1084-0699(1996)1:1(33)

[25] Drton, M., Marzban, C., Guttorp, P. and Schafer, J.T. (2003) A Markov Chain Model for Tornadic Activity. *Monthly Weather Review*, **131**, 2941-2953. https://doi.org/10.1175/1520-0493(2003)131<2941:AMCMOT>2.0.CO;2

[26] Alyousifi, Y., Masseran, N. and Ibrahim, K. (2018) Modeling the Stochastic Dependence of Air Pollution Index Data. *Stochastic Environmental Research and Risk Assessment*, **32**, 1603-1611. https://doi.org/10.1007/s00477-017-1443-7

[27] Alyousifi, Y., Ibrahim, K., Kang, W. and Zin, W.Z.W. (2019) Markov Chain Modeling for Air Pollution Index Based on Maximum a Posteriori Method. *Air Quality, Atmosphere and Health*, **12**, 1521-1531. https://doi.org/10.1007/s11869-019-00764-y

[28] Alyousifi, Y., Ibrahim, K., Kang, W. and Zin, W.Z.W. (2020) Modeling the Spatio-Temporal Dynamics of Air Pollution Index Based on Spatial Markov Chain Model. *Environmental Monitoring and Assessment*, **192**, Article No. 719. https://doi.org/10.1007/s10661-020-08666-8

[29] Alyousifi, Y., Ibrahim, K., Kang, W. and Zin, W.Z.W. (2021) Robust Empirical Bayes Approach for Markov Chain Modeling of Air Pollution Index. *Journal of Environmental Health Science and Engineering*, **9**, 343-356. https://doi.org/10.1007/s40201-020-00607-4

[30] Carlin, B.P. and Louis, T.A. (2000) Bayes and Empirical Bayes Methods for Data Analysis. 2nd Edition, Chapman and Hall/CRC, New York. https://doi.org/10.1201/9781420057669

[31] Álvarez, L.J., Fernández-Bremauntz, A.A., Rodrigues, E.R. and Tzintzun, G. (2005) Maximum a Posteriori Estimation of the Daily Ozone Peaks in Mexico City. *Journal of Agricultural, Biological, and Environmental Statistics*, **10**, 276-290. https://doi.org/10.1198/108571105X59017

[32] Rodrigues, E.R., Tarumoto, M.H. and Tzintzun, G. (2015) A Non-Homogeneous Markov Chain Model to Study Ozone Exceedances in Mexico City. In: Nejadkoorki, F., Ed., *Current Air Quality Issues*, InTech, Croatia, 375-394. https://doi.org/10.5772/59728

[33] Morales-Morillón, J.A. (2018) Cadenas de Markov no-homogéneas para el mapeo genético de poblaciones mezcladas (Admixture Mapping). Undergraduate Final Year Report, Facultad de Ciencias, Universidad Nacional Autónoma de México, Mexico. (In Spanish)

[34] Fleming, T.R. and Harrington, D.P. (1978) Estimation for Discrete Time Non-Homogeneous Markov Chains. *Stochastic Processes and Their Applications*, **7**, 131-139. https://doi.org/10.1016/0304-4149(78)90012-1

[35] Aalem, O.O. and Johansen, S. (1978) An Empirical Transition Matrix for Non-Homogeneous Markov Chains Based on Censored Observation. *Scandinavian Journal of Statistics*, **5**, 141-150.

[36] Fan, T.-H. and Tsai, C.-A. (1999) A Bayesian Method in Determining the Order of a Finite Markov Chain. *Communications in Statistics—Theory and Methods*, **28**, 1711-1730. https://doi.org/10.1080/03610929908832381

[37] Boys, R.J. and Henderson, D.A. (2002) On Determining the Order of Markov Dependence of an Observed Process Governed by a Hidden Markov Chain. *Scientific Programming*, **10**, Article ID: 683164. https://doi.org/10.1155/2002/683164

Appendix

In this appendix we present the plots of the estimated and observed transition probabilities for region SW when both the NADF-009 and NOM-172 rules are taken into account (**Figures A1-A5**). We also present a comparison between the plots of the estimated probabilities $P(X_t = k)$; $k \in \mathcal{S}$; when we take into account both NADF-009 and NOM-172 rules and data from all regions (**Figure A6** and **Figure A7**). We also have the products of the transition matrices at time $t+s$ with $t=10$ for $s=2,3,4,5$, for both the NADF-009 and NOM-172 rules when we consider data from region SW.

A.1. Estimated and Observed Transition Probabilities in the Case of NOM-172 and NADF-009 in the Case of Region SW

In this section we present the plots of the estimated and observed transition probabilities for region SW when both the NADF-009 and NOM-172 rules are taken into account.

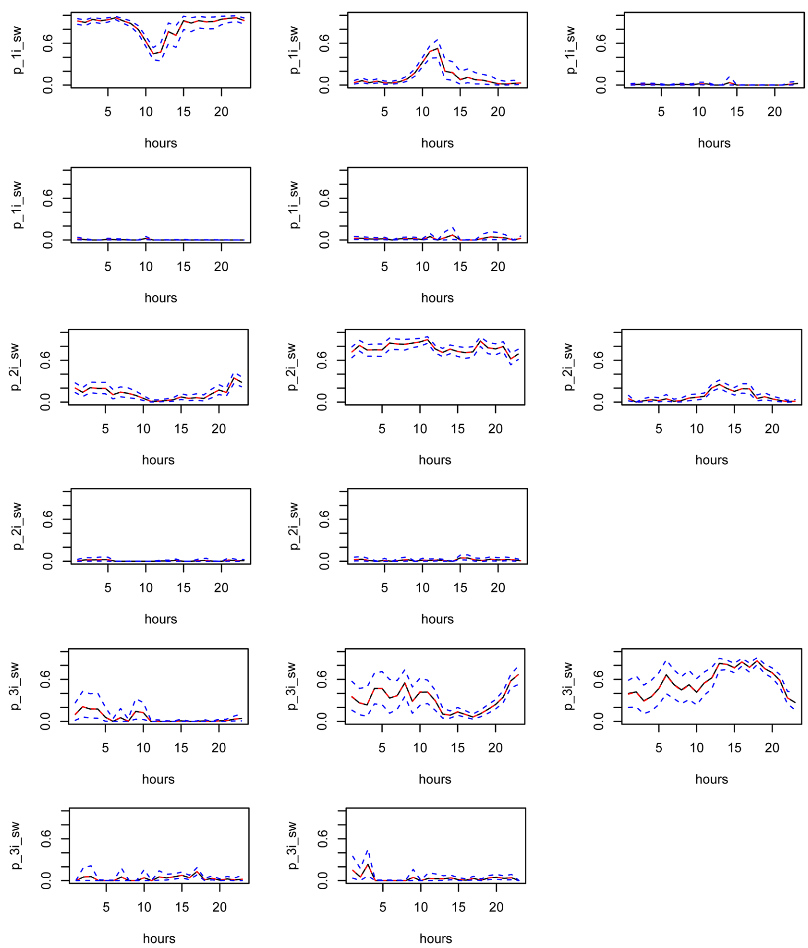

Figure A1. Estimated (dashed red lines) and observed (black lines) transition probabilities P_{1i}, P_{2i}, and P_{3i}, $i=1,2,3,4,5$ (from left to right) together with the 95% credible intervals (blue dashed lines) when data from region SW are used and the NOM-172 is taken into account.

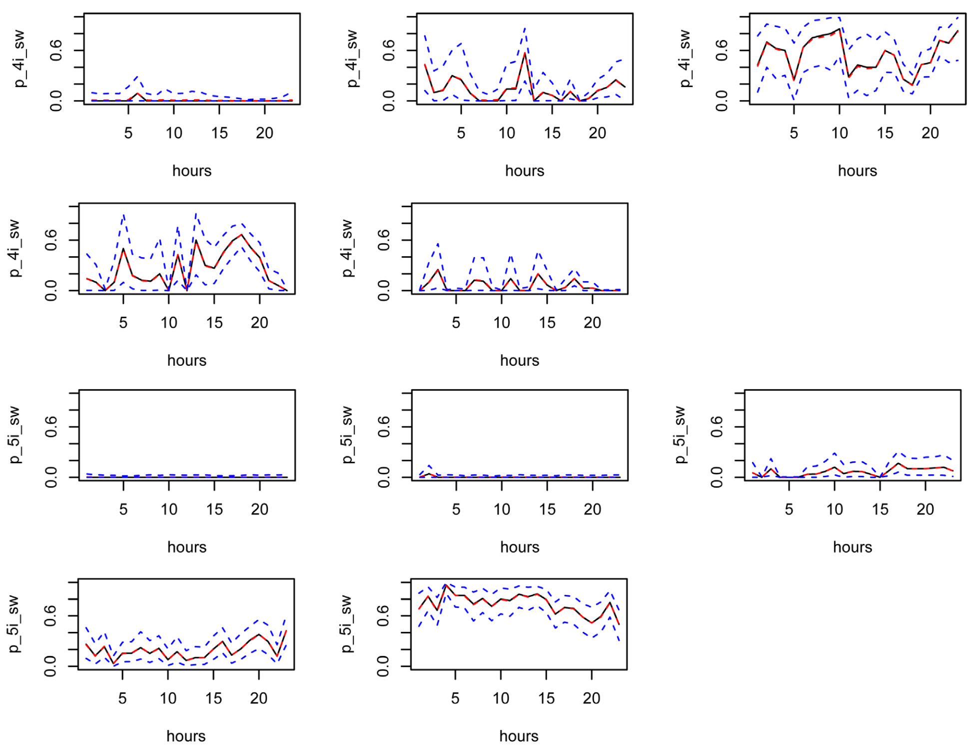

Figure A2. Estimated (dashed red lines) and observed (black lines) transition probabilities P_{4i} and P_{5i}, $i = 1,2,3,4,5$ (from left to right) together with the 95% credible intervals (blue dashed lines) when data from region SW are used and the NOM-172 is taken into account.

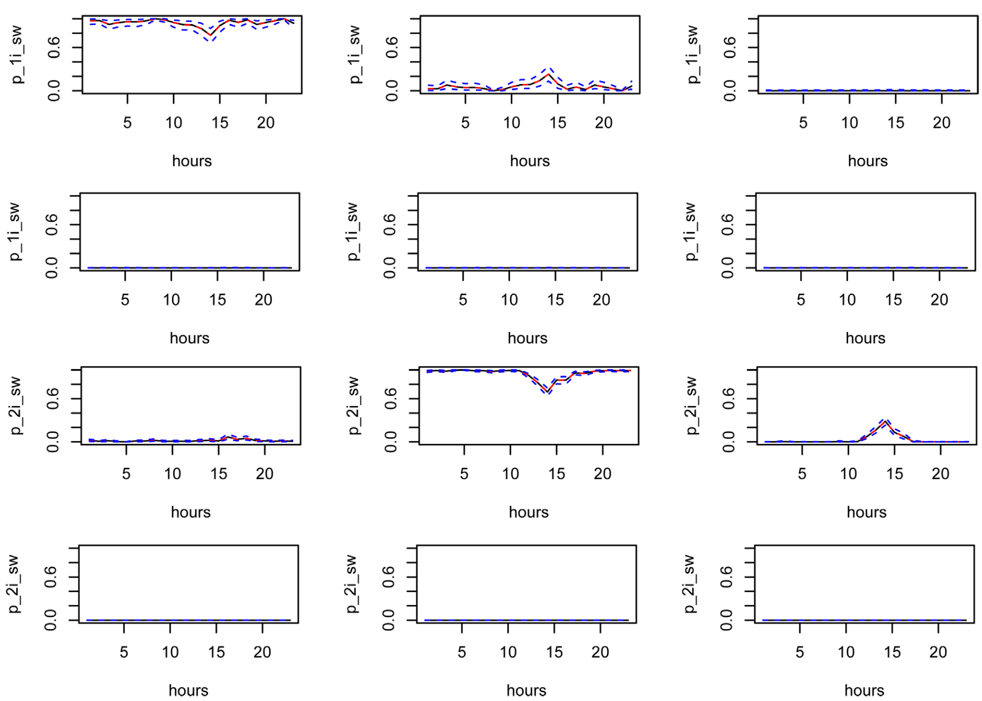

Figure A3. Estimated (dashed red lines) and observed (black lines) transition probabilities P_{1i} and P_{2i}, $i = 1,2,3,4,5,6$ (from left to right) together with the 95% credible intervals (blue dashed lines) when data from region SW are used and the NADF-009 is taken into account.

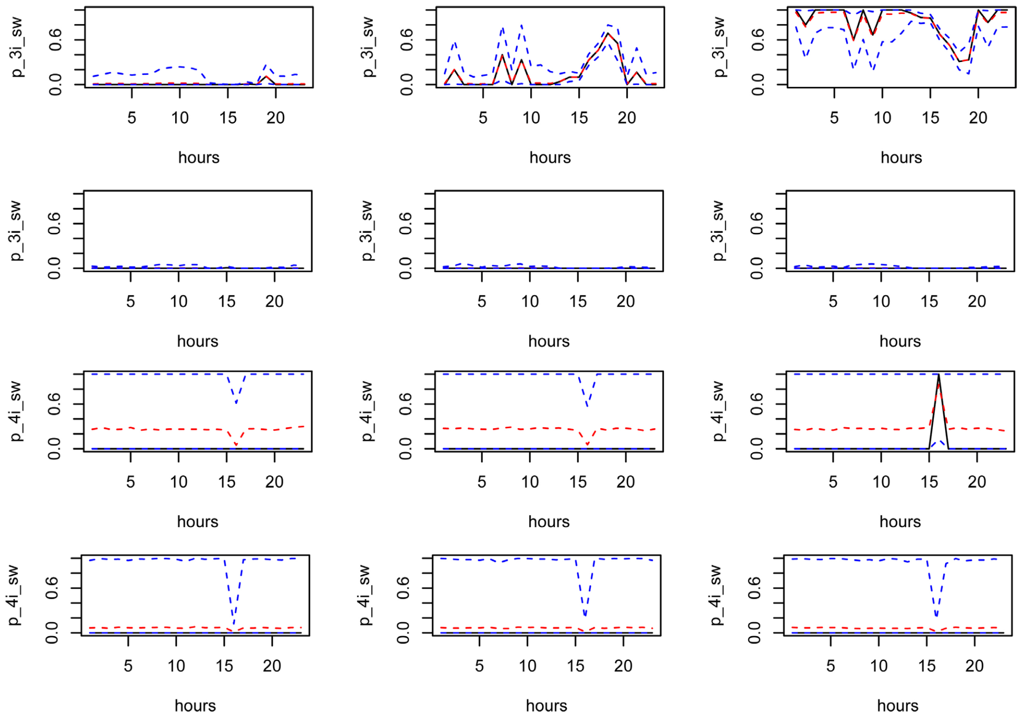

Figure A4. Estimated (dashed red lines) and observed (black lines) transition probabilities P_{3i} and P_{4i}, $i = 1, 2, 3, 4, 5, 6$ (from left to right) together with the 95% credible intervals (blue dashed lines) when data from region SW are used and the NADF-009 is taken into account.

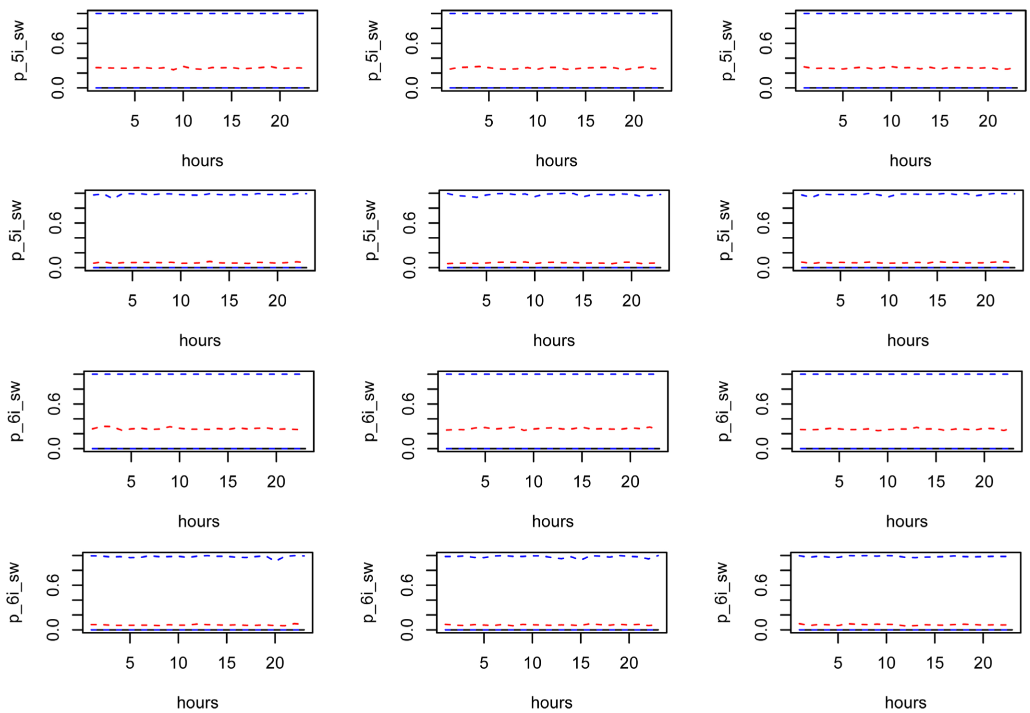

Figure A5. Estimated (dashed red lines) and observed (black lines) transition probabilities P_{5i} and P_{6i}, $i = 1, 2, 3, 4, 5, 6$ (from left to right) together with the 95% credible intervals (blue dashed lines) when data from region SW are used and the NADF-009 is taken into account.

A.2. Estimated Probabilities at Time *t*.

In this section we present the plots of the probabilities $P(X_t = k)$; $k \in \mathcal{S}$, under the two rules for all regions.

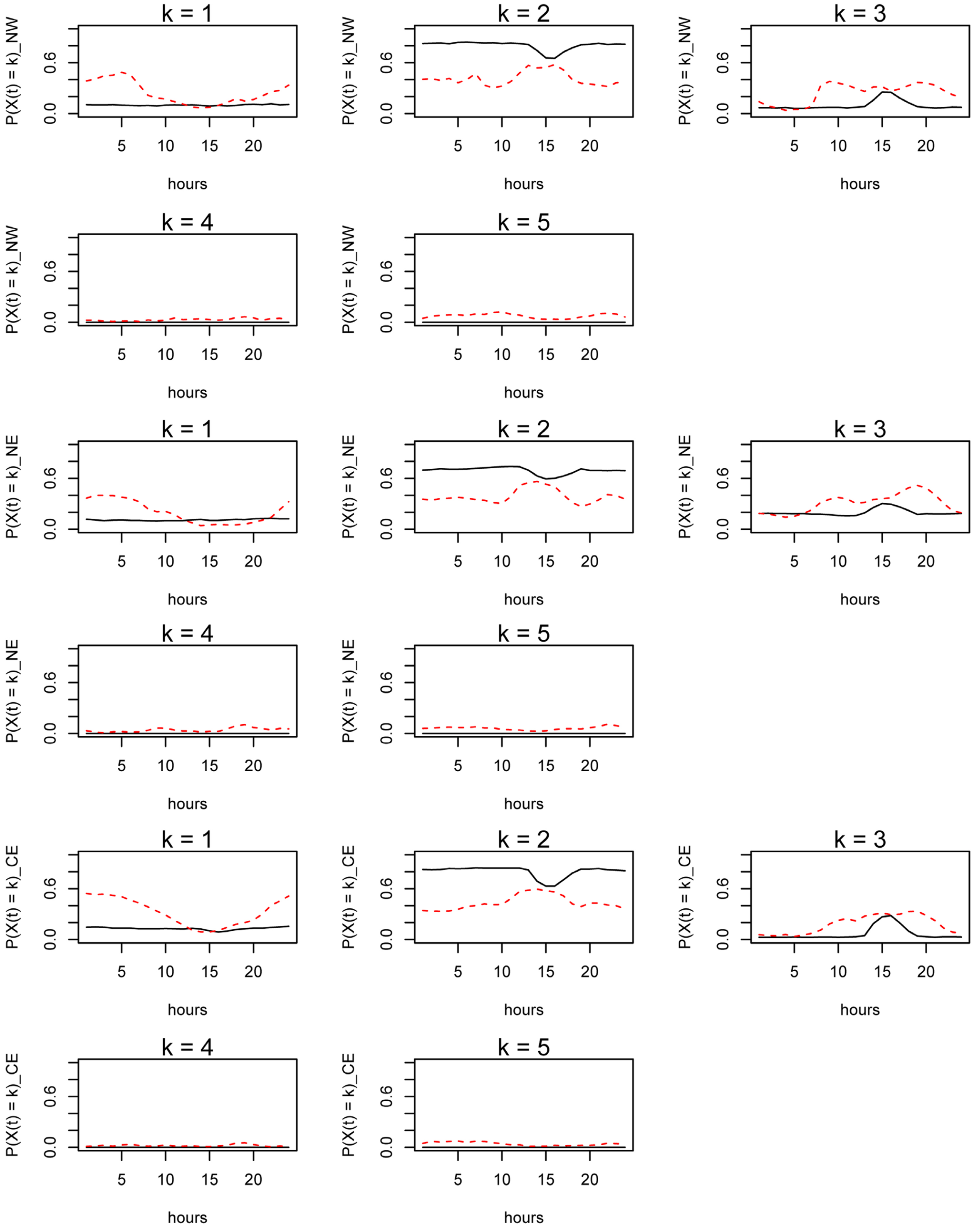

Figure A6. Estimated probabilities $P(X_t = k)$, $k \in \mathcal{S}$ under the NOM-172 (red dashed lines) and NADF-009 (black continuous lines) legislations when data from regions NW, NE, and CE are used.

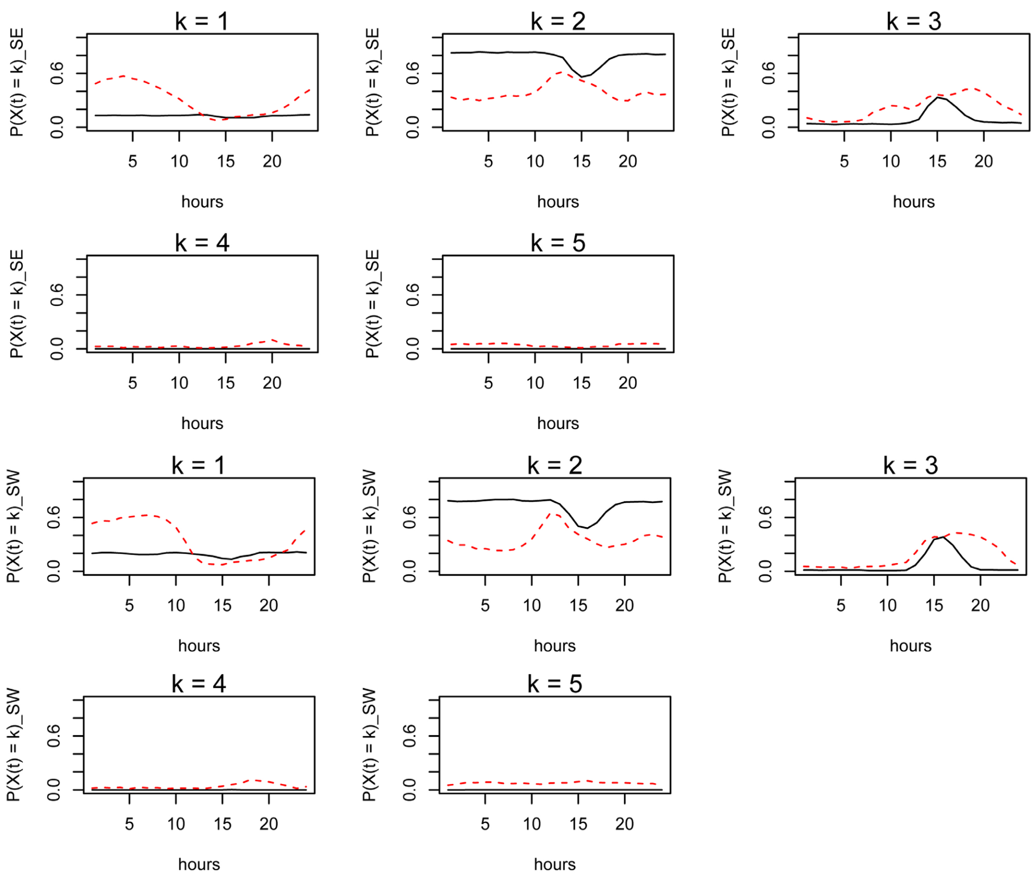

Figure A7. Estimated probabilities $P(X_t = k)$, $k \in \mathcal{S}$ under the NOM-172 (red dashed lines) and NADF-009 (black continuous lines) legislations when data from regions SE and SW are used.

A.3. Transition Matrices under NADF-009 Legislation

In this appendix we present the values of the k-step transition matrices at time $t+k$ with $t=10$ for $k=2,3,4,5$, in the case of NADF-009 legislation.

$$P_{NADF}(10)\times P_{NADF}(11)$$

$$=\begin{pmatrix} 0.86728496 & 0.1298228 & 0.002121038 & 2.879144e-04 & 2.463245e-04 & 2.369375e-04 \\ 0.01365320 & 0.9822032 & 0.003951540 & 5.186911e-05 & 7.529605e-05 & 6.488195e-05 \\ 0.03903096 & 0.0457814 & 0.898957725 & 5.823604e-03 & 5.473678e-03 & 4.932630e-03 \\ 0.29932902 & 0.3450891 & 0.312730790 & 1.368365e-02 & 1.512898e-02 & 1.403840e-02 \\ 0.31738536 & 0.3265309 & 0.317425648 & 1.237760e-02 & 1.361011e-02 & 1.267042e-02 \\ 0.31223209 & 0.3461313 & 0.296905120 & 1.426511e-02 & 1.573003e-02 & 1.473641e-02 \end{pmatrix}$$

$$P_{NADF}(10)\times P_{NADF}(11)\times P_{NADF}(12)$$

$$=\begin{pmatrix} 0.7932578 & 0.19378723 & 0.01233826 & 2.016255e-04 & 2.081614e-04 & 0.0002069464 \\ 0.0196220 & 0.90565576 & 0.07445590 & 8.758518e-05 & 6.919422e-05 & 0.0001095604 \\ 0.0538462 & 0.06322078 & 0.86970851 & 5.419730e-03 & 4.404689e-03 & 0.0034000942 \\ 0.2916582 & 0.35972633 & 0.33633613 & 4.749636e-03 & 4.079265e-03 & 0.0034504618 \\ 0.3070073 & 0.34309148 & 0.33840815 & 4.457437e-03 & 3.818362e-03 & 0.0032172889 \\ 0.3036946 & 0.36207284 & 0.32174396 & 4.820716e-03 & 4.148161e-03 & 0.0035197486 \end{pmatrix}$$

$$P_{NADF}(10)\times P_{NADF}(11)\times P_{NADF}(12)\times P_{NADF}(13)$$
$$=\begin{pmatrix} 0.68680192 & 0.2673267 & 0.04511503 & 0.0002562627 & 0.0002452247 & 0.0002548267 \\ 0.03024087 & 0.7463856 & 0.22299945 & 0.0001071835 & 0.0001426761 & 0.0001241934 \\ 0.05279191 & 0.1002298 & 0.84258347 & 0.0014552637 & 0.0015653432 & 0.0013742223 \\ 0.26060852 & 0.3517038 & 0.38429445 & 0.0011734771 & 0.0011592651 & 0.0010604524 \\ 0.27339790 & 0.3400793 & 0.38327522 & 0.0011196013 & 0.0011118267 & 0.0010161472 \\ 0.27103968 & 0.3546869 & 0.37085747 & 0.0011841772 & 0.0011644084 & 0.0010673418 \end{pmatrix}$$

$$P_{NADF}(10)\times P_{NADF}(11)\times P_{NADF}(12)\times P_{NADF}(13)\times P_{NADF}(14)$$
$$\begin{pmatrix} 0.53391704 & 0.3477369 & 0.1176640 & 0.0001952744 & 0.0002189690 & 0.0002678320 \\ 0.03967104 & 0.5471454 & 0.4128126 & 0.0001053052 & 0.0001352007 & 0.0001304123 \\ 0.04489038 & 0.1678843 & 0.7858325 & 0.0004363241 & 0.0004813575 & 0.0004751864 \\ 0.20924004 & 0.3434012 & 0.4462270 & 0.0003514340 & 0.0003857189 & 0.0003945451 \\ 0.21878027 & 0.3381339 & 0.4419779 & 0.0003431609 & 0.0003772599 & 0.0003875900 \\ 0.21731255 & 0.3465122 & 0.4350389 & 0.0003529732 & 0.0003869620 & 0.0003964589 \end{pmatrix}$$

$$P_{NADF}(10)\times P_{NADF}(11)\times P_{NADF}(12)\times P_{NADF}(13)\times P_{NADF}(14)\times P_{NADF}(15)$$
$$=\begin{pmatrix} 0.48586234 & 0.3612326 & 0.1512670 & 0.001124792 & 0.0002913879 & 0.0002219073 \\ 0.04223916 & 0.5142324 & 0.4399513 & 0.003309384 & 0.0001454586 & 0.0001223727 \\ 0.04309104 & 0.2269008 & 0.7233028 & 0.006280240 & 0.0002133945 & 0.0002117309 \\ 0.19314011 & 0.3592690 & 0.4434931 & 0.003650928 & 0.0002378757 & 0.0002089682 \\ 0.20168003 & 0.3552357 & 0.4390186 & 0.003618069 & 0.0002387894 & 0.0002088093 \\ 0.20045489 & 0.3615969 & 0.4339320 & 0.003565676 & 0.0002402178 & 0.0002103127 \end{pmatrix}$$

A.4. Transition Matrices under NOM-172 Legislation

In this appendix we present the values of the k-step transition matrices at time $t+k$ with $t=10$ for $k=2,3,4,5$, in the case of NOM-172 legislation.

$$P_{NOM}(10)\times P_{NOM}(11)$$
$$=\begin{pmatrix} 0.285894695 & 0.60830515 & 0.0525356 & 0.010681151 & 0.04258341 \\ 0.033856513 & 0.82729570 & 0.1094398 & 0.002901937 & 0.02650608 \\ 0.061911956 & 0.61773813 & 0.2720796 & 0.018161457 & 0.03010885 \\ 0.006429186 & 0.49019092 & 0.4710980 & 0.001851115 & 0.03043082 \\ 0.003974996 & 0.06809812 & 0.1226210 & 0.172475314 & 0.63283053 \end{pmatrix}$$

$$P_{NOM}(10)\times P_{NOM}(11)\times P_{NOM}(12)$$
$$=\begin{pmatrix} 0.143991494 & 0.6361475 & 0.1629568 & 0.00854453 & 0.04835969 \\ 0.026945924 & 0.6838821 & 0.2379670 & 0.01134303 & 0.03986202 \\ 0.038092906 & 0.5963030 & 0.3022516 & 0.01951440 & 0.04383815 \\ 0.010307591 & 0.5203849 & 0.3922892 & 0.02969768 & 0.04732061 \\ 0.006163374 & 0.1916220 & 0.2043570 & 0.05157009 & 0.54628758 \end{pmatrix}$$

$$P_{NOM}(10)\times P_{NOM}(11)\times P_{NOM}(12)\times P_{NOM}(13)$$
$$=\begin{pmatrix} 0.12491464 & 0.4985168 & 0.3024831 & 0.01945058 & 0.05463489 \\ 0.03654933 & 0.5169363 & 0.3769723 & 0.02347644 & 0.04606565 \\ 0.04322419 & 0.4638306 & 0.4113310 & 0.03097039 & 0.05064374 \\ 0.02042782 & 0.4139566 & 0.4706395 & 0.04069420 & 0.05428188 \\ 0.01104967 & 0.1608835 & 0.2752688 & 0.09512706 & 0.45767094 \end{pmatrix}$$

$$P_{NOM}(10)\times P_{NOM}(11)\times P_{NOM}(12)\times P_{NOM}(13)\times P_{NOM}(14)$$
$$=\begin{pmatrix} 0.10671288 & 0.4315340 & 0.3561246 & 0.03426467 & 0.07136387 \\ 0.04444095 & 0.4365171 & 0.4184575 & 0.03894169 & 0.06164268 \\ 0.04742384 & 0.4011857 & 0.4395324 & 0.04297404 & 0.06888403 \\ 0.02954652 & 0.3654060 & 0.4813401 & 0.04899905 & 0.07470838 \\ 0.01517133 & 0.1600377 & 0.3090895 & 0.09287188 & 0.42282963 \end{pmatrix}$$

$$P_{NOM}(10)\times P_{NOM}(11)\times P_{NOM}(12)\times P_{NOM}(13)\times P_{NOM}(14)\times P_{NOM}(15)$$
$$=\begin{pmatrix} 0.13251298 & 0.3737090 & 0.3580718 & 0.05113890 & 0.08456733 \\ 0.07629579 & 0.3811442 & 0.4091275 & 0.05511971 & 0.07831278 \\ 0.07663088 & 0.3589040 & 0.4222203 & 0.05928177 & 0.08296309 \\ 0.05798270 & 0.3376266 & 0.4522358 & 0.06525374 & 0.08690114 \\ 0.02922407 & 0.1669670 & 0.3159645 & 0.13466304 & 0.35318148 \end{pmatrix}$$