Journal of Data Science 21 (4), 799–817
October 2023
DOI: 10.6339/23-JDS1091
Data Science in Action

Analyzing the Rainfall Pattern in Honduras Through Non-Homogeneous Hidden Markov Models

Gustavo Alexis Sabillón[1] and Daiane Aparecida Zuanetti[1,*]
[1]*Departamento de Estatística, Universidade Federal de São Carlos, Brazil*

Abstract

One of the major climatic interests of the last decades has been to understand and describe the rainfall patterns of specific areas of the world as functions of other climate covariates. We do it for the historical climate monitoring data from Tegucigalpa, Honduras, using non-homogeneous hidden Markov models (NHMMs), which are dynamic models usually used to identify and predict heterogeneous regimes. For estimating the NHMM in an efficient and scalable way, we propose the stochastic Expectation-Maximization (EM) algorithm and a Bayesian method, and compare their performance in synthetic data. Although these methodologies have already been used for estimating several other statistical models, it is not the case of NHMMs which are still widely fitted by the traditional EM algorithm. We observe that, under tested conditions, the performance of the Bayesian and stochastic EM algorithms is similar and discuss their slight differences. Analyzing the Honduras rainfall data set, we identify three heterogeneous rainfall periods and select temperature and humidity as relevant covariates for explaining the dynamic relation among these periods.

Keywords *Bayesian approach; dynamic models; estimation and classification performance; rainfall pattern description; stochastic EM algorithm*

1 Introduction

One of the major climatic interests of the last decades has been to understand and describe the rainfall patterns of specific areas of the world and how they are influenced or explained by other climatic characteristics. Here, our main interest is to analyze these features for the historical climate monitoring data from Tegucigalpa, Honduras, using hidden Markov models (HMMs), since rainfall patterns change over time due to other climate variables and we want to describe and understand these latent associations.

These models are a statistical tool in which the system being modeled is considered a Markov process with non-observable states, that usually identify heterogeneous regimes. For each non-observable state there is a corresponding emission of an observable value that makes up the observed data set. In the context of homogeneous hidden Markov models (HHMMs), transition probabilities between the non-observable states are constant. For this reason, HHMMs may not be appropriate for modeling certain practical situations in which transition probabilities vary through time. On the other hand, transition probabilities for the non-homogeneous hidden Markov model (NHMM) vary through time and depend on covariates. This modification on the structure of homogeneous Markov models provides an interesting alternative to model processes

*Corresponding author. Email: dzuanetti@ufscar.br.

Received August 31, 2022; Accepted February 19, 2023

which have transitions between non-observable states that depend on time-varying covariates, that is, which are dynamic. In the manuscript, we present efficient and scalable methods for estimating a NHMM, identifying heterogeneous regimes in a time series and selecting relevant covariates that explain this heterogeneity for the Tegucigalpa data over the last 50 years.

A common approach for estimating the parameters of a NHMM is using maximum likelihood estimation. In general, this is done using the Expectation-Maximization (EM) algorithm (Dempster et al., 1977) also known as Baum-Welch algorithm for HMMs (Rabiner, 1989; MacDonald and Zucchini, 1997; Zucchini and MacDonald, 2009; Maruotti and Rocci, 2012; Zucchini et al., 2016), that includes forward-backward and Viterbi algorithms for gaining computacional efficiency. This algorithm has been widely used by various authors and distinct research areas. In environmental and climate studies, for example, Robertson et al. (2004) examine the probability distribution of local daily rainfall occurrence in the state of Ceará (Brazil) and identifies relationships with large-scale atmospheric circulation patterns; Betrò et al. (2008) investigate the capability of a NHMM in identifying possible recurrent patterns in the occurrence of extreme events over a small area of Central-East Sardinia (Italy); Lagona et al. (2011) estimate the probabilities of multi-pollutant exceedances, conditioning the occurrence and persistence of these exceedances on time varying factors; Neykov et al. (2012) apply the EM algorithm to link atmospheric circulation to daily precipitation at Bulgaria; among others. In other areas, Ghavidel et al. (2015) use a NHMM for mapping quantitative trait loci (QTLs) and Pennoni and Genge (2020) for identifying similar typologies of individuals sharing common perceptions according to different dimensions of trust, etc.

Despite being known as a good method for computing point estimates (Rydén, 2008), the EM algorithm is subject to several problems such as slow convergence and convergence towards local solutions (Celeux et al., 1996; Dempster et al., 1977). Because of it, many researchers have proposed improvements for the traditional method. Celeux and Diebolt (1985) present a stochastic EM algorithm to fit a finite mixture model. Malefaki et al. (2010) propose a stochastic version of the EM algorithm that achieves comparable estimates with the EM algorithm in less execution time for hidden semi-Markov models. Considering NHMM, Di Mari et al. (2016) propose a generalization in 3-steps of the original EM algorithm. Maruotti et al. (2017) model multiple air pollutant exposures and identifies different regimes through a three-step Alternating Expected Conditional Maximization (AECM) algorithm.

Under Bayesian perspective, NHMMs are usually estimated using Metropolis-within-Gibbs Markov chain Monte Carlo (MCMC) algorithms. Shirley et al. (2010) use this method for estimating a NHMM with random effects for alcoholism treatment trial data. Shen et al. (2017) present a similar method for detecting differentially methylated regions from methylation array data. Avoiding Metropolis-Hastings steps, which may not present good mixing, Holsclaw et al. (2017) use the Pólya-Gamma data augmentation approach to estimate a NHHM only using Gibbs steps. They apply their model and inference scheme to 30 years of rainfall in India, leading to a number of insights into rainfall-related phenomena in the region.

In both frequentist and Bayesian methods described below, the number of non-observable states is selected using model selection criteria. For estimating and selecting the model simultaneously, without using model selection criteria in a second stage, the Bayesian reversible jump method (Green, 1995) may be used. However, its implementation is not trivial, especially for NHMM which is a complex model, and it usually presents poor mixing. Rydén (2008) and Zuanetti and Milan (2017) propose, respectively, a reversible jump and a data-driven reversible jump for selecting and estimating a HHMM. Meligkotsidou and Dellaportas (2011) describes a reversible jump algorithm for NHMM.

In this manuscript, before analyzing Tegucigalpa climate data, we present a stochastic EM algorithm and a Bayesian method for estimation of NHMM parameters and non-observable state sequence prediction. The stochastic EM, first proposed by Celeux and Diebolt (1985, 1992) to estimate finite mixture models, has been shown to be efficient in avoiding local solutions, simplifying the implementation and presenting faster convergence in several other models. For the Bayesian algorithm, we use slice sampling algorithm (Neal, 2003) to simulate values of transition coefficients through the rjags package (Plummer, 2022) and avoiding Metropolis-Hastings steps. Although these methodologies have already been used for estimating several other statistical models, it is not the case of NHMMs which are still widely fitted by the traditional EM algorithm. Therefore, in addition to adapting these methods to estimate a NHMM, we compare their performance to the EM algorithm (Visser and Speekenbrink, 2010). The selection of the number of non-observable states and relevant variables are done through model selection criteria for all compared methods.

The manuscript is organized as follows. Model specification is described in Section 2 as well as computational details for the stochastic EM and Bayesian algorithms. Section 3 illustrates a performance analysis of the methods in simulated data sets. Section 4 applies the methods to describe rainfall patterns in Tegucigalpa, Honduras, and shows the results. Finally, Section 5 provides some final remarks.

2 Non-Homogeneous Hidden Markov Model

Among the many authors that define HMM, Rabiner and Juang (1986) explains HMMs as a doubly stochastic process with an underlying stochastic process that is not observable, but can only be observed through another stochastic process that produces the sequence of observable values. Altman (2007) goes further and states that HMMs describe the relationship between two stochastic processes: an observable process and an underlying hidden (non-observable) process. The non-observable process is assumed to follow a Markov chain, and the observable data are modeled as being conditionally independent on the sequence of non-observable states.

Here, to define a NHMM, we consider a discrete observable variable, which follows a Binomial distribution since it is the case of the real data set to be analyzed. However, it may be seamlessly adapted to a continuous variable, if necessary. The elements of an NHMM are:

1. A discrete space of non-observable states $\varphi = \{1, \ldots, K\}$;
2. A set of observable values $\omega = \{1, \ldots, O\}$;
3. A random variable S_t which assumes a value in the space of non-observable states φ, at a given time t, for $t = 1, \ldots, T$;
4. A random variable Y_t which assumes a value in the set of observable values ω, at a given time t, for $t = 1, \ldots, T$;
5. An initial probability distribution for the non-observable states $\boldsymbol{p} = \{p_i\}$, such that $p_i = P(S_1 = i)$ and $\sum_{i=1}^{k} p_i = 1$;
6. A vector $\boldsymbol{X}_t = (X_{t1}, \ldots, X_{tD})$, for $t = 1, \ldots, T$, that represents the values of D observed covariates, with $X_{t1} = 1$ for any t, that influence the transition probabilities between non-observable states at times $t - 1$ and t;
7. A coefficient matrix $\boldsymbol{\beta}$ that has entries $\boldsymbol{\beta}_{ij} = (\beta_{ij1}, \ldots, \beta_{ijD})^\top$ which contains D transition coefficients, each associated to an observed covariate. Such that β_{ij1} is considered the intercept accompanied by the covariate $X_{t1} = 1$ for all t;
8. A state transition probability distribution $\boldsymbol{A}_t = \left\{a_{ijt}\right\}$, such that $a_{ijt} = h(\boldsymbol{X}_t\boldsymbol{\beta}_{ij})$ where $h(\cdot)$

is a link function such that $0 \leqslant h(\cdot) \leqslant 1$ and $\sum_{j=1}^{K} h(\boldsymbol{X}_t\boldsymbol{\beta}_{ij}) = 1$. The link function $h(\cdot)$ will be further discussed; and

9. A probability distribution and its parameters for the observations associated to each non-observable state that specifies $P(Y_t = o \mid S_t = j)$.

A typical link function $h(\cdot)$ is the Softmax given by

$$a_{ijt} = h(\boldsymbol{X}_t\boldsymbol{\beta}_{ij}) = \frac{\exp(\boldsymbol{X}_t\boldsymbol{\beta}_{ij})}{\sum_{l=1}^{K}\exp(\boldsymbol{X}_t\boldsymbol{\beta}_{il})}, \tag{1}$$

which, according to Gao and Pavel (2017) and in multi-class logistic regression, maps a vector of covariates to an *a posteriori* probability distribution. However, it introduces a non-identifiability problem in estimating NHMM since different sets of transition coefficients lead to the same transition probabilities. For this reason, we use the multinomial-logistic link function, better known as mlogit, widely used for polytomous logistic regression model. It is a specific case of the Softmax function, in which the transition coefficients related to the first state are fixed with value equal to 0. The mlogit function is given by

$$a_{ijt} = h(\boldsymbol{X}_t\boldsymbol{\beta}_{ij}) = \begin{cases} \frac{1}{1+\sum_{l=2}^{K}\exp(\boldsymbol{X}_t\boldsymbol{\beta}_{il})}, & \text{if } j = 1 \\ \frac{\exp(\boldsymbol{X}_t\boldsymbol{\beta}_{ij})}{1+\sum_{l=2}^{K}\exp(\boldsymbol{X}_t\boldsymbol{\beta}_{il})}, & \text{if } j > 1. \end{cases} \tag{2}$$

Assume that the distribution of $Y_t \mid S_t = j$ is Binomial(n_t, θ_j), where θ_j is the probability of success related to the j-th non-observable state, $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_{K-1}, \theta_K)$, and n_t represents the known number of trials related to time t, $\boldsymbol{n} = (n_1, n_2, \ldots, n_T)$. The complete (augmented) likelihood function for the NHMM model is written as

$$\begin{aligned} \mathcal{L}(\boldsymbol{\phi} \mid \boldsymbol{D}) &= \left\{\prod_{t=1}^{T}\binom{y_t}{n_t}\right\} \\ &\times \prod_{j=1}^{K}\left\{p_j^{\mathbb{I}_{s_1}(j)}\theta_j^{\sum_{t:s_t=j} y_t}(1-\theta_j)^{\sum_{t:s_t=j}(n_t-y_t)}\prod_{i=1}^{K}\prod_{(t:t\geqslant 2,(s_{t-1},s_t)=(i,j))} h(\boldsymbol{X}_t\boldsymbol{\beta}_{ij})\right\}, \end{aligned} \tag{3}$$

where $\boldsymbol{\phi} = (\boldsymbol{\theta}, \boldsymbol{p}, \boldsymbol{\beta})$, $\boldsymbol{D} = (\boldsymbol{y}, \boldsymbol{s}, \boldsymbol{n}, \boldsymbol{X})$, $\boldsymbol{y} = (y_1, \ldots, y_T)$, $\boldsymbol{s} = (s_1, \ldots, s_T)$, $\boldsymbol{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_T)^\top$ and $\mathbb{I}_{s_1}(j)$ represents an indicator function. The log-likelihood function is given by

$$\begin{aligned} \ell(\boldsymbol{\phi} \mid \boldsymbol{D}) &= \sum_{t=1}^{T}\log\binom{n_t}{y_t} + \sum_{j=1}^{K}\Bigg\{\mathbb{I}_{s_1}(j)\log(p_j) + \sum_{t:s_t=j} y_t\log(\theta_j) \\ &+ \sum_{t:s_t=j}(n_t - y_t)\log(1-\theta_j) + \sum_{i=1}^{K}\sum_{(t:t\geqslant 2,(s_{t-1},s_t)=(i,j))}\log(h(\boldsymbol{X}_t\boldsymbol{\beta}_{ij}))\Bigg\}. \end{aligned} \tag{4}$$

2.1 Estimation Procedures

Several approaches have been adopted to perform estimations on NHMMs. Here, we present two computationally efficient algorithms and compare their performance to the EM algorithm's performance. They are the stochastic EM algorithm and a Bayesian algorithm with pre-fixed number of non-observable states, K. The EM algorithm will not be discussed here, but details are described by MacDonald and Zucchini (1997); Zucchini and MacDonald (2009); Maruotti and Rocci (2012); Zucchini et al. (2016).

2.1.1 Stochastic EM Algorithm

The stochastic EM algorithm is an interesting alternative to avoid slow convergence and local solution of the traditional EM algorithm, specially for NHMMs which are more complex models. This is due to the fact that the stochastic EM relies on random draws of the non-observable sequence $\boldsymbol{S}$ at any given iteration, replacing step E of the original algorithm. Maximum likelihood estimates are then calculated for the parameters based on the complete data, which is composed of the observed data $\boldsymbol{y}$ and the randomly drawn $\boldsymbol{s}$. The stochastic EM also significantly simplifies the necessary calculations of the EM and is much easier to implement. Although stochastic EM has already been used for estimating several other mixture models, it is not the case of NHMMs which are still widely fitted by the traditional EM algorithm.

As we have the restriction $\sum_{j=1}^{K} p_j = 1$, we use Lagrange multipliers for maximizing the log-likelihood. The log-likelihood with the Lagrange multiplier is given by

$$\begin{aligned} \ell(\boldsymbol{\phi} \mid \boldsymbol{D}) &= \sum_{t=1}^{T} \log \binom{n_t}{y_t} + \sum_{j=1}^{K} \Bigg\{ \mathbb{I}_{s_1}(j) \log(p_j) + \sum_{t:s_t=j} y_t \log(\theta_j) \\ &+ \sum_{t:s_t=j} (n_t - y_t) \log(1-\theta_j) + \sum_{i=1}^{K} \sum_{(t:t\geqslant 2,(s_{t-1},s_t)=(i,j))} \log(h(\boldsymbol{X}_t \boldsymbol{\beta}_{ij})) \Bigg\} \\ &+ \lambda_0 \Bigg(1 - \sum_{j=1}^{K} p_j \Bigg), \end{aligned} \tag{5}$$

where λ_0 is the Lagrange multiplier. Deriving this equation in terms of each parameter and finding their solution, we obtain the following maximum points

$$\hat{\theta}_j = \frac{\sum_{t:s_t=j} y_t}{\sum_{t:s_t=j} n_t} \tag{6}$$

and

$$\hat{p}_j = \mathbb{I}_{s_1}(j). \tag{7}$$

The maximum points for $\boldsymbol{\beta}_{ij}$ are obtained using numerical methods such as the Nelder-Mead method. This is due to the fact that it is not possible to find a solution for the maximum point of those parameters analytically.

If the sequence of states $\boldsymbol{S}$ was observable as $\boldsymbol{Y}$, the maximum likelihood estimates for $\boldsymbol{\theta}$, $\boldsymbol{p}$ and $\boldsymbol{\beta}$ could be found by only using the observed data and the equations derived above. However, this is not the case, and we have to recur to an iterative method which predicts $\boldsymbol{s}$ at every iteration and finds the values of the parameters which maximize the log-likelihood based on that prediction. This is the stochastic EM algorithm, described below:

1. Assign arbitrary initial values to $\boldsymbol{S}$ and $\boldsymbol{\beta}$;
2. Based on the observed values $\boldsymbol{y}$ and the assigned values of $\boldsymbol{s}$, calculate $\hat{\theta}_j$ and $\hat{p}_j$ through Equations (6) and (7) and $\hat{\boldsymbol{\beta}}_{ij}$s using numerical methods;
3. Based on the updated values of the parameters, update the values of $\boldsymbol{s}$ using Bayes theorem as

$$\begin{aligned} P(S_t = j \mid \cdots) &= \frac{P(S_t = j, Y_t = y_t \mid S_{t-1} = i, \boldsymbol{X}, \boldsymbol{n}, \boldsymbol{\theta}, \boldsymbol{p}, \boldsymbol{\beta})}{P(Y_t = y_t \mid S_{t-1} = i, \boldsymbol{X}, \boldsymbol{n}, \boldsymbol{\theta}, \boldsymbol{p}, \boldsymbol{\beta})} \\ &= \frac{a_{ijt} \binom{n_t}{y_t} \theta_j^{y_t} (1-\theta_j)^{n_t - y_t}}{\sum_{l=1}^{K} a_{ilt} \binom{n_t}{y_t} \theta_l^{y_t} (1-\theta_l)^{n_t - y_t}}, \text{ for } j = 1, \ldots, K, \end{aligned} \tag{8}$$

where $\cdots$ represents all the remaining variables and parameters of the model. Then, S_t is updated through a Discrete distribution with these probabilities, for $t = 1, \ldots, T$ and $j = 1, \ldots, K$;

4. Repeat steps 2 and 3 until convergence is reached. The convergence criterion used is a comparison of the difference in the log-likelihood from one iteration to the next. Upon attaining a difference less than a previously fixed tolerance, the algorithm stops.

2.1.2 Bayesian Algorithm

Although the stochastic EM algorithm provides solutions for some of the disadvantages of the EM algorithm, it is not without setbacks. Celeux et al. (1996) mention that neither the EM algorithm nor the stochastic EM algorithm can be confidently used for some severely overlapping mixtures. They also depend on the occurrence of all possible transitions in every iteration of the algorithm. If all possible transitions are not observed, the algorithms can not estimate the complete model, that is, it can not find the maximum values for $\boldsymbol{\beta}_{ij}$s corresponding to the transitions with no observations and the algorithm stops.

An alternative for NHMM estimation is the use of Bayesian tools such as MCMC algorithms. For the application of these Bayesian algorithms for inference on the NHMM, the posterior distribution of the parameters is given by

$$P(\boldsymbol{\theta}, \boldsymbol{p}, \boldsymbol{\beta} \mid \boldsymbol{y}, \boldsymbol{s}, \boldsymbol{n}, \boldsymbol{X}) \propto P(\boldsymbol{y} \mid \boldsymbol{s}, \boldsymbol{n}, \boldsymbol{X}, \boldsymbol{\theta}, \boldsymbol{p}, \boldsymbol{\beta}) P(\boldsymbol{s} \mid \boldsymbol{n}, \boldsymbol{X}, \boldsymbol{\theta}, \boldsymbol{p}, \boldsymbol{\beta}) P(\boldsymbol{\theta}, \boldsymbol{p}, \boldsymbol{\beta}) \tag{9}$$

where $P(\boldsymbol{y} \mid \boldsymbol{s}, \boldsymbol{n}, \boldsymbol{X}, \boldsymbol{\theta}, \boldsymbol{p}, \boldsymbol{\beta})$ and $P(\boldsymbol{s} \mid \boldsymbol{n}, \boldsymbol{X}, \boldsymbol{\theta}, \boldsymbol{p}, \boldsymbol{\beta})$ make up the complete likelihood function and $P(\boldsymbol{\theta}, \boldsymbol{p}, \boldsymbol{\beta})$ is the joint prior distribution of the parameters. If we assume conditional independence between the parameters, we have that

$$P(\boldsymbol{\theta}, \boldsymbol{p}, \boldsymbol{\beta}) = P(\boldsymbol{p}) \prod_{j=1}^{K} P(\theta_j) \prod_{i=1}^{K} \prod_{j=2}^{K} \prod_{d=1}^{D} P(\beta_{ijd}). \tag{10}$$

For the NHMM, the prior distributions considered for the parameters are

$$\begin{aligned}
&\boldsymbol{p} \sim \text{Dirichlet}(1, \ldots, 1);\\
&\theta_j \sim \text{Beta}(1, 1) \qquad 1 \leqslant j \leqslant K; \text{ and}\\
&\beta_{ijd} \sim \text{Normal}(0, 100) \qquad 1 \leqslant i \leqslant K, 2 \leqslant j \leqslant K, 1 \leqslant d \leqslant D,
\end{aligned}$$

in order to have vague (weakly informative with large variance) prior informations. Then, the full conditional posterior distributions for the parameters are given by

$$\begin{aligned}
&\boldsymbol{p} \mid \cdots \sim \text{Dirichlet}(1 + \mathbb{I}_{s_1}(1), \ldots, 1 + \mathbb{I}_{s_1}(K));\\
&\theta_j \mid \cdots \sim \text{Beta}(1 + \sum_{t:s_t=j} y_t, 1 + \sum_{t:s_t=j} (n_t - y_t)); \text{ and}\\
&\beta_{ijd} \mid \cdots \text{does not have a known and traditional form.}
\end{aligned}$$

We also want to predict the non-observable sequence $\boldsymbol{s}$ which generated the observable sequence, so S_t has a Discrete full condition posterior distribution with probabilities defined as for stochastic EM algorithm.

Using a MCMC methodology, the Bayesian algorithm is defined in the following steps:

1. Assign initial values to $\boldsymbol{s}$, $\boldsymbol{p}$, θ_i and β_{ijd} for $1 \leqslant i \leqslant K$, $2 \leqslant j \leqslant K$ and $1 \leqslant d \leqslant D$;

2. Update $\boldsymbol{p}$ from its full conditional posterior distribution;
3. Update θ_i from its full conditional posterior distribution, for $1 \leqslant i \leqslant K$;
4. Update β_{ijd} from its full conditional posterior distribution for $1 \leqslant i \leqslant K$, $2 \leqslant j \leqslant K$ and $1 \leqslant d \leqslant D$. As this distribution is not available in a known and closed form, we need to use an acceptance-rejection sampling method to simulate values of it. One of the most traditional in Bayesian estimation is the Metropolis-Hastings algorithm. Here, we choose to use the slice sampling method, implemented in rjags package (Plummer, 2022), which usually presents a better mixing and faster convergence than Metropolis-Hastings method. It is also not necessary to include additional non-observable variables, as proposed in Holsclaw et al. (2017), to obtain a closed full conditional posterior distribution for transition coefficients;
5. Update $\boldsymbol{s}$, where s_t is sampled from a Discrete distribution with probabilities for each state defined as for stochastic EM algorithm;
6. Repeat steps 2, 3, 4 and 5 until attaining the equilibrium distribution of the MCMC chain.

In the Bayesian estimation of mixture models, of which HHMM and NHMM are special cases and consider dependence between observations, we usually have the problem of label-switching. For correcting this problem, we use the algorithm called Equivalence Classes Representatives, proposed by Papastamoulis (2014), which exchange the labels for each simulated sequence $\boldsymbol{s}$ in order to minimize classification error with a specified sequence of states. This process is done using the package label.switching (Papastamoulis, 2016), available for R. Other methods for dealing with this problem are discussed by Stephens (2000).

When the number of non-observable states is unknown we may use model selection criteria to choose the best model for both frequentist and Bayesian estimation methods. Here, we suggest to estimate models with different numbers of non-observable states and select the best one using Bayesian information criterion (BIC, Schwarz, 1978) or corrected Akaike information criterion (AICc, Cavanaugh, 1997). The best model is the model with the lowest value of BIC or AICc. We illustrate their use when analyzing the climate data set.

Under Bayesian perspective, we have other model selection criteria which are most used, as Deviance Bayesian Criterion (DIC), Extended Bayesian Information Criterion (EBIC), among others. However, as a means of comparison, we only replace the maximum likelihood estimates with Bayesian point estimates for calculating BIC and AICc for Bayesian fits.

3 Simulation Examples

For comparing the estimation and classification performance of the two proposed algorithms and traditional EM, carried out using the package depmixS4 (Visser and Speekenbrink, 2010), we simulate several data sets based on the following conditions:

1. We set up 3 chain lengths: $T = 300, 600, 1200$;
2. We fix the number of non-observable states at $K = 2, 3$;
3. We use 2 covariates to influence the transition probabilities, such that $X_m \sim N(0, 1)$ for $m = 1, 2$;
4. We simulate a total of $M = 100$ data sets (replications) for each situation and estimate the models using the three procedures.

We establish two situations to effectively test and compare the estimation capabilities and performance of the algorithms. We call the first situation as Situation A. This situation contemplates unfavorable conditions for the estimation algorithms to retrieve estimates of the pa-

rameters. The second situation, Situation B, presents conditions which are favorable for the estimation algorithms to retrieve estimates.

The transition parameters, $\boldsymbol{\beta}$, chosen for Situation A produce transition probabilities which generate a non-observable state chain which is more difficult to predict and that are very similar among the states, while the transition parameters for Situation B produce probabilities which generate a more stable non-observable chain, which is easier to predict.

The success probabilities for the observable variables used in the simulation are the following. When we consider $K = 2$, for Situation A we have that $\theta_1 = 0.45$ and $\theta_2 = 0.55$ and number of trials fixed at $n_t = 50$. For Situation B, we have that $\theta_1 = 0.25$ and $\theta_2 = 0.75$ and a number of trials of $n_t = 500$. When we consider $K = 3$, for Situation A we have that $\theta_1 = 0.40$, $\theta_2 = 0.50$ and $\theta_3 = 0.60$ with a number of trials of $n_t = 300$ and, for Situation B, $\theta_1 = 0.25$, $\theta_2 = 0.50$ and $\theta_2 = 0.75$ fixing the number of trials at $n_t = 500$. These choices reinforce Situation A as a difficult case for estimation and Situation B as an easier case for estimation.

For comparing methods performance we use the mean squared error (MSE) of estimates for every models' parameters and the relative frequency of correct prediction or classification for $\boldsymbol{S}$. The MSE is calculated for each parameter based on its $M = 100$ estimates and the MSE values of transition parameters are shown together through boxplots. Then, each point represented in the boxplots is the calculated MSE for each transition parameter. The classification success rate shown as bar plots is the average of success rates in predicting $\boldsymbol{S}$ for the $M = 100$ replications. We also verify the standard deviation and bias of the estimates, but as they are in agreement with MSE we do not show these results.

The MSE for the success probabilities associated with Binomial components will be discussed briefly, given that all algorithms show excellent performance in most cases. We omit EM algorithm results for the MSE because its performance is poor, in some cases obtaining values which are two orders-of-magnitude greater than the values of the other two algorithms. This poor performance of traditional EM method, especially when compared to the stochastic EM performance since they have the same target distribution, probably is due to a convergence to local solutions in many replications.

3.1 Results

The Bayesian algorithm is run for a total of 55,000 iterations. The first 5,000 iterations are discarded as burn in. Samples of MCMC are collected every 5 iterations. The sample size of the final sub-sequence used to perform inference is 10,000. As usually done in Bayesian analysis, point estimates for the parameters are their posterior means, given by the average of the values of the sub-sequence after label-switching correction. We verify the convergence of MCMC using trace plots and Geweke's statistic for the log-likelihood value in each iteration. For the stochastic EM and original EM algorithms the tolerance is set to 0.00001 to ensure the convergence of the algorithm. The starting values for all methods are randomly chosen.

We observe in Figures 1–4 that the performance of the Bayesian and stochastic EM algorithms in the estimation of the parameters is similar under analyzed conditions. For most cases shown, the stochastic EM has lower values of MSE than those presented by the Bayesian method. In terms of predicting the non-observable state sequence, the Bayesian method achieves almost 100% accuracy in all situations considered and the stochastic EM algorithm has a similar performance. The EM algorithm performs poorly in terms of predicting the non-observable state sequence, also as a reflection of a poor parameters estimation.

The success probabilities for the Binomial distributions which generate the observable values

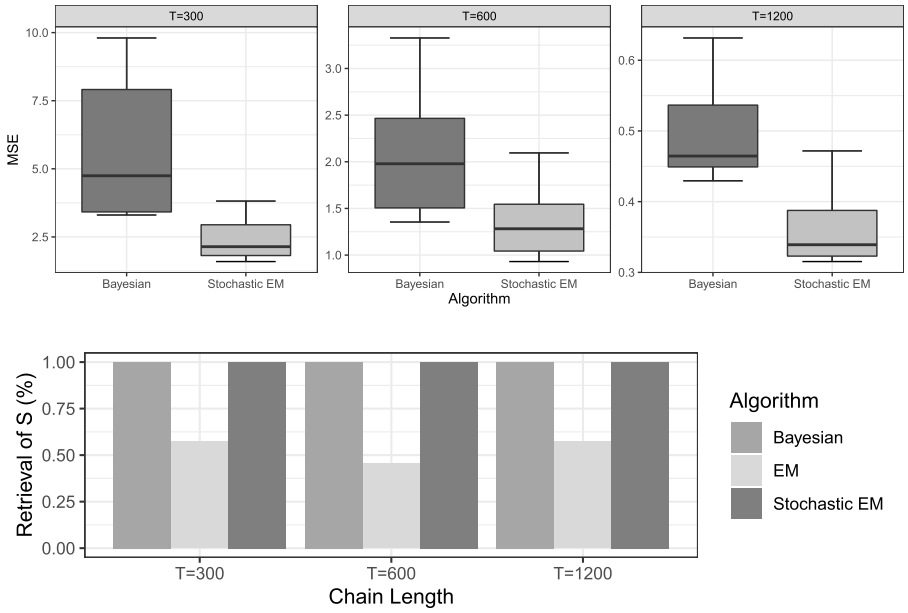

Figure 1: Mean squared errors and success rates for predicting $\boldsymbol{S}$ under Situation A and $K = 2$.

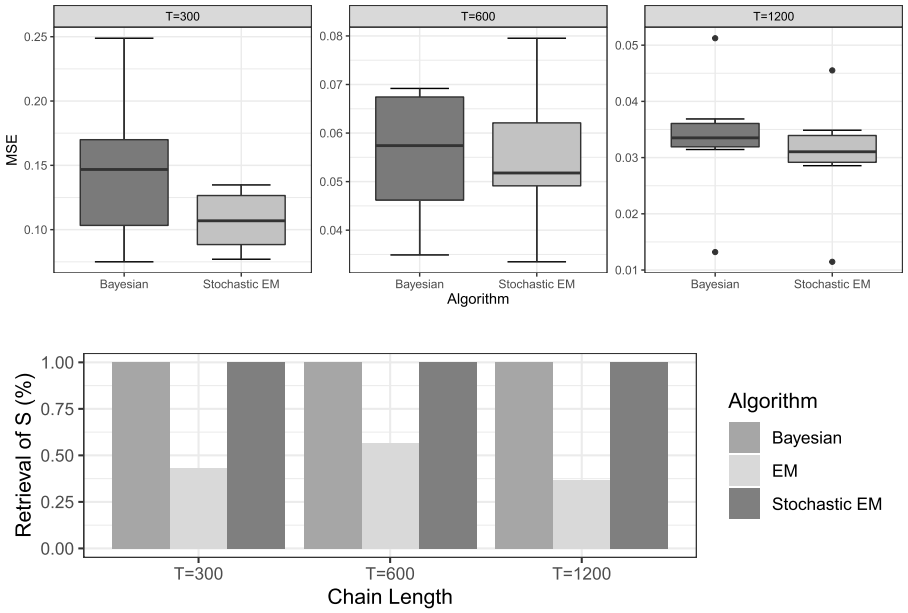

Figure 2: Mean squared errors and success rates for predicting $\boldsymbol{S}$ under Situation B and $K = 2$.

are well estimated by the three algorithms, presenting MSEs often in the thousandths decimal place, and even smaller for some specific cases. The only situation in which success probabilities are not well estimated was Situation A for $K = 2$, when the fixed success probabilities are close and the used number of trials $n_t = 50$ is too small to differentiate them.

Weaker performances of Bayesian algorithm when compared to the stochastic EM usually happen for parameters associated to transitions which are not frequently observed in the data

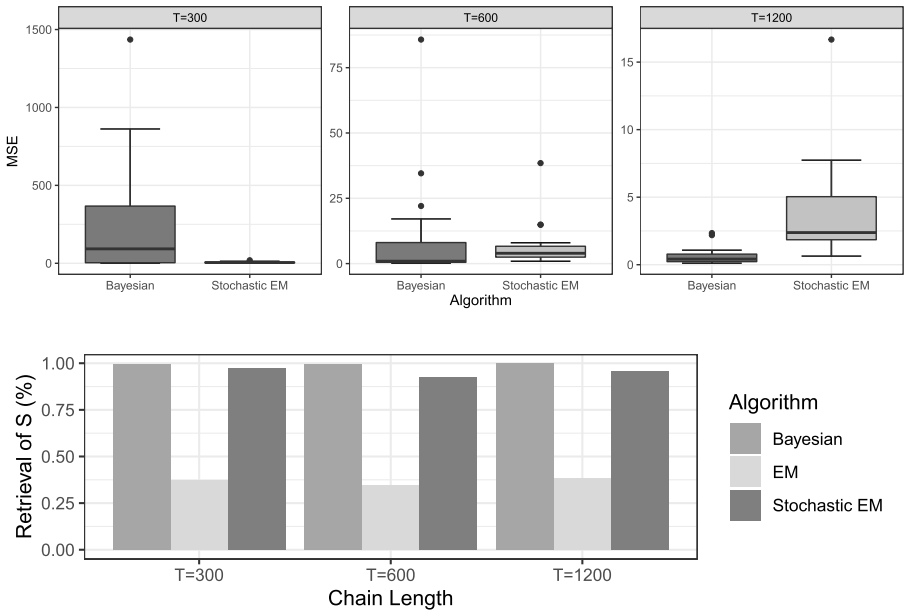

Figure 3: Mean squared errors and success rates for predicting $\boldsymbol{S}$ under Situation A and $K = 3$.

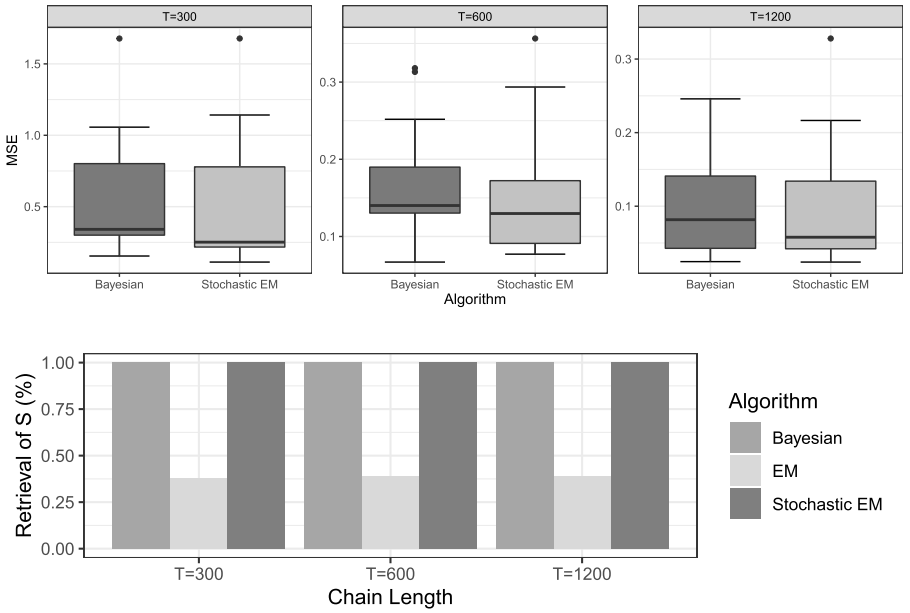

Figure 4: Mean squared errors and success rates for predicting $\boldsymbol{S}$ under Situation B and $K = 3$.

and their estimation is based almost completely on the prior distribution which intentionally has a high variability. In spite of the fact that the Bayesian algorithm shows slightly poorer performance than the stochastic EM, it has the advantage of providing interval estimates through credibility intervals for all the models' parameters. This is not true for the traditional and stochastic EM algorithms which, in general, provide confidence intervals through bootstrap or asymptotic methods. Stochastic EM, in contrast, shows lower computational time cost for

running, especially in case of fast convergence, and may be more efficient for high dimensional time series.

4 Analyzing the Rainfall Patterns of Tegucigalpa

In this section, using the algorithms previously described, we model the rainfall regime and rain probabilities associated with it in the city of Tegucigalpa, Honduras, in the time period between 1969 and 2019. The most recent data is used to verify the predictive performance of the fitted and chosen model later on.

The data was retrieved from the database of Tutiempo Network, S.L. The data was recorded daily at the weather station of Toncontin airport in Tegucigalpa. The location of this weather station is 14.05 N and 87.22 W and its WMO (World Meteorological Organization) identification number is 787200. Therefore, the studied data consists of weather time series from this only one station. It collects data on daily precipitation, relative humidity and temperature. However, as our goal is to analyze the precipitation behavior among the months, we aggregate the daily informations to obtain monthly features. Considering as a rainy day a day when the amount of observed precipitation is greater than 0.5 mm, the response variable analyzed for each month from Jan/1969 to Oct/2019 is the quantity of rainy days. As covariates for impacting the transition probabilities among periods of different chance of raining, we use the average temperature and average relative humidity for each month.

The database contains 610 observations of the response variable and the two covariates used and their descriptive statistics are shown in Table 1. We observe that the city of Tegucigalpa does not present any month without rain in the analyzed period and the monthly average temperature is reasonably stable. As it is expected for being close to the Equator line, the weather of this city does not vary much throughout the year.

We standardized the covariates and apply the three algorithms presented in earlier sections to the data set. We perform model fitting considering different values for K non-observable states as, in fact, we do not know the correct value of K. We fit three different models for the transition probabilities. Some of them include a single covariate, for which we alternate the use of the variables temperature and humidity. Other models uses both variables as covariates for the transition probabilities. The distribution we define for the data is Binomial(N, θ_i) for $i \in \{1, 2, \ldots, K-1, K\}$ and $N = 30$.

The Bayesian algorithm is run for a total of 88,000 iterations. The first 8,000 iterations are discarded as burn in. Samples of MCMC are collected every 8 iterations and, consequently, the sample size of the final sub-sequence used to perform inference is 10,000. We verify the convergence of MCMC using trace plots and Geweke's statistic for the log-likelihood value in each iteration. As usually done in Bayesian analysis, point estimate for the parameters is their posterior means, given by the average of the values of the sub-sequence. For the stochastic EM

Table 1: Descriptive statistics for the variables.

Variable	**Min.**	**Q1**	**Mean**	**Q3**	**Max.**	**SD**	**Asymmetry**
Number of rainy days (days/month)	1.00	7.00	13.00	19.00	29.00	6.38	0.14
Temperature (°C)	18.20	21.00	22.03	23.00	25.90	1.44	−0.30
Humidity (%)	43.50	65.25	70.24	75.78	90.80	7.80	−0.52

Table 2: Parameters estimates for the selected model, where SD represents the standard deviation and CI represents credibility interval.

Parameter	Estimate	SD	Asymmetry	95% CI
β_{121}	−2.61	1.25	−0.63	(−5.45, −0.53)
β_{122}	4.85	1.13	0.65	(2.97, 7.36)
β_{123}	10.60	1.80	0.47	(7.45, 14.58)
β_{131}	1.59	0.38	0.24	(0.89, 2.37)
β_{132}	1.39	0.28	0.19	(0.87, 1.98)
β_{133}	3.97	0.61	0.34	(2.87, 5.27)
β_{221}	7.75	3.33	0.88	(2.78, 15.60)
β_{222}	10.56	2.77	0.88	(6.44, 17.16)
β_{223}	8.27	2.56	0.81	(4.29, 14.11)
β_{231}	8.69	3.32	0.88	(3.73, 16.53)
β_{232}	7.18	2.71	0.93	(3.22, 13.63)
β_{233}	5.59	2.52	0.85	(1.73, 11.44)
β_{321}	−0.60	0.37	−0.20	(−1.34, 0.08)
β_{322}	2.72	0.46	0.33	(1.89, 3.67)
β_{323}	4.73	0.70	0.28	(3.44, 6.19)
β_{331}	0.15	0.20	−0.02	(−0.39, 0.68)
β_{332}	0.04	0.23	−0.01	(−0.40, 0.47)
β_{333}	1.66	0.38	0.24	(0.97, 2.44)
θ_1	0.20	0.01	0.09	(0.18, 0.21)
θ_2	0.66	0.01	−0.02	(0.65, 0.67)
θ_3	0.42	0.01	−0.001	(0.40, 0.43)

algorithm the tolerance is set to 0.00001 to ensure the convergence of the algorithm.

Under the conditions previously established and using both AICc and BIC for selecting the best model estimated for each method, the chosen model using the traditional EM has $K = 3$ regimes, considers only the temperature as covariate and presents AICc and BIC values of 3,802.5 and 3,867.9, respectively. The best model for stochastic EM algorithm also considers only the temperature as covariate, but has $K = 2$ regimes and AICc and BIC values equal to 3,833.3 and 3,859.7, respectively. Under Bayesian estimation procedure, the chosen model has $K = 3$ regimes, both temperature and humidity as covariates and shows AICc and BIC values equal to 3,611.3 and 3,702.4, respectively.

Considering the two model selection criteria, we choose the model with two covariates and three non-observable states fitted by the Bayesian algorithm as the best fit for the data set. Table 2 shows the parameters estimate for the selected model. As we choose the Bayesian estimation, in addition to point estimation we also present interval estimation.

From Table 2 we observe that the 3 non-observable states are well identified by the estimation algorithm. The first non-observable state generates binomially distributed observations with probability of success equal to 0.20. This can be interpreted as a state in which there is a low probability of observing days with precipitation for the given month. A second state generates binomially distributed observations with probability of success equal to 0.66. This can be

interpreted as a state in which there is a high probability of observing days with precipitation during the corresponding month. Argeñal (2010) mentions that there are two clearly observable seasons in the central mountainous region of Honduras, where Tegucigalpa is located. These seasons are commonly called dry season and rainy season. The two non-observable states previously described account for these two seasons. The third non-observable state generates binomially distributed observations with probability of success equal to 0.42. This state represents transitional periods between the rainy and dry seasons in which there are intermediate probabilities of observing rainy days during a given month.

The transition coefficient matrix for the model with the best fit is given in Equation (11) by

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} & \begin{pmatrix} -2.61 \\ 4.85 \\ 10.59 \end{pmatrix} & \begin{pmatrix} 1.59 \\ 1.39 \\ 3.97 \end{pmatrix} \\ \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} & \begin{pmatrix} 7.75 \\ 10.56 \\ 8.27 \end{pmatrix} & \begin{pmatrix} 8.69 \\ 7.18 \\ 5.59 \end{pmatrix} \\ \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} & \begin{pmatrix} -0.60 \\ 2.72 \\ 4.73 \end{pmatrix} & \begin{pmatrix} 0.15 \\ 0.04 \\ 1.66 \end{pmatrix} \end{bmatrix}. \tag{11}$$

In order to verify if the fitted transition coefficient matrix generates probabilities which are coherent with the precipitation phenomenon, we perform a graphical analysis of the transitions functions they are used in. Figure 5 shows the graphs of the transition functions when the covariate humidity is fixed at a value equal to 0.

The top graph of Figure 5 shows that while temperatures are much lower than the mean of the recorded temperatures, there is a high probability of the non-observable chain continuing in state 1, a state with low probability of precipitation. As the temperatures rise and become closer

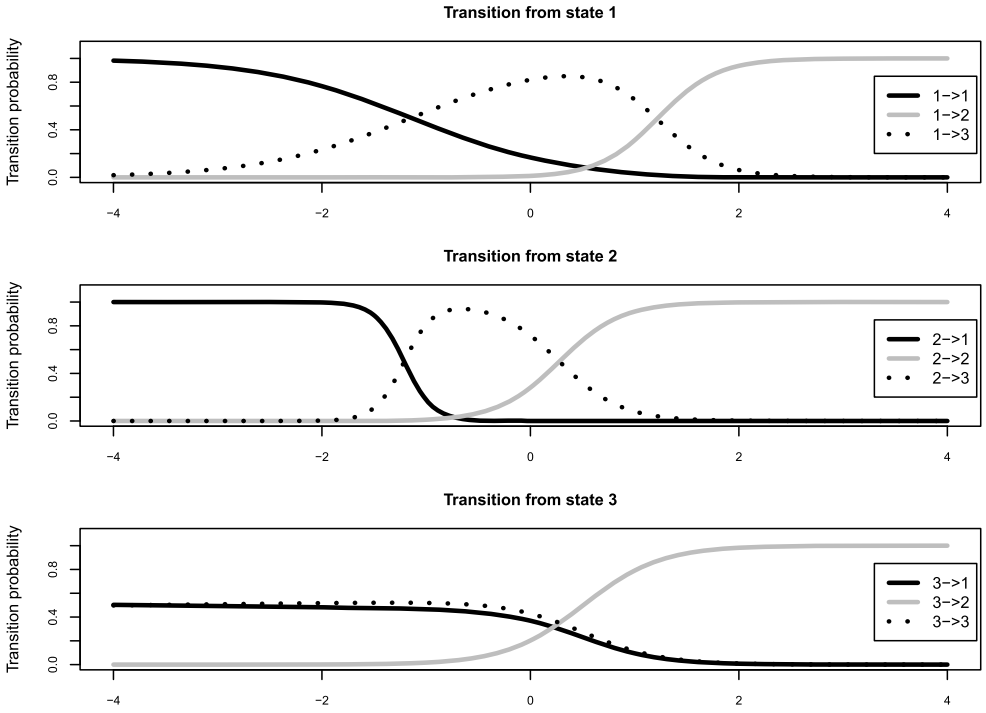

Figure 5: Transition functions when humidity is fixed to zero. On the x axis, we vary the standardized temperature.

to the mean of the observed temperatures, the probability of moving from state 1 to state 3, a state with intermediate probability of precipitation, increases. As the temperature increases and becomes greater than the mean of the recorded temperatures, then the probability of moving from state 1 to state 2, a state with high probability of precipitation, increases.

The middle graph of Figure 5 shows that while temperatures are much lower than the mean of the recorded temperatures, there is a high probability of the non-observable chain moving from state 2 to state 1, a state of low probability of precipitation. When the temperature is close to the mean of the recorded temperatures, there is a high probability of moving from state 2 to state 3, a state with intermediate probability of precipitation. As the temperatures become greater than the mean of the recorded temperatures, then the probability of remaining in state 2 increases.

The bottom graph of Figure 5 shows that while temperatures are lower than the mean of the recorded temperatures, then it is almost equally likely to transition from state 3 to state 1, a state with a low probability of precipitation, or to continue in state 3, a state with intermediate probability of rain. When temperatures increase and become greater than the mean of the observed temperatures, then the probability of moving to state 2, a state with high probability of precipitation, increases.

Figure 6 shows the graphs of the transition functions when the covariate temperature is fixed at a value equal to 0. In general, the behavior of the transition functions when the temperature is fixed is very similar to the behavior of the transition functions when the humidity is fixed. We observe the only difference in behavior in the transitions for state 3, seen in the bottom graph of Figure 6. When the humidity is lower than the mean of the recorded humidities, there is a high

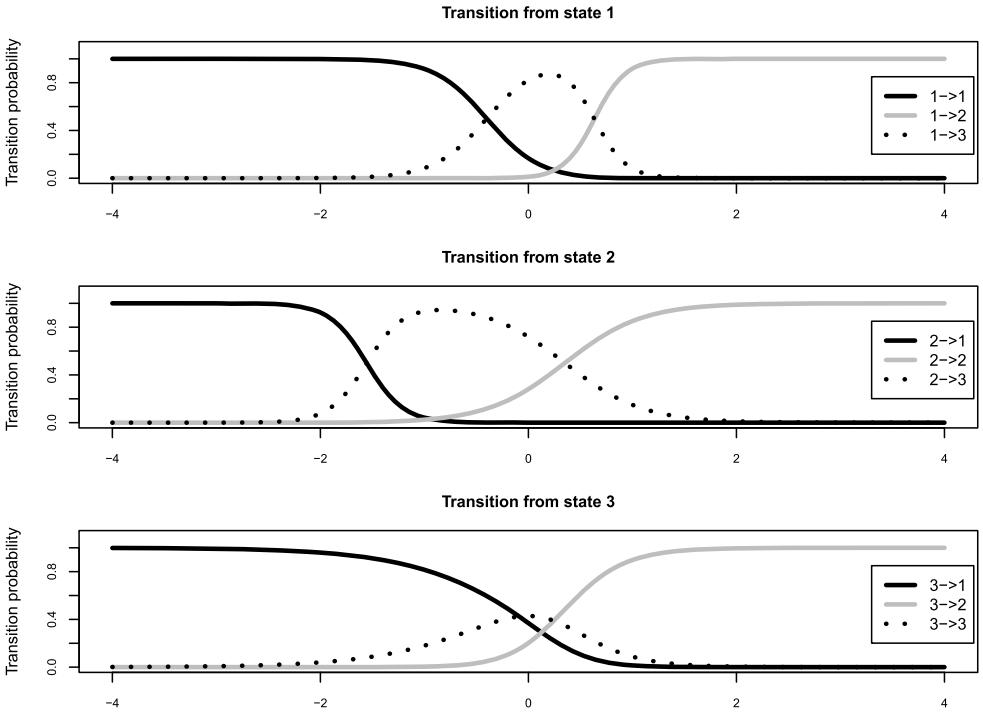

Figure 6: Transition functions when temperature is fixed to zero. On the x axis, we vary the standardized humidity.

probability of moving from state 3 to state 1, a state with low probability of precipitation. When the humidity is higher than the mean of the recorded humidities, there is a high probability of moving to state 2, a state with high probability of precipitation. Finally, when humidity is close to the mean of the observed humidities, it is almost equally like to transition to state 1 or 2, or to remain in state 3.

The transition dynamics previously described are coherent with the behavior of the precipitation patterns in the city of Tegucigalpa. As stated before, there are two clearly observable seasons: a rainy season, in which high temperatures and humidities prevail, along with a high probability of observing days with precipitation and a dry season, in which low temperatures and relatively low humidities are usually recorded, along with a low probability of observing days with precipitation. A brief transitional period between the rainy and dry seasons with intermediate probabilities of rain is related to values of temperature and humidity which are closer to the mean value of the records and accounts for the third non-observable state.

A useful tool to better understand and to visualize the precipitation phenomenon in Tegucigalpa is a graph portraying the sequence of non-observable states predicted by the algorithm along with the number of days with precipitation observed monthly. An online application developed in R Shiny is available at https://gsabillon85.shinyapps.io/PrecipitationNHMM/ to show details about these results. The application allows the user to view predicted states and data along with covariates values for each individual year for the time period studied. In general, we distinguish two periods with distinct rain probabilities. We observe that the months of February, March and April are usually linked to the non-observable state with low probability of rain. The months of June to October are mostly related to the non-observable state with high probability of rain. The non-observable states associated to the months of November, December, January and May are slightly more heterogeneous with a certain prevalence of the non-observable state with intermediate probability of rain.

From the predicted sequence of non-observable states we obtain the matrix

$$\begin{bmatrix} 128 & 7 & 63 \\ 12 & 147 & 64 \\ 59 & 68 & 61 \end{bmatrix} \tag{12}$$

which contains counts of the amount of times each possible transition has occurred. From Expression (12) we observe that the transitions that are least likely to occur are transitions between states 1 and 2. This is coherent with the phenomenon being studied, as it is unlikely to move directly from a rainy season to a dry season, or vice-versa, without experiencing a small time interval with intermediate amounts of rain. We also observe that the most frequent transitions involve remaining in state 1 or state 2. This is consistent with the phenomenon, because remaining in state 1 indicates a time period of low probability of precipitation, which reflects a dry season, and remaining in state 2 indicates a time period of high probability of rain, which would portray a rainy season.

We also verify the performance of the fitted and chosen model to predict the number of rainy days in the period from Nov/19 to Nov/22 in a total of 37 forecast points for the last three years. Using the parameters' estimate and observed average temperature and average relative humidity for each month t in this most recent period, we calculate $\hat{a}_{\hat{s}_{t-1}jt}$, for $j = 1, 2, 3$, and predict $\hat{y}_t = \hat{E}(Y_t|S_{t-1} = \hat{s}_{t-1}) = n_t \sum_{j=1}^{3} \hat{a}_{\hat{s}_{t-1}jt}\hat{\theta}_j$ and $\hat{s}_t = \arg_j \max \hat{a}_{\hat{s}_{t-1}jt}$. This predicted $\hat{s}_t$ is used to calculate $\hat{a}_{\hat{s}_t jt}$ for the observation $t + 1$ and so on. We also predicted $\hat{s}_t$ by randomly drawing from probabilities $\hat{a}_{\hat{s}_{t-1}jt}$, for $j = 1, 2, 3$, but the results are very similar.

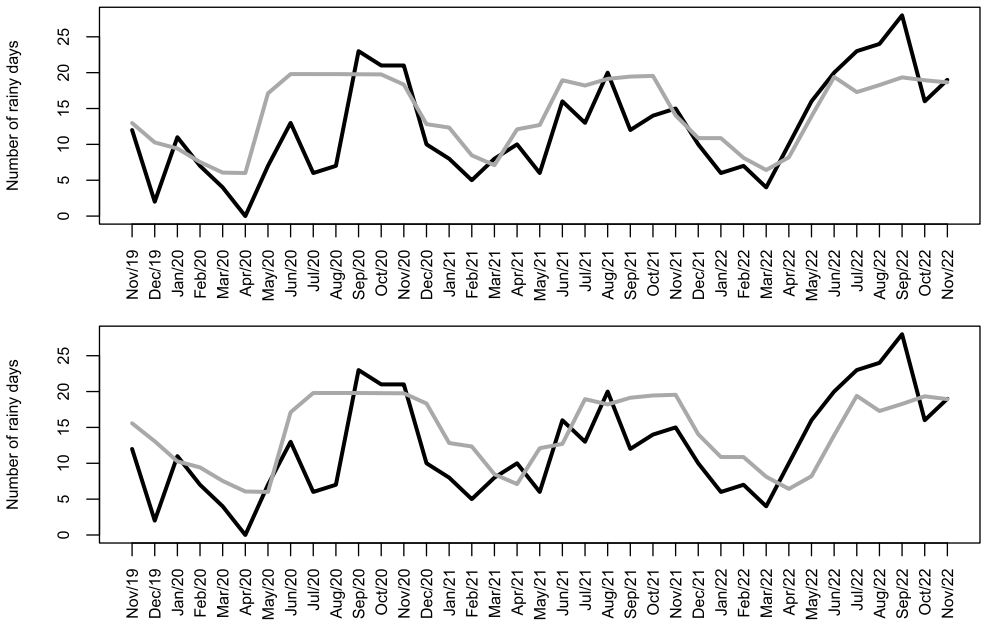

Figure 7: Observed (black) and predicted (gray) number of rainy days for each month from Nov/19 to Nov/22. The top graph considers covariates of the month of prediction, as we carry out the estimation, and the bottom graph considers covariates of the month before the prediction month, that is, covariates lagged by one month.

The top graph of Figure 7 shows the observed and predicted number of rainy days for each month using the observed values of covariates in the month of prediction. We carry out the model estimation in this way since our main goal is to understand and describe the rainfall patterns and not to make longer predictions, for which covariates values are not yet available (but forecasts for them may be used). We observe that the fitted model correctly predict months with the most and least amount of rainy days as well as the number of rainy days for each month. The mean squared error associated with those 37 predictions is 27.99.

The bottom graph of Figure 7 shows the predicted number of rainy days for each month using the covariates lagged by one month. It is useful when we want to predict the number of rainy days in advance (for the next month, in this case) using observed values of covariates for the current month since the informations for the next month are not yet available. We observe that, even if the model has not been fitted for this scenario, it still performs and predicts well. Now, the mean squared error associated with those predictions is 34.22.

5 Discussion

We study the structure, elements and functions of NHMMs in order to apply them for describing the rainfall pattern in Honduras. We present two algorithms, the stochastic EM and Bayesian algorithm, as more efficient alternatives than the EM algorithm for estimating NHMMs in terms of faster convergence, easier implementation and scalability for high dimensional data. Comparisons between these three algorithms show that the performance of the stochastic EM and Bayesian algorithms is similar, whereas the EM algorithm shows poor performance in terms of estimating

the transition parameters, mainly due to local solutions. Similar values for mean squared error are obtained by the stochastic EM and Bayesian algorithms in simulated situations, with the stochastic EM having slightly better performance in some cases.

Despite showing slightly poorer performance, the Bayesian method has the advantage of providing interval estimates, through credibility intervals, for all models' parameters whereas for traditional and stochastic EM it would be possible using bootstrap or asymptotic methods. Stochastic EM, in contrast, shows lower computational time cost for running which makes the method applicable to high dimensional data. However, it is mandatory to have observations allocated in all possible transitions for the algorithm to work properly. We observe a success rate greater than 90% in the prediction of the non-observable state sequence for these two algorithms in all considered simulations, while the traditional EM shows a rate no greater than 56% in any simulation. Although the presented methodologies have already been used for estimating several other statistical models, NHMMs are still widely fitted by the traditional EM algorithm.

We apply NHMMs and the three algorithms to 50 years of rainfall data collected in Tegucigalpa, Honduras. The best model, under considered conditions, is fitted by the Bayesian algorithm, considers 3 non-observable states and the two covariates tested, temperature and humidity. The algorithm clearly identifies 3 non-observable states. One non-observable state represents periods with a low probability of observing days with precipitation, the second non-observable state represents periods with a high probability of precipitation, and the third state represents periods with an intermediate probability of days with precipitation per month (0.42). Both variables (temperature and humidity) are important to determine the rain season and the results are consistent with the studied phenomenon. In general, an NHMM models the phenomenon well. The fitted model may also be used to predict future informations of precipitation depending on future informations of temperature and humidity. If available, many other weather and environmental features may be used in the analysis.

In this case study, we only have two covariates available and an automatic and efficient variable selection process was not necessary. However, as future studies, variable selection methods may be included in the estimation process to make it more efficient.

Supplementary Material

Data that support the findings are openly available in GitHub at https://github.com/gsabillon85/NHMM-Estimation and https://gsabillon85.shinyapps.io/PrecipitationNHMM/.

The R codes used for simulation are openly available in GitHub at https://github.com/gsabillon85/NHMM-Estimation.

References

Altman RM (2007). Mixed hidden Markov models: An extension of the hidden Markov model to the longitudinal data setting. *Journal of the American Statistical Association*, 102(477): 201–210. https://doi.org/10.1198/016214506000001086

Argeñal F (2010). Variabilidad climática y cambio climático en Honduras. In: *Secretaria e Recursos Naturales y Ambiente and Programa de las Naciones Unidas para el Desarrollo PNUD*.

Betrò B, Bodini A, Cossu QA (2008). Using a hidden Markov model to analyse extreme rainfall events in Central-East Sardinia. *Environmetrics*, 19(7): 702–713. https://doi.org/10.1002/env.944

Cavanaugh JE (1997). Unifying the derivations for the Akaike and corrected Akaike information criteria. *Statistics & Probability Letters*, 33(2): 201–208. https://doi.org/10.1016/S0167-7152(96)00128-9

Celeux G, Chauveau D, Diebolt J (1996). Stochastic versions of the EM algorithm: An experimental study in the mixture case. *Journal of Statistical Computation and Simulation*, 55(4): 287–314. https://doi.org/10.1080/00949659608811772

Celeux G, Diebolt J (1985). The SEM algorithm: A probabilistic teacher algorithm derived from the EM algorithm for the mixture problem. *Computational Statistics Quarterly*, 2: 73–82.

Celeux G, Diebolt J (1992). A stochastic approximation type EM algorithm for the mixture problem. *Stochastics: An International Journal of Probability and Stochastic Processes*, 41(1–2): 119–134.

Dempster AP, Laird NM, Rubin DB (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 39(1): 1–38.

Di Mari R, Oberski DL, Vermunt JK (2016). Bias-adjusted three-step latent Markov modeling with covariates. *Structural Equation Modeling: A Multidisciplinary Journal*, 23(5): 649–660. https://doi.org/10.1080/10705511.2016.1191015

Gao B, Pavel L (2017). On the properties of the Softmax function with application in game theory and reinforcement learning. arXiv preprint: https://arxiv.org/abs/1704.00805

Ghavidel FZ, Claesen J, Burzykowski T (2015). A non-homogeneous hidden Markov model for gene mapping based on next-generation sequencing data. *Journal of Computational Biology*, 22(2): 178–188. https://doi.org/10.1089/cmb.2014.0258

Green PJ (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82(4): 711–732. https://doi.org/10.1093/biomet/82.4.711

Holsclaw T, Greene AM, Robertson AW, Smyth P, et al. (2017). Bayesian nonhomogeneous Markov models via Pólya-gamma data augmentation with applications to rainfall modeling. *The Annals of Applied Statistics*, 11(1): 393–426. https://doi.org/10.1214/16-AOAS1009

Lagona F, Maruotti A, Picone M (2011). A non-homogeneous hidden Markov model for the analysis of multi-pollutant exceedances data. In: *Hidden Markov Models, Theory and Applications* (Przemyslaw Dymarski, ed.), 207–222. Intech, Rijeka, Croatia.

MacDonald IL, Zucchini W (1997). *Hidden Markov and Other Models for Discrete-Valued Time Series*, volume 110. Chapman and Hall, London.

Malefaki S, Trevezas S, Limnios N (2010). An EM and a stochastic version of the EM algorithm for nonparametric hidden semi-Markov models. *Communications in Statistics: Simulation and Computation*, 39(2): 240–261. https://doi.org/10.1080/03610910903411185

Maruotti A, Bulla J, Lagona F, Picone M, Martella F (2017). Dynamic mixtures of factor analyzers to characterize multivariate air pollutant exposures. *The Annals of Applied Statistics*, 11(3): 1617–1648. https://doi.org/10.1214/17-AOAS1049

Maruotti A, Rocci R (2012). A mixed non-homogeneous hidden Markov model for categorical data, with application to alcohol consumption. *Statistics in Medicine*, 31(9): 871–886. https://doi.org/10.1002/sim.4478

Meligkotsidou L, Dellaportas P (2011). Forecasting with non-homogeneous hidden Markov models. *Statistics and Computing*, 21(3): 439–449. https://doi.org/10.1007/s11222-010-9180-5

Neal RM (2003). Slice sampling. *Annals of Statistics*, 31(3): 705–767. https://doi.org/10.1214/aos/1056562461

Neykov N, Neytchev P, Zucchini W, Hristov H (2012). Linking atmospheric circulation to daily

precipitation patterns over the territory of Bulgaria. *Environmental and Ecological Statistics*, 19(2): 249–267. https://doi.org/10.1007/s10651-011-0185-9

Papastamoulis P (2014). Handling the label switching problem in latent class models via the ECR algorithm. *Communications in Statistics: Simulation and Computation*, 43(4): 913–927. https://doi.org/10.1080/03610918.2012.718840

Papastamoulis P (2016). label.switching: An R package for dealing with the label switching problem in MCMC outputs. *Journal of Statistical Software, Code Snippets*, 69(1): 1–24.

Pennoni F, Genge E (2020). Analysing the course of public trust via hidden Markov models: A focus on the Polish society. *Statistical Methods & Applications*, 29: 399–425. https://doi.org/10.1007/s10260-019-00483-9

Plummer M (2022). *rjags: Bayesian Graphical Models Using MCMC*. R package version 4-13.

Rabiner L, Juang B (1986). An introduction to hidden Markov models. *IEEE ASSP Magazine*, 3(1): 4–16. https://doi.org/10.1109/MASSP.1986.1165342

Rabiner LR (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2): 257–286. https://doi.org/10.1109/5.18626

Robertson AW, Kirshner S, Smyth P (2004). Downscaling of daily rainfall occurrence over northeast Brazil using a hidden Markov model. *Journal of Climate*, 17(22): 4407–4424. https://doi.org/10.1175/JCLI-3216.1

Rydén T (2008). EM versus Markov chain Monte Carlo for estimation of hidden Markov models: A computational perspective. *Bayesian Analysis*, 3(4): 659–688.

Schwarz G (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2): 461–464. https://doi.org/10.1214/aos/1176344136

Shen L, Zhu J, Robert Li SY Fan X (2017). Detect differentially methylated regions using non-homogeneous hidden Markov model for methylation array data. *Bioinformatics*, 33(23): 3701–3708. https://doi.org/10.1093/bioinformatics/btx467

Shirley KE, Small DS, Lynch KG, Maisto SA, Oslin DW (2010). Hidden Markov models for alcoholism treatment trial data. *The Annals of Applied Statistics*, 4(1): 366–395. https://doi.org/10.1214/09-AOAS282

Stephens M (2000). Dealing with label switching in mixture models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(4): 795–809. https://doi.org/10.1111/1467-9868.00265

Visser I, Speekenbrink M (2010). depmixS4: An R package for hidden Markov models. *Journal of Statistical Software*, 36(7): 1–21. https://doi.org/10.18637/jss.v036.i07

Zuanetti DA, Milan LA (2017). A generalized mixture model applied to diabetes incidence data. *Biometrical Journal*, 59(4): 826–842. https://doi.org/10.1002/bimj.201600086

Zucchini W, MacDonald IL (2009). *Hidden Markov Models for Time Series: An Introduction Using R*. Chapman and Hall, London.

Zucchini W, MacDonald IL, Langrock R (2016). *Hidden Markov models for time series: An introduction using R*. Chapman and Hall, Boca Raton.