A NONHOMOGENEOUS POISSON HIDDEN MARKOV MODEL FOR CLAIM COUNTS

BY

YI LU AND LEILEI ZENG

ABSTRACT

We propose a nonhomogeneous Poisson hidden Markov model for a time series of claim counts that accounts for both seasonal variations and random fluctuations in the claims intensity. It assumes that the parameters of the intensity function for the nonhomogeneous Poisson distribution vary according to an (unobserved) underlying Markov chain. This can apply to natural phenomena that evolve in a seasonal environment. For example, hurricanes that are subject to random fluctuations (El Niño-La Niña cycles) affect insurance claims. The Expectation-Maximization (EM) algorithm is used to calculate the maximum likelihood estimators for the parameters of this dynamic Poisson hidden Markov model. Statistical applications of this model to Atlantic hurricanes and tropical storms data are discussed.

KEYWORDS

Poisson hidden Markov model; Nonhomogeneous Poisson process; Seasonality; Intensity function; Hurricanes and tropical storms; Maximum likelihood estimation; EM algorithm.

1. INTRODUCTION

There is a tendency for the cost of hurricane insurance to increase and the coverage to be dropped because of the bigger-than-expected number of hurricane-related claims and the increase in related reinsurance costs. In order for insurers to keep competitive in their property-casualty insurance business, it becomes important to have a better estimation of both the frequency and the severity of damages incurred by hurricanes as well as tropical storms. Pielke et al. (2008) study the total economic damage related to 207 tropical storm and hurricane landfalls along the U.S. Gulf and Atlantic coasts from 1900-2005. According to their two normalization methods, the average annual normalized damage over 106 years is about $10 billion (2005 values), with most recent estimates being over $150 billion in years 2004 and 2005, based on a total of

Astin Bulletin 42(1), 181-202. doi: 10.2143/AST.42.1.2160740

12 hurricanes. This paper intends to discuss the stochastic modeling of claims generated from these catastrophic events, with a focus on the claim counts (number of occurrences).

Homogeneous Poisson processes are commonly used to model the claim counts in the risk theory literature. These sometimes give a crude representation as their claim intensity rate λ is constant over time. Beard et al. (1984) and Daykin et al. (1994) suggest that the risk process is often subject to continual changes in risk propensity. The model to be employed must then suitably define a time-dependent function or a stochastic process for the claims intensity, instead of the constant Poisson parameter λ. Grandell (1991) also points out that the former can be used to model "size fluctuations" in the claim intensity of a risk such as the seasonality, while the latter can be used to characterize the underlying "risk fluctuations" in the claims intensity. The corresponding claim counting processes are the Nonhomogeneous Poisson (NHP) process and the Cox process or the doubly stochastic Poisson process, respectively.

In practice, natural phenomena evolving in a periodic environment, or under seasonal conditions, affect insurance claims. For example, weather factors are known to affect automobile or fire insurance claims, while seasonal snow storms in the north and hurricanes or floods in the south affect property-casualty insurance. A periodic time-dependent intensity rate is a reasonable model for the claim frequency in such situations. Garrido and Lu (2004) propose a NHP process with a doubly periodic intensity rate, where periodicity does not repeat the exact same pattern in each short-term period; rather, its peak intensity varies over a longer period. A double-beta type intensity function was proposed. Lu and Garrido (2005) further derive the Maximum Likelihood Estimation (MLE) of the model parameters, and discuss the application of the model to the dataset of Atlantic hurricanes affecting the United States (1899-2000).

Moreover, Lu and Garrido (2007) propose a regime-switching NHP process which accounts for both, the seasonal variations and the random fluctuations in the claims intensity, in which the intensity process is of the form

$$\lambda(t) = \lambda_S(t)\,\mathbf{q}(t), \quad t \geq 0, \tag{1}$$

where $\lambda_S(t)$ is a deterministic short-term intensity function with periodicity and $\{\mathbf{q}(t);\, t \geq 0\}$ is a stochastic level process governed by a m-state Markov chain. Here, m different levels represent different risk conditions. In practice, such conditions can be slippery roads, foggy days, stormy weather, years affected by the El Niño phenomenon and so on, which affect through the claim frequency the insurance business. Under certain risk classifications, the transition probabilities of underlying Markov chain can be empirically estimated, and the MLEs of other model parameters (level parameters and the parameters for the periodic short-term intensity function) can be obtained through the partial likelihood function.

In some cases when there is no preclassification available for the underlying risks (or environmental factors), the number of states in the Markov chain and transitions between these states cannot be observed. The Poisson Hidden Markov Model (PHMM) can be used to model the dynamics of the claim counts which are affected by unobservable underlying processes. Moreover, it can address some common characteristics of count data such as serial correlation and over-dispersion. In the literature the PHMM has been used in many areas, such as cryptanalysis, speech recognition and bioinformatics (e.g., Leroux and Puterman 1992, Albert 1991, Juang and Rabiner 1991, Cooper and Lipsitch 2004, and Altman and Petkau 2005). Paroli et al. (2000) suggest a PHMM for time series of overdispersed insurance counts in non-life insurance. The MLEs of the model parameters are obtained and the application to the data of the injury frequencies are discussed. Hughes and Guttorp (1999) also suggest a nonhomogeneous hidden Markov model for precipitation occurrences where the time-dependent transition probabilities of the hidden Markov model associate with other atmospheric data at that time.

Various approaches have been proposed for modeling time series of counts, which can account for discreteness, serial correlation and over-dispersion. Depending on how serial dependence structure is introduced, the distinction is often made between observation-driven models, where the lagged values of observed counts are directly incorporated into the mean function, and parameter-driven models, where a latent (unobserved) dynamic process is assumed to govern the conditional mean function (Cox 1981 and Jung et al. 2006).

Popular members of the class of observation-driven models include integer-valued autoregressive moving average (INARMA) models, where a certain type of thinning operation is used to replace the scalar multiplications in the Gaussian ARMA framework for the integer-valued case (see McKenzie 2003 and Weiß 2008 for recent overviews). However, this class of models has limitations in dealing with the presence of over-dispersion and seasonality (Quddus 2008), and the methods of estimation can be very complex and only a limited range of models has been systematically analyzed in terms of their practical applicability (Heinen 2003, Weiß 2009). Another group of observation-driven models, recently proposed by Heinen (2003) and Ferland et al. (2006), are autoregressive conditional Poisson models (ACP) where a Poisson distribution is specified for the counts and their mean is autoregressive conditional on past observations. In contrast to the INARMA models, the ACP models can describe integer-valued process with over-dispersion and are well suited for both point and density forecasts. Note that the ACP(p, q) models are also referred to as the INGARCH(p, q) models (Ferland et al. 2006, and Weiß 2009). This class of models can also account for the seasonal variation in the mean by adding a combination of sine and cosine terms into the link function (Höhle and Paul 2008, and Freeland and McCabe 2004).

In the class of parameter-driven models, Zeger (1988) adopts the generalized linear models framework and introduces a latent multiplicative autoregressive term into the conditional mean function to account for both autocorrelation and over-dispersion. Despite the popularity of this model, especially in biometrics applications, the estimation procedure is not straightforward and often very challenging because the likelihood function requires high-dimension integration. Alternatively, as mentioned earlier, the PHMMs have been used for analyzing time series of count data (MacDonald and Zucchini 1997, and Paroli et al. 2000). It is assumed that the mean of the Poisson distribution changes according to a latent multi-state Markov chain. This class of models allows for over-dispersion relative to a Poisson distribution and for serial correlation between observed counts. It is most appropriate when the serial correlation is thought to arise through a largely unobserved process and there is a natural interpretation for what might constitute a suitable process.

For our particular interests in modeling the Atlantic hurricane and tropical storm counts, it has been suggested that there is a significant connection between the frequency of hurricanes and the El Niño phenomenon (Gray 1984, Pielke and Landsea 1999, and Katz 2002). In addition, the El Niño phenomenon can be classified into one of three states, La Niña, neutral and El Niño, according to Trenberth (1997). We are particularly interested in understanding the dynamic of the hurricane and tropical storm counts and making inferences about the unobserved process generating the autocorrelation. Moreover, the hurricane and tropical storm counts exhibit the repeated patterns of seasonality and over-dispersion which need to be taken into account (details see Section 2). These motivate us to consider the PHMM.

In this paper, we generalize the PHMM by introducing a nonhomogeneous component in the intensity to account for both seasonal variations and random fluctuations for a time series of claim counts. By choosing the continuous short-term intensity functions over piecewise constant ones, our model can better capture the continuous changing external environment. The Expectation Maximization (EM) algorithm is used for the parameters estimation and it is compared to the direct maximization approach. Statistical applications of the model to the dataset of Atlantic tropical storms and hurricanes affecting the U.S. (1899-2007) will be discussed, which would help fine-tune the model and make it more accurate and applicable for use by insurance companies exposed to climatological risks.

The rest of the paper is organized as follows. A brief description of U.S. Atlantic hurricanes and tropical storms dataset is given in Section 2. A Nonhomogeneous Poisson Hidden Markov Model (NPHMM) with a beta-type short-term periodic intensity function is introduced in Section 3. Section 4 presents the MLE of the model parameters through the EM algorithm. The application of the model to the datasets is given in Section 5, followed by concluding remarks given in Section 6.

2. Data and modeling background

Most of our data was obtained from Landreneau (2003), which included 409 hurricanes and tropical storms (168 hurricanes) that crossed or passed immediately adjacent to the Unites States coastline (Texas to Maine), for the years

1899-2002. Additional data for the years 2003-2007 were obtained from the National Oceanic and Atmospheric Administrations (NOAA), Climate Prediction Center web site[1] using the annual Atlantic Hurricane Season Climate Summary which contains 26 additional hurricanes and tropical storms (13 hurricanes) for the years 2003-2007. From the data sources above, we obtain not only the total annual counts but also the monthly counts for which the hurricanes and tropical storms made landfall. Henceforth, we call these two combined datasets "the U.S. Atlantic hurricanes and tropical storms (H&TS) data" and "the U.S. Atlantic hurricanes (HONLY) data", respectively. Thus, over the 109-year period 1899-2007, a total of 435 tropical storms and hurricanes (including 181 categories 1-5 hurricanes) crossed the Atlantic U.S. coastline at one or more points.

The official Atlantic hurricane season runs from June 1 to November 30 according to the NOAA. Note that over the 109 years, no hurricanes occurred in the non-hurricane season, and only two tropical storms occurred in the months of February and May respectively, and one in December. Our focus hence lies in characterizing the intensity of the counts during the official Atlantic hurricane season and excludes those five tropical storms from the analysis. We can easily extend our analysis by modeling the intensity of hurricane and tropical storm counts throughout the whole year and including these five data points; this will not make too much of a difference for the estimations with the intensity level being nearly zero for the time intervals [0, 5/12) and [11/12,1).

The average annual number of hurricanes and tropical storms is 3.99 (1.66 for hurricanes only) over the entire period, which is about an average of four hurricane and tropical storm (one to two hurricanes) landfalls per year. The years with a maximum number of 6 hurricanes were 1916 and 1985, while 21 out of the 109 years had no hurricanes. Furthermore, the hurricane and tropical storm season peaks from mid-August to October, with September having the most major hurricanes and tropical storms (e.g., 39.2% of all hurricanes). In this paper, we are interested in modeling the monthly hurricane and tropical storm data, not the aggregated yearly data. Figure 1 is a run chart of the monthly 435 Atlantic hurricanes and tropical storms (grey lines) with 181 hurricanes (black lines), while Figure 2 plots histograms of the monthly H&TS counts (black bars) and the HONLY counts (grey bars), reporting small counts and high percentages of zeros (61% and 76% months, respectively, without observations). The empirical mean and variance of the two monthly datasets are 0.6651 and 1.1021 for the H&TS data and 0.2768 and 0.2954 for the HONLY data, which suggest that the H&TS counts are over-dispersed compared to the HONLY counts. It should be noted that there is no formal test for over-dispersion seen in the literature when the counts show some dependence. (Tests of over-dispersion for i.i.d. counts can be found, for example, in Rao and Chakravarti 1956.)

1 http://www.cpc.ncep.noaa.gov/products/outlooks/hurricane-archive.shtml

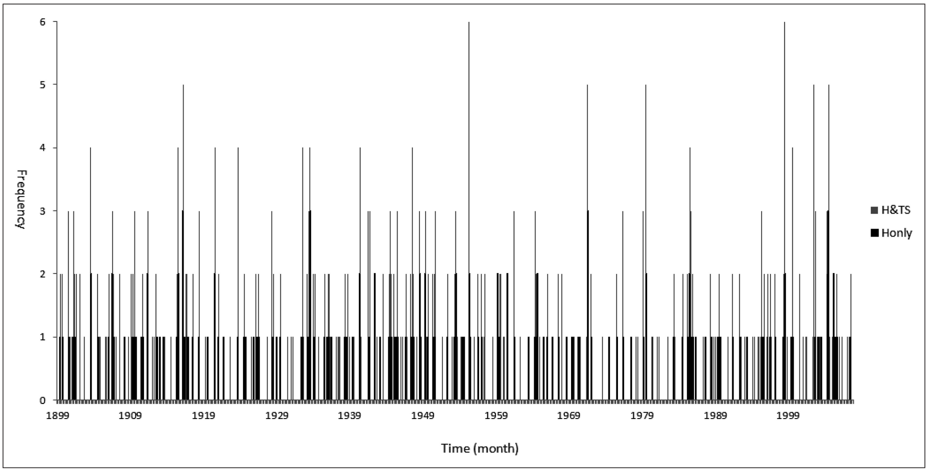

FIGURE 1: The U.S. Atlantic H&TS and the HONLY (1899-2007) monthly counts.

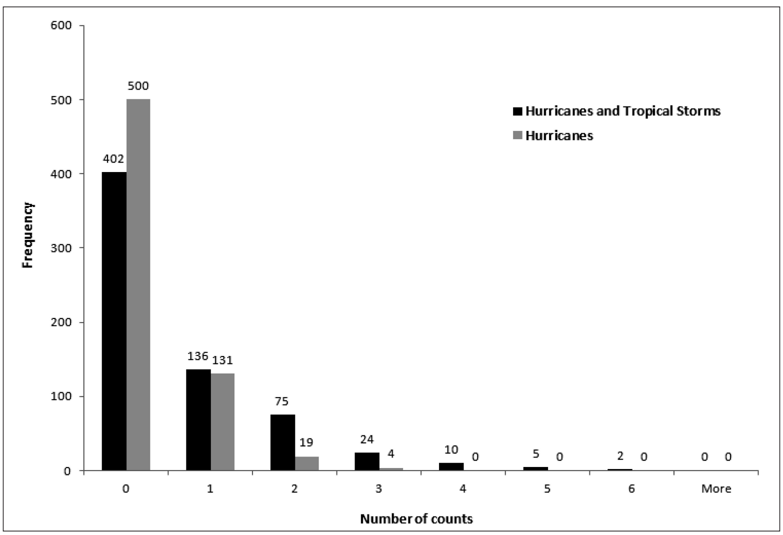

FIGURE 2: The U.S. Atlantic H&TS and the HONLY (1899-2007) monthly histograms.

Figure 3 gives their monthly frequencies (solid bars for hurricane-only data), showing clearly a left-skewed short-term (6-month) intensity pattern. The autocorrelation plots of the monthly counts from two datasets are provided in Figure 4. These graphs reveal significant autocorrelation in counts and a distinct cyclic pattern in the sign of the autocorrelation that alternates between positive and negative. The spectral plot shows a dominant peak at a period of six months which is the length of the official hurricane season (see Figure 5). All these clearly show the violation of randomness assumption and imply that

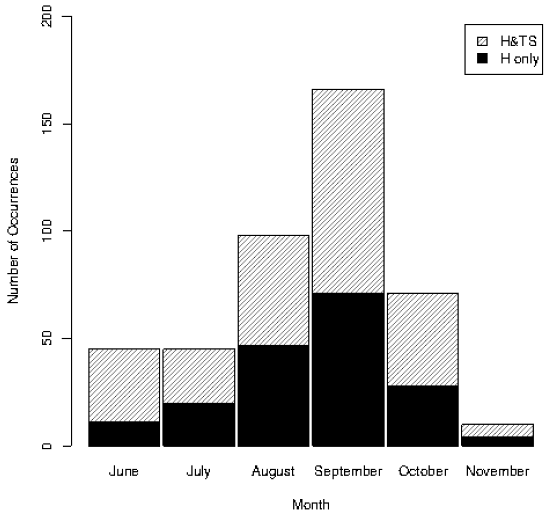

FIGURE 3: The monthly frequencies of the U.S. Atlantic H&TS and the HONLY data.

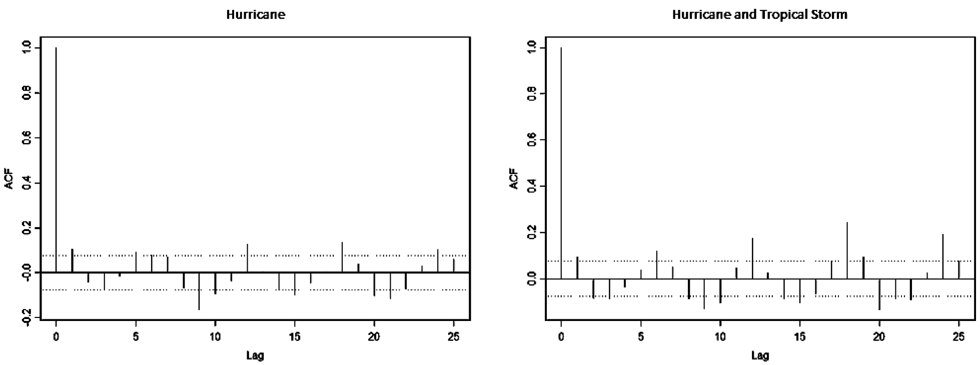

FIGURE 4: The ACFs for the U.S. Atlantic H&TS and the HONLY (monthly) data.

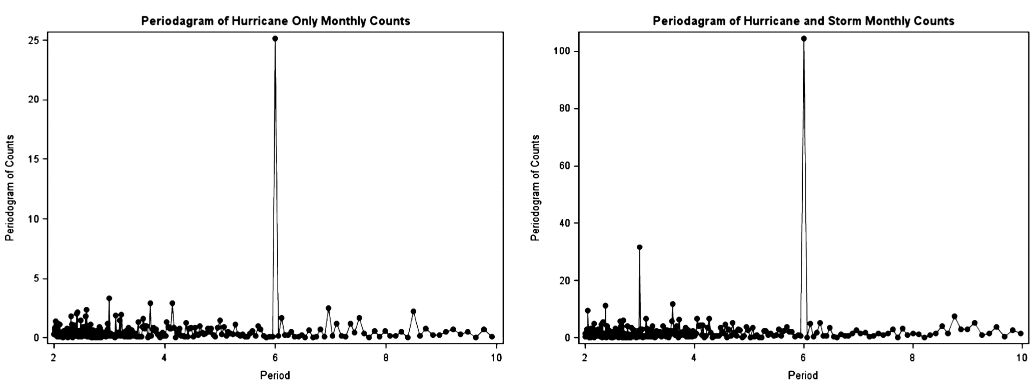

FIGURE 5: The periodograms of the monthly H&TS and the HONLY datasets.

the time series of monthly claim counts during the official hurricane season is characterized by seasonal fluctuations but less linear trend.

Climatological studies suggest that the intensity for hurricanes and tropical storms does not repeat the exact same short-term pattern every year. Rather, it varies from year to year, as in alternating the El Niño-La Niña cycles. For example, research on the tropical cyclones affecting the coast of Texas, during the El Niño/La Niña years between 1900 and 1996, shows that the highest percentage of all major hurricanes which have affected the coast of Texas occurred when El Niño was present for at least part of the given year [see Cole and Pfaff (1997)], which implies the possible variation of the annual counts over time. See, also, Pielke and Landsea (1999), for their study on the strong connection between the Atlantic hurricane landfalls in the U.S. and the El Niño Southern Oscillation (ENSO) phenomena.

The NOAA uses the Accumulated Cyclone Energy (ACE) index, in combination with the numbers of named storms, hurricanes, and major hurricanes, to categorize North Atlantic hurricane seasons as being above-normal, near-normal and below-normal. We understand that the NOAA uses a deterministic modeling approach to classify different hurricane seasons and to estimate the transmission probabilities between these states, based on the ACE index and other known information. A further generalization could be to consider a hidden process which affects the hurricane seasons, where the exact number of states of the hidden process and the movement between states are unobservable.

We propose a NPHMM for the analysis of the monthly hurricane and tropical storm counts during the hurricane season. This approach not only enables one to account for serial correlation and over-dispersion via a latent dynamic process but also to address the seasonality. As mentioned in Section 1, those who are interested in making inferences about the latent process may find this modeling approach appealing. Here we presume that the seasonality of the Atlantic hurricanes and tropical storms repeats a similar short-term pattern every year. Meanwhile, the average intensity, affected likely by the El Niño-La Niña cycles and other unknown (random) phenomena which are not clearly observed, varies over time. That is, the natural environment presumably makes transitions between states according to a stochastic process. In this paper, we assume a simple Markovian process.

3. A nonhomogeneous Poisson hidden Markov model

In this section, we first recall the PHMM defined, in MacDonald and Zucchini (1997), then generalize it by considering the seasonality in the Poisson intensity to form a NPHMM. Some characteristics of the latter model are also presented.

For a two-dimension discrete time stochastic process $\{(J_t, N_t); t \in \mathbb{N}^+\}$, let $\{J_t; t \in \mathbb{N}^+\}$ be an unobservable Markov chain process with finite state space $E = \{1, 2, \ldots, m\}$ and $\{N_t; t \in \mathbb{N}^+\}$ be the observed sequence of counts depending on $\{J_t; t \in \mathbb{N}^+\}$, where J_t denotes the state of the background environment process at time interval $[t-1,t)$ and N_t represents the corresponding counts within that interval. We assume that the unobserved process $\{J_t;\, t \in \mathbb{N}^+\}$ is a discrete, homogeneous, aperiodic, irreducible Markov chain on the finite state space E with an unknown transition probability matrix $P = [p_{i,j}]_{i,j\in E}$, where $p_{i,j} = \mathbb{P}\{J_t = j \,|\, J_{t-1} = i\}$, $i,j \in E$, such that $\sum_{j\in E} p_{i,j} = 1$. The initial distribution which is also the stationary distribution is denoted by $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_m)'$.

Given $\{J_t;\, t \in \mathbb{N}^+\}$, we assume that $\{N_t;\, t \in \mathbb{N}^+\}$ is a sequence of independent random variables. Moreover, we assume that for every t (the time index), given $J_t = j$, N_t is Poisson distributed with intensity rate λ_j, that is, the conditional distribution of N_t, $t \in \mathbb{N}^+$, is given by

$$q_{k,j} = \mathbb{P}\big\{N_t = k \,\big|\, J_t = j\big\} = e^{-\lambda_j}\frac{\lambda_j^k}{k!}, \quad k \in \mathbb{N},\ j \in E,$$

with $\sum_{k\in\mathbb{N}} q_{k,j} = 1$. These define the so-called PHMM, and the process $\{J_t;\, t \in \mathbb{N}^+\}$ can be called the (external) environment or background process.

Following from the above assumptions, the marginal distribution of N_t, $t \in \mathbb{N}^+$, then takes the form

$$\mathbb{P}\{N_t = k\} = \sum_{j=1}^{m} \mathbb{P}\{J_t = j\}\,\mathbb{P}\big\{N_t = k \,\big|\, J_t = j\big\} = \sum_{j=1}^{m} \pi_j q_{k,j}, \quad k \in \mathbb{N},$$

which is a finite mixture of Poisson distributions with a finite mean given by $\mathbb{E}[N_t] = \sum_{j=1}^{m} \pi_j \lambda_j$ as a weighted average of Poisson intensity parameters.

Now, as indicated in Sections 1 and 2, the seasonality may also affect the claim frequency within each year; in this case, we assume that, given $J_t = j \in E$, $t \in \mathbb{N}^+$, the number of counts within the time interval $[t-1,t)$ follows a NHP process with a time dependent intensity function

$$\lambda(s) = \lambda_j \beta_j(s - t + 1), \quad t - 1 \le s < t, \tag{2}$$

in which $\beta_j(s)$ is assumed to be a continuous short-term intensity function defined on [0,1) to reflect the seasonality. Furthermore, we assume that $\int_0^1 \beta_j(s)\,ds = 1$ so that the parameter λ_j is the average intensity within the time interval $[t-1,t)$. Hence for every $t \in \mathbb{N}^+$ the total number of counts between $t-1$ and t, N_t, is also Poisson distributed with parameter $\Lambda_j = \int_{t-1}^{t} \lambda(s)\,ds = \lambda_j$ and the unconditional distribution of N_t follows a finite mixture of Poisson's with probabilities

$$\mathbb{P}\{N_t = k\} = \sum_{j=1}^{m} \pi_j e^{-\lambda_j}\frac{\lambda_j^k}{k!}, \quad k \in \mathbb{N},$$

and a finite mean $\mathbb{E}[N_t] = \sum_{j=1}^{m} \pi_j \lambda_j$.

In order to precisely describe the seasonality and random fluctuation, we further divide the time interval $[t-1,t)$ into I sub-intervals $[t-1+(l-1)/I,\, t-1+l/I)$,

$l = 1, 2, \ldots, I$. For example, in practice, if t indexes a year, then I can be the number of months in a year. Let N_t^l denote the number of counts in the lth sub-interval, and $\mathbf{N}_t = (N_t^1, N_t^2, \ldots, N_t^I)$ is then the vector of observed monthly counts from the t-th year. Again, we assume that given $\{J_t; t \in \mathbb{N}^+\}$, $\mathbf{N}_t$ depends only on J_t, and $N_t^1, N_t^2, \ldots, N_t^I$ are conditionally independent random variables. Moreover, for every t and l, given $J_t = j$, N_t^l is Poisson distributed with an expected value

$$\Lambda_j^l = \int_{(l-1)/I}^{l/I} \lambda_j \beta_j(s)\, ds, \quad l = 1, 2, \ldots, I, \quad j \in E. \tag{3}$$

Then the probability of observing n_t^l events in the time interval $[t - 1 + (l - 1)/I, t - 1 + l/I)$ is

$$q_{n_t^l, j} = \mathbb{P}\left\{N_t^l = n_t^l \,\middle|\, J_t = j\right\} = e^{-\Lambda_j^l} \frac{[\Lambda_j^l]^{n_t^l}}{n_t^l!}, \quad n_t^l \in \mathbb{N},\ j \in E. \tag{4}$$

Hence, we obtain a NPHMM with $\beta_j(s)$ as the short-term (e.g., annum) intensity function.

In this paper, similar to Lu and Garrido (2005, 2007) we consider a beta density function for $\beta_j(s)$, with parameters μ_j, $v_j \geq 1$, defined on $[0, 1)$, that is,

$$\beta_j(s) = \begin{cases} \frac{s^{\mu_j - 1}(1-s)^{v_j - 1}}{B(\mu_j, v_j)}, & 0 \leq s < 1 \\ 0, & \text{otherwise} \end{cases}, \tag{5}$$

where $B(\mu_j, v_j) = \int_0^1 s^{\mu_j - 1}(1 - s)^{v_j - 1}\, ds$ is the (complete) beta function. It follows that (3) can be written as

$$\begin{aligned} \Lambda_j^l &= \frac{\lambda_j}{B(\mu_j, v_j)} \int_{\frac{l-1}{I}}^{\frac{l}{I}} s^{\mu_j - 1} (1 - s)^{v_j - 1} ds \\ &= \frac{\lambda_j}{B(\mu_j, v_j)} \left[B\left(\mu_j, v_j; \frac{l}{I}\right) - B\left(\mu_j, v_j; \frac{l-1}{I}\right)\right], \end{aligned} \tag{6}$$

where

$$B(\mu_j, v_j; t) = \int_0^t s^{\mu_j - 1}(1 - s)^{v_j - 1} ds, \quad t \in (0, 1)$$

is the incomplete beta function, and consequently $\Lambda_j = \sum_{l=1}^{I} \Lambda_j^l = \lambda_j$.

Note that here the shape of the beta function varies according to an underlying environment process to reflect the impact of the El Niño-La Niña phenomena.

The choice of the beta function allows the flexibility in the shape (the skewness) of the intensity function. In addition, it is computationally tractable and convenient due to the ready-to-use complete and incomplete beta functions. It has been shown to be a suitable candidate for the short-term period intensity function for hurricane counts (Lu and Garrido 2005, 2007).

Following the above definitions, given the environment process $\{J_t; t \in \mathbb{N}^+\}$, the number of claim counts up to (integer) time t, $\sum_{s=1}^{t}\sum_{l=1}^{I} N_s^l$, can be seen as a nonhomogeneous Poisson random variable with intensity function

$$\lambda(s) = \begin{cases} \lambda_{J_1}\beta_{J_1}(s), & 0 \le s < 1 \\ \lambda_{J_2}\beta_{J_2}(s-1), & 1 \le s < 2 \\ \cdots\cdots & \\ \lambda_{J_t}\beta_{J_t}(s-t+1), & t-1 \le s < t \end{cases}.$$

Furthermore, we can also calculate the joint conditional probability function of $\mathbf{N}_t = (N_t^1, N_t^2, \ldots, N_t^I)$, given that $N_t = \sum_{l=1}^{I} N_t^l = n_t$ as follows:

$$\mathbb{P}\left\{N_t^1 = n_t^1, N_t^2 = n_t^2, \ldots, N_t^I = n_t^I \,\middle|\, N_t = n_t\right\} = \sum_{j=1}^{m} \pi_j \left[\frac{n_t!}{n_t^1! \cdots n_t^I!} \prod_{l=1}^{I}\left(\frac{\Lambda_j^l}{\Lambda_j}\right)\right],$$

which is a finite mixture of multinomial distributions with parameters $n_t = \sum_{l=1}^{I} n_t^l$ and $\Lambda_j^1/\Lambda_j, \ldots, \Lambda_j^I/\Lambda_j$, for $j \in E$.

4. Maximum likelihood estimations

The method of the MLE can be used to estimate the parameters in the NPHMMs described above. The parameters to be estimated are the transition probabilities $p_{i,j}$ for $i, j \in E$, the initial (stationary) distribution $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_m)'$, the average intensity parameters $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_m)'$, and the parameters for beta-type function $\beta_j(s)$, $\gamma_j = (\mu_j, \nu_j)'$, for $j \in E$.

Note that the diagonal element, $p_{i,i}$, of the transition matrix P can be evaluated by the formula, $p_{i,i} = 1 - \sum_{j=1, j\neq i}^{m} p_{i,j}$, $i \in E$, and the stationary distribution $\boldsymbol{\pi}$ can be either obtained from the equation $\boldsymbol{\pi}' = \boldsymbol{\pi}' P$ using the estimated transition matrix P, or estimated directly by the following EM algorithm. We now denote by $\boldsymbol{\psi}$ the vector of parameters to be estimated, that is,

$$\boldsymbol{\psi} = (p_{1,2}, p_{1,3}, \ldots, p_{m,m-1}, \boldsymbol{\lambda}', \gamma_1', \gamma_2', \ldots, \gamma_m')',$$

and by $\boldsymbol{\Psi}$ the corresponding parameter space.

Suppose that in total T (years) random vectors of $\mathbf{N}_t = (N_t^1, N_t^2, \ldots, N_t^I)$ are observed as $\mathbf{n}_t = (n_t^1, n_t^2, \ldots, n_t^I)$, $t = 1, 2, \ldots, T$. If we also observe the underlying state information of the Markov chain $\{J_t;\, t \in \mathbb{N}^+\}$ up to time T, say, $J_1 = j_1, \ldots, J_T = j_T$, then by applying the Markovian property of process $\{J_t;\, t \in \mathbb{N}^+\}$ and the conditional independence of random vectors $\{\mathbf{N}_t;\, t \in \mathbb{N}^+\}$ we can write the complete likelihood function $\mathcal{L}_T^c$ as follows:

$$\begin{aligned}
\mathcal{L}_T^c(\boldsymbol{\psi}) &= \mathbb{P}\{\mathbf{N}_1 = \mathbf{n}_1, \ldots, \mathbf{N}_T = \mathbf{n}_T, J_1 = j_1, \ldots, J_T = j_T\} \\
&= \mathbb{P}\{J_1 = j_1, \ldots, J_T = j_T\}\, \mathbb{P}\left\{\mathbf{N}_1 = \mathbf{n}_1, \ldots, \mathbf{N}_T = \mathbf{n}_T \middle| J_1 = j_1, \ldots, J_T = j_T\right\} \\
&= \mathbb{P}\{J_1 = j_1\}\left[\prod_{t=2}^{T} \mathbb{P}\left\{J_t = j_t \middle| J_{t-1} = j_{t-1}\right\}\right] \prod_{t=1}^{T} \mathbb{P}\left\{\mathbf{N}_t = \mathbf{n}_t \middle| J_t = j_t\right\} \\
&= \pi_{j_1}\left(\prod_{l=1}^{I} q_{n_1^l, j_1}\right) \prod_{t=2}^{T}\left[p_{j_{t-1}, j_t}\left(\prod_{l=1}^{I} q_{n_t^l, j_t}\right)\right],
\end{aligned} \tag{7}$$

where $q_{n_t^l, j_t}$ is the Poisson parameter for the time interval $[t-1+(l-1)/I,\, t-1+l/I)$ with $J_t = j_t$ given by (4). Since the state information of Markov chain $\{J_t;\, t \in \mathbb{N}^+\}$ up to time T is actually missing, the summing of variables $j_1, j_2, \ldots, j_T$ over E yields the likelihood function for observed data in the form

$$\mathcal{L}_T^{\text{obs}}(\boldsymbol{\psi}) = \sum_{j_1 \in E} \cdots \sum_{j_T \in E} \pi_{j_1}\left(\prod_{l=1}^{I} q_{n_1^l, j_1}\right) \prod_{t=2}^{T}\left[p_{j_{t-1}, j_t}\left(\prod_{l=1}^{I} q_{n_t^l, j_t}\right)\right]. \tag{8}$$

Further let $q_{\mathbf{n}_t, j_t}$ be the Poisson probability for time interval $[t-1, t)$ with observation $\mathbf{n}_t$ and $J_t = j_t$, that is,

$$q_{\mathbf{n}_t, j_t} = \prod_{l=1}^{I} q_{n_t^l, j_t} = e^{-\Lambda_{j_t}} \prod_{l=1}^{I} \frac{[\Lambda_{j_t}^l]^{n_t^l}}{(n_t^l)!}, \tag{9}$$

where $\Lambda_{j_t} = \sum_{l=1}^{I} \Lambda_{j_t}^l$, and $\Lambda_{j_t}^l$ is defined by (3). Denote by $M^1 = [\pi_{j_1} q_{\mathbf{n}_1, j_1}]_{j_1 \in E}$, an $1 \times m$ vector, and by $M^t = [p_{j_{t-1}, j_t} q_{\mathbf{n}_t, j_t}]_{j_{t-1}, j_t \in E}$, an $m \times m$ matrix. Then (8) can be rewritten as the following matrix form

$$\mathcal{L}_T^{\text{obs}}(\boldsymbol{\psi}) = (M^1)' \left[\prod_{t=2}^{T} M^t\right] \mathbf{1}, \tag{10}$$

where $\mathbf{1}$ is an $m \times 1$ column vector with all components being 1.

The MLE of parameters in $\boldsymbol{\psi}$ can be found by directly maximizing the observed likelihood function $\mathcal{L}_T^{\text{obs}}$, given by (10). However, the EM algorithm has been proposed as an alternative and widely utilized for fitting the PHMM. In fact, since its introduction, the EM algorithm has become the most popular method in the literature because the calculation of the derivatives of the log-likelihood of PHMM has historically been problematic and needs to be done numerically. Although criticized for being computationally slow, the EM algorithm appears to work in a broad selection of PHMM applications. In this paper, we consider both maximizing the observed log-likelihood directly through a quasi-Newton algorithm and the maximum likelihood via the EM algorithm. Note that the regularity conditions of Wu (1983) for the convergence of EM algorithm to stationary values can be similarly verified as in Paroli et al. (2000).

The EM algorithm performs first an E-step, which calculates an expectation of the complete log-likelihood function given the observed data under current estimate of parameters, and then an M-step, which updates the estimates by maximizing the expected log-likelihood obtained at the E-step. The estimation of parameters can be obtained iteratively between the E-step and M-step until convergence. Let $\boldsymbol{\eta}_s^t = (\mathbf{n}_s, \mathbf{n}_{s+1}, \ldots, \mathbf{n}_t)'$, with $\mathbf{n}_r = (n_r^1, n_r^2, \ldots, n_r^I)$ for $s \le r \le t$, be the observation of the random column vector $\mathbf{N}_s^t = (\mathbf{N}_s, \mathbf{N}_{s+1} \ldots, \mathbf{N}_t)'$, between (years) s and t. Then $\boldsymbol{\eta}_1^T$ is the observations up to time T. Further, for implementing the EM algorithm, we need the expression of the log-likelihood function, $\log \mathcal{L}_T^c(\boldsymbol{\psi})$, obtained from (7) as follows:

$$\log \mathcal{L}_T^c(\boldsymbol{\psi}) = \log \pi_{j_1} + \sum_{t=1}^{T} \sum_{l=1}^{I} \log q_{n_t^l, j_t} + \sum_{t=2}^{T} \log p_{j_{t-1}, j_t}, \quad \boldsymbol{\psi} \in \boldsymbol{\Psi}. \tag{11}$$

The log-likelihood function consists two parts: the log-likelihood for a Markov chain depending on the transition probabilities $p_{i,j}$, $i, j \in E$ and its stationary distribution π_j, $j \in E$, and the log-likelihood for independent vectors of observations depending only on the parameters λ_j, μ_j, v_j, $j \in E$. Note that the stationary distribution $\boldsymbol{\pi}$ can also be estimated from the estimation of the transition probability matrix by the relationship $\boldsymbol{\pi} = \boldsymbol{\pi} P$.

Let $\boldsymbol{\psi}^{(k)}$ be the vector of estimates obtained at the kth iteration. Then given $\boldsymbol{\psi}^{(k)}$, the $(k + 1)$th iteration of the EM algorithm is given in the following.

E-step: computer the expectation of the log-likelihood function (11), given by

$$\begin{aligned}\mathbb{E}\left[\log \mathcal{L}_T^c(\boldsymbol{\psi}) \,\middle|\, \boldsymbol{\eta}_1^T, \boldsymbol{\psi}^{(k)}\right] &= \mathbb{E}\left[\log \pi_{j_1} \,\middle|\, \boldsymbol{\eta}_1^T, \boldsymbol{\psi}^{(k)}\right] + \sum_{t=2}^{T} \mathbb{E}\left[\log p_{j_{t-1}, j_t} \,\middle|\, \boldsymbol{\eta}_1^T, \boldsymbol{\psi}^{(k)}\right] \\ &\quad + \sum_{t=1}^{T} \sum_{l=1}^{I} \mathbb{E}\left[\log q_{n_t^l, j_t} \,\middle|\, \boldsymbol{\eta}_1^T, \boldsymbol{\psi}^{(k)}\right].\end{aligned} \tag{12}$$

M-step: find $\boldsymbol{\psi}^{(k+1)}$ which maximizes the expectation (12), that is,

$$\mathbb{E}\left[\log \mathcal{L}_T^c(\boldsymbol{\psi}^{(k+1)}) \,\middle|\, \boldsymbol{\eta}_1^T, \boldsymbol{\psi}^{(k)}\right] = \max_{\boldsymbol{\psi} \in \boldsymbol{\Psi}} \left\{\mathbb{E}\left[\log \mathcal{L}_T^c(\boldsymbol{\psi}) \,\middle|\, \boldsymbol{\eta}_1^T, \boldsymbol{\psi}^{(k)}\right]\right\}.$$

Baum et al. (1970) introduced the forward and backward probabilities which can be used to find the estimators of the parameters in the M-step of the algorithm. The so-called forward and backward probabilities for time t ($1 \le t \le T$) and state j ($j \in E$) at time t, denoted by $a_t(j)$ and $b_t(j)$, respectively, are defined as

$$a_t(j) = \mathbb{P}\left\{\mathbf{N}_1^t = \boldsymbol{\eta}_1^t,\, J_t = j\right\},$$
$$b_t(j) = \mathbb{P}\left\{\mathbf{N}_{t+1}^T = \boldsymbol{\eta}_{t+1}^T \,\middle|\, J_t = j\right\}.$$

Here $a_t(j)$ is the joint probability of the "past" observations (up to current time t) and the current state of the Markov chain, while $b_t(j)$ is the conditional probability of the "future" observations given the current state information. The forward and backward probabilities can be calculated forward and backwards, respectively, as follows:

$$a_1(j) = \pi_j q_{\mathbf{n}_1,j},\; a_t(j) = q_{\mathbf{n}_t,j} \sum_{i\in E} a_{t-1}(i) p_{i,j}, \quad t = 2,3,\ldots,T,\; j \in E,$$
$$b_T(j) = 1,\; b_t(j) = \sum_{i\in E} b_{t+1}(i) q_{\mathbf{n}_{t+1},i}\, p_{j,i}, \quad t = T-1, T-2,\ldots,1,\; j \in E,$$

where $q_{\mathbf{n}_t,j}$, given by (9), depends only on parameters λ_j, μ_j and v_j, $j \in E$.

Now in the M-step of the $(k+1)$th iteration, given $\boldsymbol{\psi}^{(k)}$ obtained from the kth iteration, we are able to compute $a_t^{(k)}(j)$ and $b_t^{(k)}(j)$ for $1 \le t \le T$ and $j \in J$. Then maximizing the first and second terms respectively in (12) by the standard maximization approach yields the MLE's for the stationary distribution and the transition probabilities, given by

$$\pi_j^{(k+1)} = \frac{a_1^{(k)}(j) b_1^{(k)}(j)}{\sum_{l=1}^m a_1^{(k)}(l) b_1^{(k)}(l)}, \quad j \in E,$$

$$p_{i,j}^{(k+1)} = \frac{p_{i,j}^{(k)} \sum_{t=2}^T a_{t-1}^{(k)}(i) b_t^{(k)}(j) q_{\mathbf{n}_t,j}}{\sum_{t=2}^T a_{t-1}^{(k)}(i) b_{t-1}^{(k)}(i)}, \quad i,j \in E.$$

Write $\Lambda_j^l = \lambda_j B_j(l)$ in (6) with $B_j(l)$ equal to

$$B_j(l) = \frac{1}{B(\mu_j, v_j)} \int_{(l-1)/I}^{l/I} s^{\mu_j - 1}(1-s)^{v_j - 1} ds = \frac{B\left(\mu_j, v_j; \frac{l}{I}\right) - B\left(\mu_j, v_j; \frac{l-1}{I}\right)}{B(\mu_j, v_j)}, \tag{13}$$

for $j \in E$ and $l = 1, \ldots, I$. Now let $B_j^{(k)}(l)$ be the corresponding function to (13) with parameters $\mu_j^{(k)}$ and $v_j^{(k)}$ obtained from the kth iteration of the EM algorithm. Further let $B'_{\mu_j}(l) = \frac{\partial B_j(l)}{\partial \mu_j}$ and $B'_{v_j}(l) = \frac{\partial B_j(l)}{\partial v_j}$, and denote by $B'^{(k)}_{\mu_j}(l)$ and $B'^{(k)}_{v_j}(l)$, the corresponding terms with parameters $\mu_j^{(k)}$ and $v_j^{(k)}$. Let

$B'^{(k)}_{\mu_j} = \sum_{l=1}^{I} B'^{(k)}_{\mu_j}(l)$ and $B'^{(k)}_{\nu_j} = \sum_{l=1}^{I} B'^{(k)}_{\nu_j}(l)$. Since the third summation in (12) can be written as

$$\sum_{t=1}^{T}\sum_{j=1}^{m} \mathbb{P}\left\{J_t = j \,\middle|\, \boldsymbol{\eta}_1^T, \boldsymbol{\psi}^{(k)}\right\} \times \left\{-\frac{\lambda_j}{B(\mu_j,\nu_j)}\int_0^1 s^{\mu_j-1}(1-s)^{\nu_j-1}\,ds \right.$$
$$\left. + \log\lambda_j \sum_{l=1}^{I} n_t^l + \sum_{l=1}^{I}\left[n_t^l \log B_j(l) - \log n_t^l!\right]\right\}$$
$$= \sum_{t=1}^{T}\sum_{j=1}^{m} \frac{a_t^{(k)}(j)\, b_t^{(k)}(j)}{\mathcal{L}_T^c(\boldsymbol{\psi}^{(k)})}\left\{-\lambda_j + n_t \log\lambda_j + \sum_{l=1}^{I}\left[n_t^l \log B_j(l) - \log n_t^l!\right]\right\},$$

then by taking the partial derivatives with respect to λ_j, μ_j and ν_j, respectively, we have the following system of equations for $\lambda_j^{(k+1)}$, $\mu_j^{(k+1)}$ and $\nu_j^{(k+1)}$, $j \in E$:

$$\begin{cases} \lambda_j^{(k+1)} = \dfrac{\sum_{t=1}^{T} n_t \cdot a_t^{(k)}(j) b_t^{(k)}(j)}{\sum_{t=1}^{T} a_t^{(k)}(j)\, b_t^{(k)}(j)}, \\ B'^{(k+1)}_{\mu_j} \sum_{t=1}^{T}\left(\sum_{i=1}^{m} a_t^{(k)}(i)\, b_t^{(k)}(i)\,\lambda_i^{(k+1)}\right) = \sum_{l=1}^{I} n_\bullet^l \, \dfrac{B'^{(k+1)}_{\mu_j}(l)}{B_j^{(k+1)}(l)}, \\ B'^{(k+1)}_{\nu_j} \sum_{t=1}^{T}\left(\sum_{i=1}^{m} a_t^{(k)}(i)\, b_t^{(k)}(i)\,\lambda_i^{(k+1)}\right) = \sum_{l=1}^{I} n_\bullet^l \, \dfrac{B'^{(k+1)}_{\nu_j}(l)}{B_j^{(k+1)}(l)}, \end{cases} \tag{14}$$

where $n_\bullet^l$ denotes the total number of observations in the lth month. Solving numerically, the system of equations (14) yields the estimations of $\lambda_j^{(k+1)}$, $\mu_j^{(k+1)}$ and $\nu_j^{(k+1)}$ for the $(k+1)$th iteration of the EM algorithm.

In this way, the E-step and M-step are repeated alternatively until the log-likelihood values $\{\mathcal{L}_T^{\text{obs}}(\boldsymbol{\psi}^{(k)})\}$ converge as k becomes sufficiently large. A typical criterion for stopping the iterations of the EM algorithm is when the successive estimations of $\boldsymbol{\psi}$ are smaller than a pre-set adequately small value. Then the MLE of the parameter set $\boldsymbol{\psi}$ is obtained.

5. Application to hurricane and the tropical storm data

With the assumption that the periodic short-term intensity function for seasonality is of the beta-type for all the states, we fit to both datasets the NPHMM. As explained in Section 2, we consider the observed counts to lie within the time interval [5/12,11/12) (i.e., from the beginning of June to the end of November) for all years as it is reasonable to assume that the short-term intensity function takes positive values only in this interval. The modified beta-type function $\beta_j(s)$ with parameters μ_j, $v_j \geq 1$ can be defined as

$$\beta_j(s) = \begin{cases} \frac{1}{D \cdot B(\mu_j, v_j)} \left(\frac{s-d_1}{D}\right)^{\mu_j - 1} \left(1 - \frac{s-d_1}{D}\right)^{v_j - 1}, & 0 \leq d_1 \leq s \leq d_2 \leq 1 \\ 0, & \text{otherwise} \end{cases}, \quad (15)$$

where $D = d_2 - d_1$. As in this case, $\int_{d_1}^{d_2} \beta_j(s)\, ds = 1$, the average intensity rate for that year (D portion of the year) is actually $D\lambda_j$. Apparently, for our datasets, $d_1 = 5/12$, $d_2 = 11/12$ and $D = 1/2$. Furthermore, observations (counts) $\mathbf{n}_v = (n_v^1, n_v^2, \ldots, n_v^I)$ for $1 \leq v \leq T$ with $I = 6$ and $T = 109$ are now considered as the observations within the interval [5/12, 11/12).

The MLEs are obtained both by maximizing the observed log-likelihood function directly and by maximizing the likelihood via the EM algorithm. Since multiple local maxima are often possible for PHMM likelihoods, the estimation procedure is sensitive to the choice of starting values. Our starting values are chosen via a grid search over a set of reasonable values. Overall, the two approaches result in comparable parameter estimates. The direct maximization of the likelihood produces MLE more quickly than the EM algorithm, but it encounters convergence problems with some initial values. See, Section 3 in Zucchini and MacDonald (2009), and especially Bulla and Berzel (2008), for details and discussions on this issue. The computation time for the EM algorithm is reasonable since our dataset is not large and our model is not too complex.

TABLE 1

THE MLES OBTAINED BY THE EM ALGORITHM FOR THE H&TS AND THE HONLY DATA*

m	Transition Probability			Stationary Distribution	Intensity Rates λ	μ	v
1	1			1	3.9909	2.4618	2.5783
	1			*1*	*1.6606*	*3.2251*	*3.1894*
2	.5428	.4572		.6311	3.9386	1.6215	2.0581
	.7823	.2177		.3689	4.0804	10.4825	7.8189
	.5392	*.4608*		*.6846*	*1.5650*	*6.7364*	*5.5373*
	1	*0*		*.3154*	*1.8683*	*1.5917*	*2.1655*
3	.2713	.6651	.0636	.4339	2.9943	1.2341	1.7782
	.4148	.3072	.2780	.4213	3.9836	8.7806	6.7634
	.9772	.0226	0	.1448	7.0033	2.0747	2.3575
	.1297	*.8703*	*0*	*.2490*	*0.6520*	*1.8661*	*2.8134*
	0	*.5945*	*.4055*	*.5343*	*1.7567*	*7.4127*	*5.9769*
	1	*0*	*0*	*.2167*	*2.5834*	*1.8154*	*2.2064*

* Roman font for the H&TS data and Italic font for the HONLY data.

The MLEs of the model parameters obtained by the EM algorithm are presented in Table 1 when the number of states of the underlying Markov chain are $m = 1, 2$, and 3, respectively, while the corresponding estimations obtained by the direct maximization are showed in Table 2. In each case, the upper part with Roman fonts is for the H&TS data, while the lower part with Italic fonts is for HONLY data. It is clear that when assuming a one-state underlying environment process ($m = 1$) there are no hidden effects and the values for $m = 1$ in Table 1 are estimated based on direct method and are listed for comparison.

It can be observed from Tables 1-2 that the estimations by both methods for $m = 1, 2$ are very close and for $m = 3$ are reasonably close. In particular, estimates of the (μ, ν) pairs vary in all cases, implying that the shape of fitted beta function does change depending on the underlying Markov chain state.

While the occurrence of tropical storms and hurricanes shows a strong yearly seasonality, the estimated average intensity parameter λ_j's for different states reveals the influence or consequence from some unpredictable effects. For example, in Table 1, the $m = 2$ case for HONLY data, with probability 0.5392 the chain remains in the low-occurrence state with an average yearly occurrence of approximately 1.5, while with a near zero probability the chain remains in the high-occurrence state with average yearly occurrences close to 2. It is also estimated that 68.46% (31.54%) of the total observed years are recognized as the low-occurrence (high-occurrence) years.

The use of the ACE index by the NOAA for classifying the North Atlantic hurricane seasons, explained in Section 2, motivates our choice of number of

TABLE 2

THE MLEs OBTAINED BY DIRECT MAXIMIZATION FOR THE H&TS AND THE HONLY DATA*

m	Transition Probability			Stationary Distribution	Intensity Rates λ	μ	ν
1	1			1	3.9909	2.4618	2.5783
	1			*1*	*1.6606*	*3.2251*	*3.1894*
2	.5422	.4578		.6306	3.9392	1.6209	2.0577
	.7815	.2185		.3694	4.0792	10.4788	7.8168
	.5410	*.4590*		*.6854*	*1.5649*	*6.7337*	*5.5355*
	1	*0*		*.3146*	*1.8693*	*1.5894*	*2.1641*
3	0	1	0	.3127	2.9195	1.0347	1.6482
	.2981	.3922	.3097	.5145	3.8454	6.9465	5.5688
	.9225	0	.0775	.1728	6.3645	1.8347	2.1795
	0	*1*	*0*	*.2055*	*0.5806*	*1.4885*	*2.3382*
	0	*.6512*	*.3488*	*.5891*	*1.7058*	*6.9730*	*5.7368*
	1	*0*	*0*	*.2055*	*2.6137*	*1.7612*	*2.1826*

* Roman font for the H&TS data and Italic font for the HONLY data.

TABLE 3

THE LIKELIHOODS OF MODELS FOR THE H&TS AND THE HONLY DATA*

m	Number of Parameters	EM Algorithm Log-likelihood	AIC	BIC
1	3	– 705.9197	1417.8394	1431.2887
2	8	– 677.0744	1370.1488	1406.0137
3	15	– 666.8919	1363.7838	1431.0304
1	3	*– 389.2729*	*784.5458*	*797.9951*
2	8	*– 382.9192*	*781.8384*	*817.7033*
3	15	*– 377.1822*	*784.3644*	*851.6110*

* Roman font for the H&TS data and Italic font for the HONLY data.

states m taking values up to 3. Here we apply two commonly used penalized likelihood methods, the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) suggested respectively in Akaike (1974) and Schwarz (1978), to better rank models with different number of the hidden Markov chain states $m = 1, 2, 3$. The new likelihoods by AIC is $-2\log \mathcal{L}_T^c + 2d_m$, where d_m is the number of parameters in the model, and by BIC is $-2\log \mathcal{L}_T^c + 2d_m \log K$, where K is the total number of observations.

Since loglikelihoods are very close in both estimation methods, Table 3 displays only values calculated based on the ones obtained by the EM estimation. For the U.S. Atlantic H&TS data, the AIC suggests $m = 3$ as the best model, while the BIC suggests $m = 2$ as the best model. For the U.S. Atlantic HONLY data, the AIC and BIC yield the minimum penalized likelihood values for $m = 2$ and $m = 1$, respectively. The "$m = 1$" result here for the HONLY data is somewhat consistent with the over-dispersion observed in Section 2. It is worth mentioning the point made in MacKay (2002), that the use of both the AIC and BIC methods in the hidden Markov model context methods has not been justified theoretically. See also Katz (1981) for the discussion on estimating the order of a Markov chain using the AIC and BIC.

As expected, the over-dispersion of the U.S. Atlantic H&TS that motivated the model choice now yields a good fit for the NPHMM. Meanwhile the lack of over-dispersion of the U.S. Atlantic HONLY data suggests a relatively poorer NPHMM where the average intensities in the 2-state model (preferred by the AIC) are not significantly different (1.56 and 1.87, respectively). Overall, the fitting results seem in favor of our initial model choice which is essentially the compound of two components: the nonhomogeneous beta-type intensity function describing the seasonality within the year, and the hidden Markov chain capturing the fluctuations from year to year due to the underlying environmental process.

Figure 6 shows three estimated $\lambda_j \beta_j(t)$, for $5/12 \le t \le 11/12$ and $j = 1, 2, 3$, in the 3-state model (preferred by AIC) for the U.S. Atlantic H&TS data using the EM estimation method, where states 1, 2, 3 correspond to low-, moderate-,

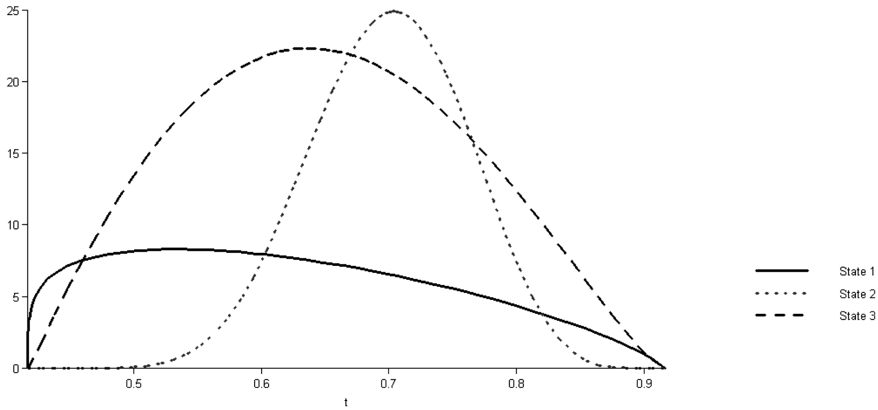

FIGURE 6: Intensity functions in the 3-state model for the H&TS data.

and high-occurrence years. As we can see the shapes of these three intensity functions are quite different. When the background process is in state 1, the solid line in Figure 6 shows a low-occurrence of the hurricanes and tropical storms with an estimated yearly average count of 2.99, and they occur near evenly in the early months of the season from June to September. When it is in an moderate-occurrence year (state 2), the dotted line implies that most events occur centrally in three months from August to October and the estimated average number of yearly occurrences is around 3.98. In state 3, the dashed line indicates a high-occurrence of these hurricanes and tropical storms with an estimated annual average count of 7 and they occur mostly in the middle four months of the season. The estimated conditional monthly counts and their corresponding proportions for each state are displayed in Table 4. The estimated unconditional yearly average count, given by $\sum_{j=1}^{3} \hat{\pi}_j \hat{\lambda}_j$, is 3.992, which is very closed to the empirical one from the data.

TABLE 4

THE ESTIMATED MONTHLY COUNTS IN THE 3-STATE MODEL FOR THE H&TS DATA

Month	Low-occurrence Estimated Counts (%)	Moderate-occurrence Estimated Counts (%)	High-occurrence Estimated Counts (%)
JUN	0.592 (19.77)	0.001 (.02)	0.592 (8.45)
JUL	0.688 (22.99)	0.124 (3.12)	1.486 (21.21)
AUG	0.640 (21.36)	1.067 (26.79)	1.836 (26.21)
SEP	0.533 (17.79)	1.945 (48.83)	1.671 (23.87)
OCT	0.379 (12.65)	0.815 (20.46)	1.096 (15.66)
NOV	0.163 (5.44)	0.031 (.78)	0.322 (4.60)
Total	2.995 (100)	3.983 (100)	7.003 (100)

We should mention the identifiability issue with respect to the model parameters encountered while running the EM or the direct maximizing estimation algorithms. There were cases when the system could not clearly identify between states and returned estimation results, for example, with the same average intensity rate for different states. This practically affects only the computational efficiency, and for prediction, the impact could be technically ignored.

6. Concluding Remarks

From the statistical application point of view, we conclude that the PHMMs are more realistic in practice than the classical Poisson processes, as they take into account the unobservable underlying environmental effect, which affects the (claim) counts. This seems to be the case for hurricane and tropical storm landfalls. Moreover, the PHMMs with a short-term claim intensity can be useful in modeling claim counts that evolve in a periodic environment. By considering the impact of seasonality, the proposed beta-type periodic claim intensity generalizes the Poisson models with constant intensity rate. The flexibility in shape of the beta function, as well as the tractability of the statistical estimation of model parameters, should make these nonhomogeneous beta-featured PHMMs easy to use in practice. We hope that the illustration of the hurricane and tropical storm dataset serves to show that NPHMMs can also be tractable if properly parameterized.

The model presented allows for over-dispersion relative to a Poisson distribution and for correlation between observations. This gives an alternative method to study the U.S. Atlantic hurricane and tropical storm data as we try to explain the role the variability of random environments plays in possibly inducing the El Niño/La Niña phenomena, which in turn affect the count distribution or process.

This dynamic NPHMM can be extended by considering the general dth order Markov process instead of the ordinary Markov chain for the state process and also the nonhomogeneous transitions for the hidden Markov chain. The Markov-modulated Poisson process or the two-step dependent PHMM in the sense of Rydén (1994) can also be studied where the distribution of the claim counts depends on the current state as well as on the previous one. By combining the loss or damage incurred from these (claim) counts, the models presented in this paper could be further developed to estimate the total insurance and/or reinsurance costs.

Acknowledgments

We are grateful to the two anonymous referees for their constructive comments for improving the manuscript. We would like to thank Prof. José Garrido for the insightful discussion at the early stage of this research project. This research was supported by the Natural Science and Engineering Research Council (NSERC) of Canada.

References

Albert, P.S. (1991) A two-state Markov mixture model for a time series of epileptic seizure counts. *Biometrics* **47**, 1371-1381.

Akaike, H. (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19(6)**, 716-723.

Altman, R.M. and Petkau, A.J. (2005) Application of hidden Markov models to multiple sclerosis lesion count data. *Statistics in Medicine* **24**, 2335-2344.

Baum, L.E., Petrie, T., Soules, G. and Weiss, N. (1970) A maximization technique occurring in the statistical estimation for probabilistic functions of Markov chains. *The Annals of Mathematical Statistics* **41**, 164-171.

Beard, R.E., Pentikäinen, T. and Pesonen, E. (1984) *Risk Theory* 3rd ed. Chapman & Hall, London.

Bulla, J. and Berzel, A. (2008) Computational issues in parameter estimation for stationary hidden Markov models. *Comp. Stat.* **23(1)**, 1-18.

Cole, J.D. and Pfaff, S.R. (1997) A climatology of tropical cyclones affecting the Texas coast during El Niño/non-El Niño years: 1990-1996. *Technical Attachment SR/SSD 97-37*, National Weather Service Office, Corpus Christi, Texas, (http://www.srh.noaa.gov/topics/attach/html/ssd97-37.htm).

Cooper, B. and Lipsitch, M. (2004) The analysis of hospital infection data using hidden Markov models. *Biostatistics* **5(2)**, 223-237.

Cox, D.R. (1981) Statistical analysis of time series: some recent developments. *Scandinavian Journal of Statistics*, **8(2)**, 93-115.

Daykin, C.D., Pentikäinen, T. and Pesonen, M. (1994) *Practical Risk Theory for Actuaries*. Chapman & Hall, London.

Freeland, R.K. and McCabe, B.P.M. (2004) Analysis of low count time series data by Poisson autoregression. *Journal of Time Series Analysis* **25(5)**, 701-722.

Freeland, R.K., Latour, A. and Oraichi, D. (2006) Integer-valued GARCH process. *Journal of Time Series Analysis* **27(6)**, 923-942.

Garrido, J. and Lu, Y. (2004) On double periodic non-homogeneous Poisson processes. *Bulletin of the Association of Swiss Actuaries* **2**, 195-212.

Gart, J.J. and Pettigrew, H.M. (1970) On the conditional moments of the k-statistics for the Poisson distribution. *Biometrika* **57(3)**, 661-664.

Grandell, J. (1991) *Aspects of Risk Theory*. Springer-Verlag, Berlin.

Gray, W.M. (1984) Atlantic seasonal hurricane frequency. Part I: El Niño and 30 mb quasi-biennial oscillation influences. *Monthly Weather Review* **112**, 1649-1668.

Heinen, A. (2003) Modelling time series count data: an autoregressive conditional Poisson model. *Center for Operations Research and Econometrics (CORE) Discussion Paper* No. 2003-63, University of Louvain, Belgium.

Höhle, M. and Paul, M. (2008) Count data regression charts for the monitoring of surveillance time series. *Computational Statistics and Data Analysis* **52**, 4357-4368.

Hughes, J.P. and Guttorp, P. (1999) A non-homogeneous hidden Markov model for precipitation occurrence. *Applied Statistics* **48**, 15-30.

Juang, B.H. and Rabiner, L.R. (1991) Hidden Markov models for speech recognition. *Technometrics* **33**, 251-272.

Jung, R.C., Kukuk, M. and Liesenfeld, R. (2006) Time series of count data: modelling, estimation and diagnostics. *Computational Statistics and Data Analysis* **51**, 2350-2364.

Katz, R.W. (1981) On some criteria for estimating the order of a Markov chain. *Technometrics* **23(3)**, 243-249.

Katz, R.W. (2002) Stochastic modeling of hurricane damage. *Journal of Applied Meteorology* **41**, 754-762.

Landreneau, D. (2001) Atlantic tropical storms and hurricanes affecting the United States: 1899-2002. *NOAA Technical Memorandum NWS SR-206 (Updated through 2002)*, National Weather Service Office, Lake Charles, Louisiana, (http://www.srh.noaa.gov/lch/research/tropical.htm).

Leroux, B.G. and Puterman, M.L. (1992) Maximum-penalized-likelihood estimation for independent and Markov-dependent mixture models. *Biometrics* **48**, 545-558.

Lu, Y. and Garrido, J. (2005) Doubly periodic non-homogeneous Poisson models for hurricane data. *Statistical Methodology* **2**, 17-35.
Lu, Y. and Garrido, J. (2007) Regime-switching periodic models for claim counts. *North American Actuarial Journal* **10**, 235-248.
MacDonald, I.L. and Zucchini, W. (1997) *Hidden Markov and Other Models for Discrete-valued Time Series*. Chapman & Hall, London.
MacKay, R.J. (2002) Estimating the order of a hidden Markov model. *The Canadian Journal of Statistics* **30(4)**, 573-589.
McKenzie, E. (2003) Discrete variate time series. In *Stochastic Processes: Modelling and Simulation* (eds Shanbhag, D.N. and Rao, C.R.). Amsterdam: Elsevier, 573-606.
Paroli, R., Redaelli, G. and Spezia, L. (2000) Poisson hidden Markov models for time series of overdispersed insurance counts. ASTIN Colloquium, Porto Cervo, Italy.
Pielke Jr., R.A., Gratz, J., Landsea, C.W., Collins, D., Saunders, M.A. and Musulin, R. (2008) Normalized hurricane damage in the United States: 1900-2005. *Natural Hazards Review* **9**, 29-42.
Pielke Jr., R.A. and Landsea, C.W. (1999) La Niña, El Niño, and Atlantic hurricane damages in the United States. *Bull. Amer. Meteor. Soc.* **80**, 2027-2033.
Quddus, M.A. (2008) Time series count data models: an empirical application to traffic accidents. *Accident Analysis and Prevention* **40**, 1732-1741.
Rao, C.R. and Chakravarti, I.M. (1956) Some small sample tests of significance for a Poisson distribution. *Biometrics* **12**, 264-282.
Rydén, T. (1994) Parameter estimation for Markov modulated Poisson processes. *Stochastic Models* **10**, 795-829.
Schwarz, G.E. (1978) Estimating the dimension of a model. *Annals of Statistics* **6(2)**, 461-464.
Trenberth, K.E. (1997) The definition of El Niño. *Bull. Amer. Meteor. Soc.* **78**, 2771-2777.
Weiss, C.H. (2008) Thinning operations for modelling time series of counts – a survey. *Advances in Statistical Analysis* **92(3)**, 319-341.
Weiss, C.H. (2009) Modelling time series of counts with overdispersion. *Statistical Methods and Applications* **18**, 507-519.
Wu, C.F.J. (1983) On the convergence properties of the EM algorithm. *The Annals of Statistics* **11**, 95-103.
Zeger, S.L. (1988) A regression model for time series of counts. *Biometrika* **75(4)**, 621-629.
Zucchini, W. and MacDonald, I.L. (2009) *Hidden Markov Models for Time Series: An Introduction Using R*, 2nd Ed. Chapman and Hall/CRC, London.

Yi Lu
Department of Statistics & Actuarial Science
Simon Fraser University
8888 University Drive
Burnaby, BC, Canada V5A 1S6
Fax: 1.778.782.4368
E-Mail: yilu@sfu.ca

Leilei Zeng
Department of Statistics & Actuarial Science
University of Waterloo
200 University Avenue West
Waterloo, Ontario, Canada N2L 3G1
Fax: 1.519.746.1875
E-Mail: lzeng@math.uwaterloo.ca