Annals of the Institute of Statistical Mathematics (2023) 75:253–280
https://doi.org/10.1007/s10463-022-00843-5

Inhomogeneous hidden semi-Markov models for incompletely observed point processes

Amina Shahzadi[1,2] · Ting Wang[1] · Mark Bebbington[3] · Matthew Parry[1]

Received: 24 April 2021 / Revised: 17 March 2022 / Accepted: 2 June 2022 /
Published online: 18 September 2022

Abstract

A general class of inhomogeneous hidden semi-Markov models (IHSMMs) is proposed for modelling partially observed processes that do not necessarily behave in a stationary and memoryless manner. The key feature of the proposed model is that the sojourn times of the states in the semi-Markov chain are time-dependent, making it an inhomogeneous semi-Markov chain. Conjectured consistency of the parameter estimators is checked by simulation study using direct numerical optimization of the log-likelihood function. The proposed models are applied to a global volcanic eruption catalogue to investigate the time-dependent incompleteness of the record by introducing a particular case of IHSMMs with time-dependent shifted Poisson state durations and a renewal process as the observed process. The Akaike Information Criterion and residual analysis are used to choose the best model. The selected IHSMM provides useful insights into the completeness of the global record of volcanic eruptions, demonstrating the effectiveness of this method.

Keywords Time-dependent missing data for point processes · Inhomogeneous semi-Markov chain · Residual analysis · Viterbi path · Global volcanic eruption record · Hazard

✉ Amina Shahzadi
aminashahzadi@gmail.com

Ting Wang
ting.wang@otago.ac.nz

[1] Department of Mathematics and Statistics, University of Otago, 362 Leith Street, Dunedin 9016, New Zealand

[2] Department of Statistics, Government College University, Katchery Road, Lahore 54000, Pakistan

[3] School of Mathematical and Computational Sciences, Massey University, Private Bag 11222, Palmerston North 4442, New Zealand

1 Introduction

Natural phenomena such as volcanic eruptions or earthquakes can cause catastrophic damage to life and infrastructure. Estimation of hazard is a primary motivation for scientific analysis of these phenomena. Statistical approaches typically involve a point process model. However, missing events are a problem common to most point process records of geophysical events where the degree of their completeness varies over time. Rapid advances in technology for detection and recording of events have substantially increased the amount of data in different fields during the last few decades, resulting in less incompleteness of event records in recent times. However, the older part of most historical records is substantially incomplete for many reasons, leading to time-inhomogeneity in the completeness of event records. A motivating example of time-dependent incomplete event records can be seen when modelling volcanic eruption records, where the missingness of events depends on time, the magnitude of eruptions and, in a complex manner, available geological and historical records (Guttorp and Thompson, 1991; Siebert et al., 2010; Deligne et al., 2010; Brown et al., 2014; Kiyosugi et al., 2015; Rougier et al., 2016). Hazard estimates from such time-inhomogeneous incomplete point process records are complicated and potentially biased, in particular underestimated if missing data is not considered in the model.

One possible way is to model incomplete data using hidden Markov models (HMMs) (Baum and Petrie, 1966). Recent years have seen an important application of HMMs in seismology (Beyreuther and Wassermann, 2008; Ibáñez et al., 2009; Wang et al., 2012, 2017; Wang and Bebbington, 2013) and volcanology (Bebbington, 2007; Wang and Bebbington, 2012). Wang and Bebbington (2012) introduced a type of homogeneous HMM to model incomplete volcanic eruption records. The model treats the number of missing events as the underlying homogeneous Markov chain and implicitly assumes that the state durations are geometrically distributed, which is often referred to as the memoryless property. The justification for using such an HMM is that there is no systematic trend(s) in the rate of observed eruptions from Taranaki volcano through time.

In real-world applications, the assumption of homogeneous HMMs with geometric state durations is often inadequate. For example, for volcanic eruption records, the presence of more observers in more geographic locations, improved literacy, communications and geological investigations help increase the proportion of recorded volcanic eruptions over time (Simkin, 1993). This leads to 'missing not at random' and we must take inhomogeneity into consideration when modelling the data in order to avoid systematic bias in hazard estimates (Rougier et al., 2016). In many volcanic eruption records, the processes that led to missed eruptions do not necessarily behave in a stationary and memoryless manner. A particular state with an arbitrary number of missing events may have persisted for a longer period in the past before the advent of written records and satellite monitoring. The assumption that the sojourn times of the HMM are geometrically distributed (memoryless) may not represent the temporal structure of missing events adequately.

To address the inadequacy mentioned above, in this study we extend HMMs to the case where, (1) the state durations of hidden Markov chain can explicitly follow a discrete distribution other than the implicit geometric distribution, e.g. a semi-Markov chain (Limnios and Oprişan, 2001) and (2) the duration distribution of a state can vary with time.

Allowing explicitly distributed state durations in any state of a first-order Markov chain produces a semi-Markov chain with the Markov property satisfied only at state transition times (Barbu and Limnios, 2008). An HMM with a semi-Markov chain as the hidden process is known as the hidden semi-Markov model (HSMM) and was originally proposed by Ferguson (1980). Subsequent studies proposed different versions of HSMMs using different distributions for modelling state durations (Russell and Moore, 1985; Levinson, 1986; Mitchell and Jamieson, 1993; Durbin et al., 1998). Over time, an extensive literature has developed on HSMMs. The computational techniques based on Derin's scheme for HSMMs can be seen in Guédon and Cocozza-Thivent (1990), Guédon (2003) with applications in plant structuring, and Bulla (2006) with applications in finance. Some essential developments of HSMMs were proposed by Barbu and Limnios (2008) with DNA applications, Trevezas and Limnios (2009) (dependence on the backward process), Malefaki et al (2010) (stochastic expectation-maximization algorithm for HSMM), Pertsinidou and Limnios (2015) (Viterbi algorithm for HSMM). Some other applications of HSMMs can be found in Sansom and Thomson (2001) and Sansom and Thompson (2003) for rainfall data analysis, Rossi et al (2015) for multiview video traffic, and Votsi et al (2018) for earthquakes. Yu and Kobayashi (2003) introduced an extension of HSMMs based on residual sojourn times. Barbu and Limnios (2008) and Yu (2015) provided comprehensive overviews of HSMMs.

However, to model the nonstationary behaviour of point process records, we need to allow explicit state durations to depend on time. Moreover, in real-world applications, the sojourn time of the last visited state may not end at the end of the observation sequence, so the duration of the last visited state in HSMMs should be right-censored.

Based on the above considerations, we propose to introduce time-varying state durations of an underlying semi-Markov chain in the right-censored HSMM of Guédon (2003). We first present some preliminaries of semi-Markov chain in Sect. 2. The proposed inhomogeneous semi-Markov chain (ISMC) and inhomogeneous hidden semi-Markov models (IHSMMs) in discrete-time are introduced in Sects. 3 and 4, respectively. A renewal process and its use in the framework of HMMs are described in Sect. 5. Section 6 then proposes an IHSMM with state durations following time-dependent shifted Poisson distributions, and the observed process following state-dependent gamma renewal processes to model time-inhomogeneous incomplete point process records. Section 7 presents a simulation study to investigate the properties of the parameter estimators. We apply the proposed model to a global volcanic eruption record in Sect. 8, followed by Discussion and Conclusion in Sects. 9 and 10, respectively. Note that HSMMs and IHSMMs in this paper are right-censored.

2 Preliminaries

Before introducing the proposed model, some notations, assumptions and definitions related to semi-Markov chains and the associated Markov renewal chains in discrete-time are needed.

2.1 Markov renewal chain

Consider a random system with a finite state space $\mathbb{S} = \{1, 2, \ldots, m\}$. Suppose that $\mathbb{N} = \{1, 2, \ldots\}$ and $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$. Let $\boldsymbol{\tau} = \{\tau_r\}_{r\in\mathbb{N}_0}$ be a sequence of discrete times with state space $\mathbb{N}$, where τ_r is the rth jump time with $1 = \tau_0 < \tau_1 < \tau_2 < \cdots < \tau_r < \tau_{r+1} < \ldots$. Let $J = \{J_r\}_{r\in\mathbb{N}_0}$ be a sequence with state space $\mathbb{S}$, where J_r is the system state at the rth jump or transition. We define the sojourn times of the states by $\mathbf{D} = \{D_r\}_{r\in\mathbb{N}}$ with state space $\mathbb{N}$, where $D_r = \tau_r - \tau_{r-1}$ for all $r \in \mathbb{N}$ is the sojourn time in state J_{r-1} before the rth jump. A two-dimensional sequence of random variables $\{\mathbf{J}, \boldsymbol{\tau}\} = \{J_r, \tau_r\}_{r\in\mathbb{N}_0}$ is said to be a Markov renewal chain if it satisfies almost surely

$$\begin{aligned}&\Pr\left(J_{r+1} = j,\, \tau_{r+1} - \tau_r = d \,|\, (J_0,\, \tau_0),\, (J_1,\, \tau_1),\, \ldots,\, (J_r = i,\, \tau_r)\right)\\ &\quad = \Pr\left(J_{r+1} = j,\, \tau_{r+1} - \tau_r = d \,|\, J_r = i\right)\Big),\end{aligned} \tag{1}$$

for all $i, j \in \mathbb{S}$ and $d \in \mathbb{N}$. This means that in a Markov renewal chain, the joint probability of the next state $J_{r+1} = j$ and the sojourn time D_{r+1} in the current state $J_r = i$ depends only on the current state. If the above expression is independent of r, then $\{\mathbf{J}, \boldsymbol{\tau}\}$ is called homogeneous, and the discrete-time semi-Markov kernel/matrix q is defined by

$$q_{ij}(d) = \Pr\left(J_{r+1} = j,\, D_{r+1} = d \,|\, J_r = i\right). \tag{2}$$

which satisfies that $0 \le q_{ij}(d) \le 1$, $\sum_{l=1}^{d} q_{ij}(l) \le 1$ for $i, j \in \mathbb{S}$ and $\sum_{d\in\mathbb{N}} \sum_{j\in\mathbb{S}} q_{ij}(d) = 1$, $i \in \mathbb{S}$. It is usually assumed that $q_{ij}(0) = 0$, $i, j \in \mathbb{S}$, that is, at any time point instant transitions are not allowed. Also, $q_{jj}(d) = 0$, that is, there is no self-transition (Barbu and Limnios, 2008; Malefaki et al., 2010; Pertsinidou and Limnios, 2015).

2.2 Semi-Markov chain

Consider a stochastic process $S = \{S_t\}_{t\in\mathbb{N}}$ associated with the Markov renewal chain $\{\mathbf{J}, \boldsymbol{\tau}\} = \{J_r, \tau_r\}_{r\in\mathbb{N}_0}$ such that $S_t = J_{n(t)}$, where $n(t) = \max\{r \in \mathbb{N}_0 \,|\, \tau_r \le t\}$ is the counting process of jumps or transitions in the interval $(1, t]$, then the process $\{S_t\}_{t\in\mathbb{N}}$ is called a semi-Markov chain. It follows that $S_t = J_r$ for $t \in [\tau_r, \tau_{r+1})$.

A semi-Markov chain is characterized by two components. The first component is an embedded first-order Markov chain $\mathbf{J}$ of the Markov renewal chain

$\{J, \boldsymbol{\tau}\}$ that models the transitions between distinct states at transition/jump times. The transition probability matrix is then $\mathbf{A} = \{a_{ij}\}_{i,j\in\mathbb{S}}$ defined by

$$a_{ij} = \Pr\left(J_{r+1} = j \,|\, J_r = i\right) = \Pr(S_{\tau_{r+1}} = j \,|\, S_{\tau_r} = i),\ r \in \mathbb{N}_0, \tag{3}$$

such that $\sum_{j=1, j\neq i}^{m} a_{ij} = 1$ and $a_{jj} = 0$, that is, self-transitions are not allowed. Therefore, in a semi-Markov chain, the Markov property is only satisfied at jump times, which is the reason for the name 'semi-Markov'. The initial probability distribution of $\mathbf{J}$ is $\boldsymbol{\pi} = \{\pi_j\}_{j\in\mathbb{S}}$ defined by

$$\pi_j = \Pr(J_0 = j) = \Pr(S_1 = j), \tag{4}$$

with $\sum_{j=1}^{m} \pi_j = 1$. The second component is arbitrary sojourn time distributions for each distinct states visited in the embedded Markov chain. The distribution of sojourn times in a given state $J_r = j$ (following the rth jump) is defined by

$$p_j(d) = \Pr(D_{r+1} = d \,|\, J_r = j),\ d \in \mathbb{N}, \tag{5}$$

where $D_{r+1} = \tau_{r+1} - \tau_r$. If the chain remains in state j for duration d starting at time, say, $t+1$ and ending at time, say, $t+d$, we write $S_{t+1} = j, \ldots, S_{t+d} = j$ and denote it by $S_{(t+1):(t+d)} = j$. Then, Eq. (5) can also be written as

$$p_j(d) = \Pr\left(S_{\tau_{r+1}} \neq j,\ S_{\tau_r:(\tau_r+d-1)} = j \,|\, S_{\tau_r} = j,\ S_{\tau_{r-1}} \neq j\right). \tag{6}$$

The probability that the embedded Markov chain is in state j at zero jump time $\tau_0 = 1$ and remains in this state for duration d before the first jump at time $\tau_1 > 1$ is

$$\begin{aligned} \Pr\left(S_{1:d} = j,\ S_{d+1} \neq j\right) &= \Pr(J_0 = j) \Pr(D_1 = d \,|\, J_0 = j,\ J_1 \neq j) \\ &= \pi_j p_j(d),\quad d \in \mathbb{N}. \end{aligned} \tag{7}$$

The next visited state can be determined by the transition probability matrix of the embedded Markov chain when the sojourn time of a given state expires.

Let $\mathbb{N}_T = \{1, 2, \ldots, T\}$ be a finite index set. Suppose that $\{S_t\}_{t\in\mathbb{N}_T}$, also written as $S_{1:T}$, is the state sequence of the semi-Markov chain (SMC) with the state space $\mathbb{S}$. Let $n(T)$ be the counting process for the number of jumps or transitions between distinct states over the sampling interval $(1, T]$. Let $J_0, \ldots, J_{n(T)}$ be the visited state sequence at jumps $0, \ldots, n(T)$ with the respective sojourn times denoted by $D_1, \ldots, D_{n(T)+1}$ satisfying $D_1 + D_2 + \cdots + D_{n(T)+1} = T$, and the jump times be defined at $\tau_r = \sum_{l=1}^{r} D_l + 1$ for $r = 1, \ldots, n(T)$ with $\tau_0 = 1$. Then the sample path of an SMC can be written as $S_{1:T} = \{(J_0, D_1), (J_1, D_2), \ldots, (J_{n(T)}, D_{n(T)+1})\}$. In this path, sojourn in the last visited state finishes at time T, which is not a desirable assumption for many practical applications. Also, it is often unknown whether the last visited state in the semi-Markov chain transits to another state after time T. We therefore consider the last sojourn time as right-censored, and denote it by $D^+_{n(T)+1}$. We have $S_{1:T} = \{(J_0, D_1), (J_1, D_2), \ldots, (J_{n(T)}, D^+_{n(T)+1})\}$, where $D^+_{n(T)+1} \geq T - \sum_{l=1}^{n(T)} D_l > T - \tau_{n(T)}$, and $n(T)$ is the counting process of the number of jumps

depending on time T. Consequently, the sum of the sojourn times in the first $n(T)$ state visits is smaller than the length of the observations (T), but the overall sojourn times in $n(T)+1$ visits may exceed T.

3 Inhomogeneous semi-Markov chain

Following the assumptions of an SMC, we define a discrete-time ISMC as a discrete-time SMC when the distributions of sojourn times in each visited states depend on time. More specifically, the semi-Markov chain $\mathbf{S} = \{S_t\}_{t\in\mathbb{N}}$ becomes an ISMC if

$$p_j(d,\, t) = \Pr(D_{r+1} = d \,|\, J_r = j,\, \tau_r = t), \quad d,\, t \in \mathbb{N}. \tag{8}$$

Here, we assume that the sojourn time distribution of state j at the rth transition depends on the rth transition time, that is the time of entering state j. However, the state duration distributions can be defined to depend on time based on a particular practical application of ISMC. Following (6), we also have

$$p_j(d,\, t) = \Pr\left(S_{\tau_{r+1}} \neq j,\, S_{\tau_r:(\tau_r+d-1)} = j \,|\, S_{\tau_r} = j,\, \tau_r = t,\, S_{\tau_{r-1}} \neq j\right). \tag{9}$$

When we consider an ISMC censored at fixed arbitrary interval $\mathbb{N}_T$, the sample path is $S_{1:T} = \{(J_0,\, D_1), (J_1,\, D_2),\, \ldots,\, (J_{n(T)},\, D^+_{n(T)+1})\}$ with joint probability

$$\begin{aligned}
&\Pr\Big\{(J_0 = j_0,\, D_1 = d_1), (J_1 = j_1,\, D_2 = d_2),\, \ldots,\\
&\qquad (J_{n(T)} = j_{n(T)},\, D^+_{n(T)+1} = d^+_{n(T)+1} > T - t_{n(T)})\Big\}\\
&\quad = \pi_{j_0}\left(\prod_{r=0}^{n(T)-1} p_{j_r}(d_{r+1},\, t_r) a_{j_r j_{r+1}}\right)\bar{P}_{j_{n(T)}}(d^+_{n(T)+1},\, t_{n(T)}),
\end{aligned} \tag{10}$$

where

$$\begin{aligned}
\bar{P}_{j_{n(T)}}(d^+_{n(T)+1},\, t_{n(T)}) = \Pr\Big(&D^+_{n(T)+1} \geq T - t_{n(T)} + 1 \,|\, J_{n(T)} = j_{n(T)},\\
&\tau_{n(T)} = t_{n(T)}\Big),
\end{aligned} \tag{11}$$

is the survival function of the censored duration in the last visited state. Note that the lower case letters j_r, d_r and t_r denote the realizations of the variables J_r, D_r and τ_r. An ISMC is described by an initial distribution, transition probability matrix of the embedded Markov chain and the parameters of time-dependent distributions for state durations.

4 Inhomogeneous hidden semi-Markov models

An IHSMM is defined as a bivariate stochastic process in which the unobservable process is an ISMC and the observable process is conditionally independent given the underlying ISMC. Suppose that the observed process has a sequence of

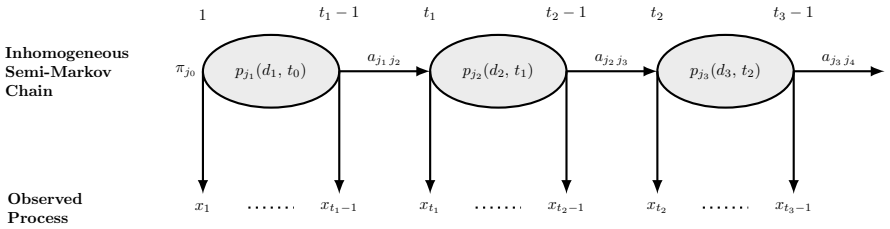


Fig. 1 The graphical structure of an IHSMM

observations $\{X_t\}_{t\in\mathbb{N}_T}$, where X_t is the observation at time t, and $\{S_t\}_{t\in\mathbb{N}_T}$ is the corresponding hidden sequence of ISMC with the state space $\mathbb{S}$ defined in Sect. 3. Figure 1 shows the schematic structure of an IHSMM.

Let $X_{t_r:(t_{r+1}-1)}$ be the sequence observed from the visited state $J_r = j_r$ during the sojourn of length $d_{r+1} = t_{r+1} - t_r$, and $f_{j_r}(x_{t_r+l-1})$ be the probability density function (PDF) of X_{t_r+l-1} given the state $S_{t_r+l-1} = j_r$ for all $l = 1, \ldots, d_{r+1}$. Then, the joint state-dependent observation probability distribution of the observed sequence $X_{t_r:(t_{r+1}-1)}$ given the state sequence $S_{t_r:(t_{r+1}-1)} = j_r$ is defined by

$$f\left(x_{t_r:(t_{r+1}-1)} \mid S_{t_r:(t_{r+1}-1)} = j_r\right) = \prod_{l=1}^{d_{r+1}} f_{j_r}(x_{(t_r+l-1)}). \tag{12}$$

We assume that the initial state starts at time 1 and the sojourn time in the last visited state is censored at time T. The likelihood function of the right-censored IHSMM, $\{X_t, S_t\}_{t\in\mathbb{N}_T}$, is

$$\begin{aligned}
\mathcal{L}(\Theta) &= \sum_{\mathbf{S}} \Pr(X_1, \ldots, X_T, S_1, \ldots, S_T \mid \Theta) \\
&= \sum_{\substack{j_0, \ldots, j_{n(T)} = 1; j_r \neq j_{r+1} \\ d_1, \ldots, d^+_{n(T)+1}}}^{m} \pi_{j_0} p_{j_0}(d_1, t_0) \prod_{l_0=1}^{d_1} f_{j_0}(x_{l_0}) \prod_{r=1}^{n(T)-1} a_{j_{r-1} j_r} p_{j_r}(d_{r+1}, t_r) \\
&\quad \times \prod_{l_r=1}^{d_{r+1}} f_{j_r}(x_{t_r+l_r-1}) a_{(j_{n(T)-1} j_{n(T)})} \bar{P}_{j_{n(T)}}(d^+_{n(T)+1}, t_{n(T)}) \\
&\quad \times \prod_{l_{n(T)}=1}^{T-t_n(T)+1} f_{j_{n(T)}}(x_{t_{n(T)}+l_{n(T)}-1}),
\end{aligned} \tag{13}$$

where Θ are the model parameters including the parameters of the embedded Markov chain, the time-dependent sojourn time distributions and the state-dependent observation probability distributions, and $n(T)$ is the counting process of the number of jumps depending on time T. The above likelihood function reduces to the likelihood function of a right-censored HSMM (Guédon, 2003) when the explicit state duration distributions are independent of time. An extension of the forward-backward

algorithm for evaluating the likelihood function of IHSMMs in (13) and discussion of related numerical issues are detailed in the supplementary file in Sections S.1 and S.3, respectively.

In recent years, direct numerical optimization algorithms have become very popular because of their speed of convergence relative to the EM algorithm. We can fit an $m-$state HSMM and IHSMM to a data set and estimate the parameters using direct numerical optimization of the log-likelihood functions that can be computed using the forward recursion in Algorithm 1 in the supplementary file. After fitting the model, the optimal state sequence for an $m-$state IHSMM can be obtained using the Viterbi path (Viterbi, 1967), which is detailed in Section S.4 and Algorithm 3 in the supplementary file.

5 Renewal process and HMMs

Let a point process be defined by $\{H_k\}_{k\in\mathbb{N}_0}$, where $H_0 < H_1 < \cdots < H_k < \ldots$ are occurrences or arrivals of random events over time. The interevent times are given by $X_k = H_k - H_{k-1}$, $k = 1, 2, \ldots$. A point process is called a renewal process when the interevent times $\{X_k\}_{k\in\mathbb{N}}$ are positive, independent and identically distributed. Each occurrence time H_k is called a renewal of the process, as the process is assumed to start over at each occurrence time. Also, each renewal term H_k can be written as $H_k = X_1 + \ldots + X_k$ with $H_0 = 0$ by convention.

The hazard rate or point process conditional intensity is defined as an instantaneous rate of occurrences of event in an interval conditional on the past. For a renewal process, it is written as

$$\lambda(h) = \frac{g(h - h_s)}{1 - G(h - h_s)}, \qquad h > h_s, \tag{14}$$

where h_s, a realization of H_s, is the time of occurrence of the most recent event before time h, $G(.)$ is some cumulative distribution function (CDF), and $g(.)$ is the corresponding PDF. The time $h - h_s$ is called the elapsed time that determines the likelihood of the next event since the last event occurred. Previous events make their contribution through the parameter estimates. If we assume $g(.)$ to be the exponential density, then the process becomes a homogeneous Poisson process.

The occurrences of volcanic eruptions can be regarded as a point process. A renewal process was first suggested by Wickman (1966) for volcanic eruptions and has since been often applied in volcanology (e.g., (Bebbington and Lai, 1996; Varley et al., 2006; Turner et al., 2008, 2009; Wang and Bebbington, 2012; Bebbington, 2013)). In this study, we assume that the complete sequence of volcanic eruptions form a gamma renewal process, that is, the interevent times are independently and identically gamma distributed (De la Cruz-Reyna and Carrasco-Núñez, 2002; Wang and Bebbington, 2012). The gamma distribution has significance because of its properties. It is flexible and generalizes the exponential distribution. Another attractive property of this distribution is the additive property: the sum of gamma random variables is a gamma random variable.

Reliable hazard estimates may ideally be obtained using completely observed volcanic eruption records. However, as discussed in Sect. 1, the missingness of events is a common problem to volcanic eruption records, and hence can affect the accuracy of hazard estimates. The number of missing events between each pair of consecutively observed events is usually unknown and can be considered as a hidden process in the framework of HMMs. For this reason, Wang and Bebbington (2012) proposed to use HMMs for incomplete volcanic eruption records. In their model, the observed process was modelled by a renewal process, and the hidden process was the first order homogeneous Markov chain with m (say) states representing 0, 1, 2, … , $m-1$ number of missing events, respectively, between each pair of consecutively observed events.

Wang and Bebbington (2012) assumed that the probability of missing events is homogeneous throughout the record, a reasonable assumption for their data which were from sedimentary cores. However, most volcanic eruption records are based on a combination of written records and/or geological outcrops, and so have an uneven distribution of missing events throughout time, as the earlier part of a volcanic record is typically more incomplete than the recent part. This inhomogeneous probability of missing events motivates us to use an inhomogeneous semi-Markov chain as the underlying process.

6 An IHSMM with gamma renewal processes

In this section, we define a particular case of IHSMMs. We assume that there are m states of the underlying ISMC. State 1 represents no missing event between each pair of consecutively observed events in a point process record, and hence is the state of completeness. States 2, … , m represent 1, … , $m-1$ number of missing events, respectively, between each pair of consecutively observed events, hence are the states of incompleteness. In State 1, suppose that the observed process is a gamma renewal process. Then, the state-dependent observation probability distribution of the interevent times X_t is a gamma distribution with PDF

$$f_1(x_t) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x_t^{\alpha-1} \exp(-x_t/\beta), \quad x_t > 0, \tag{15}$$

where α and β are shape and scale parameters, respectively, and Γ is the gamma function defined as $\Gamma(s) = \int_0^\infty w^{s-1} e^{-w} dw$, $s > 0$.

When the ISMC is in State j, $j = 2, \ldots, m$, the interevent times X_t is the sum of j interevent times $Y_1, \ldots, Y_j$. That is,

$$X_t = \sum_{k=1}^{j} Y_k \tag{16}$$

where Y_k are independent and identically gamma distributed with PDF as in (15). The state-dependent observation probability distribution in state j, $j = 2, \ldots, m$, is then obtained by the additive property of gamma distribution,

$$f_j(x_t) = \frac{1}{\beta^{j\alpha}\Gamma(j\alpha)} x_t^{j\alpha-1} \exp(-x_t/\beta), \quad x_t > 0. \tag{17}$$

Since the unobserved process of an IHSMM is an ISMC, in which the associated embedded Markov chain has zero self-transition probabilities, the duration of sojourning in a state is at least 1. Now, we need to choose the time-dependent state duration distributions $p_j(d, t)$ defined in (8) for the underlying ISMC. There can be many choices for the probability distributions of state durations, for example, the Poisson distribution (Russell and Moore, 1985; Rossi et al., 2015) and some probability mass functions from the exponential family (Mitchell and Jamieson, 1993). In order to model the mean state durations varying over time, we propose to use the shifted Poisson distribution for state durations, i.e., $p_j(d, t)$ is assumed to follow a shifted Poisson distribution with parameter $\mu_j(t)$, a function of time t. Thus, we can write

$$p_j(d, t) = \exp(-\mu_j(t)) \frac{[\mu_j(t)]^{d-1}}{(d-1)!}, \qquad d = 1, 2, \ldots, \tag{18}$$

where $\mu_j(t) > 0$, $j = 1, 2, \ldots, m$, can be interpreted as the expected sojourn time in state j dependent on time t.

Historical records of volcanic eruptions are usually more incomplete in earlier times than in latter times (e.g., (Furlan, 2010; Deligne et al., 2010; Brown et al., 2014; Rougier et al., 2016)). Also, we are assuming that a record of volcanic eruption onsets forms a renewal process in which only the elapsed time since the last eruption determines the likelihood of the next eruption. Given the time-inhomogeneity in the completeness of historical records, we define the expected sojourn time in State 1, $\mu_1(t)$, to be an increasing function of the last occurrence time H_{t-1}, where $H_{t-1} = \sum_{k=1}^{t-1} X_k$ assuming $H_0 = 0$ and the expected sojourn time in state j, $\mu_j(t)$ for $j = 2, \ldots, m$, to be a decreasing function of the last occurrence time H_{t-1}. It is quite possible that the expected sojourn time in State 1 may increase at an increasing or decreasing rate, and the sojourn time in state $j = 2, \ldots, m$ may decrease at an increasing or decreasing rate through time or some interval of time. In addition, $\mu_j(t)$, $j = 1, 2, \ldots, m$, can take any value on positive real line. For these reasons, we define $\mu_j(t)$ by a generalized logistic function (Richards, 1959), selected because of its flexibility in interpolating between different regimes. For State 1, $\mu_1(t)$ is defined by

$$\mu_1(t) = \frac{A_1}{1 + \exp(-C_1 h_{t-1} + B_1)} + D_1, \tag{19}$$

and for state $j = 2, \ldots, m$, $\mu_j(t)$ is defined by

$$\mu_j(t) = \frac{A_j}{1 + \exp(C_j h_{t-1} - B_j)} + D_j, \tag{20}$$

where $A_j, C_j, D_j > 0$ and $-\infty < B_j < \infty$ for $j = 1, 2, \ldots, m$ are the parameters. The parameters A_j and D_j control the limiting values of $\mu_j(t)$ such that $\mu_1(t)$ increases from D_1 to $A_1 + D_1$ and $\mu_j(t)$, $j = 2, \ldots, m$ decreases from $A_j + D_j$ to D_j. The

parameter C_j is the rate of increase or decrease and is usually measured in units of time; the parameter B_j controls the point of inflection. To see how the shape of $\mu_j(t)$ changes with different values of B_j and C_j, Fig. 2 illustrates $\mu_2(t)$ with $A_2 = 14$ and $D_2 = 3$ and varying values of B_2 and C_2.

In real-world applications of HMM type models, the selection of a model with an appropriate number of states is an important problem and is called the order identification problem. Since we are dealing with HMM modelling of renewal processes, we describe two methods for selecting an appropriate model for the given data.

After fitting HMMs, HSMMs and IHSMMs with different orders and estimating the parameters, we use the Akaike Information Criterion (AIC) (Akaike, 1974) to select a model, which is defined by

$$\text{AIC} = -2 \log \mathcal{L}(\hat{\Theta}) + 2k, \tag{21}$$

where k is the number of parameters to be estimated and $\log \mathcal{L}(\widehat{\Theta})$ is the numerical value of the log-likelihood of the model at its maximum point. The second term on the right-hand side of (21) is a penalty term that balances the fit against the model complexity. This is the tradeoff between underfitting and overfitting risks. A model with the lowest AIC value is considered the best approximation of the true model generating the data.

To assess whether the fitted model truly describes the stochastic nature of the data, we use residual analysis developed by Berman (1983) and Ogata (1988) based on the time-rescaling theorem. Suppose we have occurrence times, $h_0 < h_1 <, \ldots, < h_T$, from a point process with conditional intensity function $\lambda(h)$ and define the transformation

$$\tau_k = \Lambda(h_k) = \int_0^{h_k} \lambda(h)\, dh, \tag{22}$$

for $k = 1, \ldots, T$. Then the rescaled times, τ_k's, form a Poisson process with unit rate ((Daley and Vere-Jones, 2003), Theorem 7.4.I). Once a model has been fitted to the

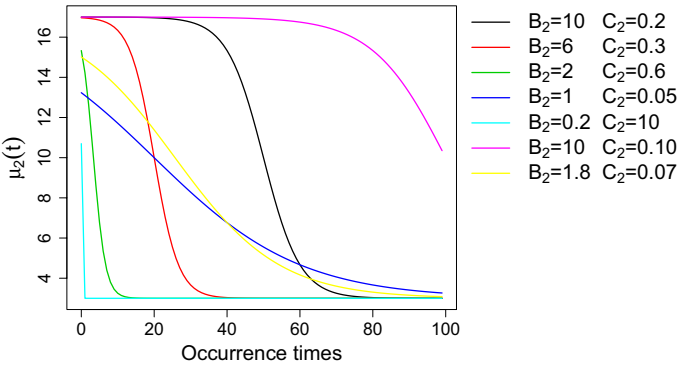


Fig. 2 Different shapes of $\mu_2(t)$ depending on different values of B_2 and C_2

Table 1 A 3−state IHSMM for simulation study

States	j	1	2	3
Initial distribution	π_j	0.23	0.60	0.17
Transition probability matrix	a_{ij}			
	1	0.00	0.65	0.35
	2	0.55	0.00	0.45
	3	0.70	0.30	0.00
Poisson distribution for	$\mu_j(t)$			
Time-dependent state durations	A_j	15	12	10
	B_j	6	4	9
	C_j	1/18	1/30	1/14
	D_j	3	2	1
Gamma distribution	α	9	18	27
	β	0.25	0.25	0.25

Table 2 Redefined values of change rate parameter, C_j, $j = 1, 2, 3$

Parameter	Sample size			
	200	500	1000	2000
C_j	$(\beta/0.4)C_j$	$(\beta/1)C_j$	$(\beta/2)C_j$	$(\beta/4)C_j$

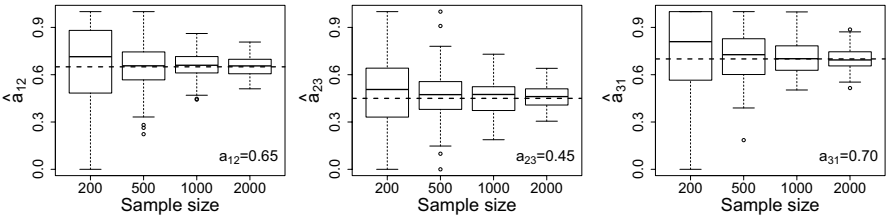
record of a point process, we can compute from its estimated conditional intensity the rescaled times $\tau_k = \Lambda(h_k)$. If the model describes the data well, then the residual point process τ_k is a stationary Poisson process with unit rate. In order to perform this, the cumulative number of events versus transformed times τ_k can be plotted with 95% and 99% confidence limits of the Kolmogrov-Smirnov (KS) statistic under the null hypothesis of uniformity to see if the cumulative curve falls within limits. Also, if the selected model is correct, then, $E_k = \tau_k - \tau_{k-1}$ are independent exponential random variables with unit mean. In other words, $U_k = 1 - \exp(-E_k)$ are independent uniform random variables on [0, 1). To check the independence of the transformed time intervals, we will use a scatter plot of E_k against E_{k+1} to observe any particular pattern along with a correlation test (Wang, 2010).

7 Simulation study

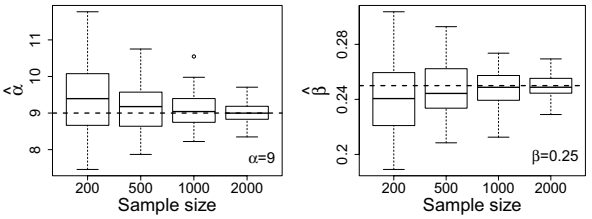
For the IHSMM defined in Sect. 6, we perform a simulation study to check the consistency of the parameter estimators. In order to obtain the maximum likelihood estimates of the model parameters from the simulated data, we perform an unconstrained numerical optimization of the negative log-likelihood in R (R Core Team, 2017) using the optimr() function of the R package optimx (Nash and Varadhan, 2011; Nash, 2014).

We consider an example of a 3-state IHSMM with a gamma renewal process using parameter values in Table 1. We consider samples of size 200, 500, 1000 and 2000. To make the rate of change in the expected sojourn times, $\mu_j(t)$, to be comparable in small and large samples, we use the redefined values of C_j for each sample size in Table 2.

For each of 100 simulations, starting from the initial state $j = 1, 2, 3$ at time $t = 1$, the parameter $\mu_j(1)$ of the shifted Poisson distribution $p_j(d_1, 1)$ depends on the last occurrence time h_0 that is taken to be zero by convention. After simulating duration d_1 in the initial state j at time $t = 1$ according to the duration distribution $p_j(d_1, 1)$ with parameter $\mu_j(1)$, the interevent times $x_1, \ldots, x_{d_1}$ are generated using the gamma distribution with parameters $(j\alpha, \beta)$. The occurrence times are, then, obtained as $h_1 = x_1$, $h_2 = x_1 + x_2$, $\ldots$, $h_{d_1} = x_1 + \cdots + x_{d_1}$. At time $d_1 + 1$, the ISMC transits to state i with the transition probability a_{ji} and stays d_2 duration according to the probability distribution $p_i(d_2, d_1 + 1)$ with parameter $\mu_i(d_1 + 1)$ depending on the last occurrence time h_{d_1}. The interevent times $x_{d_1+1}, \ldots, x_{d_1+d_2}$ are generated according to the gamma distribution with parameters $(i\alpha, \beta)$ and the occurrence times are $h_{d_1+1} = x_1 + \cdots + x_{d_1} + x_{d_1+1}$, $\ldots$, $h_{d_1+d_2} = x_1 + \cdots + x_{d_1+d_2}$. The procedure continues until the required sample size is generated. After performing the direct numerical optimization of the log-likelihood function for each of 100 simulated sequences, boxplots of the estimates of the transition probabilities and the parameters in the gamma distribution are shown in Fig. 3. We observe that by increasing the sample size, the interquartile ranges of boxplots have reduced and the parameter estimates are closer to the true values, indicating the conjecture of consistency of the estimates. To see how the estimated expected sojourn times $\hat{\mu}_j(t)$ behave, we plotted $\hat{\mu}_j(t)$, $j = 1, 2, 3$, for all 100 simulations



(a) The transition probabilities



(b) The gamma distribution

Fig. 3 MLEs of the transition probabilities and the parameters of the gamma renewal process from the simulation study. The dashed line represents the true parameter value

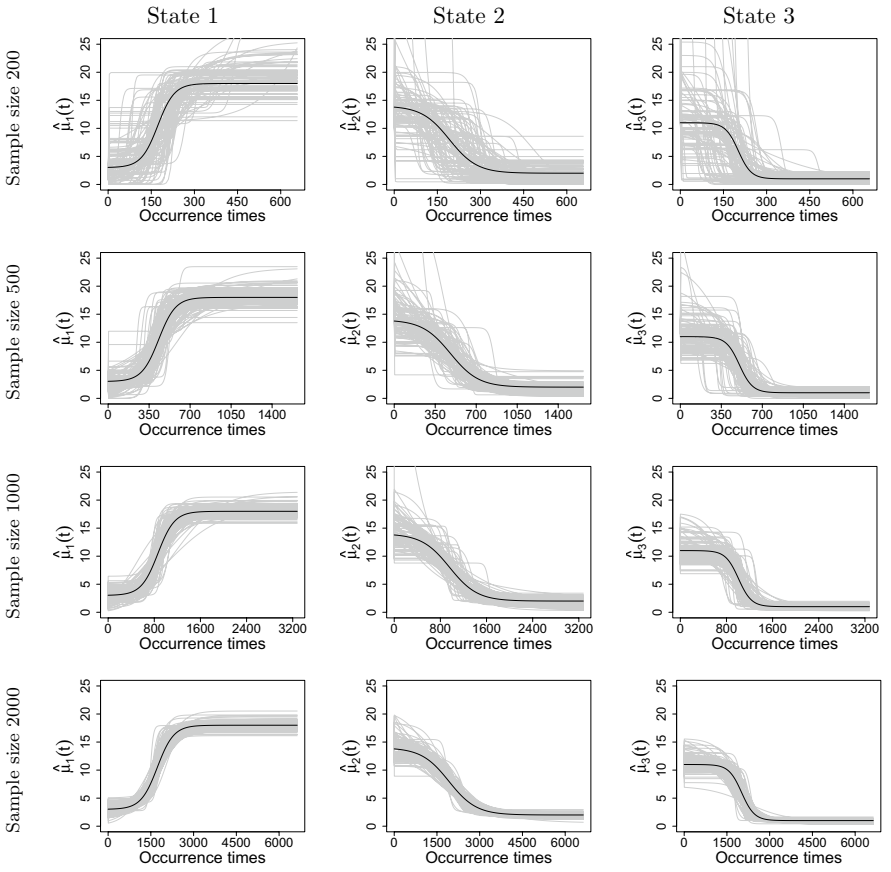


Fig. 4 MLEs of $\mu_j(t)$ from 100 simulations for sample sizes 200, 500, 1000 and 2000. The black curve is obtained from the true parameters

for each of the sample sizes in Fig. 4. We see that the estimated expected sojourn times, $\hat{\mu}_j(t)$, $j = 1, 2, 3$, have more variations for sample size 200, and become consistent around the true black curve with the increased sample size. Thus, the parameter estimates get improved with increasing number of observations.

8 Application to a global volcanic eruption record

In this section, we present an application of the proposed model described in Sect. 6 to a global volcanic eruption catalogue.

8.1 Data description

We consider the catalogue of global volcanic eruptions compiled by Hayakawa (1997). After deleting duplicates, this catalogue contains 286 volcanic eruption events with a minimum magnitude of 3.8 in the past two thousand years. The magnitude of volcanic eruption is $M = \log_{10} m - 7$, where m is the estimated mass (from both tephra and lava) in kilograms. This record has also been investigated by Coles and Sparks (2006) and Furlan (2010).

The observed eruptions are shown in Fig. 5. We see that the eruption catalogue appears relatively more complete in recent years, particularly for events with magnitude less than 5. The inhomogeneity of the record decreases as the magnitude of volcanic eruption increases. This might represent the under-reporting of small eruption events in the years before AD 1500 whose deposits might not have been seen due to erosion (Furlan, 2010; Coles and Sparks, 2006). This incompleteness depends on time. The older the record is, the more incomplete it is.

Note that we are using an aggregate of eruption events from many independent volcanoes. From moderately large eruptions and an appropriate size scale, the number of eruptions occurring per unit time should follow a Poisson distribution (Guttorp and Thompson, 1991; De la Cruz-Reyna, 1991, 1993). Accordingly, in our proposed IHSMM, the interevent times follow an exponential distribution when the observed process is complete, i.e., in State 1. In other words, the observed process in State 1 follows a gamma renewal process with parameters $(1, \beta)$; in State 2 it follows a gamma renewal process with parameters $(2, \beta)$ and so on. Thus, the Poissonian behaviour of the global volcanic eruption record has reduced the number of parameters to be estimated by one.

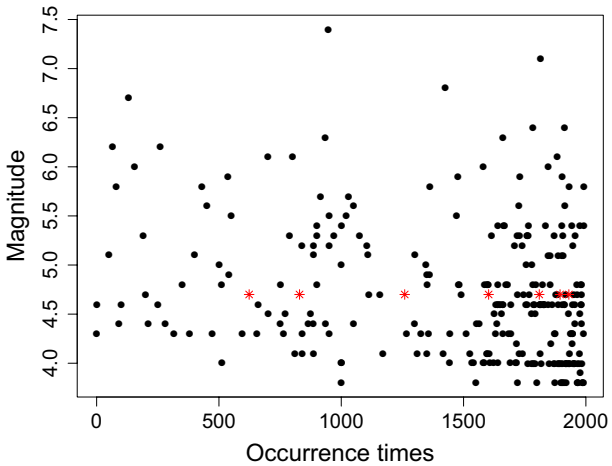


Fig. 5 Global volcanic eruption catalogue with magnitude ≥ 3.8. The red stars represent the unknown magnitudes plotted at the average magnitude

8.2 Data analysis

We consider HMMs, HSMMs and IHSMMs with 3, 4, 5, 6 and 7 hidden states to fit on the global volcanic eruption record. We used a logit-type transformation $\log\left((\phi_j - \phi_{\min})/(\phi_{\max} - \phi_j)\right)$ to restrict the possible range of the parameter estimates in order to avoid numerical instability. The models with states defined in Sect. 6 were fitted to the data. According to AIC, the 4-state and 5-state IHSMMs appeared to be the best model for this data set, as their AIC values were 0.67 apart (see Table S.1 in the supplementary file). However, the residual analysis showed that the residual process from the selected 4-state and 5-state IHSMMs deviate from a stationary Poisson process (see Figure S.2 in the supplementary file), suggesting that there might be more missing observations in some parts of the catalogue that were not captured by these models. The residual processes for the 6- and 7- state IHSMMs also deviate from a stationary Poisson process with unit rate. Increasing the order of IHSMM to a higher number of states may produce a residual process to be a stationary Poisson process. However, it will increase the number of parameters dramatically.

Moreover, the Viterbi paths for the 4-state and 5-state IHSMMs (see Figure S.1 in the supplementary file) did not give a plausible representation of the observed inter-event times. State 2 in the 4-state model and State 4 in the 5-state model had only one shortest visit over time, suggesting that these states could be combined with the other states. Also, it could be possible that a state may represent an average of many events between each pair of consecutively observed events in the record. Therefore, there is a need to redefine the states by combining some states to avoid increasing the number of parameters and to merge the effect of unnecessary states.

We call the case of state definitions in Sect. 6 as Case I. From the analysis in Case I, we merge the states linearly and nonlinearly and redefine them in Table 3 for the gamma renewal process and investigate the three new cases to select an appropriate model. Note that State 1 is the state representing complete observations and is not combined in any of these cases. The number of parameters (k), maximum log-likelihood (MLL) and AIC values for the fitted HMMs, HSMMs and IHSMMs are

Table 3 Redefinition of states in terms of number of missing events (say, N) and corresponding observed gamma renewal process (GRP)

State	Case II		Case III		Case IV	
	N	GRP	N	GRP	N	GRP
1	0	$(1, \beta)$	0	$(1, \beta)$	0	$(1, \beta)$
2	1 or 2	$(2.5, \beta)$	1, 2 or 3	$(3, \beta)$	1	$(2, \beta)$
3	3 or 4	$(4.5, \beta)$	4, 5 or 6	$(6, \beta)$	2 or 3	$(3.5, \beta)$
4	5 or 6	$(6.5, \beta)$	7, 8 or 9	$(9, \beta)$	4, 5 or 6	$(6, \beta)$
5	7 or 8	$(8.5, \beta)$	10, 11 or 12	$(12, \beta)$	7, 8, 9 or 10	$(9.5, \beta)$
6	9 or 10	$(10.5, \beta)$	13, 14 or 15	$(15, \beta)$	11, 12, 13, 14 or 15	$(14, \beta)$
7	11 or 12	$(12.5, \beta)$	16, 17 or 18	$(18, \beta)$	16, 17, 18, 19, 20 or 21	$(19.5, \beta)$

Table 4 No. of parameters (k), MLL and AIC in Cases II, III and IV

Model		Case II		Case III		Case IV	
	k	MLL	AIC	MLL	AIC	MLL	AIC
HMM							
3-state	9	−776.58	1571.16	−765.39	1548.79	−783.68	1585.36
4-state	16	−757.47	1546.95	−746.84	1525.68	−756.64	1545.29
5-state	25	−742.62	1535.25	−733.90	1517.80	−735.21	1520.42
6-state	36	−732.02	1536.03	−726.86	1525.72	−722.11	1516.23
7-state	49	−722.98	1543.96	−722.30	1542.60	−713.64	1525.29
HSMM							
3-state	9	−794.99	1607.98	−792.54	1603.07	−796.69	1611.38
4-state	16	−782.05	1596.11	−781.78	1595.57	−774.98	1581.97
5-state	25	−767.85	1585.71	−770.64	1591.27	−755.71	1561.42
6-state	36	−758.35	1588.70	−761.40	1594.81	−737.21	1546.43
7-state	49	−754.94	1607.89	−749.81	1597.62	−726.58	1551.16
IHSMM							
3-state	18	−744.63	1525.26	−731.11	1498.22	−759.28	1554.55
4-state	28	−726.83	1509.66	−715.67	1487.35	−729.96	1515.92
5-state	40	−714.99	1509.99	−704.47	1488.94	−712.03	1504.07
6-state	54	−709.73	1527.47	−700.63	1509.27	−697.94	1503.87
7-state	70	−702.11	1544.21	−697.18	1534.36	−692.88	1525.77

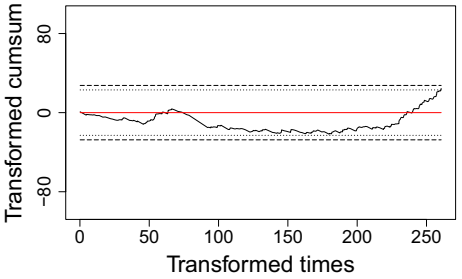


Fig. 6 The deviated cumulative number of events in the residual process from the stationary process versus the transformed times for the 4-state IHSMMs fitted to the global volcanic eruption catalogue. The central line at zero is the theoretical curve under the null hypothesis of stationary process. The dotted and dashed lines represent the two-sided 95% and 99% confidence limits of the KS statistic, respectively

listed in Table 4, which suggests that the 4-state IHSMM in Case III is the best fitted model with the smallest AIC value.

The residual analysis for the 4−state IHSMM in Case III is shown in Fig. 6. Note that in order to clearly observe the deviation of the computed residual process from the theoretical stationary process, we subtract the estimated curves from the theoretical curves. The residual process in Fig. 6 seems to be well approximated by a stationary Poisson process with unit rate. The residual process

for other higher state IHSMMs in the cases listed in Table 4 did not improve by much (see Figures S.3 to S.5 in the supplementary file). Also, the residual processes for HMMs and HSMMs did not appear to be well approximated by a stationary Poisson process in all cases (see the supplementary file). Since none of these HMMs and HSMMs have AIC values close to the smallest AIC value (from the 4-state IHSMM in Case III), we do not consider these models for further analysis. A discussion on the best models within Cases II and IV are given in the supplementary file.

We also check further assumptions of a stationary Poisson process mentioned in Sect. 6. Using the KS test of uniformity, the empirical distribution of U_k from the 4−state IHSMM in Case III is plotted in Fig. 7, which shows uniformity of U_k. Hence, there is no evidence to assert that the transformed interevent times, E_k, are not exponentially distributed. The scatter plots of E_{k+1} against E_k and U_{k+1} against U_k in Fig. 7 show no particular pattern of any association, suggesting the independence of E_k from this model. The t-test for the null hypothesis of zero correlation between E_{k+1} and E_k in this model produces a p-value of 0.203, further confirming that there is no evidence to reject the hypothesis that E_k are independent. We conclude that the residual process for the 4-state IHSMM in Case III follow a stationary Poisson process with unit rate satisfying the assumptions of independence and exponentiality.

We note that the 4−state IHSMM in Case III models a maximum number of missing events up to 9 between a pair of consecutively observed events with 28 parameters. In this 4-state model, States 2, 3 and 4 represent 2, 5, 8 missing events on average, respectively. Note that in order to redefine the states, combining the number of missing events do not affect the number of parameters to be estimated. It helps to seek the best model with an appropriate number of parameters and missing events between each pair of consecutively observed events in the volcanic eruption record. Thus, the overall analysis suggests that the 4-state IHSMM in Case III can be chosen as the best approximation of the given volcanic eruption record in terms of the number of parameters, AIC, residual analysis and the number of missing events represented by each state.

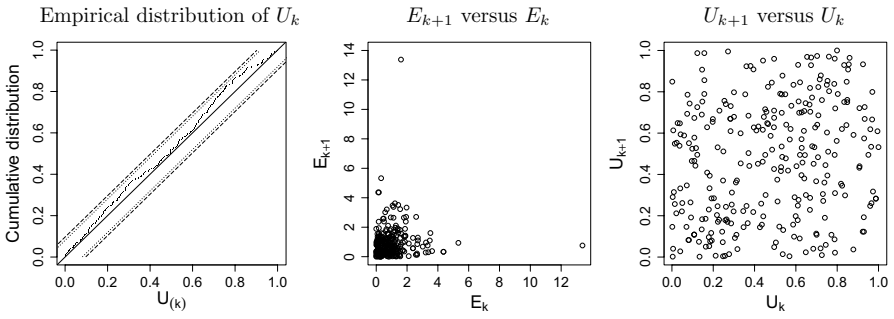


Fig. 7 Residual check for the 4-state IHSMM in Case III. Left: Empirical distribution of U_k, with the dotted and dashed lines indicating 95% and 99% confidence intervals of the KS statistic, assuming uniform distribution. Middle: Scatter plot of E_{k+1} versus E_k. Right: Scatter plot of U_{k+1} versus U_k

Table 5 Estimates of the 4- state IHSMM in Case III

States	j	1	2	3	4
Initial distribution					
	$\hat{\pi}_j$	1	0	0	0
Transition probability matrix	$\hat{a}_{ij}$				
	1	0.00	0.00	0.709	0.291
	2	0.00	0.00	0.00	1.00
	3	1.00	0.00	0.00	0.00
	4	0.314	0.686	0.00	0.00
Poisson distribution	$\hat{\mu}_j(t)$				
time-dependent state durations	$\hat{A}_j$	200.983	186.418	123.986	2.204
	$\hat{B}_j$	129.792	−4.653	17.544	12.711
	$\hat{C}_j$	0.081	0.003	0.028	0.024
	$\hat{D}_j$	1.199	0.000	0.914	$2.341e^{-5}$
Gamma distribution	α	1	3	6	9
	$\hat{\beta}$	2.930	2.930	2.930	2.930

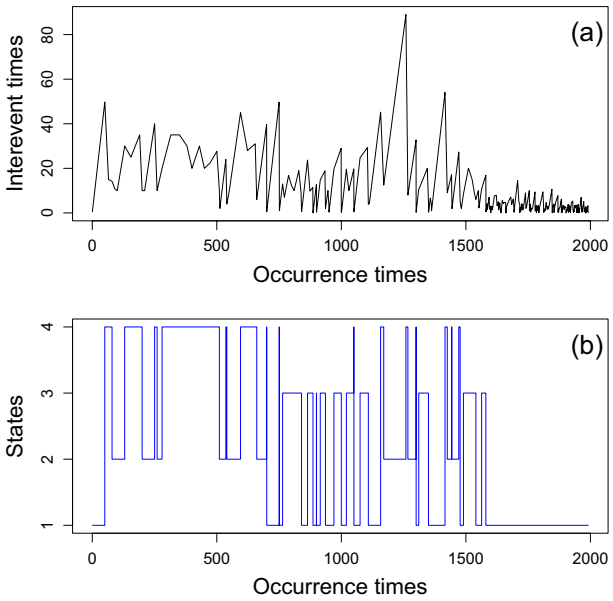


Fig. 8 **a** The observed interevent times, **b** The Viterbi path for the selected 4-state IHSMM in Case III

8.3 The selected 4-state IHSMM

The ML estimates of the best fitted 4-state IHSMM in Case III are listed in Table 5. Using these estimates the Viterbi path for the model is plotted in Fig. 8.

The estimated transition probability matrix is quite sparse, which suggests that we can classify the states into two regimes: a low-missing regime constituting the transitions between State 1 and State 3, i.e., $1 \leftrightarrow 3$, and a high-missing regime having the transition pattern $4 \rightarrow 2 \rightarrow 4$. This is apparent in the Viterbi path in Fig. 8 before AD 1580. At the begining, the record is in the state of completeness in AD 0 but is then incomplete for a long time. Note that the missingness of small eruptions (with magnitude ≤ 5.5) is more time-inhomogeneous than large eruptions as indicated in Fig. 5. Therefore, the low-missing regime may signal that it is more likely to have 4, 5 or 6 missing eruptions of small magnitudes before and after no missing eruption of large magnitude. As the large eruptions are less frequent and are less frequently missed, they are likely to be missed in State 4 of the high-missing regime. State 2 has most likely 1, 2 or 3 missed eruptions, but here it seems to be catering for 'doublets', i.e., a couple of eruptions relatively close together in long strings of missing events ($4 \rightarrow 2 \rightarrow 4$). Also, the Viterbi path suggests that the catalogue has no missing events after AD 1580.

The state-dependent gamma distributions in the selected 4−state IHSMM are plotted in Fig. 9, which shows that the state-dependent distributions for the states falling in low-missing and high-missing regimes do not overlap considerably. Notably, States 2 and 4 overlap with State 3 significantly, making it unidentifiable within a sequence of high-missing regime states. However, relatively less overlap between low-missing regime (1 and 3) and high-missing regime (2 and 4) states make these regimes identifiable. This suggests that the IHSMM balances between the features (shape and scale) of the observations in each state and the state durations and transitions.

To see how the time-dependent expected sojourn times in each state behave, we plot the estimated expected sojourn times, $\hat{\mu}_j(t)$, $j = 1, \ldots, 4$, in Fig. 10. Note that $\hat{\mu}_j(t)$ in high-missing regime states (2 and 4) decreases in earlier times and tends to become zero after AD 500. State 2 has the shortest expected sojourn time which decreases exponentially. However, the low-missing regime states (1 and 3) have lengthy estimated expected sojourn times. State 1 is more likely to have a short expected sojourn time before AD 1500, as the data in the earlier record is not

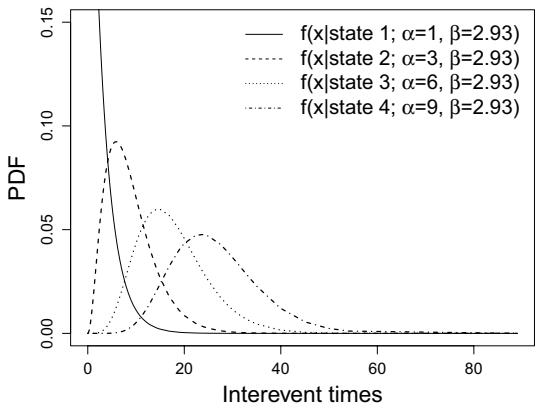


Fig. 9 Probability density function in each state of the 4-state IHSMM in Case III

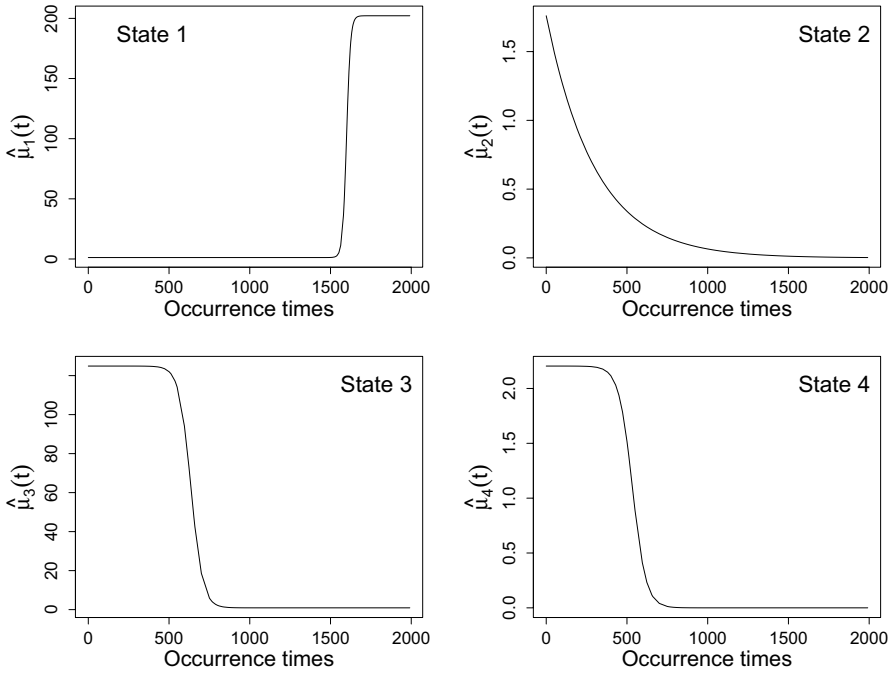


Fig. 10 Estimated expected sojourn times, $\hat{\mu}_j(t),\ j = 1, \ldots, 4$ for the 4-state IHSMM in Case III

completely observed. The sojourn time for State 1 increases from AD 1500 to AD 1600, and then persists for a long time period, which is also indicated by the Viterbi path. After a long sojourn before AD 500, the expected sojourn time of State 3 decreases between AD 500 and AD 700 and then decreases to as short as one.

One of the main objectives of the proposed model is to estimate the hazard. The hazard rate estimate defined in (14) can be obtained using the estimates of the gamma distribution in State 1 from the fitted 4−state IHSMM in Case III, which is $\hat{\lambda}(h_t) = 0.341$. For comparison, the estimated hazard rate by fitting a simple gamma renewal process with $\alpha = 1$ to the entire data set is 0.143, less than half of the estimated hazard rate from the IHSMM.

9 Discussion

We proposed a general class of IHSMMs for time-inhomogeneous processes, which models the time-inhomogeneity through time-dependent state durations of the underlying semi-Markov chain. We applied this class of models to study incomplete records of volcanic eruptions with time-varying missingness by introducing a particular IHSMM, where the hidden process is an ISMC that models the time-dependent number of missing events between successive recorded events through time-dependent Poisson state durations and the observed process is a gamma renewal

process. For the special case of IHSMMs, a simulation experiment was conducted to test a conjecture of consistency of parameter estimators. The new IHSMMs can be applied to other types of data with missing events that occur inhomogeneously in time and nonstationary processes.

The application of the proposed IHSMMs to a global volcanic eruption catalogue (Hayakawa, 1997) demonstrated the importance of using an IHSMM instead of an HMM or HSMM when the data exhibits inhomogeneity. Since the catalogue has only 286 data points, we considered 3-state to 7−state IHSMMs, HSMMs and HMMs, each with four possible cases of state definitions in order to find an acceptable model. The AIC and residual analysis were used to select the model that captures the main features of the data.

The residual analysis in Case I suggested that the IHSMM could fit the data well if models with a larger number of states were considered. In order to avoid a very large number of parameters for such a small data set, we merged the states linearly in Cases II and III and nonlinearly in Case IV. When the cases are considered individually, the 5-state IHSMMs in Cases II and IV fitted the data well as indicated by the residual analysis, suggesting that the number of missing observations in between each pair of consecutively observed events can reach up to 8 and 10 in some parts of the catalogue, respectively. However, the 4-state IHSMM in Case III performed better in terms of AIC than the 5-state models in Cases II and IV, and represented up to 9 missing events in some parts of catalogues, the average of the maximum number of missing events in the 5-state models in Cases II and IV, with less number of parameters. Overall, we found the 4-state IHSMM in Case III as the best fitted model for the catalogue. This is an acceptable model for the global volcanic eruption catalogue regarding the number of parameters, state definitions, AIC and the residual analysis. Comparing Figs. 5 and 8, we note that the catalogue is relatively incomplete before AD 1580.

A possible way of merging the states could be to assign hidden state j as having observations from a gamma distribution with parameters (α_j, β), where $\alpha_j = \alpha \sum_{k=1}^{j} a_k$ and a_k are non-negative parameters to be estimated. Implementing this, we observed that the fitted observation distributions were broader and overlapped more than those in Fig. 9. While this resulted in an improvement in likelihood and hence AIC, the selected model was a 3-state IHSMM with poor residual behaviour. Furthermore, it was difficult to interpret the results in terms of the number of missing events, and hence time-varying catalogue completeness, in our example. However, constructing the hidden states in this manner could be valuable in other applications.

By using the estimated probabilities of being in state $j = 1, \ldots, 4$ given the observation sequence $X_{1:T}$, $\hat{\gamma}_t(j)$, (the supplementary file, Section S.2) along with the median number of missed events (N_j, say) in each state we can estimate the proportion of completeness over time as

$$\hat{p}(t) = \hat{\gamma}_1(t) + \sum_{j=2}^{m} \hat{\gamma}_t(j) \big/ N_j, \tag{23}$$

which is plotted in Fig. 11. In AD 0, the record shows completeness for a very short period. Most of the time between AD 50 and AD 500, the estimated proportion of completeness of the catalogue fluctuates between 11% and 35% which indicates that

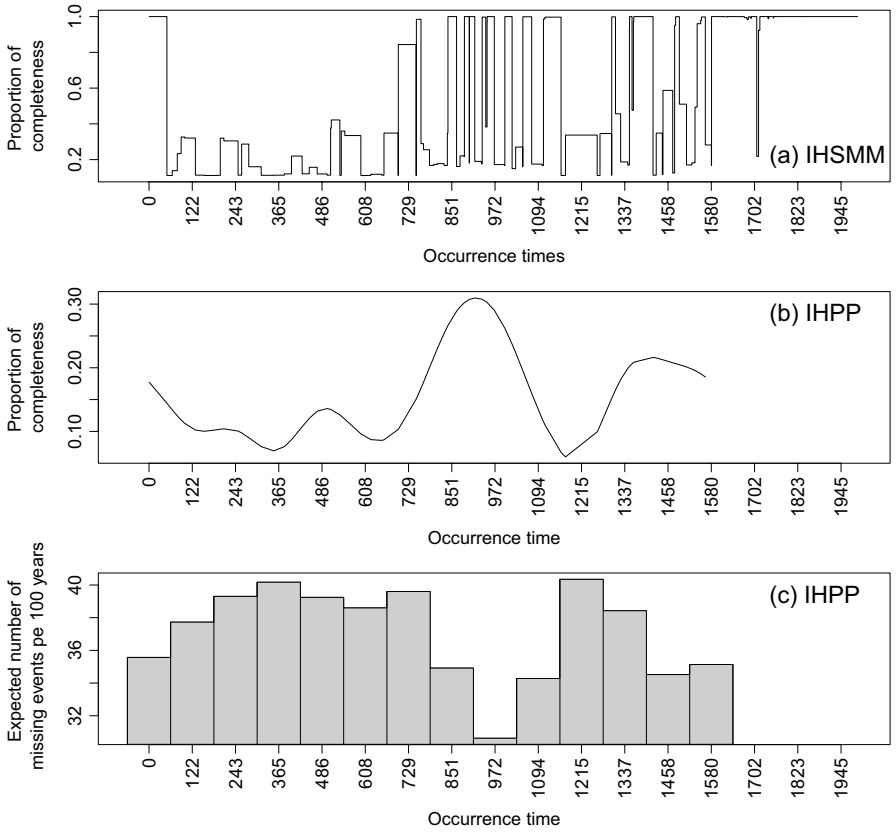


Fig. 11 **a** Estimated proportion of completeness versus occurrence times from the 4-state IHSMM in Case III, **b** The proportion of completeness per 61 years from IHPP before AD 1580, and **c** The histogram of expected number of missing events per 100 years from IHPP before AD 1580

this part of the catalogue is substantially incomplete. The estimated high completeness from AD 700 to AD 1200 in Fig. 11 appears to reflect an artefact of the Japanese record, which dominates the global record (Kiyosugi et al., 2015). Relative to the preceding 6 centuries *and* the following 6 centuries, this period of the Japanese catalogue has an anomalously high number of recorded eruptions (Hayakawa, 1997). Furlan (2010) investigated this global volcanic eruption record using a change point model to tackle the temporal behaviour of the recording bias and concluded that the under-recording of eruption events mostly disappears in the most recent 400 years.

Our results coincide with Furlan's finding as Figs. 8 and 11 show that the catalogue is complete after AD 1580. Our model assumes a correlated missingness structure. In order to compare with one in which every event can be missed randomly with time-dependent missing probabilities, we will explore fitting an

inhomogeneous Poisson process (IHPP). We fit a homogeneous Poisson process (HPP) with intensity λ to the most recent part. The MLE of the intensity of the fitted HPP to the most recent part of catalogue is $\hat{\lambda} = 0.44$. On the older part of the catalogue before AD 1580, we fit an IHPP with time-dependent intensity $\lambda(t)$ defined by

$$\lambda(t) = \sum_{p=1}^{K} c_p B_p^d(t), \tag{24}$$

where $B_p(t)$ is the *pth* cubic $B-$spline of degree d at time t, and c_p is its coefficient (Morgan et al., 2019), using R package NHPPspline (Morgan, 2021) for different numbers of knots, K. Using AIC, the IHPP with spline-based intensity is best-fitted with 20 knots for the older part of the catalogue. The estimated time-dependent intensity of missing events is then $\lambda_{incomp}(t) = \lambda - \lambda_{comp}(t)$, where $\lambda_{comp}(t)$ is the estimated time-dependent intensity for the older part of the catalogue. The expected numbers of missing events in the older part of the record are then $\int_{t_{k-1}}^{t_k} \lambda_{incomp}(s)ds$ plotted in Fig. 11 along with the proportion of completeness $\lambda_{comp}(t)/\lambda$ against time before AD 1580.

We can see that the IHSMM gives more detailed information about where the data is missing whereas the proportion of completeness estimated from IHPP is quite smooth (Fig. 11(a, b)). The estimated number of missing events per 100 years agrees with our model results, but the IHPP incorporates a large amount of smoothing even with 20 knots so we cannot see in detail which periods have more missing events. Since many causes of incompleteness in volcanic records can be traced to discrete events (e.g. European settlement, change in dynasties, wars), this is a potential impediment to interpretation.

In our proposed IHSMMs, the memoryless assumption is relaxed during the stay in each state before the transition to the next state. Nevertheless, the memoryless assumption can be trialled for many nonstationary processes. In this situation, inhomogeneous hidden Markov models (IHMMs) without explicitly specifying the state duration distributions can be used to model such processes. When the transition probabilities are time-varying, an HMM is called an IHMM. We also fitted IHMMs with time-varying transition probabilities defined by multinomial logistic functions to the given catalogue and found the performance of the 4-state IHMM to be roughly equivalent to that of the 4-state IHSMM, both models having equal number of parameters. However, the IHSMM detects that the building blocks of the pre-700 structure are distinct from those post-700, this corresponds to the anecdotal evidence of the Japanese catalogue.

HSMMs assume that the expected sojourn times for each state remain constant over time. This means that the expected sojourn times for the states representing some missing events is never going to decrease with time, which is not realistic for the global eruption record. Consequently, HSMMs may not be appropriate for the observed process of having nonstationary behaviour.

Note that we used a generalized logistic function with four parameters to formulate the time-dependent expected sojourn times, as this function has the potential

to meet the assumption that the mean sojourn times in each state may increase or decrease at an increasing or decreasing rate over time or some interval of time, particularly when there are more than two states. For example, if there are 3 states, it is quite possible for both States 1 and 2 to grow (at least for a while) as State 3 shrinks. Many other forms of logistic or exponential functions can be used to model the time-dependent sojourn times in different fields of applications. We considered the parameter of the shifted Poisson distribution to be a function of time at which the previous event occurred because we dealt with an embedded renewal process. One can simply use the jump time in the given state for time-dependency of sojourn times in different fields of applications.

Many special cases of IHSMMs can be obtained by introducing different parameterizations of time-dependent state durations for different sojourn time distributions with support on positive integers. In the proposed IHSMMs, the transition probabilities were assumed to be constant over time. We can further extend the HSMM and IHSMM using time-varying transition probabilities. This can be achieved by defining the transition probabilities as the functions of time using logistic function (Diebold et al., 1999; Filardo, 1994). However, introducing the time-varying transition probabilities in IHSMMs could make the models over-parameterized and complex from a computational point of view. Furthermore, the residual IHSMM can be easily defined as an extension of the residual HSMM proposed by Yu and Kobayashi (2003) that would be a useful model for processes for which the residual time is important.

10 Conclusions

In this paper, we proposed a general class of IHSMMs to model partially observed processes that show time-inhomogeneity, which provides a novel approach to model the time-varying structure of observed processes through time-dependent state durations of the underlying semi-Markov chain. Many historical records of point processes, e.g. volcanic eruptions, earthquakes and tsunami, are usually incomplete with time inhomogeneous missingness. We introduced a special case of IHSMM to model the time-inhomogeneity in the completeness of the historical records of point processes. The model was applied to a global volcanic eruption catalogue. The observed process was defined to follow a gamma renewal process, and the inhomogeneous number of missing events in between each pair of consecutively observed events in the catalogue was a hidden process represented by an ISMC. The ISMC was defined to be composed of a semi-Markov chain with state durations having shifted Poisson distributions which have time-dependent parameters. The AIC and residual analysis for point processes were used to select the best model. The 4-state IHSMM in Case III with state definitions in Table 3 was found to be the best fitted and satisfactory model for describing the time-dependent incompleteness of the catalogue. The IHSMMs can be used to model other types of inhomogeneous processes with or without missing data.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10463-022-00843-5.

Acknowledgements Amina Shahzadi is supported by a University of Otago PhD Scholarship. This work is supported by the Royal Society of New Zealand Marsden Fund (contract UOO1419). Mark Bebbington and Ting Wang are supported by the resilience to Nature's Challenges Volcano Programme, Grant GNS-RNC047. We thank the New Zealand eScience Infrastructure (NeSI) for providing access to the Mahuika for calculations. We also thank Wolfgang Hayek, Peter Maxwell and Alexander Pletzer the programming specialists in NeSI, for their advice on making the programs faster. We have received constructive comments from two anonymous reviewers and the Associate Editor which greatly improved our manuscript.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705.

Barbu, V., Limnios, N. (2008). *Semi-Markov Chains and hidden Semi-Markov models toward applications*: *Their use in reliability and DNA analysis. New York: Springer-Verlag*. https://doi.org/10.1007/978-0-387-73173-5.

Baum, L. E., Petrie, T. (1966). Statistical inference for probabilistic functions of finite state Markov chains. *Annals of Mathematical Statistics, 37*(6), 1554–1563. https://doi.org/10.1214/aoms/1177699147.

Bebbington, M. S. (2007). Identifying volcanic regimes using hidden Markov models. *Geophysical Journal International, 171*, 921–942. https://doi.org/10.1111/j.1365-246X.2007.03559.x.

Bebbington, M. S. (2013). Models for temporal volcanic hazard. *Statistics in Volcanology, 1*, 1–24. https://doi.org/10.5038/2163-338X.1.1.

Bebbington, M. S., Lai, C. D. (1996). On nonhomogeneous models for volcanic eruptions. *Mathematical Geology, 28*, 585–600. https://doi.org/10.1007/BF02066102.

Berman, M. (1983). Comment on "likelihood analysis of point processes and its applications to seismological data'' by Ogata. *Bulletin International Statistical Institute, 50*, 412–418.

Beyreuther, M., Wassermann, J. (2008). Continuous earthquake detection and classification using discrete hidden Markov models. *Geophysical Journal International, 175*(3), 1055–1066. https://doi.org/10.1111/j.1365-246X.2008.03921.x.

Brown, S. K., Crosweller, H. S., Stephen, R., Sparks, J., Cottrell, E., Deligne, N. I., Guerrero, N. O., Hobbs, L., Kiyosugi, K., Loughlin, S. C., Siebert, L., Takarada, S. (2014). Characterisation of the quaternary eruption record: Analysis of the Large Magnitude Explosive Volcanic Eruptions (LaMEVE) database. *Journal of Applied Volcanology, 3*(1), 5. https://doi.org/10.1186/2191-5040-3-5.

Bulla, J. (2006) *Application of hidden Markov models and hidden semi-Markov models to financial time series*. PhD-Thesis, Georg-August-Universitat Gottingen, Germany. https://mpra.ub.uni-muenchen.de/id/eprint/7675.

Coles, S. G., Sparks, R. S. J. (2006). Extreme value methods for modelling historical series of large volcanic magnitudes. In H. M. Mader, S. G. Coles, C. B. Connor, L. J. Connor, (Special Publications of IAVCEI, No. 1). *Statistics in Volcanology, Geological Society London*, pp 47–56. https://doi.org/10.1144/IAVCEI001.5.

Daley, D. J., Vere-Jones, D. (2003). *Introduction to the Theory of Point Processes*. New York: Springer. https://doi.org/10.1007/b97277.

De la Cruz-Reyna, S. (1991). Poisson-distributed patterns of explosive eruptive activity. *Bulletin of Volcanology, 54*(1), 57–67. https://doi.org/10.1007/BF00278206.

De la Cruz-Reyna, S. (1993). Random patterns of occurrence of explosive eruptions at Colima volcano, Mexico. *Journal of Volcanology and Geothermal Research, 55*(1), 51–68. https://doi.org/10.1016/0377-0273(93)90089-A.

De la Cruz-Reyna, S., Carrasco-Núñez, G. (2002). Probabilistic hazard analysis of Citlaltepetl (Pico de Orizaba) volcano, eastern Mexican volcanic belt. *Journal of Volcanology and Geothermal Research, 113*(1), 307–318. https://doi.org/10.1016/S0377-0273(01)00263-3.

Deligne, N. I., Coles, S. G., Sparks, R. S. J. (2010). Recurrence rates of large explosive volcanic eruptions. *Journal of Geophysical Research, 115*(B06), 203. https://doi.org/10.1029/2009JB006554.

Diebold, F. X., Lee, J. H., Weinbach, G. C. (1999). Regime Switching with time-varying transition probabilities. In F. X. Diebold and G. D. Rudebusch. *Business Cycles*: *Durations, Dynamics and Forecasting*, pp 144–165. Princeton Univeristy Press.

Durbin, R., Eddy, S. R., Krogh, A., Mitchison, G. (1998). Biological sequence analysis: Probabilistic models of proteins and nucleic acids. Cambridge University Press. https://doi.org/10.1017/CBO9780511790492.

Ferguson, J. D. (1980). Variable duration models for speech. *Proceedings*: *Symposium on the Application of Hidden Markov Models to Text and Speech*, pp 143–179. New Jersey: Princeton.

Filardo, A. J. (1994). Business-cycle phases and their transitional dynamics. *Journal of Business and Economics Statistics, 12*(3), 299–308. https://doi.org/10.2307/1392086.

Furlan, C. (2010). Extreme value methods for modelling historical series of large volcanic magnitudes. *Statistical Modelling, 10*(2), 113–132. https://doi.org/10.1177/1471082X0801000201.

Guédon, Y. (2003). Estimating hidden semi-Markov chains from discrete sequences. *Journal of Computational and Graphical Statistics, 12*(3), 604–639. https://doi.org/10.1198/1061860032030.

Guédon, Y., Cocozza-Thivent, C. (1990). Explicit state occupancy modelling by hidden semi-Markov models: Application of Derin's scheme. *Computer Speech and Language, 4*(2), 167–192. https://doi.org/10.1016/0885-2308(90)90003-O.

Guttorp, P., Thompson, M. L. (1991). Estimating second-order parameters of volcanicity from historical data. *Journal of the American Statistical Association, 86*(415), 578–583. https://doi.org/10.1080/01621459.1991.10475082.

Hayakawa, Y. (1997). Hayakawa's 2000-year eruption catalog. http://www.hayakawayukio.jp/catalog/2000W.

Ibáñez, J. M., Benítez, C., Gutiérrez, L. A., Cortés, G., García-Yeguas, A., Alguacil, G. (2009). The classification of seismo-volcanic signals using hidden Markov models as applied to the Stromboli and Etna volcanoes. *Journal of Volcanology and Geothermal Research, 187*(3), 218–226. https://doi.org/10.1016/j.jvolgeores.2009.09.002.

Kiyosugi, K., Connor, C., Sparks, R. S. J., Crosweller, H. S., Brown, S. K., Siebert, L., Wang, T., Takarada, S. (2015). How many explosive eruptions are missing from the geologic record? analysis of the quaternary record of large magnitude explosive eruptions in Japan. *Journal of Applied Volcanology, 4*(1), 17. https://doi.org/10.1186/s13617-015-0035-9.

Levinson, S. E. (1986). Continuously variable duration hidden Markov models for automatic speech recognition. *Computer Speech and Language, 1*(1), 29–45. https://doi.org/10.1016/S0885-2308(86)80009-2.

Limnios, N., Oprişan, G. (2001). Semi-Markov Processes and Reliability. *Statistics for Industry and Technology,* Birkhäuser Basel. https://doi.org/10.1007/978-1-4612-0161-8.

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

Mitchell, C. D., Jamieson, L. H. (1993). Modeling duration in a hidden Markov model with the exponential family. *Proceedings*: *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp 331–334. https://doi.org/10.1109/ICASSP.1993.319304.

Morgan, L. E. (2021). *NHPPspline*: *An R package*. https://github.com/morganle/NHPPspline.

Morgan, L. E., Nelson, B. L., Titman, A. C., Worthington, D. J. (2019). A Spline-based method for modelling and generating a nonhomogeneous Poisson process. *Winter Simulation Conference* (*WSC*), pp 356–367. https://doi.org/10.1109/WSC40007.2019.9004867.

Nash, J. C. (2014). On best practice optimization methods in R. *Journal of Statistical Software, 60*(2), 1–14. https://doi.org/10.18637/jss.v060.i02.

Nash, J. C., Varadhan, R. (2011). Unifying optimization algorithms to aid software system users: Optimx for R. *Journal of Statistical Software, 43*(9), 1–14. https://doi.org/10.18637/jss.v043.i09.

Ogata, Y. (1988). Statistical models for earthquake occurrence and residual analysis for point processes. *Journal of the American Statistical Association, 83*, 9–27. https://doi.org/10.2307/2288914.

Pertsinidou, C. E., Limnios, N. (2015). Viterbi algorithms for hidden semi-Markov models with application to DNA analysis. *RAIRO Operations Research, 49*, 511–526. https://doi.org/10.1051/ro/2014053.

R Core Team (2017) R: The R project for Statistical Computing. R Foundation for Statistical Computing, Auckland, New Zealand. http://www.R-project.org/.

Richards, F. J. (1959). A flexible growth function for empirical use. *Journal of Experimental Botany, 10*(2), 290–301. https://doi.org/10.1093/jxb/10.2.290.

Rossi, L., Chakaerski, J. S. (2015). A Poisson hidden Markov model for multiview video traffic. *IEEE/ACM Transactions on Networking, 23*(2), 547–558. https://doi.org/10.1109/TNET.2014.2303162.

Rougier, J., Sparks, S. R., Cashman, K. V. (2016). Global recording rates for large eruptions. *Journal of Applied Volcanology, 5*(1), 11. https://doi.org/10.1186/s13617-016-0051-4.

Russell. M., Moore, R. K. (1985). Explicit model of state duration occupancy in hidden Markov models for automatic speech recognition. *Proceedings*: *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp 5–8. https://doi.org/10.1109/ICASSP.1985.1168477.

Sansom, J., Thompson, C. S. (2003). Mesoscale spatial variation of rainfall through a hidden semi-Markov model of breakpoint data. *Journal of Geophysical Research*, 108(D8). https://doi.org/10.1029/2001JD001447.

Sansom, J., Thomson, P. (2001). Fitting hidden semi-Markov models to breakpoint rainfall data. *Journal of Applied Probability, 38*, 142–157. https://doi.org/10.1239/jap/1085496598.

Siebert, L., Simkin, T., Kimberly, P. (2010). *Volcanoes of the World*. Washington, D.C.: Smithsonian Institution; Berkeley; University of California Press. https://volcano.si.edu/learn_resources.cfm?p=4.

Simkin, T. (1993). Terrestrial volcanism in space and time. *Annual Review of Earth and Planetary Sciences, 21*(1), 427–452. https://doi.org/10.1146/annurev.ea.21.050193.002235.

Trevezas, S., Limnios, N. (2009). Maximum likelihood estimation for general hidden semi-Markov processes with backward recurrence time dependence. *Journal of Mathematical Sciences, 163*(3), 262–274. https://doi.org/10.1007/s10958-009-9675-9.

Turner, M. B., Cronin, S. J., Bebbington, M. S., Platz, T. (2008). Developing a probabilistic eruption forecast for dormant volcanoes: A case study from Mt Taranki, New Zealand. *Bulletin of Volcanology, 70*, 507–515. https://doi.org/10.1007/s00445-007-0151-4.

Turner, M. B., Bebbington, M. S., Cronin, S. J., Stewart, R. B. (2009). Merging eruption datasets: Building an integrated Holocene eruptive record of Mt Taranaki, New Zealand. *Bulletin of Volcanology, 71*, 903–918. https://doi.org/10.1007/s00445-009-0274-x.

Varley, N., Johnson, J., Ruiz, M., Reyes, G., Martin, K. (2006). Applying statistical analysis to understanding the dynamics of volcanic explosions. In H. M. Mader, S. G. Coles, C. B. Connor, L. J. Connor, (Special Publications of IAVCEI, No. 1). *Statistics in Volcanology, Geological Society London*, pp 57–76. https://doi.org/10.1144/IAVCEI001.6.

Viterbi, A. J. (1967). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory, 13*, 260–269. https://doi.org/10.1109/tit.1967.1054010.

Votsi, I., Limnios, N., Papadimitriou, E., Tsaklidis, G. (2018). *Earthquake statistical analysis through multi-state modeling. Mathematics and statistics series*. London: Wiley-ISTE.

Wang, T. (2010). *Statistical models for earthquakes incorporating ancillary data*. PhD Thesis, Massey University, Palmerston North.

Wang, T., Bebbington, M. (2013). Identifying anomalous signals in GPS data using HMMs: An increased likelihood of earthquakes. *Computational Statistics and Data Analysis, 58*, 27–44. https://doi.org/10.1016/j.csda.2011.09.019.

Wang, T., Bebbington, M. S. (2012). Estimating the likelihood of an eruption from a volcano with missing onsets in its records. *Journal of Volcanology and Geothermal Research, 243*, 14–23. https://doi.org/10.1016/j.jvolgeores.2012.06.032.

Wang, T., Bebbington, M. S., Harte, D. (2012). Markov-modulated Hawkes process with stepwise decay. *Annals of the Institue of Statistical Mathematics, 64*, 521–544. https://doi.org/10.1007/s10463-010-0320-7.

Wang, T., Zhuang, J., Obara, K., Tsuruoka, H. (2017). Hidden Markov modelling of sparse time series from non-volcanic tremor observations. *Journal of Royal Statistical Society Series C Applied Statistics, 66*(4), 691–715. https://doi.org/10.1111/rssc.12194.

Wickman, F. E. (1966). Repose-period patterns of volcanoes: Part I. volcanic eruptions regarded as random phenomena. *Arkiv for Mineralogi och Geologi, 4*, 291–367.

Yu, S. Z. (2015). *Hidden semi-Markov Model Theory*. Algorithms and Applications: Elsevier Science.

Yu, S. Z., Kobayashi, H. (2003). A hidden semi-Markov model with missing data and multiple observation sequences for mobility tracking. *Signal Processing, 83*(2), 235–250. https://doi.org/10.1016/S0165-1684(02)00378-X.