




Model Checking for Hidden Markov Models

Jodie Buckby, Ting Wang, Jiancang Zhuang & Kazushige Obara


To cite this article: Jodie Buckby, Ting Wang, Jiancang Zhuang & Kazushige Obara (2020) Model Checking for Hidden Markov Models, Journal of Computational and Graphical Statistics, 29:4, 859-874, DOI: [10.1080/10618600.2020.1743295](https://doi.org/10.1080/10618600.2020.1743295)

To link to this article: <https://doi.org/10.1080/10618600.2020.1743295>

 View supplementary material 

 Published online: 14 May 2020.

 Submit your article to this journal 

 Article views: 178

 View related articles 

 View Crossmark data 



Model Checking for Hidden Markov Models

Jodie Buckby^a, Ting Wang^a, Jiancang Zhuang^{b,c}, and Kazushige Obara^d

^aDepartment of Mathematics and Statistics, University of Otago, Dunedin, New Zealand; ^bInstitute of Statistical Mathematics, Tokyo, Japan; ^cDepartment of Statistical Science, Graduate University for Advanced Studies, Tokyo, Japan; ^dEarthquake Research Institute, University of Tokyo, Tokyo, Japan

ABSTRACT

Residual analysis is a useful tool for checking lack of fit and for providing insight into model improvement. However, literature on residual analysis and the goodness of fit for hidden Markov models (HMMs) is limited. As HMMs with complex structures are increasingly used to accommodate different types of data, there is a need for further tools to check the validity of models applied to real world data. We review model checking methods for HMMs and develop new methods motivated by a particular case study involving a two-dimensional HMM developed for time series with many null events. We propose new residual analysis and stochastic reconstruction methods, which are adapted from model checking techniques for point process models. We apply the new methods to the case study model and discuss their adequacy. We find that there is not one “best” test for diagnostics but that our new methods have some advantages over previously developed tools. The importance of multiple tests for complex HMMs is highlighted and we use the results of our model checking to provide suggestions for possible improvements to the case study model. Supplementary materials for this article are available online.

ARTICLE HISTORY

Received May 2018
Revised November 2019

KEYWORDS

Goodness of fit; Residual analysis; Stochastic reconstruction; Viterbi path; Zero-inflation

1. Introduction

An important part of any statistical study involving model development is checking the fit of the model, to assess how well the model captures the features of the data, and to provide insight into potential model improvements. However, research on model checking for HMMs is under developed and there is a need for new methods. For example, current methods of model checking, such as the pseudo-residuals described by Zucchini and MacDonald (2009) and the cumulative distribution function (CDF) plots proposed by Altman (2004), are sometimes insufficient for diagnosing specific problems with model fit.

In this article, we use the two-dimensional HMM with extra zeros developed by Wang et al. (2018) as an example to demonstrate new model checking techniques. The model in Wang et al. (2018) was developed to classify nonvolcanic tremor in the Kii and Shikoku regions of Japan into spatiotemporal segments with distinct features. Wang et al. (2018) undertook a range of model checking methods to test the validity of individual assumptions of the model. We look to develop new tools for checking the fit of HMMs that can help identify possible improvements to an HMM fitted to a specific dataset by adapting methods from other classes of models. Specifically, we investigate adapting model checking theory for point process models developed by Baddeley et al. (2005) and Zhuang (2006). Wang, Wang, and Zhuang (2018) adapted their methods for point processes and proposed similar methods for a 2-part autoregressive model. We use this work as a starting point for our new methods.

The remaining sections of this article are organized as follows. Section 2 gives an overview of HMMs in general and details the case study model, a two-dimensional HMM with extra zeros developed by Wang et al. (2018). Section 3 gives the details of methods currently used for checking the fit of HMMs. Section 4 describes some model checking methods for point processes and time series in general and shows how this methodology can be adapted for use with HMMs, including new residual analysis methods and stochastic reconstruction. In Section 5, we apply the methods to the case study model and present the results. Our findings are summarized in Section 6 and the adequacy of the methods is discussed.

2. Preliminaries

2.1. Hidden Markov Models

Hidden Markov models (HMMs) were first introduced in the late 1960s (Baum and Petrie 1966). The models were brought to further prominence by their application to speech recognition such as in Rabiner (1989). Since then, this class of models has been used in many fields of study including bioinformatics (Krogh, Mian, and Haussler 1994), finance (Hassan and Nath 2005), and seismology (Wang, Bebbington, and Harte 2012; Wang and Bebbington 2013).

An HMM is used to model time series data when the observations are dependent on an underlying unobserved Markov chain. Observations are realized from the m members of a family of distributions, $f(x_t|S_t = i)$, with $i = 1, \dots, m$, where x_t is the observed value at time $t = 1, \dots, T$, $S_t = i$ is the state of the

Markov chain at time t and m is the finite number of states. The simplest first-order HMM satisfies the conditions,

$$P(S_t = s_t | S_1 = s_1, \dots, S_{t-1} = s_{t-1}) = P(S_t = s_t | S_{t-1} = s_{t-1})$$

and

$$f(x_t | x_1, \dots, x_{t-1}, S_1 = s_1, \dots, S_t = s_t) = f(x_t | S_t = s_t).$$

The parameters to be estimated for an HMM include the parameters associated with the emission (state dependent) distribution, in addition to δ and Γ . The size m vector δ with elements δ_i contains the initial probabilities that the first observation is from each of the states. The matrix Γ with elements γ_{ij} is an $m \times m$ matrix of transition probabilities representing the probability of moving from one state at time $t - 1$ to another at time t . Thus, the number of parameters to be estimated in the model is dependent on the number of states, m . Constraints applicable to δ and Γ are such that $\delta_i \geq 0$, $\gamma_{ij} \geq 0$, $\sum_i \delta_i = 1$, and $\sum_j \gamma_{ij} = 1$. Estimation of the parameters is achieved by directly maximizing the likelihood or by using the expectation-maximization (EM) algorithm (Baum et al. 1970). After parameter estimation, the Viterbi algorithm, as proposed by Viterbi (1967), can be used to obtain the most likely sequence of hidden states over time given the data and model parameters. This algorithm maximizes the conditional probability,

$$P(S_1 = s_1, \dots, S_T = s_T | x_1, \dots, x_T) = \frac{f(x_1, \dots, x_T, S_1 = s_1, \dots, S_T = s_T)}{\sum_{s_1, \dots, s_T} P(S_1 = s_1, \dots, S_T = s_T) f(x_1, \dots, x_T | S_1 = s_1, \dots, S_T = s_T)}.$$

The obtained state sequence, the Viterbi path, classifies the observations into m states.

2.2. Motivating Example: Data and Model Description

Our study is motivated by the data and model introduced in Wang et al. (2018) for spatiotemporal classification of nonvolcanic tremor activity in the Kii region of Japan. Nonvolcanic tremor is a chain of low frequency seismic activity originating in a subduction zone. Tremor was first detected in the Nankai subduction zone of Japan (Obara 2002). Modeling nonvolcanic tremor is of interest because tremor activity is associated with slow slip events (Obara 2011). Data including the presence of tremor and the location of tremor was collated at hourly intervals using the high-sensitivity seismograph network as referred to by Maeda and Obara (2009). The nonvolcanic tremor data for the Kii region of the Nankai Subduction zone consists of 6348 hr when tremor is detected in a total of 105,192 hourly observations from January 2001 to December 2012. Observations of detected tremor are clustered in time and space interspersed with long periods of quiescence. When tremor is detected, latitude and longitude observations are given. If no tremor is detected in a given hour, a zero is present in the data.

Previously, classification of tremor data and study of migration patterns were a manual process. The HMMs developed by Wang et al. (2017, 2018) provide automation of this process with quantitative results including distinct one-dimensional or two-dimensional spatial segments with differing probabilities of tremor occurrence and transition probabilities between these segments.

The structure of the model in Wang et al. (2018) is such that given tremor occurs at time t , the two-dimensional location observations, \mathbf{Y}_t , for each state are assumed to follow a bivariate normal distribution. The HMM accounts for the many null events with a binary variable. Null events are indicated by $Z_t = 0$ and $Z_t = 1$ otherwise, with $p_i = P(Z_t = 1 | S_t = i)$ being the probability of tremor detection in state i . The joint emission distribution for each state is given below.

$$f(y_t, z_t | S_t = i) = (1 - p_i)^{1 - z_t} \left(p_i \frac{1}{2\pi |\Sigma_i|^{1/2}} \exp \left(-\frac{1}{2} (y_t - \mu_i)^T \Sigma_i^{-1} (y_t - \mu_i) \right) \right)^{z_t},$$

where $\mu_i = \mathbf{E}[\mathbf{Y}_t | Z_t = 1, S_t = i]$, $\Sigma_i = \text{cov}[\mathbf{Y}_t | Z_t = 1, S_t = i]$. Hidden states, S_t , are assumed to form a first-order stationary Markov chain, that is, $\tau \Gamma = \tau$, where τ is the stationary probability distribution.

The model was fitted using the EM algorithm for different numbers of states m . The Bayesian information criterion (BIC), (Schwarz 1978) was used to determine the number of states and a 17 state model was selected for the Kii region (Figure 1). Three different types of tremor segments are suggested by the model: episodic, weak concentration, and background. The model indicates 15 spatially distinct episodic and weak concentration segments and two background states, with the system spending most time in the background state 16, in which the proportion of tremor occurrence is about 0.002. The remaining background state (state 17) features weak sporadic tremor over the center and east of the region.

Although the 17 state model for the Kii region is considered the best from those fitted, we are interested in (1) whether this model is actually a good fit for the data, (2) whether current model checking techniques are sufficient, and (3) how to evaluate whether the assumptions of each component of the model are met. This information will provide insight into how the model can be improved and/or extended in future research.

3. Current Model Checking Methods for HMMs

The pseudo-residuals defined by Zucchini and MacDonald (2009) are often used to check the fit of HMMs and are seen as an alternative to the usual residual analysis used in regression models. In brief, the ordinary pseudo-residuals are defined as the exvisible CDF, that is, $F(x_t | x_1, \dots, x_{t-1}, x_{t+1}, \dots, x_T)$. As such, they take account of the model structure and the other data to identify extreme values, and a model with good fit should see a standard uniform distribution of these pseudo-residuals, or a standard normal distribution when transformed. Zucchini and MacDonald (2009) gave a full account of the theory behind these pseudo-residuals. However, it can be more difficult to identify good fit with pseudo-residuals when there are many discrete observations at the boundary of the sample space (Harte 2017), or when the model has a Bernoulli variable as part of the emission distribution. The ordinary pseudo-residuals for discrete observations must be defined as either intervals $[u_t^L, u_t^U]$ where the superscripts L and U represent the lower and upper limits for the interval, or as mid-points of these intervals. In our case study model, it is more difficult to identify good fit using pseudo-residuals because of the nature of the emission distribution, a

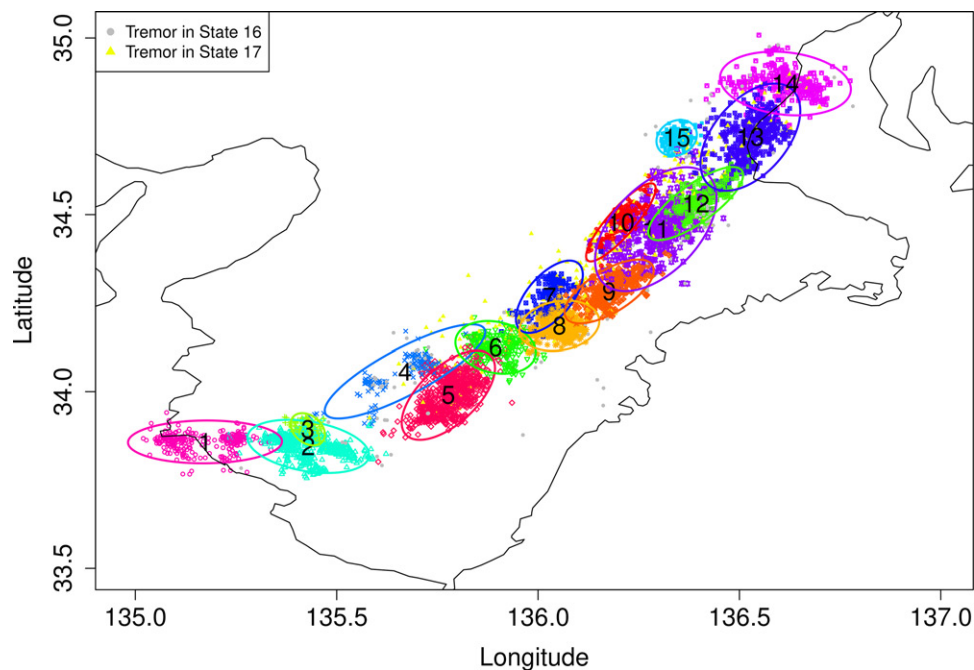


Figure 1. The fitted two-dimensional 17 state model for nonvolcanic tremor in the Kii region of south-west Japan. Observations are classified into different spatiotemporal states represented by different colors. Each state has a unique probability of tremor detection.

mixture of a discrete Bernoulli variable accounting for many null events and a continuous bivariate normal distribution.

Altman (2004) proposed an alternative method of model checking for HMMs and demonstrated that the difference between the empirical CDF and the estimated CDF of a stationary HMM should be small everywhere if the fit is good. This method can easily be extended to HMMs with more complex state-dependent emission distributions as long as the CDF is tractable, but only allows assessment of the fit of the marginal distribution of the observations, giving no information about the dependence structure. To further investigate the correlation structure of the fitted model, Altman (2004) described plotting higher dimensional CDFs of consecutive observations. Usually, the bivariate distribution is sufficient. These plots, like the univariate case, indicate good fit if they converge to the 45° line as the sample size increases. We note that one pitfall of using tests based on the CDF is that a lack of fit can remain undetected when a higher than expected empirical CDF value is followed by an interval with lower than expected values.

Beyond pseudo-residuals and CDF plots, there are few detailed methods for testing the fit of HMMs in general, although some authors have developed methods for their specific models. Titman and Sharples (2007) discussed misclassification models in health services. They compared the observed and expected number of state transitions as in the Aguirre-Hernandez and Farewell (AH/F) test (Aguirre-Hernández and Farewell 2002), which they adapted to accommodate cases where the misclassification model includes an absorbing state. Ailliot, Thompson, and Thomson (2009) assessed their proposed HMMs for precipitation by comparing simulated data from the HMMs with rainfall patterns using Q-Q plots. The capability of the model to replicate correlation between different locations was also assessed by comparing the estimated correlation to the correlation in the data. Wang et al. (2017) introduced

an HMM with Bernoulli-normal mixture emission distributions and took the approach suggested by Yang and Simpson (2010), who studied the diagnostics of zero-inflated models by testing different elements of a mixture distribution separately. However, in the context of an HMM, many of the methods are conditional on the estimated Viterbi path which adds further uncertainty.

In the following section, we look more closely at methods that are conditional on the Viterbi path relevant to our motivating example.

3.1. Methods Based on the Viterbi Path

Some simple model checking methods assume that the estimated Viterbi path is the true sequence of states, and then test the assumptions of the emission distributions. Wang et al. (2017) used this approach to test each element of the mixture emission distribution of their HMM after they obtained the Viterbi path from the data. This HMM is the one-dimensional version of the model in our motivating example and was used to study tremor migration in another region. The emission distribution in each state is a mixture of Bernoulli and normal distributions. A significant discrepancy from the assumption of a Bernoulli distribution with constant probability of event occurrence in each state, or from the assumption of a conditional normal distribution, might suggest that some important features of the observed processes are not accounted for in the model and perhaps further improvement of the model is required.

Wang et al. (2017, 2018) used an empirical method for checking the stationarity of the Bernoulli distribution. They split the entire sequence into equal time intervals and calculated the empirical proportion of tremor occurrence in each state (according to the Viterbi path) for each time period along with confidence intervals. This empirical proportion was then

compared with the estimated proportion in each state from the HMM to identify lack of fit.

Additionally, if the emission distribution given that the observation is not a structural zero is assumed to follow a normal or multivariate normal distribution, this fit can be assessed using presence residuals, defined as the standardized residuals of the observations excluding structural zeros, given the Viterbi path (Wang et al. 2017). For a multivariate normal, the standardized presence residual vector, r_t , for time t when an event occurs, is of the form,

$$r_t = |\Sigma_{\hat{s}_t}|^{-1/2}(y_t - \mu_{\hat{s}_t}), \tag{1}$$

where y_t is the multivariate observation given that an event occurs, $\mu_{\hat{s}_t}$ is the expectation of y_t given the estimated state, \hat{s}_t , from the Viterbi path and event occurrence, and $\Sigma_{\hat{s}_t}$ is the corresponding covariance matrix. We expect these residuals to have a multivariate normal distribution. We will illustrate the Viterbi path based methods using the case study model in Section 5.

4. New HMM Model Checking Methods

Various residual analysis techniques have been developed for point processes, including (1) a transformation based method (Ogata 1988), which transforms an observed point process into a standard Poisson process, (2) innovation based methods (Baddeley et al. 2005; Zhuang 2006, 2015), which compare an innovation process created from the observation to its conditional expectation, (3) thinning/complementation based methods (Schoenberg 2003; Clements, Schoenberg, and Veen 2012), which generate a Poisson process by removing some data from the observed point process or adding extra points to the observed point process according to some probability rules determined by the model, and (4) Voronoi residuals (Bray and Schoenberg 2013; Bray et al. 2014), where each Voronoi cell contains one observed event and also one expected event, given the model, if the fit is ideal. Among these residuals, the innovation based methods can be naturally extended to HMMs. Before demonstrating this, we briefly introduce the innovation based residual analysis for point processes.

Point processes model the intensity of events, often in the context of temporal or spatiotemporal data. Most point process models are specified using one of the following forms: moment intensity, conditional intensity or Papangelou intensity. For a one-dimensional (temporal) point process N , the moment intensity $\nu(t)$ is defined by,

$$\nu(t) dt = \mathbf{E} [N ([t, t + dt))],$$

where $N([t, t + dt))$ is the number of events occurring in $[t, t + dt)$. For an HMM, the pdf of the distribution for observation x_t at any time t , without knowing any information is,

$$\sum_{i=1}^m \tau_i f(x_t | S_t = i)$$

with τ_i being the i th element of the stationary probability distribution of the Markov chain. It is natural to see that this is an analogue of the moment intensity of a point process.

The conditional intensity of the temporal point process is informed by the history of the process and defined as,

$$\lambda(t) dt = \mathbf{E} [N ([t, t + dt)) | \mathcal{H}_t],$$

where \mathcal{H}_t represents knowledge of N up to time t but not including t . Finally, the Papangelou intensity is calculated as,

$$\lambda_p(t) dt = \mathbf{E} [N ([t - dt/2, t + dt/2]) | \mathcal{I}_{[t-dt/2, t+dt/2]}],$$

where $\mathcal{I}_{[t-dt/2, t+dt/2]}$ represents information of N everywhere except the time interval $[t - dt/2, t + dt/2]$. The studies of Baddeley et al. (2005) and Zhuang (2006, 2015) developed residual analysis for point process models that take account of our knowledge of the process by using the conditional intensity or the Papangelou intensity. In this study, we introduce this kind of residual analysis into the literature of HMMs.

For a temporal point process N with conditional intensity $\lambda(t)$, the following property holds: for any regular set B (countable union of intervals) and a nonnegative predictable function $h(t)$,

$$\mathbf{E} \left[\sum_{t_i \in N \cap B} h(t_i) \right] = \mathbf{E} \left[\int_B h(t) \lambda(t) dt \right]$$

since

$$\sum_{t_i \in N \cap [0, t]} h(t_i) - \int_{[0, t]} h(t) \lambda(t) dt$$

is a zero-mean martingale. Here, loosely speaking, the predictable function $h(t)$ acts as a weight and is a stochastic function determined only by previous data of the point process before time t . Thus, for any estimates $\hat{\lambda}(t)$ of $\lambda(t)$, the predictive residual is defined by

$$R(B, \hat{h}, \hat{\lambda}) = \sum_{t_i \in N \cap B} \hat{h}(t_i) - \int_B \hat{h}(t) \hat{\lambda}(t) dt,$$

where $\hat{\cdot}$ is also put on $h(t)$ for the case that h also includes estimated parameters. Similarly, by using the Georgii–Zessin–Nguyen formula, as referred to in Baddeley et al. (2005) and Zhuang (2015), the residual with respect to the Papangelou intensity, namely the exvisive residual, is defined by

$$R(B, \hat{h}, \hat{\lambda}_p) = \sum_{t_i \in N \cap B} \hat{h}(t_i) - \int_B \hat{h}(t) \hat{\lambda}_p(t) dt,$$

where $h(t)$ is an exvisible function, that is, a function determined by the occurrence pattern of N throughout time, except at time t (Baddeley et al. 2005). If the fitted model is a good approximation of the true model, then $R(B, \hat{h}, \hat{\lambda}) \approx 0$ or $R(B, \hat{h}, \hat{\lambda}_p) \approx 0$. The forms of $h(t)$ can be chosen according to the type of residual required. For raw residuals, $h(t) = 1$.

Adapting concepts from Zhuang (2006) for point processes, Wang, Wang, and Zhuang (2018) developed residuals for zero-inflated autoregressive time series models. For a time series X_t , these are calculated as,

$$R(n) = \sum_{t=1}^n \left(\hat{h}(t) X_t - \hat{h}(t) \mathbf{E} [X_t | \mathcal{H}_t] \right),$$

where \mathcal{H}_t represents the history X_1, \dots, X_{t-1} . In this case, we have discrete time t and,

$$\mathbf{E} \left[\sum_{t=1}^n \hat{h}(t) X_t \right] = \mathbf{E} \left[\sum_{t=1}^n \hat{h}(t) \mathbf{E} [X_t | \mathcal{H}_t] \right].$$

The process,

$$D(n) = \sum_{t=1}^n h(t) X_t - \sum_{t=1}^n h(t) \mathbf{E} [X_t | \mathcal{H}_t]$$

is a zero-mean martingale because,

$$\begin{aligned} & \mathbf{E} [D(n) - D(n-1) | \mathcal{H}_n] \\ &= \mathbf{E} [h(n) X_n - h(n) \mathbf{E} [X_n | \mathcal{H}_n] | \mathcal{H}_n] = 0. \end{aligned}$$

A good fit implies that $R(n) \approx D(n)$, that is, close to 0. Moreover, the asymptotic normality of $D(n)$ is asserted by the central limit theorem for martingales (Hall and Heyde 2014). The goodness of fit can thus be assessed by comparing the distribution of $R(n)$ to the asymptotic distribution of $D(n)$ when n is sufficiently large.

In the following subsection, we extend the predictive and exvisive residuals to HMMs.

4.1. Predictive Residuals

Here, we adapt the residuals developed by Baddeley et al. (2005) and Zhuang (2006, 2015) for use with HMMs. For the HMM, time t is discrete and we must consider all possible states that the system could be in at time t .

First, we consider the predictive residuals. These correspond to the conditional intensity of a point process model, that is, the expectation of the occurrence of an event in unit time given the history. In an HMM, the expectation of X_t , given the previous observations is,

$$\mathbf{E} [X_t | \mathcal{H}_t] = \mathbf{E} [X_t | X_1, \dots, X_{t-1}] = \sum_{i=1}^m \frac{\alpha_{t-1} \Gamma_i \mathbf{E} [X_t | S_t = i]}{\alpha_{t-1} \mathbf{1}'},$$

where α_{t-1} is the forward probability with the j th element,

$$\alpha_{t-1}(j) = P(X_1, \dots, X_{t-1}, S_{t-1} = j).$$

Details for calculating forward and backward probabilities within the EM algorithm can be found in Zucchini and MacDonald (2009).

Using the expectation calculated above, the raw predictive residual with $h(t) = 1$ becomes,

$$R_n^p = \sum_{t=1}^n (X_t - \mathbf{E} [X_t | \mathcal{H}_t]),$$

where the parameters are those estimated from the data. Where appropriate, other functions $h(t)$ might be introduced, as described in Section 4 for point process residual analysis. The standardized residuals for $n = 2, \dots, T$ are calculated as,

$$\bar{R}_n^p = \frac{R_n^p}{\sqrt{\sum_{t=1}^n (R_t^p - R_{t-1}^p)^2}}, \tag{2}$$

where $\sqrt{\sum_{t=1}^n (R_t^p - R_{t-1}^p)^2}$ is the standard deviation of R_n^p , as defined in Wang, Wang, and Zhuang (2018) and $R_0^p = 0$. Although \bar{R}_n^p for $n = 2, \dots, T$ are not iid, we expect to see \bar{R}_n^p fall within the 95% confidence interval of a standard normal distribution if the fitted model is close to the true model as $\bar{R}_n^p \sim N(0, 1)$ for each large enough n (Hall and Heyde 2014). The lower and upper bounds of the 95% confidence interval are calculated as $\Phi^{-1}(0.025)$ and $\Phi^{-1}(0.975)$, respectively, where Φ^{-1} is the inverse standard normal CDF.

Furthermore, we calculate standardized raw residuals $\bar{R}_{k,L}^p$ for fixed intervals of observations, where $k = 1, \dots, K$ and L is the interval size. As such,

$$R_{k,L}^p = \sum_{t=Lk+1}^{Lk+L} (X_t - \mathbf{E} [X_t | \mathcal{H}_t]) \tag{3}$$

and

$$\bar{R}_{k,L}^p = \frac{R_{k,L}^p}{\sqrt{\sum_{t=Lk+1}^{Lk+L} (R_t^p - R_{t-1}^p)^2}}. \tag{4}$$

Now, $\bar{R}_{k,L}^p$ can be considered iid for $k = 1, \dots, K$ and goodness of fit can be assessed by comparing the distribution of $\bar{R}_{k,L}^p$ to the standard normal distribution. We note that this technique requires a relatively long time series where K is a large enough sample size for comparison to a standard normal distribution and the interval L is sufficiently large to assume normality.

We demonstrate the use of raw predictive residuals using 10,000 simulated data points from a 7 state normal HMM. The data are simulated using R (R Core Team 2018) and the ‘‘HiddenMarkov’’ package (Harte 2017). The plot of \bar{R}_n^p against time in Figure 2 indicates reasonable fit, as expected with simulated data. We see that the variation of \bar{R}_n^p is larger for small n than for large n as the estimation of the variance in Equation (4) is not stable when we have few time points.

In addition, the histogram of $\bar{R}_{k,L}^p$ where $K = 99$ and $L = 100$ (Figure 3) shows that these interval residuals are close to a standard normal distribution. A Kolmogorov–Smirnov test provides no evidence to suggest that the interval residuals deviate significantly from a standard normal distribution, with $p = 0.43$.

4.2. Exvisive Residuals

We determine raw exvisive residuals in a similar way to the predictive residuals, the difference being that we consider the expectation of X_t given all of the other observations rather than just observations prior to time t . In this case, we have

$$\begin{aligned} \mathbf{E} [X_t | \mathcal{E}_t] &= \mathbf{E} [X_t | X_1, \dots, X_{t-1}, X_{t+1}, \dots, X_T] \\ &= \sum_{i=1}^m \frac{\alpha_{t-1} \Gamma_i \beta_t(i) \mathbf{E} [X_t | S_t = i]}{\sum_{j=1}^m \alpha_{t-1} \Gamma_j \beta_t(j)}, \end{aligned}$$

where $\beta_t(i) = P(X_{t+1}, \dots, X_T | S_t = i)$ is the backward probability. The raw exvisive residual becomes,

$$R_n^e = \sum_{t=1}^n (X_t - \mathbf{E} [X_t | \mathcal{E}_t]),$$

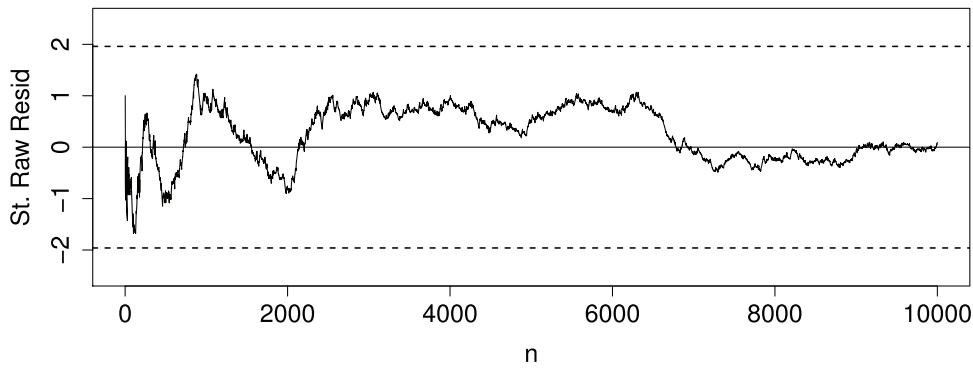


Figure 2. The standardized raw predictive residuals over time for simulated data from a 7 state normal HMM. Dashed lines indicate 95% confidence interval.

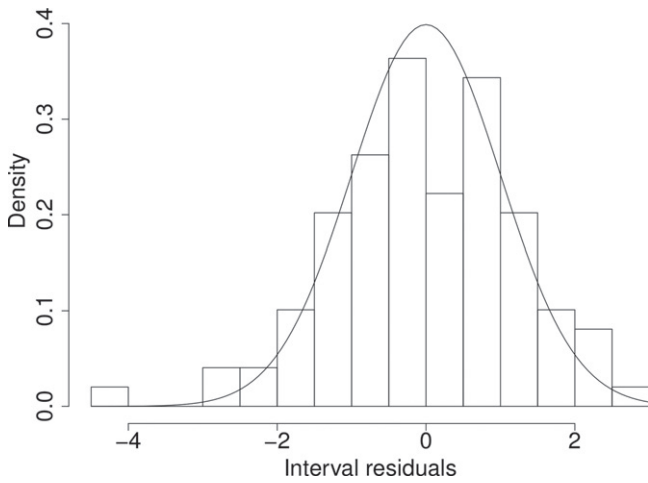


Figure 3. The standardized raw interval predictive residuals for simulated data from a 7 state normal HMM. Solid line indicates standard normal distribution.

where the parameters are those estimated from the data. To standardize these, we divide by the standard deviation of R_n^e . Although the central limit theorem for martingales is not applicable for exvisive residuals due to dependence of the increments $X_t - \mathbf{E}[X_t|\mathcal{E}_t]$, we can appeal to the central limit theorem more generally and simply include the dependence in our calculation of the standard deviation such that,

$$\bar{R}_n^e = \frac{R_n^e}{\sqrt{\sum_{t=1}^n (R_t^e - R_{t-1}^e)^2 + 2 \sum_{i \neq j} ((R_i^e - R_{i-1}^e)(R_j^e - R_{j-1}^e))}} \quad (5)$$

In practice, the dependence after a lag time of 2 or more (i.e., $X_t - \mathbf{E}[X_t|\mathcal{E}_t]$ and $X_{t-h} - \mathbf{E}[X_{t-h}|\mathcal{E}_{t-h}]$ with $h \geq 2$) is

usually negligible. We should note that the moment estimators of the covariance terms can be biased, potentially leading to some variance estimates for the denominator in Equation (5) < 0 .

Again, we expect \bar{R}_n^e for $n = 2, \dots, T$ to fall in the 95% confidence interval of a standard normal distribution but \bar{R}_n^e are not iid. We create iid exvisive interval residuals, $R_{k,L}^e$ using the same methods described for the predictive residuals.

$$R_{k,L}^e = \sum_{t=Lk+1}^{Lk+L} (X_t - \mathbf{E}[X_t|\mathcal{E}_t]) \quad (6)$$

and

$$R_{k,L}^e = \frac{R_{k,L}^e}{\sqrt{\sum_{t=Lk+1}^{Lk+L} (R_t^e - R_{t-1}^e)^2 + 2 \sum_{i \neq j} ((R_i^e - R_{i-1}^e)(R_j^e - R_{j-1}^e))}} \quad (7)$$

Now, goodness of fit is assessed by comparing the distribution of $R_{k,L}^e$ to the standard normal distribution.

In Figure 4, we plot the standardized exvisive residuals using the same simulated data and parameters from the 7 state normal HMM. The residuals fall within the standard normal confidence interval, indicating good fit.

The histogram of $R_{k,L}^e$ where $K = 99$ and $L = 100$ (Figure 5) shows that these interval residuals are reasonably close to a standard normal distribution, again indicating adequate fit. Again, the Kolmogorov–Smirnov test suggests that the interval residuals do not deviate significantly from a standard normal distribution with $p = 0.90$.

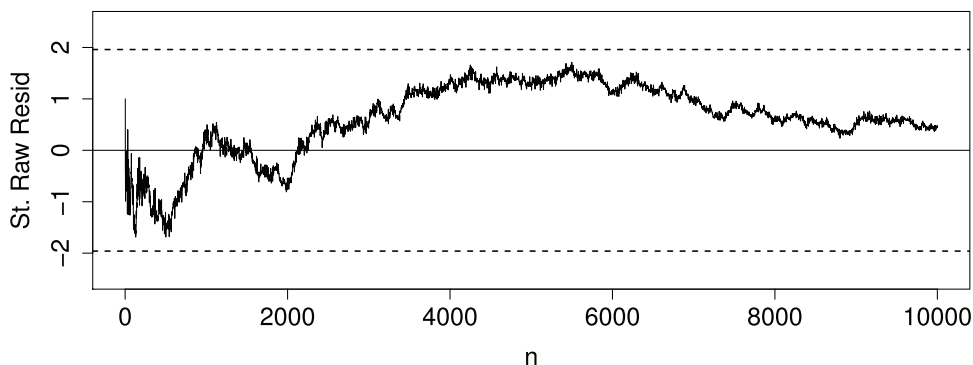


Figure 4. The standardized raw exvisive residuals over time for simulated data from a 7 state normal HMM. Dashed lines indicate 95% confidence interval.

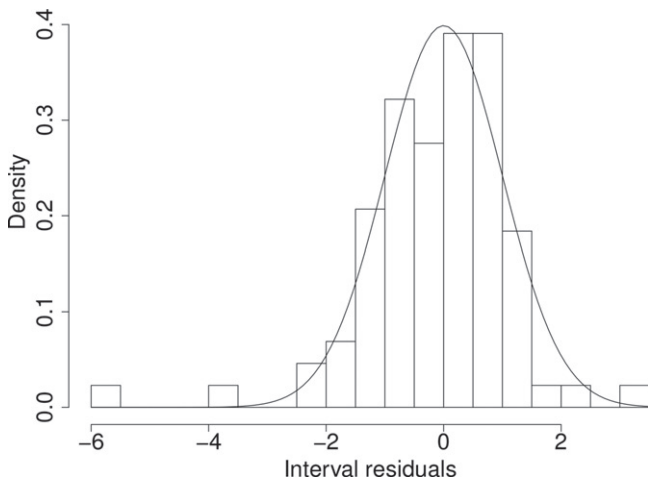


Figure 5. The standardized raw interval exvisive residuals for simulated data from a 7 state normal HMM. Solid line indicates standard normal distribution.

4.3. Stochastic Reconstruction

Here, we adapt the stochastic reconstruction of point processes (Zhuang, Ogata, and Vere-Jones 2004) for use with HMMs. The stochastic reconstruction method involves weighting the data by $\phi_{it} = P(S_t = i | X_1, \dots, X_T)$. The reconstruction for each state i

uses the formula,

$$f(x|S = i) \propto \sum_{t=1}^T \phi_{it} I(x_t \in (x - \Delta x, x + \Delta x)),$$

where Δx is a fixed arbitrary small increment in the value of x and $I(x_t \in (x - \Delta x, x + \Delta x))$ is an indicator function with value 1 if x_t lies within $(x - \Delta x, x + \Delta x)$. This weighted density for each state i can be plotted and the expected conditional density for each state calculated using the estimated parameters from the model can be overlaid for comparison. We demonstrate with the simulated data from the 7 state normal HMM used in Sections 4.1 and 4.2. The weighted density of the observations and the expected conditional density from the model are plotted in Figure 6 using the “weights” R package (Pasek 2018). As expected for simulated data we see good fit of the reconstruction to the expected densities.

In the case of a zero-inflated HMM where a Bernoulli variable models the presence of a structural zero, stochastic reconstruction can also be used as an alternative method to estimate the parameters p_i for $i = 1, \dots, m$, associated with the probability of a structural zero. We reconstruct p_i using,

$$\hat{p}_i = \frac{\sum_{t=1}^T (z_t \phi_{it})}{\sum_{t=1}^T \phi_{it}}.$$

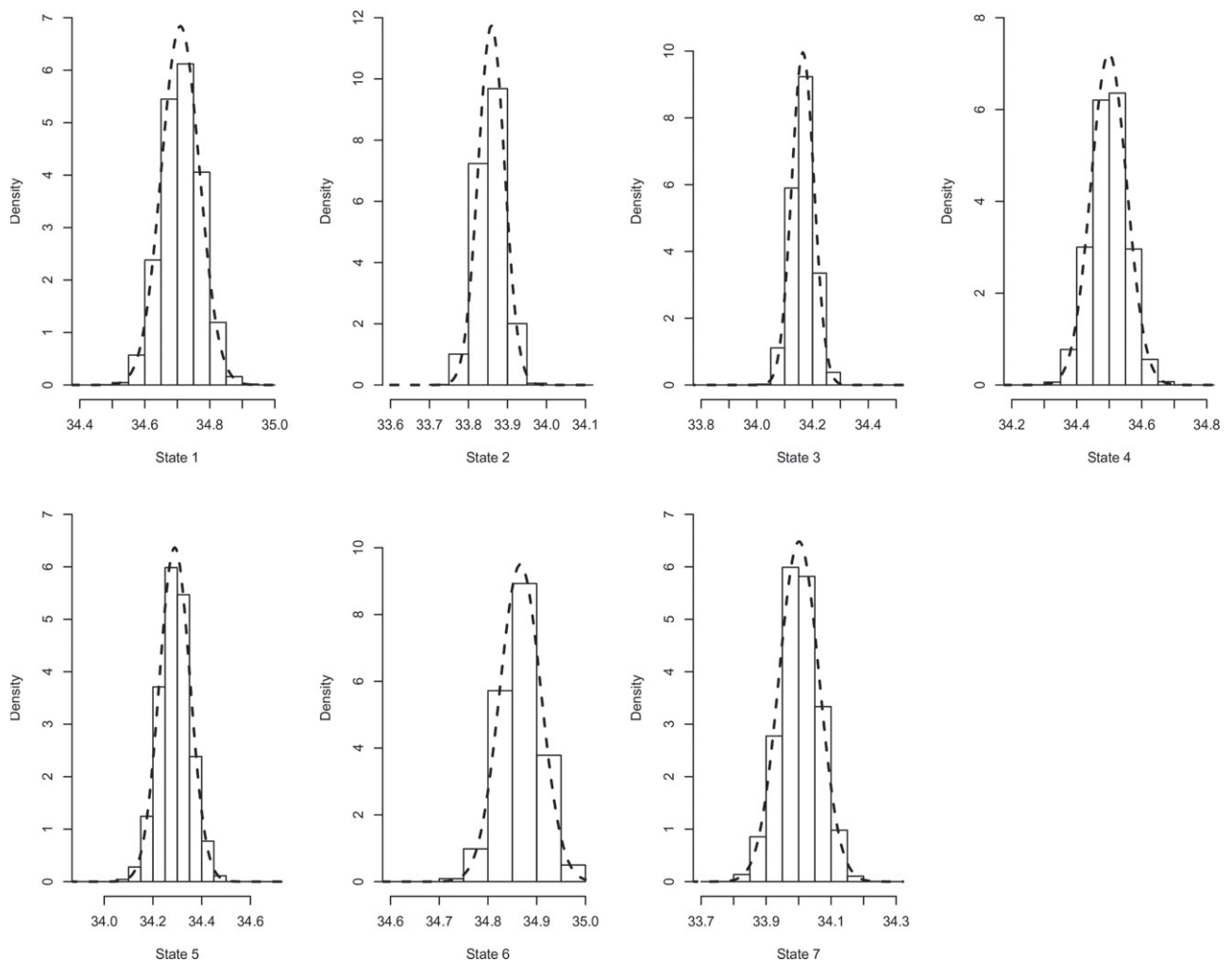


Figure 6. Weighted stochastic reconstruction for each state in a 7 state normal HMM. Dashed lines indicate expected density.

This formula can then be adapted to give estimates of p_i for different time periods. We use this to assess the variation of p_i over time. It has the added advantage, compared to the previous methods of checking the stationarity of the Bernoulli distribution discussed in Section 3.1, of not being dependent upon the estimated Viterbi path. We divide the entire time period into C intervals of equal length r . For each interval $c = 1, \dots, C$, we calculate the moving average of \hat{p}_i for each state $i = 1, \dots, m$ as follows. For the first time interval, when $c = 1$, the estimate \hat{p}_{i1} is,

$$\hat{p}_{i1} = \frac{\sum_{t=1}^r z_t \phi_{it}}{\sum_{t=1}^r \phi_{it}} \quad (8)$$

and for each subsequent value of \hat{p}_{ic} for $c = 2, \dots, C$ using a time interval shifted forward a period q ,

$$\hat{p}_{ic} = \frac{\sum_{t=q(c-1)+1}^{q(c-1)+r} z_t \phi_{it}}{\sum_{t=q(c-1)+1}^{q(c-1)+r} \phi_{it}}, \quad (9)$$

where $q < r$ and q is sufficiently large for each interval to include all states.

5. Application to the Two-Dimensional HMM With Extra Zeros

5.1. Viterbi Path Based Methods

5.1.1. Bernoulli Distribution Diagnostics

In our motivating example, evidence of non-stationarity and underestimation of some parameters p_i was found by Wang et al. (2018). We investigate this underestimation further, looking at each state. Figure 7 shows systematic bias in our model with the expected number of tremors, assuming that the Viterbi path is the true sequence of states, lower than the observed number of tremors in each state. States 3 and 15 show particularly poor fit with both \hat{p}_3 and \hat{p}_{15} values far from the confidence interval for \bar{p}_3 and \bar{p}_{15} and vice versa, where \bar{p}_i is the observed proportion of tremors in each state i given the Viterbi path. The results for states 1–2, 4–5, and 9 also suggest inadequate fit of \hat{p}_i in the 17 state model. Summing over all the states in the model, the expected number of tremors given that the Viterbi path is true is 11.33% lower than the observed number of tremors.

The underestimation of p_i is likely due to the model assumption that $P(Z_t = 1 | S_t = i)$ is constant over time. The estimator used for p_i is unbiased when $P(Z_t = 1 | S_t = i)$ is indeed constant, but bias is introduced if this assumption does not hold. Evidence from Wang et al. (2018) suggests that $P(Z_t = 1 | S_t = i)$ is a nonstationary process and supports the concept of a model in which p_i is allowed to vary over time.

5.1.2. Distribution of Presence Residuals

Here, we calculate the standardized presence residuals for our case study model. The conditional distribution of the two-dimensional tremor location given that tremor is detected and the state is known, $f(y_t | S_t = i, Z_t = 1)$, is assumed to be bivariate normal. As such, we use Equation (1) to determine the presence residuals and these are expected to follow a bivariate normal distribution if the model is a good fit to the data. In

addition, if sample sizes are large enough we can examine the presence residuals for each state separately for diagnostic purposes. There are many methods for testing bivariate normality, but here we focus on graphical methods as a tool for diagnostics. We use contour plots from the R-package “MVN” authored by Korkmaz, Goksuluk, and Zararsiz (2016). These plots show the empirical distributions of the residuals in each state, and are helpful in identifying unexpected features in the distributions that will then inform potential improvements for the model.

The standardized presence residuals for the case study model produce insightful results with clear evidence of bi-modality in the distributions of presence residuals for a number of states, particularly states 1, 2, 4, and 5. The contour plots for these states are shown in Figure 8 with plots for all individual states in a supplementary file. Additionally, we see clusters of outliers in many states, for example, in state 4. Some clustering of outliers is expected due to the migration of tremors through a segment, which is not accounted for in the model, and some outliers may occur due to misclassification in the Viterbi path. However, clear bi-modality indicates that there may be more than one process occurring within a state. In state 5, for example, one process involves migration of tremor between states 4 and 5 with the other due to migration to and from state 6.

5.2. Predictive Residuals

Presence residuals are calculated for times when a tremor occurs, which is a mere 6% of the data in the case study example, and are dependent on the estimated Viterbi path. In contrast, predictive residuals are calculated for all of the data and consider all possible state occupations. We calculate the marginal univariate standardized raw predictive residuals for the case study model using the methods described in Section 4.1. Here, $X_t^{(1)}$ is a mixed variable of zeros for null events and present latitudinal observations, $Y_t^{(1)}$, and $X_t^{(2)}$ is a mixed variable of zeros and present longitudinal observations, $Y_t^{(2)}$. The residuals are plotted against time in Figure 9. The expectations of $X_t^{(1)}$ and $X_t^{(2)}$ given the previous observations are,

$$\mathbf{E} \left[X_t^{(1)} | \mathcal{H}_t \right] = \sum_{i=1}^m \frac{\alpha_{t-1} \Gamma_{,i} p_i \mu_i^{(1)}}{\alpha_{t-1} \mathbf{1}'}$$

and

$$\mathbf{E} \left[X_t^{(2)} | \mathcal{H}_t \right] = \sum_{i=1}^m \frac{\alpha_{t-1} \Gamma_{,i} p_i \mu_i^{(2)}}{\alpha_{t-1} \mathbf{1}'},$$

where the j th element of α_{t-1} is,

$$\alpha_{t-1}(j) = P(Y_1, \dots, Y_{t-1}, Z_1, \dots, Z_{t-1}, S_{t-1} = j).$$

and $\mu_i^{(1)} = \mathbf{E} \left[Y_t^{(1)} | S_t = i, Z_t = 1 \right]$ and $\mu_i^{(2)} = \mathbf{E} \left[Y_t^{(2)} | S_t = i, Z_t = 1 \right]$. The raw predictive residuals are,

$$R_n^{p(1)} = \sum_{t=1}^n \left(X_t^{(1)} - \mathbf{E} \left[X_t^{(1)} | \mathcal{H}_t \right] \right)$$

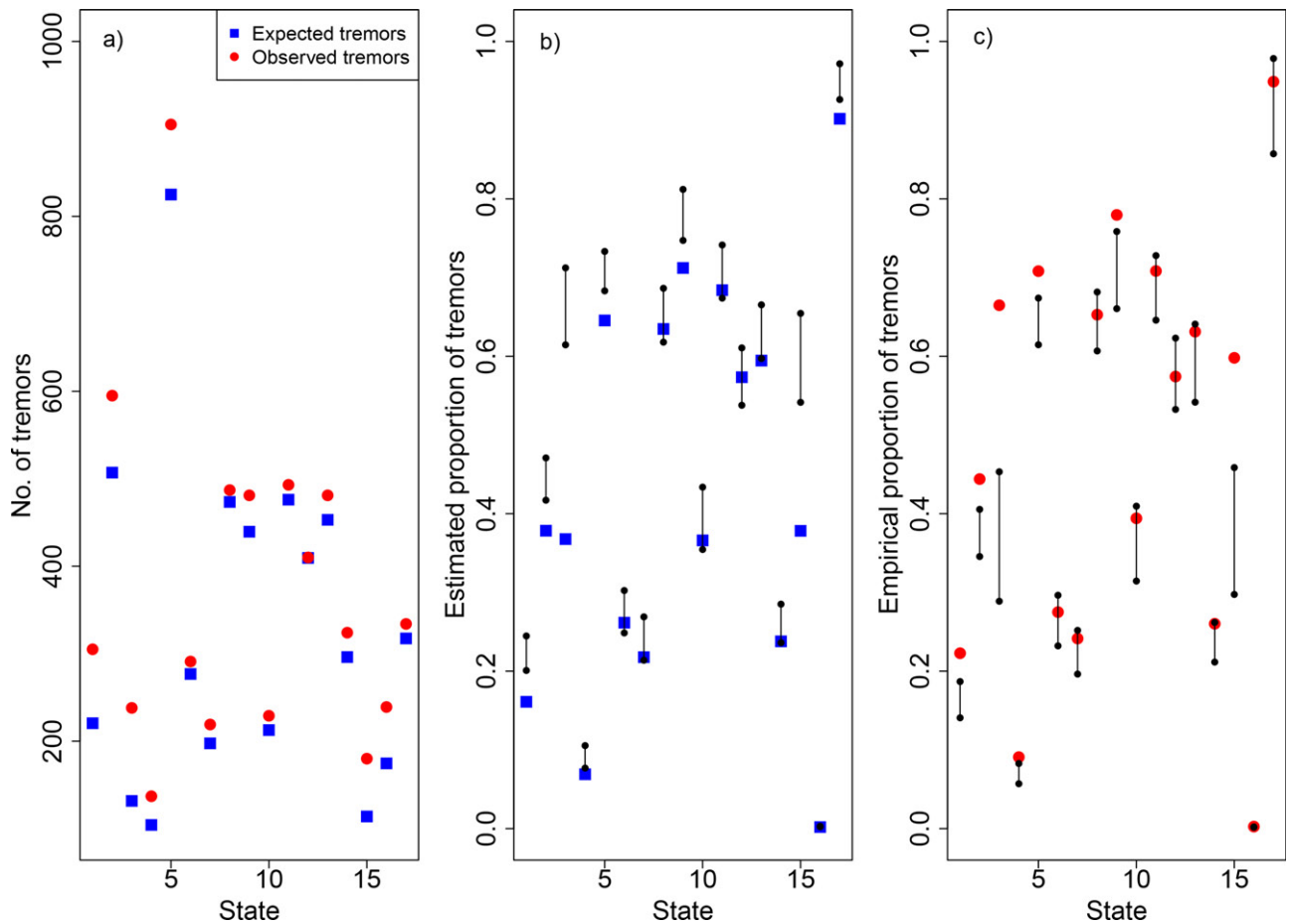


Figure 7. (a) The expected number of detected tremors for each state given the Viterbi path (squares), compared to the empirical number classified by the Viterbi path (circles). (b) Plot of \hat{p}_i with 95% confidence interval for empirical \bar{p}_i . (c) Plot of \hat{p}_i with the 95% bootstrap confidence interval for \hat{p}_i .

and

$$R_n^{p(2)} = \sum_{t=1}^n \left(X_t^{(2)} - \mathbf{E} \left[X_t^{(2)} | \mathcal{H}_t \right] \right),$$

where $\mathbf{E} \left[X_t^{(1)} | \mathcal{H}_t \right]$ and $\mathbf{E} \left[X_t^{(2)} | \mathcal{H}_t \right]$ are calculated at parameter estimates. The standardized predictive residual is as defined in Equation (2).

Figure 9 shows that the standardized residuals are quite stable after a brief run-in period and there is evidence of a lack of fit with remaining residuals significantly greater than zero. This is likely due, at least in part, to inaccuracy of estimates \hat{p}_i and future modeling should consider allowing p_i to vary over time to improve fit. The residuals for $X_t^{(1)}$ and $X_t^{(2)}$ have similar values due to the high proportion of zeros in the data.

In addition, we calculate, $\bar{R}_{k,L}^{p(1)}$ and $\bar{R}_{k,L}^{p(2)}$ as shown in Equations (3) and (4). Figure 10 shows histograms of $\bar{R}_{k,L}^{p(1)}$ and $\bar{R}_{k,L}^{p(2)}$ and we can see that there is greater variance than expected in these residuals, indicating a lack of fit. The Kolmogorov–Smirnov test indicates that the interval residuals for $X_t^{(1)}$ and $X_t^{(2)}$ deviate significantly from a standard normal distribution with $p = 0.0055$ and $p = 0.0067$, respectively.

5.3. Exvisive Residuals

An alternative to predictive residuals is the use of exvisive residuals. Here,

$$\mathbf{E} \left[X_t^{(1)} | \mathcal{E}_t \right] = \sum_{i=1}^m \frac{\alpha_{t-1} \Gamma_{,i} \beta_t(i) p_i \mu_i^{(1)}}{\sum_{j=1}^m \alpha_{t-1} \Gamma_{,j} \beta_t(j)}$$

and

$$\mathbf{E} \left[X_t^{(2)} | \mathcal{E}_t \right] = \sum_{i=1}^m \frac{\alpha_{t-1} \Gamma_{,i} \beta_t(i) p_i \mu_i^{(2)}}{\sum_{j=1}^m \alpha_{t-1} \Gamma_{,j} \beta_t(j)}$$

with $\beta_t(i) = P(Y_{t+1}, \dots, Y_T, Z_{t+1}, \dots, Z_T | S_t = i)$. The raw exvisive residuals are,

$$R_n^{e(1)} = \sum_{t=1}^n \left(X_t^{(1)} - \mathbf{E} \left[X_t^{(1)} | \mathcal{E}_t \right] \right)$$

and

$$R_n^{e(2)} = \sum_{t=1}^n \left(X_t^{(2)} - \mathbf{E} \left[X_t^{(2)} | \mathcal{E}_t \right] \right),$$

where $\mathbf{E} \left[X_t^{(1)} | \mathcal{E}_t \right]$ and $\mathbf{E} \left[X_t^{(2)} | \mathcal{E}_t \right]$ are calculated at parameter estimates. These are standardized using Equation (5). The marginal standardized exvisive residuals are plotted in Figure 11. We see some lack of fit with residuals larger than

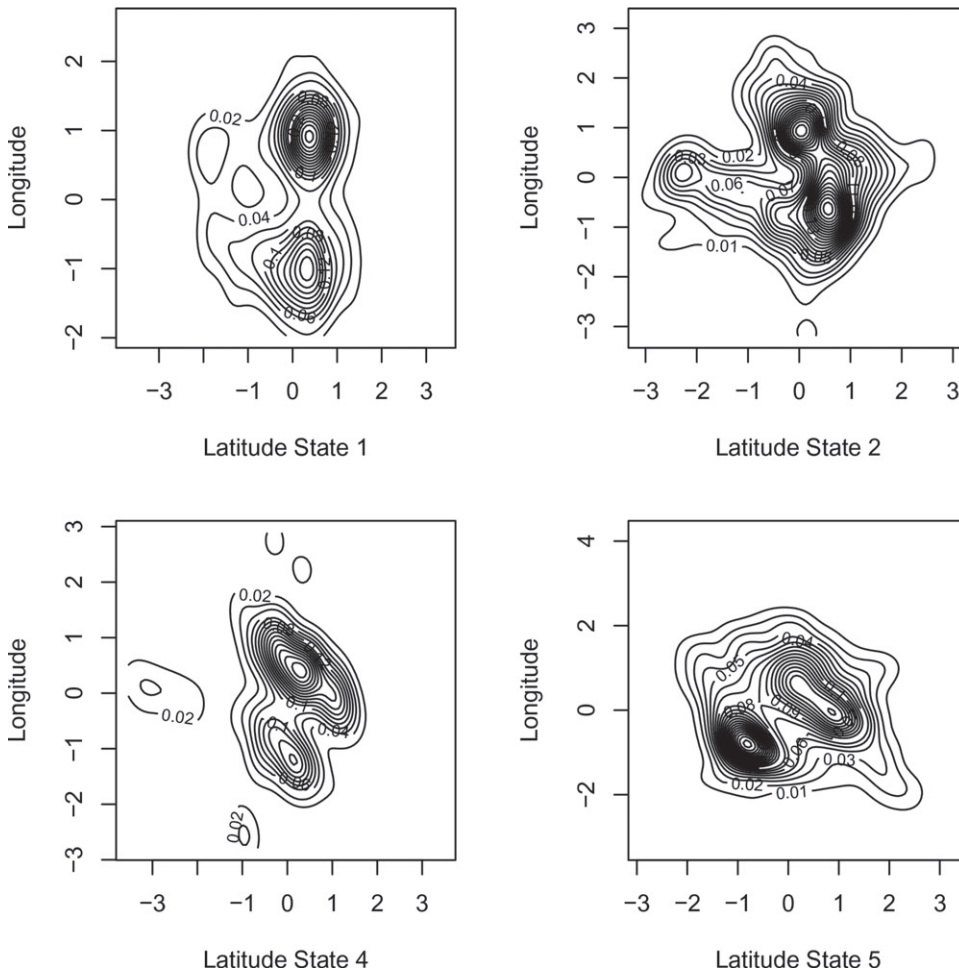


Figure 8. Contour plots of standardized presence residuals for states 1–2 and 4–5 of the case study model.

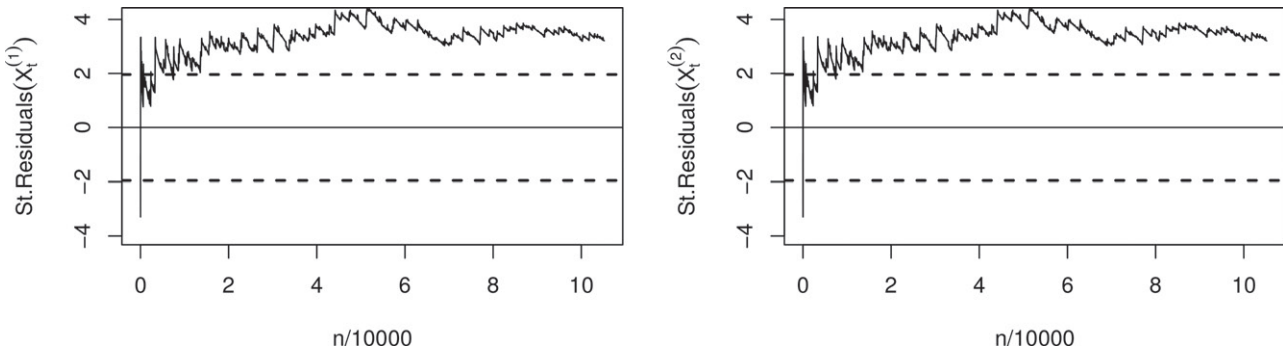


Figure 9. The marginal standardized predictive residuals plotted for $X_t^{(1)}$ (left) and $X_t^{(2)}$ (right) over time with 95% standard normal confidence interval (dashed lines).

expected. The results indicate that when we take account of all of the other data rather than the history alone the overall fit of the model is improved. This might be expected if our assumption of stationary p_i is flawed.

Again, we calculate, $\bar{R}_{k,L}^{e(1)}$ and $\bar{R}_{k,L}^{e(2)}$ as shown in Equations (6) and (7). Figure 12 shows histograms of $\bar{R}_{k,L}^{e(1)}$ and $\bar{R}_{k,L}^{e(2)}$. These also suggest a better fit than the predictive versions, with the Kolmogorov–Smirnov tests showing no significant lack of fit when comparing the distribution of interval residuals for $X_t^{(1)}$ and $X_t^{(2)}$ to a standard normal distribution ($p = 0.18$ and $p = 0.19$, respectively). However, Figure 13 still shows some evidence of greater variance than expected in these residuals.

Overall, the residuals suggest that when we take account of all of the other data in the time series the fit is adequate, but for predictive purposes the model requires improvement. Of course, we should note that our univariate residuals do not take account of the correlation between latitude and longitude observations.

5.4. Stochastic Reconstruction

As discussed in Section 4.3, the data are weighted by $\phi_{it} = P(S_t = i | \mathbf{Y}_1, \dots, \mathbf{Y}_T, Z_1, \dots, Z_T)$ where \mathbf{Y}_t is the 2D data matrix for tremor locations at time t and $Z_t = 1$ indicates tremor

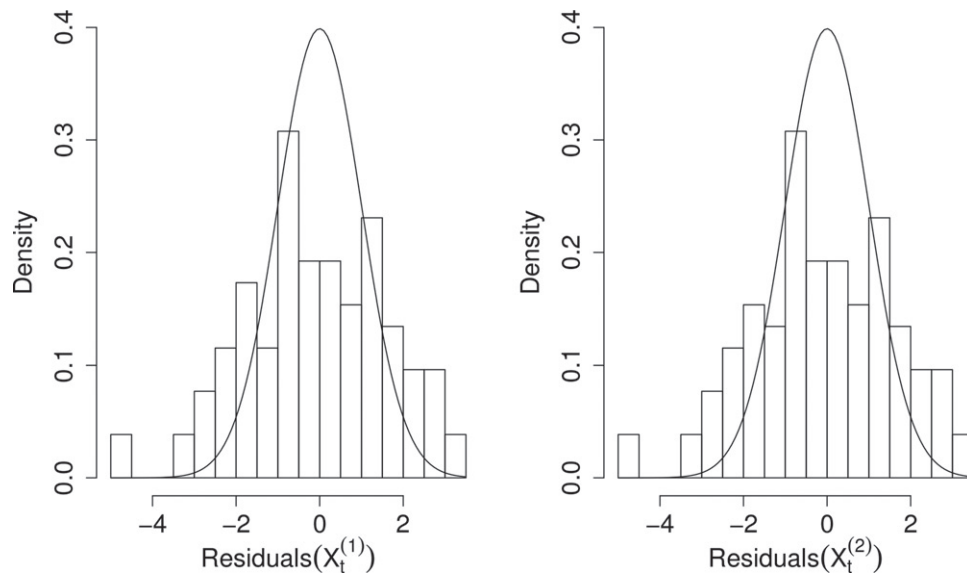


Figure 10. The marginal standardized interval predictive residuals plotted for $X_t^{(1)}$ (left) and $X_t^{(2)}$ (right). Solid line indicates standard normal distribution.

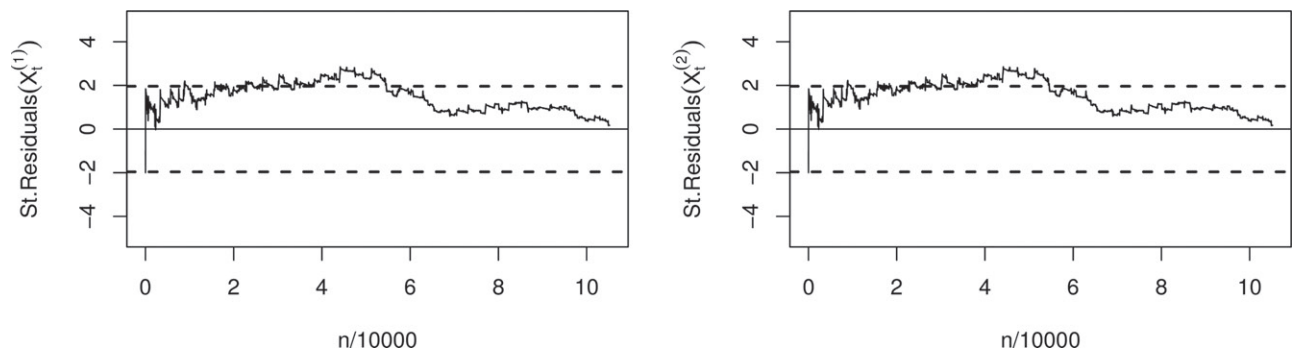


Figure 11. The marginal standardized exisive residuals plotted for $X_t^{(1)}$ (left) and $X_t^{(2)}$ (right) over time with 95% standard normal confidence interval (dashed lines).

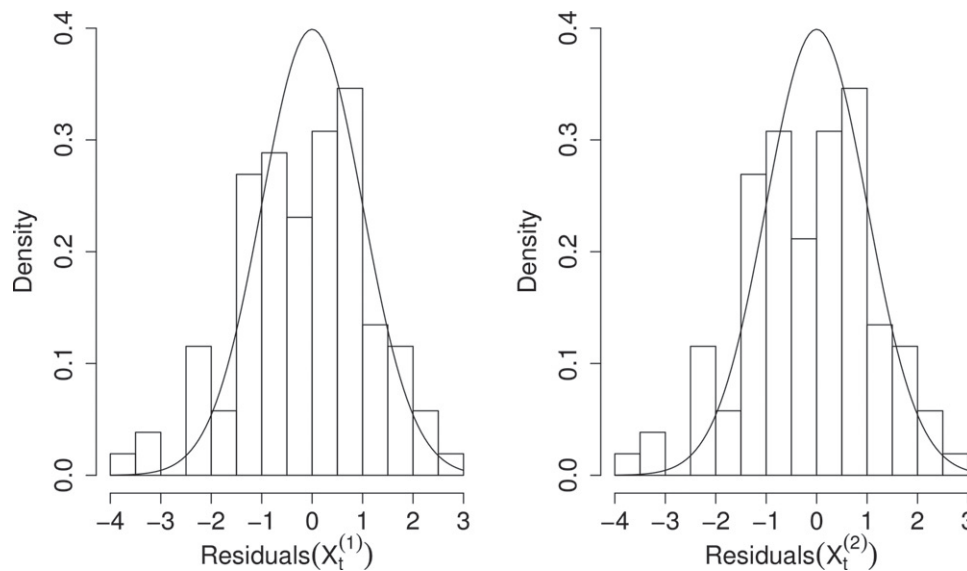


Figure 12. The marginal standardized interval exisive residuals plotted for $X_t^{(1)}$ (left) and $X_t^{(2)}$ (right). Solid line indicates standard normal distribution.

occurrence. The reconstruction uses the formula,

$$f(\mathbf{y}|S = i) \propto \sum_{t=1}^T \phi_{it} I(\mathbf{y}_t \in (\mathbf{y} - \Delta\mathbf{y}, \mathbf{y} + \Delta\mathbf{y})) I(z_t = 1), \quad (10)$$

where $\Delta\mathbf{y}$ is a fixed arbitrary small increment in the value of \mathbf{y} , $I(\mathbf{y}_t \in (\mathbf{y} - \Delta\mathbf{y}, \mathbf{y} + \Delta\mathbf{y}))$ is an indicator function with value 1 if \mathbf{y}_t lies within the rectangle $(\mathbf{y} - \Delta\mathbf{y}, \mathbf{y} + \Delta\mathbf{y})$, and $I(z_t = 1)$ is an indicator function with value 1 if a tremor occurred at

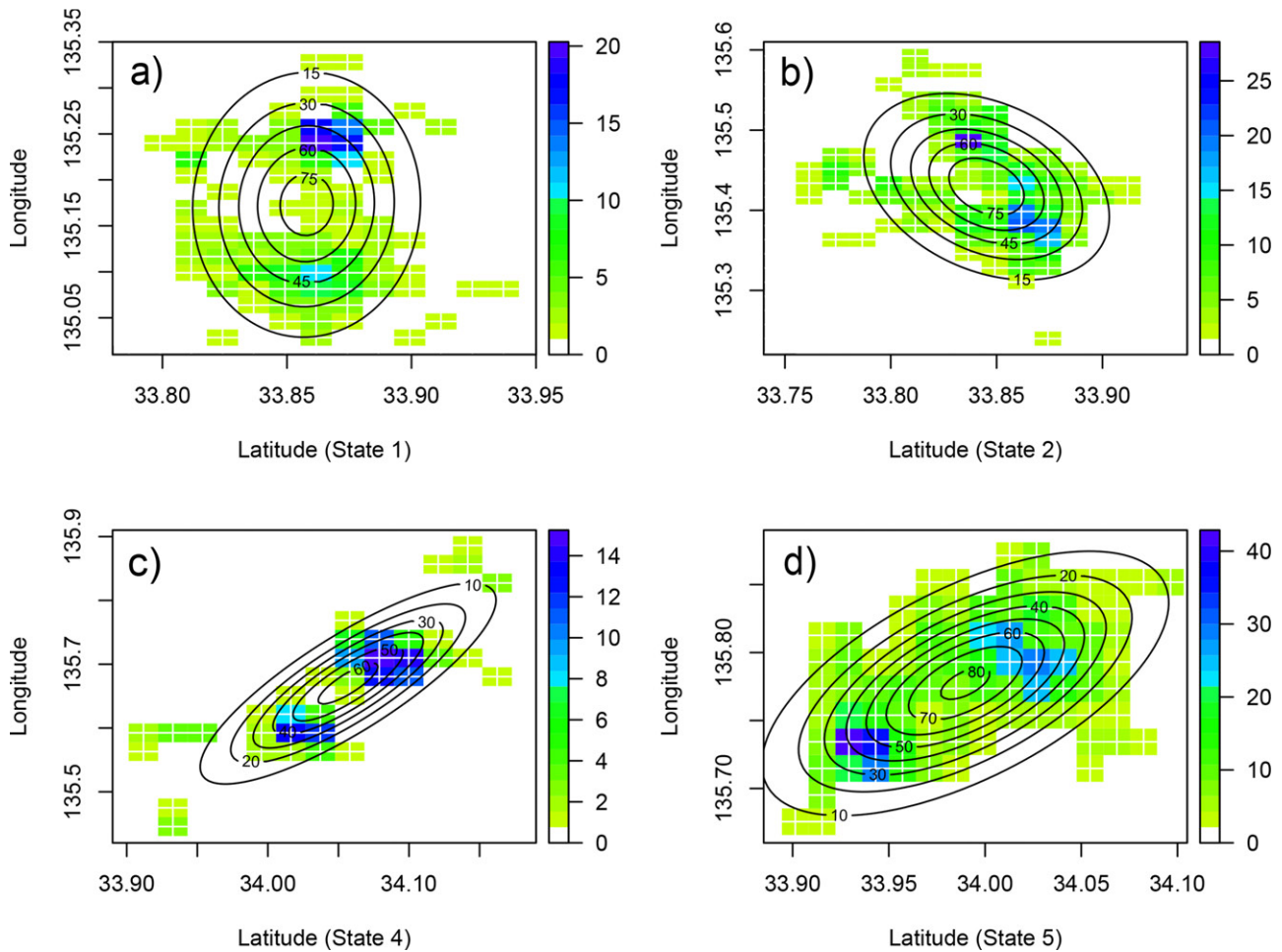


Figure 13. Heatmap representing stochastic reconstruction of tremor locations for (a) state 1, (b) state 2, (c) state 4, and (d) state 5, with the expected conditional distribution contours in black. Dark blue indicates high density identified with stochastic reconstruction.

time t . In [Figure 13](#), the weighted densities for states 1, 2, 4, and 5 are plotted in the form of a two-dimensional heatmap, using the “plot3D” R package (Soetaert 2017), and the expected conditional density contours for each state are overlaid using the estimated parameters from the model. This conditional density is,

$$f(\mathbf{y}_t | S_t = i, z_t = 1) = \frac{1}{2\pi |\hat{\Sigma}_i|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{y}_t - \hat{\boldsymbol{\mu}}_i)^T \hat{\Sigma}_i^{-1} (\mathbf{y}_t - \hat{\boldsymbol{\mu}}_i)\right).$$

The expected bivariate conditional density contours should approximately fit the heatmap of weighted latitude and longitude density for each state if the model provides a good fit, yet the results show multimodal distributions in states 1, 2, 4, and 5 similar to those found using presence residuals. The remaining stochastic reconstruction heatmaps can be found in the supplementary file.

Small spikes may occur in the reconstructed distributions when a cluster of tremors are very close in location. However, clear bi-modality indicates that our model may be improved by the addition of extra states. Here, the stochastic reconstruction identifies the same issues diagnosed by the presence residuals and this lack of fit might be resolved to some extent by using alternative model selection methods other than BIC, which is an unproven criterion in this context.

We also use stochastic reconstruction to calculate moving average values for the probability of tremor occurrence in each state, as described in Equations (3) and (4) of [Section 4.3](#). These moving average values are plotted and compared to the model estimates of p_i when $r = 8760$ (1 year) and $q = 720$ (30 days). The results are shown in [Figure 14](#). It is clear that our assumption of a stationary probability of tremor in each state across time is violated in every state except state 3, with the moving average frequently outside the 95% confidence intervals. This confirms previous findings of non-stationarity without reliance on the estimated Viterbi path and gives further weight to the proposal of a model with varying p_i over time.

6. Discussion and Conclusion

In this article, we aimed to develop new model checking methods for HMMs, particularly for more complex models such as those with two-part emission distributions. Checking the performance of a model remains an important step in our data analysis, not least to direct future model improvements.

We reviewed current methods and developed new predictive and exvisive residuals for whole model checking of HMMs, which are based on previous work in point processes. We also proposed the use of stochastic reconstruction for diagnostics

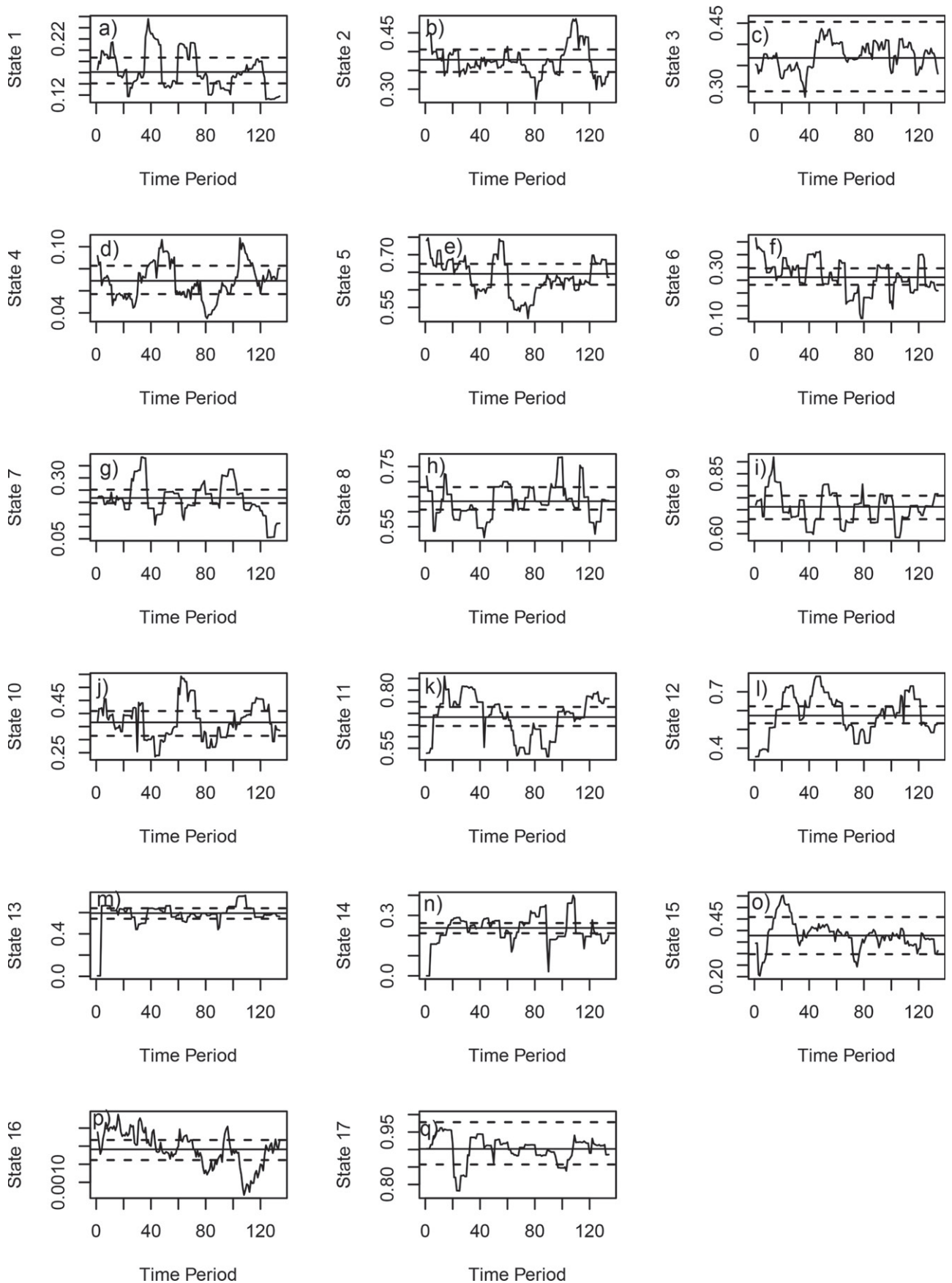


Figure 14. Moving average of the probability of tremor occurrence calculated using stochastic reconstruction for (a) state 1, (b) state 2, (c) state 3, (d) state 4, (e) state 5, (f) state 6, (g) state 7, (h) state 8, (i) state 9, (j) state 10, (k) state 11, (l) state 12, (m) state 13, (n) state 14, (o) state 15, (p) state 16, (q) state 17. Solid lines represent the model estimates of p_i with dashed 95% confidence interval. Values outside the confidence interval indicate nonstationary p_i for $i = 1, \dots, 17$.

as an alternative to previous methods. Here we discuss the advantages of our proposed methods.

6.1. Adequacy of Viterbi Path Based Diagnostics

Testing the Bernoulli and bivariate-normal elements of the emission distributions separately proved informative. Wang et al. (2018) found that the assumption of a constant probability of tremor in each state is not reasonable. We also identified a number of states where p_i is significantly underestimated (Figure 7) in Section 5.1.1. The main disadvantage of these methods is that they are based on the assumption that the Viterbi path is the true Markov chain and misclassification errors may lead to problems with these diagnostics.

The bivariate normal element of the conditional distribution was assessed through the use of presence residuals. Plotting the residuals clearly demonstrated bi-modality in residuals for some states such as 1, 2, 4, and 5 as well as clusters of outliers. These results highlight the importance of checking each state individually rather than relying on the combined presence residuals for all states. The bi-modal distributions in some states suggest more than one process occurring within these states. Wang et al. (2018) selected the two-dimensional model using BIC and we propose investigating the use of an alternative criterion. Potentially, this could lead to a better fit for the bivariate normal element of the conditional distribution by splitting some states further. Again, this method has the disadvantage of being dependent on the Viterbi path and as such is subject to misclassification errors.

6.2. Performance of Predictive and Exvisive Residuals

The predictive and exvisive residuals provide a useful whole model checking tool. They do not rely on the Viterbi path and as discussed in Section 6.1 methods dependent on the Viterbi path are unsatisfactory. We also note that our residuals can be used for stationary and even nonstationary HMMs, whereas methods such as the CDF plots proposed by Altman (2004) not only sometimes fail to identify lack of fit, but are only suitable for stationary HMMs. The new residuals allow our knowledge of the system to be considered in the same way as pseudo-residuals do, but as discussed in Section 3 pseudo-residuals can be difficult to interpret when data contains many zeros. The predictive and exvisive residuals also have the advantage that scaling can be used to calculate different types of residuals through the function $h(t)$, as described in Section 4. We focus only on $h(t) = 1$, but Wang, Wang, and Zhuang (2018) used different forms of $h(t)$ to create Pearson type residuals among others. In addition, there is the potential to develop second-order residuals (Zhuang 2006).

The residuals identified a lack of fit in the case study model, although to a varying degree depending on whether predictive or exvisive versions were used. This is because the exvisive residuals are conditional on all of the exterior data rather than just the historical observations. The lack of fit is due, at least in part, to the underestimation of p_i and the assumption that p_i is stationary, which we can confirm from the stochastic reconstruction diagnostic tests. As well as identifying a lack of fit, we

can see how the predictive and exvisive residuals change in time in Figures 10 and 12.

The predictive and exvisive residuals provide a good first step for whole model checking. However, they are not necessarily useful for diagnosing specific reasons for lack of fit and as such, the importance of testing individual features and states of the model is highlighted.

6.3. Performance of Stochastic Reconstruction

An alternative to Viterbi path based methods for diagnosing an HMM is stochastic reconstruction. Lack of fit in individual states can be identified using this method. We found that the stochastic reconstruction method identified lack of stationarity in parameters p_i and bi-modality of tremor location in some states. We conclude that the stochastic reconstruction method is preferable to Viterbi path based diagnostics as the problem of potential misclassification is avoided. Stochastic reconstruction takes into account the probability of being in any of the states at time t . We have found that there is not one “best” test for goodness of fit for complex HMMs but that a combination of tests which do not rely on the Viterbi path complement each other in identifying and diagnosing problems.

6.4. Performance of the Selected HMM for the Kii Region

The development of predictive and exvisive residuals for HMMs and the use of stochastic reconstruction in this context mark significant progress for model checking, particularly for two-part models such as described in the case study.

Our motivation for this article was to determine whether the selected 17 state two-dimensional HMM with extra zeros was a good fit for the data. We discovered that estimates of p_i were often significantly lower than the empirical values and confirmed findings in Wang et al. (2018) that the assumption of a stationary Bernoulli element of the mixture emission was violated. Future modeling should take account of this. We found that some states exhibited a clear bi-modal distribution given that a tremor occurred. We consider that this might be improved by using an alternative method of model selection. Use of BIC is common in the context of selecting the number of states in an HMM, but its performance is, as yet, unproven (MacKay 2002; Celeux and Durand 2008). Determining the number of states in an HMM is in fact an unresolved area of research and considering different model selection methods may lead to improved fit. The nonvolcanic tremor data are complex and capturing all features of the data is challenging but this article has enabled clear insight into the future direction of modeling this type of data.

6.5. Future Research

A recurring theme throughout the results of our applied model checking is the lack of fit detected in the Bernoulli element of the mixture distribution. We will look to develop a model for nonvolcanic tremor data that allows the value of p_i to vary over time, potentially with the use of splines embedded in the HMM framework. We will also consider options for model selection other than BIC, such as some derivatives of AIC (Akaike 1973),

to see which information criteria select the model that provides the best fit while still being useful in terms of inference. The current model is designed to detect large scale migration patterns and we may wish to consider a model which also takes into account small scale within state migration which currently causes some correlation in the data. This would require the addition of a covariate in the model. Also, one may consider the use of nonparametric estimation methods such as those discussed in Marsan and Lengliné (2008) and Zhuang and Mateu (2019).

There are still further avenues to be explored for new model checking methods. Zhuang (2006) and Clements, Schoenberg, and Veen (2012) discussed the use of second-order statistics in model checking for point processes to detect clustering or inhibition. Second-order residuals are calculated by comparing the variance of the first-order residuals to the expected variance. It may be possible to develop these methods for HMMs in the future, as mentioned in Section 6.2.

Supplementary Materials

The supplementary materials include contour plots of standardized presence residuals and stochastic reconstruction heatmaps for states 3 and 6–17 in the 17 state HMM to complement those in Figures 8 and 13.

Acknowledgments

We would like to thank the associate editor and two anonymous referees for their insightful comments leading to significant improvements in this article.

Funding

This work was supported by the Royal Society of New Zealand Marsden Fund (contract UOO1419). Jiancang Zhuang is partially funded by Grants-in-Aid no. 19H04073 for Scientific Research from the Japan Society for the Promotion of Science.

References

- Aguirre-Hernández, R., and Farewell, V. T. (2002), “A Pearson-Type Goodness-of-Fit Test for Stationary and Time-Continuous Markov Regression Models,” *Statistics in Medicine*, 21, 1899–1911. [861]
- Ailliot, P., Thompson, C., and Thomson, P. (2009), “Space-Time Modelling of Precipitation by Using a Hidden Markov Model and Censored Gaussian Distributions,” *Journal of the Royal Statistical Society, Series C*, 58, 405–426. [861]
- Akaike, H. (1973), “Information Theory as an Extension of the Maximum Likelihood Principle,” in *Second International Symposium on Information Theory*, Akademiai Kiado, Budapest, eds. B. N. Petrov and F. Csaki, pp. 267–281. [872]
- Altman, R. M. (2004), “Assessing the Goodness-of-Fit of Hidden Markov Models,” *Biometrics*, 60, 444–450. [859,861,872]
- Baddeley, A., Turner, R., Møller, J., and Hazelton, M. (2005), “Residual Analysis for Spatial Point Processes,” *Journal of the Royal Statistical Society, Series B*, 67, 617–666. [859,862,863]
- Baum, L. E., and Petrie, T. (1966), “Statistical Inference for Probabilistic Functions of Finite State Markov Chains,” *The Annals of Mathematical Statistics*, 37, 1554–1563. [859]
- Baum, L. E., Petrie, T., Soules, G., and Weiss, N. (1970), “A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains,” *The Annals of Mathematical Statistics*, 41, 164–171. [860]
- Bray, A., and Schoenberg, F. P. (2013), “Assessment of Point Process Models for Earthquake Forecasting,” *Statistical Science*, 28, 510–520. [862]
- Bray, A., Wong, K., Barr, C. D., and Schoenberg, F. P. (2014), “Voronoi Residual Analysis of Spatial Point Process Models With Applications to California Earthquake Processes,” *The Annals of Applied Statistics*, 8, 2247–2267. [862]
- Celeux, G., and Durand, J. (2008), “Selecting Hidden Markov Model State Number With Cross-Validated Likelihood,” *Computational Statistics*, 23, 541–564. [872]
- Clements, R. A., Schoenberg, F. P., and Veen, A. (2012), “Evaluation of Space-Time Point Process Models Using Super-Thinning,” *Environmetrics*, 23, 606–616. [862,873]
- Hall, P., and Heyde, C. C. (2014), *Martingale Limit Theory and Its Application*, New York: Academic Press. [863]
- Harte, D. (2017), “HiddenMarkov: Hidden Markov Models,” R Package Version 1.8-11, Statistics Research Associates, Wellington. [860,863]
- Hassan, M. R., and Nath, B. (2005), “Stock Market Forecasting Using Hidden Markov Model: A New Approach,” in *5th International Conference on Intelligent Systems Design and Applications*, pp. 192–196. [859]
- Korkmaz, S., Goksuluk, D., and Zararsiz, G. (2016), “MVN: An R Package for Assessing Multivariate Normality,” Department of Biostatistics, Faculty of Medicine, Hacettepe University, Ankara. [866]
- Krogh, A., Mian, I. S., and Haussler, D. (1994), “A Hidden Markov Model That Finds Genes in *E. coli* DNA,” *Nucleic Acids Research*, 22, 4768–4778. [859]
- MacKay, R. J. (2002), “Estimating the Order of a Hidden Markov Model,” *The Canadian Journal of Statistics*, 30, 573–589. [872]
- Maeda, T., and Obara, K. (2009), “Spatiotemporal Distribution of Seismic Energy Radiation From Low-Frequency Tremor in Western Shikoku, Japan,” *Journal of Geophysical Research*, 114, B00A09. [860]
- Marsan, D., and Lengliné, O. (2008), “Extending Earthquakes’ Reach Through Cascading,” *Science*, 319, 1076–1079. [873]
- Obara, K. (2002), “Non-Volcanic Deep Tremor Associated With Subduction in South-West Japan,” *Science*, 296, 1679–1681. [860]
- (2011), “Characteristics and Interactions Between Non-Volcanic Tremor and Related Slow Earthquakes in the Nankai Subduction Zone, South-West Japan,” *Journal of Geodynamics*, 52, 229–248. [860]
- Ogata, Y. (1988), “Statistical Models for Earthquake Occurrences and Residual Analysis for Point Processes,” *Journal of the American Statistical Association*, 83, 9–27. [862]
- Pasek, J. (2018), “weights: Weighting and Weighted Statistics,” R Package Version 1.0. [865]
- R Core Team (2018), *R: A Language and Environment for Statistical Computing*, Vienna, Austria: R Foundation for Statistical Computing. [863]
- Rabiner, L. R. (1989), “A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition,” *Proceedings of the IEEE*, 77, 257–286. [859]
- Schoenberg, F. P. (2003), “Multidimensional Residual Analysis of Point Process Models for Earthquake Occurrences,” *Journal of the American Statistical Association*, 98, 789–795. [862]
- Schwarz, G. (1978), “Estimating the Dimension of a Model,” *Annals of Statistics*, 6, 461–464. [860]
- Soetaert, K. (2017), “plot3D: Plotting Multi-Dimensional Data,” R Package Version 1.1.1. [870]
- Titman, A., and Sharples, L. D. (2007), “A General Goodness-of-Fit Test for Markov and Hidden Markov Models,” *Statistics in Medicine*, 27, 2177–2195. [861]
- Viterbi, A. J. (1967), “Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm,” *IEEE Transactions on Information Theory*, 13, 260–269. [860]
- Wang, T., and Bebbington, M. (2013), “Identifying Anomalous Signals in GPS Data Using HMMs: An Increased Likelihood of Earthquakes?,” *Computational Statistics and Data Analysis*, 58, 27–44. [859]
- Wang, T., Bebbington, M., and Harte, D. (2012), “Markov-Modulated Hawkes Process With Stepwise Decay,” *Annals of the Institute of Statistical Mathematics*, 64, 521–544. [859]
- Wang, T., Zhuang, J., Buckby, J., Obara, K., and Tsuruoka, H. (2018), “Identifying the Recurrence Patterns of Non-Volcanic Tremors Using a 2D Hidden Markov Model With Extra Zeros,” *Journal of Geophysical Research: Solid Earth*, 123, 6802–6825. [859,860,861,866,872]

- Wang, T., Zhuang, J., Obara, K., and Tsuruoka, H. (2017), “Hidden Markov Modelling of Sparse Time Series From Non-Volcanic Tremor Observations,” *Journal of the Royal Statistical Society, Series C*, 66, 691–715. [860,861,862]
- Wang, Y., Wang, T., and Zhuang, J. (2018), “Modelling Continuous Time Series With Many Zeros and an Application to Earthquakes,” *Environmetrics*, 29, e2500. [859,862,863,872]
- Yang, Y., and Simpson, D. G. (2010), “Conditional Decomposition Diagnostics for Regression Analysis of Zero-Inflated and Left-Censored Data,” *Statistical Methods in Medical Research*, 21, 393–408. [861]
- Zhuang, J. (2006), “Second-Order Residual Analysis of Spatiotemporal Point Processes and Applications in Model Evaluation,” *Journal of the Royal Statistical Society, Series B*, 68, 635–653. [859,862,863,872,873]
- (2015), “Weighted Likelihood Estimators for Point Processes,” *Spatial Statistics*, 14, 166–178. [862,863]
- Zhuang, J., and Mateu, J. (2019), “A Semiparametric Spatiotemporal Hawkes-Type Point Process Model With Periodic Background for Crime Data,” *Journal of the Royal Statistical Society, Series A*, 182, 919–942. [873]
- Zhuang, J., Ogata, Y., and Vere-Jones, D. (2004), “Analyzing Earthquake Clustering Features by Using Stochastic Reconstruction,” *Journal of Geophysical Research*, 109, B05301. [865]
- Zucchini, W., and MacDonald, I. (2009), *Hidden Markov Models for Time Series: An Introduction Using R*, Boca Raton, FL: CRC Press. [859,860,863]