

# Residual analysis for a zero-inflation hidden Markov model

庄建倉 モデリング研究系 准教授

## Abstract

Residual analysis is important and powerful model diagnostics and improvement. This research illustrates how to improve the formulation of a zero-inflation hidden Markov model for tremor occurrences in the Kii region through constructing corresponding residual statistics.

**Keywords:** innovation, residual analysis, hidden Markov model, tremor

## Introduction and Motivation

Statistical modelling has been increasingly and widely used in many research fields. Associated statistical inference techniques, such as model selection, specification, and evaluation, parameter estimation, and testing goodness-of-fit have been developed. Among them residual analysis can be used to evaluate a model and assess the goodness-of-fit without proposing unnecessary new models whose implementation may include heavy programming and computation tasks (Zhuang, 2006; Baddeley et al., 2005). Residual analysis is also helpful for understanding the advantages of a model even if its overall performance is not ideal.

For hidden Markov models, research on model checking for them is under developed and there is a need for new methods. Recently, Buckby et al. (2020) developed a similar innovation based residuals by extending the framework of residual analysis in Zhuang (2015). In Buckby et al. (2020), residual analysis based stochastic reconstruction methods are proposed to verify whether the model formulation is proper for describing data and how to modify the model formulation if it is not.

## Hidden Markov model

An HMM is used to model time series data when the observations are dependent on an underlying unobserved Markov chain. Observations are realised 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  $t = 1, \dots, T$ ,  $S_t$  is the state of the Markov chain at time  $t$  and  $m$  is the finite number of states. The simplest HMM satisfies the conditions,

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

and

$$f(x_t | x_1, \dots, x_{t-1}, S_1 = s_1, S_2 = s_2, \dots, S_t = s_t) = f(x_t | S_t = s_t) \quad (2)$$

## Innovation based residual analysis

In an HMM, the expectation and pdf of  $X_t$ , given the previous observations are, respectively,

$$\mathbf{E}[X_t | \mathcal{H}_t] := \mathbf{E}[X_t | X_1, \dots, X_{t-1}] = \frac{\sum_{k=1}^m \alpha_{t-1}(k) \Gamma_{ki} \mathbf{E}[X_t | S_t = i]}{\sum_{k=1}^m \alpha_{t-1}(k)} \quad (3)$$

$$f(x_t | \mathcal{H}_t) := f(x_t | X_1, \dots, X_{t-1}) = \sum_{k=1}^m \frac{\alpha_{t-1}(k) \Gamma_{ki} f(x_t | S_t = i)}{\sum_{k=1}^m \alpha_{t-1}(k)} \quad (4)$$

where  $\alpha_t(k)$  is the forward probability defined by

$$\alpha_t(k) := P(X_1, \dots, X_t, S_t = k). \quad (5)$$

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

Using the above expectation, the predictive residual with a predictable function  $h(t, x)$ , where  $h$  is a predictable function of  $t$  for every fixed  $x$ , becomes,

$$R_n := \sum_{t=1}^n (h(t, X_t) - \int h(t, x_t) f(x_t | \mathcal{H}_t) dx_t) \quad (6)$$

If  $h(t, x) = h(t)x$ , the above formula becomes the predictive residual proposed in Buckby et al. (2020)

$$R_n = \sum_{t=1}^n (h(t)X_t - h(t) \mathbf{E}[X_t | \mathcal{H}_t]). \quad (7)$$

The exvisive residuals can be defined in a similar way to the predictive residuals. That is,

$$\mathbf{E}[X_t | \mathcal{E}_t] := \mathbf{E}[X_t | X_1, \dots, X_{t-1}, X_{t+1}, \dots, X_T] = \frac{\sum_{k=1}^m \alpha_{t-1}(k) \Gamma_{ki} \mathbf{E}[X_t | S_t = i] \beta_t(i)}{\sum_{k=1}^m \alpha_{t-1}(k) \Gamma_{ki} \beta_t(i)} \quad (8)$$

$$f(x_t | \mathcal{E}_t) := \frac{\sum_{k=1}^m \alpha_{t-1}(k) \Gamma_{ki} f(x_t | S_t = i) \beta_t(i)}{\sum_{k=1}^m \alpha_{t-1}(k) \Gamma_{ki} \beta_t(i)} \quad (9)$$

where  $\beta_t(i) = P(X_{t+1}, \dots, X_T | S_t = i)$  is the backward probability. Given a function  $h(t, x)$  which is exvisible as a function of  $t$  for every  $x$ , the exvisive residual becomes

$$R_n^e := \sum_{t=1}^n \left[ h(t, X_t) - \int h(t, x_t) f(x_t | \mathcal{E}_t) dx_t \right]. \quad (10)$$

If  $h(t, x) = xh(t)$ , where  $h(t)$  is an exvisible function, the above residual becomes the form proposed by Buckby et al. (2020)

$$R_n^e = \sum_{t=1}^n (\hat{h}(t)X_t - \hat{h}(t) \mathbf{E}[X_t | \mathcal{E}_t]). \quad (11)$$

It is easy to see  $\mathbf{E}[R_n] = 0$  and  $\mathbf{E}[R_n^e] = 0$ .

## Zero-inflation HMMs for tremors

**Model formulation** The HMM example is the zero-inflation model used by Wang et al. (2018) for modelling the non-volcanic 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 Nanaki subduction zone of Japan by Obara (2002). Modelling nonvolcanic tremor activity is of interest because it associated with slow slip events. The occurrence of tremor and their locations are detected each hour by using a high-sensitivity seismograph network.

In the HMM in Wang et al. (2018), the observation  $X_t$  at time  $t$  is from a space of  $\{0\}$  (null events) and  $\{1\} \times \mathbb{R}^2$  (tremors and locations). Null events are indicated by  $Z_t = 0$ . The joint emission distribution, which belongs to so-called zero-inflation distribution, for each state at each time is given by

$$f(y_t, z_t | S_t = i) = (1 - p_i)^{1-z_t} \left( p_i \frac{1}{2\pi|\Sigma_i|^{1/2}} \exp \left[ -\frac{1}{2}(y_t - \mu_i)^T \Sigma_i^{-1} (y_t - \mu_i) \right] \right)^{z_t} \quad (12)$$

where  $\mu_i = \mathbf{E}[Y_t | Z_t = 1, S_t = i]$ ,  $\Sigma_i = \mathbf{Var}[Y_t | Z_t = 1, S_t = i]$ . The hidden states,  $S_t$ , take value from 1 to  $m$ . The model was fitted using the EM algorithm for different numbers of states. A 17 state model was used for the Kii region (Figure 1). Three types of tremor segments can be classified according to the results: **episodic**, **weak concentration**, and **background**.

**Residual analysis** To verify the emission distribution, with each data point being weighted by

$$\varphi_{it} := P(S_t = i | Y_1, \dots, Y_T, Z_1, \dots, Z_T) = \frac{\alpha_t(i) \beta_t(i)}{\sum_{k=1}^m \alpha_t(k) \beta_t(k)},$$

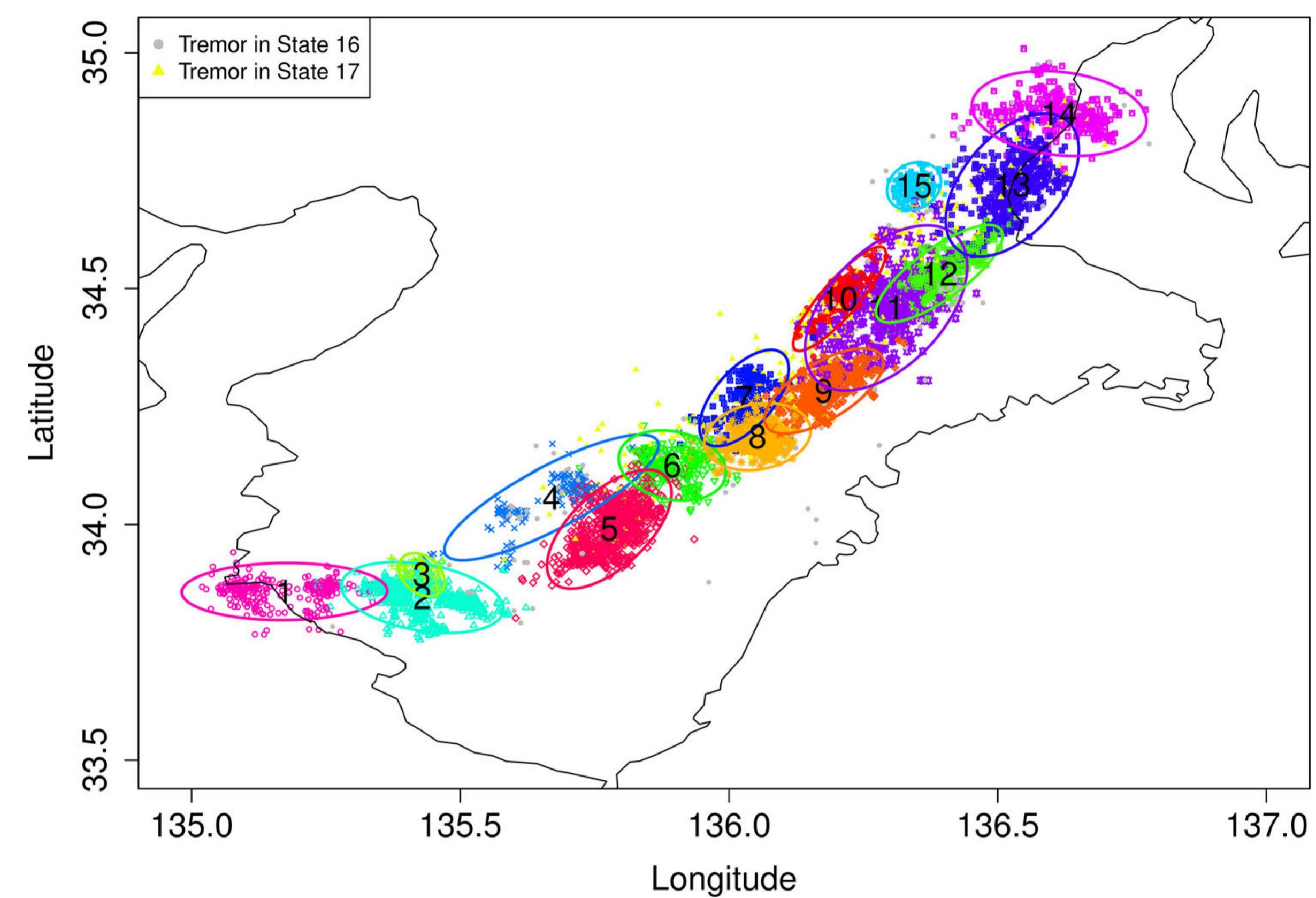
consider the following density function

$$\tilde{f}(y | S = i, Z = 1) \propto \sum_{t=1}^T \varphi_{it} I(y_t \in (y - \Delta y, y + \Delta y)) I(Z_t = 1), \quad (13)$$

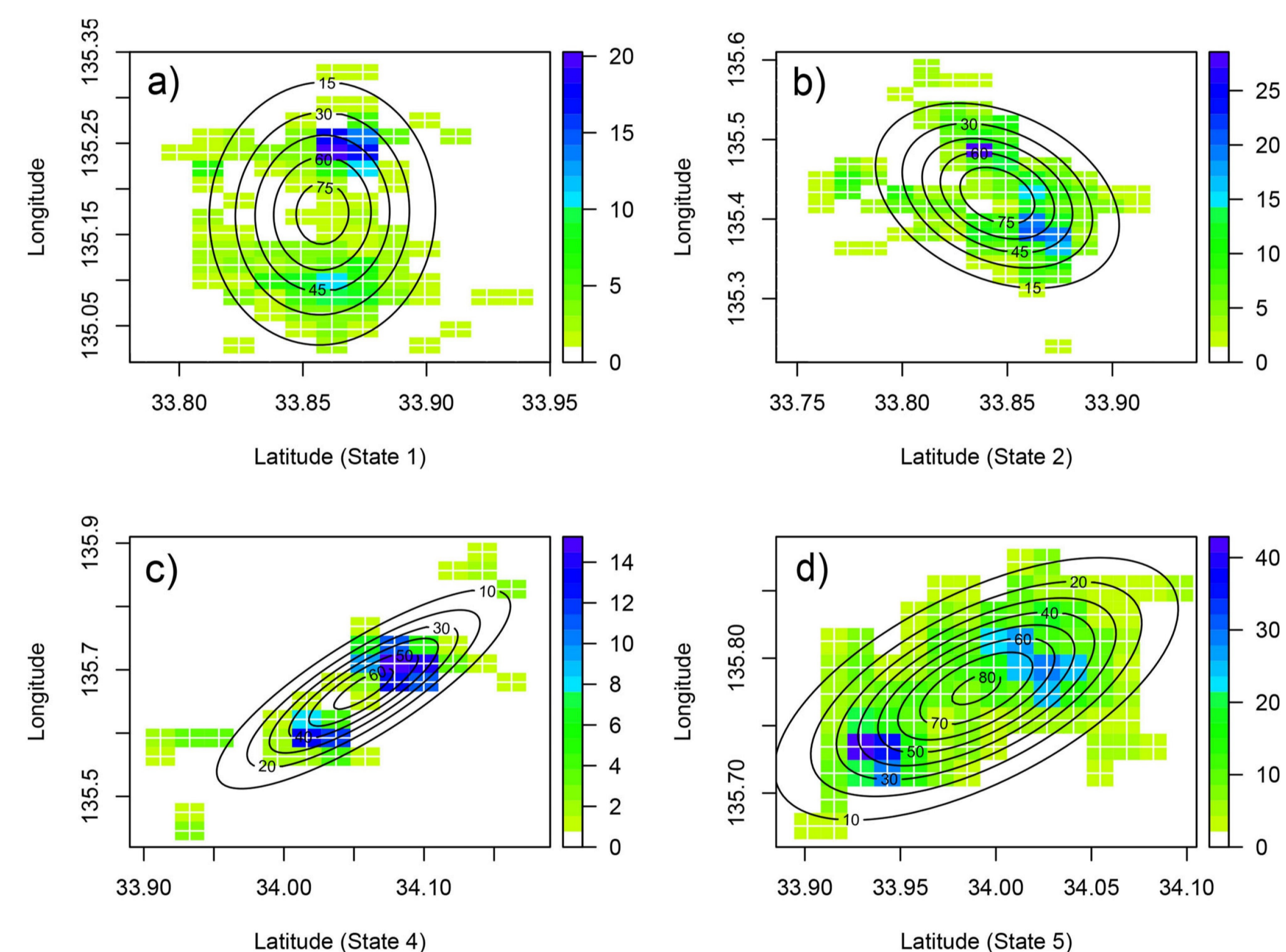
where  $\Delta y$  is a fixed arbitrary small increment in the value of  $y$ ,  $I$  is the indicator function. By equation (13),

$$\tilde{f}(y | S = i, Z = 1) \approx f(y_t | S_t = i, Z_t = 1) = \frac{1}{2\pi|\Sigma_i|^{1/2}} \exp \left[ -\frac{1}{2}(y_t - \mu_i)^T \Sigma_i^{-1} (y_t - \mu_i) \right], \quad (14)$$

That is,  $\tilde{f}(y | S = i, Z = 1)$  is a reconstructed version of  $f(y_t | S_t = i, Z_t = 1)$ . Figure 2 plots the reconstructed densities for States 1, 2, 4, and 5, in the form of heat-map, and the corresponding densities in the fitting results are superposed in contour lines. However, the heat-maps indicate the reconstructed density are bimodal, indicating that BIC might have selected fewer number of states.



**Figure 1:** (c.f. Buckby et al., 2020) The fitted 17 state model for nonvolcanic tremors in the Kii region. Observations are classified into different spatiotemporal states represented by different colors. Each state has a unique probability of tremor occurrence.



**Figure 2:** (c.f. Buckby et al., 2020) Reconstructed density functions (heatmaps) for tremor locations for (a) state 1, (b) state 2, (c) state 3, and (4) state 5, and their corresponding densities (contour lines) obtained from fitting the model to data with an EM algorithm.

## Summary

Based on the theory of residual analysis, the stochastic reconstruction technique for hidden Markov models is helpful for us to find a proper direction in which to improve on the model formulation, in a data-driven manner.

## References

- Baddeley, A., Turner, R., Møller, J., and Hazelton, M. (2005). Residual analysis for spatial point processes (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(5):617–666.
- Buckby, J., Wang, T., Zhuang, J., and Obara, K. (2020). Model checking for hidden Markov models. *Journal of Computational and Graphical Statistics*, 0(0):1–16.
- Wang, T., Zhuang, J., Buckby, J., Obara, K., and Tsuruoka, H. (2018). Identifying the recurrence patterns of nonvolcanic tremors using a 2-D hidden Markov model with extra zeros. *Journal of Geophysical Research: Solid Earth*, 123(8):6802–6825.
- Zhuang, J. (2006). Second-order residual analysis of spatiotemporal point processes and applications in model evaluation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(4):635–653.
- Zhuang, J. (2015). Weighted likelihood estimators for point processes. *Spatial Statistics*, 14(B):166–178.
- Zucchini, W. and MacDonald, I., editors (2009). *Hidden Markov models for time series: An Introduction Using R*. CRC Press, Boca Raton, 2nd edition.