

A hidden Markov model for earthquake declustering

Zhengxiao Wu¹

Received 8 August 2009; revised 18 August 2009; accepted 9 November 2009; published 12 March 2010.

[1] The hidden Markov model (HMM) and related algorithms provide a powerful framework for statistical inference on partially observed stochastic processes. HMMs have been successfully implemented in many disciplines, though not as widely applied as they should be in earthquake modeling. In this article, a simple HMM earthquake occurrence model is proposed. Its performance in declustering is compared with the epidemic-type aftershock sequence model, using a data set of the central and western regions of Japan. The earthquake clusters and the single earthquakes separated using our model show some interesting geophysical differences. In particular, the log-linear Gutenberg-Richter frequency-magnitude law (G-R law) for the earthquake clusters is significantly different from that for the single earthquakes.

Citation: Wu, Z. (2010), A hidden Markov model for earthquake declustering, *J. Geophys. Res.*, 115, B03306, doi:10.1029/2008JB005997.

1. Introduction

[2] Earthquakes often occur in clusters. Conventionally, the largest earthquake in a cluster is called the main shock, those before it are called foreshocks, and those after it are called aftershocks (hence they can only be identified after the earthquake cluster has died out [Hough and Jones, 1997]). In this view aftershocks do not have aftershocks of their own. An understanding of earthquake clusters provides the opportunity to characterize stress relaxation after a large earthquake as well as possibly to provide information important for hazards analysis. Given an earthquake data set, it is often the first step to classify the earthquakes as members of different clusters or as single earthquakes. Zhuang *et al.* [2002, 2004, 2005] address the importance of earthquake declustering and demonstrated its usage.

[3] Traditional declustering methods in seismology include window-based methods [Utsu, 1969; Gardner and Knopoff, 1974; Keilis-Borok and Kossobokov, 1986] and link-based methods [Reasenber, 1985; Frohlich and Davis, 1990; Davis and Frohlich, 1991]. A common feature of these methods is that the choice of parameter values is subjective. Often no guidelines are available for how to optimally choose the model parameters.

[4] This problem is solved by Zhuang *et al.* [2002]. Zhuang *et al.* [2002, p. 369] suggested that “the declustering method should be based on a stochastic model to objectively quantify the observations, so that any given event has a probability to be either a background event or an offspring (cluster) event generated by others.” Hence, the choice of parameters is made by maximum likelihood estimation

(MLE), whose usage has been well justified in the statistical literature.

[5] Zhuang *et al.*’s [2002] method is based on a space-time version of the epidemic-type aftershock sequence (ETAS) model proposed by Ogata [1998]. The ETAS model is a branching process with immigrants, with each existent earthquake (“ancestor”) producing offspring independently. In ETAS, Ogata [1998] carefully selects the functional form of the intensity of the process to accommodate the Gutenberg-Richter frequency-magnitude law and the empirical Omori law for aftershock sequences.

[6] Zhuang *et al.* [2002] classify earthquakes into background events and offspring events. The background events are the immigrants of the branching process, or earthquakes without an ancestor. They are interpreted as members of the global seismic activity in the ETAS model. Offspring are produced by each existent earthquake independently. In the ETAS model, the offspring are named aftershocks. This new definition of aftershock is based only on the occurrence time of each earthquake, neglecting its magnitude. Under such a definition, one can identify the aftershocks in real time and aftershocks can have their own aftershocks. Branching models such as ETAS are currently well accepted in the seismological community.

[7] We find that declustering based on ETAS (as in the work by Zhuang *et al.* [2002, 2004, 2005]) is not satisfactory for three reasons. First, the decision made by the method does not adapt itself to the new observations. Suppose an earthquake sequence X_1, X_2, \dots, X_n occurred at times $\tau_1, \tau_2, \dots, \tau_n$. In the determination of the probability that the earthquake X_i is a background event, the computation is solely based on the previous earthquakes X_1, X_2, \dots, X_{i-1} and $\tau_1, \tau_2, \dots, \tau_{i-1}$, ignoring the later observations. Therefore, the information in the data set is not fully utilized.

[8] Second, for a given earthquake catalog and a fixed set of declustering parameters, the method gives a different partition each time it is run. The reason for this is that

¹Department of Statistics and Applied Probability, National University of Singapore, Singapore.

random numbers are generated in the procedure (for example, algorithm 2 and algorithm 3 of Zhuang *et al.* [2002]) and the final partition depends on the realization of these random numbers. This is why it is called a “stochastic declustering.” In practice, though, a unique declustering is preferred.

[9] Third, as mentioned, the ETAS model operates on the assumption that each individual earthquake produces offspring independently. As a consequence, it could lead to a supercritical branching process [Helmstetter and Sornette, 2002], which happens when the expected number of offspring for each earthquake is strictly greater than one. In the supercritical regimes, the number of earthquakes grows exponentially. This characteristic is certainly not observed for real earthquake sequences, which always end.

[10] As the mechanism of earthquake sequence triggering is not yet fully understood, it is more preferable to model earthquake sequences with the simple intuition behind the window-based methods (that is, the earthquake cluster is formed by earthquakes which are near one another in time and space) than to impose questionable assumptions as are employed by the ETAS model (a discussion of some of these assumptions can be found in section 5).

[11] This motivates the filtering model (an introduction to filtering model theory is given by Xiong [2008]). Algorithms associated with this model are rigorously developed in terms of continuous time-marked point processes of Wu [2009] from a probabilist’s perspective. In this article, we present the filtering model as a discrete time hidden Markov chain model so that it is more approachable to geophysical scientists. Some generality is lost, but we believe that it illustrates the model.

[12] The paper is organized as follows: Section 2 describes the hidden Markov model (HMM) model; section 3 illustrates its application and compares numerical results from the model with the results of Zhuang *et al.* [2002]; section 4 discusses some of the geophysical findings produced by the HMM; section 5 summarizes the conclusions and outlines future work; Appendix A contains the detailed dynamics of the hidden Markov chain constructed in section 2.

2. Description of the Model

[13] The HMM and related algorithms [Baum *et al.*, 1970; Viterbi, 1967] provide a powerful framework for statistical inference on partially observed stochastic processes. They have been widely applied in many disciplines including signal processing, biology, genetics, etc. However, only recently have there been a few applications in seismology [Granat and Donnellan, 2002; Ebel *et al.*, 2007].

[14] A HMM consists of a sequence of observations O_1, O_2, \dots, O_n and a sequence of hidden states h_1, h_2, \dots, h_n . The distribution of observations O_i only depends on the hidden state h_i , while the hidden state sequence h_1, h_2, \dots forms a Markov chain. Two clever algorithms make statistical inference on HMM possible. The forward-backward algorithm [Baum *et al.*, 1970] computes the conditional distribution $P(h_i | O_1, \dots, O_n)$ and the likelihood function $P(O_1, \dots, O_n)$ of the model (hence searching for the MLE is just an optimization problem); the Viterbi algorithm [Viterbi, 1967] computes the most likely hidden state sequence. A brief introduction of these algorithms is given by Granat and Donnellan [2002].

[15] We propose to model our observations as a superposition of two independent processes: single earthquakes and earthquake clusters. The single earthquake occurrences follow a time-homogeneous (not necessarily space-homogeneous) Poisson process, while clusters are randomly initiated and eventually die out. The first earthquake in a cluster is the “mother earthquake.” When the cluster is active, “offspring” occur in the neighborhood of the mother earthquake where the neighboring area is defined by a bivariate Gaussian distribution. Each occurrence of an offspring earthquake has a probability p of deactivating the cluster (in other words, each birth has a probability p of sterilizing the mother). Further, there can be no more than one active cluster at a time, and hence a new cluster starts only after the old mother earthquake is sterilized. Initially, our model does not include the magnitudes of the events.

[16] Rather than make assumptions about how earthquakes reproduce, the model employs only the intuition that the earthquakes within a cluster are close in space and time to each other, and thus offspring within the cluster occur near the mother earthquake. Note that although it is called the mother earthquake, the first earthquake in a cluster is not necessarily the sole trigger of the offspring. The cause of the offspring occurrences is abstracted as “the cluster is active” in the HMM, or the offspring are triggered by the combined loading and action of all previous earthquakes in the cluster. So the mother earthquake in our model is but a formal mother, or formally the only earthquake that reproduces. Offspring, on the other hand, can be formally considered infertile. Thus, if offspring are seen as “aftershocks” here, then they do not have their own aftershocks. This coincides with the implication of the conventional definition of aftershocks. However, as with the ETAS model, magnitudes do not come into play in the “family structure” of the earthquake cluster in the HMM. The mother earthquake does not have to be the main shock unless, of course, it turns out to be the largest one in the cluster. In other words, it could be a foreshock in the conventional sense if it has an offspring with a greater magnitude.

[17] Suppose the data set contains n earthquakes. Earthquake i ($1 \leq i \leq n$) has the observation $O_i = (x_i, y_i, t_i)$, which are the longitude and latitude of the epicenter and the origin time, respectively. The hidden state is $h_i = (x_i, y_i, t_i, J_i, C_i, A_i)$, where J_i is the index of the most recent mother quake up to i , i.e., $J_i = \max\{k \leq i: \text{earthquake } k \text{ is a motherquake}\}$. C_i and A_i are two indicator functions: $C_i = 1$ if earthquake i is a cluster earthquake and $C_i = 0$ otherwise; $A_i = 1$ if the mother earthquake J_i is fertile at time t_i (the cluster is still active); $A_i = 0$ otherwise. O_i is a subset of h_i . The detailed specification of the transition probabilities $p(h_{i+1} | h_i)$ can be found in Appendix A for the data set used in section 3.

3. Application to an Earthquake Data Set

[18] The same data set used by Zhuang *et al.* [2002] is used in this study: the earthquakes occurred in the period 1926–1995, in the rectangular area 33° – 39° N and 131° – 140° E, with magnitudes greater than 4.0 and depths less than 100 km.

[19] Applying the forward-backward algorithm and the Nelder-Mead simplex method (an optimization procedure [see Lagarias *et al.*, 1998]), MLEs are found: $\hat{\gamma} = 0.1070$,

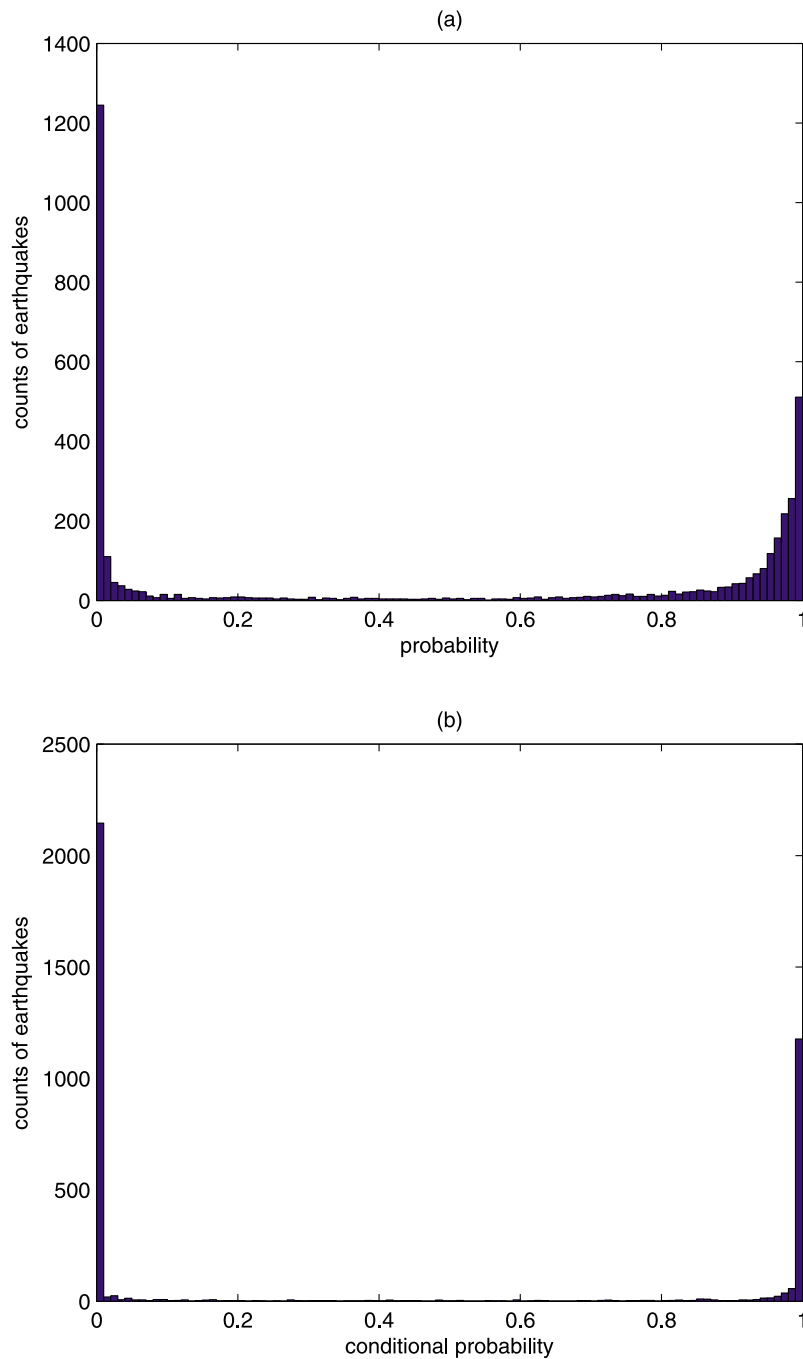


Figure 1. Histograms of (a) aftershock probabilities under ETAS model; (b) conditional probabilities to be in a cluster under the HMM model.

$\hat{\lambda} = 1.3274$, $\hat{\epsilon} = 0.0126$, $\hat{d} = 0.0070$, $\hat{p} = 0.2035$. We refer to Appendix A for the definitions of the parameters.

[20] The MLEs are used in the forward-backward algorithm once more, yielding $P(C_i = 1|O_1, \dots, O_n)$ for all earthquakes i , $1 \leq i \leq n$. This output provides the conditional probability of an earthquake i being a cluster earthquake. These conditional probabilities are plotted in histogram form and compared with the results of Zhuang *et al.* [2002]. Figure 1a is the histogram of the aftershock(offspring) probabilities of Zhuang *et al.* [2002] and Figure 1b shows our result).

[21] As mentioned, Zhuang *et al.* [2002] classified the earthquakes into background events and offspring, while the HMM separates the earthquakes into single earthquakes and earthquake clusters. A correspondence can be established between these two classifications, as the single earthquakes from the HMM can be considered the same as the background events without offspring in ETAS, and the mother earthquakes in the HMM are the same as the background events with offspring in ETAS. However, there is no way to rigorously compare the performance of the two declustering schemes on the same footing because a unique declustering is not available in ETAS model. Nevertheless, both histograms

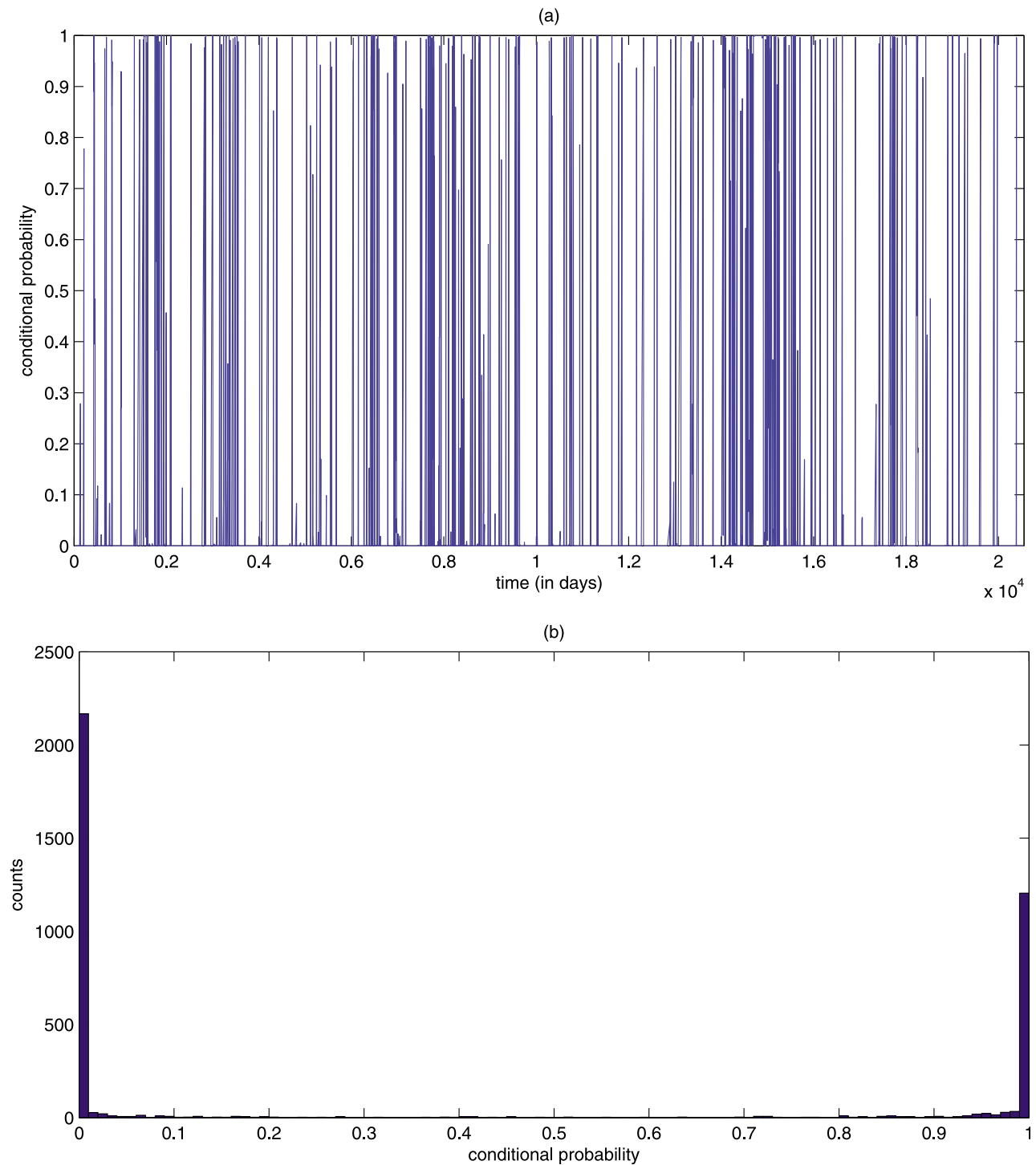


Figure 2. (a) The conditional probability that whether a cluster was alive versus time (in days). (b) Histogram of the conditional probability.

in Figure 1 have counts concentrated near probability 0 or 1. This bimodal pattern indicates that both schemes can classify most of the events with high statistical confidence. In Figure 1a, 18.3% of the events have probability of being an aftershock ranging from 0.1 to 0.9, and 4.9% of earthquakes have probability of being in clusters ranging from 0.1 to 0.9 in Figure 1b. Very few earthquakes are uncertain as to their single or cluster status. The main reason why the HMM

model gives a more decisive answer than ETAS is discussed in section 1. The HMM model makes full use of the data set: the decision about earthquake k is based on all the earthquakes, while the ETAS is not adaptive: the decision about earthquake k only depends on earthquake 1, 2, ..., $k-1$.

[22] The conditional probabilities $P(A_i = 1 | O_1, \dots, O_n)$ ($1 \leq i \leq n$) are computed to illustrate the status of clustering at different times. Figure 2a gives us the conditional

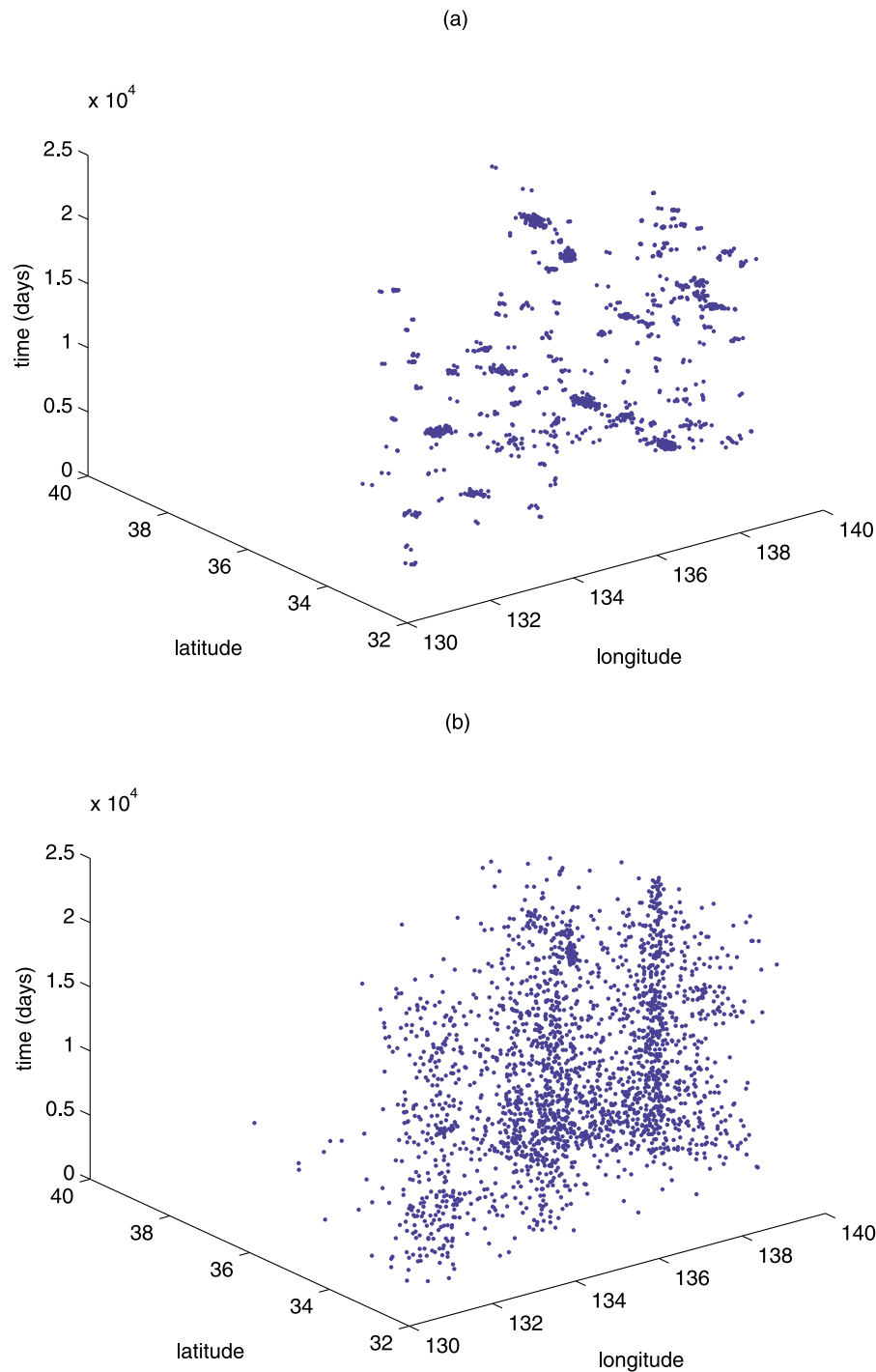


Figure 3. Plots of (a) 237 earthquake clusters and (b) 2327 single earthquakes.

probability that an earthquake cluster was active in this region over time. Here we again can see from Figure 2 that a mere 4.4% of the conditional probabilities range from 0.1 to 0.9.

4. The Most Likely Hidden State Sequence and Data Analysis

[23] The most likely cluster sequence for the HMM can be selected by the Viterbi algorithm. It is the logical choice if a

unique partition is preferred. The resulting partition is shown in Figure 3: 237 clusters that contain 1444 cluster earthquakes are in Figure 3a; and the 2327 single earthquakes are in Figure 3b. Data analysis in this section is based on this declustering. The results discussed in sections 4.1–4.3 imply that our model is geophysically sensible.

4.1. Waiting Time Between Earthquakes

[24] Poisson processes have been widely used in modeling earthquakes. The simplest case would be assuming a Pois-

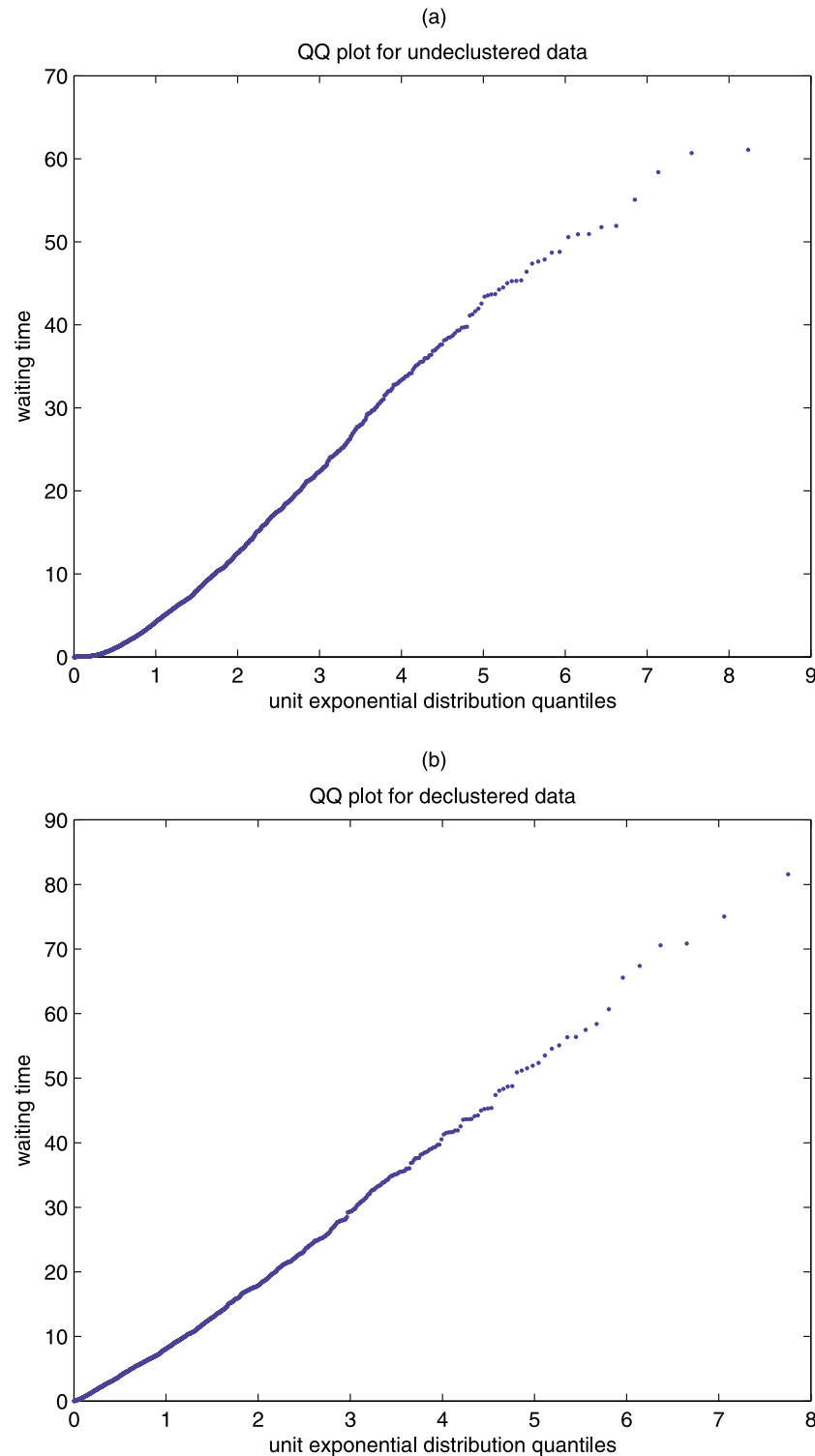


Figure 4. (a) QQ plot (exponential distribution) of waiting time before declustering. (b) QQ plot (exponential distribution) of waiting time for declustered data set.

son process with a constant intensity. The waiting time between two events in a Poisson process is exponentially distributed. We now compare our data to such a distribution.

[25] A simple and intuitive tool, a QQ plot [Wilk and Gnanadesikan, 1968], is applied to check this distributional assumption. The quantiles of the observed waiting times

are plotted against the expected quantiles of the exponential distribution (Figure 4). A strong linear pattern implies a good fit.

[26] Before the declustering, a straight line does not fit the QQ plot well, especially at the lower end. This discrepancy is caused by the existence of earthquake clusters. After the

earthquake clusters are removed, it can be seen that the constant intensity Poisson process models the single earthquakes better.

4.2. Magnitude

[27] Note that the HMM does not incorporate any information of the earthquake magnitude, only the locations of the epicenters. If the model does not make any sense geophysically, then the partitioned events would resemble two random samples (one composed of single earthquakes and the other composed of earthquake clusters) from the given earthquake catalog, in which case no statistically significant difference would be found for the two categories' magnitudes.

[28] Recall the log-linear Gutenberg-Richter(G-R) frequency-magnitude law of earthquake occurrence:

$$\log_{10} N_M = a - bM$$

where N_M is the number of earthquakes which are greater than M . Figure 5 plots $\log_{10} N_M$ against M for single earthquakes and cluster earthquakes. A simple linear regression analysis is done on both plots: the 99.5% confidence intervals for a are [7.1629, 7.2115] (single earthquakes) and [5.9214, 5.9564] (cluster earthquakes), respectively, and the 99.5% confidence intervals for b are [0.9478, 0.9586] (single earthquakes) and [0.6879, 0.6954] (cluster earthquakes), respectively. Therefore, the G-R law for the single earthquakes is statistically different from that for the cluster earthquakes.

[29] Furthermore, three Wilcoxon rank sum tests (*Wilcoxon* [1945]) are conducted to compare the median magnitudes of different earthquake categories. In the first test, the magnitudes of the cluster earthquakes are significantly greater than those of the single earthquakes (p value = 7.1875×10^{-17}). The median of the magnitudes is 4.5 for cluster earthquakes, and it is 4.2 for the single earthquakes.

[30] It is also interesting to see how the mother earthquakes differ in magnitudes from their offspring in clusters. As one would expect, the second Wilcoxon test concludes that the mother earthquakes are significantly bigger than the offspring (p value = 1.0845×10^{-16} ; the median magnitude of mother earthquakes is 4.8; the median magnitude of offspring is 4.4).

[31] The third test shows that even though the offspring are smaller than the mother earthquakes within a cluster, they are still significantly larger than single earthquakes (p value = 1.6256×10^{-7} in the Wilcoxon rank test).

4.3. Earthquake Intensity

[32] In Appendix A, for simplicity a space-homogeneous intensity γ (γ is a constant) is used in modeling the single earthquakes. However, the dynamics of the HMM can be extended to the case where the intensity of the single earthquakes is space-inhomogeneous (*Wu* [2007]), that is, it takes the form of $\gamma(x, y)$, where x, y is the longitude and latitude.

[33] The following iterative scheme is used to estimate the real space-inhomogeneous single earthquake intensity. More details are given by *Wu* [2007].

[34] 1. Set $k(x, y) = 1$.

[35] 2. Find the MLEs of $c, \lambda, \epsilon, d, p$, where c satisfies $\gamma(x, y) = ck(x, y)$.

[36] 3. Run the Viterbi algorithm and pick out the most likely single earthquakes.

[37] 4. Smooth out these single earthquakes to get an $k_{new}(x, y)$ (hence, $k_{new}(x, y)$ is an updated estimate of the single earthquake intensity up to a constant).

[38] 5. Go to step 2 until convergent.

[39] The algorithm converges after 3 iterations in the sense that the third iteration picks out the same set of single earthquakes as the second iteration. At the end of the iterations, the partition is very similar to the result when a space-homogeneous intensity γ is assumed. Repeat the data analysis on it, almost the same conclusions are drawn. Some of the geophysical findings are even more distinct. For example, the median of the mother earthquakes' magnitudes is 4.9. The intensity of single earthquakes (up to a constant) is plotted in Figure 6. Figure 7 shows the distribution of the mother earthquake epicenters. They describe how likely the single earthquakes and the cluster earthquakes would occur in different locations. They are certainly crucial for after-shock prediction and hazard assessment.

[40] The ETAS model assumes that the background events, regardless of their single or cluster status, are generated by a common intensity. Figures 6 and 7 cast doubt on such an assumption. The intensity of the background events with or without offspring could differ. In the data set, there are 2327 single earthquakes and 237 mother earthquakes. The ETAS model mixes the mother earthquakes with the large number of single earthquakes, hence it will never catch the distinct distribution of mother earthquakes, which offers important clues to the global activity of earthquake clusters.

5. Discussion

[41] The goal of declustering earthquake catalogs is to identify the earthquake clusters, thereby various hypotheses of the underlying physical mechanism concerning cluster and earthquake generation can be tested. For declustering, the HMM in this paper starts with the basics. The main assumption is nothing more than an intuition, namely, that the earthquakes in an earthquake cluster are close to each other in space and time (hence they are close to the first earthquake in the cluster). Data analysis based on the HMM brings into question the undifferentiated treatment of background events with and without offspring. As shown in section 4, not only is the G-R law for single earthquakes distinct from that for the mother earthquakes (Figure 5), the distributions of their locations are also different (Figures 6 and 7).

[42] In essence, two of the ETAS's assumptions are relaxed in the HMM. First, where there is only one process (namely, the branching process) assumed for the whole earthquake occurrence sequence in the ETAS model, there are two independent processes (for single earthquakes and for earthquake clusters) in the HMM. This allows more detailed information about the stochastic process to be drawn as the intensity functions of these two processes are estimated separately. For example, Figures 6 and 7 show that some areas are prone to earthquake clusters while others are prone to single earthquakes. Such disparity in location is not likely to be uncovered by the ETAS model.

[43] Second, by loosely defining the cause of offspring occurrences as "the cluster is active," the HMM avoids

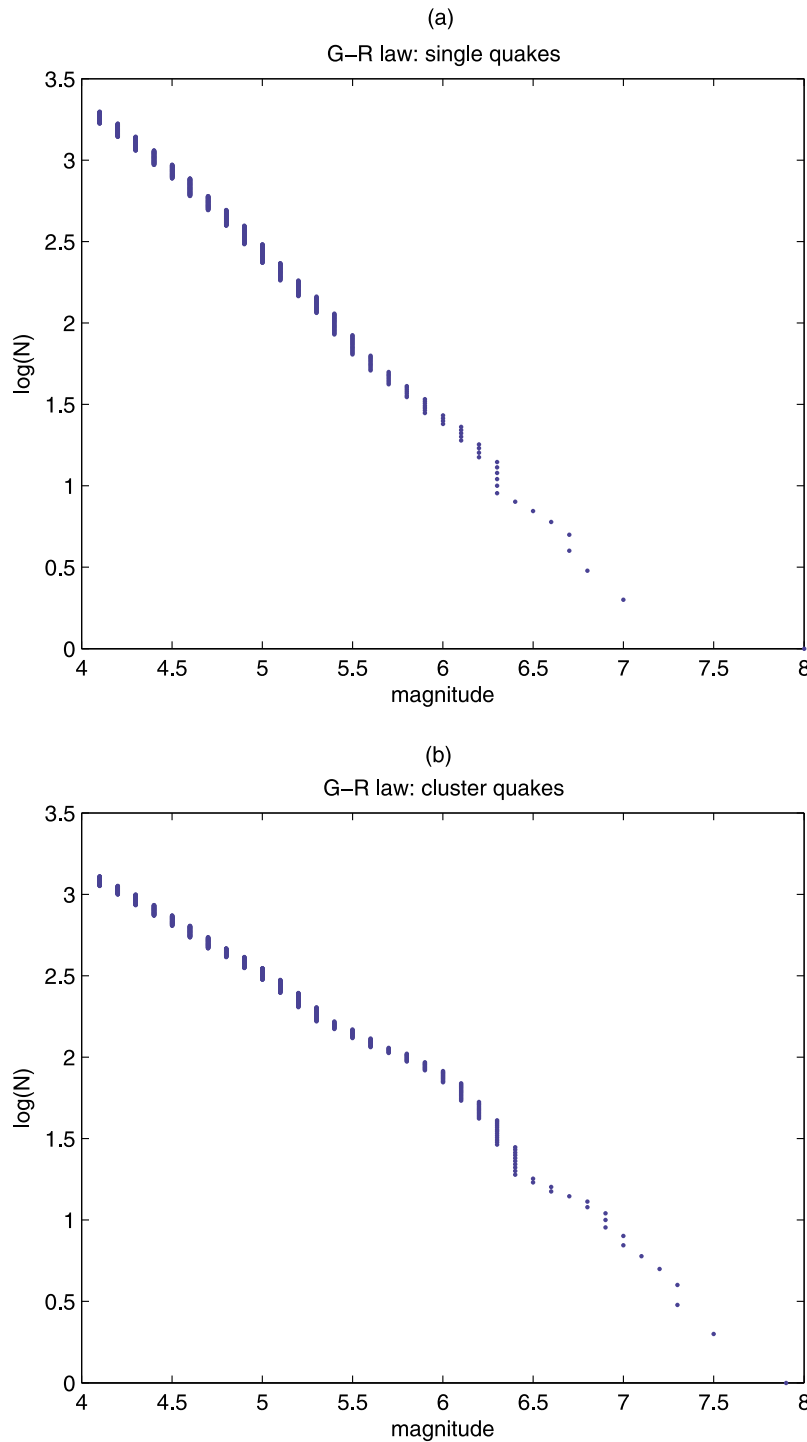


Figure 5. Plots of (a) $\log_{10} N_M$ versus M for single earthquakes and (b) $\log_{10} N_M$ versus M for cluster earthquakes.

specifying the ancestor(s) of an earthquake, whereas the ETAS model has a single, unique ancestor for each event and not several ancestors, which was noted as “an important defect” of the model of *Helmstetter and Sornette* [2002]. The HMM does not suffer from this defect as it focuses on the precedence rather than the causality of the events.

[44] In this article, a simple HMM is presented. It outdoes the ETAS model in three ways: it is adaptive to new ob-

servations so that it does a better job of declustering; it has a best guess of the hidden sequence and hence gives a unique partition; and the number of earthquakes in the model will not grow exponentially for any choice of model parameters.

[45] This model is only a first attempt to model earthquake occurrence using HMMs. It can be refined in many ways. For one thing, the assumption of no more than one active cluster at a time can be relaxed by letting the hidden

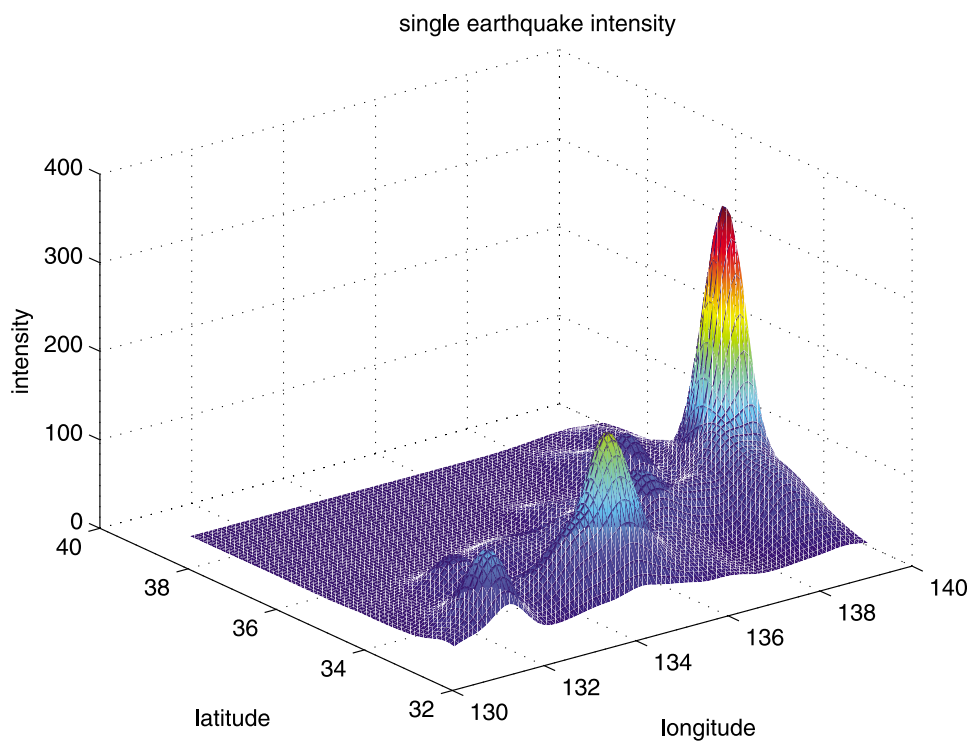


Figure 6. Unnormalized single earthquake intensity.

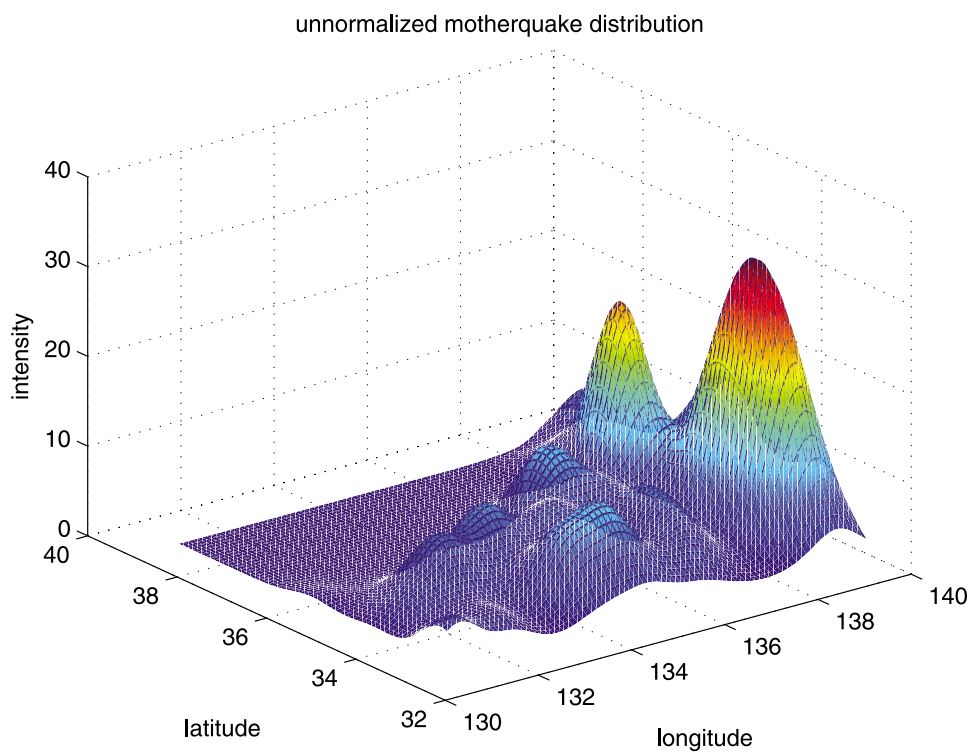


Figure 7. Unnormalized mother earthquake distribution.

state record more indicators. For another, magnitudes can be built into the model, accommodating the G-R law and the empirical Omori law. The HMM can be generalized to the extent that the ETAS is a special case of it.

[46] More importantly, other information (such as the location of fault lines), geophysical hypotheses and theories (about the underlying tectonics, say) can be incorporated and tested in the HMM framework. For instance, model selection based on empirical data is often a big concern for seismologists. Various model selection criteria have been developed in statistical literature [Burnham and Anderson, 2003]. Most of them involve the likelihood of the models, whose computation is routine for HMMs. Thus, HMMs can combine the power of scientific modeling and deep statistical theory. We believe that the HMM would serve as a suitable alternative to the ETAS model for earthquake modeling.

Appendix A: Dynamics of the Hidden Markov Chain

[47] Recall that the hidden state $h_i = (x_i, y_i, t_i, J_i, C_i, A_i)$, recording the longitude and the latitude of the epicenter, the occurrence time, the index of the most recent mother earthquake up to i , the indicator of whether or not the earthquake i is a cluster earthquake, and the indicator of whether or not a cluster is active.

[48] Five parameters are introduced in our model: γ is the intensity of the point process for single earthquakes, λ is the extra intensity when a cluster is active, ϵ is the intensity of the initiation of a new cluster, d is the variance parameter of the bivariate Gaussian distribution, p is the probability that the mother earthquake becomes sterile after giving birth to one more offspring. Note that the mother earthquake is born reproductive, which guarantees that each cluster contains at least two earthquakes. In the data set, the earthquakes occurred in the rectangular area $33^\circ\text{--}39^\circ\text{N}$ and $131^\circ\text{--}140^\circ\text{E}$. The area of the rectangle is 54 square degrees.

[49] To construct a Markov chain for h_1, h_2, \dots , it suffices to specify the conditional distribution of $p(h_{i+1}|h_i)$, which satisfies

$$p(h_{i+1}|h_i) = p(J_{i+1}, C_{i+1}, A_{i+1}|h_i)p(x_{i+1}, y_{i+1}, t_{i+1}|J_{i+1}, C_{i+1}, A_{i+1}, h_i).$$

The precise conditional distributions are listed in two cases and each case has different scenarios:

[50] 1. The first case: $A_i = 0$, i.e., no cluster is active at time t_i . This case has two scenarios. (1) Earthquake $i + 1$ is a mother quake; it starts a new active cluster; J_{i+1} is equal to $i + 1$.

$$p(J_{i+1} = i + 1, C_{i+1} = 1, A_{i+1} = 1|h_i) = \frac{\epsilon}{\epsilon + \gamma}$$

$$p(x_{i+1}, y_{i+1}, t_{i+1} | J_{i+1} = i + 1, C_{i+1} = 1, A_{i+1} = 1, h_i) = \frac{\mathbf{1}_{[131,140]}(x_{i+1})\mathbf{1}_{[33,39]}(y_{i+1})(\epsilon + \gamma) \exp\{-(t_{i+1} - t_i)(\epsilon + \gamma)\}}{54}$$

(2) Earthquake $i + 1$ is a single quake; still no active cluster at time t_{i+1} ; J_{i+1} keeps the same value as J_i .

$$p(J_{i+1} = J_i, C_{i+1} = 0, A_{i+1} = 0|h_i) = \frac{\gamma}{\epsilon + \gamma}$$

$$p(x_{i+1}, y_{i+1}, t_{i+1} | J_{i+1} = J_i, C_{i+1} = 0, A_{i+1} = 0, h_i) = \frac{\mathbf{1}_{[131,140]}(x_{i+1})\mathbf{1}_{[33,39]}(y_{i+1})(\epsilon + \gamma) \exp\{-(t_{i+1} - t_i)(\epsilon + \gamma)\}}{54}$$

[51] 2. The second case: $A_i = 1$, i.e., there is one cluster active at time t_i . This case has three scenarios. (1) Earthquake $i + 1$ is a single quake; the cluster is still active at time t_{i+1} ; J_{i+1} keeps the same value as J_i .

$$p(J_{i+1} = J_i, C_{i+1} = 0, A_{i+1} = 1|h_i) = \frac{\gamma}{\lambda + \epsilon + \gamma}$$

$$p(x_{i+1}, y_{i+1}, t_{i+1} | J_{i+1} = J_i, C_{i+1} = 0, A_{i+1} = 1, h_i) = \frac{\mathbf{1}_{[131,140]}(x_{i+1})\mathbf{1}_{[33,39]}(y_{i+1})(\lambda + \epsilon + \gamma) \exp\{-(t_{i+1} - t_i)(\lambda + \epsilon + \gamma)\}}{54}$$

(2) Earthquake $i + 1$ is a cluster quake (offspring); it does not sterilize its mother; J_{i+1} keeps the same value as J_i .

$$p(J_{i+1} = J_i, C_{i+1} = 1, A_{i+1} = 1|h_i) = \frac{(1 - p)(\lambda + \epsilon)}{\lambda + \epsilon + \gamma}$$

$$p(x_{i+1}, y_{i+1}, t_{i+1} | J_{i+1} = J_i, C_{i+1} = 1, A_{i+1} = 1, h_i) = \frac{\exp\{-\frac{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2}{2d}\}(\lambda + \epsilon + \gamma) \exp\{-(t_{i+1} - t_i)(\lambda + \epsilon + \gamma)\}}{2\pi d}$$

(2) Earthquake $i + 1$ is a cluster quake (offspring); it sterilizes its mother, J_{i+1} keeps the same value as J_i .

$$p(J_{i+1} = J_i, C_{i+1} = 1, A_{i+1} = 0|h_i) = \frac{p(\lambda + \epsilon)}{\lambda + \epsilon + \gamma}$$

$$p(x_{i+1}, y_{i+1}, t_{i+1} | J_{i+1} = J_i, C_{i+1} = 1, A_{i+1} = 0, h_i) = \frac{\exp\{-\frac{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2}{2d}\}(\lambda + \epsilon + \gamma) \exp\{-(t_{i+1} - t_i)(\lambda + \epsilon + \gamma)\}}{2\pi d}$$

[52] **Acknowledgments.** Jiancang Zhuang kindly sent me his numerical results, and I have benefited from the conversation with him. I also thank Tom Kurtz for his insightful comments and Ying Luo for her careful proofreading.

References

- Baum, L. E., T. Petrie, G. Soules, and N. Weiss (1970), A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains, *Ann. Math. Stat.*, 41, 164–171, doi:10.1214/aoms/1177697196.
- Burnham, K. P., and D. R. Anderson (2003), *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd ed., Springer, New York.

- Davis, S., and C. Frohlich (1991), Single-link cluster analysis, synthetic earthquake catalogs, and aftershock identification, *Geophys. J. Int.*, 104, 289–306, doi:10.1111/j.1365-246X.1991.tb02512.x.
- Ebel, J., D. Chambers, A. Kafka, and J. Baglivo (2007), Non-poissonian earthquake clustering and the hidden Markov model as bases for earthquake forecasting in California, *Seismol. Res. Lett.*, 78(1), 57–65, doi:10.1785/gssrl.78.1.57.
- Frohlich, C., and S. Davis (1990), Single-link cluster analysis as a method to evaluate spatial and temporal properties of earthquake catalogs, *Geophys. J. Int.*, 100(1), 19–32, doi:10.1111/j.1365-246X.1990.tb04564.x.
- Gardner, J., and L. Knopoff (1974), Is the sequence of earthquakes in southern California, with aftershock removed, poissonian?, *Bull. Seismol. Soc. Am.*, 64, 1363–1367.
- Granat, R., and A. Donnellan (2002), A hidden markov model based tool for geophysical data exploration, *Pure Appl. Geophys.*, 159(10), 2271–2283, doi:10.1007/s00024-002-8735-6.
- Helmstetter, A., and D. Sornette (2002), Subcritical and supercritical regimes in epidemic models of earthquake aftershocks, *J. Geophys. Res.*, 107(B10), 2237, doi:10.1029/2001JB001580.
- Hough, S. E., and L. M. Jones (1997), Aftershocks: Are they earthquakes or afterthoughts?, *Eos Trans. AGU*, 78(45), 505, doi:10.1029/97EO00306.
- Keilis-Borok, V. I., and V. Kossobokov (1986), Time of increased probability for the great earthquakes of the world, *Comput. Seismol.*, 19, 45–58.
- Lagarias, J., J. A. Reeds, M. H. Wright, and P. E. Wright (1998), Convergence properties of the Nelder-Mead simplex method in low dimensions, *SIAM J. Optim.*, 9(1), 112–147, doi:10.1137/S1052623496303470.
- Ogata, Y. (1998), Space-time point-process models for earthquake occurrences, *Ann. Inst. Stat. Math.*, 50, 379–402, doi:10.1023/A:1003403601725.
- Reasenber, P. (1985), Second-order moment of central California seismicity, *J. Geophys. Res.*, 90(B7), 5479–5495, doi:10.1029/JB090iB07p05479.
- Utsu, T. (1969), Aftershock and earthquake statistics (i): Some parameters which characterize an aftershock sequence and their interrelations, *J. Fac. Sci. Hokkaido Univ., Ser. VII*, 3, 129–195.
- Viterbi, A. J. (1967), Error bounds for convolutional codes and an asymptotically optimum decoding algorithm, *IEEE Trans. Inf. Theory*, 13, 260–269, doi:10.1109/TIT.1967.1054010.
- Wilcoxon, F. (1945), Individual comparisons by ranking methods, *Biometrics*, 1, 80–83, doi:10.2307/3001968.
- Wilk, M. B., and R. Gnanadesikan (1968), Probability plotting methods for the analysis of data, *Biometrika*, 55, 1–17.
- Wu, Z. (2007), A filtering approach to abnormal cluster identification, Ph.D. thesis, Univ. of Wis.-Madison, Madison.
- Wu, Z. (2009), A cluster identification framework illustrated by a filtering model for earthquake occurrences, *Bernoulli*, 15(2), 357–379, doi:10.3150/08-BEJ159.
- Xiong, J. (2008), *An Introduction to Stochastic Filtering Theory*, Oxford Univ. Press, Oxford, U. K.
- Zhuang, J., Y. Ogata, and D. Vere-Jones (2002), Stochastic declustering of space-time earthquake occurrences, *J. Am. Stat. Assoc.*, 97, 369–380, doi:10.1198/016214502760046925.
- Zhuang, J., Y. Ogata, and D. Vere-Jones (2004), Analyzing earthquake clustering features by using stochastic reconstruction, *J. Geophys. Res.*, 109, B05301, doi:10.1029/2003JB002879.
- Zhuang, J., C.-P. Chang, Y. Ogata, and Y.-I. Chen (2005), A study on the background and clustering seismicity in the Taiwan region by using a point process model, *J. Geophys. Res.*, 110, B05S18, doi:10.1029/2004JB003157.

Z. Wu, Department of Statistics and Applied Probability, National University of Singapore, 6 Science Dr. 2, 117546, Singapore. (stawz@nus.edu.sg)