A Statistical Approach for Assessing Seismic Transitions Associated with Fluid Injections

Submitted in partial fulfillment of the requirements for

the degree of

Doctor of Philosophy

in

Civil & Environmental Engineering

Pengyun Wang

B.S., Environmental Engineering, Sun Yat-sen University
M.S., Environmental Engineering, Carnegie Mellon University

Carnegie Mellon University
Pittsburgh, PA

December 2016

Acknowledgements

I want to express my first thanks to my advisors Professor Mitchell J. Small, Professor Matteo Pozzi and Professor William Harbert for their inspiring advice, kind guidance and constant support. I also owe a great gratitude to Professor Jacob Bielak for his commitment and time to serve on my dissertation committee. I thank Dustin Crandall, Martin Chapman, Ariel Conn, Ernest Linder, Timothy Murin, Hema Siriwardane, and Paul Vincent for insightful discussions. I also thank Sebastian Hainzl and Corinne Bachmann for technical instructions. This work is not possible without the support provided by: the National Energy Technology Laboratory's Regional University Alliance (NETL-RUA), a Carnegie Mellon University-College of Engineering Dean's Fellowship, the H. John Heinz Professorship of Environmental Engineering, and the Scott Institute Seed Grants from the Wilton E. Scott Institute for Energy Innovation at Carnegie Mellon University. My gratitude also goes to the staff of the CEE department, who provides their service to facilitate all aspects of my life here. In particular, I want to thank Andrea Rooney, Cornelia Moore, and Maxine Leffard for kindly offering help when it is needed. I want to acknowledge my colleagues and friends at Carnegie Mellon University for creating a good time for me during my study, including Avi Mersky, Corey Harper, Duygu Altintas, Hanqi Chen, Eric McGivney, Jingkun Gao, Nizette Consolazio, Xuechen Lei, Xin Wang, Lawrence Wong, Siyuan Liu, Zhou Ye, George Lederman, and Peng Gong. Finally and most importantly, I want to acknowledge my family for their forever support and love for me; especially my wife who has sacrificed herself tremendously by carrying most of the burden of nursing our son and providing me with emotional and material supports beyond measure. I simply cannot accomplish what I have accomplished without her.

Abstract

The wide application of fluid injection has caused a concern of the potential critical risk associated with induced seismicity. To help clarify the concern, this dissertation proposes a statistical approach for assessing seismic transitions associated with fluid injections by scientifically analyzing instrumental measures of seismic events. The assessment problem is challenging due to the uncertain effects of wastewater injections on regional seismicity, along with the limited availability of seismic and injection data. To overcome the challenge, three statistical methods are developed, with each being focused on a different aspect of the problem. Specifically, a statistical method is developed for early detection of induced seismicity, with the potential of allowing for site managers and regulators to act promptly and preparing communities for the increased seismic risk; the second method aims for addressing the further need of quantitatively assessing the transition of induced seismicity, which can reveal the underlying process of induced seismicity and provide data to support probabilistic seismic hazard analysis; and finally, the third method steps further to characterize the process of spatial distribution of induced seismicity, which accounts for spatial evolution of induced seismicity. All the proposed methods are built on the principles of Bayesian technique, which provides a flexible inference framework to incorporate domain expertise and data uncertainty. The effectiveness of the proposed methods is demonstrated using the earthquake dataset for the state of Oklahoma, which shows a promising result: the detection method is able to issue warning of induced seismicity well before the occurrence of severe consequences; the transition model provides a significantly better fit to the dataset than the classical model and sheds light on the underlying transition of induced seismicity in Oklahoma; and the spatio-temporal model provides a most comprehensive characterization of the dataset in terms of its spatial and temporal properties and is shown to have a much better short-term forecasting performance than the "naïve methods". The proposed methods can be used in combination as a decision-making support tool to identify areas with increasing levels of seismic risk in a quantitative manner, supporting a comprehensive assessment to decide which risk-mitigation strategy should be recommended.

Contents

List of Tables

List of Figures

Chapter 1 Overview

1. 1 Introduction

Underground fluid injection has long been in practice as a low-cost approach to dispose of fluid wastes and to facilitate mining and oil and gas production. According to Underground Injection Control (United States Environmental Protection Agency, N.D), there are six types of injection wells with each being responsible for injecting a different kind of fluid. In particular, Class II wells are responsible for injecting fluids (primarily brines) associated with oil and natural gas production. They fall into one of three categories, including Disposal, Enhanced recovery and Hydrocarbon storage wells. They have been widely applied across the state: approximately 180,000 Class II wells are in operation in the United States and, over 2 billion gallons of brine are injected in the United States every day. Most oil and gas injection wells are in Texas, California, Oklahoma and, Kansas.

As one consequence of the wide application of this technology, an increasing number of cases of induced seismicity have emerged and raised concerns among stakeholders, including local communities under exposure, operators and managers of disposal wells and scientists in the seismic-risk community. The concern of potentially triggering large and physically and/or psychologically damaging seismic events can be well justified by many well-documented cases of induced earthquakes associated with waste fluids injection, including at the Rocky Mountain Arsenal (RMA), Colorado, in the 1960s (Healy et al., 1968); Ashtabula, Ohio, in the 1980s (Seeber et al., 2004); and Paradox Valley, Colorado, in the 1990s (Ake et al., 2005), among many others. The Most recent cases can be found in the midcontinent region of the United States (Horton, 2012; Ellsworth, 2013; Frohlich et al., 2014; Rubinstein et al., 2014), especially in the state of Oklahoma, where much evidence has been proposed by geoscientists to link the 2011 Mw 5.7 earthquake to the fluid-injection activities in the region (Keranen et al., 2013).

It is worth noting that the proportion of injection wells that have shown to induce felt earthquakes is small (United States Geological Survey, N.D). A combination of many factors is necessary for injection to induce felt earthquakes, e.g. the injection rate and total volume injected; the presence of faults that are large enough to produce felt earthquakes; stresses that are large enough to produce earthquakes; and the presence of pathways for the fluid pressure to travel from the injection point to faults. However, the concern for the potential risk is considerable due to the potential critical consequence of induced earthquakes.

This concern is responsible for the intense research activity in the field of induced seismicity, with research efforts devoted to the study of related inducing mechanisms (Talwani and Acree, 1984; Hickman et al., 1995; Streit and Hillis, 2004; McClure and Horne, 2011; Goertz-Allmann and Wiemer, 2012), and to the development of models for risk monitoring and forecasting (Bachmann et al., 2011; Convertito et al., 2012; Brodsky and Lajoie, 2013; Llenos et al. 2013; Mena et al., 2013). These efforts are vital for providing data to decision-makings on site-selection, regulation and risk-management of injection wells. However, there is still a lack of efforts of quantifying the transition of induced seismicity (especially when the injection data is unavailable or the correlation between induced seismicity and a specific injection well is difficult to identify), which is important for providing data to probabilistic seismic risk analysis (PSHA). In this dissertation, the problem is addressed using a statistical approach.

1.2 Problem Statement

To determine whether the concern of induced seismicity can be justified, and to guide risk-management of wastewater injections, it is beneficial to clarify the risks of induced seismicity. One of the crucial steps is to detect and quantify the seismic transition of induced seismicity (i.e., when, how much and where induced seismicity is occurring). An early detection allows for prompt response from well regulators and operators, while the quantification is an important step for accurate seismic hazard assessment.

In this dissertation, a series of statistical methods are presented to provide a solution to the challenge by scientifically analyzing instrumental measures of seismic events, including magnitude, location and time of occurrence. Firstly, an early detection method is proposed as a simple and computationally less expensive tool to monitor an injection site for early signs of induced seismicity. If the model confirms a detection of induced seismicity, two progressively more sophisticated methods are proposed and applied to quantify the level of induced seismicity and its spatial evolution. The objectives and methodologies of the three models are briefly described in the following section.

1.3 Objectives and Methods

The main objective of the thesis is to provide a solution to the detection and quantification of induced seismicity, or when, where and to what extent induced seismicity occurs. Specifically, it will be achieved by the following three sub-objectives.

(1) **A statistical method for early detection of changes in seismic rate**. The objective of the detection method is to use instrumental measures of seismic activity to allow for empirical early detection of symptoms of change. The corresponding early warning is critical to allow site managers and regulators to act promptly, revising the injection activity and/or preparing communities for the increased seismic risk. The method adopts a statistical hypothesis testing procedure in which data from test period are tested against a baseline model. The test period is extended gradually to collect just sufficient evidence to reject the null hypothesis, which provides a basis for early detection. The effectiveness of the method is demonstrated using a dataset from the Oklahoma state.

(2) **A Bayesian approach for assessing seismic transitions associated with fluid injection**. The objective of this effort is to provide a more general methodology using a statistical model that considers not only the time and magnitude but also the form of seismic rate transitions. With subsequent linkage to information on ground motions, exposure, fragility and consequences, the developed method can also provide an initial decision-support tool to identify areas with increasing levels of induced events, updating

seismic hazard estimates (Petersen et al., 2015), and supporting a comprehensive assessment to decide which risk-mitigation strategy should be recommended (Bommer et al., 2015). The method adopts and modifies the Epidemic Type Aftershock Sequences (ETAS) model (Ogata, 1988), which provides a popular framework for statistically modelling seismic events. The quantitative assessment of seismic activities follows the paradigm of Bayesian modeling, in which the prior uncertainty of the model can be updated upon acquisition of new information. The performance of the model is investigated in an application to the Oklahoma dataset.

(3) **A Bayesian approach for assessing spatio-temporal evolution of seismic event rate**. This effort aims for providing a comprehensive solution to the challenge by assessing seismic transitions in both time and space. The developed method can be used to provide data for monitoring and periodically updating the regional seismic hazard under uncertain effects of fluid injection. It involves updating the spatial event rate based on the previous rate and the current observation. Due to the lack of analytical solutions, the inference of the model is carried out using Particle Filter method, in which a set of weighted samples/particles are periodically updated to represent our belief about the state of the event rate. The method is applied to the Oklahoma dataset to evaluate its performance.

Each method focuses on one or more aspects of the challenge (i.e. when, where and to what extent) and has its unique advantage, but provides a relatively more comprehensive solution than the previous methods. The relative advantage of each method is discussed in Chapter 6, along with the possibility of using them in combination to address the challenge.

1.4 Dissertation Layout

The first chapter serves as an introduction to the research problem and the specific objectives to be addressed. Chapter 2 describes the earthquake dataset of the Oklahoma state, which is used to demonstrate the performance of the proposed methods. Chapters 3 to 5 are used to describe each research objective, proposed method, and application results, respectively. The last chapter is devoted to summarizing the achievements of the dissertation and discussing the outlook for future research.

Chapter 2 The State of Oklahoma and Its Earthquake Dataset

Abstract

Facts are presented in this chapter regarding the injection history in the state of Oklahoma and the general trend of its earthquake activity. And also, an earthquake dataset (last assessed in April 2016) from the state is described in terms of temporal and spatial properties of seismic events, frequency and magnitude relationships, and magnitude of completeness. This dataset will be subsequently used to demonstrate the performance of each proposed method.

2.1 Injection and induced seismicity in the Oklahoma State

Oklahoma is a region that is well known for a long history of wastewater injection with numerous injection wells across the state and, has seen a remarkable increase in the regional seismic activity after decades of absence of induced earthquakes since the first injection wells were deployed (although it has been suggested that induced earthquakes have occurred in Oklahoma during these decades, but at lower frequency and magnitude (Hough, 2015)). The second largest earthquake M_L 5.7 in the history of the state struck its central region in 2011 and damaged nearby infrastructures. More recently, the largest earthquake M5.8 in the history of the state shook Pawnee on September 2016, which is shortly followed by a M5.0 event that struck near Cushing, Oklahoma and building damages have been reported.

The M_L5.7 earthquake sequence initiated very close to a pair of wastewater injection wells where disposal operations began almost 20 years earlier (Oklahoma Corporation Commission Well Data System). There has been some evidence to suggest that this earthquake was induced by nearby wastewater injection activities (Keranen et al., 2013).

Some of the geological and geophysical characteristics related to the recent earthquake activity within Oklahoma, documented by the Oklahoma Geological Survey, include: a) the seismicity rate in 2013 was 70 times greater than the background seismicity rate observed in Oklahoma prior to 2008; b) the majority of earthquakes in central and north-central Oklahoma occur as earthquake swarms and not in the typical foreshock-mainshock-aftershock sequences that are characteristic of naturally occurring earthquake sequences; c) these earthquake swarms are occurring over a large area, about 15% of the area of Oklahoma, that has experienced a significant increase in wastewater disposal volume over the last several years; d) most of the earthquakes are occurring within the crystalline basement, deeper than most oil and gas operations; and e) the majority of wastewater disposal is targeted for injection in the Arbuckle formations, which closely overlie the crystalline basement.

2.2 The Oklahoma earthquake dataset

The dataset roughly spans the time period from January 1975 to March 2016. The total number of recorded seismic events during this period is 20,595, with the largest being M_L 5.7 and the smallest M_L-1.2. The Entire-Magnitude-Range method (Woessner and Stefan, 2005) reveals that the magnitude completeness (M_c) of this dataset is M_L2.4 +/- 0.06 (uncertainties are calculated by bootstrapping). In this study, a more conservative value of M_L2.5 is used for M_c to mitigate the effect of missing occurring, but unrecorded seismic events below this threshold level and the b-value is set at 1.0, as commonly assumed in literature (Frohlich and Scott, 1993). As a consequence of the M_c selection, a total number of 7010 events, constituting a complete catalog, remain for subsequent analysis. The temporal and spatial properties of the catalog are illustrated in Figure 1.

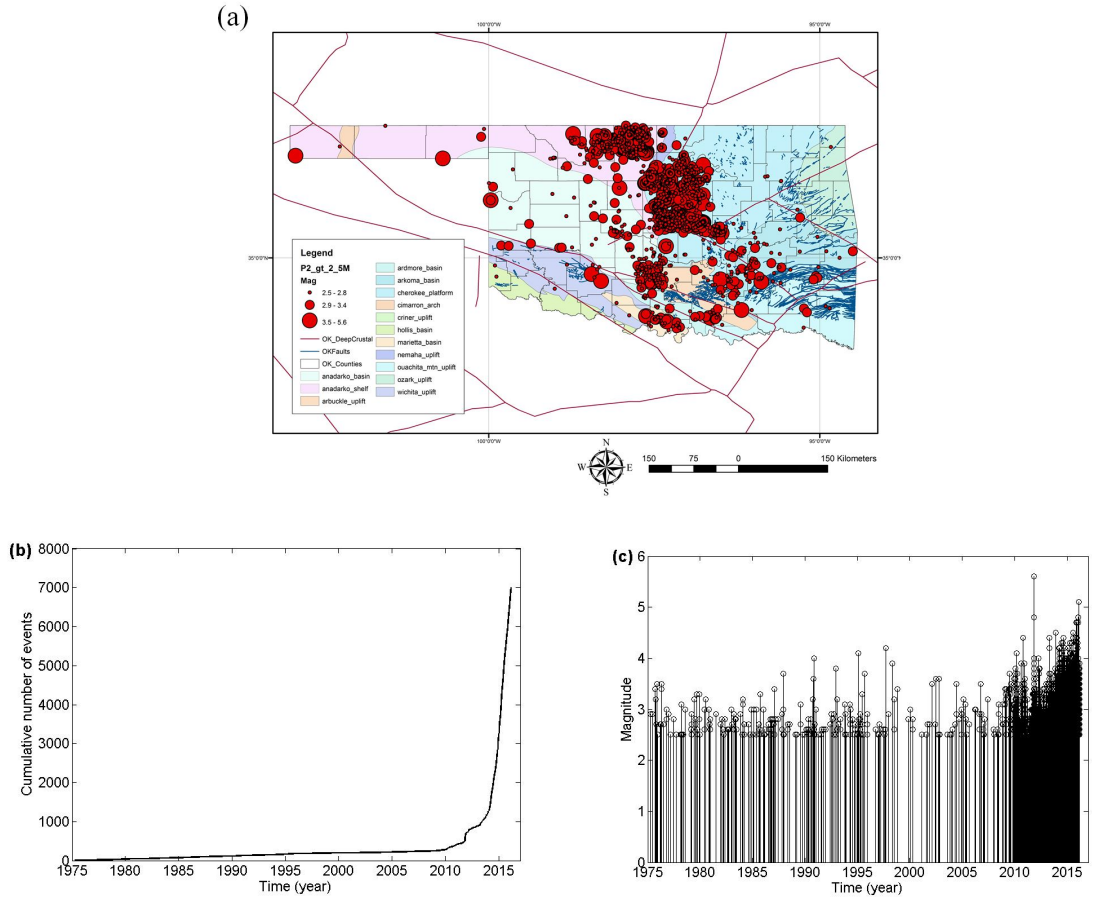

Figure 1 Spatial and temporal illustrations of the complete catalog (M ≥ 2.5). Plot (a) shows a map of Oklahoma (shaded section) and the epicenters and magnitude ranges of the earthquake events. Faults are taken from Heran et al., (2003) and Sims et al., (2008), geological provinces are taken from Northcutt and Campbell (1995); plot (b) graphs the cumulative behavior of the catalog time series and (c) displays the event occurrence times and associated magnitudes.

As shown in plot 1(a), two major areas of seismic events are the central and northern border parts of the Oklahoma state. Plot 1(b) shows that Oklahoma has undergone a substantial increase in the seismicity rate beginning around year 2010 or earlier, for events with local magnitude 2.5 and above.

This dataset are used to investigate the performance of each proposed method as documented in the following sections. Note that due to the development times of the earlier methods (i.e. method 1 and 2), their analysis does not incorporate the latest trend of the earthquake activity in the state, but the general conclusions would not be affected.

Chapter 3 A statistical approach for early detection of induced seismicity

Abstract

In this chapter, a statistical approach is proposed to detect increments in seismic rate, accounting for model uncertainty, which is particularly acute when the monitoring period is short, and interdependence among events. The approach is composed of two steps: first, stochastic earthquake declustering identifies main shocks and, second, the hypothesis of a constant rate of main shocks is statistically tested. The method is applied to the analysis of the Oklahoma injection region, demonstrating that it is able to detect an increment in seismic rate before the change is large enough to produce severe consequences. The statistical power of the method is investigated using synthetic data simulated for a wide range of scenarios.

3.1 Introduction

The objective of a seismic monitoring-detection system is to use instrumental measures of seismic activity to allow for empirical early detection of symptoms of change. The corresponding early warning is critical to allow site managers and regulators to act promptly, revising the injection activity and/or preparing communities for the increased seismic risk. Detecting changes is challenging because of: the possible long delay between the commencement of disposal operations and the onset of a change in seismicity (Kerenan et al., 2013); the aleatoric randomness affecting seismic productivity; and the scarcity of available seismicity data. Statistical modeling is needed to properly account for these effects.

To address this need, this Chapter proposes a hypothesis test for early detection of any abnormal increment in the rate of the seismic activity, as a proxy for the seismic state. The test compares the number of seismic events observed during a past baseline period,

unaffected by the fluid injection, with that during a test period that may be affected or not. Baseline seismic rate is inferred by the first datum, and the likelihood of change by comparison with the second one.

The analysis is complicated by the interdependence among the occurrence of events (Ogata, et al. 1998). To deal with this phenomenon, we propose a two-step procedure: i) to pre-process the occurrence data using a declustering algorithm, so that declustered events can be considered as main independent shocks, and ii) to adopt a Bayesian approach based on the Gamma/Poisson model to perform an hypothesis test on the sequence of main shocks. The outcome of the procedure is the probability (or p-value) for the null hypothesis that declustered rates in the baseline and testing period are the same (Gelman et al. 2003). To implement the method, the baseline period data are first analyzed to characterize the uncertainty in the Poisson rate of main events and the predictive distribution of the number of events in the test period given the baseline model. If the probability (i.e., the p-value) of obtaining as many events as are observed during the test period is too low under the baseline model, then the null hypothesis is rejected.

We demonstrate the effectiveness of the technique by analyzing the Oklahoma earthquake catalog, showing that after the detection of a critical rate shift, a reliable warning could have been sent well before the occurrence of the first large induced earthquake (i.e. the first event above M4) occurred. By a power analysis, we also investigate the effectiveness of the method in terms of the probability of detection within a specified time interval, depending on the magnitude and rapidity of the change in seismic rate induced by the injection.

The rest of the Chapter is organized as follows: firstly, we report the technical details of the declustering algorithm and the proposed hypothesis test; subsequently, the analysis of the Oklahoma dataset is reported. A power analysis tailored to the Oklahoma baseline condition is then illustrated, followed by conclusions and the identification of further research needs.

3.2 Methods

Event declustering and the stochastic algorithm. Following the ETAS model (Ogata, et al. 1998), we classify earthquakes in two categories: independent events, or main shocks; and dependent events, or aftershocks. Earthquakes are main shocks if their occurrence is due to tectonic loading and fluid intrusions, etc., unrelated to the occurrence of other events. In contrast, aftershocks are triggered by static or dynamic stress changes and/or seismically-activated fluid flows, and are at least partially related to previous earthquakes. The process of classifying events into these two categories is known as "earthquake declustering". We focus on main shocks only, whose rate we intend as the main feature for detecting symptoms of change due to fluid injection, because: i) earlier papers have linked changes in the ETAS main shock rate to changes in fluid flow (Hainzl and Ogata, 2005; Matsu'ura and Karakama, 2005; Lombardi et al., 2010; Daniel et al., 2011); and ii) although it has been hypothesized that fluid injection could affect the rate of aftershocks (Llenos et al. 2013), no model has been generally accepted for this phenomenon. Therefore declustering is a crucial step for our analysis, as it eliminates the fluctuations of the seismic rate due to interdependence.

There exist a few approaches for earthquake declustering, such as the deterministic methods proposed by Gardner and Knopoff (1974) and Reasenberg (1985). In this study, we adopt the stochastic algorithm proposed by Zhuang et al. (2002), which properly accounts for the uncertainty related to the inference procedure. The input to the algorithm is the complete earthquake catalogue, reporting magnitude and occurrence time for all events, while its outcome is a realization of the possible catalogue of main shocks only. The algorithm relies on models for the baseline activity and for the clustering structure. In our approach, these are both provided by the ETAS model, which models the non-declustered seismic events as a Poisson process whose rate Λ, as a function of time t, is:

$$\Lambda\,(t|\,H,\theta) = \lambda_0 + \sum_{t_i<t} \frac{K_i}{(t-t_i+c)^{p_{aft}}} \qquad (1)$$

where dataset $H=\{t_1, \dots, t_N; M_1, \dots, M_N\}$ includes the magnitude and occurrence time for all events in the catalogue (but note that the summation accounts for past events only), $\Theta: \{\lambda_0; K_0; \alpha_{ETAS}; c; p_{aft}\}$ is the set of ETAS model parameters, λ_0 is the background rate for main events; p_{aft} is the decay rate of aftershocks in the modified Omori law (T Utsu et al., 1995), c is a calibrating constant and parameter K_i is the productivity of the i-th main shock, given by:

$$K_i = K_0 \, e^{\alpha_{ETAS}(M_i - M_c)} \tag{2}$$

where K_0 and α_{ETAS} control the productivity of parent events, a larger magnitude M_i event (parent) is expected to generate more aftershocks (children); and Mc is the magnitude threshold for the inclusion of events in the catalogue.

According to the thinning theory, the probability φ_i that event i is an independent/background event is

$$\varphi_i = \frac{\lambda_0}{\Lambda(t_i | H, \theta)} \tag{3}$$

Therefore, the background earthquake sequence is realized by selecting each event i with probability φ_i. The outcome is intrinsically stochastic, as different sequences are generated from different runs of the algorithm with different initial random seed numbers.

In practical applications, ETAS parameters θ are not known, and we estimate their values by maximizing the following log-likelihood function [Ogata et al., 1998]:

$$\mathcal{L}(\theta) = \sum_{i=1}^{N} \log \Lambda(t_i|H, \theta) - \int_{t_0}^{t_f} \Lambda(t|H, \theta) dt \tag{4}$$

where t_0 and t_f indicate the start and end of the period of analysis. Optimization is performed numerically using a standard nonlinear technique (Fletcher and Powell, 1963).

Note that if a single ETAS model is estimated for both the baseline and test periods, an increment of seismicity in the test period can be masked by a thinning rate that is determined primarily by the baseline seismic record. To avoid this, we recommend estimating independent ETAS models for the training and the testing periods. To allow for further flexibility, the test period can be divided into sub-periods, independently estimating an ETAS model for each sub-period. However, each sub-period must be long enough to allow a reliable estimate of the corresponding model.

Considering that it might be difficult to robustly separate out temporal variations in background rate and triggered rate (Touati et al., 2011; Touati et al., 2014), an alternative approach in which triggering parameters were fixed at their estimates from the baseline period was also explored. This yielded only very small differences in detection times and performance. Since the productivity of earthquakes might also have changed from the base period to the testing period (Llelnos and Michael, 2013), allowing for separate estimation of the triggering parameters in each provides a more general and flexible approach.

Gamma-Poisson hierarchical model for hypothesis testing. After declustering, we assume that the occurrence of independent events follows a homogeneous Poisson process, until the cumulative effect of injected fluids crosses a certain critical threshold. After that, the rate can increase due to a change in the pore pressure around the active faults. To detect this change, we first establish a baseline condition for the seismicity rate (λ_{base}) by modeling the seismicity in the period believed to be under a normal regime (baseline period) by a Bayesian hierarchical Gamma-Poisson model. The number of seismic events y_{test} observed in the testing period is then compared via hypothesis testing to the expected distribution of the number of events if baseline conditions still prevailed. The distribution for this number of events is determined by a negative binomial predictive distribution with parameters that depend on the length of the baseline period,

the number of events in the baseline period (these also determine the posterior uncertainty distribution for the baseline rate), and the length of the test period.

The Gamma-Poisson model assumes that events occur as a Poisson process, but that the occurrence rate λ is uncertain, with its uncertainty represented by a gamma distribution f_y:

$$f_\lambda(\lambda) = \frac{1}{\Gamma(\alpha)\beta^\alpha}\lambda^{\alpha-1} e^{-\lambda/\beta}; \qquad \lambda \geq 0; \quad \alpha, \beta > 0 \tag{5}$$

The gamma is a particularly effective choice for modeling a Poisson rate, since it is the conjugate distribution to the Poisson, allowing computationally effective Bayesian updating. If y_{base} main shocks are observed over the baseline period of duration T_{base}, the posterior distribution of λ_{base} is also Gamma with updated parameters calculated as follows:

$$\begin{aligned} \alpha' &= \alpha^o + y_{base} \\ \beta' &= \frac{\beta^o}{1+\beta^o T_{base}} \end{aligned} \tag{6}$$

Prior parameters α^o and β^o can be assigned depending on the available information on the seismic rate. As can be noted from Eq.5, we can model a flat, informationless prior by assigning α^o=1 and $\beta^o \rightarrow \infty$. The posterior parameters are then computed as:

$$\begin{aligned} \alpha' &= y_{base} + 1 \\ \beta' &= \frac{1}{T_{base}} \end{aligned} \tag{7}$$

We assume declustered seismic events during the testing period follow a Poisson process with unknown rate λ_{test}. We assess whether or not the number of main shocks y_{test} observed during the test period of duration T_{test} is statistically consistent with the posterior distribution of λ_{base}. This calculation is facilitated by the fact that, if λ_{test} and

λ_{base} are equal, the predictive distribution for y_{test} is a negative binomial, with probability mass function:

$$P_Y(y_{test}|\alpha', \beta', T_{test}) = \frac{\left(\frac{1}{\beta'}\right)^{y_{test}} * (T_{test})^{\alpha'}}{\left(\frac{1}{\beta'} + T_{test}\right)^{[\alpha' + y_{test}]}} \cdot \frac{\Gamma(y_{test} + \alpha')}{\Gamma(\alpha') y_{test}!}; \quad y_{test} = 0,1,2,\cdots \qquad (8)$$

where α' and β' are given by Equation 7. The cumulative distribution function (cdf) of y_{test}, $F_Y(y_{test}|\alpha', \beta', T_{test})$, can be evaluated numerically; facilitated by the availability of internal functions for the negative binomial distribution in a number of statistical packages.

Formally, we define the null hypothesis H_0 and complementary hypothesis H_1 as follows:

$$\begin{aligned} H_o: &\quad \lambda_{test} \leq \lambda_{base} \\ H_1: &\quad \lambda_{test} > \lambda_{base} \end{aligned} \qquad (9)$$

According to the null hypothesis seismicity does not increase from the baseline to the test period, and we only reject H_o if the outcome y_{test} (or larger) is very unlikely to occur under it. The probability of this occurrence (y_{test} or larger) is referred to as the "p-value" of the test.

For any value of y_{test}, we can compute the corresponding p-value for the test as:

$$\text{p-value} = 1 - F_Y(y_{test} - 1|\alpha', \beta', T_{test}) \qquad (10)$$

To properly account for randomness in the declustering, a set of independent executions of the following steps are performed:

1. Decluster the earthquake record related to the baseline period and count the number of main events y_{base};
2. Compute the parameters of the posterior distribution of λ_{base}, according to Eq.7;
3. Independently decluster each sub-interval of the test period, and count the number of main events y_{test} for the entire test period;
4. Calculate the p-value of the test of hypothesis, according to Eq.10.

An empirical distribution function for the p-value is thereby determined across the set of random declustered samples. If the 95th percentile (or another selected quantile) of the p-value distribution is below a specified level of significance (e.g., below 5% or 1%), then we can reject the null hypothesis with high confidence, and infer that injection has affected an increase in the seismic rate.

This test follows the paradigm of computing posterior predictive p-values for Bayesian modeling (Gelman et al. 2003) and may be characterized as a hybrid Bayesian/Frequentist test. Purely frequentist approaches have been also formulated for tests of Poisson rates (Lehmann et al., 2006). We chose the former over the latter because we consider the hybrid approach more flexible, since it allows including any information external to the data in the prior distribution on λ_{base}. When an informationless prior is adopted, as in Eq.7, hybrid and frequentist tests are in close agreement.

It is worth noting that the proposed hypothesis-testing algorithm can also be applied to detection of decreases in seismic rate by changing the null and complementary hypotheses, and correspondingly the calculation of the p-value. This could also be potentially relevant for induced seismicity, in instances where injection has stopped, to help estimate when/how quickly the seismic rate starts to decay back to its background level.

3.3 Application to the Oklahoma Earthquake Dataset

The approach presented in the previous section is applied to the study of the earthquake catalog of Oklahoma. The study period of this application roughly spans from Jan. 1975 to April 2014. As mentioned in Chapter 2, the magnitude of completeness is M2.5, which might influence the analysis result in this study. Its effect is investigated and shown in Appendix A1.

Temporal delineation of the catalogue for hypothesis testing. As described in Statistical Methods, our proposed hypothesis testing technique requires a pre-activation period of normal seismicity to establish a baseline model, which is subsequently tested against newly observed data to infer whether the new period exhibits a significant change. In application to the earthquake sequence for the Oklahoma region, we divide the dataset into two periods, with the first one being the baseline period, initially chosen to span the interval from Jan. 1975 through Dec. 2008, with the remaining observations constituting the test period. The precise transition point from the baseline to the test period is not apparent by itself, but can be varied as part of a sensitivity analysis, as shown later in the paper. In order to establish a prospective detection of a shift in the seismic regime for the Oklahoma earthquake dataset, test periods of increasing duration are evaluated, with an incremental step of 2-months. Hence, starting from the beginning of 2009, test periods of 2 months, 4 months, 6 months and up to 64 months are considered, until the testing is terminated when either the longest test period reaches the end of the dataset or the evaluated p-value reaches below 1×10^{-10}. This threshold is chosen to represent a point at which overwhelming evidence is present to reject H_o, and further reductions in the p-value contribute no further insight or inference in this regard. A critical value for first rejecting H_o (and implementing some type of management response) might occur at typical critical p-values adopted in statistical practice, e.g., $\alpha = 0.05$ or 0.01.

If a significant increase in seismic rate has occurred, the p-value should decline as more evidence of a shift in seismic regime is uncovered by acquiring more data and lengthening the duration of the test period. Consistent with the incremental step of the test period, we divide the test period into separate sub-intervals of 2-months and decluster

them independently. We perform 1,000 parallel runs of the declustering algorithm, and compute the 95th percentile of the p-value. The faster the p-value reaches and then remains below a critical value (e.g., $\alpha = 0.01$), the more rapidly the shift in seismicity rate can be detected and confirmed.

Modeling Results. First, we elucidate the relation between the ETAS fitting and the Bayesian model for the rate λ of the declustered events. Specially, parameter λ_0 of the ETAS model defines the background independent shock rate, and can be related to λ. As an example, in Figure 2, we show how the posterior distribution of λ_{base} is consistent with the estimated value of λ_0 for the baseline period. Here λ_0 is estimated as 0.0147 #events/day. And then we obtain 1,000 posterior distributions of λ_{base}, one distribution per realization of the declustered sequence. Each of the 1000 distributions describe the uncertainty in λ_{base} based on its posterior gamma distribution (Equations 5 and 7). As indicated in Figure 2, the resulting distributions of the 10th, 50th and 90th percentiles of the posterior distributions of λ_{base} show that λ_0 is covered by the 80% confidence interval of λ_{base} with high probability.

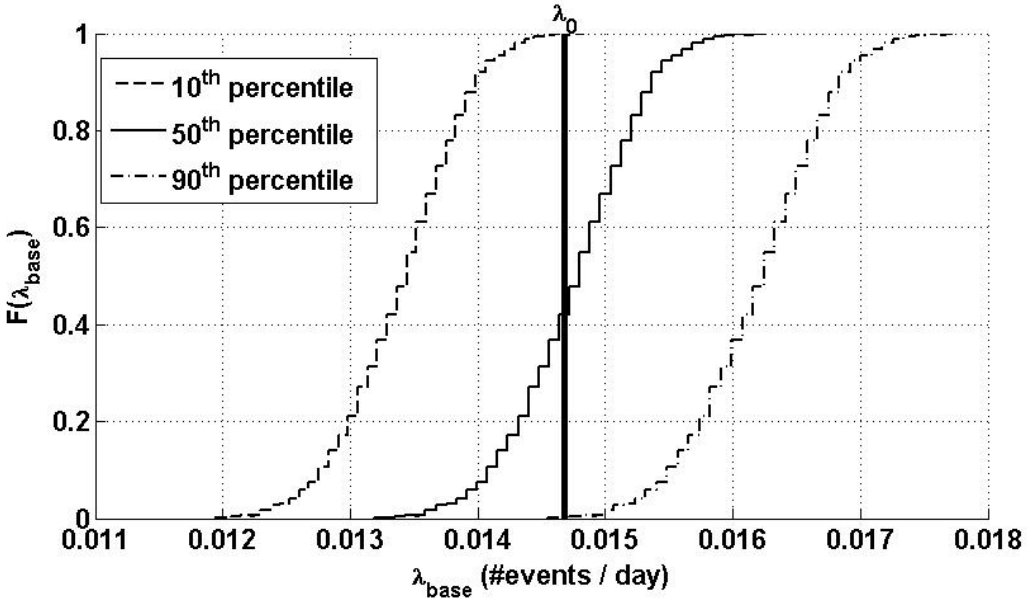

Figure 2 The comparison between 10th, 50th and 90th percentiles of each resulting posterior distribution of λ_{base}, and the MLE point estimate of the fixed λ_0 (0.0147 # events/day) in the declustering ETAS model as indicated by the vertical bold line in the plot.

To illustrate the effect of the time-dependent declustering algorithm, Figure 3 compares the sequence of original and declustered events. Sequences are represented by the bi-monthly rate, and the declusted rate is averaged along the 1,000 runs of the algorithm.

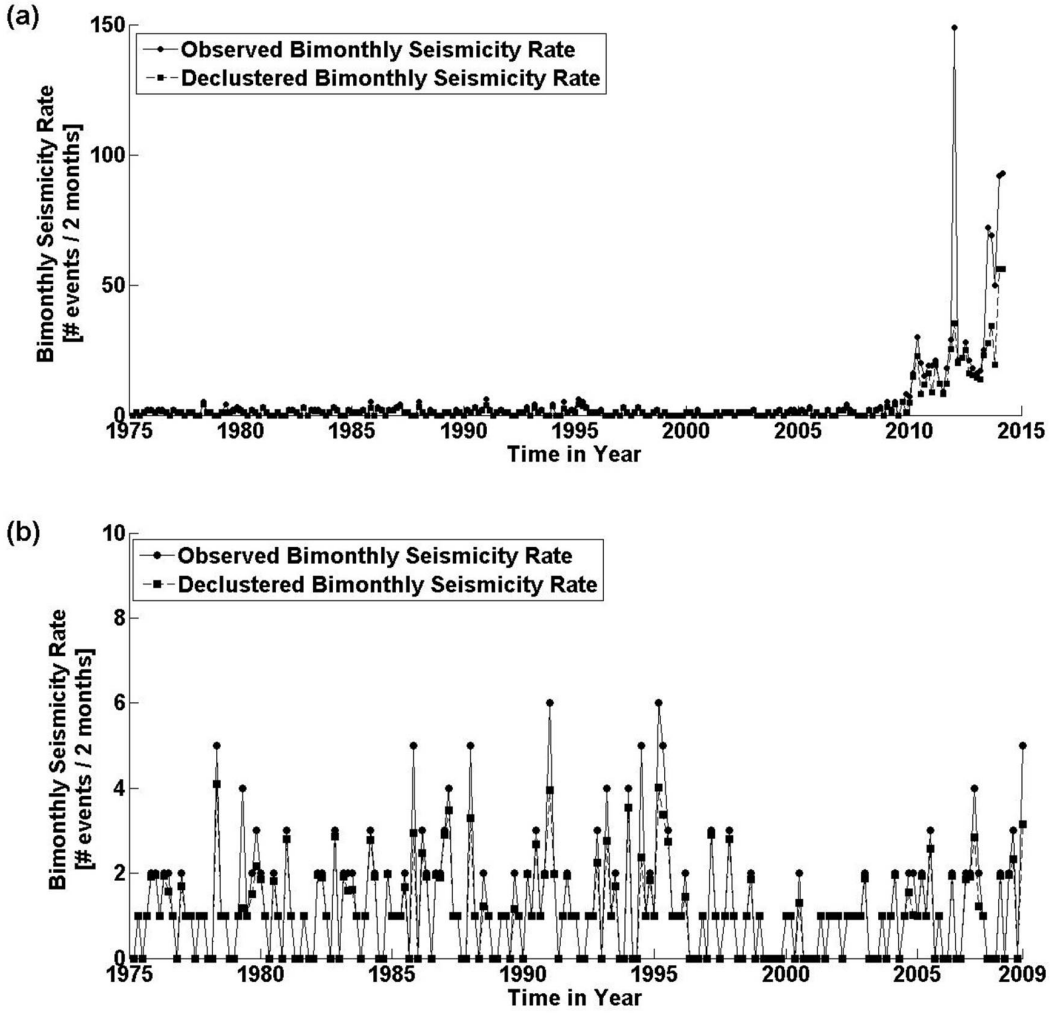

Figure 3 Oberved bimonthly seismicity rate versus the average bimonthly seismicity rate for 1000 samples of the declustered earthquake sequence for Oklahoma – (a) displays the evolution of seismicity for the entire studied period, while (b) shows that only for the period before 2009 but at a larger scale.

Figure 4 shows the corresponding temporal evolution of the p-value for the hypothesis test assuming a test period that ensues at the beginning of 2009, with increasingly longer subsequent testing periods. The 5^{th}, 50^{th} and 95^{th} percentile values of the p-value are computed from the 1000 tests conducted using the 1000 realizations of the ETAS declustering algorithm for the baseline and test periods. Generally speaking, the result is

consistent with our expectation that, for a location where a change in seismic rate has occurred, the p-value declines consistently with time. As such, by the end of August, 2009, when the testing period is 8 months long, the p-value becomes low enough to reject the baseline seismic rate with high confidence. This detection is achieved well before the occurrence of the first in the sequence of most-recent large earthquakes (M>4.0, indicated with stars in Figure 4) in March 2010. As shown in Figure 1c, three such events had occurred previously in the record, from 1990 through 1998. A model for the occurrence of (formerly) rare events requires specification of the distribution of event magnitudes as well as frequency. The occurrence of the March 2010 event, and the sequence of large events that have followed, could be reflective of transitions in both frequency and magnitude, though the latter is not addressed in this analysis. Furthermore, the specific timing of these events is influenced by the inherent randomness of the event processes, even as the events become more or less frequent.

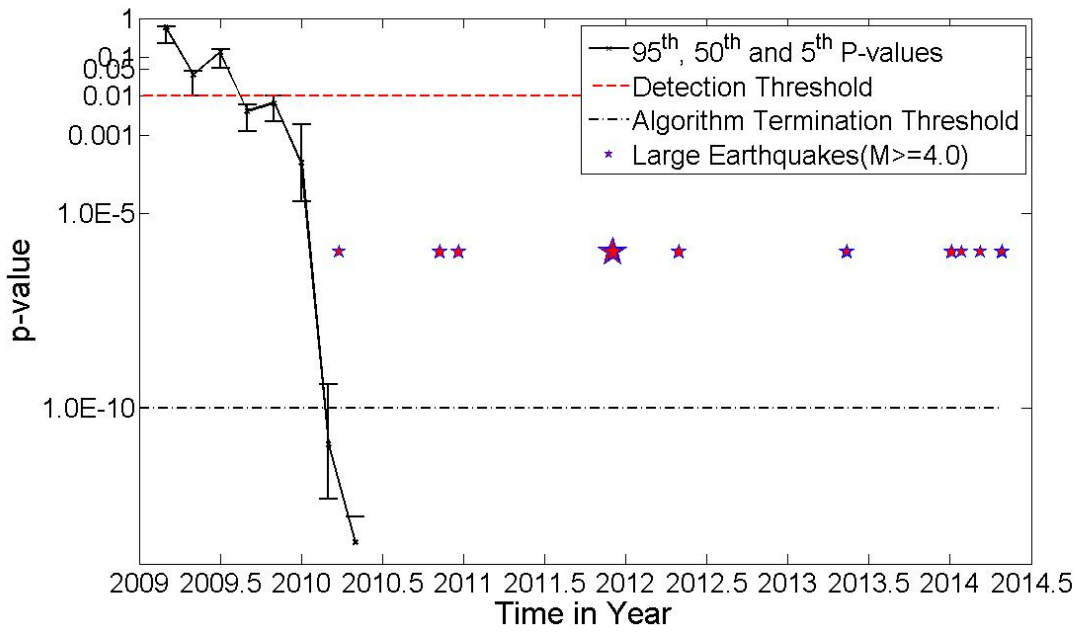

Figure 4 P-value evaluated based on the declustered seismic sequence at 2-month intervals with test origin at year 2009. At each point in time the algorithm is reiterated 1000 times, once for each realization of the ETAS algorithm for the baseline and test period, and the 5th, 50th and 95th percentiles are displayed. The dash line indicates the 0.01 significance level, while the dash dot line shows the threshold of 1×10^{-10} for the p-value, below which the hypothesis-testing algorithm terminates evaluation. In addition, the stars point out when large earthquakes (M>4.0) occurred.

The need for an 8-month period for the shift to be detected is determined in part by the significance level chosen for the test (here $\alpha = 0.01$), but also by the magnitude and rapidity of the shift in seismic event frequency and the length of the baseline period. These factors are considered in a general manner in the power analysis presented in the following section. However for the specific earthquake sequence in Oklahoma, it is first worth considering how the detection results are affected by choosing a different start date for the beginning of the test period.

To evaluate the effect of different transition dates from the baseline to the testing period, we move the assumed transition point from the beginning of the year 2009 to the beginning of 2000, 2005, 2007, or 2008, and repeat the analysis shown in Figure 4. In each scenario, the baseline period starts at the beginning of the catalogue and ends at the transition point to the test period. The evolution of the p-value for each test origin scenario is shown in the Appendix A2. The results are plotted together in Figure 5 to compare how quickly detection occurs in each case. Regardless of the choice for the origin of the test period, the p-value does not decline significantly until the year 2009 or beyond, indicating that the time of detection is relatively insensitive to our initial determination of the onset of the baseline period (Jan. 1975 to Dec. 2008). However, some differences are notable. In particular, initiating the test period at the beginning of 2008 or 2009 yields very similar results; in both cases detection is made before the end of 2009. However, selecting a start date for the test period that is too early can cause the more recent shift in seismicity to be diluted by the inclusion of a portion of the baseline period of record in the data for the test period. This is shown in Figure 5, where the selection of 2007 and 2005 as the test period initiation dates cause small, but increasing delays in the time of detection, while beginning the test period in 2000 (clearly well before the shift in seismic frequency) delays the detection significantly, until 2010.

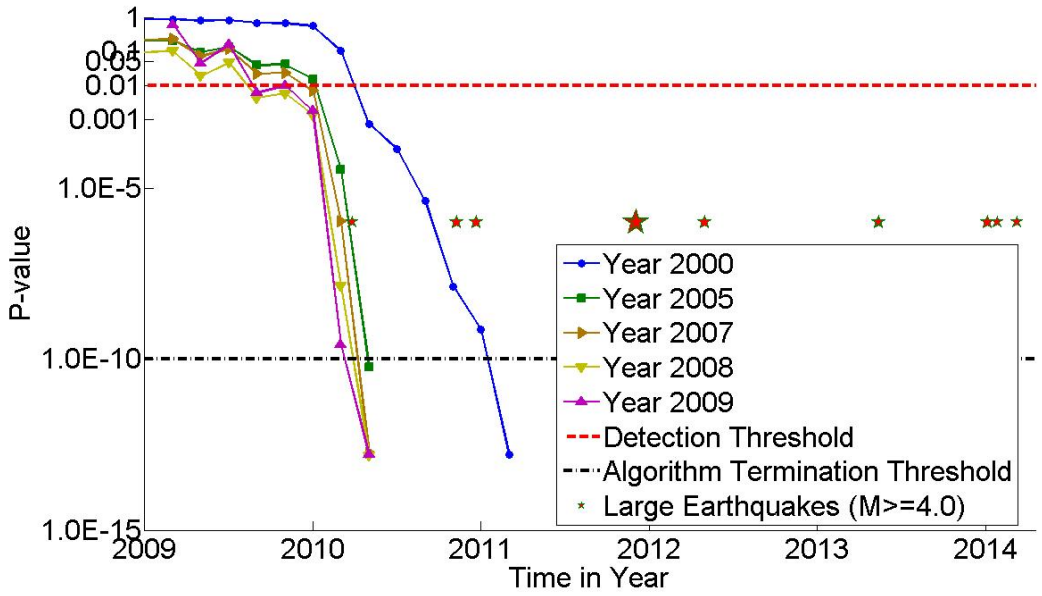

Figure 5 Evolution of the 95th percentile of the p-value under different time origins for test period. The dash line indicates the critical value of 0.01 for the p-value, while the dash dot line shows the threshold of 1×10^{-10} for the p-value, below which the hypothesis-testing algorithm terminates evaluation.

In practice, alternative transition dates may be considered for the statistical test of seismic rate. If one or more of these tests yields a statistically significant increase, it is likely that some significant increase in seismic rate has occurred in the time interval since the corresponding hypothesized transition, especially if the other transition times and tests yield results that are suggestive, e.g., with p-values between 0.01 and 0.2, even if not below the critical value for the test (in this case, 0.01). Since determination of an exact time of transition is not the objective of the method, this type of sensitivity analysis allows for consideration of the impact of transition time uncertainty on possible misallocation of portions of the measurement data to either the background or the test period, or the occurrence of gradual transitions.

3.4 Power Analysis

The power to positively detect the change in seismic rate is characterized by the probability within a specified time period that the null hypothesis in Equation 9 will be rejected, assumed here to occur when the 95th percentile of the p-value (across the multiple declustering outcomes) drops below a specified level of significance α. This probability is dependent on several factors, including:

a) The amount of information available to define the baseline seismicity rate for the baseline period – power increases with a longer T_{base}, which results in lower variances for the gamma distribution of λ_{base} and for the negative binomial distribution of y_{test};

b) The degree to which the seismicity rate is increased in the testing period due to the impact of fluid injection – bigger increases are easier to detect; and

c) The critical p-value chosen for the test of hypothesis – the lower the critical p-value (equivalent to the false positive rate for the test) the lower the power.

As such, the successful detection of a modest or a slight increase in seismicity, with a short period to define the baseline frequency, and with a high standard of proof (low critical p-value), is less likely within a specified time period.

To illustrate the factors influencing the power for detection, different scenarios are considered. For the baseline period, we investigate durations of 1, 10 and 20 years. For the magnitude of the change in the seismic rate, we assume injection increases the baseline rate by factors ranging from 2 to 10 during the test period, and we investigate different growth patterns, including an instantaneous increase and a gradual linear growth. For the former case we model the transition as a step function increasing from 1 to 2, 5, or 10 times the baseline seismic rate during the test period; for the latter, we model it as a ramp from 1 to 5 times the baseline seismic rate, over a duration of 1, 10 or 20 years (again, starting at the origin of the test period), followed by a constant value at 5 times the baseline rate.

Consistent with the procedure applied to the Oklahoma dataset, we use the ETAS model for declustering event sequences as a first step in the power analysis. The parameters

adopted for simulation of the synthetic event sequences for the baseline period are the maximum likelihood estimates (MLE) derived from the Oklahoma catalogue over the interval from 1975 to 2009, and are summarized in Table 1. Occurrence times of events are simulated iteratively: suppose t_A defines the time of the last event, time t_B of the next one is obtained by solving the following equation:

$$\int_{t_A}^{t_B} \Lambda\,(t|\,H,\theta)dt = -\log r \qquad (11)$$

where r is randomly generated from a uniform distribution between 0 and 1. For each event, its magnitude is randomly selected according to the Gutenberg-Richter law: in our study, we fix the b-value at 1.0, as is usually assumed in the literature (Frohlich and Scott, 1993).

λ_0 (Events/day)	K_0 (Events/Day)	α_{ETAS}	c (Days)	p_{aft}
0.0147	**0.012**	**0.8059**	**0.0030**	**0.9199**

Table 1 Parameter Values of the Epidemic Type Aftershock Sequence (ETAS) Model for the Simulation of Synthetic Data

As noted above, we simulate the influence of injection by changing the background rate λ_0, leaving the other parameters unchanged. Figure 6 depicts the assumed variation in the main event rate λ_0 as a function of time in the different scenarios explored in the power analysis.

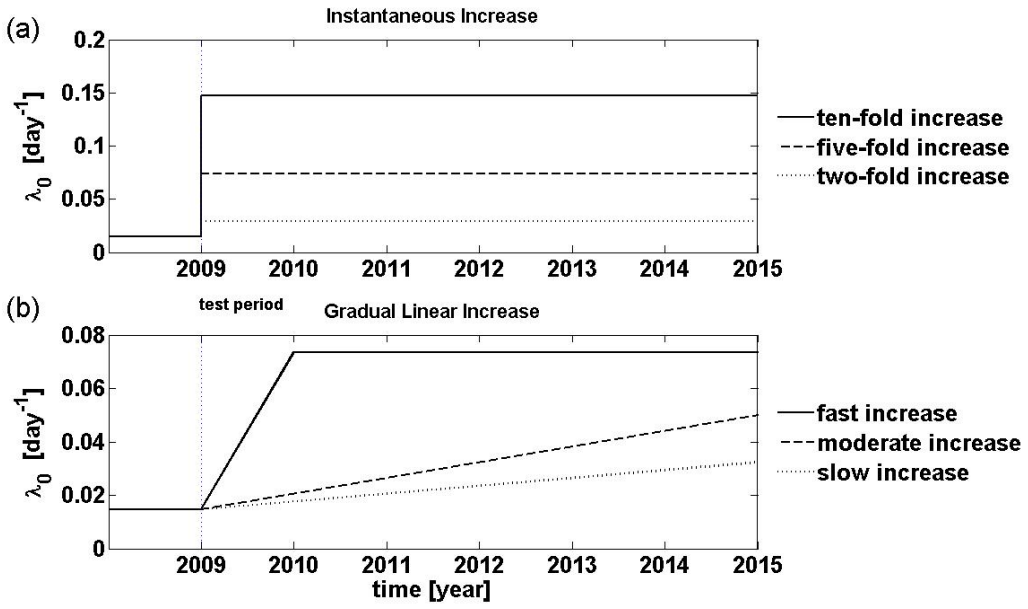

Figure 6 Illustrations of how λ_0 changes with time: a) for the instantaneous increase case, and b) for the gradual linear growth case. The beginning horizontal segment stands for the baseline period of a certain length.

To account for statistical uncertainties inherent in the power analysis, in each scenario 100 independent synthetic event catalogues are generated. We process each synthetic dataset in the same way as in analyzing the Oklahoma dataset (i.e., repeating the evaluation of the p-value every 2 months), and identify the corresponding time to detection, defined by the point at which the 95th percentile of the p-value becomes lower than 1%.

Figures 7 and 8 display the cumulative distribution functions of the time to detection for the instantaneous increase scenarios and the gradual linear increase scenarios, respectively. As shown, if we compare the distributions of time to detection resulting from different duration scenarios of the baseline period, with other modeling parameters held equal, a longer baseline period enables our algorithm to more likely accomplish a successful detection for a specified test period. In practice, this finding suggests that seismic monitoring should start as soon as possible in a region of fluid injection, to reduce uncertainty in the baseline seismicity rate, allowing changes in seismicity in the region to be detected more rapidly.

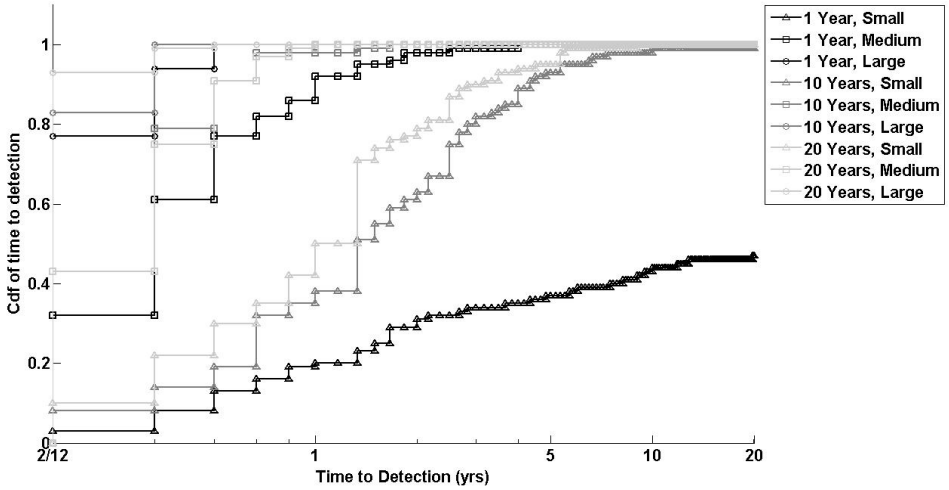

Figure 7 Cumulative distribution functions of time to detection (based on the 95th percentile of the p-value) resulting from the instantaneous increase scenarios. Each line style stands for a duration scenario of baseline period—1, 10 or 20 years. Subsequently, each type of marker represents a level of increase—2-, 5- or 10-fold the baseline seismicity rate.

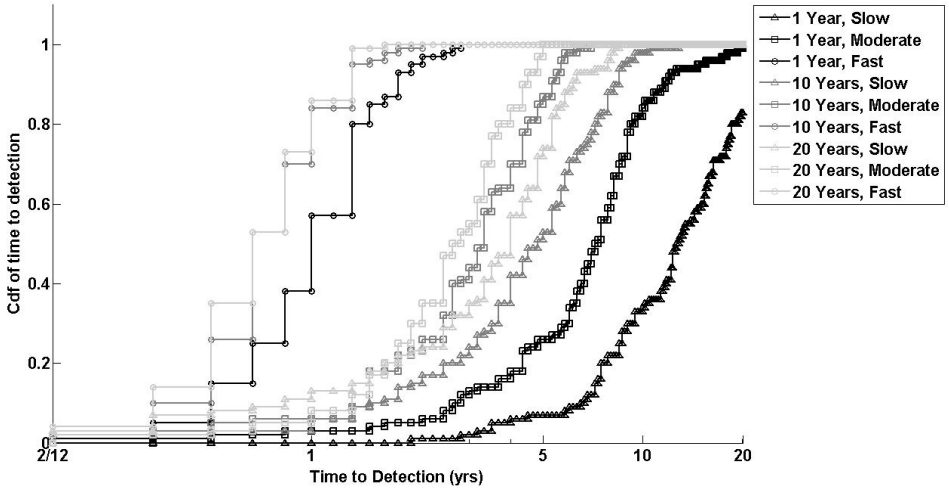

Figure 8 Cumulative distribution functions of time to detection (based on the 95th percentile of the p-value) resulting from the gradual linear increase scenarios. Each line style stands for a duration scenario of baseline period—1, 10 or 20 years. Subsequently, each type of marker represents a level of increase pace—1, 10 or 20 years to linearly reach the peak, e.g. 5-fold the baseline seismicity rate.

As expected, the degree and the pace of change in the seismicity rate play a major role in determining the power of the proposed detection method. If the increase is drastic and rapid as represented by the ten-fold increase scenario that occurs instantaneously, the change can be quickly confirmed by the statistical test, as shown: the probability is

around 0.8 for the change to be detected by the first 2-month test period. For the moderate increase scenario in Figure 7, the probability of detection is around 0.4 by the first test period, but it increases well above 0.5 when the test period is 4-months long, while for the small increase scenario (2-fold increase), the probability of detection is as low as 0.1 in the first test period, and does not rise above 0.5 until after one and a half years of monitoring when the baseline period is 10-20 years in length, and until after 10 years when a short (1-year) record is used to define the baseline condition.

As for the pace of change in seismicity, if we compare the ramp increase scenarios against the instantaneous increase scenarios (specifically, the 5-fold instantaneous increase scenarios), the probability of positive detection in the first test period drops below 0.1 for the fast increase scenarios in Figure 8, as opposed to the power of 0.4 that is generally achieved for the corresponding instantaneous case. For the moderate ramp scenarios, the change in seismic frequency can barely be detected in the first test period, and it requires over 3 years to be detected with a probability greater than 0.5 when the baseline period is 10 or 20 years in length, and requires approximately 10 years when the baseline period is limited (1-year). For the slow ramp scenarios (20 years to linearly reach the peak rate at 5-fold the baseline value), one year of monitoring cannot render a probability of detection better than 0.1, and it requires over 5 years for the probability to reach 0.5 when the baseline period is 10 or 20 years in length, and well over 10 years when only a short (1-year) record is available for the baseline period. Additionally, the comparison between the instantaneous increase scenarios and the ramp increase scenarios indirectly implies that if a 'normal period' is mistakenly included in the test period, the power of the detection method will be attenuated because the change that occurs during the true test period is diluted and masked by the inclusion of occurrence data from the normal period. To avoid this outcome, it is suggested that multiple tests be implemented at a given point in time with different time origins selected to initiate the algorithm to see if one or more yield a statistically positive signal.

3.5 Conclusions & Discussions

Detection of an increment in seismic activity due to deep-well wastewater disposal is of critical importance when assessing and managing risk and hazard. In this paper we propose a detection method based on earthquake declustering and statistical hypothesis testing. The method properly accounts for interdependency in the events, uncertainty in the independent shock rates, variability in the realized number of independent events, and the option to choose alternative hypothesis test specifications, including the critical level of significance used for the test. Application to the earthquake catalogue of the Oklahoma region shows that the proposed method is able to reliably detect a change before stronger evidence of an actual change (including the occurrence of a large earthquake) occurred. By simulating synthetic event sequences for different scenarios of change, we show how the power of detection depends on several factors, including the amount of historical data available for constraining and inferring the baseline seismicity rate, the extent and pattern of the change induced by injection, and the significance level chosen for the statistical test.

We believe that our proposed statistical method for seismic change detection can provide a preliminary decision-support tool to guide management and operations of disposal wells by providing information at an early stage. After such detections, more extensive seismic studies can be conducted to corroborate whether a change has in fact occurred, pinpoint locations of increased pore pressure and fault enhancement, and better assess future seismic risks for different mitigation strategies that might be adopted. More generally, as the detection method does not rely on injection data, it can also be applied to cases where the background rates may have changed due to natural fluid flow, dike intrusions, slow slip events, and other (natural) driving processes.

For future work, we are also interested in using the proposed method to help design an array of seismic stations in conjunction with a value of Mc (estimated with estimated Maximum Likelihood b-value) able to satisfy a pre-determined Time to Detection (yrs.) with a given probability for each criteria level of increased seismicity.

Chapter 4 A Bayesian approach for assessing seismic transitions associated with wastewater injections

Abstract

In this chapter, we develop a statistical method for modeling a seismic sequence involving non-stationary induced seismicity. It is composed of two steps: first, we select a model for the integrated seismicity (i.e., natural and induced) within the framework of the Epidemic Type Aftershock Sequence via Bayesian model comparison. Second, we perform Bayesian inference within that model, to assess the seismic activity and associated parameters.

The method is applied to the analysis of the events from Oklahoma, demonstrating that it is able to provide a consistent representation of the occurrence of the dataset. Results show that the overall seismic rate (including main and aftershock events) for events with local magnitudes (M_L) above 2.5 has been escalated by a factor of over 100, from 0.05 to over 5 events per day, between January 1975 and August 2014. For this overall increase, the contribution of the main events is estimated to be approximately 56%. Assuming the b-value of the Gutenberg-Richter law is 1.0, the probability of exceeding M_L 5.0 in a two-month period is predicted to have increased from about 0.05 to over 0.5 during the study period. A sensitivity analysis is presented to show how the probabilistic inference is affected by the assumed b-value and the assumed maximum event magnitude.

4.1 Introduction

To promote sustainable and safe management of wastewater injection, it is necessary to clarify the hazards caused by the combination of induced and natural seismicity. This paper provides a general methodology using a statistical model that considers not only the time and magnitude but also the form of seismic rate transitions, and associated event probabilities. It is worth mentioning that this methodology can also be employed to

model seismicity driven by other nonstationary processes, such as aseismic creep, magmatic intrusions, and natural fluid intrusion.

To develop the methodology, we adopt and modify the Epidemic Type Aftershock Sequences (ETAS) model (Ogata, 1988), which accounts for both main (independent) and aftershock (dependent) events. We consider aftershocks in addition to main events in our analysis for two reasons: (i) due to the concern that they can cause significant damages, and (ii) for a more complete interpretation of the collected data when estimating seismic rates and their transitions.

The ETAS model has been well established in statistical seismology; however, Llenos and Michael (2013) demonstrated the inadequacy of a stationary ETAS model for fitting an earthquake sequence involving induced seismicity (i.e., the catalog for the state of Oklahoma) via various statistical tests, and concluded that the rate of main earthquakes must have increased in 2009 to explain the more recent seismic history. In order to capture the non-stationarity due to induced seismicity, we modify the ETAS model by parametrically encoding induced seismicity into the model, enabling it to learn both the frequency of natural earthquakes and the timing and magnitude of transitions in the rate of induced events. The resultant model is called non-stationary ETAS model. Making use of the Gutenberg-Richter law (GR law) (Gutenberg and Richter, 1944), the inferred intensity of natural and induced earthquakes combined can then be related to a hazard projection, as the probability of an event exceeding a given magnitude threshold in a specified time window.

It is worth noting that various non-stationary ETAS models have been the subjects of many previous research efforts studying temporal characteristics of seismicity patterns. For example, they have been employed to detect and monitor aseismic forcing (such as fluid signals) in seismicity data (Hainzl and Ogata, 2005; Marsan and Helmstetter, 2013) by focusing on the variation of main event rates. In another example, Matsu'ura and Karakama (2005) investigated the effect of water on earthquake occurrence by statistically comparing alternative functional forms for the main event rate. In a third

example, Kumazawa and Ogata (2013) assumed a non-stationary process to quantitatively describe remotely triggered seismic activities. Not only have non-stationary ETAS models been used to study natural earthquake occurrences, they have also been used to analyze and monitor induced earthquakes at geothermal fields, based on the correlation between event occurrence rate and operational parameters (Bachmann et al., 2011; Brodsky and Lajoie, 2013; Mena et al., 2013); and to study temporal features of earthquakes induced by water injection in a gas field (Lei et al., 2008). Although all the existing ETAS models have the potential for modeling non-stationary main event rates associated with wastewater injection, the methodology proposed in this paper offers several advantages, as discussed in the "Discussions and conclusions" section.

The quantitative assessment of seismic activities follows the paradigm of Bayesian modeling, in which the prior uncertainty of the non-stationary ETAS model can be updated upon acquisition of new information, and posterior credible intervals for seismic occurrence rates can then be identified. Such a structure is flexible, particularly since it allows processing instrumental seismic records and the incorporation of expert geophysical knowledge, helping to constrain the prior parameter ranges of random variables involved in the inference process.

We investigate the performance of the non-stationary ETAS model in an application to the earthquake catalog for the state of Oklahoma, a region that is well known for a long history of wastewater injection and has seen a remarkable increase in the regional seismic activity after decades of absence of induced earthquakes since the first injection wells were deployed (although it has been suggested that induced earthquakes have occurred in Oklahoma during these decades, but at lower frequency and magnitude (Hough, 2015)). In analyzing this dataset, we compare parametric functions for modeling the effects of induced seismicity, including step, linear ramp, and logistic functions, through Bayesian modeling comparison (MacKay, 2003). As shown in a later section, the model with three superimposed logistic functions provides the optimal fitting to the dataset, and subsequently is selected to reconstruct the historical transitions in seismic events in Oklahoma.

The rest of this chapter is organized as follows: firstly, we describe in detail the proposed integrated model and the Bayesian procedure for inferring its parameters; subsequently, the Oklahoma earthquake catalog data used for demonstration are presented and the corresponding analysis of seismic transition is reported. A sensitivity analysis is then conducted for the probability of events of different magnitude occurring in different time intervals, exploring the effects of selected parameter values, followed by further discussion and conclusions.

4.2 Statistical models and methods

Non-stationary ETAS model. The general ETAS model has been presented by Eq. 1-2 in Chapter 3. It has a wide application in statistical seismology. In the context of induced seismicity, however, the cumulative effects of underground wastewater disposal can lead to non-stationarity, so that λ_0 can vary with time. Note that λ_0 now includes both background and induced main events. To model this behavior, we propose to represent λ_0 as a sum of the natural main-event rate μ_0 and an induced main-event time function rate $\Delta\mu_0(t)$. Instead of imposing a fixed parametric form for $\Delta\mu_0(t)$, we assume a set of possible alternative functions, so that the appropriate model can be selected through data processing, including a jump function, a linear ramp, and a logistic function. More generally, $\Delta\mu_0(t)$ can be assumed to be the sum of several components, each modeling the activation of induced seismicity in one area, at a specific time, in a large region with numerous operating injection wells. To illustrate one possible form of λ_0, assume that $\Delta\mu_0(t)$ includes one logistic function:

$$\lambda_0(t) = \mu_o + \frac{\Delta}{1+e^{-\eta(t-t_o-t_d)}} \tag{12}$$

where t_o stands for the start of the study period and t_d is a model parameter indicating the delay between t_o and the time when induced seismicity reaches half of its full intensity Δ; η controls the slope of the logistic curve – the bigger the η, the faster the induced seismicity increases at $t_o + t_d$. According to Eq. 3, λ_0 is initially dominated by μ_0,

transitions upward at an increasing rate until time $t_o + t_d$, and then asymptotically approaches a value of $\mu_0 + \Delta$ as t increases further.

Table 2 lists the alternative model classes investigated in this study, including the stationary (i.e., with no induced seismicity) ETAS model E_0. Models $E_{1\text{-}3}$ assume one increase of various forms in seismicity, while models $E_{4\text{-}6}$ allow for more than one logistic increase to occur in the dataset. For the models with more than one increase, parameter t_{di} is defined as the time delay between: t_o and the 1^{st} increase in λ_0 for i=1; the $(i\text{-}1)^{th}$ and i^{th} increase for i>1.

Name	Assumptions	Formula
E_0	No induced seismicity	$\lambda_0(t) = \mu_o$
E_1	Step jump in seismicity	$\lambda_0(t) = \mu_o + \Delta * I(t \geq t_o + t_d)$
E_2	Linear ramp in seismicity	$\lambda_0(t) = \mu_o + \min(\Delta, \eta(t - t_d) * I(t \geq t_o + t_d))$
E_3	1 logistic increase in seismicity	$\lambda_0(t) = \mu_o + \sum_{i=1}^{n_l=1} \frac{\Delta_i}{1 + e^{-\eta_i(t - t_o - \sum_{j=1}^{i} t_{dj})}}$
E_4	2 logistic increases in seismicity	$\lambda_0(t) = \mu_o + \sum_{i=1}^{n_l=2} \frac{\Delta_i}{1 + e^{-\eta_i(t - t_o - \sum_{j=1}^{i} t_{dj})}}$
E_5	3 logistic increases in seismicity	$\lambda_0(t) = \mu_o + \sum_{i=1}^{n_l=3} \frac{\Delta_i}{1 + e^{-\eta_i(t - t_o - \sum_{j=1}^{i} t_{dj})}}$
E_6	4 logistic increases in seismicity	$\lambda_0(t) = \mu_o + \sum_{i=1}^{n_l=4} \frac{\Delta_i}{1 + e^{-\eta_i(t - t_o - \sum_{j=1}^{i} t_{dj})}}$

Table 2 Summary and designation of all the alternative models investigated in this study with respect to λ_0. E_0 is the stationary (i.e., no induced seismicity) ETAS model. Models $E_{1\text{-}3}$ assume one increase of various forms in seismicity, while models $E_{4\text{-}6}$ allow for more than one logistic increase to occur in dataset. *I* stands for the indication function. For the models with more than one increase, parameter t_{di} is defined as the time delay: between t_o and the 1st increase in λ_0 for i=1; between the (i-1)th and ith increase for i>1.

The selection of the best form to model induced seismicity should be based on the specific earthquake data under study and, in principle, simple models are preferred over complex models unless the complexity can be justified by the likelihood of the data. A detailed method for model selection is provided later in the section on "Bayesian model comparison".

The fit of an inferred ETAS model can be assessed by first converting the occurrence times t_i to transformed times τ_i using the following theoretical cumulative function (Ogata, 1988):

$$\tau_i = \int_0^{t_i} \Lambda(t)dt. \quad (13)$$

The transformed times τ_i thus represent the mean number of events that the model predicts should occur in the time interval $[0, t_i]$. If the model fits the data well, then the events in the transformed time behave as a homogenous Poisson process with unit rate. Therefore, the observed cumulative number of events should grow linearly with unit slope as a function of the transformed times. Positive or negative deviations from the unit-slope line indicate that the model under- or over- predicts, respectively, the seismic rate for the corresponding intervals of the study period.

If change in the main event rate, $\Delta\mu_0(t)$, represents the direct effect of external forces on induced seismicity; and $\Lambda(t)$ represents the total event rate, including total main event rate and aftershocks of all previous earthquakes, then the direct contribution of external forces on seismicity transitions from time t_0 to t_f is given by

$$\frac{n_{\Delta\mu_0}}{n_\lambda} = \frac{\int_{t_0}^{t_f} \Delta\mu_0(t)dt}{\int_{t_0}^{t_f} \Lambda(t)dt} \quad (14)$$

Bayesian inference for the non-stationary ETAS model. Inference for the ETAS model consists of updating the assumed prior parameter distribution to obtain the posterior distribution based on observations, using Bayes' formula:

$$p(\boldsymbol{\theta}|H) \propto \pi(\boldsymbol{\theta})p(H|\boldsymbol{\theta}) \quad (15)$$

where dataset $H=\{t_1, \dots, t_N; m_1, \dots, m_N\}$ includes the magnitude and occurrence time for all events in the sequence, $\boldsymbol{\theta}=\{\mu_0, K_0, \alpha_{ETAS}, c, p, \Delta, \eta, t_d\}$ is the set of model parameters, $\pi(\boldsymbol{\theta})$ is the prior distribution and $p(H|\boldsymbol{\theta})$ is the likelihood for observation H given $\boldsymbol{\theta}$, which considers all of the N events in the time interval $[t_o, t_f]$ and takes the form of Eq. 4.

As all the parameters are non-negative, we assume the prior of each parameter θ_i in $\boldsymbol{\theta}$ is an independent lognormal distribution:

$$\theta_i \sim \ln\mathcal{N}\ (a_i, b_i), \qquad i = 1, \dots, n_\theta \quad (16)$$

where n_θ is the number of parameters in $\boldsymbol{\theta}$; and hyperparameters a_i and b_i can be selected according to ETAS parameter estimates across different global tectonic zones (Chu et al., 2011), or with flatter distributions with higher uncertainty to allow for unusual behavior at the targeted site; to be subsequently updated by the observed data in the calculation of the posterior $p(\boldsymbol{\theta}|H)$. Note that for more general applications, the value of parameter Δ can also be negative by choosing a proper distribution.

Since the posterior in Eq. 15 has a complicated form and cannot be solved analytically for this application, we numerically approximate that distribution using the Markov Chain Monte Carlo (MCMC) method (Rasmussen, 2013). Specifically, the Metropolis-Hastings algorithm (Chib and Greenberg, 1995) is used to simulate the parameter chain: in each sampling step, a new proposal $\tilde{\boldsymbol{\theta}}$ is drawn based on a random walk on a transformed standard normal distribution, $\boldsymbol{U} = \frac{\log(\boldsymbol{\theta}) - \boldsymbol{a}}{\boldsymbol{b}}$, and then the acceptance ratio AR is calculated based on Eq. 7 :

$$AR = \frac{\pi(\tilde{\boldsymbol{\theta}})}{\pi(\boldsymbol{\theta})} \frac{p(H|\tilde{\boldsymbol{\theta}})}{p(H|\boldsymbol{\theta})} \quad (17)$$

The new proposal $\widetilde{\boldsymbol{\theta}}$ is accepted if AR ≥ r, r being uniformly drawn from [0, 1], and rejected if otherwise. The sampling process is repeated until the chain reaches a sufficient length, with stable statistical properties for the fitted joint distribution of the model parameters.

Bayesian model comparison. Each form of the function for $\lambda_0(t)$ defines a parametric model class for the non-stationary ETAS model. In the Bayesian paradigm, model classes can be compared probabilistically. As in Eq. 15, we calculate the posterior probability of each alternative model E_i, using Bayes' formula (MacKay, 2003):

$$P(E_i|H) = \frac{P(E_i)p(H|E_i)}{P(H)} \quad (18)$$

where H is the observed catalog; $P(E_i)$ is the prior probability assigned to model E_i; $p(H|E_i)$ is the global likelihood on H for model E_i with parameter $\boldsymbol{\theta}_{\mathbf{Ei}}$, given by:

$$p(H|E_i) = \int \pi(\boldsymbol{\theta}_{\boldsymbol{Ei}}|E_i)p(H|\boldsymbol{\theta}_{\boldsymbol{Ei}}, E_i)d\boldsymbol{\theta}_{\boldsymbol{Ei}} \quad (19)$$

and, P(H) is a normalizing factor.

The global likelihood $p(H|E_i)$ measures the probability that randomly selected parameter values from the model E_i would generate dataset H. Model classes that are too simple are unlikely to generate the dataset. Model classes that are too complex can generate many possible datasets; so they may generate that particular data set H at random. Therefore $p(H|E_i)$ can provide a balance between model complexity and data likelihood.

If we refer to the stationary ETAS model as E_0, the goodness of fit of other alternatives can be assessed by their posterior probability ratio with respect to E_0 as:

$$O_{i0} = \frac{p(E_i|H)}{p(E_0|H)} = \frac{P(E_i)}{P(E_0)}\frac{p(H|E_i)}{p(H|E_0)} \quad (20)$$

If the posterior probability ratio (O_{i0}) is above unity, it means the alternative model class (E_i) is more probable, a posteriori, than E_0. Since the calculation of the global likelihood

in Eq. 19 is analytically infeasible for alternative models, we resort to parallel tempering MCMC (Gregory, 2005) to approximate its value, as described in Appendix B.

Seismic transition analysis. Seismic transition analysis with the ETAS model consists of determining the event occurrence rate and then estimating the probability of events exceeding a specified magnitude threshold, based on the GR law, in a moving time window. Moving time windows can be chosen to be separate and, for example, 2-months in length. Since estimates with the ETAS model include secondary aftershock triggering, the frequency of occurrence is simulated via Monte Carlo Simulation (MCS). Occurrence times of events are simulated iteratively as follows. Suppose t_A defines the time of the last event, then t_B is the time of the next one and is obtained by solving the following equation:

$$\int_{t_A}^{t_B} \Lambda\,(t|\,H,\theta)dt = -\log r \tag{21}$$

where r is randomly generated from a uniform distribution between 0 and 1. For each event, its magnitude is randomly selected according to the GR law, with a certain fixed b-value. Uncertainties in the ETAS model are taken into account by replicating the simulation process with random parameter vectors sampled from the posterior MCMC chain, and thereby determining an empirical distribution function for the occurrence rate.

Subsequently, the rate distribution is utilized to calculate credible intervals for the probability of occurrence of events with a given magnitude. This is defined as the probability of exceeding a certain magnitude *M*, given the number of events *n* that occur in the studied time window and the b-value in the GR law:

$$P\left(\widetilde{M} > M\right) = 1 - \left(1 - \frac{10^{-b*M} - 10^{-b*Mmax}}{10^{-b*Mc} - 10^{-b*Mmax}}\right)^{n} \tag{22}$$

where M_c and M_{max} represent the assumed minimum and maximum event magnitudes, respectively.

4.3 Application to the Oklahoma earthquake catalogue

The model and method proposed in the previous section are applied to the study of the earthquake catalog for Oklahoma. The study period for this application spans the time period from January 1975 to September 2014. The following shows modeling approaches and application results.

Modeling approaches. As aforementioned, we assume the priors for each parameter θ_i in $\boldsymbol{\theta}$ are independent lognormal distributions, with parameters chosen to yield relatively broad prior distributions, as documented in Appendix C.

The model is selected based on the posterior probability ratio relative to model E_0, calculated using Eq. 20. We select a uniform prior because no discriminative knowledge is available for the alternative models. Subsequently, the posterior distribution of the parameters of that model is obtained based on the entire Oklahoma seismic dataset, using the Metropolis-Hasting MCMC algorithm with a total number of 1×10^5 simulation steps. Five hundred of these steps are randomly selected after the burn-in phase (the first 2×10^4 values of the chain) through down-sampling to represent the posterior uncertainties of the model.

The selected ETAS model is then deployed to reconstruct the historical transitions in seismicity for the Oklahoma dataset by determining seismicity rates within 2-month moving windows for a magnitude bin of M_L = [2.5, 6.0]. Note that since the estimation of the model parameters has already been completed, the following seismic transition analysis is not very sensitive to the length of the moving windows. Also note that the upper bound of the magnitude range, M_{max}, is somewhat arbitrary as there is no certain knowledge about how to set it (McGarr, 2014). Beginning in January 2000 through the end of the study period (August 2014), for each 2-month interval, we first simulate 20 sequences of seismic events for each parameter sample, using MCS with the 500 selected parameter vectors for event occurrence and the Gutenberg-Ritchter law with the assumed value of b for event magnitude. Then, the counts of events across all the simulated seismic sequences are collected and treated as the empirical distribution for the number

of events to occur in the studied interval. The corresponding credible interval for the magnitude probability is calculated using Eq. 22 with the assumed b-value and M_{max}.

Model comparisons. Figure 9 shows the posterior expected value of λ_0 and Λ as a function of time, along with model fits as a function of transformed time τ_i, for three of the alternative models: E_0 (no induced seismicity), E_3 (1 logistic transition) and E_5 (3 logistic transitions). These models encompass the key range of outputs and alternatives, so are emphasized here. The results for all seven of the alternative models are presented in Appendix B. As shown, models $E_{1\text{-}6}$ all begin showing signs of an increase in λ_0 in late 2009. Furthermore, comparing the relationships between the observed cumulative number of events and the transformed times τ_i for the alternative models, models $E_{4\text{-}6}$ (i.e., the models with more than one logistic increase) appear to better match the observed data. Specifically, models $E_{4\text{-}6}$ predict that λ_0 increases multiple times from approximately 0.01 events per day to well above 2 events per day during the study period.

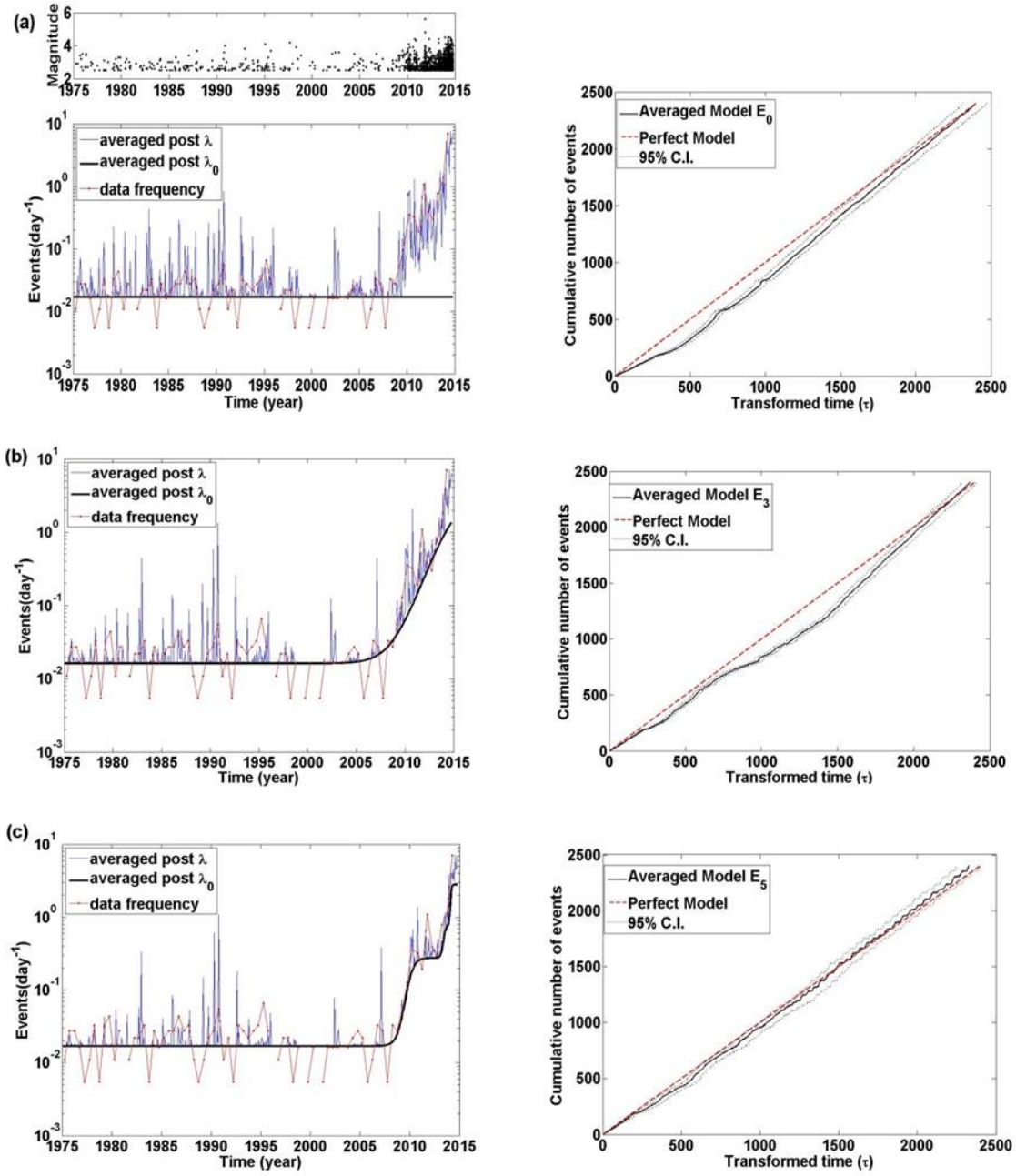

Figure 9 Expected values of λ_0 as a function of time on the left, and the characteristic fit as a function of transformed time τ_i on the right, for model E_0 in (a), model E_3 in (b) and model E_5 in (c), respectively.

Table 3 lists posterior probability ratios in favor of each alternative model over the stationary model E_0 (details of the calculation are illustrated in Appendix B). While model E_0 is clearly dominated by all other models considered, models $E_{4\text{-}6}$ have the highest posterior probability ratios (i.e., approximately 1×10^{90} more likely than model E_0). These results strongly indicate that there was more than one increase in seismicity in

the data. Furthermore, Bayes' formula justifies the complexity of models $E_{4\text{-}6}$ and regards them as the best models to represent the seismic behavior of the Oklahoma dataset. However, increasing complexity does not necessarily lead to improved performance, as the performance of the models peaks at E_5 and then decreases at E_6 due to the penalty imposed for the additional parameters. Model E_5 provides the best balance of model complexity and data likelihood, and therefore is selected for further analysis of the seismic activity in Oklahoma. It is worth mentioning that the result is the same under other information criteria, such as the AIC and BIC.

Model	E_0	E_1	E_2	E_3	E_4	E_5	E_6
Posterior probability ratio (O_{io})	1	2.0×10^{44}	4.0×10^{48}	1.4×10^{56}	3.6×10^{89}	1.4×10^{91}	5.3×10^{89}

Table 3 Posterior probability ratios in favor of each alternative model over the stationary model E_0 with regard to the Oklahoma dataset.

Parametric inference in model E_5. Features of the five hundred samples from the posterior parameter distribution of model E_5 are summarized in Table 4, including their means, standard deviations, 5^{th}, 50^{th} and 95^{th} percentiles (a more complete description is presented in Figure C1). As shown, the first increment in seismic rate occurred approximately 35 years after the beginning of the study period (i.e., around January 2010 in real time, consistent with the result found by Llenos and Michael (2013) and Wang et al., (2015)), followed by a second increment 3.5 years later (i.e., June 2013) and a third increment around 6 months after the second one (i.e., January 2014). The magnitudes of the three increments are 0.28, 0.78 and 2.16 events per day, respectively. The corresponding uncertainty in $\lambda_0(t)$ is shown in Figure 10. Besides, the direct contribution of external forces on the seismicity transitions is estimated to be 56% on average, calculated using Eq. 14 and the selected parameter samples. The upper and lower bound of the 95% confidence interval for this quantity is 51% and 59%, respectively.

	Mean	Std. Dev	5th percentile	50th percentile	95th percentile
μ_0 (Events/day)	0.02	0.001	0.015	0.017	0.019
t_{d1} (years)	35.06	0.16	34.84	35.03	35.37
Δ_1 (Events/day)	0.28	0.02	0.24	0.27	0.31
η_1 (days^{-1})	0.008	0.002	0.005	0.008	0.011
t_{d2} (years)	3.49	0.19	3.15	3.49	3.78
Δ_2 (Events/day)	0.78	0.26	0.44	0.71	1.37
η_2 (days^{-1})	0.02	0.01	0.01	0.02	0.04
t_{d3} (years)	0.58	0.13	0.37	0.57	0.78
Δ_3 (Events/day)	2.16	0.25	1.68	2.18	2.54
η_3(days^{-1})	0.13	0.12	0.05	0.08	0.42
K_0(Events/day)	0.03	0.01	0.02	0.03	0.05
α	1.79	0.09	1.65	1.78	1.94
c (days)	0.23	0.06	0.15	0.22	0.35
p	1.83	0.11	1.67	1.82	2.04

Table 4 Statistics of the marginal posterior distribution of the parameters of model E_5 based on the posterior samples.

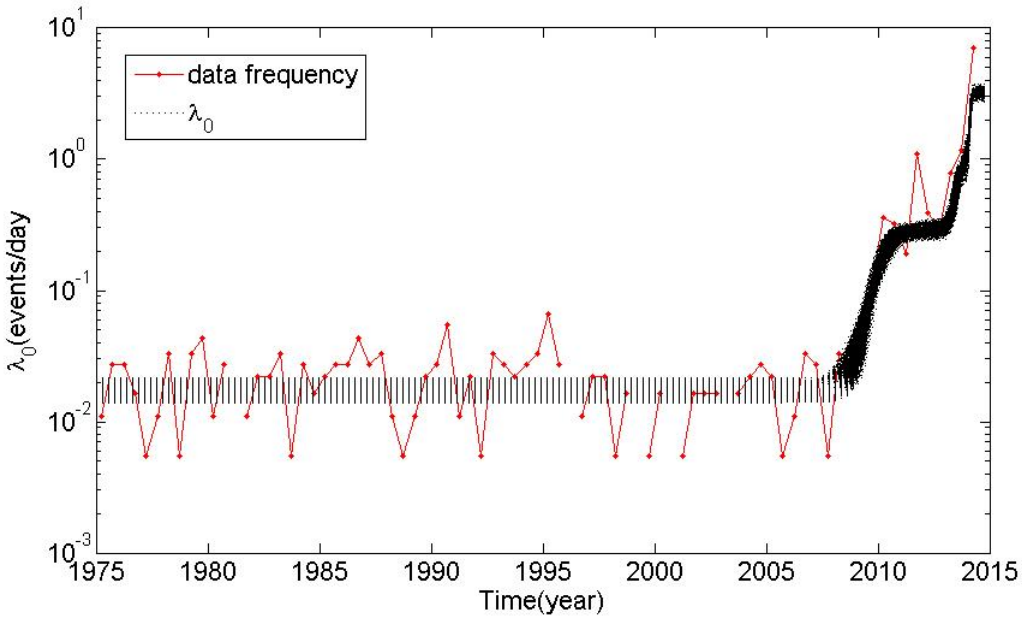

Figure 10 Posterior distribution of λ_0 (t) of model E_5.

Our study is limited in its ability to provide insight into spatial variations of the seismic rate, which we intend to address in our further research efforts. Nonetheless, the identified three seismic rate increments in Oklahoma seemingly correspond to a complex system of faults that have been activated (McNamara et al., 2015). Linking our results to the geophysical observations made by McNamara et al. (2015), it may be noted that the first identified increment corresponds to an increased seismicity rate in the vicinity of Jones and of Prague, which began in 2010 and late 2011, respectively; the second increment corresponds to the vicinity of Guthrie and Langston, where the seismic rate increased significantly in mid-2013; and the third increment corresponds to the seismicity in north-central Oklahoma (increasing significantly since late 2013) and in the vicinity of Cushing (beginning to emerge in late 2014).

It is also of interest to examine the correlation between the logistic model parameters. Consider the three important parameters that affect the timing and upward shift of the logistic curve $\{t_d, \Delta, \eta\}$. For the first increment, t_d and Δ are positively correlated as shown in Table 4, meaning that the later the time that the model assigns as the midpoint of the increase in the induced seismicity, the higher is the amount of increase in λ_0

predicted necessary to compensate for the delayed timing. On the other hand, t_d is negatively correlated with η because a large value of t_d requires a small value of η to account for previous increases in the induced seismicity. Similarly, Δ and η are negatively correlated so that under-estimating η can lead to over-estimating Δ. As for the other sets of logistic parameters, we can draw similar insights from the full correlation coefficient matrix for the parameters of model E_5 presented in Table C2 in the appendix.

Assessment of seismic transitions. Figure 11 shows the inferred rate, as a function of time estimated to have occurred in each 2-month interval from January 2000 through August 2014 (including natural, induced and aftershock events with magnitude M_c and above). The results suggest a period of stationary natural seismicity from January 2000 until late 2009. Thereafter, a distinct increasing seismicity began to occur and the region experienced two more increments in λ_0, resulting in an overall seismic rate increase of a factor of more than 100, from 0.05 to over 5 events per day (across the state of Oklahoma) by August 2014. As indicated, the past trends of seismicity modeled by model E_5 are generally in agreement with the observed record.

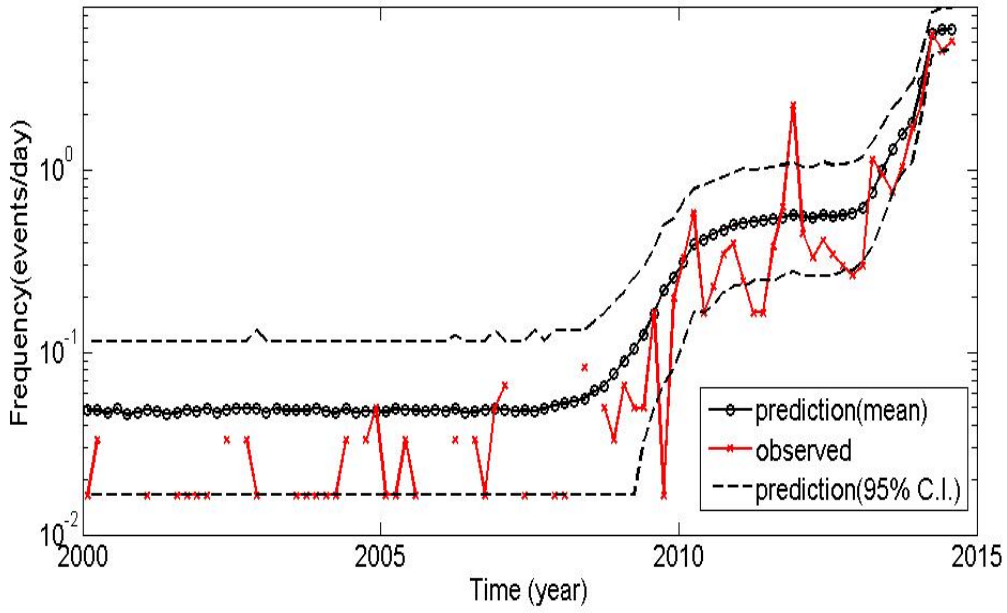

Figure 11 Empirical distribution of event frequency (including natural, induced and aftershock events) to occur in each 2-month interval from Jan. 2000 through Aug. 2014, reconstructed using model E_5.

Figure 12 shows the corresponding result for the predicted probability of exceeding a given magnitude in a given time interval. We evaluate it as the probability of exceeding M_L 4 or 5 for a moving 2-month window according to Eq. 22. Based on the GR law, the resulting magnitude probability for the state of Oklahoma exhibits a similar trend to the seismic rate: the probability of occurrence of an M_L 4.0 event or larger in a two-month period increases from less than 0.1 in January 2000 to nearly 1.0 (i.e., almost surely to occur) by late 2014. For an M_L 5.0 event the probability of exceedance increases from about 0.01 in 2000 to over 0.5 in late 2014. Note that these results are subject to variation under different b-values, as shown in a subsequent sensitivity analysis.

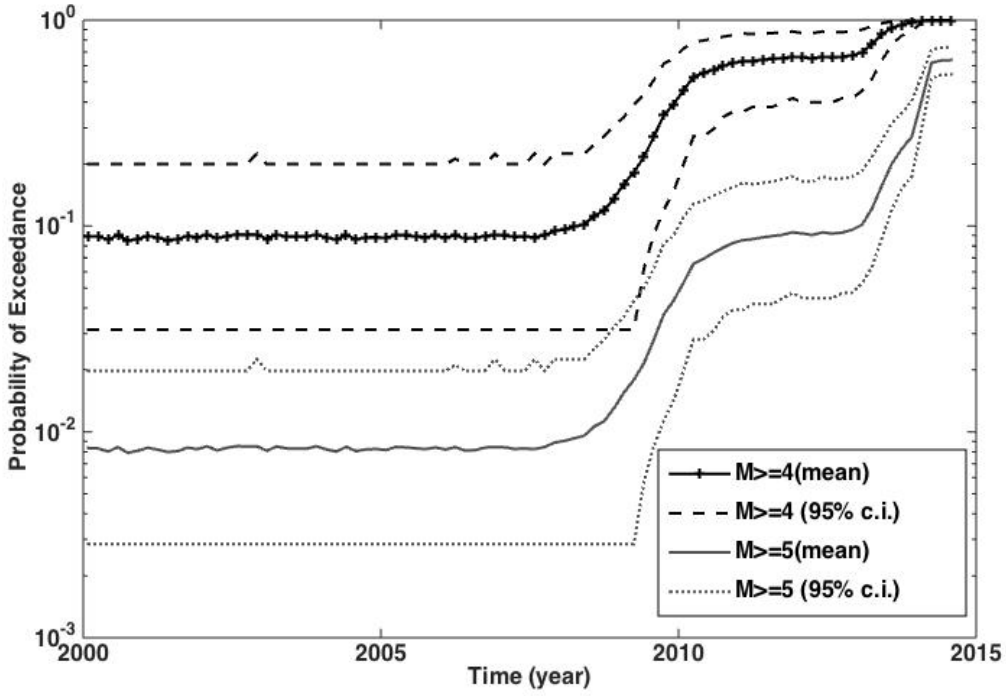

Figure 12 Empirical distribution of magnitude probability for each 2-month interval from Jan. 2000 through Aug. 2014, reconstructed using model E_5.

Sensitivity of predicted probabilities of event magnitude to the assumed b-value and M_{max}. As well known, b-value and M_{max} play an important role in estimating the probability of events exceeding a certain magnitude. It is thus important to examine the impact of these two parameters, especially since we do not estimate their values by processing data. A range of values for the b-value and M_{max} are investigated – allowing the b-value to equal 0.9, 1.0 or 1.1, and M_{max} to increase from M_L 6.0 to 7.0 and 8.0 - and the results are displayed for the probability of exceeding M_L 5.0 with other parameters held equal in Fig. 13 (note that in the previous analysis, the b-value is fixed at 1.0 and M_{max} at M_L 6.0). As shown, the magnitude probability is quite sensitive to the b-value - a lower b-value generates a higher probability; while the choice of M_{max} is less influential on the estimated probability.

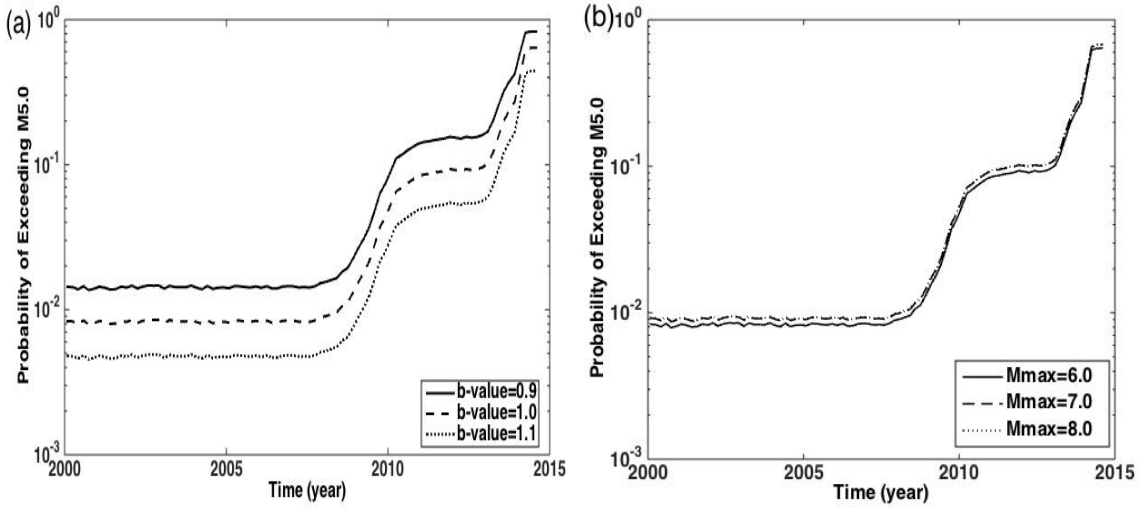

Figure 13 (a) Sensitivity analysis of the b-value for the magnitude probability analysis for the Oklahoma dataset. (b) Sensitivity analysis of M_{max}. Note that in each analysis scenario, the value is set at 1.0 and M6.0 for the b-value and M_{max} respectively, if not otherwise specified.

4.4 Conclusions and Discussions

Modeling the effect of external forces on a seismic activity is critical for providing direct inputs to a probabilistic seismic hazard analysis. In this paper, we propose a method based a parametric extension of the ETAS model, Bayesian inference, Monte Carlo Simulation, and the GR law. The method properly accounts for interdependence among events, comparison of alternative parametric functions for modeling induced seismicity, and uncertainty in both the natural event rate and the timing and magnitude of increases in the event rate associated with induced seismicity.

The methodology in this paper offers several advantages, compared to the existing non-stationary ETAS models. First of all, instead of using windowing techniques, our method estimates model parameters based on the entire dataset, providing a consistent and fully-informed inference. This is also the case for Kumazawa and Ogata (2013) and Marsan and Helmstetter (2013), but they adopted non-parametric inference methods, and these tend to introduce additional complexity and computational burden to the inference, compared to the parametric method used in this paper. And finally, because our method does not use injection data, it can be applied more generally to other non-stationary processes, allowing for the presence of physical processes and random outcomes that make it difficult to specify a direct correlation between event rates and operational parameters.

Application to the Oklahoma region shows that the proposed method is able to provide consistent inferences on seismic transitions, identifying an appropriate model and its parameters. According to the identified model, the direct contribution of external forces on the seismicity transitions in Oklahoma is approximately 56%. By reconstructing past seismicity, we show that the overall seismic rate increased by a factor of more than 100, from 0.05 to over 5 events per day during the study period. Accordingly, the probabilities of large magnitude events have been considerably affected in this region. Through a sensitivity analysis, we show how these event probabilities depend on the b-value in GR law and the M_{max} to occur.

Wastewater injection data (e.g. injection rates) for Oklahoma are not available in this study. Such data could be used to better interpret the inferred transition times or to provide prior statistical likelihood to time periods when major changes in injection rates occur. If such correlations are identified, this could provide important insight into the development of the pore pressure level of the injection site, helping to identify a critical pore pressure level at which the strength of local faults has been significantly weakened and susceptible to pore-pressure perturbation. These correlations can also provide guidance to future operations and management of injection wells in terms of minimizing seismic hazards. Injection data can also be used to help choose functional forms for the non-stationary ETAS model. For example, if the injection scheme is periodic and the corresponding correlation is evident, we may consider using periodic functions for modeling seismicity transitions.

We believe that the proposed method can provide a statistical basis for modeling and assessing the transitions of seismicity. With subsequent linkage to information on ground motions, exposure, fragility and consequences, the method can provide an initial decision-support tool to identify areas with increasing levels of induced events, updating seismic hazard estimates (Petersen et al., 2015), and supporting a comprehensive assessment to decide which risk-mitigation strategy should be recommended (Bommer et al., 2015).

Chapter 5 A Bayesian approach for assessing spatio-temporal evolution of seismic event rate associated with fluid injections

Abstract

In this chapter, we develop a statistical method for periodically updating the spatial seismic event rate based on the previous event rate and the current observation. The method accounts for uncertainty in the initial state and the transition of the event rate from the previous to the current time step. The transition model contains a correlation structure for the changes in event rate in different locations, the parameter of which can be tuned to achieve the optimal correlation function. The Monte Carlo sequential Bayesian inference method is employed to estimate the parameters involved in the inference process.

Application of this approach to the Oklahoma dataset shows the model is able to well characterize the spatial distribution of seismic rate as a function of time. It shows that a considerable part (i.e., the middle and the northern part) of the Oklahoma state has seen significant increase in seismic rate. The timing and magnitude of the increase varies for different locations, with the timing ranging from year 2010 to 2015 and the magnitude ranging from 0.1 to 1 events per day. The model is also investigated for its short-term forecasting ability. In general, its forecasting performance is satisfactory both in itself and in comparison to a naïve method.

5.1 Introduction

An increased event rate implies an increased seismic hazard (McGarr et al., 2015). Characterization of event rates and locations is necessary for performing probabilistic seismic hazard analysis (e.g., Cornell, 1968; McGuire, 2004). There are existing methods

for developing seismic source characterizations of tectonic earthquakes (e.g., Field et al., 2014; Lawrence et al., 2014; Moschetti et al., 2015). However, due to the highly non-stationary nature of induced earthquakes, these existing methods for tectonic events are not suitable for characterizing the evolution of induced seismic event rate in time and space.

This study presents a general sequential Bayesian inference method for characterizing where, when and to what extent induced seismicity increases, by recursively updating the spatial event rate based on the previous rate and the current observation. The method relies discretizing the region into a grid and the time into time intervals. As a result, the spatial event rate is represented by points on the grid and the evolution is represented by the correlated transition of those points from one time step to the next. Due to lack of analytical solutions, the inference of the model is carried out using Monte Carlo method, in which a set of weighted samples/particles are periodically updated to represent our belief about the state of the event rate. The overall Bayesian framework provides a flexible inference structure since it allows processing instrumental seismic records and the incorporation of expert geophysical knowledge, helping to constrain the prior parameter ranges of random variables involved in the inference process.

The performance of the proposed model is investigated in an application to the Oklahoma earthquake dataset. In order to apply the model, the dataset needs to be preprocessed properly as shown in the "data preprocessing" section. And then, the model is tuned for a hyper-parameter, which controls the correlation strength in the transition of event rates on the grid points, based on the entire processed data using Bayesian model selection method. The tuned model is subsequently employed to perform inference showing how seismic event rate evolves across the Oklahoma state. The model is also investigated in terms of its short-term event rate forecasting ability for the dataset, which is compared to a naïve forecasting model. It is shown that the proposed model considerably outperforms the naïve model.

The rest of this chapter is organized as follows: firstly, we describe in detail the proposed model and the Monte Carlo procedure for inferring its parameters; subsequently, the analysis of the seismic evolution is reported. Finally, the result for the predicting performance of the proposed model is presented both in itself and in relative to a naïve method, followed by further discussion and conclusions.

5.2 Models and Methods

Spatio-temporal Point Process Model. We consider independent earthquake occurrences (without aftershocks) as a Spatio-temporal point process (STPP). STPP is a random collection of points, where each point represents the time and location of an event. Such a process is generated by a statistical model characterized by its associated event rate function λ, with the rate at time t and spatial coordinate $\mathbf{z} = [x \quad y]^{\mathrm{T}}$ given by $\lambda(\mathrm{t}, \mathbf{z})$. If λ is known, we can calculate the probability density of the occurrence of a sequence of events. If given a sequence of data and a uncertain prior knowledge about λ, we can update our belief through Bayesian theorem based on the prior knowledge and the likelihood function induced on the data. The formula for the log of the likelihood function is given by:

$$\log \mathrm{p}(\boldsymbol{H}|\lambda) = \sum_{i=1}^{N} \log[\lambda(t_i, \mathbf{z}_i)] - \int_0^T \int_{\mathbf{z}\in A} \lambda(t, \mathbf{z}) d\mathbf{z}\, dt \qquad (23)$$

where $\boldsymbol{H} = \{(t_i, \mathbf{z}_i) : i = 1, .., N\}$ contains the sequence of times and space coordinates of Nevents occurring in region A, during time interval $[0, T]$.

The key component in a STPP model is λ. In the following, we will specify its form used in this study and how to learn its parameters using Bayesian estimation methods.

Specification of $\boldsymbol{\lambda}$. To specify the form of λ, we start by discretizing space into a grid, and time into equal-duration time intervals. Subsequently, let vector $\boldsymbol{\lambda}_k$ contain the event rates on each point on the grid at time interval t_k. As a result, λ can be represented by a sequence of $\boldsymbol{\lambda}_k$, given information about the spatio-temporal discretization. Correspondingly, let $\boldsymbol{H}_k$ contain events occurring during t_k.

The sequence of $\boldsymbol{\lambda}_k$ can be treated as a sequence of the state of a system and it can be estimated using Bayesian sequential updating methods, in which we sequentially update the current system state based on one or more prior states and the current observation. We assume the sequence follows a first-order Markov process, meaning the current state depends on the historical states only through the previous state. This process is graphically displayed in Figure 14.

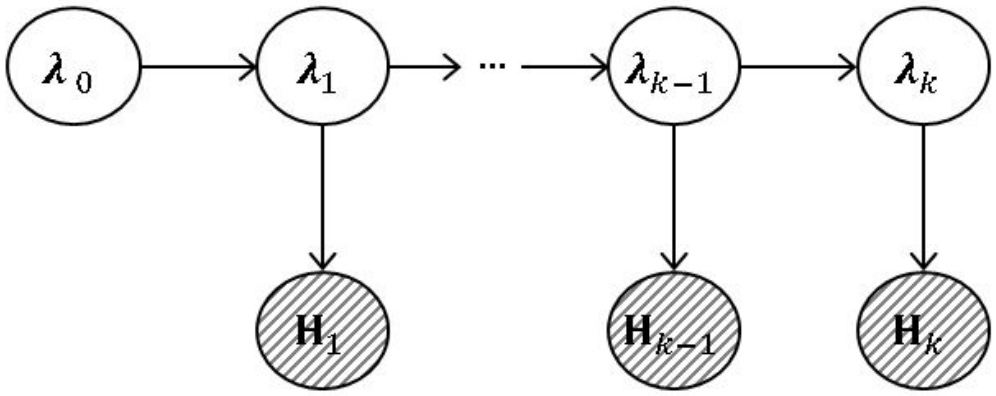

Figure 14 Graphical representation of the process for the state and observation of the earthquake system.

To complete the updating process, we need to specify the initial uncertainty and state transition function for the system. Let the initial distribution $\boldsymbol{\lambda}_0$ be modeled by a multivariate lognormal distribution:

$$\boldsymbol{\lambda}_0 \sim \mathcal{LN}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) \tag{24}$$

where $\boldsymbol{\mu_0}$ is a vector and $\boldsymbol{\Sigma_0}$ is a covariance matrix, the combination of which specifies the statistical property of $\boldsymbol{\lambda}_0$. In the covariance matrix, the correlation between two components is calculated using the squared-distance exponential function:

$$\rho_{ij} = e^{-\frac{(z_i - z_j)'(z_i - z_j)}{2l^2}} \tag{25}$$

where $\boldsymbol{z}_i$ and $\boldsymbol{z}_j$ stand for the spatial coordinates of the components, $\left(\boldsymbol{z_i} - \boldsymbol{z_j}\right)'$ is the transpose of $\left(\boldsymbol{z_i} - \boldsymbol{z_j}\right)$ and, hyper-parameter l controls the decay of the correlation as a function of distance. The smaller l is, the faster the decay.

The state transition is assumed to take the following form

$$\boldsymbol{\lambda}_k = \mathbf{A}_k + \boldsymbol{\lambda}_{k-1} \quad (26)$$

where $\mathbf{A}_k$ is a random vector, encoding the uncertain effect of induced seismicity. For simplicity, the distribution of $\mathbf{A}_k$ is assumed to be stationary in time so that $\mathbf{A}_k = \mathbf{A}$ for any k. Let random variable C takes value in $\{0, 1\}$ with $\mathrm{C} = 0$ standing for no change in the state and $\mathrm{C} = 1$ for the occurrence of an increase. We assume the effect of induced seismicity is such that the state of the system experiences an increase with a probability $\mathrm{P}(\mathrm{C} = 1)$, specifically in terms of $\mathbf{A}$

$$p(\mathbf{A}) = \mathrm{P}(\mathrm{C} = 0) \times \delta(\mathbf{A} - \mathbf{0}) + \mathrm{P}(\mathrm{C} = 1) \times \mathcal{LN}(\boldsymbol{\mu}_\mathbf{A}, \boldsymbol{\Sigma}_\mathbf{A}) \quad (27)$$

where $\delta(\mathbf{A} - \mathbf{0})$ is the delta function centered on $\mathbf{0}$, $\boldsymbol{\mu}_\mathbf{A}$ is a vector and $\boldsymbol{\Sigma}_\mathbf{A}$ is a covariance matrix for the components in $\mathbf{A}$. The correlation between the components in also calculated according to Eq. 25. In this case, parameter l is responsible for controlling correlation among the changes in different state components. The choice of its value is vital for optimizing model performance. In principle, the optimal value can be identified via model selection as shown later.

As a result, $p(\boldsymbol{\lambda}_k | \boldsymbol{\lambda}_{k-1})$ can be represented by

$$p(\boldsymbol{\lambda}_k | \boldsymbol{\lambda}_{k-1}) = \mathrm{P}(\mathrm{C} = 0) \times \delta(\boldsymbol{\lambda}_k - \boldsymbol{\lambda}_{k-1}) + \mathrm{P}(\mathrm{C} = 1) \times \mathcal{LN}(\boldsymbol{\lambda}_k - \boldsymbol{\lambda}_{k-1}; \boldsymbol{\mu}_\mathbf{A}, \boldsymbol{\Sigma}_\mathbf{A}) \quad (28)$$

where $\delta(\boldsymbol{\lambda}_k - \boldsymbol{\lambda}_{k-1})$ is the delta function centered on $\boldsymbol{\lambda}_{k-1}$.

With the initial uncertainty and transition function, we can now estimate the sequence of $\boldsymbol{\lambda}_k$ based on the set of all available observation vectors $\boldsymbol{H}_{1:k} = \{\boldsymbol{H}_1, \boldsymbol{H}_2, \ldots, \boldsymbol{H}_k\}$. The estimation problem can be understood in a Bayesian sense by turning the problem into an estimation of the conditional posterior $p(\boldsymbol{\lambda}_k | \boldsymbol{H}_{1:k})$. Our goal is to recursively evaluate this density on the arrival of new observations. For interested readers, the description of the recursive procedure is in documented in Appendix D.

Updating and Particle Filter. Analytical methods have been developed for solving such recursive Bayesian estimation problems, including Kalman Filter and its variants. However, they cannot be applied to our case because of the non-linear transition and the non-Gaussian (i.e. non-homogenous Poisson) likelihood used in this study. Although at the expense of more intense computation, the Particle Filter (PF) method provides a more general framework to solve the estimation problem, requiring no requirements on linearity and Gaussian likelihood. Therefore, it is employed in this study.

The PF method is a general Monte Carlo method for approximating sequential distributions that are analytically intractable. It is traditionally based on Sequential Importance Sampling (Liu and Chen, 2001), which, at time $k-1$, aims at approximating $p(\boldsymbol{\lambda}_{k-1}|\boldsymbol{H}_{1:k-1})$ with a set of weighted samples $\{\boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}}, w_{k-1}^{i}\}_{i=1}^{N_s}$, also known as particles, and recursively update these particles to obtain an approximation $\{\boldsymbol{\lambda}_{\boldsymbol{k}}^{\boldsymbol{i}}, w_{k}^{i}\}_{i=1}^{N_s}$ to the posterior distribution at the next time step $p(\boldsymbol{\lambda}_k|\boldsymbol{H}_{1:k})$. Under certain assumptions specified below, the particles are updated in a way that:

$$\boldsymbol{\lambda}_{\boldsymbol{k}}^{\boldsymbol{i}} \sim q\left(\boldsymbol{\lambda}_{\boldsymbol{k}}^{\boldsymbol{i}}|\boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}}, H_k\right) \tag{29}$$

$$w_k^i = w_{k-1}^i \frac{p(H_k|\boldsymbol{\lambda}_{\boldsymbol{k}}^{\boldsymbol{i}})\, p(\boldsymbol{\lambda}_{\boldsymbol{k}}^{\boldsymbol{i}}|\boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}})}{q(\boldsymbol{\lambda}_{\boldsymbol{k}}^{\boldsymbol{i}}|\boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}}, H_k)} \tag{30}$$

where $q\left(\boldsymbol{\lambda}_{\boldsymbol{k}}^{\boldsymbol{i}}|\boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}}, H_k\right)$ is a importance distribution for propagating particles at the current time step, which is assumed only dependent on the previous state and the current observation, and $p\left(\boldsymbol{\lambda}_{\boldsymbol{k}}^{\boldsymbol{i}}|\boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}}\right)$ is the predicative density. $\{w_k^i\}_{i=1}^{N_s}$ is normalized in each time step to sum to 1.

According to Eq. 28, the distribution of $\mathbf{x}_{\boldsymbol{k}}^{\boldsymbol{i}}|\mathbf{x}_{\boldsymbol{k-1}}^{\boldsymbol{i}}$ follows a mixture of two probability densities

$$p\left(\boldsymbol{\lambda}_{\boldsymbol{k}}^{\boldsymbol{i}}|\boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}}\right) = \mathrm{P}(\mathrm{C}=0)\times\delta\left(\boldsymbol{\lambda}_{\boldsymbol{k}}^{\boldsymbol{i}} - \boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}}\right) + \mathrm{P}(\mathrm{C}=1)\times\mathcal{LN}\left(\boldsymbol{\lambda}_{\boldsymbol{k}}^{\boldsymbol{i}} - \boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}};\, \boldsymbol{\mu}_{\mathbf{A}}, \boldsymbol{\Sigma}_{\mathbf{A}}\right) \tag{31}$$

where $\delta(\boldsymbol{\lambda}_{\boldsymbol{k}}^{\boldsymbol{i}} - \boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}})$ is the delta function centered at $\boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}}$.

If $q\left(\boldsymbol{\lambda}_{\boldsymbol{k}}^{\boldsymbol{i}}|\boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}}, H_k\right)$ is chosen to be equal to $p\left(\boldsymbol{\lambda}_{\boldsymbol{k}}^{\boldsymbol{i}}|\boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}}\right)$, then Eq. 30 reduces to

$$w_k^i = w_{k-1}^i p\left(H_k|\mathbf{x}_{\boldsymbol{k}}^{\boldsymbol{i}}\right) \tag{32}$$

This variant of PF method is referred to as the Boot Strap Particle Filter. It is one of the easiest to implement, and thus one of the most widely used. However in the conflict between the predicative density and the likelihood, particles generated from the predicative density are not able to well explore the important region of the posterior space. This problem is particularly severe when the predicative density is broad and the dimension of $\boldsymbol{\lambda}_k$ is high. In order to draw samples effectively, one needs to design an informative importance density, considering both the predicative density and the likelihood. Specifically, we propose $q\left(\boldsymbol{\lambda}_{\boldsymbol{k}}^{\boldsymbol{i}}|\boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}}, H_k\right)$ to be of a similar form as the predicative density:

$$q\left(\boldsymbol{\lambda}_{\boldsymbol{k}}^{\boldsymbol{i}}|\boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}}, H_k\right) = \mathrm{P}(\mathrm{C}=0|\boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}}, H_k)\times\delta(\boldsymbol{\lambda}_{\boldsymbol{k}}^{\boldsymbol{i}} - \boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}}) + \mathrm{P}(\mathrm{C}=1|\boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}}, H_k)\times\mathcal{LN}\left(\boldsymbol{\lambda}_{\boldsymbol{k}}^{\boldsymbol{i}} - \boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}}; \hat{\mathbf{x}}_{k|\boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}}, H_k, \mathrm{C}=1}^{i}, \beta\boldsymbol{\Sigma}_{\mathbf{A}}\right) \tag{33}$$

where $\hat{\mathbf{x}}_{k|\boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}}, H_k, \mathrm{C}=1}^{i}$ is the point estimate for the posterior mode of the log of $\boldsymbol{\lambda}_{\boldsymbol{k}}^{\boldsymbol{i}}$ in the case of change, $\boldsymbol{\Sigma}_{\mathbf{A}}$ is the same covariance matrix as for the predicative density, β is a scalar controlling the spread of the importance density, and $\mathrm{P}(\mathrm{C}=0|\boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}}, H_k)$ and $\mathrm{P}(\mathrm{C}=1|\boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}}, H_k)$ represent the posterior coefficients of the mixture. The posterior mode estimate can be obtained using Newton methods.

The formula for $\mathrm{P}(\mathrm{C}=0|\boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}}, H_k)$ and $\mathrm{P}(\mathrm{C}=1|\boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}}, H_k)$ are:

$$\mathrm{P}\left(\mathrm{C}=0|\boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}}, H_k\right) = \frac{p\left(H_k|\boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}}, \mathrm{C}=0\right)\mathrm{P}\left(\mathrm{C}=0|\boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}}\right)}{p\left(H_k|\boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}}, \mathrm{C}=0\right)\mathrm{P}\left(\mathrm{C}=0|\boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}}\right)+p\left(H_k|\boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}}, \mathrm{C}=1\right)\mathrm{P}\left(\mathrm{C}=1|\boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}}\right)}$$

$$= \frac{p\left(H_k|\boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}}, \mathrm{C}=0\right)\mathrm{P}(\mathrm{C}=0)}{p\left(H_k|\boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}}, \mathrm{C}=0\right)\mathrm{P}(\mathrm{C}=0)+p\left(H_k|\boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}}, \mathrm{C}=1\right)\mathrm{P}(\mathrm{C}=1)} \tag{34}$$

$$\mathrm{P}(\mathrm{C}=1|\boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}}, H_k) = 1 - \mathrm{P}(\mathrm{C}=0|\boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}}, H_k) \tag{35}$$

where $p\left(H_k|\boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}}, \mathrm{C}=0\right)$ can be easily calculated by

$$p\left(H_k|\boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}}, \mathrm{C}=0\right) = p\left(H_k|\boldsymbol{\lambda}_{\boldsymbol{k}}^{\boldsymbol{i}} = \boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}}\right) \tag{36}$$

while $p\left(H_k|\boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}}, \mathrm{C}=1\right)$ is difficult to compute but can be expanded and approximated as an Gaussian-density weighted integral, given by

$$p\left(H_k|\boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}}, C=1\right) = \int p\left(H_k|\boldsymbol{\lambda}_{\boldsymbol{k}}^{\boldsymbol{i}} = \mathbf{A} + \boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}}\right)\mathcal{LN}(\boldsymbol{A};\, \boldsymbol{\mu}_{\mathbf{A}}, \boldsymbol{\Sigma}_{\mathbf{A}})d\boldsymbol{\lambda}_{\boldsymbol{k}}^{\boldsymbol{i}}$$

$$= \int \frac{p\left(H_k|\boldsymbol{\lambda}_{\boldsymbol{k}}^{\boldsymbol{i}}=\mathbf{A}+\boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}}\right)}{\prod \mathbf{A}}\mathcal{N}(\log(\boldsymbol{A})\,;\, \boldsymbol{\mu}_{\mathbf{A}}, \boldsymbol{\Sigma}_{\mathbf{A}})d\boldsymbol{\lambda}_{\boldsymbol{k}}^{\boldsymbol{i}}$$

$$= \int \frac{p\left(H_k|\boldsymbol{\lambda}_{\boldsymbol{k}}^{\boldsymbol{i}}=e^{\mathbf{x}}+\boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}}\right)}{e^{\sum \mathbf{x}}}\mathcal{N}(\mathbf{x};\, \boldsymbol{\mu}_{\mathbf{A}}, \boldsymbol{\Sigma}_{\mathbf{A}})d\boldsymbol{\lambda}_{\boldsymbol{k}}^{\boldsymbol{i}}$$

$$= \int f(\mathbf{x})\mathcal{N}(\mathbf{x};\, \boldsymbol{\mu}_{\mathbf{A}}, \boldsymbol{\Sigma}_{\mathbf{A}})d\boldsymbol{\lambda}_{\boldsymbol{k}}^{\boldsymbol{i}} \tag{37}$$

where $\mathbf{x} = \log(\boldsymbol{A})$ and $f(\mathbf{x}) = \frac{p\left(H_k|\boldsymbol{\lambda}_{\boldsymbol{k}}^{\boldsymbol{i}}=e^{\mathbf{x}}+\boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}}\right)}{e^{\sum \mathbf{x}}}$.

A common technique to approximate such a weighted integral is the Sigma Point method, which can provide an estimate of $p\left(H_k|\boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}}, C=1\right)$ using a discrete sum of the form

$$\sum_{j=0}^{2n_x} w_j f\left(\boldsymbol{\chi}^{(j)}\right) \tag{38}$$

where $\left\{\boldsymbol{\chi}^{(j)}\right\}$ is a set of sigma points, $\{w_j\}$ is the set of the corresponding weights, and n_x is the dimension of the state vector $\mathbf{x}$. The specific rule for determining the sigma points and the corresponding weights can be found in Appendix E.

In practice, the iterations of the update equations in Eq. 29-30 lead to a degeneracy problem where only a few particles possess the majority of the total weight. One common way to overcome it is resampling (Liu and Chen, 2001). In resampling, a new set of N_s particles is generated according to the discrete approximation to the distribution $p(\boldsymbol{\lambda}_{\boldsymbol{k}}|H_{1:k})$, provided by the weighted particles:

$$p(\boldsymbol{\lambda}_{\boldsymbol{k}}|H_{1:k}) \approx \sum_{i=1}^{N_p} w_k^i \delta\left(\boldsymbol{\lambda}_{\boldsymbol{k}} - \boldsymbol{\lambda}_{\boldsymbol{k}}^{\boldsymbol{i}}\right) \tag{39}$$

where $\delta\left(\boldsymbol{\lambda}_{\boldsymbol{k}} - \boldsymbol{\lambda}_{\boldsymbol{k}}^{\boldsymbol{i}}\right)$ is the delta function centered at $\boldsymbol{\lambda}_{\boldsymbol{k}}^{\boldsymbol{i}}$. After resampling, the weight of each particle should be set to $\frac{1}{N_s}$. Thus resampling effectively deals with the degeneracy problem by getting rid of particles with very small weights.

We have completely described the steps necessary for the sequential update procedure; which is now summarized in Table 5.

A. Initialize filter	
1. Initialize state vector samples	$\boldsymbol{\lambda}_o^i \sim p(\boldsymbol{\lambda}_o)$
2. Initialize weights	$w_o^i = 1 \,/\, N_s$
3. Initialize time step k	$k = 1$
B. Importance sampling	
For i = 1,…,N_s	
a. Calculate the j Sigma points	$\boldsymbol{\chi}_k^{(i,j)} = \boldsymbol{\mu}_{\mathbf{A}} + \mathbf{Lc}^{(j)}, j = 0,1,\dots,2n_x$
b. Calculate the $p\left(H_k \middle\vert \mathbf{x}_{\boldsymbol{k-1}}^{\boldsymbol{i}}, C = 1\right)$	$p\left(H_k \middle\vert \boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}}, C = 1\right) = \sum_{j=1}^{2n_x} w_j l\left(\boldsymbol{\chi}_k^{(i,j)}\right)$
c. Calculate the posterior coefficients	$\mathrm{P}\left(\mathrm{C} = 0 \middle\vert \boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}}, H_k\right) = \frac{p\left(H_k \middle\vert \boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}}, \mathrm{C}=0\right)\mathrm{P}(\mathrm{C}=0)}{p\left(H_k \middle\vert \boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}}, \mathrm{C}=0\right)\mathrm{P}(\mathrm{C}=0)+p\left(H_k \middle\vert \boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}}, \mathrm{C}=1\right)\mathrm{P}(\mathrm{C}=1)}$
	$\mathrm{P}\left(\mathrm{C} = 1 \middle\vert \boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}}, H_k\right) = 1 - \mathrm{P}\left(\mathrm{C} = 0 \middle\vert \boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}}, H_k\right)$
d. Obtain the estimate of the posterior mode for the log of $\boldsymbol{\lambda}_{\boldsymbol{k}}^{\boldsymbol{i}}$	$\hat{\mathbf{x}}_{k \middle\vert \boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}}, H_k, \mathrm{C}=1}^{i}$
e. Propagate particles	$\boldsymbol{\lambda}_{\boldsymbol{k}}^{\boldsymbol{i}} \sim q\left(\boldsymbol{\lambda}_{\boldsymbol{k}}^{\boldsymbol{i}} \middle\vert \boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}}, H_k\right)$
C. Update the importance weights	$\widetilde{w}_k^i = w_{k-1}^i \frac{p\left(H_k \middle\vert \boldsymbol{\lambda}_{\boldsymbol{k}}^{\boldsymbol{i}}\right) p\left(\boldsymbol{\lambda}_{\boldsymbol{k}}^{\boldsymbol{i}} \middle\vert \boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}}, H_k\right)}{q\left(\boldsymbol{\lambda}_{\boldsymbol{k}}^{\boldsymbol{i}} \middle\vert \boldsymbol{\lambda}_{\boldsymbol{k-1}}^{\boldsymbol{i}}, H_k\right)}$
	$w_k^i = \widetilde{w}_k^i \,/\, \sum_{i=1}^{N_S} \widetilde{w}_k^i$
D.Resample	
G. Moment Calculations	
H. Time step update and return to B	$k = k + 1$

Table 5 Particle Filter procedure.

Bayesian model selection. The parameter l in the correlation function is a hyper-parameter for the model. Different choices of l correspond to models of different complexity, which can considerably affect the modeling performance. Therefore, it is of interest to identify the optimal choice of l and hence the optimal model. The model with $l \to 0$ corresponds to the most complicate model as it allows the components in $\boldsymbol{\lambda}_k$ to change independently; while the model with $l \to \infty$ corresponds to the simplest model since all the components in the system state are forced to change by the same amount as

if there is only one parameter. Therefore, the selection method should prefer models with large l, unless the complexity induced by small l can be justified by data likelihood.

To optimize for l, Bayesian model selection technique is employed. A set of alternative models, each with a different value of l, are proposed and compared. In the Bayesian paradigm, models can be compared probabilistically: the posterior probability of each alternative model S_j is calculated using Bayes' formula (MacKay, 2003):

$$P(S_j|H_{1:k}) = \frac{P(S_j)p(H_{1:k}|S_j)}{P(H_{1:k})} \quad (40)$$

where $H_{1:k}$ is the full input dataset; $\mathrm{P}(S_j)$ is the prior probability assigned to model S_j; $p(H_{1:k}|S_j)$ is the global likelihood on $H_{1:k}$ for model S_j, given by:

$$p(H_{1:k}|S_j) = p(\mathrm{H}_1|S_j) \dots p(H_k|S_j, H_{1:k-1}) \dots p(H_K|S_j, H_{1:K-1}) \quad (41)$$

where $p(H_k|S_j, H_{1:k-1})$ can be approximated by

$$p(H_k|S_j, H_{1:k-1}) = \sum_{i=1}^{N_s} w_{k-1}^{j,i} p(H_k|\boldsymbol{\lambda}_{k-1}^{j,i}) \quad (42)$$

where $\boldsymbol{\lambda}_{k-1}^{j,i}$ represents the ith particle generated at time step k under model S_j, and $p(H_k|\boldsymbol{\lambda}_{k-1}^{j,i}) = \sum_{c=0}^{1} \mathrm{P}(C = c) \times p(H_k|\boldsymbol{\lambda}_{k-1}^{j,i}, C = c)$.

In the absence of prior preference over alternative models, the model yielding the highest global likelihood $p(H_{1:k}|S_j)$ is considered the best model. One way to interpret the global likelihood is to treat it as the probability density of generating the earthquake data by the model. The model that is mostly likely to generate the data should be considered the optimal. It is worth noting that model selection through global likelihood provides a balance between model complexity and data likelihood, which is a desirable property.

Thinning algorithm. In order to apply the above model, earthquake datasets must be first declustered as it typically contains both independent and dependent events. We adopt the same stochastic algorithm as in Chapter 3. The algorithm relies on models for the

independent activity and that for the clustering structure. They both can be provided by the non-stationary ETAS model proposed in Chapter 4. The parameters $\boldsymbol{\theta}$ in the ETAS model can be estimated using Bayesian inference method.

After obtaining the posterior distribution of $\boldsymbol{\theta}$, we select the predicted mean $\widehat{\boldsymbol{\theta}}$ to perform declustering using the thinning method presented in Chapter 3. Therefore, the independent earthquake sequence is realized by selecting each event i with probability φ_i. Although the outcome is intrinsically stochastic, as different sequences are generated from different runs of the algorithm, we focus on only one of the possible declustered sequences. The associated stochastic effect on our analysis can be investigated by comparing results from different declustered sequences.

5.3 Application to the Oklahoma Earthquake Dataset

Data pre-processing. The model and method proposed in the previous section are applied to the study of the earthquake catalog for Oklahoma. The study period roughly spans the time period from January 1975 to March 2016.

Since the catalog contains both independent and clustering events, the declustering algorithm is employed to remove the clustering events. As one realization, the number of resulted independent events is 4327. The declustered events are illustrated in Figure 15, along with the complete data and the $\hat{\lambda}_0(t)$. As expected, the frequency of the declustered data is below that of the complete data, but close to the employed independent event rate. These events are used to infer the evolution of the seismic state for the study region.

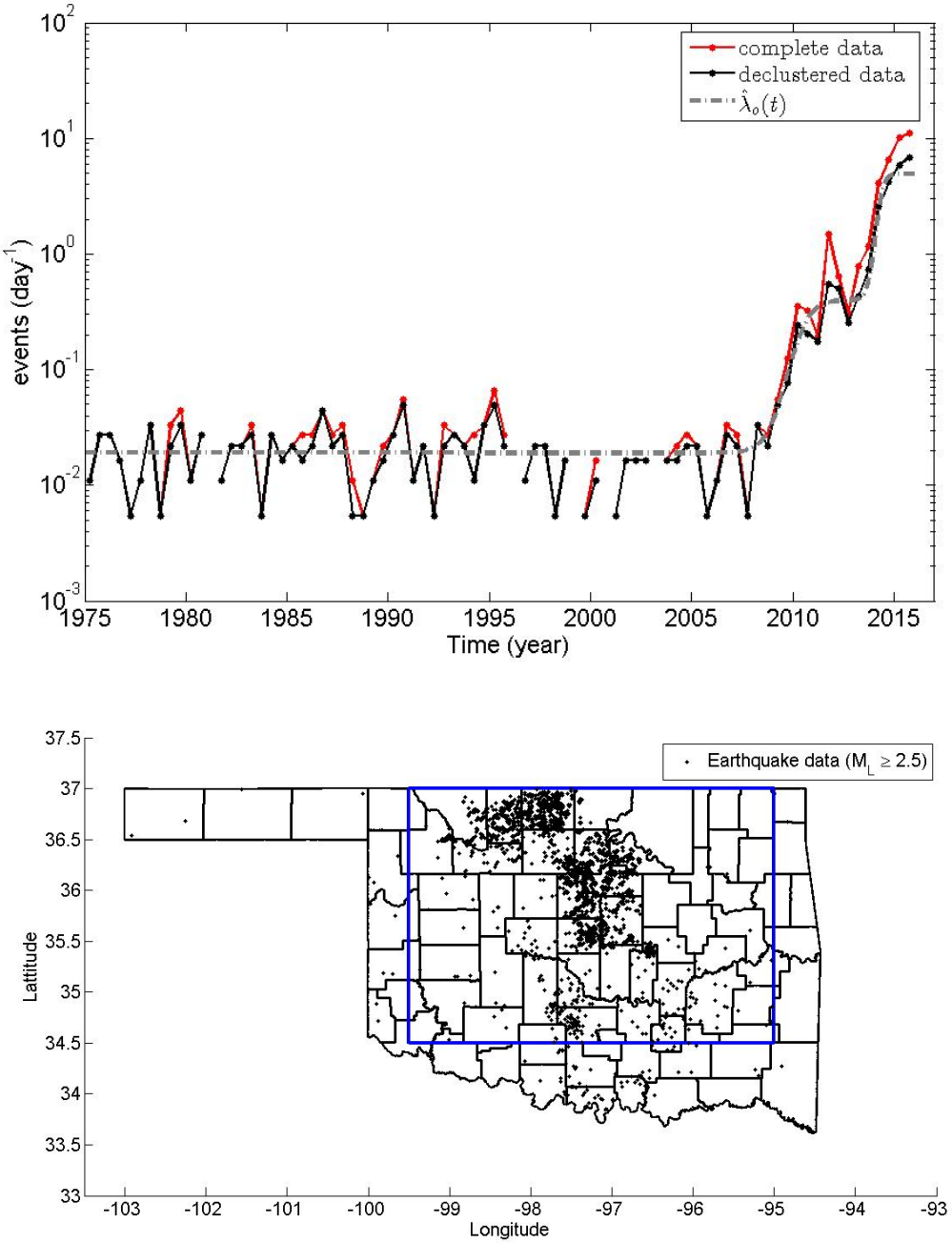

Figure 15 Spatial and temporal illustrations of the declustered catalog (M ≥ 2.5). Plot (a) shows the frequency of the declustered data along with that of the complete data and the estimated independent rate; (b) shows a map of Oklahoma, the study region (inside the rectangle), and the epicenters of the declustered earthquake events.

Modeling approaches. The dataset period is discretized by 2-month intervals, resulting in 247 time steps in total (i.e., K = 247). In order to apply the proposed method in a regular setting, we focus on the events in a rectangular geographical region within the Oklahoma

boarder as shown by the rectangle in Figure 15(b). The study region goes from 99.5°W to 95°W in longitude, and from 34.5°N to 37°N in latitude. It is divided into cells of 0.5°×0.5°, generating a 5 by 9 grid and thus a state vector with 45 components. The location of each point on the grid is represented by the center of the corresponding cell. Figure 16 shows the grid on the Oklahoma map. Figure 17 shows the cumulative number of events functions over time for each cell. Since some of the cells experience more events during the study period than the others, they are labeled out in both figures for highlighting purposes. As a result, there are 16 cells to be used to highlight the model performance.

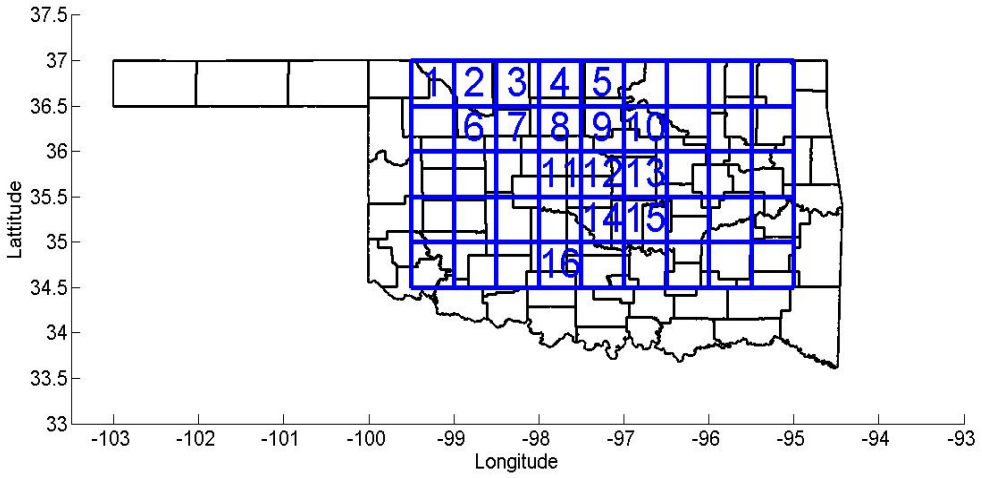

Figure 16 The 5 by 9 grid on the map of Oklahoma. The cells with labels will be used to show model performance in details.

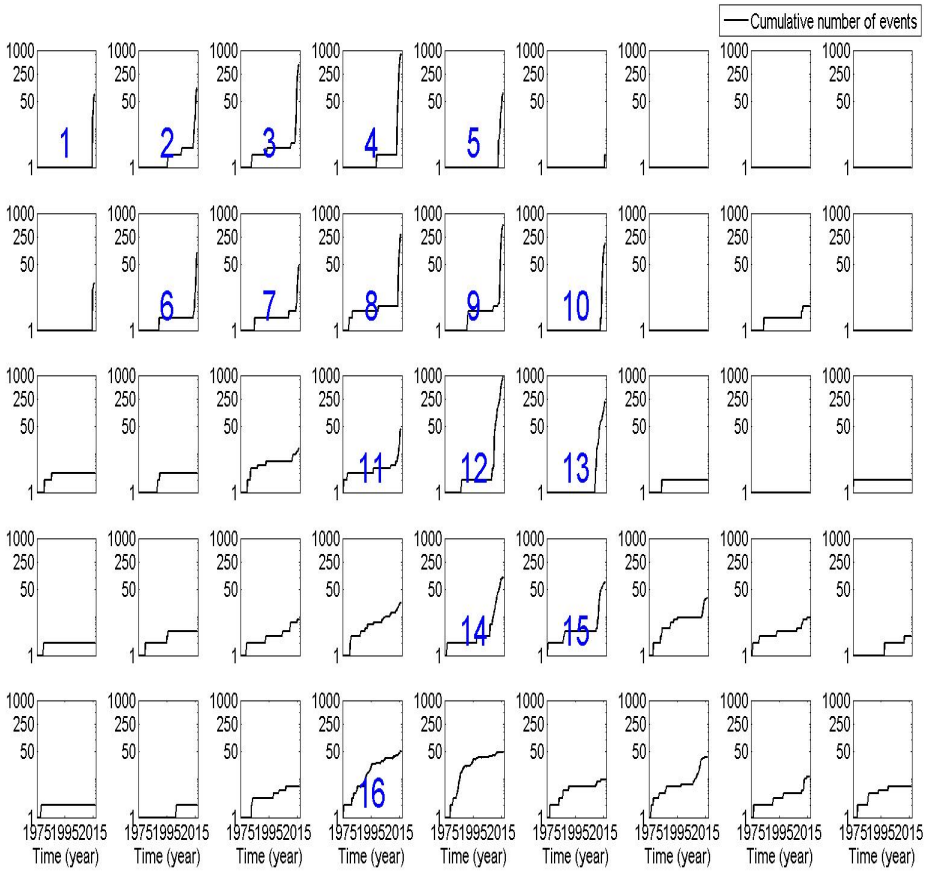

Figure 17 Cumulative number of events functions over time for each cell. As the vertical axis is displayed in log scale, 1 is added to each point on the curves to display zero values.

The values for $\{\boldsymbol{\mu_0}, \Sigma_0, \boldsymbol{\mu_A}, \Sigma_A\}$ are chosen to render broad distributions for the initial uncertainty and the transition functions. $\boldsymbol{\mu_0}$ and $\boldsymbol{\mu_A}$ are a vector of -8 and -5, respectively. Σ_0 and Σ_A are specified by two components: the standard deviation vector and the correlation matrix. For both of them, the standard deviation is assumed to be **1** and the correlation matrix is obtained through the correlation function in Eq. 22. We assume the probability of experiencing a change $P(C = 1) = 0.1$.

Without prior preference over alternative models, the optimal model should yield the highest global likelihood on the dataset. We examine a range of values for the hyper-parameter l, corresponding to different models, as shown in Table 6. The model deemed optimal is then selected to carry out the sequential inference on $\boldsymbol{\lambda}_k$.

Model	S_1	S_2	S_3	S_4	S_5	S_6
l (km)	1	50	100	150	200	250

Table 6 Summary of alternative models with different values for the hyper-parameter l.

Model Comparison and Inference Results. Figure 18 displays the log global likelihood for the model comparison. As we can see, the model performance is optimized by S_2 with $l = 50km$, which therefore is chosen for the following analysis.

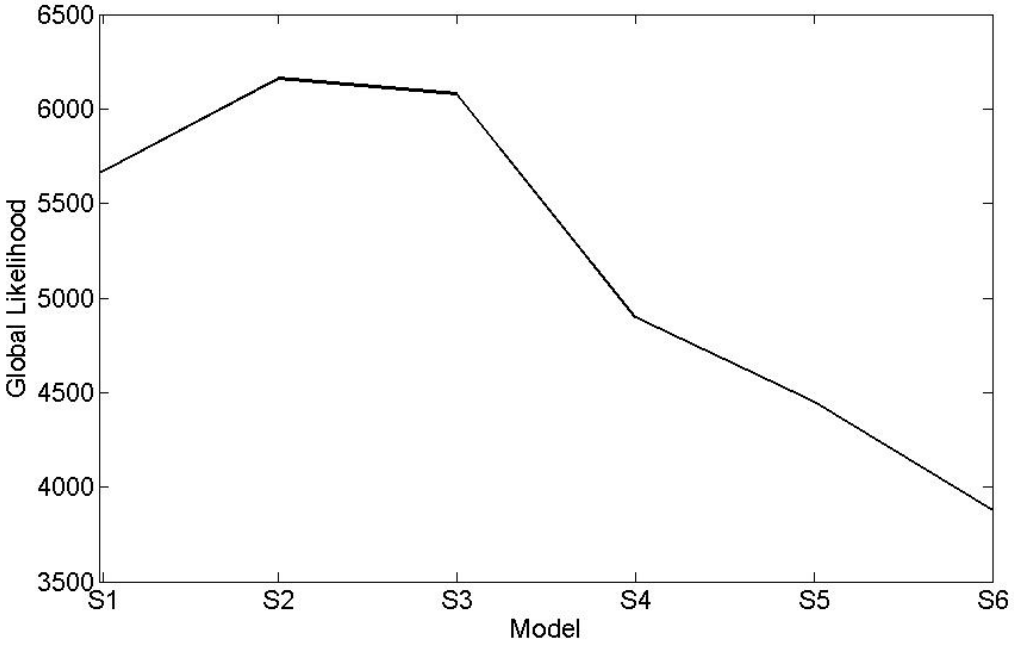

Figure 18 Log global likelihood for each alternative model.

The inference results from model S_2 are displayed in several different ways for the Oklahoma dataset. First of all, Figure 19 shows the estimated event rate as a function of time for the whole study region. It suggests the event rate experiences a sequence of increases rising from 0.01 to over 7 events per day initiating at 2010 in Oklahoma, which is consistent with results from existing literature on the earthquake activity in Oklahoma State. It is notable that the uncertainty in the inference changes with time. For a period of stationary earthquake process, the inference uncertainty reduces as more data becomes available, which is the case for the time period from 1975 to 2010. However, the

uncertainty increases right after changes occurring in the system as in 2010, 2011 and 2014. This is caused by the limited availability of data to infer the new state, but as more data becomes available later on, the uncertainty reduces until new changes in the event rate. The uptick in the estimated event rate for the end of the dataset seems in conflict with the decline in the overall observed event rate, but it can be explained by the increase in event rate in individual cells including cell 2, 3, 6 and 7, as shown in Figure 20.

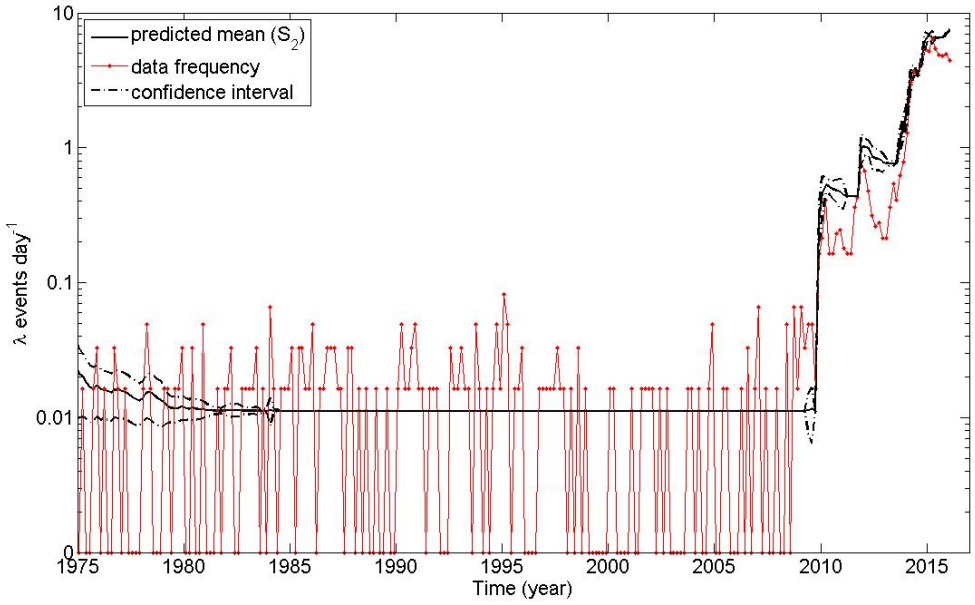

Figure 19 Average estimated event rate as a function of time for the whole study region, along with its confidence interval and the observed event rate.

Figure 20 shows how the mean estimated event rate evolves for each of the highlighting cells (a full display for each cell is in Figure F1 in Appendix F). Basically, the estimated event rate reacts not only according to the observation in the cell but also is affected by the data in the nearby cells. If the number of observed events in a cell or its surrounding cells increases significantly, the event rate in the cell is predicted to jump accordingly; if there is no increase or the increase is insignificant, the system state stays stationary. For those cells rarely seeing earthquakes, their states are also predicted to jump slightly because of the correlation dictated by the model as shown in the non-labeled cells in the full graph. Another visualization of the inference result is shown in Figure 21, which shows the mean estimated event rate in each cell for the first two months of 2016.

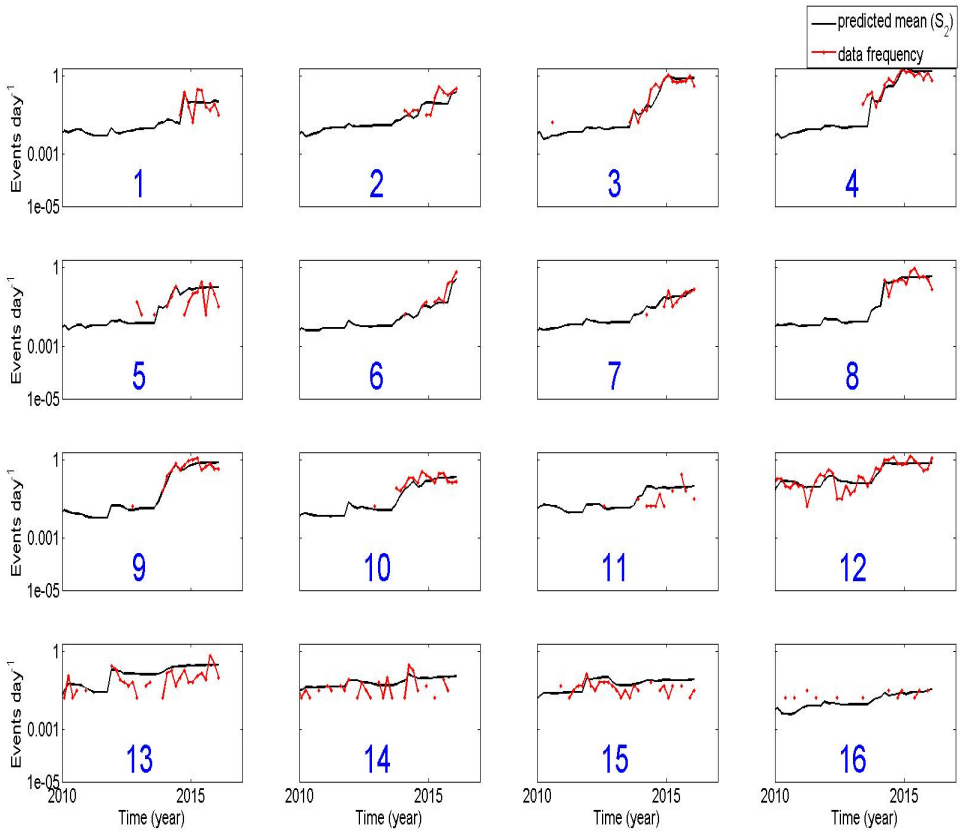

Figure 20 Average estimated event rate as a function of time for each highlighting cell along with the observed event rate.

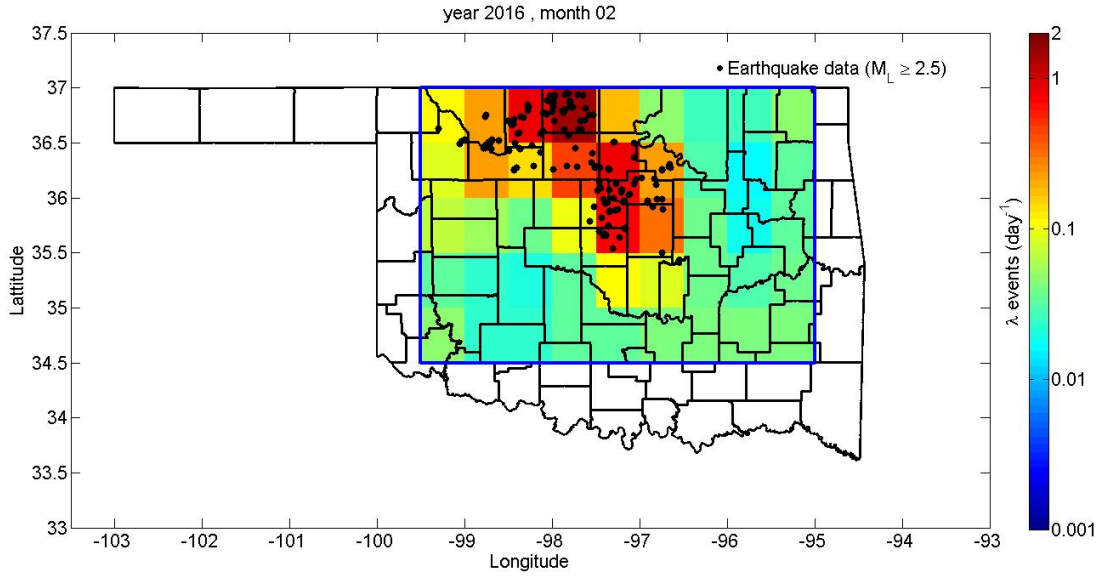

Figure 21 Mean estimated event rate in each cell for the first two months of 2016.

So far, we have shown the model's ability to sequentially update the event rate for the region based on the previous state and the current observation. It is also of interest to examine its predicting performance. First of all, we need to specify to how perform predictions for future earthquakes using the model. Unlike traditional earthquake predictions in which the prediction is carried out for an extensive future period (e.g. 50yrs) , we are focused on short term predictions due to the changing nature of induced earthquake. Specifically, the prediction is based on the average estimated event rate resulting from the current time step and it is used to predict the number of events to occur in the next time step. This prediction procedure is applied to each cell for each time step. Note that one time step is 2 months.

The predicted number of events to occur in each time step for each cell is shown in Figure 22, along with the number of observed events. Note that the display is only for the critical period from 2009 to 2016, and the highlighting cells; a display for the entire region is in Figure F2. An ideal prediction result is such that the number of observed events equals the number of predicted events, as indicted by the curve in each plot. As we can see, although the model's prediction is not ideal, generally it is satisfactory. Due to the absence of an ideal model, it is more interesting to conduct comparative analysis to evaluate the model's predicting performance in relative to "naïve models" as defined in the following.

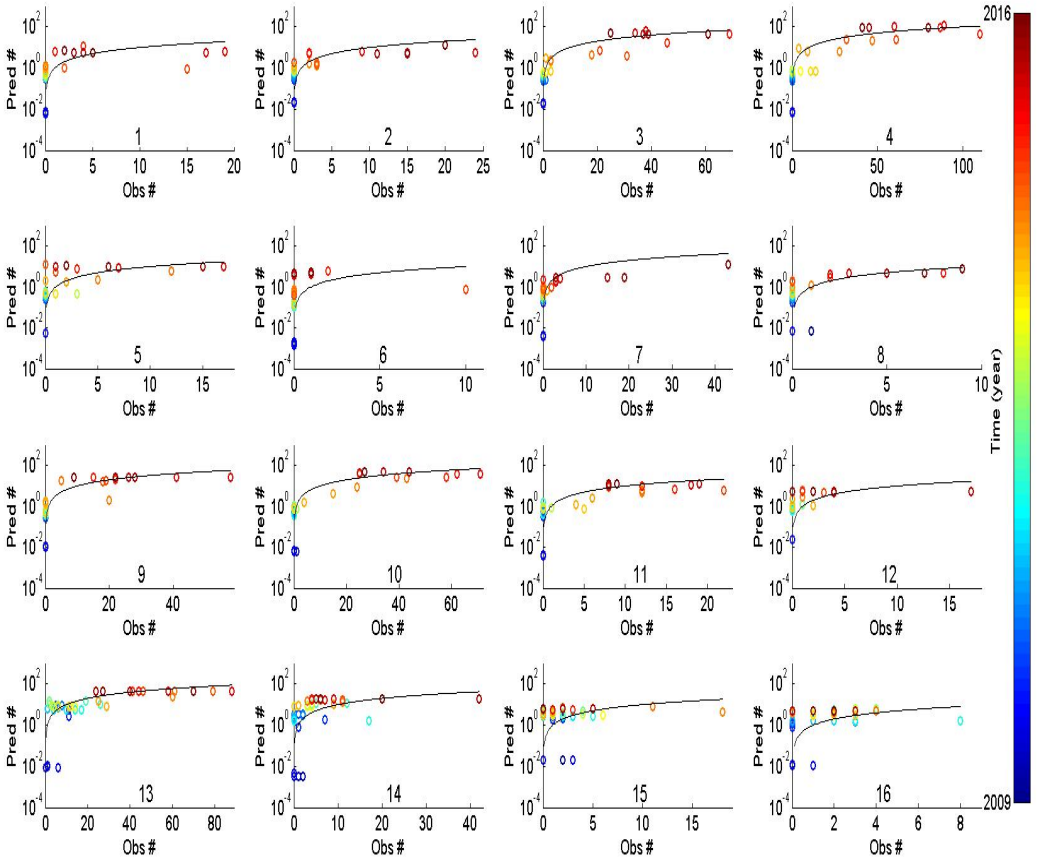

Figure 22 The predicted number of events to occur in each time step for each cell, along with the number of observed events. The timing of the prediction is indicated by the color of the circle with blue representing the beginning of the period and red for the most recent time. The curve in each subplot represents an ideal prediction, on which the predicted number of events equal the observed number.

At each time step, a naïve model uses the average event rate in the past observational period of a certain length to predict the number of events to occur in the next time step. The performance of the naïve model depends on the length chosen for the past observational period – none of the extremely short and long values, but only a moderate value can yield a good forecast. Due to this consideration, a range of values are investigated from 2 months up to 20 months to identify the optimal naïve model.

The comparison of predicting performance is based on the measurement of the predicting likelihood, which can be calculated similarly as the global likelihood. Figure 23 shows

the log ratio of the likelihood of model S_2 to that of each naïve model. A positive log ratio indicates a better predicting ability of the proposed model. As shown, regardless of the choice of the observational length, the log ratio is well above 1 in favor of our model. In another words, the proposed model is able to provide a significantly better forecast than the naïve model.

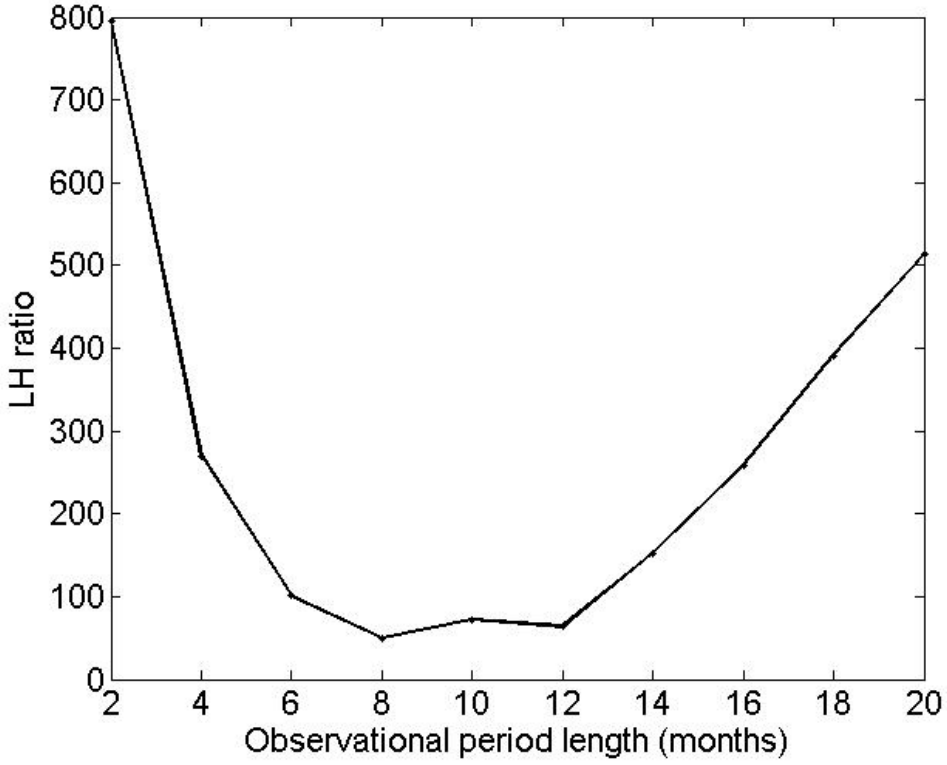

Figure 23 Log ratio of the predicting likelihood of model S_2 to that of each naïve model with a different value for the observational length.

5.4 Conclusions and Discussions

Spatially mapping event rates as a function of time is critical for providing direct inputs to temporal probabilistic seismic hazard analysis of induced earthquakes. In this chapter, we propose a general model for modeling the evolution of event rate in time and space. The model consists of spatial and temporal discretization and periodic updating of the event rate, which is based on the previous state, transition model and current observation.

Application to the Oklahoma dataset shows that the proposed method is able to well characterize the spatial distribution of the increased seismicity as a function of time. In general, a considerable part (e.g., the middle and the northern part) of the region has seen significantly increase in seismic rate. The exact timing and magnitude of the increase varies for different subareas, with the timing ranges from 2010 to 2014 and the magnitude ranges from 0.1 to 1 events per day. The model is also shown to have arguably satisfactory short-term forecasting ability, both in itself and in comparison to a naïve method.

We believe the model is able to provide insight into the underlying process of the spatial distribution of induced earthquakes. It can also be used as a tool to provide data for monitoring and periodically updating the regional seismic hazard influenced by the uncertain effect of fluid injections. Besides, its ability to perform short-term earthquake forecasting can be potentially used for guiding the operation of injection wells.

Chapter 6 Summary and Future Research

The dissertation results in three main achievements:

(1) A statistical method for the early detection of induced seismicity. The early detection of symptoms of change is critical to allow well managers and regulators to act promptly, revising the injection activity and/or preparing communities for the increased seismic risk.

(2) A general methodology using a statistical model that considers not only the time and magnitude but also the form of seismic rate transitions. The proposed method can provide a statistical basis for modeling and assessing the transitions of seismicity. With subsequent linkage to information on ground motions, exposure, fragility and consequences, the developed method can also provide an initial decision-support tool to identify areas with increasing levels of induced events, updating seismic hazard estimates, and supporting a comprehensive assessment to decide which risk-mitigation strategy should be recommended.

(3) A comprehensive method for assessing seismic transitions in both time and space. The model is able to provide insight into the underlying process of the spatial distribution of induced earthquakes. It can also be used as a tool to provide data for monitoring and periodically updating the regional seismic hazard influenced by the uncertain effect of fluid injection. Besides, its ability to perform short-term earthquake forecasting can be potentially used for the planning of the operation of injection wells.

Each achievement has its unique advantage in providing solutions to the research challenge. The early detection method is simple and computationally less expensive, and therefore it is recommended as a preliminary tool to monitor an injection site for early signs of induced seismicity. After a detection of change in seismic rate is predicted using the detection method, sophisticated methods and tools (e.g., the 2nd/3rd method) can be

employed to more closely monitor and assess the process of induced seismicity. The temporal assessment model is able to parametrically describe the transition of seismic rate, using a limited number of parameters and therefore less complicate in relative to the spatio-temporal method. If knowledge regarding the parametric form of induced seismicity is available, one would prefer a parametric model over a non-parametric model such as the proposed spatio-temporal model. When such knowledge is not available, the spatio-temporal model is recommended due to its flexibility. It is expected that such flexibility usually comes at the expense of complexity, which makes a practical solution difficult when the dimensionality of the problem (i.e. the grid resolution) is high and no analytical solution is available.

There are several open questions regarding how to use and improve the proposed methods:

- For example, we are interested in using the proposed detection method to help design an array of seismic stations in conjunction with a value of Mc (estimated with estimated Maximum Likelihood b-value) able to satisfy a pre-determined Time to Detection (yrs.) with a given probability for each criteria level of increased seismicity.
- We are also interested in using the spatio-temporal model to handle problems of high dimensionality. One possible solution to this issue might be using certain approximation technique to convert the Poisson likelihood to Gaussian likelihood and manipulating the transition model to be linear. Then Kalman Filter can be applied to the model to provide analytical solutions, requiring significantly less computation and making high dimensional problems feasible. Another possible solution is to employ parallel programming and computer clusters to provide powerful computations to meet the challenge of high dimensionality.
- In addition, it is valuable to modify the correlation function in the spatio-temporal model to greater flexibility. Currently, the correlation function is assumed universal across the study region, which does not account for variations in correlation strength for different locations.

Beyond the improvement of the proposed methods, it is important to combine them with other models such as frequency-magnitude distribution, ground motion and exposure models to evaluate seismic risk under the framework of PSHA. The combined effort is valuable for providing a decision-support tool to identify areas with increasing levels of seismic risk in a quantitative manner, supporting a comprehensive assessment to decide which risk-mitigation strategy should be recommended.

Finally, it is worth mentioning that the proposed methods can also be employed to model seismicity driven by non-stationary processes other than fluid injection, such as aseismic creep, magmatic intrusions, and natural fluid intrusion.

Data and Resources

Seismicity data were obtained from Oklahoma Geological Survey, the University of Oklahoma, at http://www.ou.edu/content/ogs/research/earthquakes/catalogs.html (last accessed March 2016). The operation history of wastewater injection wells was obtained from Oklahoma Corporation Commission Well Data System, at http://www.occpermit.com/WellBrowse (last assessed October 2014). The geological and geophysical characteristics related to the recent earthquake activity within Oklahoma, documented by the Oklahoma Geological Survey, can be found at http://wichita.ogs.ou.edu/documents/OGS_Statement-Earthquakes-4-21-15.pdf (last assessed May 2015). W.D. Heran, G.N. Green, and D.B. Stoeser's 2003 geologic map database ("A digital geologic map database for the state of Oklahoma, No. 2003-247") is available from https://pubs.er.usgs.gov:443/publication/ofr03247 (last assessed May 2015).

Reference

Ake, J., K. Mahrer, D. O'Connell, and L. Block (2005). Deep-injection and closely monitored induced seismicity at Paradox Valley, Colorado, *Bull. Seismol. Soc. Am. 95*, no. 2, 664-683.

Bachmann, C. E., S. Wiemer, J. Woessner, and S. Hainzl (2011). Statistical analysis of the induced Basel 2006 earthquake sequence: introducing a probability-based monitoring approach for Enhanced Geothermal Systems, *Geophys. J. Int.* **186,** no. 2, 793-807.

Baker, J. W. and A. Gupta (2015). A Bayesian change point model to detect changes in event occurrence rates, with application to induced seismicity, *ICASP12–12th International Conference on Applications of Statistics and Probability in Civil Engineering*.

Bommer, J. J., H. Crowley, and R. Pinho (2015). A risk-mitigation approach to the management of induced seismicity, *J. Seismol.* **19,** no.2, 623-646.

Brodsky, E. E., and L. J. Lajoie (2013). Anthropogenic seismicity rates and operational parameters at the Salton Sea Geothermal Field, *Science* **341,** doi:10.1126/science.1239213.

Chib, S., and E. Greenberg (1995). Understanding the metropolis-hastings algorithm, *The American Statistician* **49,** no. 4, 327-335.

Chu, A., F. P. Schoenberg, P. Bird, D. D. Jackson, and Y. Y. Kagan (2011). Comparison of ETAS parameter estimates across different global tectonic zones. *Bull. Seismol. Soc. Am.* **101,** no.5, 2323-2339.

Convertito, V., Maercklin, N., Sharma, N., and Zollo, A (2012). From induced seismicity to direct time-dependent seismic hazard, *Bull. Seismol. Soc. Am.* **102**, no. 6: 2563-2573.

Cornell, C. A. (1968). Engineering seismic risk analysis, *Bull. Seismo. Soc. Am.* **58**, no. 5, 1583–1606.

Daniel, G., Prono, E., Renard, F., Thouvenot, F., Hainzl, S., Marsan D., Helmstetter A. et al (2011). Changes in effective stress during the 2003–2004 Ubaye seismic swarm, France, *J. Geophys. Res.* 116, no. B1.

Ellsworth, W. L. (2013). Injection-induced earthquakes, *Science* **341,** no. 6142, doi: 10.1126/science.1225942.

Ellsworth, W. L., Hickman, S. H., Llenos, A. L., McGarr, A., Michael, A. J. and Rubinstein, J. L. (2012). Are seismicity rate changes in the midcontinent natural or manmade? presentation at the Seismological Society of America 12.

Field, E. H., Arrowsmith, R. J., Biasi, G.P., Bird, P., Dawson, T.E., Felzer, K. R., Jackson, D.D., Johnson, K.M., Jordan, T.H., Madden, C., Michael, A.J., et al. (2014). Uniform California Earthquake Rupture Forecast, Version 3 (UCERF3)—The time independent model, *Bull. Seismol. Soc. Am.* **104**, no. 3, 1122–1180.

Fletcher, R., and M. J. Powell (1963). A rapidly convergent descent method for minimization, *Comput. J.* 6, no. 2, 163-168.

Frohlich, C., and S. D. Davis (1993). Teleseismic b values; or, much ado about 1.0, *J. Geophys. Res.* **98,** no. B1, 631–644.

Frohlich, C., W. Ellsworth, W. A. Brown, M. Brunt, J. Luetgert, T. MacDonald, and S. Walter (2014). The 17 May 2012 M4. 8 earthquake near Timpson, East Texas: An event possibly triggered by fluid injection, *J. Geophys. Res.* **119,** no. B1, 581-593.

Gardner, J. K., and L. Knopoff (1974). Is the sequence of earthquakes in southern California, with aftershocks removed, Poissonian, *Bull. Seismol. Soc. Am.* 64, no. 5, 1363-1367.

Gischig, V. S., and S. Wiemer (2013). A stochastic model for induced seismicity based on non-linear pressure diffusion and irreversible permeability enhancement, *Geophys. J. Int.* 194*,* no. 2, 1229-1249.

Goertz-Allmann, B. P. & Wiemer, S. (2012). Geomechanical modeling of induced seismicity source parameters and implications for seismic hazard assessment. Geophysics, 78(1), KS25-KS39.

Gregory, P. (2005). *Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with Mathematica® Support*, Cambridge University Press.

Gutenberg, B., and C. F. Richter (1944). Frequency of earthquakes in California, *Bull. Seismol. Soc. Am***. 34,** no. 4, 185-188.

Hainzl, S., and Y. Ogata (2005). Detecting fluid signals in seismicity data through statistical earthquake modeling, *J. Geophys. Res.* **110,** no. B5, doi:10.1029/2004JB003247.

Healy, J. H., W. W. Rubey, D. T. Griggs and C. B. Raleigh (1968). The Denver earthquakes, *Science* **161,** no. 3848, 1301-1310.

Heran, W. D., G. N. Green, and D. B. Stoeser (2003). A digital geologic map database for the state of Oklahoma, No. 2003-247.

Hickman, S., R. Sibson, and R. Bruhn (1995). Introduction to special section: Mechanical involvement of fluids in faulting, *J. Geophys. Res.* 100, no. B7, 12831-12840.

Horton, S. (2012). Disposal of hydrofracking waste fluid by injection into subsurface aquifers triggers earthquake swarm in central Arkansas with potential for damaging earthquake, *Seismol. Res. Lett.* 83, no. 2, 250-260.

Hough, S. E., and M. Page (2015). A century of induced earthquakes in Oklahoma?,*Bull. Seismol. Soc. Am.***105**, no. 6, 2863-2870.

Julier, Simon J., and Jeffrey K. Uhlmann (1996). A general method for approximating nonlinear transformations of probability distributions. Technical report, Robotics Research Group, Department of Engineering Science, University of Oxford.

Julier, Simon J., Jeffrey K. Uhlmann, and Hugh F. Durrant-Whyte (1995). A new approach for filtering nonlinear systems, *American Control Conference, Proceedings of the 1995*. Vol. 3.

Karvounis, D. C., V. S. Gischig, and S. Wiemer (2014). Towards a real-time forecast of induced seismicity for enhanced geothermal systems, *SHALE ENERGY ENGINEERING 2014*, 246.

Keranen, K. M., H. M. Savage, G. A. Abers, and E. S. Cochran (2013). Potentially induced earthquakes in Oklahoma, USA: Links between wastewater injection and the 2011 Mw 5.7 earthquake sequence, *Geology* 41, no. 6, 699-702.

Kumazawa, T., and Y. Ogata (2013). Quantitative description of induced seismic activity before and after the 2011 Tohoku-Oki earthquake by nonstationary ETAS models, *J. Geophys. Res.* **118,** no. B12, 6165-6182.

Lawrence, M., McCann, M., Ostenaa, D., Wong, I., Unruh, J., Hanson, K., Olig, S., Clague, J., LaForge, R., Lettis, W., Swan, B., Zachariasen, J., Youngs, R. and Addo, K. (2014), The BC Hydro SSHAC Level 3 seismic source model, *Proc. of the 10th Natl. Conf. on Earthquake Eng., Earthquake Eng. Res. Inst.*, Anchorage, AK.

Lehmann, E. L., and J. P. Romano (2006). *Testing statistical hypotheses*, Springer Science & Business Media.

Lei, X., G. Yu, S. Ma, X. Wen, and Q. Wang (2008). Earthquakes induced by water injection at ~3 km depth within the Rongchang gas field, Chonquing, China, *J. Geophys. Res.* **113,** no. B10, doi:10.1029/2008JB005604.

Liu, J. S., and Chen, R. (1998). Sequential Monte Carlo methods for dynamic systems, *Journal of the American statistical association* **93**, no. 443, 1032-1044.

Liu, J. S., Chen, R. and Logvinenko, T. (2001). A theoretical framework for sequential importance sampling with resampling, *Sequential Monte Carlo methods in practice*, 225-246, Springer New York.

Llenos, A. L., and A. J. Michael (2013). Modeling Earthquake Rate Changes in Oklahoma and Arkansas: Possible Signatures of Induced Seismicity, *Bull. Seismol. Soc. Am.* **103,** no. 5, 2850-2861.

Lombardi, A. M., M. Cocco, and W. Marzocchi (2010). On the Increase of Background Seismicity Rate during the 1997–1998 Umbria-Marche, Central Italy, Sequence: Apparent Variation or Fluid-Driven Triggering?, *Bull. Seismol. Soc. Am.* 100, no. 3, 1138-1152.

MacKay, D. J. (2003). *Information theory, inference, and learning algorithms*, Cambridge University Press, Vol. 7.

Majer, E., J. Nelson, A. Robertson-Tait, J. Savy, and I. Wong (2012). *Protocol for addressing induced seismicity associated with enhanced geothermal systems*, US Department of Energy.

Marsan, D., E. Prono, and A. Helmstetter (2013). Monitoring aseismic forcing in fault zones using earthquake time series, *Bull. Seismol. Soc. Am.* **103,** no. 1, 169-179, doi:10.1785/0120110304.

Matsu'ura, R. S., and I. Karakama (2005). A point process analysis of the Matsushiro earthquake swarm sequence: The effect of water on earthquake occurrence, *Pure Appl. Geophys.* **162,** no. 6-7, 1319-1345.

McClure, M. W., and R. N. Horne (2011). Investigation of injection-induced seismicity using a coupled fluid flow and rate/state friction model, *Geophysics* **76,** no. 6, WC181-WC198.

McGarr, A. (2014). Maximum magnitude earthquakes induced by fluid injection, *J. Geophys. Res.* **119,** no. B2, 1008-1019.

McGarr, A., et al (2015). Coping with earthquakes induced by fluid injection, *Science* **347**, no. 6224, 830-831.

McGuire, R. K. (2004). Seismic Hazard and Risk Analysis, *Earthquake Eng. Res. Inst.*, Oakland, Calif.

McNamara, D. E., H. M. Benz, R. B. Hermann, E. A. Bergman, P. Earle, A. Holland, R. Baldwin, and A. Gassner (2015). Earthquake hypocenters and focal mechanisms in central Oklahoma reveal a complex system of reactivated subsurface strike-slip faulting, *Geophys. Res. Lett.* **42**, no. 8, 2742-2749, doi: 10.1002/2014GL062730.

Mena, B., S. Wiemer, and C. Bachmann (2013). Building robust models to forecast the induced seismicity related to geothermal reservoir enhancement, *Bull. Seismol. Soc. Am.* **103,** no. 1, 383-393.

Moschetti, M. P., et al. (2015). Seismic source characterization for the 2014 update of the U.S. national seismic hazard model, *Earthquake Spectra* **31**, S1, S31–S57.

Musso, C., Oudjane, N., and LeGland, F. (2001). Improving regularised particle filters, *Sequential Monte Carlo methods in practice*, 247-271, Springer New York.

Northcutt, R. A., and J. A. Campbell (1995). Geologic provinces of Oklahoma, *Oklahoma Geological Survey, Open-File Report OF 5-95*, available at http://www.ogs.ou.edu/geolmapping/Geologic_Provinces_OF5-95.pdf (last accessed May 2015).

Northcutt, R. A., and J. A. Campbell (1996). Geologic provinces of Oklahoma, 128-134.

Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes, *J. Am. Stat. Assoc.* 83, no. 401, 9-27.

Ogata, Y. (1998). Space-time point-process models for earthquake occurrences, *Annals of the Institute of Statistical Mathematics* **50**, no. 2, 379-402.

Ogata, Y. and Zhuang, J. (2006). Space–time ETAS models and an improved extension, *Tectonophysics* **413**, no. 1, 13-23.

Petersen, M. D., C. S. Mueller, M. P. Moschetti, S. M. Hoover, J. L. Rubinstein, A. L. Llenos, A. J. Michael, W. L. Ellsworth, A. F. McGarr, A. A. Holland, and J. G. Anderson (2015). Incorporating induced seismicity in the 2014 United States National Seismic

Hazard Model: results of the 2014 workshop and sensitivity studies, *US Geological Survey*, no. 2015-1070.

Rasmussen, J. G. (2013). Bayesian inference for Hawkes processes, *Meth. Comput. Appl. Probab.* **15,** no. 3, 623-642.

Reasenberg, P. (1985). Second-order moment of central California seismicity, 1969–1982, *J. Geophys. Res.* 90, no. B7, 5479-5495.

Rubinstein, J. L., W. L. Ellsworth, A. McGarr, and H. M. Benz (2014). The 2001–Present Induced Earthquake Sequence in the Raton Basin of Northern New Mexico and Southern Colorado, *Bull. Seismol. Soc. Am.* **104**, no. 5, doi:10.1785/0120140009.

Seeber, L., J. G. Armbruster, and W. Kim (2004). A fluid-injection-triggered earthquake sequence in Ashtabula, Ohio: Implications for seismogenesis in stable continental regions, *Bull. Seismol. Soc. Am.* 94, no. 1, 76-87.

Sims, P. K., R. W. Saltus, and E. D. Anderson (2008). Precambrian basement structure map of the continental United States-An interpretation of geologic and aeromagnetic data, *U.S. Geol. Surv. Scientif. Investig. 3012*, scale: 1:8,000,000. .

Streit, J. E., and R. R. Hillis (2004). Estimating fault stability and sustainable fluid pressures for underground storage of CO2 in porous rock, *Energy* 29, no. 9, 1445-1456.

Talwani, P., and S. Acree (1984). Pore pressure diffusion and the mechanism of reservoir-induced seismicity, *Pure Appl. Geophys.* 122*,* no. 6, 947-965.

Touati, S., M. Naylor, and I. G. Main (2014). Statistical modeling of the 1997–1998 Colfiorito earthquake sequence: Locating a stationary solution within parameter uncertainty, *Bull. Seismol. Soc. Am.* 104, no. 2, 885-897.

Touati, S., M. Naylor, I. G. Main, and M. Christie (2011). Masking of earthquake triggering behavior by a high background rate and implications for epidemic-type aftershock sequence inversions, *J. Geophys. Res.* 116, no. B3.

Utsu, T., Y. Ogata and R. S. Matsu'ura (1995). The centenary of the Omori formula for a decay law of aftershock activity, *J. Phys. Earth* **43,** no. 1, 1-33.

Wang, P., M. Pozzi, M. J. Small, and W. Harbert (2015). Statistical method for early detection of changes in seismic rate associated with wastewater injections, *Bull. Seismol. Soc. Am.* **105,** no. 6, doi: 10.1785/0120150038.

Wang, P., Small, M. J., Harbert, W. and Pozzi, M. (2016). A Bayesian Approach for Assessing Seismic Transitions Associated with Wastewater Injections, *Bull. Seismol. Soc. Am.***106**, no. 3, 832-845.

Wiemer, S. (2001). A software package to analyze seismicity: ZMAP, *Seismol. Res. Lett.* 72, no. 3, 373-382.

Woessner, J., and S. Wiemer (2005). Assessing the quality of earthquake catalogues: Estimating the magnitude of completeness and its uncertainty, *Bull. Seismol. Soc. Am.* 95, no. 2, 684-698.

Zhuang, J., Y. Ogata, and D. Vere-Jones (2002). Stochastic declustering of space-time earthquake occurrences, *J. Am. Stat. Assoc.* 97, no. 458, 369-380.

Appendices

Appendix A

To investigate the effect of different values of Mc, we applied the detection algorithm to three Mc scenarios - Mc=2.0, 2.5, or 3.0, and repeat the p-value analysis with the test origin at the beginning of 2009. The evolution of the 95th percentile of the p-value for each Mc scenario is shown in Figure A1. As shown, when Mc is M2.0, the detection is achieved later than when Mc equals to M2.5; while if it is M3.0, the detection is made earlier.

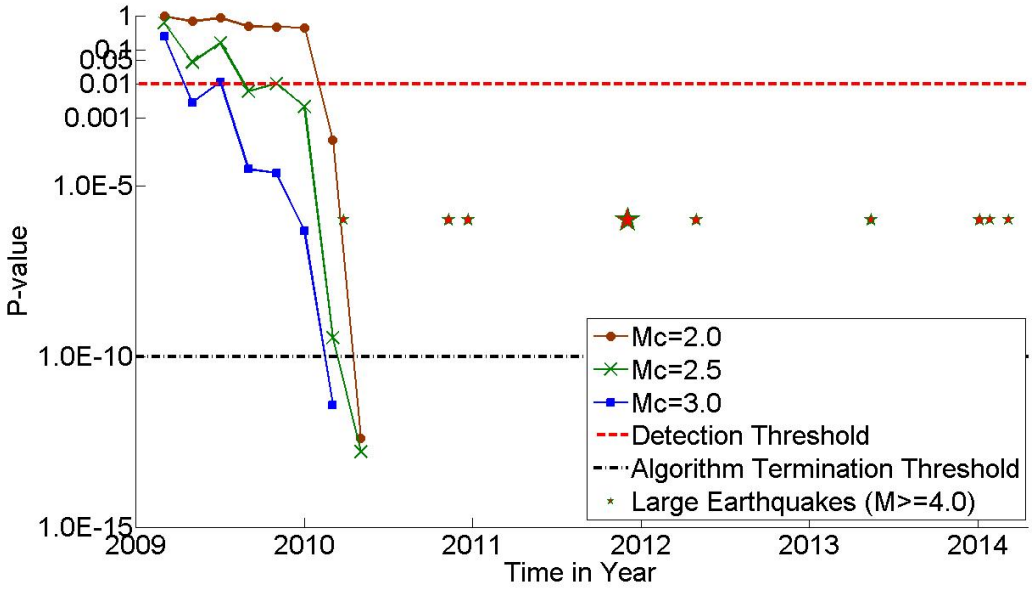

Figure A1 Evolution of the 95th percentile of the p-value under different Mc scenarios, with the test origin at the beginning of 2009. The dash line indicates the critical value of 0.01 for the p-value, while the dash dot line shows the threshold of 1×10^{-10} for the p-value, below which the hypothesis-testing algorithm terminates evaluation.

To evaluate the effect of different transition dates from the baseline to the testing period, we move the assumed transition point from the beginning of the year 2009 to the beginning of 2000, 2005, 2007, or 2008, and repeat the p-value analysis shown in Figure 4. In each scenario, the baseline period starts at the beginning of the catalogue and ends at the transition point to the test period. The evolution of the p-value for each test origin

scenario is shown in Figure A2. As shown, regardless of the choice for the point of time to initiate the testing algorithm, the p-value does not decline significantly (e.g. below 0.01 as indicated by the dash line) until the year 2009 or beyond.

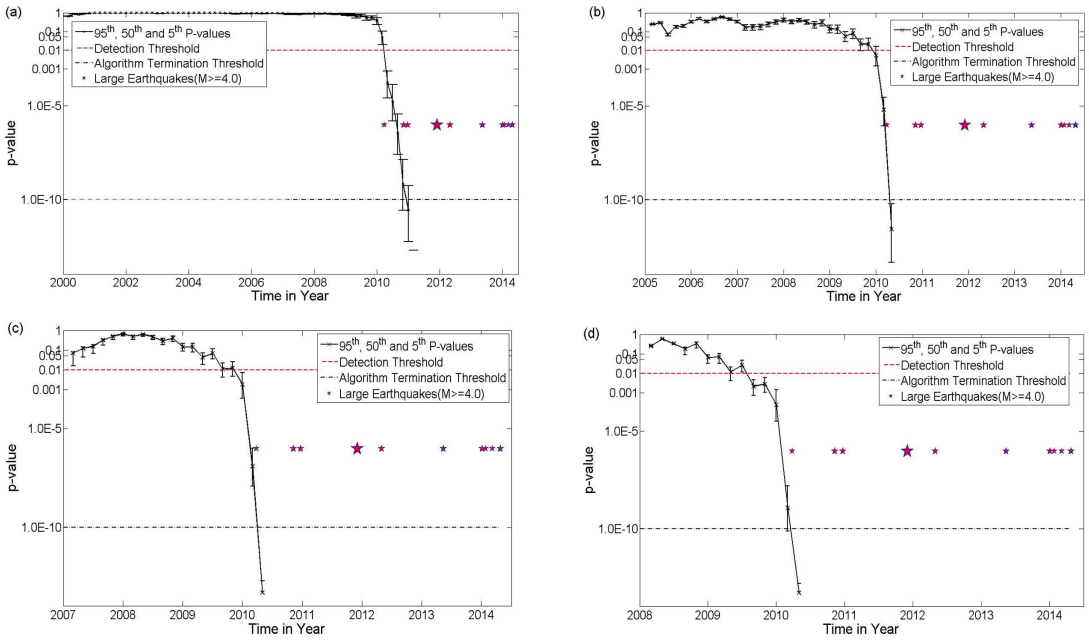

Figure A2 P-value evolutions under distinct time origins for the test period. Plots (a) to (d) display the p-value evolution for the test origin at Jan. 2000, Jan. 2005, Jan. 2007 and Jan. 2008, respectively. The dash dot line shows the threshold of 1×10^{-10} for the p-value, below which the hypothesis-testing algorithm terminates evaluation.

Appendix B. Parallel Tempering MCMC & Model Comparison Results

In Bayesian model comparison, since the calculation of the global likelihood p(H|E) for alternative models, is analytically infeasible for the model considered in this study, we resort to parallel tempering MCMC (Gregory, 2005) to approximate its value.

To evaluate p(H|E) using parallel tempering MCMC, we first define a partition function

$$Z(\beta) = \int d\theta\, p(\theta|E)p(H|E,\theta)^{\beta} = \int d\theta\, \exp\{\ln[p(\theta|E)] + \beta\ln[p(H|E,\theta)]\} \quad \text{(B1)}$$

where β is the tempering parameter [0, 1]. Now take the derivative of $\ln[Z(\beta)]$

$$\frac{d}{d\beta}\ln[Z(\beta)] = \frac{1}{Z(\beta)}\frac{d}{d\beta}Z(\beta) \quad \text{(B2)}$$

$$\frac{d}{d\beta}Z(\beta) = \int d\theta\, \ln[p(H|E,\theta)]\times\exp\{\ln[p(\theta|E)] + \beta\ln[p(H|E,\theta)]\} = \int d\theta\, \ln[p(H|E,\theta)]p(\theta|E)p(H|E,\theta)^{\beta} \quad \text{(B3)}$$

Substituting Eq. B3 into Eq. B2, we obtain

$$\frac{d}{d\beta}\ln[Z(\beta)] = \frac{\int d\theta\, \ln[p(H|E,\theta)]p(\theta|E)p(H|E,\theta)^{\beta}}{\int d\theta\, p(\theta|E)p(H|E,\theta)^{\beta}} = \mathbb{E}_{\beta}(\ln[p(H|E,\theta)]) \quad \text{(B4)}$$

where $\mathbb{E}_{\beta}(\ln[p(H|E,\theta)])$ is the expectation value of the $\ln[p(H|E,\theta)]$. This quantity can be evaluated from the MCMC results which consist of sets of parameter θ_t samples, one set for each value of the tempering parameter β. Let $\{\theta_{t,\beta}\}$ represent the samples for tempering parameter β.

$$\mathbb{E}_{\beta}(\ln[p(H|E,\theta)]) = \frac{1}{n}\sum_t \ln\left[p(H|E,\theta_{t,\beta})\right], \quad \text{(B5)}$$

where n is the number of samples in each set after the burn-in period. From Eq. B4 we can write

$$\int_0^1 \mathrm{d}\ln[Z(\beta)] = \ln[Z(1)] - \ln[Z(0)] = \int \mathrm{d}\beta\ \mathbb{E}_\beta(\ln[p(H|E,\theta)]) \quad \text{(B6)}$$

Now from Eq. B1

$$Z(1) = \int \mathrm{d}\theta\, p(\theta|E)p(H|E,\theta)^1 = p(H|E), \quad \text{(B7)}$$

and

$$Z(0) = \int \mathrm{d}\theta\, p(\theta|E). \quad \text{(B8)}$$

From Eq. B6, B7, and B8 we can write

$$\ln[p(H|E)] = \ln[Z(0)] + \int \mathrm{d}\beta\ \mathbb{E}_\beta(\ln[p(H|E,\theta)])\ . \quad \text{(B9)}$$

For a proper prior, $Z(0) = 1$ and Eq. B9 reduces to

$$\ln[p(H|E)] = \int \mathrm{d}\beta\ \mathbb{E}_\beta(\ln[p(H|E,\theta)]) \quad \text{(B10)}$$

In comparing alternative models for the Oklahoma dataset, the right side of Eq. B10 is approximated by first evaluating Eq. A5 for a set of different values of β (here the set is {0.01, 0.2575, 0.505, 0.7525, 1}), based on the whole Oklahoma dataset of magnitude Mc and above. Subsequently, an interpolating function is generated using MATLAB and integrated over the interval [0 1]. Thus p(H|E) can be solved and the posterior probability ratios of the alternative models can be calculated. The detailed results of the model comparison are displayed as follows in Table B1. Figure B1 displays the expected value of λ0 and λ as a function of time for posterior alternative model E0-6, according to the Oklahoma dataset, as well as their characteristic fit as a function of transformed times τi.

Model Name	$\mathbb{E}_{\beta}(ln[p(H\|E,\theta)])$					Global Likelihood	Posterior Probability Ratio (O_{i0})
	β=0.01	β=0.2575	β=0.505	β=0.7525	β=1		
E_0	-2406.9	-1715.9	-1697.7	-1693.4	-1688.4	-1791.3	1
E_1	-2378.5	-1614.2	-1596.5	-1588.0	-1587.4	-1698.3	2.0×10^{44}
E_2	-2390.3	-1588.4	-1572.0	-1567.2	-1565.0	-1679.4	4.0×10^{48}
E_3	-2351.0	-1585.0	-1556.5	-1547.5	-1543.8	-1662.0	1.4×10^{56}
E_4	-2410.5	-1474.9	-1461.2	-1456.9	-1455.7	-1585.1	3.6×10^{89}
E_5	-2451.0	-1473.5	-1447.9	-1443.6	-1440.3	-1581.4	1.4×10^{91}
E_6	-2479.2	-1477.0	-1447.8	-1440.0	-1438.1	-1584.7	5.3×10^{89}

Table B1 Evaluated Eq. B5 for different values of β and the corresponding global likelihood and posterior probability ratios Oi0 (in relative to the stationary ETAS model) for each alternative model based on the Oklahoma dataset.

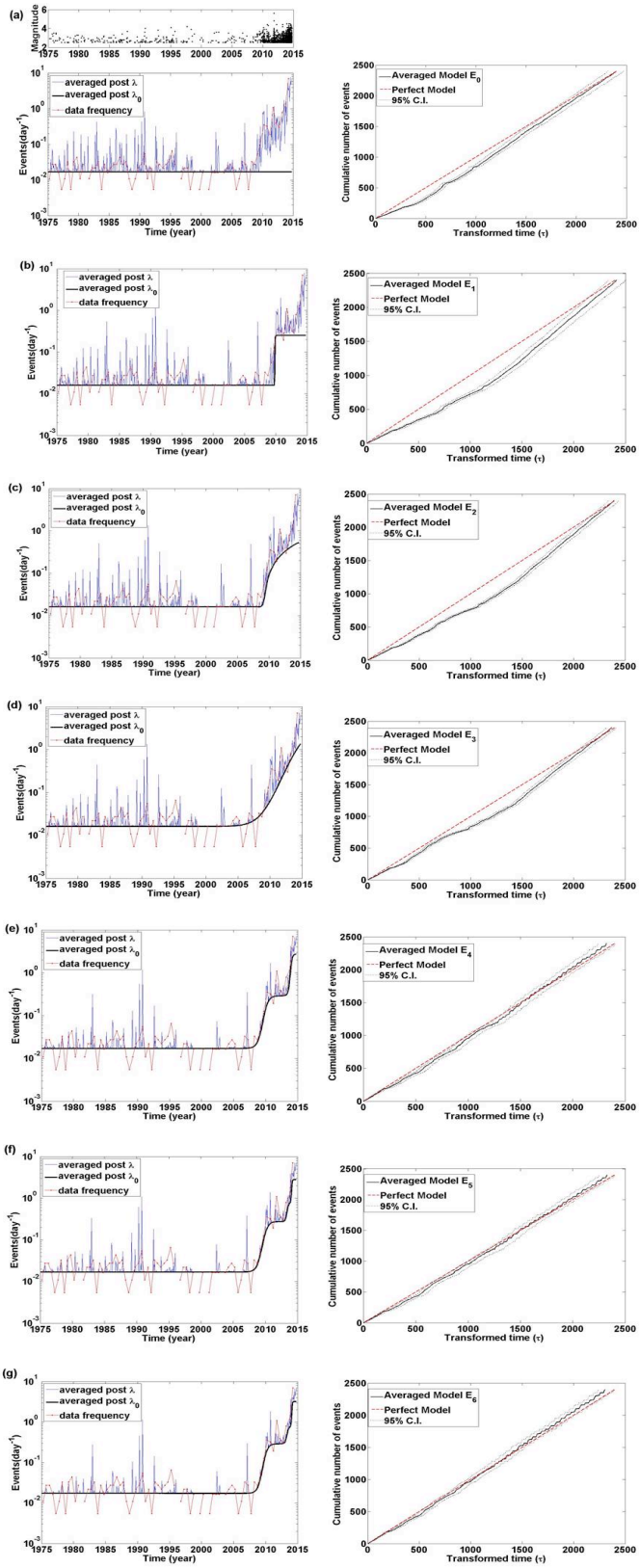

Figure B1 Expected values of λ_0 as a function of time on the left, and the characteristic fit as a function of transformed times τ_i on the right, for model E_0 - E_6 in (a) – (g), respectively.

Appendix C. Prior for Parameters of Alternative Models in Chapter 4 and Inference of model E_5

Priors are an important component of the Bayesian methodology: if a considerable amount of prior knowledge is available, an informative prior should be used on the parameters; while in the case of absence a non-informative prior is more appropriate. In analyzing the Oklahoma dataset, we base the choice of the prior on literature (Chu et al., 2011) for the basic parameters of the ETAS model (i.e., $\{\mu_0, K_0, \alpha_{ETAS}, c, p\}$); and that for the induced seismicity parameters (i.e., $\{\Delta, t_d, \eta\}$) is based on expertise. Considering parameter t_d as an example, its prior is set broadly distributed but not completely flat, as it is barely reasonable to assign equal probability to all times in the infinite future for the activation of induced seismicity. The statistics of the prior adopted in this study for the parameters of each alternative model is documented in Table C1.

A full exhibition of the selected samples from the posterior parameter distribution of model E_5 is shown in Figure C1, as well as the corresponding complete correlation coefficient matrix in Table C2.

Parameters		Model			
		E_0	E_1	E_2	$E_{3\text{-}6}$
μ_0 (Events/day)	mean	0.05	0.05	0.05	0.05
	coefficient of variation	2	2	2	2
K_0 (Events/Day)	mean	0.001	0.001	0.001	0.001
	coefficient of variation	5	5	5	5
α	mean	2	2	2	2
	coefficient of variation	2	2	2	2
c (Days)	mean	0.001	0.001	0.001	0.001
	coefficient of variation	2	2	2	2
p	mean	1	1	1	1
	coefficient of variation	1	1	1	1
t_d(years)	mean	NA	20	20	20
	coefficient of variation	NA	2	2	2
Δ (Events/day)	mean	NA	0.5	0.5	0.5
	coefficient of variation	NA	0.5	0.5	0.5
η (day^{-1})	mean	NA	NA	0.0001	0.01
	coefficient of variation	NA	NA	10	2

Table C1 The statistics of the prior adopted in this study for the parameters of each alternative model. Note that for models $E_{3\text{-}6}$, each set of induced seismicity parameters $\{t_{di}, \Delta_i, \eta_i\}$ shares a common prior represented by $\{t_d, \Delta, \eta\}$ in the table.

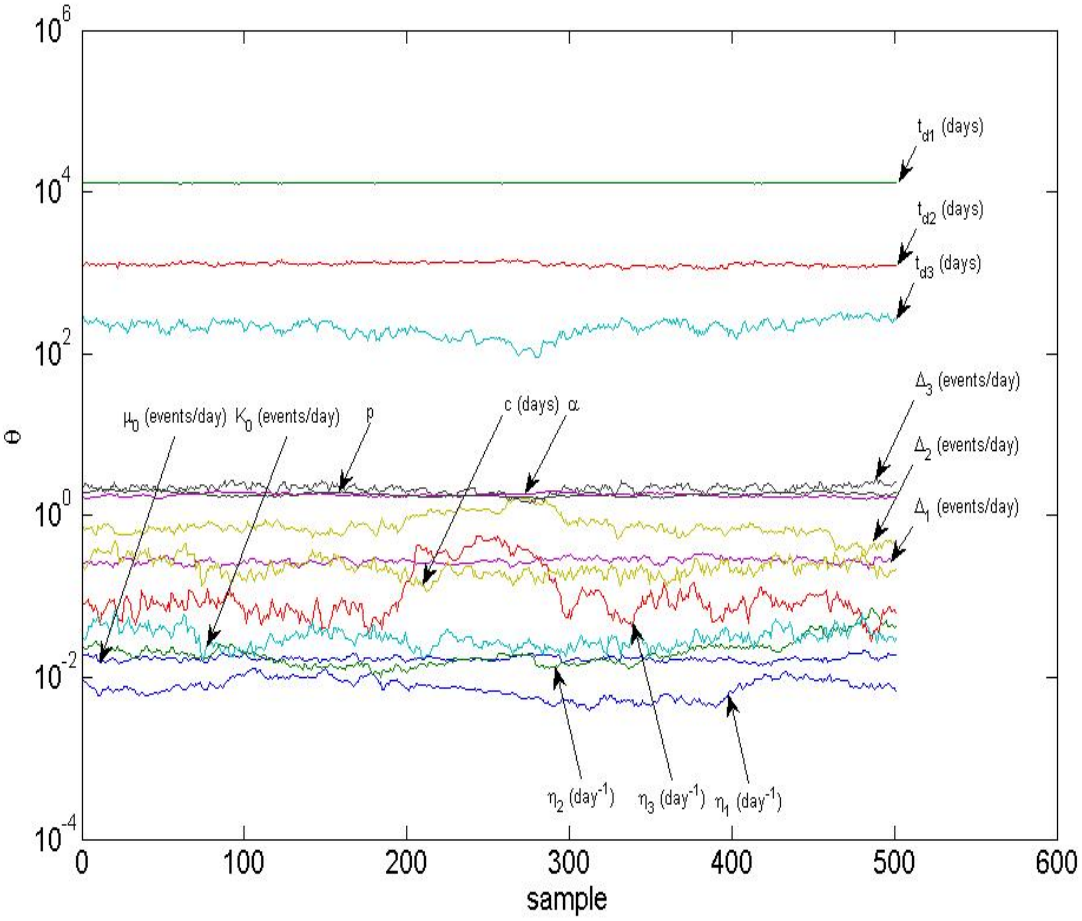

Figure C1 Selected samples from the posterior distribution of model E_5 parameters, as the representation of the uncertainties of model E_5.

Corr. Coeff. Matrix	μ_0 (Events/day)	t_{d1} (years)	Δ_1 (Events/day)	η_1 ($days^{-1}$)	t_{d2} (years)	Δ_2 (Events/day)	η_2 ($days^{-1}$)	t_{d3} (years)	Δ_3 (Events/day)	η_3($days^{-1}$)	K_0(Events/day)	α	c (days)	p
μ_0 (Events/day)	1	0.04	-0.21	-0.07	-0.04	0.09	0.23	0.00	0.00	0.04	-0.05	0.03	-0.10	-0.17
t_{d1} (years)		1	0.54	-0.60	-0.73	0.06	-0.07	-0.12	-0.15	-0.09	-0.14	0.15	-0.13	-0.16
Δ_1 (Events/day)			1	-0.29	-0.38	-0.14	-0.02	-0.10	0.01	-0.19	-0.12	-0.06	-0.12	0.00
η_1 ($days^{-1}$)				1	0.30	-0.31	0.11	0.26	0.32	-0.21	0.18	-0.17	0.17	0.22
t_{d2} (years)					1	0.52	-0.35	-0.58	-0.36	0.50	-0.12	0.09	-0.13	-0.16
Δ_2 (Events/day)						1	-0.46	-0.80	-0.75	0.78	-0.43	0.38	-0.41	-0.52
η_2 ($days^{-1}$)							1	0.58	0.38	-0.22	0.29	-0.49	0.18	0.28
t_{d3} (years)								1	0.69	-0.59	0.32	-0.29	0.31	0.40
Δ_3 (Events/day)									1	-0.58	0.12	-0.17	0.16	0.30
η_3($days^{-1}$)										1	-0.30	0.14	-0.34	-0.43
K_0(Events/day)											1	-0.54	0.93	0.73
α												1	-0.32	-0.51
c (days)													1	0.79
p														1

Table C2 Complete correlation coefficient matrix for the parameters of model E5.

Appendix D. Recursive Procedure for the Sequential Bayesian Updating Process

This appendix shows briefly how a sequential Bayesian updating process can be implemented recursively. Let us consider to estimate the system state **x** at time *k* based on the previous state and all the available observations. In Bayesian paradigm, this problem can be formulated by $p(\mathbf{x}_k|\boldsymbol{H}_{1:k})$ where $\boldsymbol{H}_{1:k}$ stands for the set of available observations.

Using Bayes law, $p(\mathbf{x}_k|\boldsymbol{H}_{1:k})$ can be expanded as

$$p(\mathbf{x}_k|\boldsymbol{H}_{1:k}) = \frac{p(\boldsymbol{H}_{1:k}|\mathbf{x}_k)p(\mathbf{x}_k)}{p(\boldsymbol{H}_{1:k})} \tag{D1}$$

Because the set $\boldsymbol{H}_{1:k}$ can be written as $\{\boldsymbol{H}_k, \boldsymbol{H}_{1:k-1}\}$, Eq. D1 can be rewritten as

$$p(\mathbf{x}_k|\boldsymbol{H}_{1:k}) = \frac{p(\boldsymbol{H}_k,\boldsymbol{H}_{1:k-1}|\mathbf{x}_k)p(\mathbf{x}_k)}{p(\boldsymbol{H}_k,\boldsymbol{H}_{1:k-1})} \tag{D2}$$

Using the chain rule of probability, Eq. D2 becomes

$$p(\mathbf{x}_k|\boldsymbol{H}_{1:k}) = \frac{p(\boldsymbol{H}_k|\,\boldsymbol{H}_{1:k-1},\mathbf{x}_k)p(\boldsymbol{H}_{1:k-1}|\mathbf{x}_k)p(\mathbf{x}_k)}{p(\boldsymbol{H}_k|\boldsymbol{H}_{1:k-1})p(\boldsymbol{H}_{1:k-1})} \tag{D3}$$

Applying Bayes law to $p(\boldsymbol{H}_{1:k-1}|\mathbf{x}_k)$, and reducing the resulting equations, yields the following progression

$$\begin{aligned} p(\mathbf{x}_k|\boldsymbol{H}_{1:k}) &= \frac{p(\boldsymbol{H}_k|\,\boldsymbol{H}_{1:k-1},\mathbf{x}_k)p(\mathbf{x}_k|\boldsymbol{H}_{1:k-1})p(\boldsymbol{H}_{1:k-1})p(\mathbf{x}_k)}{p(\boldsymbol{H}_k|\boldsymbol{H}_{1:k-1})p(\boldsymbol{H}_{1:k-1})p(\mathbf{x}_k)} \\ &= \frac{p(\boldsymbol{H}_k|\,\boldsymbol{H}_{1:k-1},\mathbf{x}_k)p(\mathbf{x}_k|\boldsymbol{H}_{1:k-1})}{p(\boldsymbol{H}_k|\boldsymbol{H}_{1:k-1})} \\ &= \frac{p(\boldsymbol{H}_k|\mathbf{x}_k)p(\mathbf{x}_k|\boldsymbol{H}_{1:k-1})}{p(\boldsymbol{H}_k|\boldsymbol{H}_{1:k-1})} \end{aligned} \tag{D4}$$

where $p(\boldsymbol{H}_k|\,\boldsymbol{H}_{1:k-1},\mathbf{x}_k) \rightarrow p(\boldsymbol{H}_k|\mathbf{x}_k)$ because observation at time t_k is assumed to be only dependent on the current state $\mathbf{x}_k$.

One last step is needed to create a completely recursive form for the conditional probability density function equations. The Chapman-Kolmogorov equation provides a

link between the *prior density*, defined as $p(\mathbf{x}_k|\boldsymbol{H}_{1:k-1})$, and the previous posterior density

$$p(\mathbf{x}_k|\boldsymbol{H}_{1:k-1}) = \int p(\mathbf{x}_k|\mathbf{x}_{k-1}, \boldsymbol{H}_{1:k-1}) p(\mathbf{x}_{k-1}|\boldsymbol{H}_{1:k-1})\, d\mathbf{x}_{k-1}$$
$$= \int p(\mathbf{x}_k|\mathbf{x}_{k-1}) p(\mathbf{x}_{k-1}|\boldsymbol{H}_{1:k-1})\, d\mathbf{x}_{k-1} \quad \text{(D5)}$$

where $p(\mathbf{x}_k|\mathbf{x}_{k-1}, \boldsymbol{H}_{1:k-1}) \rightarrow p(\mathbf{x}_k|\mathbf{x}_{k-1})$ is due to the property of the first-order Markov process.

In light of this, Eq. D4 can be rewritten as

$$p(\mathbf{x}_k|\boldsymbol{H}_{1:k}) = \frac{p(\boldsymbol{H}_k|\mathbf{x}_k) \int p(\mathbf{x}_k|\mathbf{x}_{k-1}) p(\mathbf{x}_{k-1}|\boldsymbol{H}_{1:k-1}) d\mathbf{x}_{k-1}}{p(\boldsymbol{H}_k|\boldsymbol{H}_{1:k-1})} \quad \text{(D6)}$$

Now, from Eq. D6, a recursive link has been established between the previous posterior $p(\mathbf{x}_{k-1}|\boldsymbol{H}_{1:k-1})$ and the current posterior $p(\mathbf{x}_k|\boldsymbol{H}_{1:k})$ that requires the specification of the *predictive density* given by $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ and the *likelihood function* $p(\boldsymbol{H}_k|\mathbf{x}_k)$.

Appendix E. Sigma Point Integration Method

Appendix E describes the Sigma Point Integration Method for approximating Guassian-weighted integrals. To introduce the method, let us first consider a general multidimensional Gaussian-weighted integral

$$\hat{\mathbf{f}}(\mathbf{x}) = \int \mathbf{f}(\mathbf{x})\, \mathcal{N}(\mathbf{x}; \hat{\mathbf{x}}, \boldsymbol{\Sigma}) d\mathbf{x} \quad \text{(E1)}$$

Apply affine transformation

$$\mathbf{x} = \hat{\mathbf{x}} + \mathbf{L}\mathbf{c} \quad \text{(E2)}$$

where $\mathbf{L}$ is defined by the matrix square root equation

$$\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}^{\mathbf{T}} \quad \text{(E3)}$$

As a result, $\mathbf{c} \sim \mathcal{N}(\mathbf{c}; \mathbf{0}, \mathbf{I})$ and Eq. E1 can be rewritten as

$$\hat{\mathbf{f}}(\mathbf{x}) = \int \tilde{\mathbf{f}}(\mathbf{c})\, \mathcal{N}(\mathbf{c}; \mathbf{0}, \mathbf{I}) d\mathbf{c} \quad (\mathbf{E}4)$$

For a general nonlinear function $\tilde{\mathbf{f}}(\mathbf{c})$, the integral in Eq. B4 cannot be solved analytically. Numerical methods to solve integrals of this kind are called *multiple integration rules* and each rule is designed to integrate a specific class of multidimensional polynomial approximations of $\tilde{\mathbf{f}}(\mathbf{c})$. The application of affine transformation is very important as it allows for evaluating integrals weighted by a Gaussian density that is symmetric about zero in all dimensions. This can be shown to greatly simplify numerical methods for approximating Gaussian-weighted integrals of nonlinear functions. The integration rules that involve polynomial approximations typically require calculation of Jacobian and Hessian differential matrices, which itself is a difficult task. An alternative is to replace the differential matrices by their multidimensional central finite difference approximations, which leads to a class of methods called the Sigma Point method.

Basically, Sigma Point method approximate the integral in Eq. E4 by a discrete sum of the form

$$\hat{\mathbf{f}}(\mathbf{x}) \cong \sum_{j=1}^{n_s} w_j \tilde{\mathbf{f}}\left(\mathbf{c}^{(j)}\right) \quad \text{(E5)}$$

The exact form of w_j, $\mathbf{c}^{(j)}$ and n_s depends on the choice of the integration rules. The value of n_s typically depends on the dimension of the state vector n_x. The integration rule used in this study is as follows

$$w_j = \begin{cases} w_0, \ j = 0 \\ \frac{1-w_0}{2n_x}, \ j = 1,2,\dots,2n_x \end{cases} \quad \text{(E6)}$$

$$\mathbf{c}^{(j)} = \sqrt{\frac{n_x}{1-w_0}}\mathbf{r}^{(j)}, j = 0,1,\dots,2n_x \quad \text{(E7)}$$

where $\mathbf{r}^{(j)}$ is a unit vector along one of dimensional axes, such as $[0,\dots,1,\dots,0]^T$, except for $\mathbf{r}^{(0)} = \mathbf{0}$.

From Eq. E2 and E4, for each vector integration point $\mathbf{c}^{(j)}$ we can write

$$\tilde{\mathbf{f}}\left(\mathbf{c}^{(j)}\right) = \mathbf{f}\left(\hat{\mathbf{x}} + \mathbf{L}\mathbf{c}^{(j)}\right), j = 0,1,\dots,2n_x \quad \text{(E8)}$$

Defining the sigma points as

$$\boldsymbol{\chi}^{(j)} = \hat{\mathbf{x}} + \mathbf{L}\mathbf{c}^{(j)}, j = 0,1,\dots,2n_x \quad \text{(E9)}$$

the approximate integral in Eq. E5 becomes

$$\hat{\mathbf{f}}(\mathbf{x}) \cong \sum_{j=0}^{2n_x} w_j \mathbf{f}\left(\boldsymbol{\chi}^{(j)}\right) \quad \text{(E10)}$$

Appendix F. Full Inference and Forecasting Results.

Appendix F contains the full display of the inference and forecast results of the proposed model in Chapter 5.

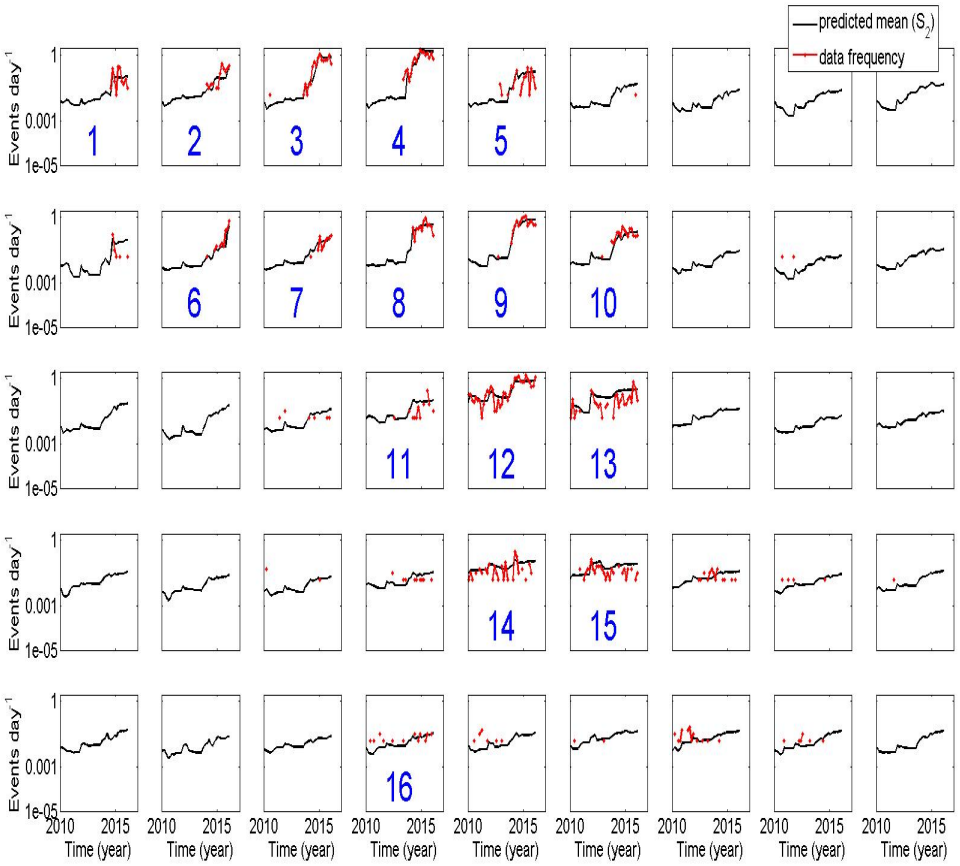

Figure F1 Average estimated event rate as a function of time for each cell along with the observed event rate.

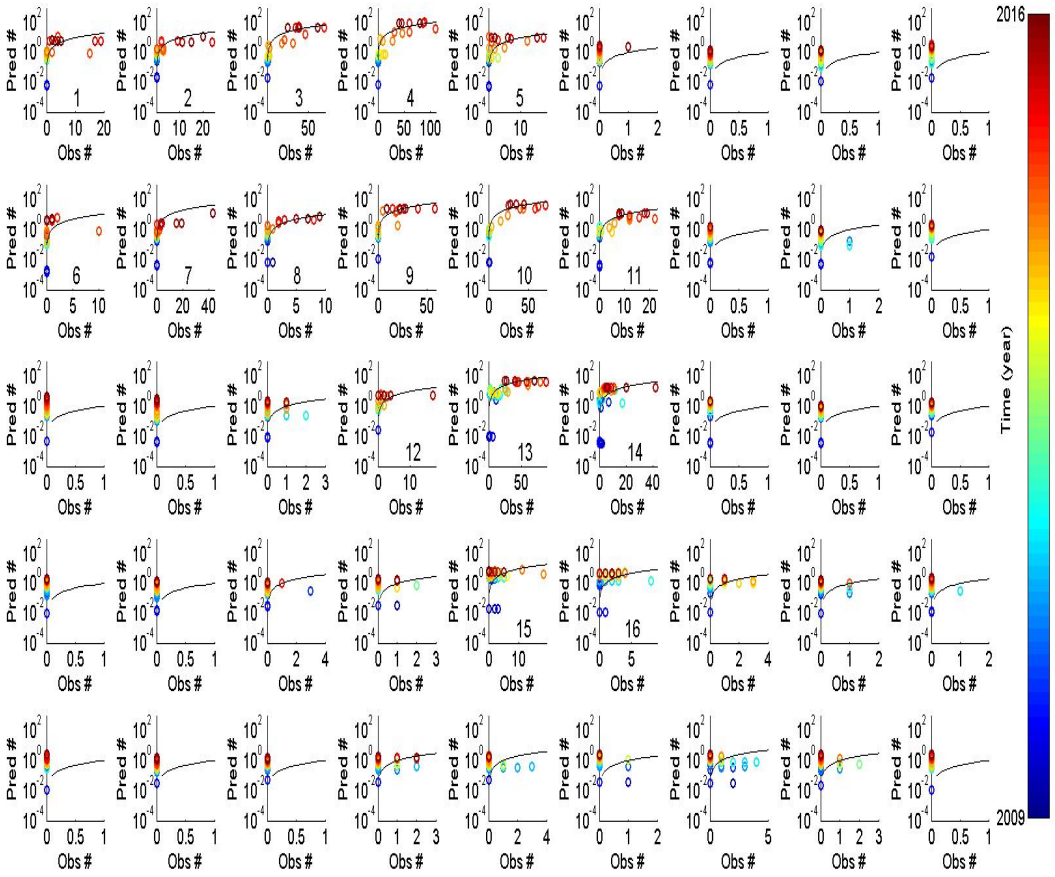

Figure F2 The predicted number of events to occur in each time step for each cell, along with the number of observed events. The timing of the prediction is indicated by the color of the circle with blue representing the beginning of the period and red for the most recent time. The curve in each subplot represents an ideal prediction, on which the predicted number of events equal the observed number.