Estimating Spatially Varying Event Rates with a Change Point using Bayesian Statistics: Application to Induced Seismicity

Abhineet Gupta[a], Jack W. Baker[a]

[a] *Civil & Environmental Engineering Department, Stanford University, Stanford, CA 94305, USA*

Abstract

We describe a model to estimate event rates of a non-homogeneous spatio-temporal Poisson process. A Bayesian change point model is described to detect changes in temporal rates. The model is used to estimate whether a change in event rates occurred for a process at a given location, the time of change, and the event rates before and after the change. To estimate spatially varying rates, the space is divided into a grid and event rates are estimated using the change point model at each grid point. The spatial smoothing parameter for rate estimation is optimized using a likelihood comparison approach. An example is provided for earthquake occurrence in Oklahoma, where induced seismicity has caused a change in the frequency of earthquakes in some parts of the state. Seismicity rates estimated using this model are critical components for hazard assessment, which is used to estimate seismic risk to structures. Additionally, the time of change in seismicity can be used as a decision support tool by operators or regulators of activities that affect

Email addresses: `abhineet@stanford.edu` (Abhineet Gupta), `bakerjw@stanford.edu` (Jack W. Baker)

seismicity.

Keywords: change point method, Bayesian inference, induced seismicity, spatio-temporal process

1. Introduction

In this paper, we estimate the rates of a non-homogeneous spatio-temporal Poisson process. The rates vary spatially with the possibility of an independent temporal change at any point in space. We use a Bayesian estimation approach and describe a change point model to detect temporal changes. We describe a likelihood comparison methodology to estimate spatially-varying event rates using the change point model. The results from the model are regions of estimated change, times of change, and spatially varying event rates. The model is demonstrated through an application to induced seismicity in Oklahoma.

Similar approaches for change detection have been used previously, for example, a Bayesian model was developed for Poisson processes to assess changes in intervals between coal-mining disasters [1]. A model was proposed to detect early changes in seismicity rates based on earthquake declustering and hypothesis testing [2]. While there is some precedence, the problem described in this paper is different than the previous ones because the event rates vary spatially in addition to the possibility of a temporal change. Estimation of these spatially varying rates requires an appropriate rate smoothing procedure, which is also described here.

The motivation for this paper is the significant increase in seismicity that has been recently observed in the Central and Eastern US (CEUS) [3]. For

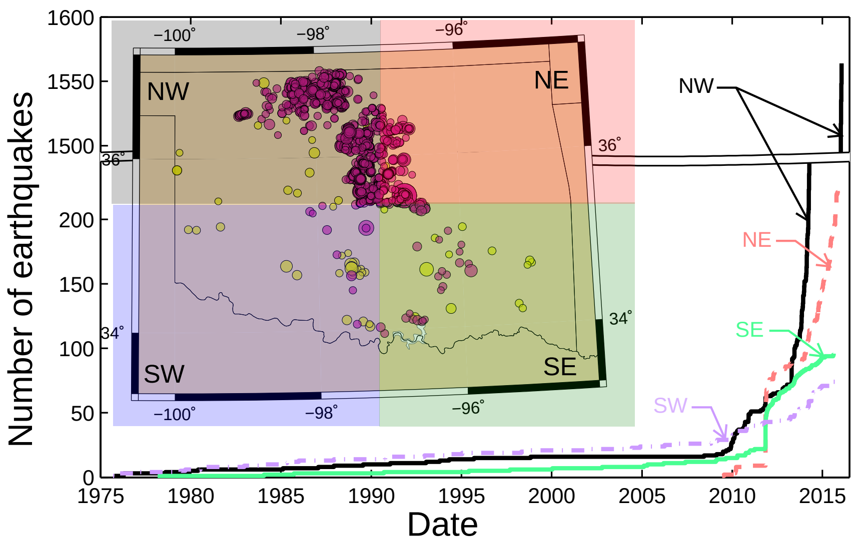

Figure 1: Cumulative number of earthquakes in four quadrants of Oklahoma with magnitude ≥ 3 from 1974 through Dec 31, 2015. The earthquakes post 2008 are shown in pink on the map, and the size of the circles is proportional to the earthquake magnitude. We have omitted the western panhandle of Oklahoma in this and all subsequent maps, since no seismicity increase has been observed in this region, and to draw focus to the remainder of the state.

example in 2014 and 2015, more earthquakes were observed in Oklahoma than in California. There is a possibility that this increased seismicity is a result of underground wastewater injection [e.g., 3, 4, 5]. Seismicity generated as a result of human activities is referred to as induced or triggered seismicity. Figure 1 shows the cumulative number of earthquakes with magnitude ≥ 3 since 1974 for four quadrants of Oklahoma. There is a significant increase in seismicity rate starting around 2008, though the date and magnitude of rate increase varies among the different regions. Hence, the times of change and the seismicity rates need to be estimated individually for this spatio-temporal process.

There is a need to understand and manage the induced seismicity hazard and risk [6, 7]. The increased seismicity due to anthropogenic processes affects the safety of buildings and infrastructure, especially since seismic loading has historically not been the predominant design force in most CEUS regions. This makes the seismicity rate a critical component for hazard assessment [8]. The work in this paper will aid in effective risk assessment through better future prediction of earthquakes in a local region using the estimated spatially-varying seismicity rates. These rates would aid in development of hazard maps, which are commonly used to estimate the seismic loading during the structural design process. Additionally, identifying changes in seismicity rates can be used as a decision support tool by stakeholders and regulators to monitor and manage the seismic impacts of human activities [2].

The structure of the paper is divided into the description of the model and its application on induced seismicity. In section 2, we describe a Bayesian change point model that is used to identify changes in event rates, and to estimate the event rates before and after the change. In section 3, we present a methodology to estimate event rates for a spatio-temporal non-homogeneous Poisson process. In section 4, we apply this methodology to estimate spatially-varying earthquake rates in Oklahoma. In section 5, we address some model limitations with examples from the application in Oklahoma.

2. Bayesian model for change point detection

In this section, we describe a Bayesian change point model to detect changes in event rates for a non-homogeneous Poisson process with one change point. We also describe the algorithmic implementation of the model.

2.1. Model

A Bayesian change point model to detect a change in event rates is described by [1] and [9]. This model uses time between events to detect a change in rates. Given a dataset of inter-event times, the Bayes factor [10] is calculated to indicate whether a change in event rates occurred. The Bayes factor is defined here as the ratio of the likelihood of a model with no change to the likelihood of a change point model, given the observed data.

$$B_{01}(\boldsymbol{t}) = \frac{\mathcal{L}(H_0 \mid \boldsymbol{t})}{\mathcal{L}(H_1 \mid \boldsymbol{t})} \tag{1}$$

where $B_{01}(\boldsymbol{t})$ is the Bayes factor, $\boldsymbol{t}$ is a vector of inter-event times, and H_0 and H_1 represent the models with no change and a change, respectively. $\mathcal{L}(H \mid \boldsymbol{t})$ defines the likelihood of model H given some observed data $\boldsymbol{t}$. The two models, H_0 and H_1, are described below and the final formulation of the equation to calculate the Bayes factor is given later in equation 21.

Values smaller than one for the Bayes factor indicate that the model with change is favored over the model with no change. The threshold value of the Bayes factor that indicates strong preference for one or the other model can be selected based on the the required degree of confidence, but typically values less than 0.01 or larger than 100 are used to favor one or the other model. If a change is detected in the data, the time of change and event rates before and after the change are subsequently calculated.

For a sequence of events in a non-homogeneous Poisson process with a single change, the unknown variables of interest are the time of change τ, the event rate before the change λ_1, and the event rate after the change λ_2.

$$\lambda(s) = \begin{cases} \lambda_1, & 0 \leq s \leq \tau \\ \lambda_2, & \tau < s \leq T \end{cases} \tag{2}$$

where the observation period for events is defined as $[0, T]$. Assume that the zeroth event in the event sequence occurs at time 0 and the n^{th} event occurs at time T. The inter-event times are defined as

$$\boldsymbol{t} = \{t_1, t_2, \ldots, t_n\} \quad s.t. \quad \sum_i t_i = T \tag{3}$$

where t_i denotes the time between occurrences of the $i - 1^{th}$ and the i^{th} events.

Since the events follow a Poisson distribution with different rates before and after the change, the inter-event times are exponentially distributed and can be expressed as

$$f^X_{\lambda(s)}(x) = \lambda(s)\mathrm{e}^{-\lambda(s)\,x} \tag{4}$$

where $f^X(x)$ denotes a probability distribution function of X, $\lambda(s)$ is the parameter for the distribution (the event rate), and X is the random variable (the inter-event time).

For the Bayesian framework, conjugate priors are defined for λ_j as gamma distributions with parameters k_j and θ_j [11]. Then the prior probability distribution of the rates $\pi(\lambda_j)$ is written as

$$\pi(\lambda_j) \propto \lambda_j^{k_j-1} \mathrm{e}^{-\lambda_j/\theta_j} \tag{5}$$

where $\propto$ is the proportionality symbol.

The time of change τ is assumed to be equally likely at any time during the observation period. Hence, its prior $\pi(\tau)$ is assumed to be uniformly distributed.

$$\pi(\tau) = \frac{1}{T}, \quad 0 \leq \tau \leq T \tag{6}$$

The likelihood function $\mathcal{L}$ for the unknown parameters $\{\tau, \lambda_1, \lambda_2\}$ given the inter-event times $\boldsymbol{t}$ is written as the product of the probability distributions for events following the Poisson distribution, and occurring before and after time τ.

$$\mathcal{L}(\tau, \lambda_1, \lambda_2 \mid \boldsymbol{t}) = \lambda_1^{N(\tau)} \mathrm{e}^{-\lambda_1 \tau} \lambda_2^{N(T)-N(\tau)} \mathrm{e}^{-\lambda_2 (T-\tau)} \tag{7}$$

where $N(t)$ represents the number of events between $[0, t]$. Assume that the time of change τ, event rate before change λ_1, and event rate after change λ_2, are mutually independent. Then the posterior density $\pi(\tau, \lambda_1, \lambda_2 \mid \boldsymbol{t})$ for all the unknown parameters is calculated as

$$\begin{aligned} \pi(\tau, \lambda_1, \lambda_2 \mid \boldsymbol{t}) &\propto \mathcal{L}(\tau, \lambda_1, \lambda_2 \mid \boldsymbol{t}) \pi(\lambda_1, \lambda_2, \tau) \\ &= \mathcal{L}(\tau, \lambda_1, \lambda_2 \mid \boldsymbol{t}) \pi(\lambda_1) \pi(\lambda_2) \pi(\tau) \end{aligned} \tag{8}$$

The marginal distributions for each of τ, λ_1, and λ_2 are obtained by integrating the above posterior density over the remaining two variables. The marginal posterior distribution of τ is calculated as

$$\begin{aligned}
\pi(\tau \mid \boldsymbol{t}) &\propto \int_0^\infty \int_0^\infty \pi(\lambda_1, \lambda_2, \tau \mid \boldsymbol{t}) \, \mathrm{d}\lambda_1 \, \mathrm{d}\lambda_2 \\
&= \pi(\tau) \int_0^\infty \left(\int_0^\infty \lambda_1^{N(\tau)+k_1-1} \mathrm{e}^{-\lambda_1\left(\tau+\frac{1}{\theta_1}\right)} \, \mathrm{d}\lambda_1 \right) \lambda_2^{N(T)-N(\tau)+k_2-1} \mathrm{e}^{-\lambda_2\left(T-\tau+\frac{1}{\theta_2}\right)} \, \mathrm{d}\lambda_2 \\
&= \frac{1}{T} \cdot \frac{\Gamma(r_1(\tau))\Gamma(r_2(\tau))}{S_1(\tau)^{r_1(\tau)} S_2(\tau)^{r_2(\tau)}} \qquad (9)
\end{aligned}$$

where $\Gamma(x)$ is the gamma function, and

$$\begin{array}{lllllll}
r_1(\tau) & = & N(\tau) + k_1 & & S_1(\tau) & = & \tau + \frac{1}{\theta_1} \\
r_2(\tau) & = & N(T) - N(\tau) + k_2 & & S_2(\tau) & = & T - \tau + \frac{1}{\theta_2}
\end{array} \qquad (10)$$

Equation 9 is written in log space for implementation of the algorithm, described in section 2.2.

$$\begin{aligned}
\log \pi(\tau \mid \boldsymbol{t}) \propto & -\log T + \log\left(\Gamma(r_1(\tau))\right) + \log\left(\Gamma(r_2(\tau))\right) \\
& - r_1(\tau) \log\left(S_1(\tau)\right) - r_2(\tau) \log\left(S_2(\tau)\right) \qquad (11)
\end{aligned}$$

Similarly, the marginal distribution of λ_1 is calculated as shown below. A closed form solution for integration over τ does not exist. Hence, to evaluate the probability distribution, the time range is discretized over a uniform Δt and summed over to approximate the marginal distribution.

$$\begin{aligned}
\pi(\lambda_1 \mid \boldsymbol{t}) &\propto \int_0^T \int_0^\infty \pi(\lambda_1, \lambda_2, \tau \mid \boldsymbol{t}) \, \mathrm{d}\lambda_2 \, \mathrm{d}\tau \\
&= \int_0^T \left(\int_0^\infty \lambda_2^{r_2(\tau)-1} \mathrm{e}^{-\lambda_2 S_2(\tau)} \, \mathrm{d}\lambda_2 \right) \pi(\tau) \lambda_1^{r_1(\tau)-1} \mathrm{e}^{-\lambda_1 S_1(\tau)} \, \mathrm{d}\tau \\
&\approx \sum_{\tau=0}^{T} \frac{1}{T} \lambda_1^{r_1(\tau)-1} \mathrm{e}^{-\lambda_1 S_1(\tau)} \Gamma(r_2(\tau)) S_2(\tau)^{r_2(\tau)} \qquad (12)
\end{aligned}$$

This equation is also converted to log domain for algorithmic implementation.

$$\log \pi(\lambda_1 \mid \boldsymbol{t}) \propto \log \left(\sum_{\tau=0}^{T} \mathrm{e}^{z_1} \right) \tag{13}$$

where

$$z_1 = -\log T + (r_1(\tau) - 1) \log \lambda_1 - \lambda_1 S_1(\tau) + \log \left[\Gamma(r_2(\tau)) \right] + r_2(\tau) \log \left(S_2(\tau) \right) \tag{14}$$

The marginal distribution of λ_2 is calculated and approximated similarly as

$$\pi(\lambda_2 \mid \boldsymbol{t}) \propto \sum_{\tau=0}^{T} \frac{1}{T} \lambda_2^{r_2(\tau)-1} \mathrm{e}^{-\lambda_2 S_2(\tau)} \Gamma(r_1(\tau)) S_1(\tau)^{r_1(\tau)} \tag{15}$$

and

$$\log \pi(\lambda_2 \mid \boldsymbol{t}) \propto \log \left(\sum_{\tau=0}^{T} \mathrm{e}^{z_2} \right) \tag{16}$$

where

$$z_2 = -\log T + (r_2(\tau) - 1) \log \lambda_2 - \lambda_2 S_2(\tau) + \log \left[\Gamma(r_1(\tau)) \right] + r_1(\tau) \log \left(S_1(\tau) \right) \tag{17}$$

We now describe the constant rate model. For a sequence of events in a homogeneous Poisson process with no change, the unknown variable of interest is the event rate λ_0. Assume this event rate has a gamma distribution prior with parameters k_0 and θ_0, similar to the prior for parameters λ_1 and λ_2. Then the posterior distribution of the event rate $\pi(\lambda_0 \mid \boldsymbol{t})$ follows the gamma distribution with the following parameters [11]

$$k_{posterior} = k_0 + N(T) \qquad \theta_{posterior} = \left(\tfrac{1}{\theta_0} + T \right)^{-1} \tag{18}$$

With the above results, the Bayes factor can be calculated. The likelihood of the change point model, H_1, given the observed events is obtained by integrating the posterior distribution given in equation 8 over τ, λ_1, and λ_2.

$$\mathcal{L}(H_1 \mid \boldsymbol{t}) = \int_0^\infty \int_0^\infty \int_0^T \pi(\tau, \lambda_1, \lambda_2 \mid \boldsymbol{t})\,\mathrm{d}\tau\,\mathrm{d}\lambda_1\,\mathrm{d}\lambda_2 \tag{19}$$

The likelihood of a constant rate model H_0 given the observed events is similarly obtained as

$$\mathcal{L}(H_0 \mid \boldsymbol{t}) \propto \int_0^\infty \lambda_0^{k_0+N(T)-1} \mathrm{e}^{-\lambda_0(1/\theta_0+T)}\,\mathrm{d}\lambda_0 \tag{20}$$

Equations 19 and 20 each require a proportionality factor, and this factor is different for the two equations. Hence, in the calculation for the Bayes factor, we multiply the ratio of likelihoods with a constant term $c(T)$, to correctly convert the proportionality in the likelihood calculations. When $\pi(\tau) = 1/T$, and $k_1 = k_2$, $c(T)$ can be computed by equating the Bayes factor to 1 for a boundary condition of a single event occurring half-way through the observation period [1]. If the value of parameters for the gamma conjugate priors are $k_j = 0.5$ and $\theta_j \to \infty$ for $j = 0, 1, 2$, then the Bayes factor can be written as [1]

$$B_{01}(\boldsymbol{t}) = \frac{4\sqrt{\pi} T^{-n} \Gamma(n + 1/2)}{\sum_{\tau=0}^{T} \Gamma(r_1(\tau))\Gamma(r_2(\tau)) S_1(\tau)^{-r_1(\tau)} S_2(\tau)^{-r_2(\tau)}} \tag{21}$$

2.2. Algorithm

Since all the unknown variables in the model described above, $\{\tau, \lambda_1, \lambda_2\}$, are continuous, they are discretized for algorithmic implementation. Additionally, the algorithm is susceptible to arithmetic overflow (i.e., the condition when a calculation produces a result that is greater in magnitude than which can be represented in computer memory), for instance when computing $\Gamma(x)$ for large x (e.g., $\Gamma(200) = 3.94 \times 10^{372}$). To prevent overflow, the computations are performed in the log domain and reverted back at the end. We represent in our algorithm, the largest finite floating-point number on a computer as REAL_MAX ($= 1.7977 \times 10^{308}$ on a 64-bit machine). Algorithms 1 and 2 describe how to obtain the posterior densities of τ and λ, respectively.

Algorithm 1 Estimating the distribution of τ using the change point model

1: Discretize τ uniformly into x_i for $i = 1, \ldots, p$ over its domain $[0, T]$.

2: At each x_i, calculate $log\_prob_i = \log(\pi(x_i \mid \boldsymbol{t}))$, using equation 11.

3: To exponentiate the log probability, find the smallest *scale* such that $\sum_i \mathrm{e}^{log\_prob_i - scale} \leq$ REAL_MAX. The *scale* ensures that the final result can be represented in the computer memory.

4: Calculate $prob_i = \mathrm{e}^{log\_prob_i - scale}$.

5: Normalize $pdf_i = \dfrac{prob_i}{\sum prob_i \times (x_{i+1} - x_i)}$ to obtain the probability density function value at each x_i.

3. Assessing spatially varying event rates with a change point

In this section, we present a methodology to estimate spatially varying event rates using the model from the previous section. We also describe a likelihood comparison method based on the approach described by [12], to optimize the model's spatial averaging parameter.

Algorithm 2 Estimating the distribution of λ using the change point model

1: Discretize the variable of interest, λ_1 or λ_2, into x_i for $i = 1, \ldots, q$ over its domain. Since the domain is $[0, \infty)$, select a large enough range such that probability of observing a rate less than the smallest value, and greater than the largest value, is negligible.

2: Discretize τ uniformly into τ_j for $j = 1, \ldots, p$ over its domain $[0, T]$.

3: At each x_i, calculate z_{ij} for all $j = 1, \ldots, p$, using equation 14, or equation 17.

4: At each x_i, calculate $sum_i = \sum_j \mathrm{e}^{z_{ij} - scale_i}$ using the smallest $scale_i$ such that $sum_i \leq$ REAL_MAX. The $scale_i$ ensures that the final result can be represented in the computer memory.

5: At each x_i, calculate $log\_prob_i = \log(sum_i)$, using equation 13, or equation 16.

6: Follow steps 3 through 5 in algorithm 1 to obtain the probability density function value at each x_i.

3.1. Estimating event rates over a spatial grid

Given a two dimensional space where discrete event sources cannot be identified, we divide the region into a uniform grid, and calculate event rates at each grid point (see figure 2). The spacing of the grid can be determined using prior knowledge about the physics of the process under consideration, or optimized using the approach described in section 3.2.

At each grid point, the change point model of section 2 is implemented on the events observed in a circular region of radius $\mathfrak{r}$ around the grid point. For the events observed in this circular region, the inter-event times are calculated to be used as input for the model. If a change is not detected using the Bayes factor for the sequence of events, then the event rate at the grid point is estimated using the no-change model. If a change is detected, then the post-change rate is used as the current event rate. This estimated rate is divided by $\pi\mathfrak{r}^2$ to compute the event rate per unit area. Based on the properties of the event process and the application of the rates, the post-change rate can be selected as the posterior mean, mode or median of the posterior distribution, or the complete distribution can be selected.

The size of the circular region affects the smoothing of the spatially-varying rates. As shown in figure 2, if the radius $\mathfrak{r}$ of the circular region is too small compared to the grid size, then some observed events in the space will not be considered in estimating the event rates. When $\mathfrak{r}$ is large, there will be some events that will be included more than once in the rate calculations due to overlapping circular regions at adjacent grid points. It is not possible to weigh the events according to distance, since the above change point analysis uses inter-event times between events. However, this

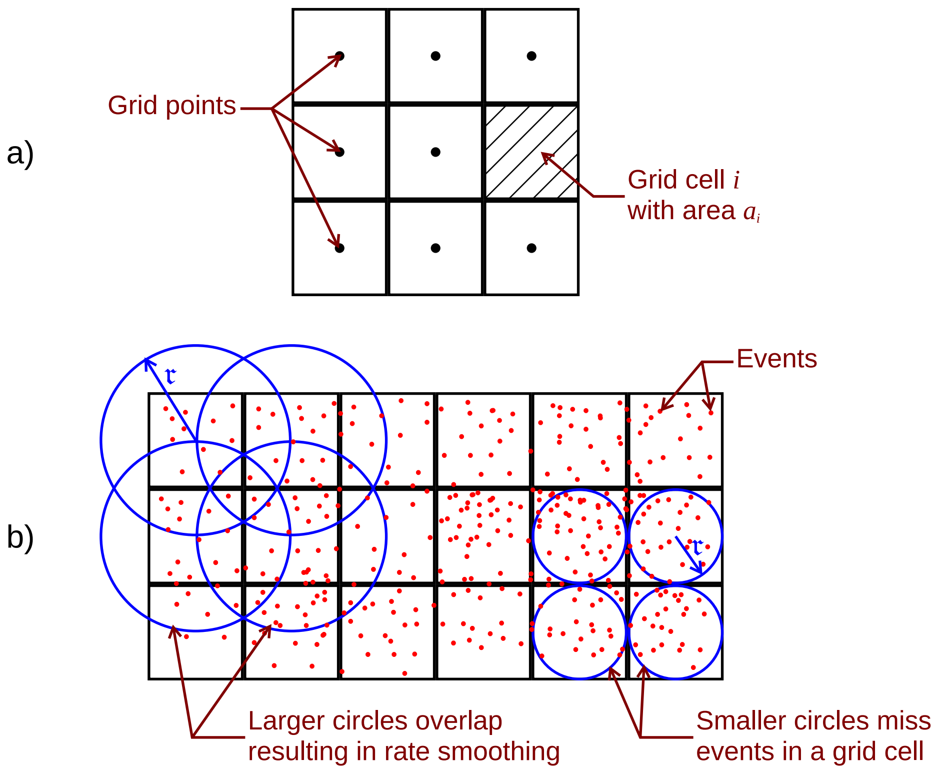

Figure 2: a) A two-dimensional space divided into a uniform grid, showing grid points and grid cells, and b) circular regions of different sizes showing the influence of radius $\mathfrak{r}$ on rate smoothing.

multi-counting of events does not artificially increase the event rates over the entire space since the rates estimated at the grid points are normalized to a rate per unit area, and are only applicable over the corresponding grid cells. A larger value of $\mathfrak{r}$ increases the number of common events between adjacent grid points, and thus has the effect of smoothing the estimated rates. The desired smoothing of the event rates is difficult to determine a priori, so we use a likelihood comparison methodology described below to select $\mathfrak{r}$.

3.2. Optimizing the parameters of the model

In this section, we determine the radius $\mathfrak{r}$ described above by maximizing the likelihood of the model associated with observing future events, for varying $\mathfrak{r}$. We use a modified version of the likelihood comparison methodology described by [12].

We first formulate the likelihood of the model. Let there be m grid cells, each associated with a grid point (a grid cell is the rectangle formed by midpoints of grid intersections associated with a grid point, as shown in figure 2). Let the event rate per unit area per unit time at grid point i be represented by λ_i for $i = 1, \ldots, m$. Let $\mathcal{C}_f$ be some future catalog of events with a catalog duration t_f. Let the number of events observed in the future catalog within grid cell i be n_i. Let the area of grid cell i be given by a_i. Then using the fact that events belong to a Poisson process, the likelihood $\mathcal{L}_i$ of the model for grid cell i associated with events n_i is computed as -

$$\mathcal{L}_i = \frac{(\lambda_i a_i t_f)^{n_i} \mathrm{e}^{-\lambda_i a_i t_f}}{n_i!} \tag{22}$$

The likelihood over the entire space $\mathcal{L}$ is calculated by multiplying the likelihood for all m grid cells.

$$\mathcal{L} = \prod_{i=1}^{m} \mathcal{L}_i = \prod_{i=1}^{m} \frac{(\lambda_i a_i t_f)^{n_i} \mathrm{e}^{-\lambda_i a_i t_f}}{n_i!} \tag{23}$$

We compute the log-likelihood ℓ by taking the log.

$$\ell = \sum_{i=1}^{m} n_i \log(\lambda_i a_i t_f) - t_f \sum_{i=1}^{m} \lambda_i a_i + c_f \tag{24}$$

where $c_f = -\sum_i \log n_i!$ is a constant term which depends on the future catalog, but not on the event rates. This term can be disregarded when comparing the log-likelihoods of two models for the same future catalog.

If there are two different models, M_1 and M_2 with corresponding log-likelihoods ℓ_1 and ℓ_2 respectively, they are compared by calculating the probability gain G_{12} per event. If $G_{12} > 1$, it implies that M_1 has a higher likelihood associated with the events in $\mathcal{C}_f$, and that M_1 is a better estimator of events the larger the gain is.

$$G_{12} = \exp\left(\frac{\ell_1 - \ell_2}{\sum_i n_i}\right) \tag{25}$$

This probability gain calculation is similar to that of [12], except that we do not normalize the event rates in a grid cell with the total number of events in the future catalog. Normalization of event rates is useful when examining the spatial distribution of events, and assuming that the cumulative event rate remains constant over time. When implementing the change point analysis however, we expect that event rates may change for some regions in the space. Hence, our calculations omit the normalization step.

The likelihood comparison approach will be used for comparison of models with different radii of the circular region. However, this approach is versatile and can be used to compare the performance of any two models that estimate rates over a spatial grid, for a given future catalog.

4. Application in Oklahoma

In this section, we implement the above calculations to detect and quantify changes in seismicity rates in Oklahoma due to induced seismicity. We first consider a single location, then apply the model throughout the state, and finally optimize the spatial smoothing parameter.

We use the Oklahoma Geological Survey earthquake catalog, for magnitudes $M \geq 3$ earthquakes from January 01, 1974 to December 31, 2015 [13]. Earthquakes are typically assumed to behave as a Poisson process when an earthquake catalog is declustered [e.g., 14, 15]. We decluster the catalog using the Reasenberg approach described by [16], using parameters developed for California since these parameters have not been determined for Oklahoma. The minimum magnitude for catalog completeness is set to magnitude 3. The original catalog contains 1708 $M \geq 3$ events, and 1051 mainshocks remain after declustering. We note that there has not been a conclusive study identifying the best declustering methodology to use for regions of induced seismicity. Since declustering is done independently of the model implementation, other declustering techniques like Gardner-Knopoff [14] may be utilized while maintaining the model framework described in this paper.

4.1. Application at a single location

We first implement the Bayesian change point analysis described in Section 2 for a site at 96.7° W and 35.6° N. We consider a circular region of radius

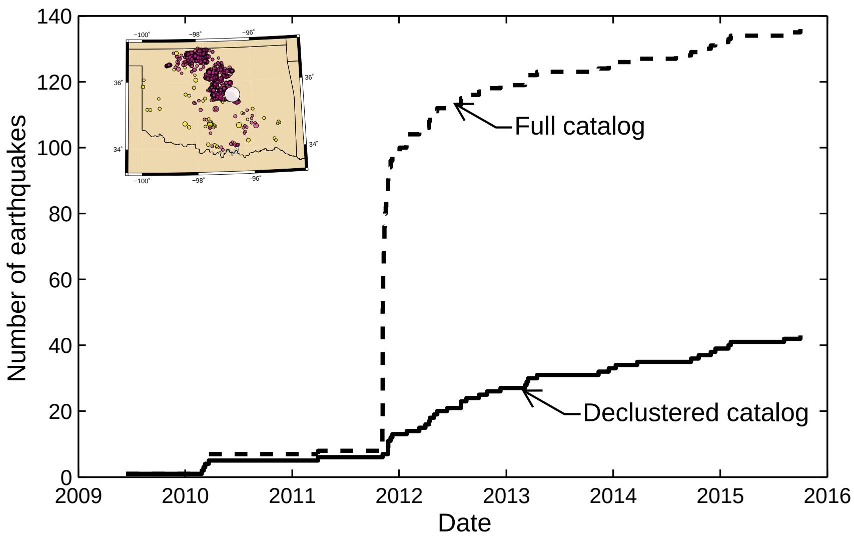

Figure 3: The non-declustered (full) and declustered catalogs of events within 25 km of 96.7° W and 35.6° N. The white circle on the inset marks the circular region on the map of Oklahoma. There were no observed earthquakes from 1974 to 2009, hence the date range has been shortened.

$\mathfrak{r} = 25$ km around this site. The radius size is optimized later. The earthquakes observed in this region since 1974 are shown in figure 3. This region includes the largest recently recorded earthquake in Oklahoma of magnitude 5.6 at Prague on November 06, 2011. From the figure, it is visually apparent that a change in seismicity rate occurred around 2009, but we would like to identify this change using our model.

We first determine whether the inter-event times between earthquakes support a change point model. We use the following hyper-parameter values for the priors: $k_j = 0.5$ and $\theta_j \to \infty$ for $j = 0, 1, 2$. For our application, we reduce the Bayes factor threshold to a value of less than 1×10^{-3} to require a strong preference for the change model before inferring that a change occurred. This is done for numerical stability, and to minimize accidental change detection when running multiple analyses at different grid points. A Bayes factor of 7×10^{-32} is computed for this data, suggesting strongly that a change point model better describes the data than a constant rate model.

The posterior distributions for the time of change τ, and the rates before the change λ_1 and after the change λ_2 are then computed. Figure 4 shows a high probability density that a change in seismicity rates occurred between December 20, 2008 and February 24, 2010 with the highest density on June 13, 2009. This matches the expected range for time of change from a visual inspection. Figure 5 shows the posterior distributions of seismicity rates before and after the change. The maximum a posteriori (MAP) estimators of the distributions indicate that the post-change seismicity rate is about 300 times the pre-change rate. We also observe a narrower probability distribution for the post-change rate due to the occurrence of more earthquakes, and hence more data, after 2009.

One advantage of a Bayesian model is that it provides posterior probability distributions for the parameters, like the time of change τ and rate λ_2, as shown in figures 4 and 5. These distributions can be utilized in risk estimation to account for uncertainties in parameter estimates.

4.2. Spatially varying seismicity rates

We now apply the model over the entire state to identify those regions where seismicity rates have changed, and to estimate the current seismicity rates.

United States Geological Survey (USGS) divides a region with unmapped

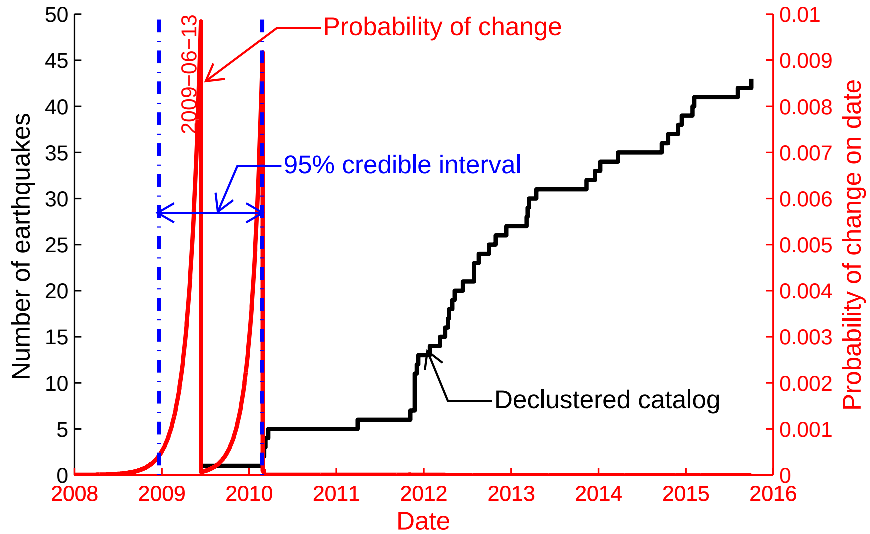

Figure 4: The probability of change on any given date τ, along with the 95% credible interval between December 20, 2008 and February 24, 2010. The MAP estimator is at June 13, 2009

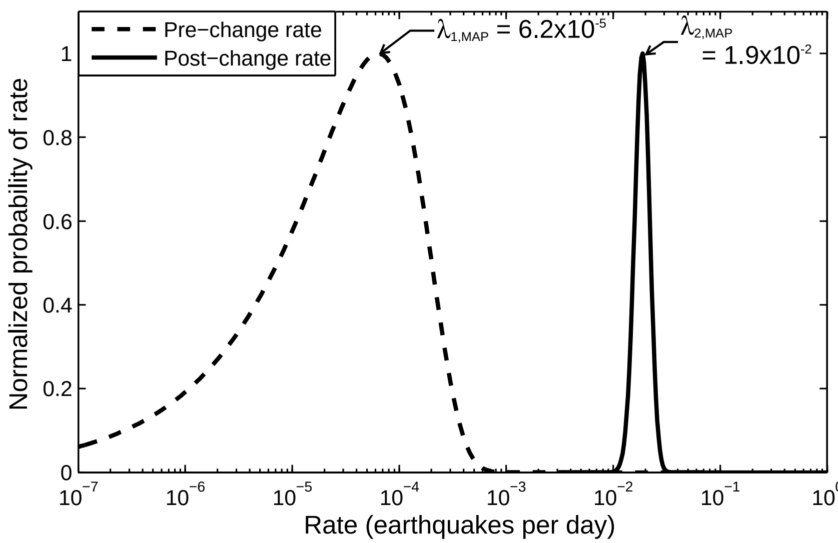

Figure 5: The normalized probability distribution of λ_1 and λ_2 for the selected location along with their MAP estimators.

seismic faults into a 0.1° latitude by 0.1° longitude grid (approximately 10 km by 10 km) to estimate the rate at each grid point for their hazard maps generation [17], and for developing smoothed seismicity models for induced earthquakes [8]. We use the same uniform grid. At each grid point, we use the earthquakes observed within a circular region of radius $\mathfrak{r} = 25$ km to estimate the seismicity rate at that grid point. The choice of a circular region is made so that the earthquakes considered in the change point model are within the same maximum distance from a grid point, however, the model can be implemented on any arbitrary shape. If the earthquakes support the change point model, we compute the MAP estimators for the time of change τ and the post-change rate λ_2. Otherwise, we compute the MAP estimator for the constant rate model λ_0. We designate this rate as the current rate of seismicity at the grid point.

Figures 6 and 7 show the MAP estimators for time of change in the state, and the current seismicity rate, respectively. For clarity, only the regions with rates greater than 0.001 earthquakes per year per km^2 are shown in figure 7.

The regions of seismicity change generally agree with regions identified by others as having anomalously high earthquake activity [18], and the dates of change agree with other general observations of a statewide seismicity increase in 2009 [19, 2].

4.3. Model optimization

The model optimization approach described in section 3.2 uses future events to select the model with the maximum likelihood. To simulate future events, we extract two mutually exclusive subsets from the earthquake cata-

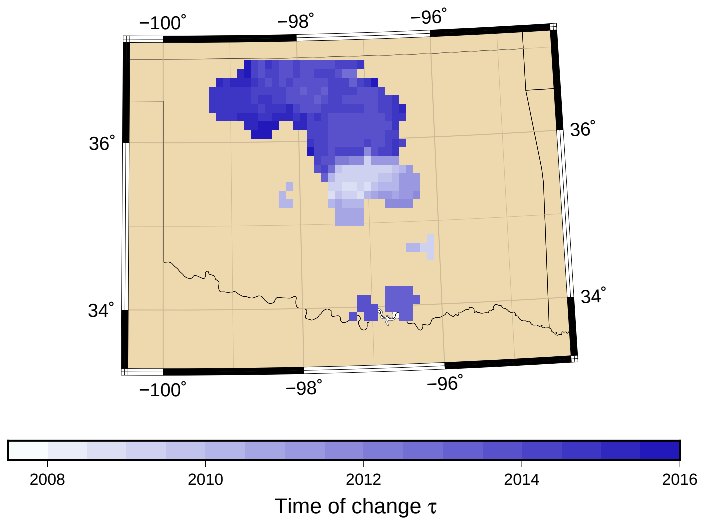

Figure 6: Time of change τ for those parts of the state where change is detected using a 25 km radius region.

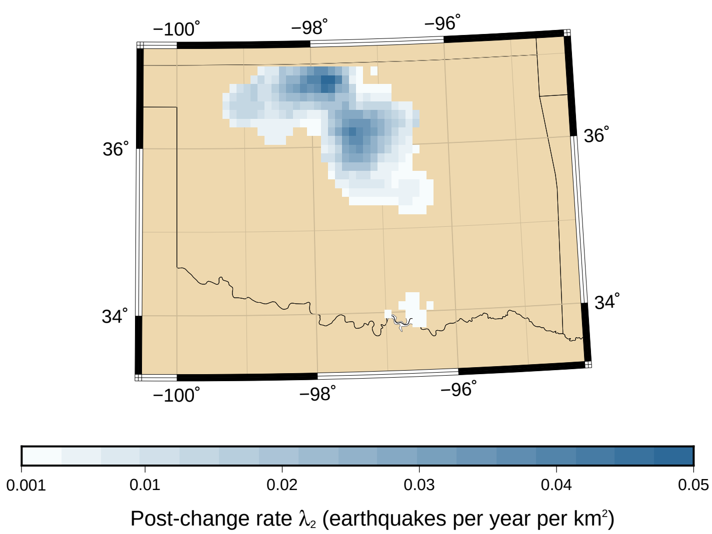

Figure 7: Current seismicity rates at grid points using a 25 km radius region. For clarity, only regions with rates greater than 0.001 are shown.

log, estimate the rates for a model on one subset, and calculate the likelihood of this model given the events in the other subset. The former subset is called the training catalog, and the latter the test catalog. This is similar to the cross-validation approach used to develop machine learning models [20].

The training catalog consists of observations from 1974 up to a varying end date. Observed earthquakes in the training catalog are used to estimate the seismicity rates, and then these rates are used to make predictions of seismicity in the next 0.5 year or 1 year. Hence, our test catalogs contain the earthquake observations over 0.5 year or 1 year duration following the end of each training catalog.

The probability gain per event, described in equation 25, is computed with ℓ_2 corresponding to a reference uniform rate model that estimates equal seismicity rates at all grid points in the state by dividing the observed number of earthquakes in the training catalog by the number of grid points. This reference model is compared to the Bayesian change point models with different radii $\mathfrak{r}$ of the circular region. The model with radius that yields the highest probability gain for the events in the test catalog is selected as the optimum model.

The probability gains G_{12} for 0.5 year and 1 year test catalogs for several choices of $\mathfrak{r}$ and training catalog are shown in figures 8 and 9, respectively. It is observed that for most $\mathfrak{r}$ and for all the recent test catalogs, the G_{12} values are larger than 1, indicating that the likelihood of the Bayesian models is higher than the uniform rate models for their respective test catalogs.

The highest probability gain (G_{12}) is typically observed for a radius $\mathfrak{r}$ of 25 km to 35 km across all training catalogs. The highest probability gains

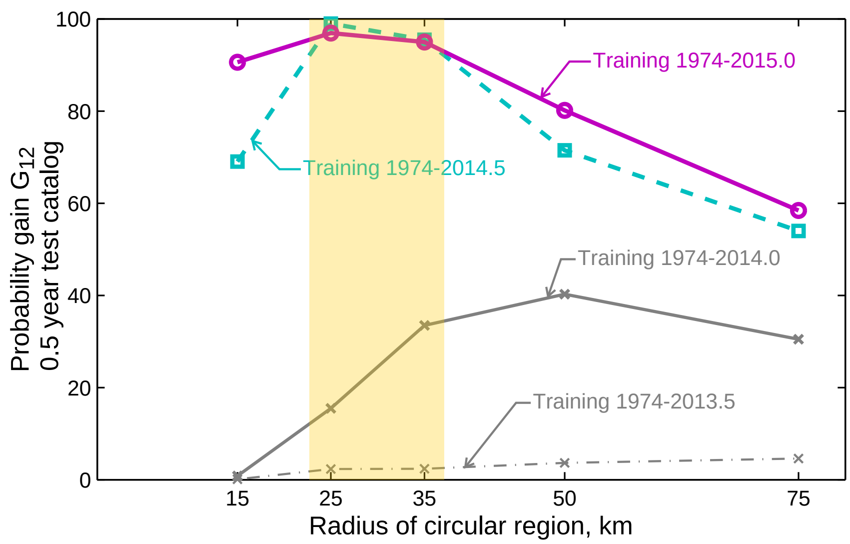

Figure 8: Probability gain per earthquake with respect to a uniform rate model for different training catalogs and radii of circular region. The test catalog is 0.5 year duration post end of training catalog. The highlighted region emphasizes the typically higher G_{12} for a radius $\mathfrak{r}$ of 25 to 35 km.

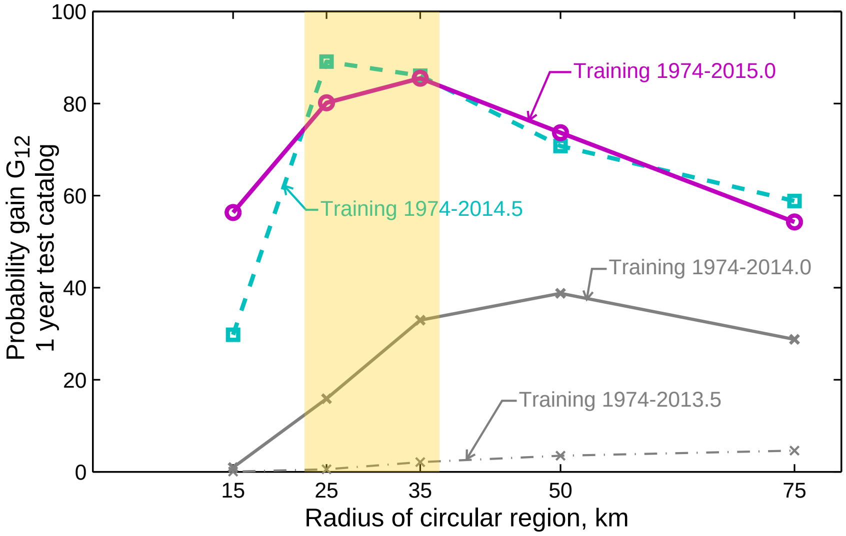

Figure 9: Probability gain per earthquake with respect to a uniform rate model for different training catalogs and radii of circular region. The test catalog is 1 year duration post end of training catalog. The highlighted region emphasizes the typically higher G_{12} for a radius $\mathfrak{r}$ of 25 to 35 km.

across all catalogs are obtained for the two longest training catalogs. This indicates that a radius of the circular region in the range of 25 km to 35 km is best suited for this application of estimating spatially-varying seismicity rates using the Bayesian change point model for induced seismicity in Oklahoma. As a result, our previous analysis using a radius of 25 km corresponds to a model that is expected to be effective in predicting future earthquakes. This optimal radius may vary in other regions of induced seismicity. The optimal radius is 25 km to 35 km in this case due to the 0.1° by 0.1° grid size, and likely due to the uncertainty in earthquake locations resulting from limited seismic recordings.

Comparing the probability gains per earthquake (G_{12}) of figure 8 with figure 9, the gain is generally higher for the 0.5 year test catalogs, than for the 1 year test catalogs. Hence, the model indicates better future predictions of earthquakes over shorter timespans, as expected for a dynamic phenomenon like this.

5. Model limitations

We discuss here two limitations associated with the model described in previous sections.

5.1. Choice of priors

The choice of hyper-parameters for the prior distribution affects the results obtained from a Bayesian model. However, we expect that significantly different pre-change and post-change rates will limit the impacts of the choice of hyper-parameters on the posterior distributions. Data-rich regions are also expected to be less impacted by the choice of priors since the posterior distributions are controlled to a greater extent by the data as the sample size increases [11]. The hyper-parameter values selected in this paper simulate an infinite variance prior distribution or an uninformative prior, where the user imposes no prior beliefs about the process [11]. We study the impacts of alternate parameter choices below through application on Oklahoma data.

We first utilize the previous example of a single location described in section 4.1 to analyze the impact of choices on our priors. Figure 10 shows the MAP estimators with 95% credible intervals for the time of change and the rates for different prior values.

We observe from the figure that different hyper-parameter values yield slightly different posterior distributions. For the time of change τ, the posterior distributions have little variation. This is because of the significant change in seismicity rates around mid 2009. The credible intervals for the pre-change rate λ_1 are generally large. This is because no events are observed before 2009 in our data. Hence, the posterior distribution of λ_1 has large variance and is more sensitive to the choice of the prior distribution. When this is contrasted with the data-rich post-change rate λ_2, it is observed that the confidence intervals and the MAP estimators show little variation with different prior values.

We also compare the previous statewide results with results obtained when using hyper-parameters $k_j = 0.05, \theta_j = 0.1$, in figures 11 and 12. The results are in good agreement, with differences at the boundaries of regions with change, and in regions of low post-change rates. The boundaries are impacted because fewer earthquakes are observed in these regions.

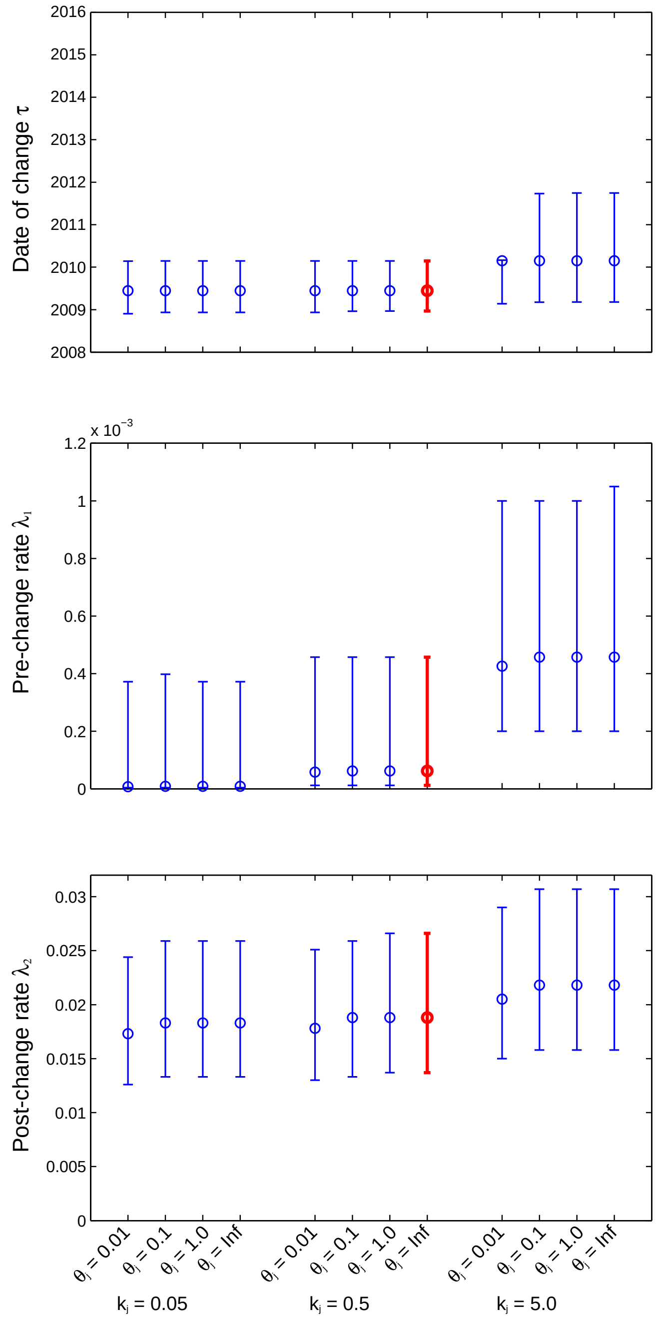

Figure 10: MAP estimators and 95% credible intevals for τ, λ_1 and λ_2 for different hyper-parameter values. The circles are the MAP estimators at each hyper-parameter value, and the wings represent the lower and upper limits for the 95% credible interval. The red line marks the results for the default hyper-parameters used in this paper.

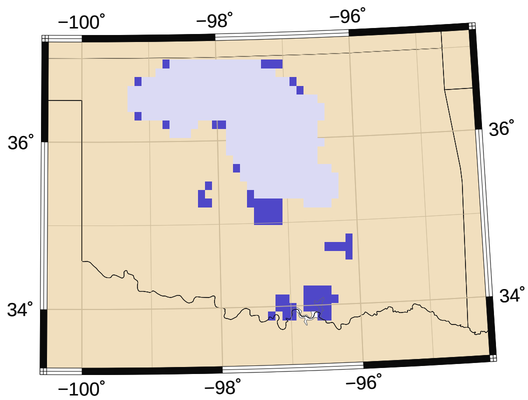

Figure 11: Regions of seismicity change for $k_j = 0.5, \theta_j \to \infty$, overlain with those for $k_j = 0.05, \theta_j = 0.1$. The light-shaded regions are common to both hyper-parameters, while the dark-shaded regions are noted only for the former hyper-parameters.

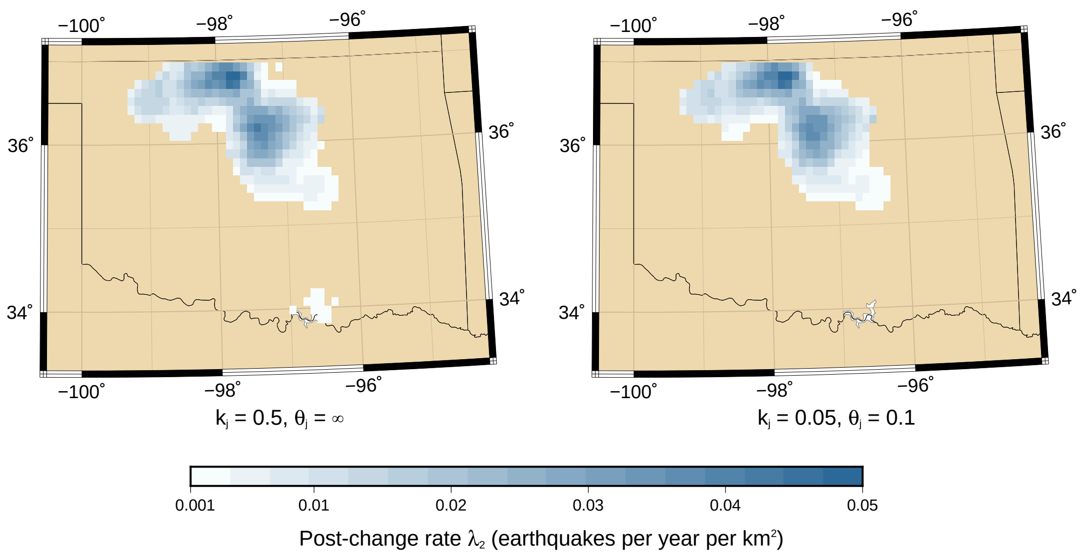

Figure 12: Current seismicity rates for $k_j = 0.5, \theta_j \to \infty$ (left), and with for $k_j = 0.05, \theta_j = 0.1$ (right).

The regions of low post-change rates are impacted because the pre-change and post-change rates are similar to each other.

Based on the comparisons, we observe that different choices of prior distributions affect posteriors, but there is limited impact in data-rich locations. Due to a large increase in seismicity rates at many locations, and many earthquakes being observed in the post-change periods, there is little impact from choice of hyper-parameters on the posterior distributions for this application on the parameters of interest, τ and λ_2.

5.2. Assumption of a single change

The other limitation of the change point model is that it assumes a single change in the rate of events (equation 2). A multiple change point analysis is possible using Gibbs sampling [21], which we do not describe here. However, every change point adds two independent parameters to the model (additional time of change, and rate), which can introduce overfitting, and requires more data to reduce the variance of the posterior distributions for uninformative priors. One possible approximation to the multiple change point analysis is to sequentially bisect the catalog at the maximum a posteriori (MAP) estimator of the previous change point, until the Bayes factor indicates a support for a no change model on all the branches. This process of sequential bisection is not the same as a complete multiple change point analysis since each subsequent branch is conditioned on the location of the previous change point. However, this method could serve as a rudimentary check to determine whether the process should be instead modeled with a multiple-change point model.

As an example, we evaluate whether a multiple change point model might be better applicable to the same single location from section 4.1. We use the method of sequential bisection at the MAP estimator of change point τ. Here, the Bayes factor for the post-change branch is 9.1×10^{-2}. The pre-change branch has no events, hence has no observable change. Since both branches have Bayes factors larger than our selected threshold of 1×10^{-3}, we state that a single change point model is an acceptable model for this example.

6. Conclusions

We presented a Bayesian change point methodology to detect a change in event rates for a non-homogeneous Poisson process, and evaluated spatially-varying event rates for this process. The Bayesian methodology enables us to develop probability distributions for the time of change, and for the event rates before and after the change.

We evaluated the spatially varying event rates for a process by dividing the space into a grid and evaluating the rate at each grid point. Rates were evaluated based on the events observed in a circular region of radius $\mathfrak{r}$ around each point. We also presented a likelihood comparison methodology to optimize the radius $\mathfrak{r}$ for best future predictions of event probabilities.

We demonstrated the application of the Bayesian change point methodology on the spatially varying earthquake rates associated with induced seismicity in Oklahoma. We optimized the radius $\mathfrak{r}$ and concluded that a radius between 25 and 35 km yields the highest probability of observing future earthquakes in Oklahoma.

The model implementation in Oklahoma identified the regions in the state where seismicity rates have changed. We also estimated the current seismicity rates using the model. Our results were in general agreement with other studies on time of seismicity change [19, 2], and regions of seismicity change [18]. The current seismicity rates can be used to make short-term future predictions of earthquakes in the state. We observed that there is better prediction over the next 0.5 year duration compared to the next 1 year duration. In a future publication, we will compare the performance of our model for future earthquake predictions with other rate estimation models, using the Collaboratory for the Study of Earthquake Predictability (CSEP) tests [22].

The occurrence of seismicity change combined with estimated seismicity rates can serve as a risk mitigation tool for operations that affect seismicity, for example, to prepare prioritization plans for infrastructure inspections [23]. This information can be used in seismic hazard and risk assessments for the region [24].

One of the possible extensions to this model for its application on induced seismicity could be the combination of the Bayesian change point methodology with an earthquake catalog declustering approach like the epidemic type aftershock sequence (ETAS) model [25]. Combining the declustering model with the change point model would allow estimation of declustering parameters, in addition to seismicity rates, for the local conditions. In this paper, the earthquake catalog declustering was done independently of the change point model implementation. This allowed for the development of numerical algorithms to solve the change point model. Solving the combined declustering and change point model would require random state generation algorithms like Markov Chain Monte Carlo (MCMC) methods.

The Bayesian change point model presented here, along with the methodology to assess spatially varying event rates, is a versatile model that can be used to estimate current event rates for any spatially varying non-homogeneous Poisson process. Change point models have been used to study DNA sequence segmentation [26], species extinction [27], financial markets [28], and software reliability [29]. Some of the other applications where this spatio-temporal change point model can be used are assessing spread of diseases, and identifying changes in climate patterns. This model enables stakeholders to make real-time decisions about the impact of changes in event rates.

7. Resources

Earthquake catalog declustering is performed using the code by [30]. Matlab source code to perform change point calculations for Oklahoma is available at `https://github.com/abhineetgupta/BayesianChangePoint`.

8. Acknowledgements

Funding for this work came from the Stanford Center for Induced and Triggered Seismicity. We would like to thank Max Werner and Morgan Moschetti for providing feedback on the spatial rate assessment methodology.

9. References

[1] A. Raftery, V. Akman, Bayesian analysis of a Poisson process with a change-point, Biometrika 73 (1) (1986) 85–89.

[2] P. Wang, M. Pozzi, M. J. Small, W. Harbert, Statistical Method for Early Detection of Changes in Seismic Rate Associated with

Wastewater Injections, Bulletin of the Seismological Society of Americadoi:10.1785/0120150038.
URL http://www.bssaonline.org/content/early/2015/10/05/0120150038

[3] W. L. Ellsworth, Injection-Induced Earthquakes, Science 341 (6142) (2013) 1225942. doi:10.1126/science.1225942.
URL http://www.sciencemag.org/content/341/6142/1225942

[4] K. M. Keranen, M. Weingarten, G. A. Abers, B. A. Bekins, S. Ge, Sharp increase in central Oklahoma seismicity since 2008 induced by massive wastewater injection, Science 345 (6195) (2014) 448–451. doi:10.1126/science.1255802.
URL http://www.sciencemag.org/content/345/6195/448

[5] M. Weingarten, S. Ge, J. W. Godt, B. A. Bekins, J. L. Rubinstein, High-rate injection is associated with the increase in U.S. mid-continent seismicity, Science 348 (6241) (2015) 1336–1340. doi:10.1126/science.aab1345.
URL http://www.sciencemag.org/content/348/6241/1336

[6] N. R. Council, Induced Seismicity Potential in Energy Technologies, The National Academies Press, Washington, DC, 2013.
URL http://www.nap.edu/catalog/13355/induced-seismicity-potential-in-energy-technologies

[7] R. J. Walters, M. D. Zoback, J. W. Baker, G. C. Beroza, Characterizing and Responding to Seismic Risk Associated with Earthquakes Poten-

tially Triggered by Saltwater Disposal and Hydraulic Fracturing. URL http://www.gwpc.org/sites/default/files/files/Walters%20et%20al%20%20SRL%20paper%20on%20risk_submitted.pdf

[8] M. D. Petersen, 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, Incorporating Induced Seismicity in the 2014 United States National Seismic Hazard Model - Results of 2014 Workshop and Sensitivity Studies, Open-File Report 20151070, U.S. Geological Survey, Reston, Virginia (2015).

[9] A. Gupta, J. W. Baker, A Bayesian change point model to detect changes in event occurrence rates, with application to induced seismicity, in: 12th International Conference on Applications of Statistics and Probability in Civil Engineering (ICASP12), 2015. URL https://circle.ubc.ca/handle/2429/53235

[10] R. E. Kass, A. E. Raftery, Bayes Factors, Journal of the American Statistical Association 90 (430) (1995) 773–795. doi:10.1080/01621459.1995.10476572. URL http://amstat.tandfonline.com/doi/abs/10.1080/01621459.1995.10476572

[11] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Rubin, Bayesian data analysis, 3rd Edition, Taylor & Francis, 2014. URL http://www.tandfonline.com/doi/full/10.1080/01621459.2014.963405

[12] M. J. Werner, A. Helmstetter, D. D. Jackson, Y. Y. Kagan, High-Resolution Long-Term and Short-Term Earthquake Forecasts for California, Bulletin of the Seismological Society of America 101 (4) (2011) 1630–1648. doi:10.1785/0120090340.
URL http://www.bssaonline.org/content/101/4/1630

[13] Oklahoma Geological Survey, University of Oklahoma, Earthquake catalog available on the World Wide Web.
URL http://www.ou.edu/content/ogs/research/earthquakes/catalogs.html

[14] J. K. Gardner, L. Knopoff, Is the sequence of earthquakes in Southern California, with aftershocks removed, Poissonian?, Bulletin of the Seismological Society of America 64 (5) (1974) 1363–1367.
URL http://bssa.geoscienceworld.org/content/64/5/1363

[15] T. van Stiphout, J. Zhuang, D. Marsan, Seismicity declustering, Community Online Resource for Statistical Seismicity Analysis 10.
URL http://www.corssa.org/articles/themev/van_stiphout_et_al

[16] P. Reasenberg, Second-order moment of central California seismicity, Journal of Geophysical Research 90 (B7) (1985) 5479. doi:10.1029/JB090iB07p05479.
URL http://doi.wiley.com/10.1029/JB090iB07p05479

[17] M. D. Petersen, M. P. Moschetti, P. M. Powers, C. S. Mueller, K. M. Haller, A. D. Frankel, Y. Zeng, S. Rezaeian, S. C. Harmsen, O. S. Boyd,

N. Field, R. Chen, K. S. Rukstales, N. Luco, R. L. Wheeler, R. A. Williams, A. H. Olsen, Documentation for the 2014 update of the United States national seismic hazard maps, Tech. rep., U.S. Geological Survey Open-File Report 20141091 (Jul. 2014).
URL http://pubs.usgs.gov/of/2014/1091/

[18] W. Ellsworth, A. Llenos, A. McGarr, A. Michael, J. Rubinstein, C. Mueller, M. Petersen, E. Calais, Increasing seismicity in the U. S. midcontinent: Implications for earthquake hazard, The Leading Edge 34 (6) (2015) 618–626. doi:10.1190/tle34060618.1.
URL http://library.seg.org/doi/full/10.1190/tle34060618.1

[19] A. L. Llenos, A. J. Michael, Modeling Earthquake Rate Changes in Oklahoma and Arkansas: Possible Signatures of Induced Seismicity, Bulletin of the Seismological Society of America 103 (5) (2013) 2850–2861. doi:10.1785/0120130017.
URL http://www.bssaonline.org/content/103/5/2850

[20] T. M. Mitchell, Machine Learning, 1st Edition, McGraw-Hill, Inc., New York, NY, USA, 1997.

[21] D. A. Stephens, Bayesian Retrospective Multiple-Changepoint Identification, Journal of the Royal Statistical Society. Series C (Applied Statistics) 43 (1) (1994) 159–178. doi:10.2307/2986119.
URL http://www.jstor.org/stable/2986119

[22] J. D. Zechar, D. Schorlemmer, M. Liukis, J. Yu, F. Euchner, P. J. Maechling, T. H. Jordan, The Collaboratory for the Study of Earth-

quake Predictability perspective on computational earthquake science, Concurrency and Computation: Practice and Experience 22 (12) (2010) 1836–1847. doi:10.1002/cpe.1519.
URL http://onlinelibrary.wiley.com/doi/10.1002/cpe.1519/abstract

[23] C. Hill, Oklahoma DOT hires engineering firm to create earthquake response protocol for bridge inspections (Apr. 2015).
URL http://www.equipmentworld.com/oklahoma-dot-hires-engineering-firm-to-create-earthquake-response-protocol-fo

[24] J. W. Baker, A. Gupta, Bayesian Treatment of Induced Seismicity in Probabilistic Seismic Hazard Analysis., Bulletin of the Seismological Society of America (in press).

[25] Y. Ogata, J. Zhuang, Spacetime ETAS models and an improved extension, Tectonophysics 413 (12) (2006) 13–23. doi:10.1016/j.tecto.2005.10.016.
URL http://www.sciencedirect.com/science/article/pii/S0040195105004889

[26] J. V. Braun, H.-G. Muller, Statistical Methods for DNA Sequence Segmentation, Statistical Science 13 (2) (1998) 142–162.
URL http://www.jstor.org/stable/2676755

[27] A. R. Solow, Inferring Extinction from Sighting Data, Ecology 74 (3) (1993) 962–964. doi:10.2307/1940821.
URL http://www.jstor.org/stable/1940821

[28] X. Liu, X. Wu, H. Wang, R. Zhang, J. Bailey, K. Ramamohanarao, Mining distribution change in stock order streams, in: 2010 IEEE 26th International Conference on Data Engineering (ICDE), 2010, pp. 105–108. doi:10.1109/ICDE.2010.5447901.

[29] F.-Z. Zou, A change-point perspective on the software failure process, Software Testing, Verification and Reliability 13 (2) (2003) 85–93. doi:10.1002/stvr.268.
URL http://onlinelibrary.wiley.com/doi/10.1002/stvr.268/abstract

[30] S. Wiemer, A Software Package to Analyze Seismicity: ZMAP, Seismological Research Letters 72 (3) (2001) 373–382. doi:10.1785/gssrl.72.3.373.
URL http://srl.geoscienceworld.org/content/72/3/373