Structural Safety 65 (2017) 1–11

Contents lists available at ScienceDirect

Structural Safety

journal homepage: www.elsevier.com/locate/strusafe

Estimating spatially varying event rates with a change point using Bayesian statistics: Application to induced seismicity

Abhineet Gupta *, Jack W. Baker

Civil & Environmental Engineering Department, Stanford University, Stanford, CA 94305, USA

ARTICLE INFO

Article history:
Received 9 March 2016
Received in revised form 9 November 2016
Accepted 16 November 2016

Keywords:
Change point method
Bayesian inference
Induced seismicity
Spatio-temporal process

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 seismicity.

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 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., 3–5]. Seismicity generated as a result of human activities is referred to as induced or triggered seismicity. Fig. 1 shows the cumulative number of earthquakes with magnitude $\geqslant 3$ 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.

* Corresponding author.
E-mail addresses: abhineet@stanford.edu (A. Gupta), bakerjw@stanford.edu (J.W. Baker).

http://dx.doi.org/10.1016/j.strusafe.2016.11.002

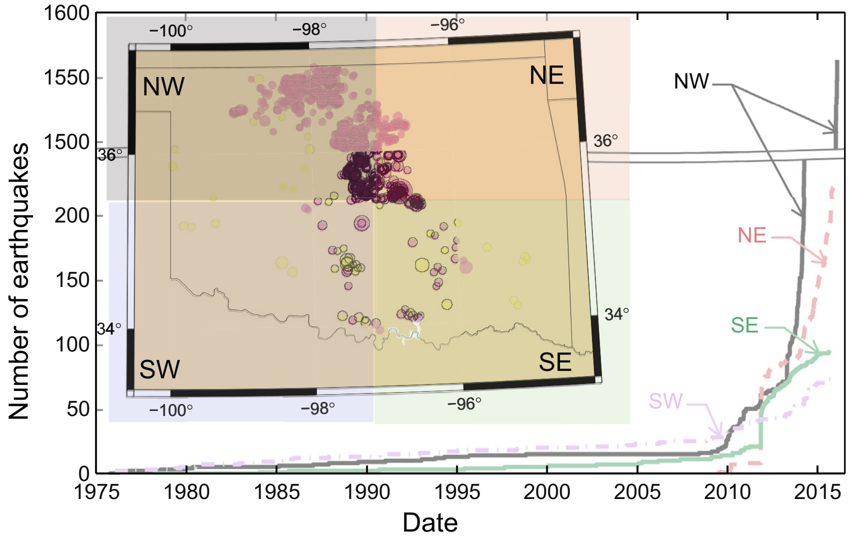

Fig. 1. Cumulative number of earthquakes in four quadrants of Oklahoma with magnitude ⩾3 from 1974 through Dec 31, 2015. The earthquakes post 2008 are shown in pink on the map, and the size of the circles is proportional to the earthquake magnitude. We have omitted the western panhandle of Oklahoma in 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.

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 [1,9]. 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.

$$B_{01}(\boldsymbol{t}) = \frac{\mathscr{L}(H_0|\boldsymbol{t})}{\mathscr{L}(H_1|\boldsymbol{t})} \quad (1)$$

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. $\mathscr{L}(H|\boldsymbol{t})$ 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 Eq. (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 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.

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

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 nth event occurs at time T. The inter-event times are defined as

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

where t_i denotes the time between occurrences of the $i-1$th and the ith 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

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

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

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

where $\propto$ is the proportionality symbol.