



# Including Non-Stationary Magnitude–Frequency Distributions in Probabilistic Seismic Hazard Analysis

MAURICIO REYES CANALES<sup>1</sup>  and MIRKO VAN DER BAAN<sup>1</sup>

**Abstract**—We describe a first principles methodology to evaluate statistically the hazard related to non-stationary seismic sources like induced seismicity. We use time-dependent Gutenberg–Richter parameters, leading to a time-varying rate of earthquakes. We derive analytic expressions for occurrence rates which are verified using Monte Carlo simulations. We show two examples: (1) a synthetic case with two seismic sources (background and induced seismicity); and (2) a recent case of induced seismicity, the Horn River Basin, Northeast British Columbia, Canada. In both cases, the statistics from the Monte Carlo simulations agree with the analytical quantities. The results show that induced seismicity affects seismic hazard rates but that the exact change greatly depends on both the duration and intensity of the non-stationary sequence as well as the chosen evaluation period. The developed methodology is easily extended to handle spatial source distributions as well as ground motion analysis in order to generate a complete methodology for non-stationary probabilistic seismic hazard analysis.

**Key words:** Non-stationary seismicity, time-dependent Gutenberg–Richter parameters, Monte-Carlo simulations, induced seismicity, Horn River Basin, Canada.

## 1. Introduction

Several studies (Ellsworth 2013; Atkinson et al. 2015; Petersen et al. 2016; Atkinson et al. 2016; van der Baan and Calixto 2017) have shown increased seismicity in geologically stable basins in North America, thought to be associated with hydraulic fracturing treatments and/or salt water disposals. Some of these studies have delineated methods to assess the hazard for induced events, based on modified versions of the traditional probabilistic seismic hazard analysis (PSHA). Atkinson et al. (2015) made

a preliminary hazard evaluation for the Fox Creek area (Alberta), based on earthquake catalogs that contain the induced events, Ground Motion Prediction Equations (GMPEs) suited for induced seismicity, while assuming stationarity as a first approach. Petersen et al. (2016, 2017) elaborate a one year hazard forecast for the central and eastern United States, based on catalogs with recorded induced events, together with different sets of GMPEs. However, these studies are limited to short-term hazard predictions due to the assumption of temporal stationarity, which presumes that the induced seismicity sequence remains of unchanged intensity during both the observation and forecasting period.

Conversely, Langenbruch and Zoback (2016) and van der Baan and Calixto (2017) show that the rate of earthquakes in Oklahoma first strongly increased but now greatly subsided in line with salt-water disposal volumes. It is evident that the induced seismicity has a non-stationary behaviour as it is strongly dependent on human activities. The traditional PSHA, which assumes stationarity, has to be modified in order to properly assess the hazard due to changing seismicity rates over time.

The Monte Carlo simulation method for PSHA (Musson 2000; Assatourians and Atkinson 2013; Bourne et al. 2014, 2015) is flexible enough to deal efficiently with non-stationary seismicity. This method consist of two main steps: (1) generation of synthetic earthquake catalogs and (2) generation of ground motion catalogs by using ground motion prediction equations (GMPEs). Even though traditional PSHA has been previously adapted to allow for time-dependency [e.g. Convertito et al. (2012)], the Monte Carlo method for PSHA has several advantages over traditional PSHA based on analytic

<sup>1</sup> Department of Physics, University of Alberta, Edmonton, Canada. E-mail: reyescan@ualberta.ca; mirko.vanderbaan@ualberta.ca

expressions (Cornell 1968), including its adaptability to different seismicity models, its ability to handle uncertainty, and its ease of implementation to compute a variety of hazard statistics (Musson 2000). One of the shortcomings of the Monte Carlo simulation method is, however the increased number of calculations and thus increased computation times, compared with direct evaluation of analytic equations.

In this study we develop a methodology based on Monte Carlo simulations for the generation of non-stationary earthquake catalogs using time-dependent Gutenberg–Richter parameters. We also derive analytical expressions for various occurrence likelihoods such as annual rates of exceedance, in the case of non-stationarity. We verify these expressions using Monte Carlo simulations. These simulated catalogs are intended to be used in further PSHA steps to develop a complete methodology for non-stationary seismic hazard analysis including the assessment of expected ground motion. To exemplify the applicability of the developed methodology and evaluate the implications of non-stationary seismicity in the hazard analysis, we study two examples. The first example is a synthetic case with two seismic sources (background and induced) and arbitrary seismic parameters. The second example comprises a recent case of induced seismicity: the Horn River Basin, Northeast British Columbia, Canada.

## 2. Theory

### 2.1. Non-Stationary Magnitude–Frequency Distributions

We assume that the magnitude-frequency distribution of earthquakes is described by the Gutenberg–Richter (GR) distribution given by (Gutenberg and Richter 1944):

$$\log(N) = a - bm, \quad (1)$$

where  $N$  is the number of earthquakes with a magnitude greater than  $m$ . The  $b$ -value indicates the ratio of small and large magnitude events and the  $a$ -value is related to the number  $N_0$  of earthquakes with a non-

negative magnitude per unit time duration (e.g., month or year). The latter is given by:

$$N_0 = 10^a. \quad (2)$$

Equation 1 can be used to compute the discrete cumulative distribution function (CDF) as Baker (2008):

$$F_M(m) = \frac{1 - 10^{-b(m-M_{min})}}{1 - 10^{-b(M_{max}-M_{min})}}, \quad (3)$$

where  $F_M(m)$  denotes the cumulative distribution function for magnitude  $m$ ,  $M_{max}$  is the maximum magnitude and  $M_{min}$  is the minimum magnitude considered for the synthetic catalog. Taking the derivative, the probability density function (PDF) can be computed. The discrete probability for a magnitude  $m$  to occur within the range  $[m_j, m_{j+1})$ , a magnitude bin, is given by the subtraction of the two boundary CDF values (Baker 2008):

$$P(m_j \leq m < m_{j+1}) = F_M(m_{j+1}) - F_M(m_j), \quad (4)$$

with  $m_j$  and  $m_{j+1}$  respectively the lower and upper magnitude and  $j$  is an integer index to create magnitude bins. To calculate the rate of earthquakes  $\lambda(m_j \leq m < m_{j+1})$  per unit time duration for a magnitude bin, we multiply the probability of occurrence  $P(m_j \leq m < m_{j+1})$  of that magnitude bin, Eq. 4, by the total expected number of earthquakes  $N(M_{min} \leq m \leq M_{max})$  per unit time duration in the range  $m = [M_{min}, M_{max}]$ , yielding:

$$\lambda(m_j \leq m < m_{j+1}) = P(m_j \leq m < m_{j+1})N(M_{min} \leq m \leq M_{max}), \quad (5)$$

The expected number of earthquakes  $N(M_{min} \leq m \leq M_{max})$  per unit time duration in the range  $m = [M_{min}, M_{max}]$  is derived from the Gutenberg–Richter relation, Eq. 1, as:

$$N(M_{min} \leq m \leq M_{max}) = 10^{a-bM_{min}} - 10^{a-bM_{max}}. \quad (6)$$

If the time unit is a year, then  $\lambda(m_j \leq m < m_{j+1})$  and  $N(M_{min} \leq m \leq M_{max})$  refer to the annual rate of earthquakes for a magnitude bin and the annual number of earthquakes in the range  $m = [M_{min}, M_{max}]$ , respectively.

Generally the earthquake rate  $\lambda(m_j \leq m < m_{j+1})$ , Eq. 5, is assumed to be stationary (that is, time-

invariant). In this case the intercept  $a$  and slope  $b$  in the GR distribution, Eq. 1, are constant. For non-stationary sequences the rate  $\lambda(m_j \leq m < m_{j+1}; t)$  is still given by Eqs. 5 and 6, but now the GR parameters,  $a(t)$  and  $b(t)$ , are understood to vary with time  $t$ , that is:

$$\lambda(m_j \leq m < m_{j+1}; t) = \frac{P(m_j \leq m < m_{j+1}; t)}{N(M_{min} \leq m \leq M_{max}; t)}, \quad (7)$$

where the expected number of earthquakes  $N(M_{min} \leq m \leq M_{max}; t)$  per unit time duration in the range  $m = [M_{min}, M_{max}]$ , is redefined as:

$$N(M_{min} \leq m \leq M_{max}; t) = 10^{a(t)-b(t)M_{min}} - 10^{a(t)-b(t)M_{max}}. \quad (8)$$

Similar expressions hold for Eqs. 3 and 4 to compute the probability of occurrence of a magnitude bin  $P(m_j \leq m < m_{j+1})$  in a time-varying frame:

$$P(m_j \leq m < m_{j+1}; t) = F_M(m_{j+1}; t) - F_M(m_j; t), \quad (9)$$

where  $F_M(m)$  is redefined as:

$$F_M(m; t) = \frac{1 - 10^{-b(t)(m-M_{min})}}{1 - 10^{-b(t)(M_{max}-M_{min})}}. \quad (10)$$

The considered  $M_{min}$  and  $M_{max}$  magnitudes in Eqs. 8 and 10 are however kept fixed. These non-stationary earthquake rates will be used in the generation of synthetic earthquake catalogs and to verify the occurrence statistics.

## 2.2. Poisson Distribution and Derived Statistical Quantities for the Seismic Hazard Analysis

The Poisson distribution describes the number of events within a certain time interval for stationary earthquake rates, and it has been traditionally assumed to describe the temporal distribution of earthquakes (Cornell 1968; Assatourians and Atkinson 2013; Baker 2013; Anagnos and Kiremidjian 1988). The stationary Poisson distribution is defined as (Cornell 1968):

$$P[N = n; t_a, t_b] = \frac{\lambda^n (t_b - t_a)^n e^{-\lambda(t_b - t_a)}}{n!}, \quad (11)$$

where  $P[N = n; t_a, t_b]$  is the probability of  $n$  occurrences happening in a time interval  $\Delta t = t_b - t_a$ , for

start and end times  $t_a$  and  $t_b$  respectively, and  $\lambda$  is the rate of occurrence of events per unit time duration. By definition  $0! = 1$ . Note that Eq. 11 only depends on the time interval  $\Delta t = t_b - t_a$ , instead of the individual start and end times. For the stationary case,  $\lambda$  is equivalent to the rate of earthquakes  $\lambda(m_j \leq m < m_{j+1})$  per unit time duration for a magnitude bin, Eq. 5. Thus,  $\lambda(t_b - t_a)$  equals the number of events of this magnitude bin within the considered time interval. The probability of at least one event happening in a time interval  $P[N > 0; t_a, t_b]$  is defined as (Baker 2013):

$$P[N > 0; t_a, t_b] = 1 - e^{-\lambda(t_b - t_a)}. \quad (12)$$

The non-stationary Poisson model has a rate of occurrence that varies with time. In this case, we use the mean  $m_\lambda(t_a, t_b)$  of the time-dependent rate, instead of a constant rate of occurrence. The non-stationary Poisson distribution is defined as (Sigman 2013):

$$P[N = n; t_a, t_b] = \frac{m_\lambda^n(t_a, t_b)(t_b - t_a)^n e^{-m_\lambda(t_a, t_b)(t_b - t_a)}}{n!}, \quad (13)$$

where  $m_\lambda(t_a, t_b)$  is the mean of the time-varying rate of occurrence  $\lambda(t)$  in the time interval  $t = [t_a, t_b]$ , defined as (Sigman 2013):

$$m_\lambda(t_a, t_b) = \frac{\int_{t_a}^{t_b} \lambda(s) ds}{(t_b - t_a)}. \quad (14)$$

For instance,  $\lambda(t)$  could be equivalent to the time-varying rate of earthquakes  $\lambda(m_j \leq m < m_{j+1}; t)$  per unit time duration for a magnitude bin, Eq. 7, and  $m_\lambda(t_a, t_b)$  results in the mean rate of earthquakes  $m_\lambda(m_j \leq m < m_{j+1}; t_a, t_b)$  for a magnitude bin in the time interval  $\Delta t = t_b - t_a$ .  $\lambda(t)$  could also be the rate of earthquakes  $\lambda(M_{min} \leq m \leq M_{max}; t)$  per unit time duration for the entire range  $m = [M_{min}, M_{max}]$ , that is:

$$\lambda(M_{min} \leq m \leq M_{max}; t) = \frac{P(M_{min} \leq m \leq M_{max}; t)}{N(M_{min} \leq m \leq M_{max}; t)}, \quad (15)$$

where  $N(M_{min} \leq m \leq M_{max}; t)$  is given by Eq. 8 and  $P(M_{min} \leq m \leq M_{max}; t)$  is a modification of Eq. 9, resulting in:

$$P(M_{min} \leq m \leq M_{max}; t) = F_M(M_{max}; t) - F_M(M_{min}; t). \quad (16)$$

Replacing  $\lambda(t)$  by  $\lambda(M_{min} \leq m \leq M_{max}; t)$  in Eq. 14,  $m_{\lambda}(t_a, t_b)$  results in the mean rate of earthquakes  $m_{\lambda}(M_{max} \leq m \leq M_{min}; t_a, t_b)$  for the entire range  $m = [M_{min}, M_{max}]$ , in the time interval  $\Delta t = t_b - t_a$ .

Analogous to Eq. 12, the probability of at least one event for the non-stationary Poisson distribution is defined as:

$$P[N > 0; t_a, t_b] = 1 - e^{-m_{\lambda}(t_a, t_b)(t_b - t_a)}, \quad (17)$$

The non-stationary Poisson distribution, Eq. 13, is also applicable if multiple independent sequences occur such as a constant background seismicity and non-stationary induced seismicity. In this case the rate  $\lambda(t)$  in Eq. 14 to compute the mean rate of occurrence  $m_{\lambda}(t_a, t_b)$  simply becomes the sum of the rates of all sequences, that is:

$$m_{\lambda, tot}(t_a, t_b) = \frac{\int_{t_a}^{t_b} (\lambda_{bg}(s) + \lambda_{ind}(s)) ds}{(t_b - t_a)}, \quad (18)$$

where the background and induced rate of earthquakes are given by  $\lambda_{bg}(t)$  and  $\lambda_{ind}(t)$  respectively. Eqs. 11–18 can be used to compute analytic expectations given known or estimated (non-)stationary GR magnitude-frequency distributions. For instance, Eqs. 12 and 17 can be used to determine the likelihood that an event with a magnitude between 4.5 and 5.5 occurs in the next five years.

The rate of exceedance  $\lambda_{exc}(m \geq m_j; t)$  per unit time duration for a magnitude level is another useful statistical variable for seismic hazard analysis. It represents the number of events at time  $t$  in excess of a certain magnitude per unit time duration. It is defined as:

$$\lambda_{exc}(m \geq m_j; t) = P(m_j \leq m \leq M_{max}; t) N(M_{min} \leq m \leq M_{max}; t), \quad (19)$$

where  $N(M_{min} \leq m \leq M_{max}; t)$  is given by Eq. 8, and  $P(m_j \leq m \leq M_{max}; t)$  is the time-varying probability of occurrence of a magnitude  $m$  occurring in the range  $m = [m_j, M_{max}]$ , that is:

$$P(m_j \leq m \leq M_{max}; t) = F_M(M_{max}; t) - F_M(m_j; t). \quad (20)$$

The cumulative distribution function  $F_M$  is again given by Eq. 10. If required, we can define a separate rate of exceedance per unit time duration for the background seismicity  $\lambda_{exc, bg}$  and the induced seismicity  $\lambda_{exc, ind}$ .

Equations 19 and 20 are extensions of Eqs. 4, 5 and 7 for the rates of events within a single magnitude bin. It is possible to define a mean rate of exceedance  $m_{\lambda, exc}(m \geq m_j; t_a, t_b)$  for a magnitude level in a time interval  $\Delta t = t_b - t_a$ , by inserting the appropriate exceedance rate  $\lambda_{exc}(m \geq m_j; t)$  into Eq. 14. Note also that by inserting the mean rate of exceedance  $m_{\lambda, exc}(m \geq m_j; t_a, t_b)$  into Eq. 17, and then inverting the resulting expression, we obtain:

$$m_{\lambda, exc}(m \geq m_j; t_a, t_b) = \frac{-\ln(1 - P[N > 0; t_a, t_b])}{t_b - t_a}. \quad (21)$$

Equation 21 is used to relate the probability  $P[N > 0; t_a, t_b]$  of at least one event to exceed a magnitude  $m$  in a time interval  $[t_a, t_b]$  to a mean annual rate of exceedance  $m_{\lambda, exc}(m \geq m_j; t_a, t_b)$ . This is useful since probability  $P[N > 0; t_a, t_b]$  is often provided in seismic hazard analyses as will be shown later. This expression is valid for stationary and non-stationary sequences. However, for non-stationary sources, the time interval  $[t_a, t_b]$  has to be identical for both quantities.

If the GR parameters have known uncertainties then it is possible to compute upper, lower and middle (average) curves, reflecting for instance one times the standard deviation. This will be described in more detail in the implementation section.

### 2.3. Generation of Synthetic Earthquake Catalogs Using the Monte Carlo Method

The quantities shown previously can also be obtained from synthetic earthquake catalogs by counting, creating simultaneously a verification procedure to ensure all computations are correct. On the other hand, the generation of these synthetic earthquake catalogs are the first step in the PSHA

methodology using the Monte Carlo simulation method.

The generation of synthetic earthquake catalogs using the Monte Carlo method can be summarized in two main steps: (1) simulation of a temporal point process to obtain event origin times; (2) simulation of earthquake magnitudes. For the non-stationary case, the simulation of temporal point process is given by the thinning method for a non-stationary Poisson process (Sigman 2013; Zhuang and Touati 2015). This method provides the event origin times of the synthetic events, in the context of non-stationarity.

The simulation of a stationary Poisson process is performed by applying Monte Carlo sampling to the inverse cumulative distribution function (CDF) of the Poisson distribution with constant rate  $\lambda$  (Zhuang and Touati 2015):

$$\tau = \frac{-\ln(r)}{\lambda}, \quad (22)$$

where  $\tau$  is a random temporal variate,  $\lambda$  is the rate of occurrence, and  $r$  is a random number obtain from a uniform distribution between  $[0,1]$ . In order to generate a sequence of events in the time frame  $[t_a, t_b]$ , we define the following algorithm:

**Algorithm 1:** Simulation of event times of a stationary Poisson process with rate  $\lambda$  between times  $t_a$  and  $t_b$  (Zhuang and Touati 2015):

1. Set  $t = t_a$ ,  $K = 0$ .
2. Generate  $r$ .
3.  $t = t + \frac{-\ln(r)}{\lambda}$ . If  $t \geq t_b$ , then stop.
4. Set  $K = K + 1$  and set  $t_K = t$ .
- 5 Go to step 2

where  $t_K$  is the vector that contains the desired event origin times and  $K$  the number of event origin times. However, for the thinning Poisson process to handle non-stationary sequences, an extra 'rejection' step is added and the rate of occurrence of events  $\lambda$  is a function of time.

For the thinning process, we simulate a stationary Poisson process at rate  $\lambda^*$ , where  $\lambda^* \geq \max(\lambda(t))$ . The rate  $\lambda^*$  is larger than needed for the actual process, so for each simulated time arrival, we independently generate another random number  $r$  to decide whether to keep it or reject it. If  $r \leq \lambda(t)/\lambda^*$ , we keep the arrival time (Sigman 2013).

**Algorithm 2:** Simulation of event times of a non-stationary Poisson process with rate  $\lambda(t)$  between times  $t_a$  and  $t_b$  (Sigman 2013; Zhuang and Touati 2015):

1. Consider  $\lambda^*$  such that  $\lambda^* \geq \max(\lambda(t))$ .
2. Set  $t = t_a$ ,  $K = 0$ .
3. Generate  $r$ .
4.  $t = t + \frac{-\ln(r)}{\lambda^*}$ . If  $t \geq t_b$ , then stop.
5. Generate  $r$ .
6. If  $r \leq \lambda(t)/\lambda^*$ , then set  $K = K + 1$  and set  $t_K = t$ .
7. Go back to step 3.

Note that time  $t$  keeps advancing in this algorithm irrespective if an event is accepted or rejected. For the generation of synthetic earthquake catalogs in the entire range  $m = [M_{min}, M_{max}]$ , the rate of occurrence  $\lambda(t)$  is equivalent to the time-dependent rate of earthquakes  $\lambda(M_{min} \leq m \leq M_{max}; t)$ , Eq. 15.

Once we generate the  $K$  event origin times in the time period  $[t_a, t_b]$ , we sample the GR distributions considering that the event origin times are grouped in intervals equivalent to the time samples used to describe the temporal evolution of the GR parameters. For the generation of magnitudes  $m$ , we apply Monte Carlo sampling to the inverse cumulative distribution function (CDF) of the GR distribution (Zhuang and Touati 2015):

$$m = \frac{-\ln(r)}{b(t)\ln(10)} + M_{min}, \quad (23)$$

where  $r$  is again a random number obtain from a uniform distribution between  $[0,1]$ . During the sampling, any  $m > M_{max}$  is excluded (Truncated GR distribution), and the sampling is repeated until we get  $K$  values with  $m \leq M_{max}$ . We pair the event origin times in a time sample  $t'$ , with magnitudes sampled from the GR distribution at time  $t'$ , defined specifically by  $a(t')$  and  $b(t')$ .

By repeating the simulation of event origin times and sampling of the GR distributions, we create multiple realizations of the synthetic earthquake catalog. The use of multiple independent realizations is useful since it allows for computing more robust statistics in particular for short or rapidly varying sequences.

It is possible to simulate multiple independent processes. For instance, we could have stationary

background seismicity combined with a time-limited induced seismicity sequence, each described by its own set of GR distributions, Eq. 1. In this case, the background seismicity would have an intercept  $a_{bg}$  and slope  $b_{bg}$  which are time independent, whereas the induced seismicity sequence would have an intercept  $a_{ind}(t)$  and slope  $b_{ind}(t)$  leading to different earthquake rates  $\lambda_{bg}$  and  $\lambda_{ind}$  in Eqs. 5 and 15.

The key is to create each synthetic catalog independently and separately. Once created they can be combined for further analysis. As a practical note we recommended using the same time units, time durations and again the same  $M_{min}$  and  $M_{max}$  magnitudes. How to deal with uncertainties in estimated GR parameters will be described in the next section.

### 3. Implementation

The generation and evaluation of the synthetic magnitude catalogs is done in three steps, namely: (1) computation of the relevant earthquake rates, (2) Monte Carlo simulation, and (3) computation of the analytic expectancies and verification.

#### 3.1. Computation of the Earthquake Rates

For a given set of seismic parameters ( $a(t)$ - and  $b(t)$ -values,  $M_{min}$  and  $M_{max}$ ), we calculate the time-dependent rate of earthquakes  $\lambda(M_{min} \leq m \leq M_{max}; t)$ , Eq. 15. The rates are calculated for each time sample  $t$  in a given range (e.g.  $t = [t_a; t_b]$ ), where the time samples are defined by the used time unit (e.g. day, week, month, year).

We will assume that the seismic parameters are known, for instance from historic catalogs in the case of natural (background) seismicity or by evaluating current and past induced seismic catalogs. If the appropriate seismic parameters are unknown as is likely for future induced seismicity then the proposed methodology still allows for evaluating scenarios where for instance the GR intercept  $a_{ind}$  is twice that of the background seismicity  $a_{bg}$  for a limited time-frame.

One caveat is that it is important to ensure that the derived  $a$ -value is normalized per unit area when

comparing different catalogs such as for natural and induced seismicity. This explains for instance the role of the activation parameter as used by (Ghofrani and Atkinson 2016), which plays a normalization role instead of representing a likelihood of occurrence.

#### 3.2. Monte Carlo Simulation

For the generation and evaluation of synthetic earthquake catalogs, we define five steps: (1) generation of event origin times, (2) generation of magnitudes, (3) multiple processes, (4) inclusion of uncertainties, and (5) extract relevant statistics.

*Generation of event origin times* Following the thinning method for non-stationary Poisson process (algorithm 2), we can generate a single realization of event origin times for the rate of earthquakes  $\lambda(M_{min} \leq m \leq M_{max}; t)$ . Again, these will be the event origin times for the magnitudes in the full range  $m = [M_{min}, M_{max}]$ . We can repeat this process until we generate  $N_r$  realizations of event origin times, see Fig. 1. In the case of stationary sources, we simply apply the stationary Poisson simulation (algorithm 1), keeping a constant rate of earthquakes  $\lambda(M_{min} \leq m \leq M_{max})$ .

*Generation of magnitudes* Once the event origin times have been generated, we group them into time intervals that correspond to the different GR distributions per time sample. Then, we sample each GR distribution, using Eq. 23, and pair the sampled magnitudes with the corresponding event origin times contained in that time sample. The grouping is purely done to reflect that observed GR parameters are always estimated within certain time interval in historical earthquake catalogs. This step is not required for continuous distributions.

To make comparisons between the simulations and the statistical quantities possible, we group the synthetic magnitudes in magnitude bins, given a defined bin size. This is needed since all analytical occurrence statistics are computed for magnitude ranges (e.g. Eq. 5).

The resulting synthetic earthquake catalog contains event time, realization number and magnitude, for a seismic source defined by either stationary or non-stationary GR parameters. Such catalogs are equally needed for analysis of ground motion in a full

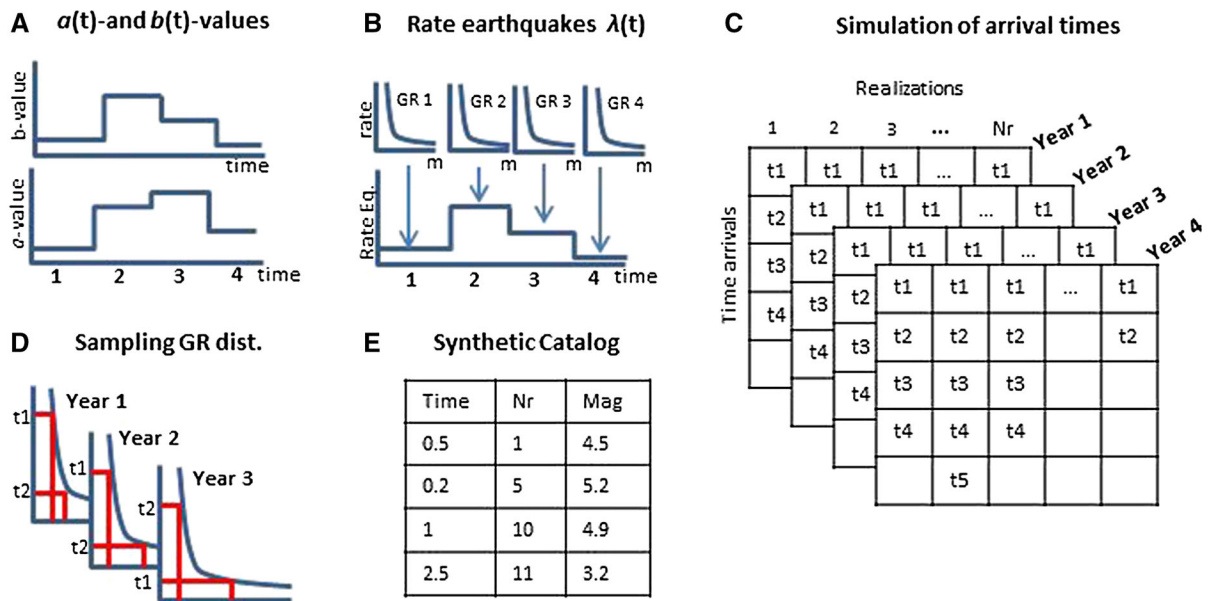


Figure 1

Sketch of the Monte Carlo simulation methodology for the generation of non-stationary earthquake catalogs. **a** Definition of the time-dependent  $a(t)$ -and  $b(t)$ -values, and **b** equivalent rate of earthquakes. Notice how each temporal sample might correspond to a different GR distribution. **c** Generation of multiple realizations containing different event origin times, using the thinning method for non-stationary Poisson process. **d** Sampling of the GR distributions. **e** Synthetic catalog containing time, realization and magnitude

seismic hazard analysis (Assatourians and Atkinson 2013). Figure 1 shows a sketch of the Monte Carlo simulation methodology for the generation of non-stationary earthquake catalogs.

*Multiple processes* If multiple processes occur such as natural and induced seismic sources with different statistical properties, then we simulate each process independently using their appropriate GR parameters, creating two or more separate synthetic catalogs. The processes may have different time durations or activity levels (See synthetic example, Fig. 2). The synthetic catalogs can be combined in order to extract the statistics and study the related hazard.

*Inclusion of uncertainties* During the estimation of the GR parameters from historical catalogs, we take the uncertainties in the  $a$ -and  $b$ -values into account. Hazard studies (Halchuk et al. 2014) rely on the error in the  $b$ -value by defining 3 sets of GR parameters, namely upper ( $b$ -value + error), lower ( $b$ -value – error), and middle curves ( $b$ -value). The corresponding  $a$ -values are correlated to the  $b$ -values,

and they are simply calculated from the  $b$ -values and  $N$ , Eq. 1.

PSHA uses the logic tree approach, which gives a weight to each set of GR parameters (Assatourians and Atkinson 2013). Generally, the middle curve gets a weight of 0.68 (Halchuk et al. 2014), considering that in a Gaussian distribution 68% of the data values are within one standard deviation of the mean. In order to include the weights, we multiply the number of earthquakes  $N_0$  of each set of GR parameters, Eq. 2, by its corresponding weight. Next, we simulate independently each weighted set of GR parameters (Upper, lower and middle curves) generating 3 different catalogs. Finally, we combine the 3 catalogs to generate a unique synthetic catalog that counts for the uncertainties in the GR parameters. The same three sets of GR parameters are used to compute the analytic expressions, yielding again upper, lower and middle curves.

*Extract relevant statistics* From the synthetic earthquake catalogs we can extract important statistical quantities for the hazard analysis. The mean rate of earthquakes  $m_i(m_j \leq m < m_{j+1}; t_a, t_b)$  in a time

interval can be obtained by counting the number of events within that magnitude bin, and then dividing this count by the length of the time interval  $t = [t_a, t_b]$  and the number of realizations  $N_r$ .

The probability  $P[N = n; t_a, t_b]$  of  $n$  occurrences with magnitudes between  $M_{min}$  and  $M_{max}$  in a time interval  $\Delta t = t_b - t_a$  can be obtained from the synthetic catalogs by counting the number of realizations with 0, 1, 2, ...  $n$  occurrences and dividing by the number of realizations ( $N_r$ ). Clearly this can be done for any magnitude range. Likewise, by dividing the relevant time duration it is possible to obtain expectancies that annually  $n$  events occur in a specified magnitude range for the considered time interval.

Finally, the mean rate of exceedance  $m_{\lambda,exc}(m \geq m_j; t_a, t_b)$  for a magnitude level in a time interval can be obtained by counting the number of events with magnitude  $m$  bigger than a certain magnitude level, and again dividing by the length of the time interval  $\Delta t = t_b - t_a$  and the number of realizations  $N_r$ . It is important to use multiple realizations for non-stationary processes in particular

for those that are very limited in duration, since all derived statistical quantities have estimation variances that are inversely proportional to the number of realizations. Contrary to stationary sequences, it is not possible in these cases to average over longer time durations.

### 3.3. Analytical Expectancies and Verification

In order to verify the Monte Carlo simulation results, we will compare some statistical quantities derived from the simulations with the equivalent analytical quantities. These analytical quantities are: (1) the probability  $P[N = n; t_a, t_b]$  of  $n$  occurrences with magnitudes between  $M_{min}$  and  $M_{max}$  in a time interval  $\Delta t = t_b - t_a$ ; (2) the mean rate of earthquakes  $m_{\lambda}(m_j \leq m < m_{j+1}; t_a, t_b)$  for a magnitude bin in a time interval  $\Delta t = t_b - t_a$ ; and (3) the mean rate of exceedance  $m_{\lambda,exc}(m \geq m_j; t_a, t_b)$  for a magnitude level in a time interval  $\Delta t = t_b - t_a$ .

The probability  $P[N = n; t_a, t_b]$  of  $n$  occurrences with magnitudes between  $M_{min}$  and  $M_{max}$  in a time interval  $\Delta t = t_b - t_a$ , is calculated theoretically using

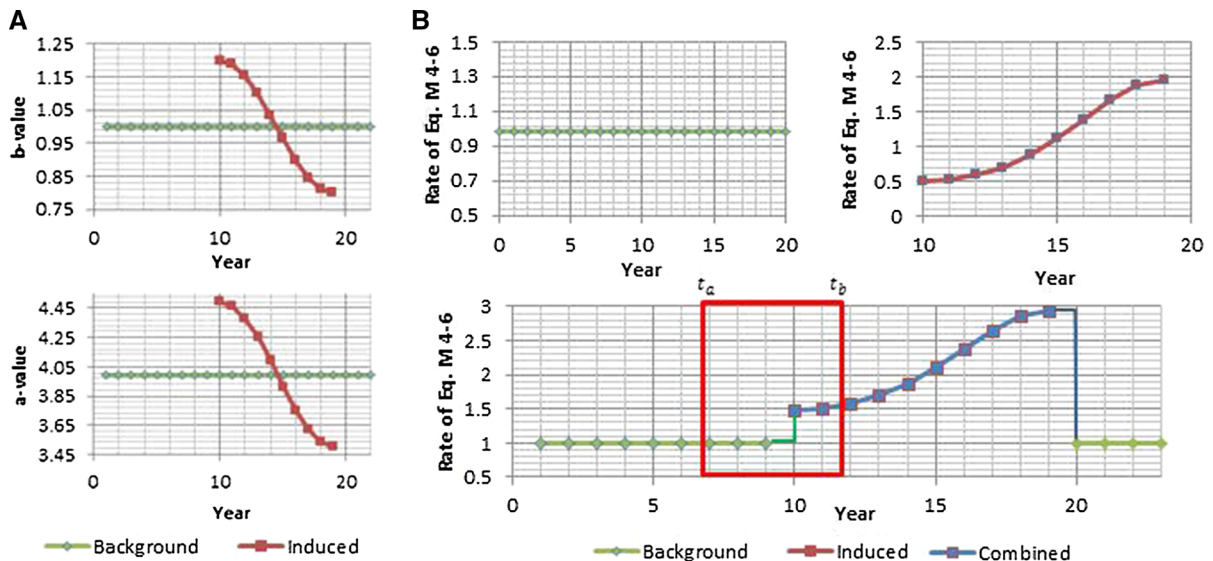


Figure 2

**a** Temporal evolution of the annual  $a(t)$ - and  $b$ -values for the background and the induced seismicity. **b** Equivalent annual rate of earthquakes in the range  $m = [4.0, 6.0]$  for the background, induced and combined seismicity. The total seismicity rate is the sum of rates for background and induced seismicity, Eq. 18. However induced seismicity is limited to the years 10 through 19. All seismicity returns to the natural pattern starting year 20



Eqs. 13–15. Likewise, the mean rate of earthquakes  $m_\lambda(m_j \leq m < m_{j+1}; t_a, t_b)$  for a magnitude bin in a time interval  $\Delta t = t_b - t_a$ , is given by inserting the rate of earthquakes  $\lambda(m_j \leq m < m_{j+1}; t)$  per unit time duration for a magnitude bin, Eq. 7, into Eq. 14. The rate  $\lambda(m_j \leq m < m_{j+1}; t)$  is given by the expected number of earthquakes  $N(M_{min} \leq m \leq M_{max}; t)$  per unit time duration in the range  $m = [M_{min}, M_{max}]$ , Eq. 8, and the probability of occurrence  $P(m_j \leq m < m_{j+1}; t)$  of a magnitude bin, Eq. 9, where the probability  $P(m_j \leq m < m_{j+1}; t)$  depends on the cumulative distribution function  $F_M(m; t)$ , Eq. 10. The expected number of earthquakes  $N(M_{min} \leq m \leq M_{max}; t)$  and the cumulative distribution function  $F_M(m; t)$  are both directly determined by the GR parameters, and the minimum and maximum magnitude,  $M_{min}$  and  $M_{max}$ , Eqs. 8 and 10.

The mean rate of exceedance  $m_{\lambda,exc}(m \geq m_j; t_a, t_b)$  for a magnitude level in a time interval  $\Delta t = t_b - t_a$ , is given by inserting the rate of exceedance  $\lambda_{exc}(m \geq m_j; t)$  per unit time duration for a magnitude level, Eq. 19, into Eq. 14. Simultaneously, the rate of exceedance  $\lambda_{exc}(m \geq m_j; t)$  is given by the expected number of earthquakes  $N(M_{min} \leq m \leq M_{max}; t)$ , Eq. 8 and the probability of occurrence  $P(m_j \leq m \leq M_{max}; t)$  of a magnitude  $m$  occurring in the range  $m = [m_j, M_{max}]$ , Eq. 20, where the probability  $P(m_j \leq m \leq M_{max}; t)$  depends on the cumulative distribution function  $F_M(m; t)$ , Eq. 10.

These analytical quantities can be directly compared with the same quantities derived from the Monte Carlo simulations for verification. All variables can be extended to include spatial variations in the GR parameters such that ground motion predictions can be made, which is the ultimate goal of probabilistic seismic hazard analysis (Cornell 1968; Assatourians and Atkinson 2013).

#### 4. Synthetic Example

For illustration purposes we consider the following situation. A region has a stationary natural seismicity with GR parameters:  $a_{bg} = 4$ , producing  $N_0 = 10,000$ , Eq. 2, and  $b_{bg} = 1$ . An induced seismicity sequence occurs between the years 10 through 19. Starting year 20, only natural seismicity occurs.

While the induced seismicity is active, the GR parameters are equal to:

$$b(t) = 1 + \frac{\cos(\frac{\pi(t-10)}{9})}{5}, \quad (24)$$

$$a(t) = 4 + \frac{\cos(\frac{\pi(t-10)}{9})}{2}, \quad (25)$$

where  $t$  is a discrete temporal sample, starting at  $t = 10$  years. The time unit used in this example is year. We consider a minimum and maximum magnitude of  $M_{min} = 4.0$ ,  $M_{max} = 6.0$  respectively and a magnitude bin size of  $M_{bin} = 0.1$ . Figure 2a shows the temporal variation of the GR values for the background and induced sources. Figure 2b shows the rate of earthquakes per year in the range  $m = [4, 6]$ , for both sources and the resulting combined seismicity in the period [10, 19]. It also shows a hypothetical moving time interval with limits  $t_a$  and  $t_b$ . This time interval will be used to illustrate the changes in the hazard statistics given by the temporal variations in the source parameters.

We compute 5 independent simulations each comprising  $N_r = 10,000$  realizations with a time duration of 30 years. Both sources were simulated independently and their catalogs were combined for the statistical analysis. The number of realizations is chosen because of the low rates for both the lowest and highest magnitudes of interest, combined with the short duration of the induced seismicity sequence. This also allow us to validate the theoretical predictions for non-stationary sequences.

We evaluate the importance of the start  $t_a$  and end  $t_b$  times of the time interval, and how the inclusion of sections with active induced seismicity alters the statistics. Figure 3a, b show how the mean annual rate of earthquakes  $m_\lambda(m_j \leq m < m_{j+1}; t_a, t_b)$  and the mean annual rate of exceedance  $m_{\lambda,exc}(m \geq m_j; t_a, t_b)$ , change for 10-year length intervals with different start times. The total rate while both processes are active is simply the sum of both processes. As expected, the mean rates increase with the duration of induced seismicity in the interval. For instance, the annual rate of earthquakes  $\lambda(4 \leq m < 4.1; t)$  for the magnitude bin  $m = [4, 4.1)$  equals 0.2, 0.27 and 0.42 events per year for the periods of  $t = [0, 9]$ ,  $t = [5, 14]$ , and  $t = [10, 19]$ , respectively (Fig. 3a).

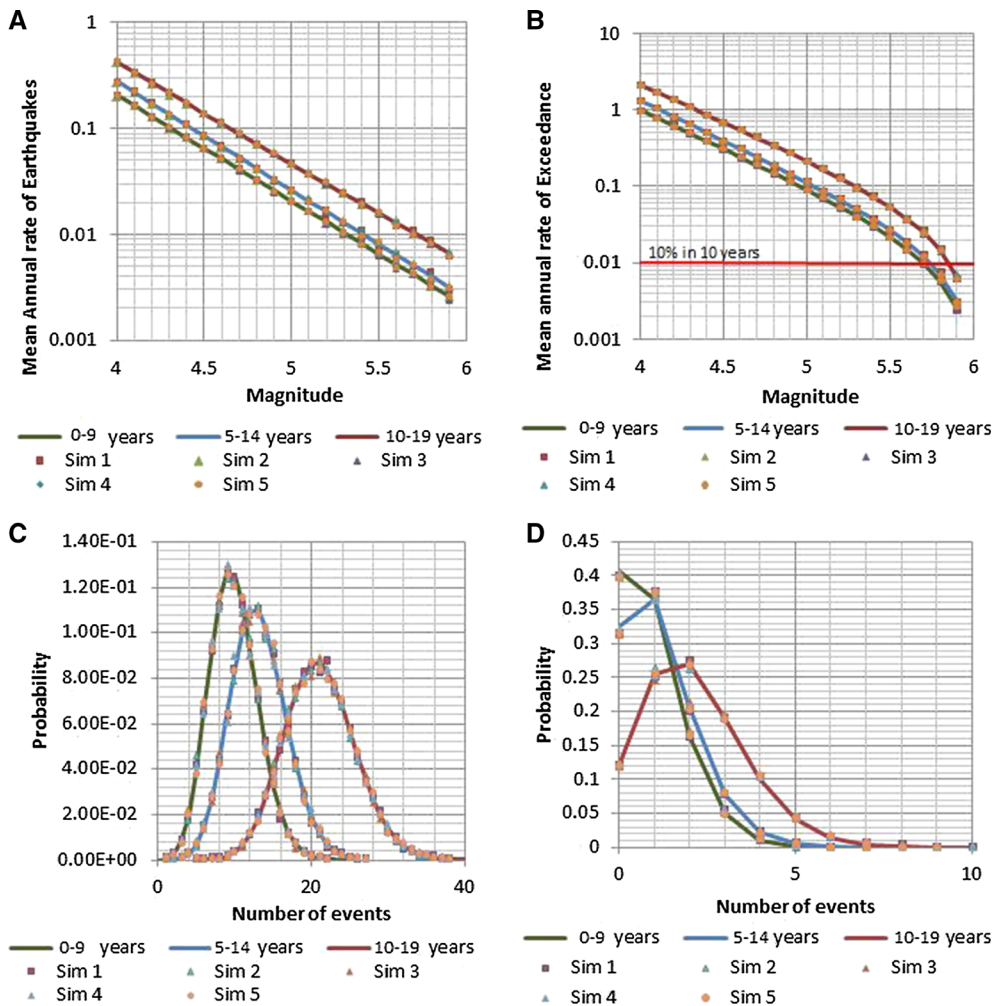


Figure 3

**a** Mean annual rate of earthquakes as a function of magnitude within a 10 year interval starting respectively at 0, 5 and 10 years. **b** Corresponding annual rate of exceedance as a function of magnitude. The likelihood of occurrence of 10% in 10 years is indicated in the plot. **c** Probability as a function of number of occurrences within the magnitude range  $m = [4, 6]$  for the 10 year time interval starting at 0, 5 and 10 years. **d** Equivalent probability as a function of number of occurrences within the magnitude range  $m = [5, 6]$ . The (partial) presence of induced seismicity clearly affects the occurrence statistics. *Theo* theoretical values, *Sim* simulation number, each using 10,000 realizations

The spread in the predictions for different realizations is indicative of the estimation variances of the Monte Carlo simulations. Both theoretical (solid lines) and simulated predictions (dots) agree very well in all cases.

Note that the mean rate of exceedance  $m_{\lambda,exc}(m \geq m_j; t_a, t_b)$  tail off towards the end and deviate from a straight line (Fig. 3b). This occurs because we limited the maximum magnitude  $M_{max}$  to 6. In other words, magnitudes in excess of 6 are not possible, and the annual rate  $\lambda(5.9 \leq m < 6)$  and the

rate of exceedance  $\lambda_{exc}(m \geq 5.9)$  for the largest magnitude bin are thus identical. This indicates in turn that although we compute all occurrence probabilities assuming a Gutenberg-Richter distribution for event magnitudes, generated synthetic event catalogs by Monte Carlo simulation by design do not follow this distribution for the largest magnitude events.

We can relate a probability of exceedance  $P[N > 0; t_a, t_b]$  in a time interval  $[t_a, t_b]$  to an annual rate of exceedance  $\lambda_{exc}(m \geq m_j; t)$  for a magnitude

level, using Eq. 21. In Fig. 3b we highlight the probability of exceedance of 10% in 10 years, which is equivalent to  $\lambda_{exc}(m \geq m_j; t) = 0.01$ . This value results from Eq. 21, using  $P[N > 0; t_a, t_b] = 0.1$  and  $\Delta t = 10$  years. We use a 10 year period since this is the length of the considered time interval. For the 10% probability in 10 years, we find for this hypothetical example  $m = 5.68$ ,  $m = 5.72$  and  $m = 5.85$  for the periods of  $t = [0, 9]$ ,  $t = [5, 14]$ , and  $t = [10, 19]$ , respectively.

Figure 3c shows the probability of  $n$  occurrences within the magnitude range  $m = [4, 6]$ , for different intervals with constant length of 10 years with start times  $t_a$  of respectively 0, 5 and 10 years. The first time interval  $t = [0, 9]$  has no induced seismicity (no overlap), the second interval  $t = [5, 14]$  half overlaps the induced seismicity period, and the third interval  $t = [10, 19]$  full overlaps the period of induced seismicity. The theoretical values, Eqs. 13 and 18, appear as a solid line, and the Monte Carlo simulation results appear as dots. Figure 3d shows the probability of  $n$  occurrences within the magnitude range  $m = [5, 6]$ , for the same intervals with constant length of 10 years. Both figures show how the induced seismicity significantly impacts the magnitude-frequency distributions. For instance, the likelihood of exactly 3 events in the magnitude range  $m = [5, 6]$  to occur is respectively 0.16, 0.20, and 0.26 for the periods of  $t = [0, 9]$ ,  $t = [5, 14]$ , and  $t = [10, 19]$  (Fig. 3d). As we can see in these examples, the amount and duration of induced seismicity contained in the time interval alters the hazard statistics.

To illustrate the influence of the length of the time interval  $[t_a, t_b]$ , the duration of the predictions, we evaluate different lengths while both sources are active. The time intervals are changing to respectively 1, 3, 5 or 10 years length, all starting at  $t_a = 10$  years. As before, we compute both the theoretical expectations and five Monte Carlo simulations using 10,000 realizations each.

Figure 4a, b show the mean annual rate of earthquakes  $m_\lambda(m_j \leq m < m_{j+1}; t_a, t_b)$  and the mean annual rate of exceedance  $m_{\lambda,exc}(m \geq m_j; t_a, t_b)$ , for the different time interval lengths. As we expected, the mean annual rate of earthquakes increases while the time interval includes increasing amounts of accelerating induced seismicity. For instance, the annual

rate of earthquakes for the magnitude bin  $m = [4.5, 4.6]$  equals 0.094, 0.095, 0.10 and 0.13 events per year for time intervals with 1, 3, 5 and 10 years length, respectively (Fig. 4a). Given the design of this synthetic example, The curves in Fig. 3 for the time interval  $t = [10, 19]$ , are equivalent to the curves in Fig. 4 for the 10-years time interval. Both cases share the same interval length and period, and thus seismicity.

Figure 4c, d show the probability of  $n$  occurrences within the magnitude range  $m = [4, 6]$  and  $m = [5, 6]$ , respectively, for the same four models. Again, the solid lines show the theoretical values and the dots show the results from the Monte Carlo simulations. As in Fig. 3d, c, the probability to have at least one occurrence increases while the time interval includes period of combined induced and background seismicity. For instance, the most likely number of events in the range  $m = [4, 6]$  is 1, 4, 8 and 21 for the time intervals with 1, 3, 5 and 10 years length, respectively (Fig. 4c).

## 5. Horn River Basin Case

### 5.1. Area of Study and Data

We study the impact of induced seismicity on hazard analysis, using the recent activity in the Horn River Basin, Northeast B.C. as an example. Several studies have been made in the Horn River Basin due to the significant increase of seismicity related to the hydraulic fracturing activities conducted between Dec. 2006 and Dec.2011 (BC oil and gas commission 2012; Farahbod et al. 2015a, b), particularly in the Etsho area (Fig. 5a, b). The detected seismicity in the area was very low prior to 2006, but with an important increase since Dec. 2006, particularly between 2010 and 2011 in line with the amount of human activity.

Due to the lack of recorded seismicity at the Horn River Basin, we assume that the GR parameters before Dec. 2006, for natural seismicity, are based on the GR parameters described by the 2015 National seismic-hazard model of Canada (Halchuk et al. 2014). For the period between Dec. 2006 and Dec. 2011, we assume that the GR parameters are based on

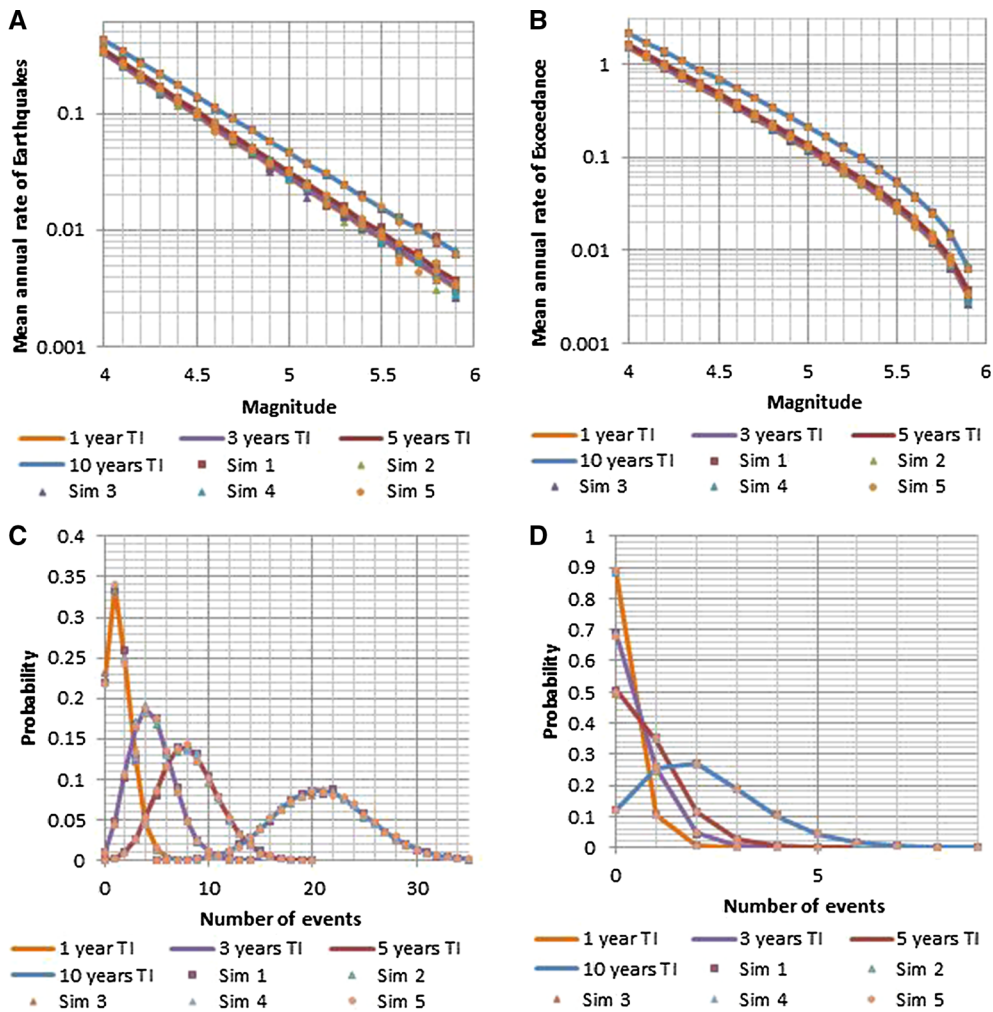


Figure 4

**a** Mean annual rate of earthquakes as a function of magnitude within 1, 3, 5 and 10 year length intervals, starting at year 10 (First year of induced seismicity). **b** Corresponding annual rate of exceedance as a function of magnitude. **c** Probability as a function of number of occurrences within the magnitude range  $m = [4, 6]$  for the 1, 3, 5 and 10 year length intervals. **d** Equivalent probability as a function of number of occurrences within the magnitude range  $m = [5, 6]$ . The time interval length, as well as the inclusion of induced seismicity, affects the occurrence statistics. *Theo* theoretical values, *Sim* simulation number, each using 10,000 realizations. *TI* time interval length

calculations made using the catalog from Farahbod et al. (2015b), which contains induced earthquakes in the Horn River Basin. This catalog consists of 338 events recorded between Dec. 2006 and Dec. 2011, with magnitudes ranging between  $m = 1.0$  and  $m = 3.6$ .

Analysing the catalog from Farahbod et al. (2015b), it is possible to distinguish 2 periods where the induced seismicity has clearly different recurrence statistics: a first period with lower earthquake

rates between Dec. 2006 and Dec. 2009, and a second period with higher rates between Dec. 2009 and Dec. 2011. Figure 5c shows the earthquake magnitudes vs. time, as well as the indicated two periods with different recurrence statistics. The difference between periods is thought to be related to a considerable increase in the injection rates after Dec. 2009 (Farahbod et al. 2015a), see Fig. 5d, leading to increased earthquake activity. After Dec. 2011 we

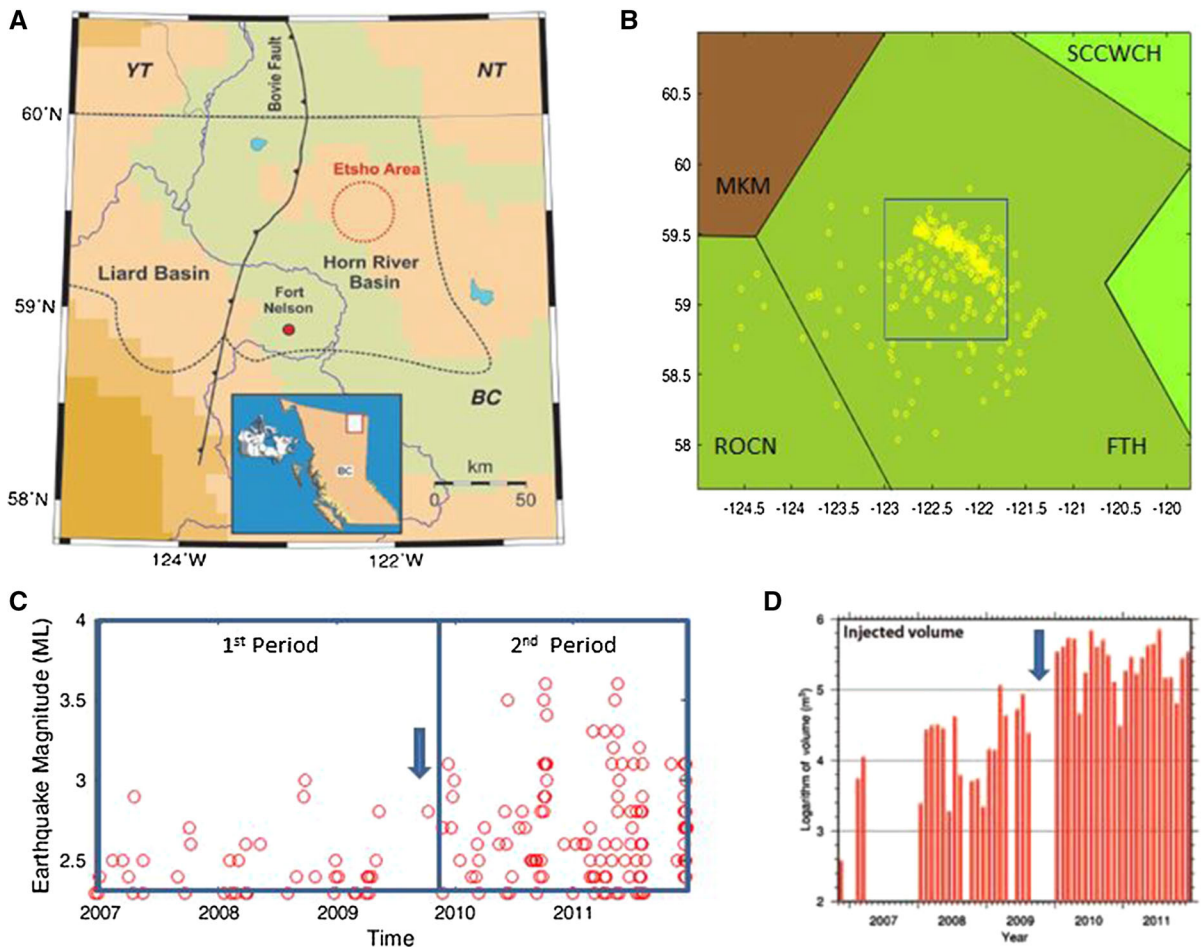


Figure 5

**a** Location of the Horn River Basin, Northeast British Columbia. The Etsho area (circle in red) is described as the area with much seismic activity in the basin [from Farahbod et al. (2015b)]. **b** Location of the seismic events (yellow dots) using the catalog from Farahbod et al. (2015b). The blue square (IS) indicates the seismic source area used in this study. The polygons in the picture represent some of the natural sources defined by the 2015 seismic-hazard model of Canada (Halchuk et al. 2014). *FTH* foothills, *ROCN* Rocky Mountain fold/thrust belt north, *MKM* Mackenzie Mountains, *SCCWCH* stable cratonic core western Canada H. **c** Earthquake magnitudes as a function of time, using the catalog from Farahbod et al. (2015b). The two distinctive periods of induced seismicity are shown. **d** Monthly injection rates in the Horn River Basin. Notice the substantial increase in the injection volumes after Dec. 2009 (vertical arrows). From Farahbod et al. (2015a)

assume that the seismicity rates return to their natural state.

### 5.2. Seismic Parameters from the Different Time Periods

We focus on the Etsho area within the Horn River Basin (Fig. 5a, b) where much of the induced seismicity occurred between 2006 and 2011 (BC oil and gas commission 2012). We first compute relevant GR parameters for the natural and induced sequences.

Because natural and induced seismicity cover different areas, one must normalize the derived  $a$ -values per unit area before computing occurrence statistics.

For the GR parameters of natural seismicity, before Dec. 2006 and after Dec. 2011, we use the Foothills source area, as described by Halchuk et al. (2014). The total Foothills area is  $308,349.10 \text{ km}^2$ . To count for the uncertainty in the GR parameters, Halchuk et al. (2014) uses a mix of three GR distributions for this area. Table 1 shows the GR parameters for the Foothills area and the rescaled

Table 1

GR parameters for the natural seismicity at the Foothills area and the rescaled values for the Etsho area. These are annual  $a_{nat}$ -values.  $w$ : weight of each GR distribution

	Lower curve ( $w = 0.16$ )	Middle curve ( $w = 0.68$ )	Upper curve ( $w = 0.16$ )
GR foothills	$a_{nat,3} = 1.84$ $N_0 = 69.18$ $b_{nat,3} = 0.64$	$a_{nat,1} = 2.43$ $N_0 = 269.15$ $b_{nat,1} = 0.8685$	$a_{nat,2} = 3.034$ $N_0 = 1081.43$ $b_{nat,2} = 1.090$
GR etsho	$a_{etsho,nat,3} = 0.26$ $N_0 = 1.837$ $b_{nat,3} = 0.64$	$a_{etsho,nat,1} = 0.85$ $N_0 = 7.12$ $b_{nat,1} = 0.8685$	$a_{etsho,nat,2} = 1.44$ $N_0 = 28.17$ $b_{nat,2} = 1.090$

values for the Etsho area. The weights of each distribution are shown in the table. The rescaled  $N_0$  are obtained by first calculating the density of earthquakes and then by multiplying these densities with the area of induced earthquakes. Using Eq. 2, we have the following earthquake densities: (1)  $8.88 \times 10^{-4}$ , (2)  $3.51 \times 10^{-3}$ , and (3)  $2.29 \times 10^{-4}$  earthquakes/km<sup>2</sup>/year. The Etsho area is 8,262.57 km<sup>2</sup> (blue square Fig. 5b).

Finally when combining all three sets, the  $N_0$  values of each set are first multiplied with their respective weights. For the analytical results, we calculate the rates from the weighted sets, and then sum the results. In the case of the Monte Carlo simulations, we simulate each weighted set, and then combine all simulation cubes to obtain a unique catalog. It is important to emphasize that using a somewhat different natural seismicity rate does not greatly affect the hazard computations during the period of induced seismicity since the latter dominates then.

To study the temporal evolution of the GR parameters during the periods of induced seismicity, we define a moving time window to screen the catalog. In this case, we define a 1-year sliding time window. Then, the GR values are assigned to the corresponding central month in each time window. This process is repeated by moving the entire time window one month ahead every time, until the central month equals the last month of the catalog. For the end points, the time window will include sections of the natural events.

For the calculation of the GR parameters in each time interval, we use only the earthquakes with

magnitude equal or bigger than the magnitude of completeness ( $M_c$ ) estimated for the Horn River Basin. According to the calculations of Farahbod et al. (2015b), the magnitude of completeness equals  $M_c = 2.4$  for the induced events in the Horn River Basin between Dec.2006 and Dec.2011. To estimate the  $b$ -values, we use a modified version of the maximum likelihood method [MLM, Aki (1965), Wiemer and Wyss (1997)], that accounts for the magnitude binning. The annual  $a$ -values are calculated from the  $b$ -values and Eq. 1. However, we convert annual  $a$ -values to corresponding monthly ones. This is done by dividing their  $N_0$  values by 12, and then transforming these back using Eq. 2.

The uncertainties in the GR parameters are also included. This is done by adding or subtracting the errors estimated for the  $b$ -values, resulting in 3 curves for both the  $a$ - and  $b$ -values. For the estimations of the  $b$ -value error, we use the method described by Shi and Bolt (1982). For the middle set of GR parameters, we assign a weight of 0.68 for the Monte Carlo simulations. For both the upper and lower set obtained by adding or subtracting the errors in the  $b$ -values, we assign a weight of 0.16.

Figure 6a, b shows the temporal evolution of the  $a$ - and  $b$ -values in the Horn River Basin. Notice the constant GR parameters before and after the induced seismicity period and the decrease of the  $b$ -values around Dec. 2010, likely associated with the increased injection rates. Again, the temporal evolution of GR parameters indicates 2 distinctive periods of induced seismicity. For all periods, we consider a minimum and maximum magnitude of  $M_{min} = 2.5$ , and  $M_{max} = 5.0$  respectively, based on the range of magnitude earthquakes recorded in the catalog.

We choose  $M_{max} = 5.0$  based on previous hazard assessments for induced seismicity in Western Canada. For instance, Atkinson et al., (2015) made a preliminary hazard evaluation for Fox Creek (Alberta, Canada) using values for the maximum magnitude between  $M_{max} = 4.5$  and  $M_{max} = 6.5$ . These values are smaller than the maximum magnitude for natural seismicity in the area, which range between  $M_{max} = 6.5$  and  $M_{max} = 7.5$ . The largest earthquake recorded in the Horn River Basin is  $m = 3.6$ . Based on these observations, we consider that a  $M_{max} = 5.0$  would be appropriate for the hazard

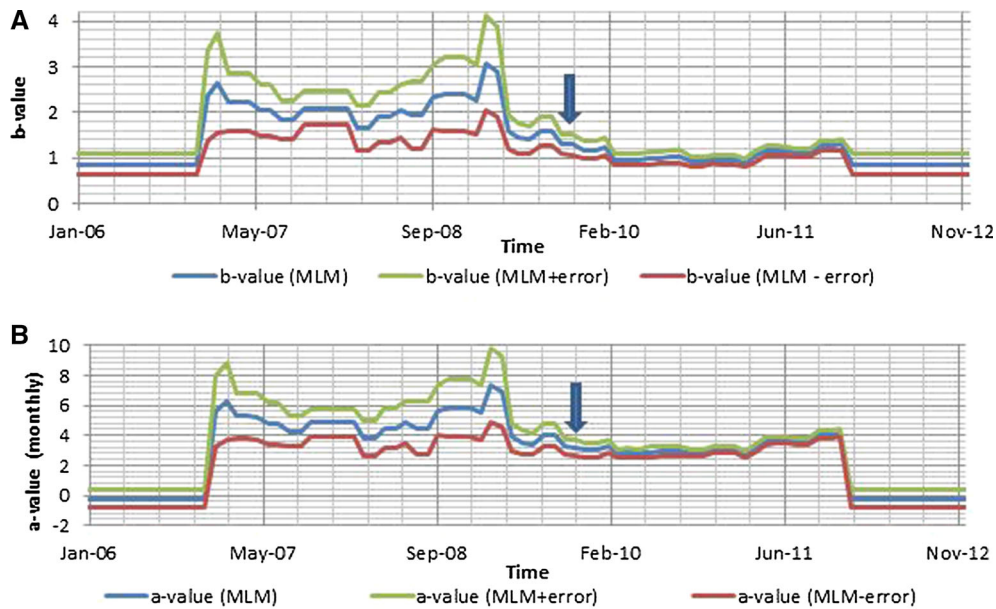


Figure 6

Temporal evolution of the **a** *b*-values, and **b** monthly *a*-values in the Horn River Basin. For the periods before Dec. 2006 and after Dec. 2011, only background seismicity is assumed. The induced seismicity parameters were calculated from the observed catalog (Fig. 5c). The uncertainties in the GR parameters are also included, by adding or subtracting the errors in the *b*-values. The green, blue and red curves describe the set of GR parameters for the upper, middle and lower curves in the occurrence statistics. Vertical arrow indicates change in injection volume, separating first and second period of induced seismicity. MLM = Maximum likelihood method

analysis instead of  $M_{max} = 7.5$  as used in the fifth generation seismic hazard model for Canada (Halchuk et al. 2014). We will briefly address this point further in the discussion section.

Contrary to the synthetic example we do not simulate the natural and induced sequence separately but assume a single time-varying GR distribution in both the analytic computations and Monte Carlo simulations. This simplifies the calculations but also means we do not need to identify natural from induced events between Dec. 2006 through Dec. 2011.

### 5.3. Hazard Analysis

We compute 5 independent simulations each comprising  $N_r = 10,000$  realizations with a time duration of  $N_t = 120$  months (Dec. 2004 up to Dec. 2014), based on the seismicity and GR parameters described before. We consider a magnitude bin size of  $M_{bin} = 0.1$ , and monthly rates of earthquakes. In the case of the background seismicity, the

transformation to monthly rates leads to the following *a*-values:  $a_{etsho,nat,1} = -0.226$ ,  $a_{etsho,nat,2} = 0.37$ ,  $a_{etsho,nat,3} = -0.81$ . For consistency purposes, we later transform the analytical and simulated monthly rates to annual rates of earthquakes by multiplying by 12.

Figure 7a, b show the predictions for the mean annual rate of earthquakes  $m_\lambda(m_j \leq m < m_{j+1}; t_a, t_b)$ , and the mean annual rate of exceedance  $m_{\lambda,exc}(m \geq m_j; t_a, t_b)$ , for the periods before and after the induced events, as well as the first and second period of the induced seismicity. Both the theoretical predictions and the quantities derived from the five Monte Carlo simulations are plotted. The upper, middle and lower analytic curves are included in order to show the variable range of rates generated by the uncertainties in the GR parameters. The occurrence statistics of these curves are calculated by using the time-dependent GR parameters described in Fig. 6 and the expressions in the theory section. These curves are not necessarily straight lines, because they

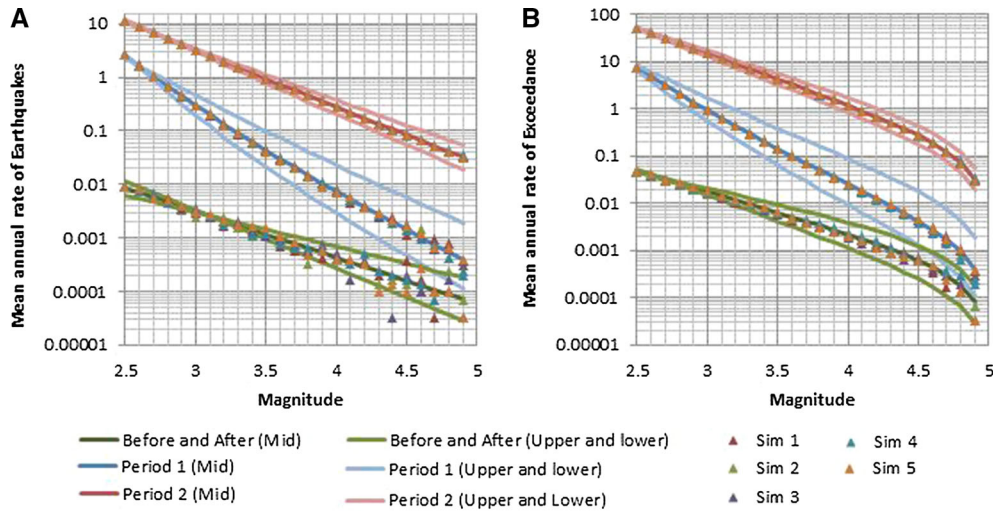


Figure 7

**a** Mean annual rate of earthquakes as a function of magnitude for the periods before and after the induced seismicity, as well as the first and second period of induced seismicity. **b** Corresponding mean annual rate of exceedance as a function of magnitude. The continuous lines show the analytical values for the upper, middle and lower curves resulted from the averaging of the GR parameters in Fig. 6. *Sim* simulation number, each using 10,000 realizations

are obtained by averaging multiple curves resulting from time-dependent GR parameters.

For instance, the mean annual rate of earthquakes  $m_\lambda(2.5 \leq m < 2.6; t)$ , for the magnitude bin  $m = [2.5, 2.6)$ , is equal to 0.0088 events per year for the periods before and after the induced seismicity, and increases to 2.59 and 11.44 events per year for the first and second period of induced seismicity, respectively. On the other hand, the annual rate of exceedance  $m_\lambda(m \geq 2.5)$  for magnitudes  $m \geq 2.5$ , is equal to 0.048 events per year for the periods before and after the induced seismicity, and increases to 7.39 and 51.49 events per year for the first and second period of induced seismicity, respectively. The catalog contains annually respectively 7 and 48.5 events of  $m \geq 2.5$  for the same periods. This confirms the substantial increase in seismicity in the period Dec. 2006 through Dec. 2011 in the Horn River Basin, as shown in Fig. 5c.

Using different time intervals  $[t_a, t_b]$ , we compute the probability of  $n$  occurrences within the magnitude range  $m = [2.5, 4)$ , in order to study the impact of the induced seismicity in the Horn River Basin. We choose the ranges  $m = [2.5, 4)$  to properly compare the predictions with the recorded catalogs, since the

largest recorded magnitude was  $m = 3.6$  (Fig. 5c). Figure 8a shows the probability of  $n$  occurrences (Eq. 17), within the magnitude ranges  $m = [2.5, 4)$ , for a 3-year time interval for the natural seismicity, that is, before Dec. 2006 and after Dec. 2011. This plot shows, for instance, that the likelihood of zero natural events in the magnitude range  $m = [2.5, 4)$  to occur within three years equals 0.87.

Likewise, Fig. 8b shows the probability of  $n$  occurrences within the magnitude range  $m = [2.5, 4)$ , for the 3-year interval in the first period of induced seismicity (Dec. 2006–Dec. 2009). The most likely number of events, within the magnitude range  $m = [2.5, 4)$ , is 21 events with a probability of 0.088. This prediction is similar to the actual number of earthquakes  $m = [2.5, 4)$  recorded in the catalog of Farahbod et al. (2015b) between Dec. 2006 and Dec. 2009, which was 21 events.

Finally, Fig. 8c shows the probability of  $n$  occurrences within the magnitude range  $m = [2.5, 4)$ , for the 2-year interval in the second period of induced seismicity (Dec. 2009–Dec. 2011). The most likely number of events within the magnitude range  $m = [2.5, 4)$ , is 99 events with a probability of 0.040. This prediction is similar to the actual number



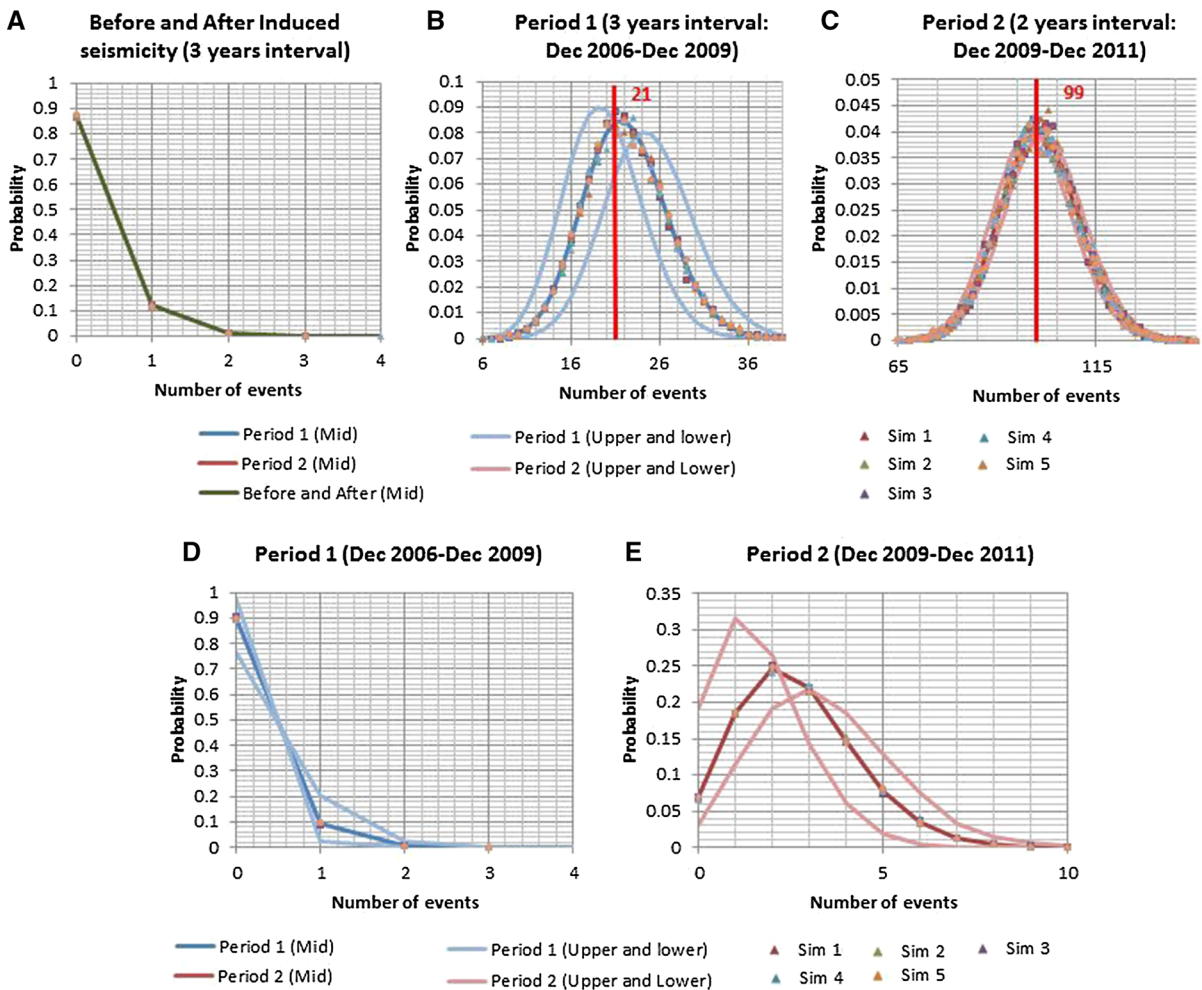


Figure 8

Probability as a function of the number of occurrences within the magnitude range  $m = [2.5, 4.0]$  for **a** the periods before and after induced seismicity, **b** the first period of induced seismicity and **c** the second period of induced seismicity. Once more, the presence of induced seismicity clearly affects the occurrence statistics. The actual number of events in the first and second period of induced seismicity are indicated with a red line. Probability as a function of the number of occurrences within the magnitude range  $m = [4.0, 5.0]$  for **d** the first period of induced seismicity, and **e** the second period of induced seismicity. These results confirm the change in the hazard statistics between the two periods of induced seismicity. The continuous lines show the analytical values for the upper, middle and lower curves resulting from the uncertainties in the GR parameters. *Sim* simulation number, each using 10,000 realizations

of earthquakes  $m = [2.5, 4)$  recorded between Dec. 2009 and Dec. 2011, which was also 99 events.

We can also obtain the likelihood of larger-magnitude events, for instance, the probability of  $n$  occurrences within the magnitude range  $m = [4.0, 5.0]$ . Figure 8d shows the probability of  $n$  occurrences within the magnitude range  $m = [4.0, 5.0]$ , for the 3-year interval in the first period of induced seismicity (Dec. 2006–Dec. 2009). The most likely number of events was 0, with a

probability of 0.90. Figure 8e shows the probability of  $n$  occurrences within the magnitude range  $m = [4.0, 5.0]$ , for the 2-year interval in the second period of induced seismicity (Dec. 2009–Dec. 2011). There is a predicted probability of 0.07 to have no earthquakes within the magnitude range  $m = [4.0, 5.0]$ . Fortunately, no events in this magnitude range actually occurred. We can enumerate a couple of reasons why this prediction seems to fail. First, it is possible that the observations fell in the

0.07 probability of no occurrence. However, it is more likely that the discrepancy between observed and predicted likelihood for an  $m = [4.0, 5.0]$  earthquake is caused by the large uncertainties in the predictions for moderate to larger magnitude events (Figs. 7, 8d).

Traditional PSHA is typically based on earthquake catalogs that cover several decades. Any magnitude events that have average recurrence periods that are well contained (sampled) within the length of available observations will have reasonable estimation variances in terms of their occurrence rates. Conversely, magnitudes that occur only rarely or are not observed will have large uncertainties in estimated occurrence rates. Their occurrence rates are in practice estimated by extending (extrapolating) the frequency-magnitude statistics beyond the range of well-constrained observations. In this case the largest observed magnitude is 3.6. Therefore obtaining reliable estimates for the occurrence likelihoods of events with magnitudes  $m = [4.0, 5.0]$  is very challenging. These uncertainties are further compounded if we are dealing with non-stationary sequences since the magnitude-frequency statistics will now vary over time, thus making it even more difficult to reliably establish the occurrence periods of the largest events of interest.

## 6. Discussion

Traditional seismic hazard analysis has always assumed that the earthquake rates are stationary such that long-term predictions become feasible (Cornell 1968; Baker 2013). Clearly, induced seismicity is determined by anthropogenic patterns and is thus likely strongly correlated to the amount of industrial activity (Brodsky and Lajoie 2013; Langenbruch and Zoback 2016; van der Baan and Calixto 2017; Convertito et al. 2012; Bourne et al. 2014). Treating induced seismicity as a stationary process is thus likely to lead to biased long-term predictions. This is one of the reasons why Petersen et al. (2016) and Petersen et al. (2017) opted for one-year hazard predictions.

The developed analytical expressions and Monte Carlo simulations can handle both stationary and non-

stationary sequences thus allowing for a true assessment of the likelihood of larger magnitude events to occur within a certain timeframe. The good match between the actual number of earthquakes in the catalog and the prediction from the non-stationary Poisson model supports the use of the Poisson model for injection induced seismicity, as suggested by Langenbruch et al. (2011) and Langenbruch and Zoback (2016).

The use of the Poisson model has allowed for computing hazard statistics analytically (Cornell 1968; Baker 2013). The Poisson model assumes however that the earthquakes occur randomly in time and space. This is not accurate since earthquakes tend to cluster temporally and spatially (e.g. as seen in aftershock sequences). Conversely, mainshocks have been shown to be temporally independent Gardner and Knopoff (1974), leading some authors (Gardner and Knopoff 1974; Reasenber 1985) to strongly advocate that earthquake catalogs are declustered by (1) identifying mainshocks, and (2) removing all associated aftershock sequences. The GR parameters are then computed from declustered catalogs, and subsequently used in hazard predictions (e.g., Petersen et al. (2016)). The GR parameters of the declustered catalogs tend to have smaller  $a$ -values and thus a reduced number of predicted earthquakes, Eq. 2. On the other hand, the  $b$ -value is often enlarged, indicating a larger likelihood for the occurrence of larger magnitude events, Eq. 3. Declustered catalogs thus tend to increase the predicted hazard. Other authors argue that declustering is only needed to minimize spatial distortions in earthquake occurrences and may lead to significant underestimation of the true seismic hazard if no compensation is applied to correct for the removed seismicity (Marzocchi and Taroni 2014). Furthermore, final predictions can depend strongly on the used declustering algorithm (van Stiphout et al. 2012).

The Monte Carlo simulation method described here can be changed to handle the occurrence of aftershocks, following mainshocks. However, this would require extensive knowledge of the recurrence patterns both in space and time. This may not be feasible in practice but it would allow for testing the hypothesis if using mainshock/aftershock sequences

instead of a random temporal occurrence has a substantive influence on hazard predictions on the timescales of years to decades. It is important to emphasize however that the short-term non-stationarity due to the occurrence of mainshock/aftershock sequences is different from the intermediate to long-term non-stationarity considered in this paper since the former only have a minor influence on the GR parameters describing long-term time scales. Conversely, induced seismicity can strongly fluctuate as it is determined by the amount of industrial activity (Brodsky and Lajoie 2013; Langenbruch and Zoback 2016; van der Baan and Calixto 2017).

Further work is necessary to create a complete hazard analysis for induced seismicity. Some of the future aspects include:

1. The GR parameters are not known beforehand; however, the developed methodology allows us to evaluate different hazard scenarios from a diverse set of time-varying GR parameters. The GR parameters might be obtained from alternative sources of information, for instance, the seismogenic index (Shapiro et al. 2010), models based on compaction (Bourne et al. 2014, 2015, 2018), or the in-situ state of stress Roche et al. (2015). Future approaches may relate injection rates with  $a$ -values, as proposed by Shapiro et al. (2010). This might be specially useful for current activities in order to forecast hazard seismicity, as applied in Oklahoma by Langenbruch and Zoback (2016).

It is unclear if the total or only mainshock seismicity is proportional to the net injection volumes. For instance, various studies related to salt-water disposal in Texas and Oklahoma, USA, find that the total seismicity, including mainshocks and aftershocks, is proportional to the injected volumes (Keranen et al. 2014; Hornbach et al. 2015; Langenbruch and Zoback 2016). Conversely, Brodsky and Lajoie (2013) determine a direct correlation between the net volume (injected minus produced) and the seismic activity in the Salton Sea Geothermal field after they remove aftershocks using the model of Ogata and Zhuang (2006). Total seismicity and net or injection volume are uncorrelated in their case history.

They thus postulate that only the level of mainshock seismicity is proportional to net volume. Clearly in order to predict temporal changes in seismic hazard it will be very important to establish what causal relationship is most appropriate for a specific region and/or type of industrial activity.

2. An important aspect is the specification of the maximum magnitude  $M_{max}$ . One of the reasons to define a maximum magnitude  $M_{max}$  is due to geological conditions since the magnitude is related to the fault area (Wyss 1979; Scholz 1982). For instance, we would not expect an earthquake larger than a certain magnitude  $m$  if there are no faults of sufficient size. The second reason is related to the very low likelihood of the large magnitude events. Estimation variances in Monte Carlo simulations are proportional to the number of realizations and inversely proportional to the likelihood of occurrence. In other words, very rare events have large estimation uncertainties, in that a single drawn event can greatly influence final predictions. This explains for instance the increasing deviation from the theoretical curves in Fig. 7a, b for the rarest events. To circumvent this issue, Monte Carlo simulations often impose a maximum magnitude to stabilize predictions.
3. The above methodology is very flexible and devised such that it can be easily extended to handle also spatial variations in seismicity in order to generate a full seismic hazard analysis in terms of expected peak ground motion within a certain timeframe. In case of the Monte Carlo simulations this implies defining spatial occurrence statistics and incorporating appropriate ground motion prediction equations, using a similar numerical scheme as used by Assatourians and Atkinson (2013) and Musson (2000).

## 7. Conclusions

A method to compute synthetic earthquake catalogs and associated occurrence earthquake statistics is developed for non-stationary seismicity, using Monte

Carlo simulations. The Poisson model remains relevant for analysing and computing non-stationary induced seismicity. However, non-stationary Gutenberg-Richter (GR) parameters have to be included in order to properly assess the hazard for this type of seismicity. In both examples, tests showed excellent agreements between analytical predictions and numerical results.

In the simulated forecasts, we assume that the GR induced parameters are known. The next steps will include incorporating relationships between earthquake parameters and injection volumes, and extensions to handle spatial source distributions as well as ground motion evaluation in order to generate a complete methodology for non-stationary probabilistic seismic hazard analysis.

#### Acknowledgements

The author would like to thank the sponsors of the Microseismic Industry Consortium for financial support, and Honn Kao for providing an updated event catalog for the Horn River area. The event catalog used in this study is available at: <https://doi.org/10.4095/299419>. We thank Clayton Deutsch for discussion on the Monte Carlo simulation method. We also thank anonymous reviewers for their careful reading and suggestions.

#### REFERENCES

- Aki, K. (1965). Maximum likelihood estimate of  $b$  in the formula  $\log N = a - bM$  and its confidence limits. *Bull. Earthq. Res. Inst.*, *43*, 237–239.
- Anagnos, T., & Kiremidjian, A. (1988). A review of earthquake occurrence models for seismic hazard analysis. *Probabilistic Engineering Mechanics*, *3*, 3–11.
- Assaturians, K., & Atkinson, G. (2013). EqHaz: An open-source probabilistic seismic hazard code based on the Monte Carlo simulation approach. *Seismological Research Letters*, *84*, 516–524.
- Atkinson, G., Ghofrani, H., & Assaturians, K. (2015). Impact of induced seismicity on the evaluation of seismic hazard: Some preliminary considerations. *Seismological Research Letters*, *86*, 1009–1021.
- Atkinson, G., Eaton, D., Ghofrani, H., Walker, D., Cheadle, B., Schultz, R., et al. (2016). Hydraulic fracturing drives induced seismicity in the western Canada sedimentary basin. *Seismological Research Letters*, *87*, 631–647.
- Baker, J. W. (2008). An introduction to probabilistic seismic hazard analysis (PSHA). *Version*, *1*(3), 2017. [https://web.stanford.edu/~bakerjw/Publications/Baker\\_\(2008\)\\_Intro\\_to\\_PSHA\\_v1\\_3.pdf](https://web.stanford.edu/~bakerjw/Publications/Baker_(2008)_Intro_to_PSHA_v1_3.pdf). Accessed Dec.
- Baker, J. W. (2013). Probabilistic Seismic Hazard Analysis. White Paper Version 2.0.1. [https://web.stanford.edu/~bakerjw/Publications/Baker\\_\(2013\)\\_Intro\\_to\\_PSHA\\_v2.pdf](https://web.stanford.edu/~bakerjw/Publications/Baker_(2013)_Intro_to_PSHA_v2.pdf). Accessed Dec 2017.
- B.C Oil and Gas Commission (2012). Investigation of observed seismicity in the Horn River Basin, technical report. [www.bcogc.ca/node/8046/download?documentID=1270](http://www.bcogc.ca/node/8046/download?documentID=1270). Accessed Dec 2017.
- Bourne, S. J., Oates, S. J., van Elk, J., & Doornhof, D. (2014). A seismological model for earthquake induced by fluid extraction from a subsurface reservoir. *Journal of Geophysical Research: Solid Earth*, *119*, 8991–9015.
- Bourne, S. J., Oates, S. J., Bommer, J. J., van Elk, J., & Doornhof, D. (2015). A Monte Carlo method for probabilistic hazard assessment of induced seismicity due to conventional natural gas production. *Bulletin of the Seismological Society of America*, *105*, 1721–1738.
- Bourne, S. J., Oates, S. J., & van Elk, J. (2018). The exponential rise of induced seismicity with increasing stress levels in the Groningen gas field and its implications for controlling seismic risk. *Geophysical Journal International*, *213*, 1693–1700.
- Brodsky, E. E., & Lajoie, L. J. (2013). Anthropogenic seismicity rates and operational parameters at the Salton Sea geothermal field. *Science*, *341*, 543–546.
- Cornell, C. (1968). Engineering seismic risk analysis. *Bulletin of the Seismological Society of America*, *58*, 1583–1606.
- Convertito, V., Maercklin, N., Sharma, N., & Zollo, A. (2012). From induced seismicity to direct time-dependent seismic hazard. *Bulletin of the Seismological Society of America*, *102*, 2563–2573.
- Ellsworth, W. (2013). Injection induced earthquakes. *Science*, *341*, 142–145.
- Farahbod, A., Kao, H., Cassidy, J., & Walker, D. (2015a). How did hydraulic-fracturing operations in the Horn River Basin change seismicity patterns in the northeastern British Columbia, Canada? *The Leading Edge*, *34*, 658–663.
- Farahbod, A., Kao, H., Cassidy, J., & Walker, D. (2015b). Investigation of regional seismicity before and after hydraulic fracturing in the Horn River Basin, northeast British Columbia. *Canadian Journal of Earth Sciences*, *52*, 112–122.
- Gardner, J. K., & Knopoff, L. (1974). Is the sequence of earthquakes in Southern California, with aftershocks removed, Poissonian? *Bulletin of the Seismological Society of America*, *64*(5), 1363–1367.
- Ghofrani, H., & Atkinson, G. M. (2016). A preliminary statistical model for hydraulic fracture-induced seismicity in the Western Canada Sedimentary Basin. *Geophysical Research Letters*, *43*, 164–172.
- Gutenberg, R., & Richter, C. F. (1944). Frequency of earthquakes in California. *Bulletin of the Seismological Society of America*, *34*, 185–188.
- Halchuk, S., Allen, T. I., Adams, J., & Rogers G. C. (2014). Fifth generation seismic hazard model input files as proposed to produce values for the 2015 National Building Code of Canada. *Geol. Surv. Canada*, Open-File Report. 7576, <https://doi.org/10.4095/293907>.

- Hornbach, M. J., DeShon, H. R., Ellsworth, W. L., Stump, B. W., Hayward, C., Frohlich, C., et al. (2015). Causal factors for seismicity near Azle, Texas. *Nature Communications*, 6, 6728.
- Keranen, K. M., Weingarten, M., Abers, G. A., Bekins, B. A., & Ge, S. (2014). Sharp increase in central Oklahoma seismicity since 2008 induced by massive wastewater injection. *Science*, 345, 448–451.
- Langenbruch, C., Dinske, C., & Shapiro, S. A. (2011). Inter event times of fluid induced earthquakes suggest their Poisson nature. *Geophysical Research Letters*, 38, L21302.
- Langenbruch, C., & Zoback, M. D. (2016). How will induced seismicity in Oklahoma respond to decreased saltwater injection rates? *Science Advance*, 2, e1601542. <https://doi.org/10.1126/sciadv.1601542>.
- Marzocchi, W., & Taroni, M. (2014). Some thoughts on declustering in probabilistic seismic hazard analysis. *Bulletin of the Seismological Society of America*, 104, 1838–1845.
- Musson, R. M. W. (2000). The use of Monte Carlo simulations for seismic hazard assessment in UK. *Annali di Geofisica*, 43, 1–9.
- Ogata, Y., & Zhuang, H. C. (2006). Space-time ETAS models and an improved extension. *Tectonophysics*, 413, 13–23.
- Petersen, M. D., Mueller, C. S., Moschetti, M. P., Hoover, S. M., Llenos, A. L., Ellsworth, W. L., et al. (2016). 2016 One-year seismic hazard forecast for the Central and Eastern United States from induced and natural earthquakes. U.S: Geological Survey, Open-File Report. <https://doi.org/10.3133/ofr20161035>.
- Petersen, M. D., Mueller, C. S., Moschetti, M. P., Hoover, S. M., Shumway, A. M., McNamara, D. E., et al. (2017). 2017 one-year seismic hazard forecast for the central and Eastern United States from induced and natural earthquakes. *Seismological Research Letters*, 88, 772–783.
- Reasenber, P. (1985). Second-order moment of central California seismicity, 1969–82. *Journal of Geophysical Research: Solid Earth*, 90, 5479–5495.
- Roche, V., Grob, M., Eyre, T., & van der Baan, M. (2015). Statistical characteristics of Microseismic events and in-situ stress in the Horn Basin. *Geoconvention 2015: New Horizons*.
- Scholz, C. H. (1982). Scaling laws for large earthquakes: Consequences for physical models. *Bulletin of the Seismological Society of America*, 72, 1–14.
- Shapiro, S., Dinske, C., & Langenbruch, C. (2010). Seismogenic index and magnitude probability of earthquakes induced during reservoir fluid stimulations. *The Leading Edge*, 29, 936–940.
- Shi, Y., & Bolt, B. A. (1982). The standard error of the magnitude-frequency *b*-value. *Bulletin of the Seismological Society of America*, 72, 1677–1687.
- Sigman, K. (2013). Non-stationary Poisson processes and Compound (batch) Poisson processes. *Columbia Edu.*, 2018, <http://www.columbia.edu/~ks20/4404-Sigman/4404-Notes-NSP.pdf>. Last accessed Feb.
- van Stiphout, T., Zhuang, J., & Marsan, D. (2012). Seismicity declustering, Community online resource for statistical seismology. <https://doi.org/10.5078/corssa52382934>. Available at: <http://www.corssa.org>. Last accessed Feb 2018.
- van der Baan, M., & Calixto, F. J. (2017). Human-induced seismicity and large-scale hydrocarbon production in the USA and Canada. *Geochemistry, Geophysics, Geosystems*, 18, 2467–2485.
- Wiemer, S., & Wyss, M. (1997). Mapping the frequency-magnitude distribution in asperities: An improved technique to calculate recurrence times? *Journal of Geophysical Research*, 102, 15115–15128.
- Wyss, M. (1979). Estimating maximum expectable magnitude of earthquakes from fault dimensions. *Geology*, 7, 336–340.
- Zhuang, J., D. Harte, M.J. Werner, S. Hainzl, & S. Zhou. (2012). Basic models of seismicity: temporal models, Community Online Resource for Statistical Seismicity Analysis, <https://doi.org/10.5078/corssa-79905851>. <http://www.corssa.org>. Last accessed Feb 2018.
- Zhuang, J. & Touati S. (2015). Stochastic simulation of earthquake catalogs, Community Online Resource for Statistical Seismicity Analysis. <https://doi.org/10.5078/corssa-43806322>. <http://www.corssa.org>. Last accessed Feb 2018.

(Received October 2, 2018, revised January 19, 2019, accepted January 24, 2019, Published online February 5, 2019)