Probabilistic Seismic Hazard Analysis for Induced Seismicity

by

Mauricio Reyes Canales

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Geophysics

Department of Physics
University of Alberta

Abstract

In recent times, there has been increased seismicity in geologically stable basins in North America, thought to be associated with hydraulic fracturing and/or waste water disposal activities. Induced events could generate a seismic hazard higher than natural seismicity, particularly in areas with small-to-moderate natural background seismicity, leading to increasing concerns for operators, regulators, and the public in general. Therefore, it has been necessary to quantify the seismic hazard related to anthropogenic activities. In this thesis, we develop a Probabilistic Seismic Hazard Analysis to estimate the seismic hazard related to induced seismicity. Some of the main challenges to adapt Probabilistic Seismic Hazard Analysis to induced seismicity include the non-stationary behavior of the induced events, the prediction of future recurrence rates, and determine when seismicity declustering is appropriated for the hazard analysis.

First, we developed a methodology to compute synthetic earthquake catalogs for non-stationary seismicity using Monte Carlo simulations. We found that the Poisson model remains relevant for analyzing non-stationary induced seismicity. We define non-stationary Gutenberg-Richter parameters to describe time-dependent seismicity rates and to assess the hazard for this type of seismicity. Then, we apply two methodologies to predict the Gutenberg-Richter parameters related to injection-induced seismicity, in particular, temporal changes in the a-values. Our results show that these short-term predictions are able to describe the changes in the overall seismicity patterns within a reasonable level. However, they make inaccurate predictions for specific magnitude ranges, resulting in an under- or overestimation of the hazard, among other things, due

to the assumption of unchanged b-values.

We also identify the cases when seismicity declustering is recommended. If mainshocks and aftershocks have considerably different b-values, declustering leads to improved hazard assessments, since it allows for better estimations of the magnitude-frequency distribution of the largest events. On the other hand, if mainshocks and aftershocks have similar b-values, declustering is not recommended since it eliminates large magnitude events that significantly contribute to the seismic hazard. We also show that assuming Poissonian distributions in hazard predictions does not lead to inaccurate long-term hazard predictions, even if time-varying aftershock sequences are present in the catalog.

Finally, I describe a methodology to build seismicity scenarios in areas prone to induced earthquakes. This type of scenario building, based on projected operations in susceptible areas to induced seismicity, can give us insights into future seismicity patterns. Building seismicity scenarios is one of the first steps to fully forecast seismic hazard for induced seismicity. I also recommend the assessment of retrospective annual seismic hazard analysis in areas prone to induced earthquakes, like the province of Alberta. These annual seismic hazard assessments can provide useful insights into the temporal evolution of the seismic hazard related to anthropogenic activity.

Preface

This dissertation is submitted for the degree of Doctor of Philosophy in Geophysics at the University of Alberta. The research described herein is original, and neither this nor any substantially similar dissertation was or is being submitted for any other degree or other qualification at any other university.

A version of **chapter 3** has been published as: Reyes Canales, M., and van der Baan, M., 2019, Including Non-stationary Magnitude-Frequency distributions in Probabilistic Seismic Hazard Analysis: Pure and Applied Geophysics, 176, 2299-2319. I was responsible for code development, data analysis, figure preparation and manuscript writing.

A manuscript including parts of **chapter 4** of this thesis is in preparation for submission to *Pure and Applied Geophysics*. I am responsible for code development, data analysis, figure preparation, and manuscript writing.

A version of **chapter 5** has been submitted to *Journal of Geophysical Research*. I was responsible for code development, data analysis, figure preparation and manuscript writing.

To God almighty,

thank you for all your blessings and lessons.

To my Mom and all my wonderful family,

thank you for your love and support.

Acknowledgments

I would like to express my gratitude to my supervisor, Dr. Mirko van der Baan. His continuous support, patience, motivation, and knowledge were fundamental to make this project possible. I have been inspired by his vast experience and I have greatly benefited by his insightful comments and suggestions. I highly appreciate him for investing his time and efforts in my PhD.

I would like to acknowledge the members of my PhD final examination committee, Dr. Jack Baker, Dr. Claire Currie, Dr. Moritz Heimpel, and Dr. Bruce Sutherland, and for their feedback which helped in improving the overall quality of my thesis. I also would like to thank Dr. Clayton Deutsch for discussion on the Monte Carlo simulation method, and Dr. Robert Shcherbakov for discussion on ETAS simulation method.

I would like to thank all my past and present colleagues at the Microseismic Industry Consortium for their support. I would also like to thank all my friends and colleagues from Geophysics and Physics at the University of Alberta.

Finally, I am indebted to my family, my mom, my aunts, my uncles and cousins for their unconditional love and support, and the numerous sacrifices they made to help me succeed in my academic career.

Contents

List of Tables

List of Figures

List of Abbreviations

CDF	Cumulative Distribution Function
BC	Province of British Columbia
ETAS	Epistemic-Type Aftershock Sequence
GMPEs	Ground-Motion Prediction Equations
GR	Gutenberg-Richter
LSM	Least-Square Method
MLM	Maximum Likelihood Method
MP	Mainshock Poissonian
PDF	Probability Distribution Function
PMF	Probability Mass Function
PGA	Peak Ground Acceleration
PSHA	Probabilistic Seismic Hazard Analysis
SA	Spectral acceleration
USA	United States of America
USGS	United State Geological Survey

Chapter 1

Introduction

1.1 Background

Induced seismicity refers to earthquakes caused by human activities. The concept behind induced seismicity is that certain industrial activities change the state of stress of the earth's crust, resulting in the potential failing of critically stressed faults (Ellsworth, 2013). Some of the industrial activities associated with induced seismicity include: impoundment of reservoirs, surface and underground mining, withdrawal of fluids and gas from the subsurface, and injection of fluids into underground formations (McGarr et al., 2002; Ellsworth, 2013). Seismicity associated with petroleum production has been recorded since the 1920s, with reservoir impoundment since the 1930s, and with high-pressure liquid injection since the 1960s (McGarr et al., 2002). However, the number of induced seismicity cases has increased over time, particularly with the development of unconventional hydrocarbon shales and the extraction of geothermal heat (Ellsworth, 2013).

Extensive use of hydraulic fracturing and waste water injection wells have been required to develop shale oil and gas plays. Recent studies (Atkinson et al., 2015; Ellsworth, 2013; 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 waste water disposal wells. In some cases like Oklahoma (USA), waste water disposal activities have

been linked to events as large as magnitude $M = 5.6$ (2016 Pawnee earthquake, Langenbruch and Zoback (2016)) resulting in injury and damage of buildings (Hincks et al., 2018).

In this light, induced seismicity associated with shale oil and gas production has become an increasing concern for operators, regulators, and the public in general. Due to the increase in human-induced seismicity associated with shale oil and gas production, it is necessary to quantify the seismic hazard related to these activities. The seismic hazard caused by induced seismicity could be higher than the natural seismic hazard, especially in areas with small-to-moderate natural background seismicity (Atkinson et al., 2015). This is the case of many places in North America, including the Western Sedimentary Basin of Canada, where hydraulic fracturing and waste water disposal activities may have increased seismic hazard, with implications for building codes, safety and public perception.

Government and industry have used Probabilistic Seismic Hazard Analysis (PSHA) for assessing hazard related to natural seismicity, and it can be used to delineate the hazard caused by human-induced events. PSHA quantifies the possible ground motion at one location, in a period of time, caused by earthquake shaking (Cornell, 1968; Baker, 2008). Some recent studies have delineated methods to assess the hazard related to induced events, based on modified versions of the 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, and Ground Motion Prediction Equations (GMPEs) suited for induced seismicity. Petersen et al. (2016), Petersen et al. (2017) and Petersen et al. (2018) 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. These studies assume stationarity as a first approach, in accordance with the traditional PSHA for natural seimsicity.

However, these preliminary studies still require to solve some of the main challenges in the implementation of PSHA for induced seismicity. In this thesis, we develop a PSHA to estimate the seismic hazard related to induced seismicity, proposing novel methodologies to solve some of the main challenges of PSHA

for induced seismicity, including the non-stationary behavior of the induced events, the prediction of future recurrence rates, and determine when seismicity declustering is appropriated for the hazard analysis.

1.2 Motivation and contribution

The primary purpose of this thesis is to develop a model able to quantify the seismic hazard related to induced events. PSHA was originally designed for natural seismicity, and there are multiple challenges in the implementation of this method for induced events. Through this thesis, we describe the changes required to adapt PSHA for induced seismicity. Simultaneously, we investigate some of the driving science questions related to induced seismicity and its impact on the hazard analysis:

1. If we have non-stationary earthquake sequences like induced seismicity, can we still develop a PSHA model for the hazard estimations?

2. Is assuming a Poisson distribution still valid for the seismic hazard analysis, in particular, if we are dealing with non-stationary seismicity?

3. If we know beforehand specific geological and operational parameters in a seismogenic area, can we predict the future rate of earthquakes related to human-induced activity? If so, can we incorporate this information in PSHA?

4. If we decluster an earthquake catalog, are we getting accurate estimations of the seismic hazard, particularly an appropriate estimate of the recurrence rate for large magnitude events?

Understanding these questions is fundamental to characterize induced events and properly adapt induced seismicity into PSHA. To answer these questions, we delineated novel methodologies, and conducted examples which ultimately contribute to the development of seismic hazard analysis for induced seismicity. The most relevant innovations and contributions of this thesis are enumerated as follows:

1. We derive analytic expressions for non-stationary occurrence rates, which are verified using Monte Carlo simulations. The use of time-dependant Gutenberg-Richter parameters is a particular relevant innovation in this thesis.

2. We develop a PSHA methodology that allows non-stationary seismicity like induced seismicity. The use of non-stationary seismicity rates in the PSHA methodology is a novel approach delineated in this thesis.

3. We show that, in terms of seismic hazard analysis, declustering is only recommended in catalogs with a large number of earthquakes or in catalogs where the b values of the mainshocks are significantly different than the b-values of the complete catalog.

4. We show that the Poisson statistics and simulations lead to a representative assessment of long-term seismic hazard, even if aftershocks have a non-Poissonian distribution. The use of non-stationary Poisson distributions in PSHA is one of the innovations in this thesis.

5. We incorporate physics-based models into PSHA to forecast the future rate of earthquakes for induced seismicity.

1.3 Thesis overview

Chapter 2 describes the main steps of PSHA using the Monte Carlo simulation method, as well as identifying the main challenges of implementing PSHA for induced seismicity.

Chapter 3 shows a methodology to evaluate statistically the hazard related to non-stationary seismic sources like induced seismicity. To delineate non-stationary seismicity rates, we define time-dependent Gutenberg-Richter (GR) parameters. We show two examples: (1) a synthetic case with two seismic sources; and (2) a recent case of induced seismicity, the Horn River Basin, Northeast British Columbia, Canada.

Chapter 4 describes two methodologies designed to predict the earthquake recurrence rates related to induced seismicity: the Seismogenic Index and the

Hydromechanical Nucleation model. We apply both methods to the Horn River Basin induced seismicity case. We compare predictions of both models versus observed seismicity. To make a complete seismic hazard analysis, we include the generation of synthetic ground motion catalogs and seismic hazard maps.

Chapter 5 investigates if separating mainshocks and aftershocks using a temporal declustering approach, leads to more accurate hazard assessments, particularly for large-magnitude events. We apply a simulation method based on the Epistemic-Type Aftershock Sequence (ETAS) to generate mainshock (declustered) and complete (non-declustered) synthetic earthquake catalogs. We show two examples: two ynthetic examples (short and long-term seismicity) and a case of induced seismicity, Oklahoma, USA.

Chapter 6 presents an overview of the reported induced seismicity cases in Alberta. Then, I propose a methodology to build seismicity scenarios in areas susceptible to induced earthquakes. This type of scenario building, based on projected operations in susceptible areas to induced seismicity, can give us insights into future seismicity patterns. Building seismicity scenarios is one of the first steps to fully forecast seismic hazard for induced seismicity.

Chapter 7 presents the conclusions of the thesis and suggested possible directions for future research.

Chapter 2

Probabilistic Seismic Hazard Analysis and challenges for induced seismicity

2.1 Introduction

Probabilistic seismic hazard analysis (PSHA) has been largely used for assessing hazard related to natural seismicity, and it can be used to delineate the hazard caused by human-induced events. PSHA quantifies the possible ground motion at one location, in a period of time, caused by earthquake shaking (Cornell, 1968; Baker, 2008). Two of the main PSHA products are: (1) the seismic hazard curve, that is, the annual rate of exceedance vs. ground motion intensity; and (2) the seismic hazard maps, which describe spatially the likelihood to reach or exceed a ground motion in a period of time.

PSHA products are used by governments and industry in applications for life and property safety, such as developing building code requirements, deciding the security criteria for critical facilities like dams, hydroelectric plants, nuclear plants, and determining earthquake insurance rates (Baker, 2008; Mulargia et al., 2017). Seismic hazard maps are routinely updated to account for more accurate earthquake recurrence parameters and a more refined understanding of the seismic sources. Figure 2.1 shows the 2015 National Building Code of

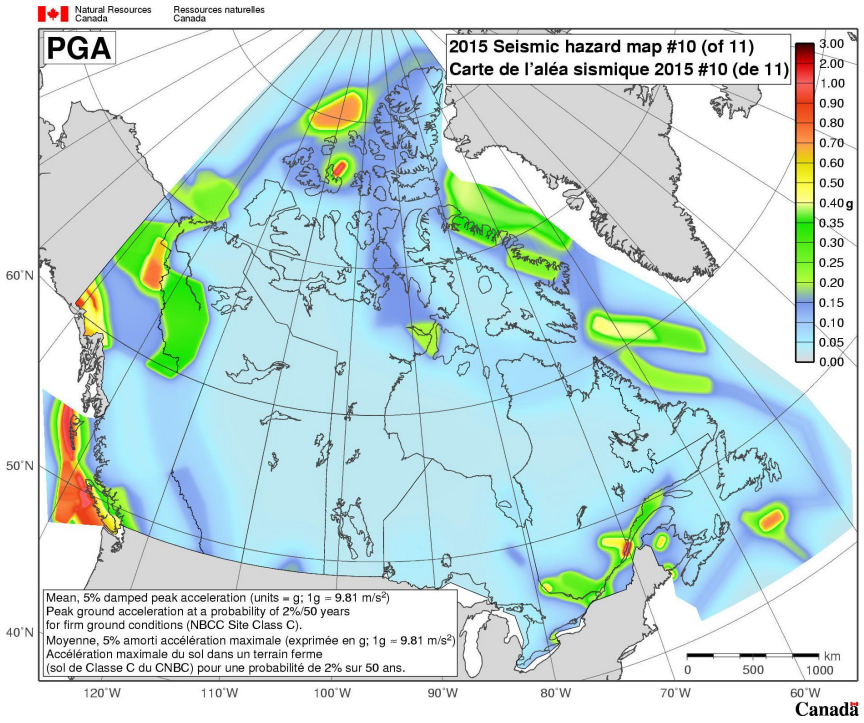

Figure 2.1: 2015 National Building Code of Canada seismic hazard maps. This map shows the 2 % likelihood to reach or exceed a given PGA (g) in the next 50 years. In this map, g represents the gravitational acceleration. From Natural Resources Canada (2015).

Canada seismic hazard map (Natural Resources Canada, 2015). This map shows 2 % likelihood to reach or exceed a given ground motion (Peak Ground Acceleration, PGA) in the next 50 years. As expected, the areas with higher seismicity correspond to the areas with a higher likelihood of ground motions, for instance, the pacific coast of Canada.

The PSHA principles were defined first by Cornell (1968), with subsequent modifications and additions over the years, including the use of Monte Carlo simulations for the generation of synthetic earthquake catalogs (Musson, 2000). PSHA using the Monte Carlo method has several advantages over conventional PSHA based on analytic 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). In this chapter, we describe the main steps of PSHA using the Monte Carlo simulation method, as well as identifying the main challenges of implementing

PSHA for induced seismicity. Some of these challenges include: (1) The non-stationary behavior of the induced seismicity; (2) The prediction of future recurrence rates related to induced seismic sources; (3) the debate of whether or not declustering is appropriate for seismic hazard analysis.

2.2 Probabilistic Seismic Hazard Analysis using the Monte Carlo simulation method

PSHA using the Monte Carlo simulation approach generates synthetic earthquake catalogs based on defined recurrence parameters (Musson, 2000). This method follows a very similar procedure to the conventional PSHA, with the difference that the hazard statistics are extracted from synthetic earthquake catalogs rather than analytical calculations. The Monte Carlo approach assumes the occurrence of earthquake events as a stationary process following a Poisson distribution in time (Assatourians and Atkinson, 2013). However, this method is flexible enough to introduce a non-Poissonian and/or non-stationary behavior (Musson, 2000).

PSHA using the Monte Carlo simulation method consists of the following steps: (1) Define a seismic source area; (2) Obtain the Gutenberg-Richter (GR) parameters (a-and b-values) from observed earthquake catalogs; (3) Generation of synthetic earthquake catalogs; (4) Generation of ground motion catalogs by using Ground Motion Prediction Equations (GMPEs); (5) Generation of seismic hazard curves and maps. Figure 2.2 shows a sketch of PSHA using the Monte Carlo simulation method.

1. **Define earthquake source areas**: The seismic source areas are defined as the regions that contain the seismic events (Baker, 2008). One of the most important steps consists of defining faults, areas or volumes that could generate seismic events. The potential sources are identified based on recorded seismic events that constitute the historical seismicity. When individual seismogenic faults are not possible to identify, the earthquake sources are described as areas where earthquakes could occur anywhere

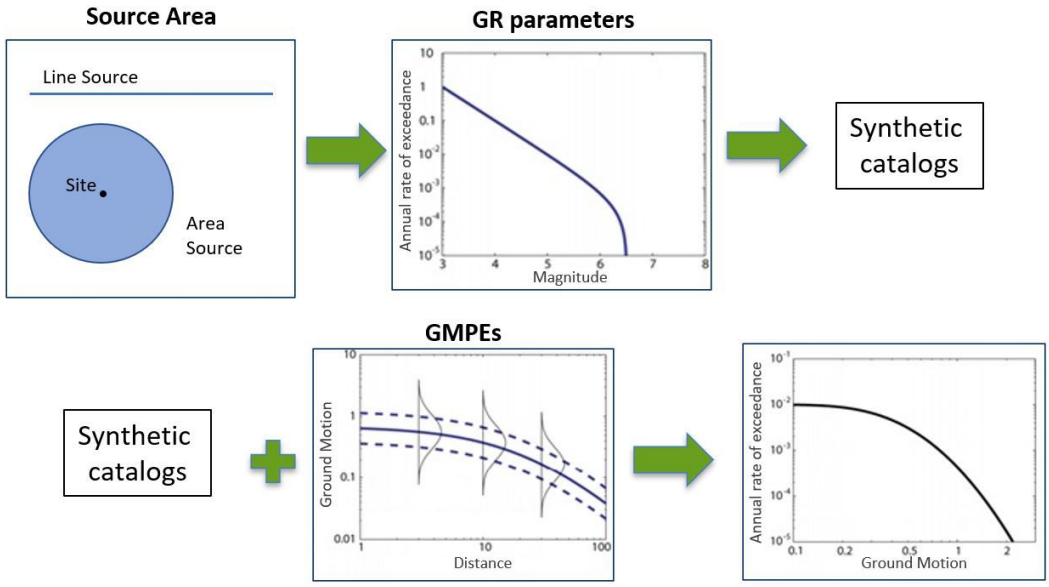

Figure 2.2: Main steps of PSHA using the Monte Carlo simulation method: (1) Define a seismic source area, (2) from recorded earthquake catalogs, obtain GR parameters (a-and b-values), (3) simulate synthetic earthquake catalogs, (4) use GMPEs in order to get Ground Motion catalogs, (5) generate seismic hazard curve. Based on the PSHA steps from Baker (2008)

(Baker, 2008). Conventional PSHA assumes an unchanged seismic source area over time, delimited by using historical earthquake catalogs.

2. **Define the seismic source parameters**: The Gutenberg-Richter recurrence law (1944) describes the earthquake size distribution for a particular region. This law states the relationship between the magnitude and the total number of earthquakes in a given region, following the relation:

$$\log(\lambda_m) = a - bm, \tag{2.1}$$

where λ_m represents the rate of earthquakes with magnitude greater than m, in other words, the total number of events N greater than m divided by the time duration of the catalog. The a-value specifies the overall rate of earthquakes in a region, and the b-value indicates the ratio of small and large magnitude events. The a- and b-values are commonly obtained by plotting, on a semi-log scale, the rate of earthquakes vs. magnitude from observed catalogs, and then fitting a curve with slope b and Y-axis inter-

cept a, to the observed data points. Two of the most common methods used to obtain the GR parameters, a- and b-values, from observed catalogs are the least square method (LSM) and the Maximum Likelihood Method (MLM, Aki (1965)).

Seismicity declustering is a standard method applied to the earthquake catalogs before the estimation of the GR parameters described above. This method consists of generating a mainshock catalog by removing foreshocks and aftershocks from the observed catalog (Gardner and Knopoff, 1974). Therefore, the focus relies on the recurrence parameters of the mainshocks, which, in turn, are the large magnitude earthquakes. As expected, the large magnitude events are related to higher ground motions and, therefore, higher potential damage. More details about seismicity declustering can be found in chapter 5.

The a- and b-value together with the Maximum Magnitude (M_{max}) and the Minimum Magnitude (M_{min}) characterize the seismic source parameters. M_{max} gives the upper bound of the linear recurrence law, based on observations (Kramer, 1996), and M_{min} is the lowest magnitude event taken in the catalog. Often, the M_{max} boundary value is not well defined; for that reason, some models select a range of M_{max} values and then assign different weights to each value, following the logic-tree format (Kramer, 1996). In other words, several hazard curves are calculated using a range of M_{max} values and then these results are combined depending on their weight and likelihood. This logic-tree format can be also applied to pairs of a- and b-values to account for the uncertainty in the GR parameters.

Frequently, the Minimum Magnitude M_{min} is equivalent to the Magnitude of Completeness (M_c) of one area. The Magnitude of completeness is the minimum magnitude above which all earthquakes within a certain region are reliably recorded (Wiemer and Wyss, 1997). Therefore, the Minimum Magnitude M_{min} does not necessarily correspond with the lowest magnitude recorded in the earthquake catalog. Estimations of a- and b-values are usually performed only considering events above the Magnitude of Completeness M_c value.

3. **Generation of synthetic earthquake catalogs**: The generation of synthetic earthquake catalogs using the Monte Carlo simulation method can be summarized in two main steps: (1) Simulation of a temporal point process to obtain arrival times for a single realization; (2) Simulation of earthquake magnitudes m. To generate arrival times for a single realization, a Monte Carlo sampling is performed to the inverse cumulative distribution function (CDF) of the Poisson distribution with stationary rate λ. In this step, the earthquake recurrence λ is defined by the GR parameters, eq.2.1, and the average number of samples in a realization is given by the recurrence rates and the period of time to simulate. To generate magnitudes for each time arrival, another Monte Carlo sampling is performed to the inverse CDF of the GR recurrence law, eq.2.1. Consequently, each pair of time arrivals and magnitudes represent a synthetic event. Finally, the location of these synthetic events is assumed to be uniformly random distributed within the seismic source area.

   Each realization represents a possible scenario of seismicity in a forthcoming period of time. By repeating multiple times, different realizations are simulated, resulting in the synthetic earthquake catalog to analyze. From these synthetic catalogs, it is possible to extract the statistics and study the seismic hazard in one area, in a period of time. One of the advantages of the Monte Carlo simulation is that, by generating a large number of realizations, it is possible to calculate the likelihood of rare but potentially damaging events. A more detailed description of the generation of synthetic earthquake catalogs using the Monte Carlo simulation method can be found in chapter 3, including the changes necessary to allow non-stationary seismicity.

4. **Generation of ground motion catalogs**: To generate ground motion catalogs, it is necessary to define first the ground motion prediction equations (GMPEs) for one area. The GMPEs predict a probability distribution of ground motion intensity (E.g. Peak ground acceleration (PGA), Spectral acceleration (SA), etc.) as a function of several variables such as magnitude, distance, and site characterization (Baker, 2008; Bourne et al., 2014). Each magnitude-distance pair is related to a predicted

ground motion intensity. The ground motion prediction equations are generally developed using statistical regression on thousands of observed ground motion intensities from several earthquakes (Baker, 2008).

Once the synthetic earthquake catalog has been generated, we calculate the distances between a selected site and the locations of the synthetic events. Given the distance and magnitude of the synthetic events, we use GMPEs to estimate the ground motions that would cause these synthetic events in the selected site. These estimated ground motions result in a catalog of ground motions. To count for the aleatoric variability in the ground motions, a random number drawn from the standard normal distribution is multiplied by the sigma value (variability of the GMPE model), and added to the log of the ground motions (Assatourians and Atkinson, 2013).

5. **Generating seismic hazard curves and maps**: From the ground motion catalogs, we can obtain statistical results for the hazard analysis. For instance, the rate of exceeding a ground motion in a period of time, $\lambda(gm \geq gm_j; t_a, t_b)$, is given by counting all the ground motions larger than gm_j, and dividing this count by the time duration of the simulations $(t_b - t_a)$ and the number of realizations of the synthetic catalog. Then, the seismic hazard curve is generated by plotting the rate of exceedance as a function of ground motions.

   The seismic hazard maps are built by calculating first the seismic hazard curves at different sites. Then, a common rate of exceeding a ground motion $\lambda(gm \geq gm_j; t_a, t_b)$ is defined for all seismic curves and the related ground motion is selected. This yields a ground motion value for each considered site, which in turn can be interpolated to generate a seismic hazard map. Assuming a Poisson process, the rate of exceeding a ground motion $\lambda(gm \geq gm_j; t_a, t_b)$ is equivalent to the probability $P[N > 0; t_a, t_b]$ to reach or exceed an event in a period of time $(t_b - t_a)$ (Baker, 2008):

$$\lambda(gm \geq gm_j; t_a, t_b) = \frac{-\ln(1 - P[N > 0; t_a, t_b])}{t_b - t_a}, \tag{2.2}$$

where $P[N > 0; t_a, t_b]$ is the probability to reach at least one event ($N > 0$, N is the number of events) in a period of time ($t_b - t_a$). For instance, the 2% probability of exceedance in 50 years is equivalent to the rate of exceedance $\lambda(gm \geq gm_j; t_a, t_b) = 0.000404$, or 1 event with a recurrence of $2,475$ years. More details about the generation of synthetic ground motion catalogs and seismic hazard maps can be found in chapters 4 and 6.

2.3 Challenges in PSHA for induced seismicity

There are numerous challenges in the implementation of PSHA for induced seismicity. First, traditional PSHA (Cornell, 1968) has been developed assuming a temporal stationary Poisson process, which fits well for natural mainshock seismicity. However, induced seismicity is conditioned in time and space by the well operations, therefore, assuming temporal stationarity and a Poisson process will have significant effects on the hazard calculations (Atkinson et al., 2015). Furthermore, the common time durations and likelihoods for evaluating the probability of exceedance in the natural hazard maps (e.g. 2% probability of exceedance in 50 years) are no longer representative, because the induced seismicity is unlikely to be temporally stationary in that time frame, and relevant human processes may last less than 50 years (e.g. hydraulic fracturing treatments and even waste-water injection). To deal with these issues, we develop a methodology to evaluate statistically the hazard related to non-stationary seismic sources like induced seismicity. We use time-dependent GR parameters, which lead to a time-varying rate of earthquakes. We also derive analytic expressions for occurrence rates which are verified using Monte Carlo simulations. Further details can be found in chapter 3 of this thesis.

Second, traditional seismic hazard analysis assumes future rates based on GR parameters from historical seismicity. On the other hand, prediction of GR parameters for induced seismicity is still a major challenge due to the lack of recorded events and its non-stationary behaviour. In chapter 4 we describe two physics-based models in order to address this issue: The Seismogenic Index (Shapiro et al., 2010; Langenbruch and Zoback, 2016) and the Hydromechani-

cal Nucleation model (Norbeck and Rubinstein, 2018). The Seismogenic Index (Shapiro et al., 2010) modifies the classical GR relationship to allow time dependent cumulative a-values. This cumulative a-value increases depending on the total volume injected and the Seismogenic Index, which is related to the volume concentration of preexisting faults and the state of stress in one particular area (Shapiro et al., 2010). The Hydromechanical Nucleation model (Norbeck and Rubinstein, 2018) relies on the empirical seismicity rate model from Dieterich (1994) and Segall and Lu (2015), which relates the changes in the Coulomb stress with changes in the seismicity rates. Also, Segall and Lu (2015) assume that the Coulomb stress change is proportional to the pressurization rate instead of the total injected volume and thus pore pressure change.

Third, some studies have discouraged the application of seismicity declustering in the seismic hazard evaluation, as it might reduce the predicted hazard by excluding a large number of small-to-medium magnitude-size earthquakes (Atkinson et al., 2015; Langenbruch and Zoback, 2016). Seismicity declustering classifies the events from an earthquake catalog into foreshocks, mainshocks and aftershocks (Van Stiphout et al., 2012). On the one hand, the background or mainshock events are a consequence of tectonic loading, and are considered temporally independent. On the other hand, aftershocks, foreshocks or triggered earthquakes are temporally dependent, they are consequence of a parent event, either a background or another triggered event (Van Stiphout et al., 2012).

Time-dependency might not be the only difference between mainshock and aftershock sequences. Some studies suggest different b-values between mainshock (declustered) and complete (non-declustered) earthquake catalogs (Utsu, 1966; Suyehiro and Sekiya, 1972). In contrast, other authors suggest that the observed difference is not statistically significant (Knopoff et al., 1982) or it is simply an statistical artifact (Lombardi, 2003). If mainshocks and aftershocks have different b-values, then this is likely to impact the hazard assessment, because generally a single b-value is assumed within a source region. The b-value is used to extrapolate occurrence rates to earthquake magnitudes that may not be well represented within the observed catalog. As a consequence, assuming averaged b-values may lead to biased hazard assessment. In chapter 5,

we use synthetic earthquake catalogs based on the Epistemic-Type Aftershock Sequence (ETAS) methodology (Ogata, 1988, 1998; Ogata and Zhuang, 2006) to determine the implications of seismicity declustering in the hazard analysis. We also study a case when it is appropriate to decluster an earthquake catalog.

Other challenges have to be properly addressed, including the prediction of the maximum magnitude for induced events, the development of GMPEs for induced seismicity, or even more accurate physics-based models for the estimation of earthquake occurrence. In chapter 7, we provide a further description of these pending issues, as well as future steps to generate a fully adapted seismic hazard analysis for induced seismicity.

Chapter 3

Including Non-stationary Magnitude-frequency distributions in Probabilistic Seismic Hazard Analysis [1]

3.1 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

[1] A version of this chapter has been published as: Reyes Canales, M., and van der Baan, M., 2019, Including Non-stationary Magnitude-Frequency distributions in Probabilistic Seismic Hazard Analysis: Pure and Applied Geophysics, 176, 2299-2319.

et al. (2016) and Petersen et al. (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 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.

3.2 Theory

3.2.1 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):

$$\log(N) = a - bm, \tag{3.1}$$

where N is the number of earthquakes with a magnitude greater than m. Notice that log refers to logarithm base 10. 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:

$$N_0 = 10^a. \tag{3.2}$$

Eq. 3.1 can be used to compute the discrete cumulative distribution function (CDF) as Baker (2008):

$$F_M(m) = \frac{1 - 10^{-b(m-M_{min})}}{1 - 10^{-b(M_{max}-M_{min})}}, \tag{3.3}$$

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):

$$P(m_j \leq m < m_{j+1}) = F_M(m_{j+1}) - F_M(m_j), \tag{3.4}$$

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. 3.4, 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:

$$\lambda(m_j \leq m < m_{j+1}) = P(m_j \leq m < m_{j+1})N(M_{min} \leq m \leq M_{max}). \tag{3.5}$$

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.3.1, as:

$$N(M_{min} \leq m \leq M_{max}) = 10^{a-bM_{min}} - 10^{a-bM_{max}}. \tag{3.6}$$

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.3.5, is assumed to be stationary (that is, time-invariant). In this case the intercept a and slope b in the GR distribution, eq.3.1, are constant. For non-stationary sequences the rate $\lambda(m_j \leq m < m_{j+1}; t)$ is still given by eqs. 3.5 and 3.6, but now the GR parameters, $a(t)$ and $b(t)$, are understood to vary with time t, that is:

$$\lambda(m_j \leq m < m_{j+1}; t) = P(m_j \leq m < m_{j+1}; t)N(M_{min} \leq m \leq M_{max}; t), \tag{3.7}$$

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:

$$N(M_{min} \leq m \leq M_{max}; t) = 10^{a(t)-b(t)M_{min}} - 10^{a(t)-b(t)M_{max}}. \tag{3.8}$$

Similar expressions hold for eqs. 3.3 and 3.4 to compute the probability of occurrence of a magnitude bin $P(m_j \leq m < m_{j+1})$ in a time-varying frame:

$$P(m_j \leq m < m_{j+1}; t) = F_M(m_{j+1}; t) - F_M(m_j; t), \tag{3.9}$$

where $F_M(m)$ is redefined as:

$$F_M(m; t) = \frac{1 - 10^{-b(t)(m-M_{min})}}{1 - 10^{-b(t)(M_{max}-M_{min})}}. \tag{3.10}$$

The considered M_{min} and M_{max} magnitudes in eq. 3.8 and 3.10 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.

3.2.2 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):

$$P[N = n; t_a, t_b] = \frac{\lambda^n (t_b - t_a)^n e^{-\lambda(t_b - t_a)}}{n!}, \tag{3.11}$$

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 λ is the rate of occurrence of events per unit time duration. By definition $0! = 1$. Note that eq. 3.11 only depends on the time interval $\Delta t = t_b - t_a$, instead of the individual start and end times. For the stationary case, λ is equivalent to the

rate of earthquakes $\lambda(m_j \leq m < m_{j+1})$ per unit time duration for a magnitude bin, eq. 3.5. 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):

$$P[N > 0; t_a, t_b] = 1 - e^{-\lambda(t_b - t_a)}. \tag{3.12}$$

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):

$$P[N = n; t_a, t_b] = \frac{m_\lambda^n(t_a, t_b)(t_b - t_a)^n e^{-m_\lambda(t_a, t_b)(t_b - t_a)}}{n!}, \tag{3.13}$$

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):

$$m_\lambda(t_a, t_b) = \frac{\int_{t_a}^{t_b} \lambda(s) ds}{(t_b - t_a)}. \tag{3.14}$$

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.3.7, 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:

$$\lambda(M_{min} \leq m \leq M_{max}; t) = P(M_{min} \leq m \leq M_{max}; t) N(M_{min} \leq m \leq M_{max}; t), \tag{3.15}$$

where $N(M_{min} \leq m \leq M_{max}; t)$ is given by eq. 3.8 and $P(M_{min} \leq m \leq M_{max}; t)$ is a modification of eq. 3.9, resulting in:

$$P(M_{min} \leq m \leq M_{max}; t) = F_M(M_{max}; t) - F_M(M_{min}; t). \tag{3.16}$$

Replacing $\lambda(t)$ by $\lambda(M_{min} \leq m \leq M_{max}; t)$ in eq. 3.14, $m_\lambda(t_a, t_b)$ 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. 3.12, the probability of at least one event for the non-stationary Poisson distribution is defined as:

$$P[N > 0; t_a, t_b] = 1 - e^{-m_\lambda(t_a,t_b)(t_b - t_a)}, \tag{3.17}$$

The non-stationary Poisson distribution, eq. 3.13, 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. 3.14 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:

$$m_{\lambda,tot}(t_a, t_b) = \frac{\int_{t_a}^{t_b} (\lambda_{bg}(s) + \lambda_{ind}(s)) ds}{(t_b - t_a)}, \tag{3.18}$$

where the background and induced rate of earthquakes are given by $\lambda_{bg}(t)$ and $\lambda_{ind}(t)$ respectively. Eqs. 3.11 - 3.18 can be used to compute analytic expectations given known or estimated (non-)stationary GR magnitude-frequency distributions. For instance, eqs. 3.12 and 3.17 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:

$$\lambda_{exc}(m \geq m_j; t) = P(m_j \leq m \leq M_{max}; t) N(M_{min} \leq m \leq M_{max}; t), \tag{3.19}$$

where $N(M_{min} \leq m \leq M_{max}; t)$ is given by eq.3.8, 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:

$$P(m_j \leq m \leq M_{max}; t) = F_M(M_{max}; t) - F_M(m_j; t). \tag{3.20}$$

The cumulative distribution function F_M is again given by eq.3.10. 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 $\lambda_{exc,ind}$.

Eqs. 3.19 and 3.20 are extensions of eqs. 3.4, 3.5 and 3.7 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. 3.14. Note also that by inserting the mean rate of exceedance $m_{\lambda,exc}(m \geq m_j; t_a, t_b)$ into eq. 3.17, and then inverting the resulting expression, we obtain:

$$m_{\lambda,exc}(m \geq m_j; t_a, t_b) = \frac{-\ln(1 - P[N > 0; t_a, t_b])}{t_b - t_a}. \tag{3.21}$$

Eq. 3.21 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.

3.2.3 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 λ (Zhuang and Touati, 2015):

$$\tau = \frac{-ln(r)}{\lambda}, \tag{3.22}$$

where τ is a random temporal variate, λ 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 λ 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 λ is a function of time.

For the thinning process, we simulate a stationary Poisson process at rate λ^*, where $\lambda^* \geq max(\lambda(t))$. The rate λ^* 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). Any λ^* that satisfies the condition $\lambda^* \geq max(\lambda(t))$ works for the thinning process; however, we recommend a λ^* similar -but still larger- than $max(\lambda(t))$, in order to avoid excessive computational cost.

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 λ^* 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. 3.15.

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):

$$m = \frac{-ln(r)}{b(t)ln(10)} + M_{min}, \tag{3.23}$$

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. 3.1. 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 λ_{bg} and λ_{ind} in eqs. 3.5 and 3.15.

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.

3.3 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.

3.3.1 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. 3.15. 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 at one specific site. 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.

3.3.2 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 figure 3.1. 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. 3.23, 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.3.5).

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 seismic hazard analysis (Assatourians and Atkinson, 2013). Figure 3.1 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, figure 3.2). 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. 3.1.

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. 3.2, 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 $m_\lambda(m_j \leq m < m_{j+1}; t_a, t_b)$ 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.

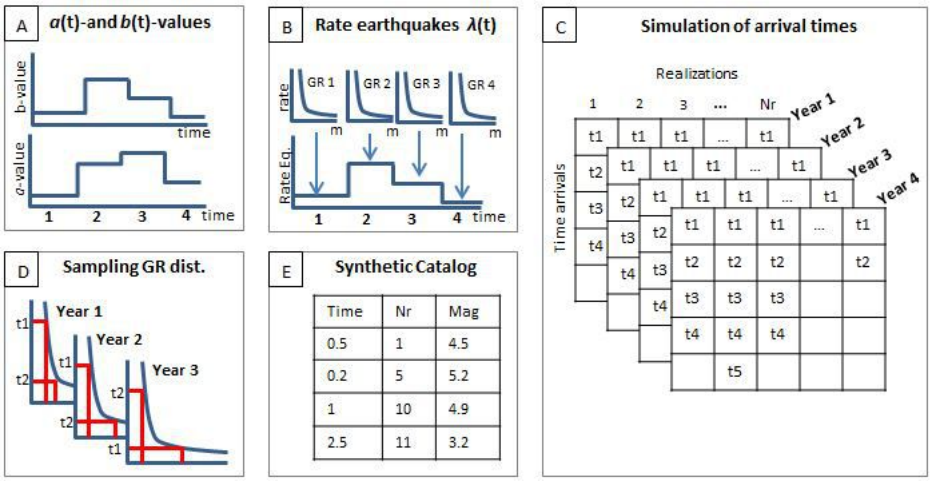

Figure 3.1: Sketch of the Monte Carlo simulation methodology for the generation of non-stationary earthquake catalogs. (A) Definition of the time-dependent $a(t)$-and $b(t)$-values, and (B) equivalent rate of earthquakes. Notice how each temporal sample might correspond to a different GR distribution. (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.

3.3.3 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 eqs. 3.13-3.15. 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.3.7, into eq. 3.14. 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. 3.8, and the probability of occurrence $P(m_j \leq m < m_{j+1}; t)$ of a magnitude bin, eq. 3.9, where the probability $P(m_j \leq m < m_{j+1}; t)$ depends on the cumulative distribution function $F_M(m; t)$, eq.3.10. 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. 3.8 and 3.10.

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.3.19, into eq. 3.14. 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. 3.8 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. 3.20, where the probability $P(m_j \leq m \leq M_{max}; t)$ depends on the cumulative distribution function $F_M(m; t)$, eq.3.10.

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).

3.4 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.3.2, and $b_{bg} = 1$. An induced seismicity sequence occurs between the years 10 through 19. Starting year 20, only natural seismicity occurs. The first time interval $t = [0, 9]$ has no induced seismicity. While the induced seismicity

is active, the GR parameters are equal to:

$$b(t) = 1 + \frac{cos(\frac{\pi(t-10)}{9})}{5}, \tag{3.24}$$

$$a(t) = 4 + \frac{cos(\frac{\pi(t-10)}{9})}{2}, \tag{3.25}$$

where t is a discrete temporal sample, starting at $t = 10$ years, when the induced seismicity appears. 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 3.2 (A) shows the temporal variation of the GR values for the background and induced sources. Figure 3.2 (B) 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.

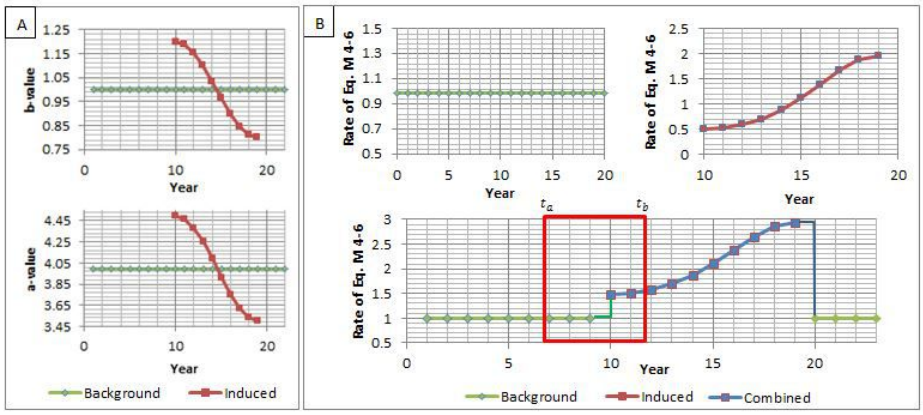

Figure 3.2: (A) Temporal evolution of the annual $a(t)$-and b-values for the background and the induced seismicity. (B) Equivalent annual rate of earthquakes $\lambda(M_{min} \leq m \leq M_{max}; t)$ in the range $m = [4.0, 6.0]$, for the background, induced and combined seismicity. The total seismicity rate is the sum of rates for background and induced seismicity, eq. 3.18. However induced seismicity is limited to the years 10 through 19. All seismicity returns to the natural pattern starting year 20.

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. Figures 3.3 (A) and (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 (Figure 3.3(A)). 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. 3.3 B). 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. 3.21. In figure 3.3 (B) 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.3.21, 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

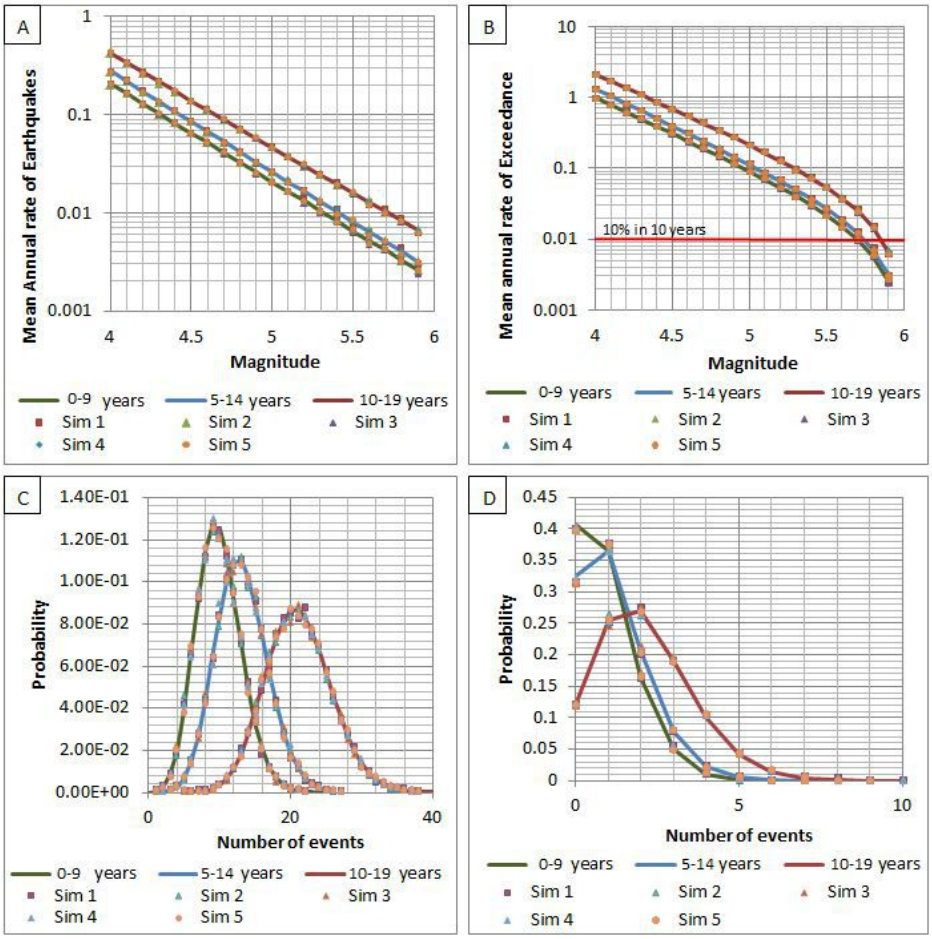

Figure 3.3: (A) Mean annual rate of earthquakes $m_\lambda(m_j \leq m < m_{j+1}; t_a, t_b)$, as a function of magnitude within a 10 year interval starting respectively at 0, 5 and 10 years. (B) Corresponding annual rate of exceedance $m_{\lambda,exc}(m \geq m_j; t_a, t_b)$, as a function of magnitude. The likelihood of occurrence of 10% in 10 years is indicated in the plot. (C) Probability as a function of number of occurrences, $P[N = n; t_a, t_b]$ of n, within the magnitude range $m = [4, 6]$ for the 10 year time interval starting at at 0, 5 and 10 years. (D) Equivalent probability as a function of number of occurrences, $P[N = n; t_a, t_b]$ of n, 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.

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 3.3 (C) 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. Again, 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. 3.13 and eq.3.18, appear as a solid line, and the Monte Carlo simulation results appear as dots. Figure 3.3 (D) 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 $t = [10, 19]$ (Figure 3.3 (D)). 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.

Figures 3.4 (A) and (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 (Figure 3.4 (A)). Given the design of this synthetic example, The curves in figure 3.3 for the time interval $t = [10, 19]$, are equivalent to the curves in figure 3.4 for the 10-years time interval. Both cases share the same interval length and period, and thus seismicity.

Figures 3.4 (C) and (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 figures 3.3 (D) and (C), the probability to have at least one occurrence increases while the time interval

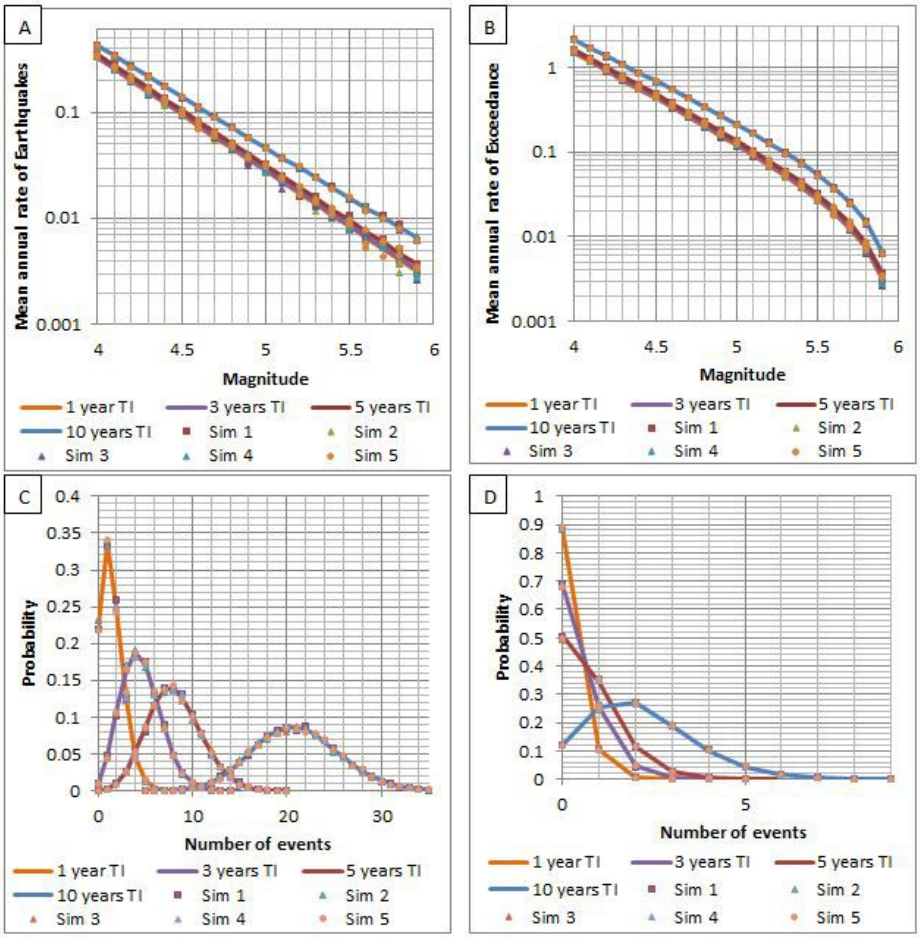

Figure 3.4: (A) Mean annual rate of earthquakes $m_\lambda(m_j \leq m < m_{j+1}; t_a, t_b)$, as a function of magnitude within 1, 3, 5 and 10 year length intervals, starting at year 10 (First year of induced seismicity). (B) Corresponding annual rate of exceedance $m_{\lambda,exc}(m \geq m_j; t_a, t_b)$, as a function of magnitude. (C) Probability as a function of number of occurrences, $P[N = n; t_a, t_b]$ of n, 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, $P[N = n; t_a, t_b]$ of n, 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.

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 (Figure 3.4 (C)).

3.5 Horn River Basin case

3.5.1 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 (Figures 3.5 (A) and (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 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 3.5 (C) 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 figure 3.5 (D), leading to increased earthquake activity. After Dec. 2011 we assume that the seismicity rates return to their natural state.

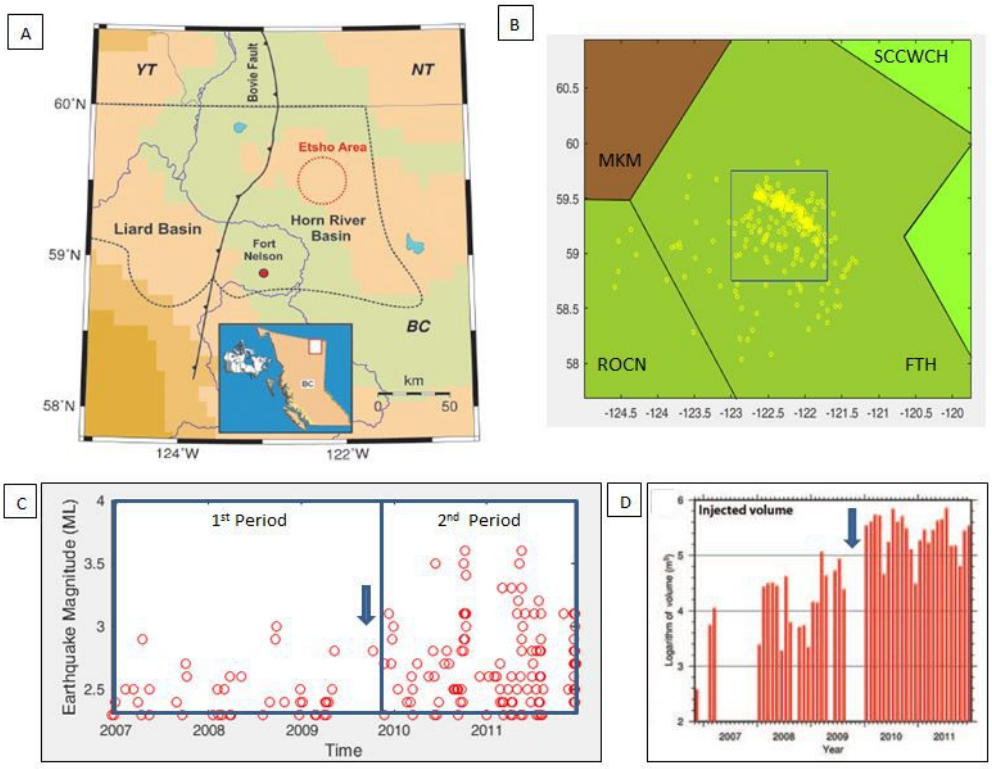

Figure 3.5: (A) Location of the Horn River Basin, Northeast British Columbia. The Etsho area (circle in red) is described as the area with much seismic activity in the basin (From Farahbod et al. (2015b)). (B) Location of the seismic events (yellow dots) using the catalog from Farahbod et al. (2015b). 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, SCCWHC: 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 subtantial increase in the injection volumes after Dec. 2009 (Vertical arrows). From Farahbod et al. (2015a).

3.5.2 Seismic parameters from the different time periods

We focus on the Etsho area within the Horn River Basin (Figures 3.5 (A) and (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

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

Table 3.1: 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.

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 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 3.1 shows the GR parameters for the Foothills area and the rescaled 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.3.2, we have the following earthquake densities: (1) 8.88x10^{-4}, (2) 3.51x10^{-3}, and (3) 2.29x10^{-4} earthquakes/km^2/year. The Etsho area is 8,262.57 km^2 (Blue square fig.3.5 (B)).

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. 3.1. 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. 3.2.

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.

Figures 3.6 (A) and (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 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.

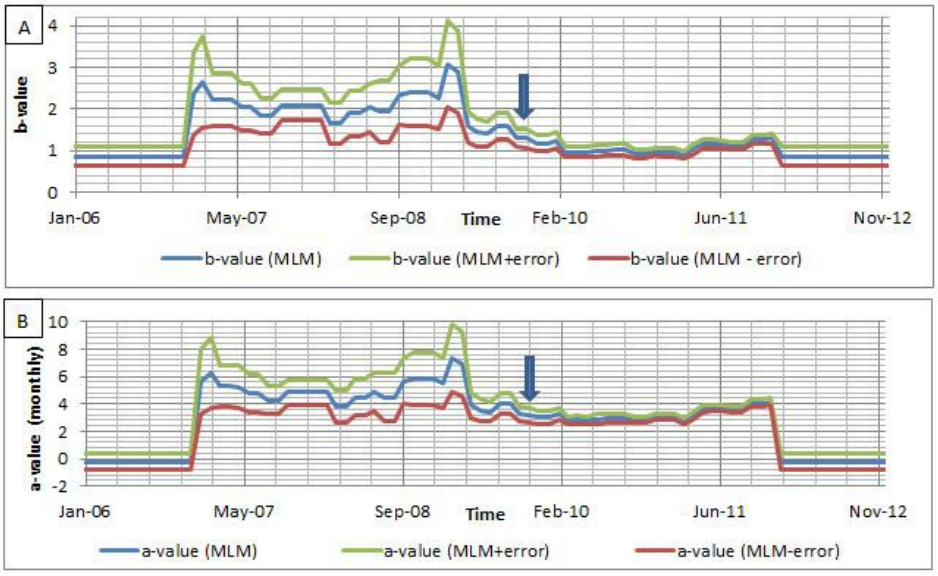

Figure 3.6: Temporal evolution of the (A) b-values, and (B) monthly a-values in the Horn River Basin. For the periods before Dec. 2006 and after Dec 2011, only background seismicity is assumed. The induced seismicity parameters were calculated from the observed catalog (Fig. 3.5 (C)). The uncertainties 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.

3.5.3 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.

Figures 3.7 (A) and (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 figure 3.6 and the expressions in the theory section. These curves are not necessarily straight lines, because they 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 figure 3.5 (C).

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. 3.5(C)). Figure 3.8 (A) shows the probability of n occurrences (Eq. 3.17), 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, figure 3.8(B) 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 - Dec. 2009, which was 21 events.

Finally, figure 3.8 (C) 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 of earthquakes $m = [2.5, 4)$ recorded between Dec. 2009 - 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 3.8 (D) 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 3.8 (E) 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 enu-

merate 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 (Figures 3.7 and 3.8 (D)).

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.

3.6 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 as-

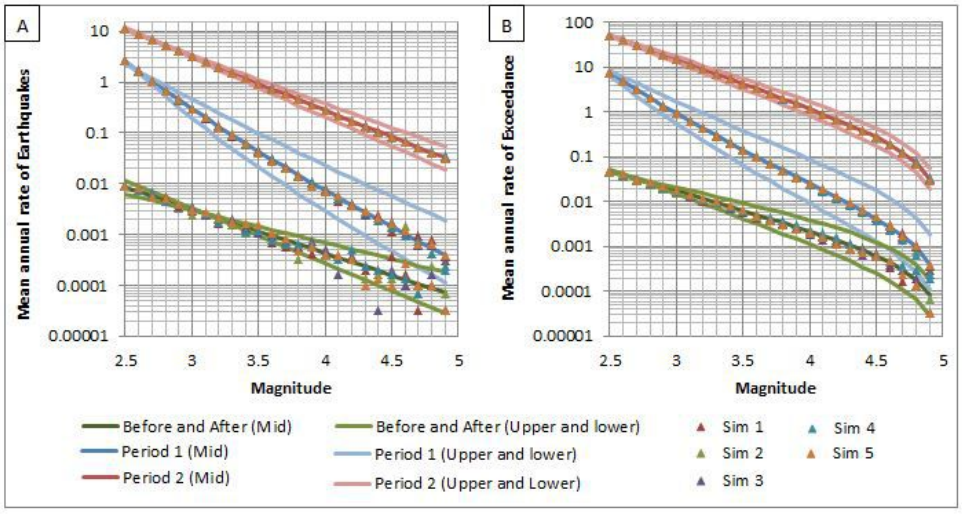

Figure 3.7: (A) Mean annual rate of earthquakes $m_\lambda(m_j \leq m < m_{j+1}; t_a, t_b)$, as a function of magnitude for the periods before and after the induced seismicity, as well as the first and second period of induced seismicity. (B) Corresponding mean annual rate of exceedance $m_{\lambda,exc}(m \geq m_j; t_a, t_b)$, 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 figure 3.6. Sim=simulation number, each using 10,000 realizations.

sessment 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; Reasenberg, 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

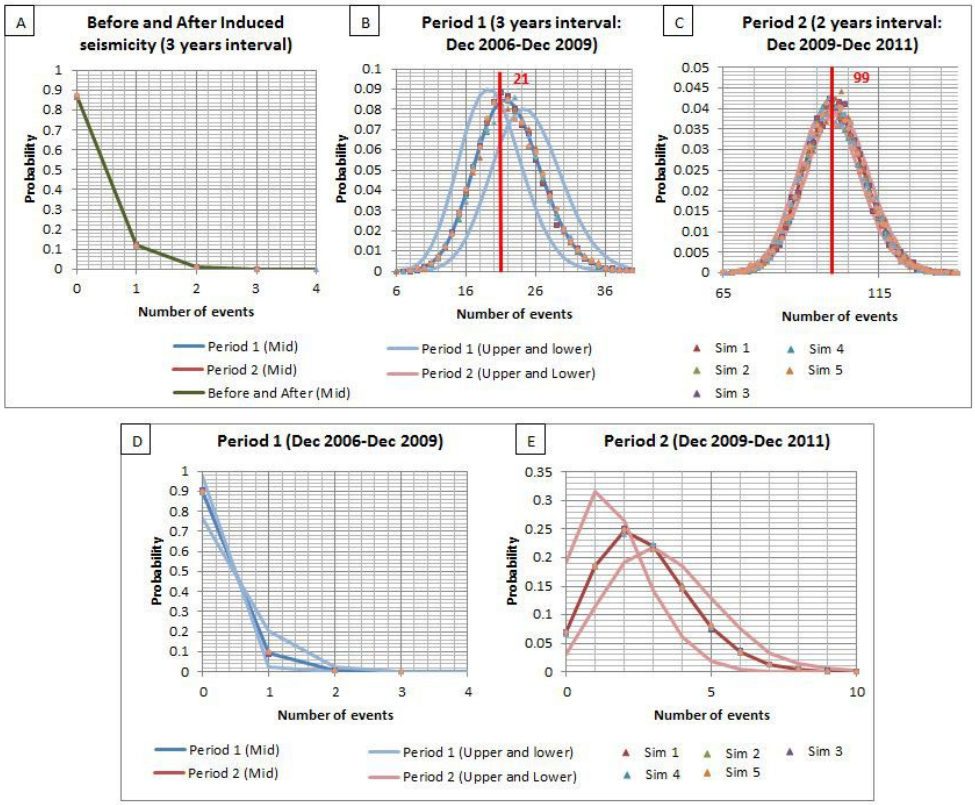

Figure 3.8: Probability as a function of the number of occurrences, $P[N = n; t_a, t_b]$ of n, within the magnitude range $m = [2.5, 4.0)$ for (A) the periods before and after induced seismicity, (B) the first period of induced seismicity and (C) the second period of induced seismicity. Once more, the 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.

have smaller a-values and thus a reduced number of predicted earthquakes, eq.3.2. On the other hand, the b-value is often enlarged, indicating a larger likelihood for the occurrence of larger magnitude events, eq.3.3. 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 figures 3.7 (A) and (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 predition equations, using a similar numerical scheme as used by Assatourians and Atkinson (2013) and Musson (2000).

3.7 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.

Chapter 4

Forecasting of earthquake rates using physics-based models for probabilistic seismic hazard analysis: a case study [1]

4.1 Introduction

Probabilistic seismic hazard analysis (PSHA) has been largely used for assessing hazards related to natural seismicity. PSHA quantifies the possible ground motion at one location, in a period of time, caused by earthquake shaking (Cornell, 1968; Baker, 2008). PSHA outputs (e.g., seismic hazard curves and seismic hazard maps) are used by governments and industry in applications for life and property safety, such as developing building code requirements, deciding the security criteria for critical facilities like dams, hydroelectric plants, nuclear plants, and determining earthquake insurance rates (Baker, 2008; Mulargia et al., 2017).

Recent studies (Atkinson et al., 2015; Ellsworth, 2013; Van der Baan and

[1] A manuscript including parts of this chapter is in preparation for submission to *Pure and Applied Geophysics*

Calixto, 2017) have shown increased seismicity in geologically stable basins in North America, thought to be associated with hydraulic fracturing treatments and/or waste water disposal wells. As a result, it has been necessary to quantify the seismic hazard related to induced activities associated with shale oil and gas production. However, there are some challenges in the implementation of PSHA for induced seismicity, including the estimation of future induced seismicity rates and its non-stationary behavior. Reyes Canales and Van der Baan (2019) developed a methodology to include non-stationary seismicity rates in the seismic hazard analysis, where the changing rates result from time-dependent Gutenberg-Richter (GR) parameters. However, estimating and forecasting GR parameters for induced seismicity is still a major challenge due to the lack of recorded events and the time dependency of rates. In contrast, this is a lesser problem for natural seismicity, since seismic hazard analysis assumes stationary GR parameters based on long-term historical catalogs.

Petersen et al. (2016) and Petersen et al. (2017) elaborate a one-year hazard forecast for the central and eastern United States, based on catalogs with recorded induced events. 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. On the other hand, some physics-based models like the Seismogenic Index (Shapiro et al., 2010) and the Hydromechanical Nucleation Approach (Norbeck and Rubinstein, 2018) have been implmented in order to forecast time-dependent GR parameters related to induced seismicity. In the context of this study, physics-based models refer to the approaches that include physical properties (geological and operational parameters) in the forecasts of human-induced events.

The Seismogenic Index (Shapiro et al., 2010) modifies the classical GR relationship to allow time-dependent cumulative a-values. This cumulative a-value increases depending on the total volume injected and the Seismogenic Index, which is related to the volume concentration of preexisting faults and the state of stress in one particular area (Shapiro et al., 2010). On the other hand, the Hydromechanical Nucleation model (Norbeck and Rubinstein, 2018) relies on the empirical seismicity rate model from Dieterich (1994) and Segall

and Lu (2015), which relates the changes in the Coulomb stress with changes in the seismicity rates. Also, Norbeck and Rubinstein (2018) assume that the Coulomb stress change is proportional to the pressurization rate instead of the total injected volume and thus pore pressure change.

In this study, we evaluate the performance of both physics-based approaches using two tests: (1) how accurate are the rate predictions (number of earthquakes exceeding a certain magnitude per month); (2) how appropriate are the predictions for specific magnitude ranges (given by forecasted GR parameters). The first test is important since it describes the changes in the overall seismicity patterns. The results from the second test are used to predict the exceedance probability and the likelihood for a certain magnitude event to occur within a fixed time period. We apply both methodologies to one area with recent induced seismicity: the Horn River Basin, Northeast British Columbia (BC). We generate synthetic earthquake catalogs using the Monte Carlo simulation approach (Reyes Canales and Van der Baan, 2019; Assatourians and Atkinson, 2013; Musson, 2000). From these catalogs, we can obtain relevant statistics for the hazard analysis, and compare both models using the two test criteria. In order to make a complete Probabilistic Seismic Hazard Analysis (PSHA), we also include the generation of synthetic ground motion catalogs and ground motion predictions. For the first test, our results show that the predictions follow the observed induced seismicity patterns: increase and decrease of the number of earthquakes, in line with the anthropogenic activities, and a similar number of earthquakes per month. However, for the second test, the predictions for specific magnitude ranges may both under- or overestimate the hazard, among other things, due to the complexity in the evolution of the seismicity and possible changes in the GR parameters that are not seen by the models. The latter is especially relevant for the b-value, which may change over time but is assumed constant for both models.

4.2 Theory

4.2.1 Forecasting Gutenberg Richter parameters: Physics-based models

One of the most difficult challenges in the seismic hazard analysis for induced seismicity is the prediction of the Gutenberg-Richter (GR) parameters (a-and b-values). In this chapter, we describe two physics-based models to address this issue: The Seismogenic Index (Shapiro et al., 2010; Langenbruch and Zoback, 2016), and the Hydromechanical Nucleation model (Dieterich, 1994; Segall and Lu, 2015; Norbeck and Rubinstein, 2018).

Seismogenic Index: Shapiro et al. (2010) modify the classical Gutenberg-Richter magnitude-frequency distribution to include fluid injection-induced earthquakes at hydrocarbon and geothermal reservoirs. The Gutenberg-Richter magnitude-frequency distribution is given by (Gutenberg and Richter, 1944):

$$\log(N) = a_c - bM, \tag{4.1}$$

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_c-value is related to the cumulative number of N_{0c} earthquakes with a non-negative magnitude. The latter is given by:

$$N_{0c} = 10^{a_c}. \tag{4.2}$$

On the other hand, Shapiro et al. (2010) define the following magnitude-frequency distribution:

$$\log(N) = \log(Q_c(t)) + \Sigma - bm = a'_c(t) - bm, \tag{4.3}$$

where $Q_c(t)$ is the cumulative volume injected up to time t, Σ is the Seismogenic Index, and $a'_c(t)$-value is related to the cumulative number N_0c of earthquakes with a non-negative magnitude up to time t. Notice how the $a'_c(t)$-value is time-dependent and is determined by the cumulative volume injected and the Seismogenic Index. In order to get the annual $a'(t)$-value, we simply divide the

equivalent cumulative number N_{0c}, equation 4.2, by the time duration of the catalog.

The Seismogenic Index Σ incorporates the volume concentration of pre-existing faults and the state of stress in one area (Shapiro et al., 2010). It is supposed to be a constant parameter with time (Shapiro et al., 2010; Langenbruch and Zoback, 2016). In practice, the Seismogenic Index is obtained by calculating the cumulative a_c-value from a catalog with induced earthquakes and the cumulative volume injected in that time, as follows:

$$\Sigma = a_c - \log(Q_c(t)). \tag{4.4}$$

Once the Seismogenic Index is calculated for one area, it is possible to predict the changes in the cumulative $a'_c(t)$-value by adding the log of the future volume to inject to the Seismogenic Index:

$$a'_c(t) = \log(Q_c(t)) + \Sigma. \tag{4.5}$$

Notice that the assumption of constant Seismogenic Index relies on the assumption of constant b-value. Temporal changes in the b-value could alter the estimations of the Seismogenic Index and $a'_c(t)$-value.

The Seismogenic Index has been calculated and applied in different induced seismicity cases with promising results, including seismicity related to salt-water injection, hydraulic fracturing treatments, and geothermal activities (Shapiro et al., 2010; Langenbruch and Zoback, 2016).

Hydromechanical Nucleation model: Dieterich (1994) and Segall and Lu (2015) develop an empirical seismicity rate model that relates changes in the Coulomb stress with changes in the seismicity rates. The temporal evolution of seismicity rate can be described by using the following ordinary differential equation:

$$\dot{R(t)} = \frac{R(t)}{t_c}\Big(\frac{\dot{s}}{\dot{s_0}} - R(t)\Big), \tag{4.6}$$

where $R(t)$ is the ratio between the seismicity rate $r(t)$ resulting from the injection, and the background seismicity rate r_0:

$$R(t) = r(t)/r_0. \tag{4.7}$$

The stressing rate $\dot{s}$ is the Coulomb stressing rate on the faults, and $\dot{s_0}$ represents the tectonic stressing rates. The Coulomb stress is defined as:

$$s = \tau - f(\sigma - p), \tag{4.8}$$

where τ is the shear stress, f is the fault friction coefficient, σ is the normal stress, and p is the fluid pressure (Pollard and Fletcher, 2005).

Finally, the parameter t_c is the characteristic decay time and is defined by:

$$t_c = \frac{\bar{a}\bar{\sigma}}{\dot{s_0}}, \tag{4.9}$$

where $\bar{a}$ is the direct-effect parameter in the rate-and-state friction formulation, and $\bar{\sigma}$ is the effective normal stress. The parameter t_c reflects a characteristic timescale over which the seismicity $r(t)$, resulting from the injection, tends to the background seismicity rate r_0 (Norbeck and Rubinstein, 2018). The $\bar{a}$ is a friction parameter that represents the direct velocity strengthening magnitude, and it is obtained through laboratory experiments (Segall and Lu, 2015).

Equation 4.6 can be solved numerically using an explicit Runge-Kutta method with adaptive time-stepping, and embedded error estimates (Norbeck and Rubinstein, 2018). For the purpose of this study, we use ode45 in MATLAB to integrate the ordinary differential equation, as used by Segall and Lu (2015). However, in order to solve this ordinary differential equation, a model to describe the changes in the Coulomb stressing rate $\dot{s}$ is required.

Norbeck and Rubinstein (2018) assume that the changes in the Coulomb stressing rates $\dot{s}$ are approximately equivalent to the pressurization rate $\dot{p}$, then $\dot{s} \simeq \dot{p}$. The pressurization rate $\dot{p}$ in response to injection is moderated by the compressibility of the system (Horne, 1995):

$$\dot{p} = \frac{Q(t)}{V\phi\beta}, \tag{4.10}$$

where $Q(t)$ is the injection rate per time unit, V is the reservoir bulk volume, ϕ is the rock porosity, and β is the total reservoir compressibility, obtained in

laboratory experiments. These reservoir parameters are assumed unchanged for the entire stimulated reservoir, evidently resulting in a oversimplification of the reservoir properties. By solving equation 4.10, we can model the Coulomb stressing rate $\dot{s}$ necessary to solve equation 4.6, and finally obtain the parameter $R(t)$, which ultimately reflects the change in the a-value. Re-arranging equation 4.7, we have:

$$r(t) = r_0 R(t). \tag{4.11}$$

The background seimicity rate r_0 can be re-written using the notation from Reyes Canales and Van der Baan (2019), where the rate of earthquakes $\lambda(M_{min} \leq m < M_{max})$ for a magnitude range $m = [M_{min}, M_{max}]$ is given by:

$$\lambda(M_{min} \leq m < M_{max}) = P(M_{min} \leq m < M_{max}) N(M_{min} \leq m \leq M_{max}), \tag{4.12}$$

where M_{min} and M_{max} are the minimum and maximum magnitudes considered, respectively. The total expected number of earthquakes $N(M_{min} \leq m \leq M_{max})$ is given by:

$$N(M_{min} \leq m \leq M_{max}) = 10^{a - bM_{min}} - 10^{a - bM_{max}}. \tag{4.13}$$

In this equation, a and b are the parameters defined by the Gutenberg-Richter distribution (Gutenberg and Richter, 1944). Finally, the probability of occurrence $P(M_{min} \leq m < M_{max})$ is given by:

$$P(M_{min} \leq m < M_{max}) = F_M(M_{min}) - F_M(M_{max}), \tag{4.14}$$

where $F_M(m)$ denotes the cumulative distribution function for magnitude m, and is defined by:

$$F_M(m) = \frac{1 - 10^{-b(m - M_{min})}}{1 - 10^{-b(M_{max} - M_{min})}}. \tag{4.15}$$

Subsequently, the background seimicity rate r_0 in the magnitude range $m = [M_{min}, M_{max}]$ is equivalent to:

$$r_0 = \lambda(M_{min} \leq m \leq M_{max}). \tag{4.16}$$

Then, the seismicity rate resulting from the injections $r(t)$, eq. 4.11, can be rewritten as:

$$r(t) = \lambda(M_{min} \leq m \leq M_{max})R(t). \tag{4.17}$$

The parameter $\lambda(M_{min} \leq m \leq M_{max})$ can be expressed in terms of the probabilities $P(M_{min} \leq m \leq M_{max})$ and number of earthquakes $N(M_{min} \leq m \leq M_{max})$ for the magnitude range $m = [M_{min}, M_{max}]$:

$$\begin{aligned} r(t) = \ & P(M_{min} \leq m \leq M_{max}) * \\ & N(M_{min} \leq m \leq M_{max})R(t). \end{aligned} \tag{4.18}$$

Then, we can develop the expression $N(M_{min} \leq m \leq M_{max})R(t)$, taking into account the parameter $R(t)$:

$$\begin{aligned} N(M_{min} \leq m \leq M_{max})R(t) = \ & R(t)10^{a-bM_{min}} \\ & - R(t)10^{a-bM_{max}}. \end{aligned} \tag{4.19}$$

Then, by applying properties of logarithms, we have:

$$\begin{aligned} N(M_{min} \leq m \leq M_{max})R(t) = \ & 10^{\log(R(t))+a-bM_{min}} \\ & - 10^{\log(R(t))+a-bM_{max}}. \end{aligned} \tag{4.20}$$

It turns out that $\log(R(t))$ is simply added to the background a-value, resulting in a new, time-dependent $a'(t)$-value:

$$a'(t) = \log(R(t)) + a. \tag{4.21}$$

In other words, the $a'(t)$-value can be estimated by knowing the $R(t)$ parameter, which depends on the injection rates and other stress and tectonic parameters from the site, as well as the background a-value. Thus, the GR magnitude-frequency distribution for the Hydromechanical Nucleation model becomes:

$$\log(N) = \log(R(t)) + a - bm = a'(t) - bm. \tag{4.22}$$

The Hydromechanical Nucleation model has been applied in different induced seismicity cases, including seismicity related to salt-water injection in Oklahoma (Norbeck and Rubinstein, 2018). Both the Seismogenic Index and the Hydromechanical model rely on constant b-values and time-dependent a-values. In other words, the assumption of constant b-values can result in differences between the forecast and actual seismicity, even if the predicted a-values are reasonable.

4.2.2 Generation of synthetic ground motion catalogs using the Monte Carlo simulation method

Reyes Canales and Van der Baan (2019) derive analytical expressions for the seismic hazard analysis of non-stationary seismic sources. Some of these expressions include the rate of exceedance $\lambda_{exc}(m \geq m_j; t)$, and the probability $P[N = n; t_a, t_b]$ of n occurrences happening in a time interval $\Delta t = t_b - t_a$. For a non-stationary source, the rate of exceedance $\lambda_{exc}(m \geq m_j; t)$ is defined as (Reyes Canales and Van der Baan, 2019):

$$\lambda_{exc}(m \geq m_j; t) = P(m_j \leq m \leq M_{max}; t) N(M_{min} \leq m \leq M_{max}; t), \quad (4.23)$$

where $P(m_j \leq m \leq M_{max}; t)$ is the time-varying probability of occurrence of a magnitude m occurring in the range $m = [m_j, M_{max}]$, and is defined as:

$$P(m_j \leq m \leq M_{max}; t) = F_M(M_{max}; t) - F_M(m_j; t). \quad (4.24)$$

On the other hand, for a non-stationary source, the probability $P[N = n; t_a, t_b]$ of n occurrences happening in a time interval $\Delta t = t_b - t_a$ is given by:

$$P[N = n; t_a, t_b] = \frac{m_\lambda^n(t_a, t_b)(t_b - t_a)^n e^{-m_\lambda(t_a, t_b)(t_b - t_a)}}{n!}, \quad (4.25)$$

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):

$$m_\lambda(t_a, t_b) = \frac{\int_{t_a}^{t_b} \lambda(s)ds}{(t_b - t_a)}, \tag{4.26}$$

where, $\lambda(t)$ could be equivalent to the rate of earthquakes $\lambda(M_{min} \leq m < M_{max})$ for a magnitude range $m = [M_{min}, M_{max}]$, eq.4.12. Equation eq.4.19 is known as the non-stationary Poisson distribution.

Likewise, we can obtain similar expressions for ground motion values in the context of non-stationary seismicity. An essential expression in the seismic hazard analysis is the rate of exceeding a ground motion $\lambda(gm \geq gm_j; t)$, which in the context of non-stationary seismic sources is given by (Baker, 2008, 2013):

$$\begin{aligned} \lambda(gm \geq gm_j; t) = \; & \lambda(M_{min} \leq m \leq M_{max}, t) * \\ & \int_{M_{min}}^{M_{max}} \int_0^r P(gm > gm_j \mid m, r) * \\ & f_m(m, t) f_R(r) dr dm, \end{aligned} \tag{4.27}$$

where $P(gm > gm_j \mid m, r)$ is the probability to exceed a ground motion gm_j in a time t, given a magnitude m and distance r. This probability is given by the Ground Motion Prediction Equations (GMPEs), which require as input the magnitude and distance of the event in order to predict the ground motion. $f_m(m, t)$ is the probability density function (PDF) of the magnitudes, and $f_R(r)$ is the PDF of the distance. Notice that the time-dependency in the rate of exceeding a ground motion $\lambda(gm \geq gm_j; t)$ is given by the time-dependency in the rate of earthquakes $\lambda(M_{min} \leq m \leq M_{max}, t)$ and magnitude PDF $f_m(m, t)$, consequence of time-dependent GR parameters described by Reyes Canales and Van der Baan (2019). The plot of the rate of exceedance as a function of ground motions is also known as the seismic hazard curve (Baker, 2008, 2013).

Like the rate of exceeding a ground motion $\lambda(gm \geq gm_j; t)$ might change over time, the mean rate of exceeding a ground motion $m_\lambda(gm \geq gm_j; t_a, t_b)$ in a time interval $t = [t_a, t_b]$, is an useful expression in the hazard analysis for non-stationary sources:

$$m_\lambda(gm \geq gm_j; t_a, t_b) = \frac{\int_{t_a}^{t_b} \lambda(gm \geq gm_j; s)ds}{(t_b - t_a)}. \tag{4.28}$$

However, solving analytically these expressions can be rather complicated. Fortunately, the Monte Carlo simulation method can easily obtain numerical results for these expressions (Musson, 2000; Assatourians and Atkinson, 2013).

First, a synthetic earthquake catalog has to be generated. This methodology has been described in Reyes Canales and Van der Baan (2019). The generation of synthetic earthquake catalogs can be summarized in two main steps: (1) Simulation of a temporal point process to obtain arrival times for a single realization; (2) Simulation of earthquake magnitudes m (Musson, 2000; Assatourians and Atkinson, 2013; Reyes Canales and Van der Baan, 2019). To generate arrival times for a single realization, a Monte Carlo sampling is performed to the inverse cumulative distribution function (CDF) of the Poisson distribution with rate λ. In this step, the earthquake recurrence λ is defined by the GR parameters, and the average number of samples in one realization is given by the recurrence rates and the period of time to simulate. To generate magnitudes for each time arrival, another Monte Carlo sampling is performed to the inverse CDF of the GR recurrence law. Consequently, each pair of time arrivals and magnitudes represent a synthetic event. Finally, the location of these synthetic events is assumed to be uniformly random distributed within the seismic source area.

The synthetic earthquake catalogs could be given by either stationary or non-stationary seismicity rates. Then, the distance between a given site and the location of a synthetic event is calculated. Given the distance and magnitude of the synthetic events, we use a set of GMPEs to estimate the corresponding ground motion at one site for that event. We repeat this process for all synthetic events, resulting in a catalog of ground motions. In order to count for the aleatoric variability in the ground motions, a random number drawn from the standard normal distribution is multiplied by the sigma value (variability of the GMPE model) and added to the log of the ground motions (Assatourians and Atkinson, 2013).

Finally, from these ground motion catalogs, we can obtain statistical results

for the hazard analysis. For instance, the mean rate of exceeding a ground motion $m_\lambda(gm \geq gm_j; t_a, t_b)$ is given by counting all the ground motions greater than gm_j, and dividing this count by the time duration of the simulations $(t_b - t_a)$ and the number of realizations N_r of the synthetic catalog:

$$m_\lambda(gm \geq gm_j; t_a, t_b) = \frac{N_{(gm \geq gm_j)}}{N_r(t_b - t_a)}, \tag{4.29}$$

where $N_{(gm \geq gm_j)}$ is the number of events with a ground motion greater than gm_j.

Notice that the ground motion predictions in a non-stationary seismicity case depend on the variable seismicity rates as well as the time interval $(t_b - t_a)$ to evaluate.

In the case of the ground motion analysis, it is possible to relate the probability $P[N > 0; t_a, t_b]$ of at least one event to exceed a ground motion gm in a time interval $[t_a, t_b]$ to a mean rate of exceeding a ground motion $m_\lambda(gm \geq gm_j; t_a, t_b)$:

$$m_{\lambda,exc}(gm \geq gm_j; t_a, t_b) = \frac{-\ln(1 - P[N > 0; t_a, t_b])}{t_b - t_a}. \tag{4.30}$$

This is useful since the probability $P[N > 0; t_a, t_b]$ is often provided in seismic hazard analyses. For instance, using equation 4.30, the 1% probability to reach or exceed a given ground motion in 1 year is equivalent to the mean rate of exceeding a ground motion $m_\lambda(gm \geq gm_j; t_a, t_b)$ of 0.01. 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.

Other useful results in the seismic hazard analysis are the seismic hazard maps (Baker, 2008; Assatourians and Atkinson, 2013). These maps describe the probability of reaching or exceeding a ground motion in a period of time. A seismic hazard map is built by calculating first the seismic hazard curves at different sites. Then, a common mean rate of exceeding a ground motion $m_\lambda(gm \geq gm_j; t_a, t_b)$ is defined for all seismic curves, and the related ground motion is selected. This yields a ground motion value for each considered site, which in turn can be interpolated to generate a seismic hazard map. Figure 4.1 shows a sketch of the methodology for the generation of synthetic ground

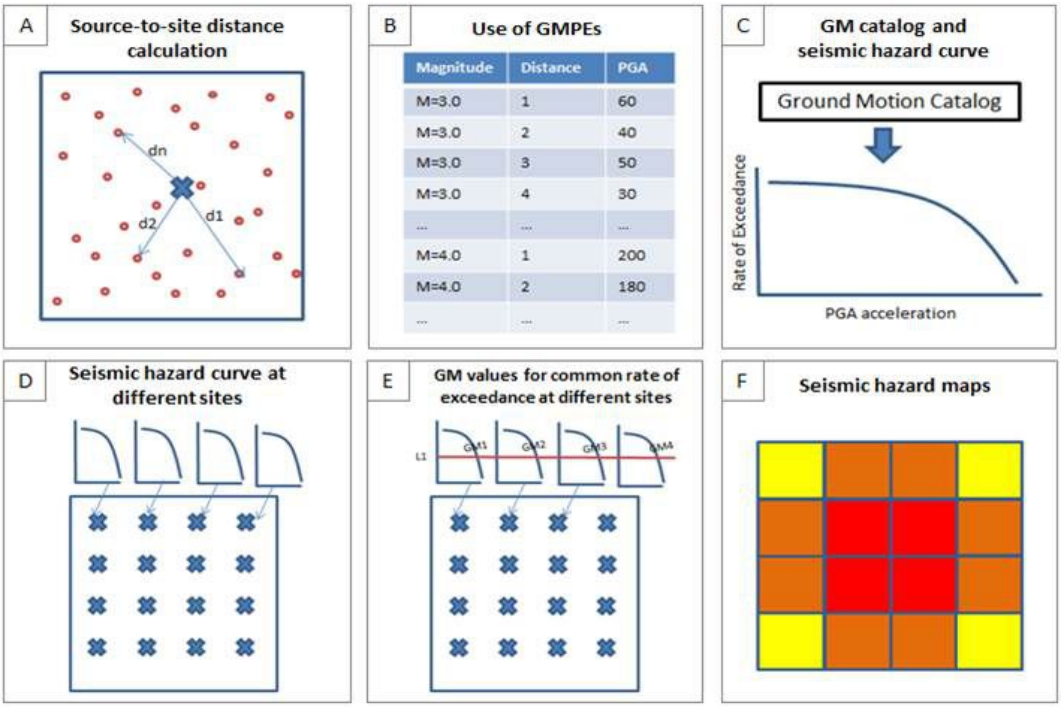

Figure 4.1: Sketch of the methodology for the generation of synthetic ground motion catalogs, seismic hazard curves, and seismic hazard maps. (A) Calculate all distances between a given site and the location of the synthetic events (Source-to-site distance). (B) Given the distances and magnitudes of the synthetic events, we use the GMPEs to estimate the corresponding ground motions at one site. (C) Seismic hazard curves (Rate of exceedance as a function of ground motions) from ground motion catalogs. (D) Seismic hazard curves calculated at different sites. (E) Define a common rate of exceedance level and obtain the corresponding ground motion for each seismic hazard curve. (F) Knowing the ground motion at each site, it is possible to build a seismic hazard map.

motion catalogs, seismic hazard curves, and seismic hazard maps.

4.3 Implementation

1. **Forecasting GR parameters based on physics-based models**: We rely on the Seismogenic Index and the Hydromechanical Nucleation model to forecast the GR parameters associated with induced earthquakes and, therefore, forecast the seismic hazard.

   In the case of the Seismogenic Index, once the induced seismic activ-

ity has started, we can calculate the GR parameters from the recorded earthquake catalog and reported injected volumes. There are different methodologies to calculate GR parameters from an earthquake catalog, including the maximum likelihood method (MLM, Aki (1965), Wiemer and Wyss (1997)). We calculate the total volume injected $Q_c(t)$, and by knowing the corresponding cumulative a_c-value, we can use equation 4.4 to obtain the Seismogenic Index value. Assuming a constant b-value and Seismogenic Index, and knowing the future injections plans $Q_c(t)$, we use equation 4.6 to calculate the future cumulative $a'_c(t)$- value. While the injection plans and induced seismicity continues, we can regularly re-calculate the GR parameters and Seismogenic Index to improve the seismic hazard predictions. This is especially important, considering that the assumption of constant GR parameters might not be appropriate.

For the Hydromechanical Nucleation model, we first calculate the pressurization rate $\dot{p}$ (eq. 4.10), using the catalog of injection rate per time unit $Q(t)$ and the required reservoir parameters. Then, assuming that the Coulomb stressing rate is approximate to the pressurization rate, $\dot{s} \simeq \dot{p}$, we solve equation 4.10 to obtain $R(t)$. Finally, by using 4.21, we add the $\log(R(t))$ to the background a-value in order to obtain the $a'(t)$-value resulting from the injection rates. In theory, we could make a forecast of the GR parameters even before the start of the injection. However, we require some unknown parameters at the site such as the background stressing rates, the effective normal stress, and the total perturbed reservoir volume. In practice, the parameters are calibrated in order to fit the model prediction with the induced seismicity pattern. Then, by knowing future injection plans, we could have more reliable forecasts from this calibrated model.

2. **Generation of synthetic earthquake catalogs**: Once the GR parameters have been defined, we proceed to generate synthetic earthquake catalogs. A description of the Monte Carlo simulation for the generation of synthetic earthquake catalogs has been done by Reyes Canales and Van der Baan (2019). For the generation of synthetic earthquake catalogs, we require the GR parameters and uncertainties to calculate seismicity rates

(can be stationary or non-stationary), define a source area, the duration of induced catalogs, the number of realizations N_r, and define the range of magnitudes to simulate $m = [M_{min}, M_{max}]$. From these catalogs we can obtain useful statistics for the hazard analysis, including the mean annual rate of earthquakes $m_\lambda(m_j \leq m < m_{j+1}; t_a, t_b)$, the mean annual rate of exceedance $m_{\lambda,exc}(m \geq m_j; t_a, t_b)$, and the probability of n occurrences for a magnitude range.

3. **Generation of synthetic ground motion catalogs**: Once the synthetic earthquake catalogs have been simulated, we proceed to generate synthetic ground motion catalogs. First, we calculate the source-to-site distances. Then, together with the magnitudes of each synthetic events, we use the GMPEs to estimate the corresponding ground motion for each event, resulting in a catalog of ground motions. From these synthetic ground motion catalogs, we can obtain useful statistics for the hazard analysis, including the seismic hazard curves (Equation 4.29) and seismic hazard maps. In the case of multiple GMPEs, PSHA uses the logic tree approach, which gives a weight to each set of GMPEs (Assatourians and Atkinson, 2013). We split proportionally the synthetic earthquake catalog based on the weight of each GMPEs. Next, we calculate the ground motions for each synthetic event using the corresponding GMPE, resulting in a ground motion catalog per branch.

4.4 Horn River Basin case

To evaluate the Seismogenic Index and the Hydromechanical Nucleation models for the forecasting of seismic hazard, we apply these methods in one area with recent induced seismicity: the Horn River Basin, Northeast BC. Previous studies (Reyes Canales and Van der Baan, 2019) have shown how the seismic hazard change over time in response to the inclusion of injection-induced earthquakes. In this study, we compare the hazard given by actual earthquake catalogs (Observations) with the hazard provided by physics-based models (Predictions). This comparison is made in terms of the number of forecasted earthquakes in

a magnitude range, GR predictions, and predicted ground motions.

4.4.1 Area of study and data

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). The detected seismicity in the area was very low before 2006, but with a significant increase since Dec. 2006, particularly between 2010 and 2011 in line with increasing injection rates. The target reservoirs were the shale plays related to the Muswa-Otter Park and Evie Formations. Figure 4.2 (A) shows the location of the seismic events from the Farahbod et al. (2015b) catalog.

To calculate the Seismogenic Index index in the area, we use the earthquake catalog from Farahbod et al. (2015b), which contains induced earthquakes in the Horn River Basin. This earthquake catalog consists of 338 events recorded between Dec. 2006 and Dec. 2011, with magnitudes ranging between $m = 1.0$ and $m = 3.6$. We also require catalogs of injected volumes for both the Seismogenic Index and the Hydromechanical Nucleation model. For the injected volumes, we use the catalog from Farahbod et al. (2015a), which contains the volumes injected per month at the Horn River basin between Dec. 2006 and Dec. 2011. Figure 4.2 (B) shows the injected volumes per month at the Horn River Basin from Dec. 2006 up to Dec. 2011 (Farahbod et al., 2015a).

4.4.2 Forecasting seismic hazard using the Seismogenic Index

To calculate the Seismogenic Index and study the temporal evolution of the GR during the period of induced seismicity, we define a time window starting in Dec. 2006 with a moving endpoint. The earthquakes in that time window are selected to calculate the GR parameters, which are then assigned to the endpoint of that time window. This process is repeated by moving the endpoint of the time window one month ahead every time, until the end of the catalog

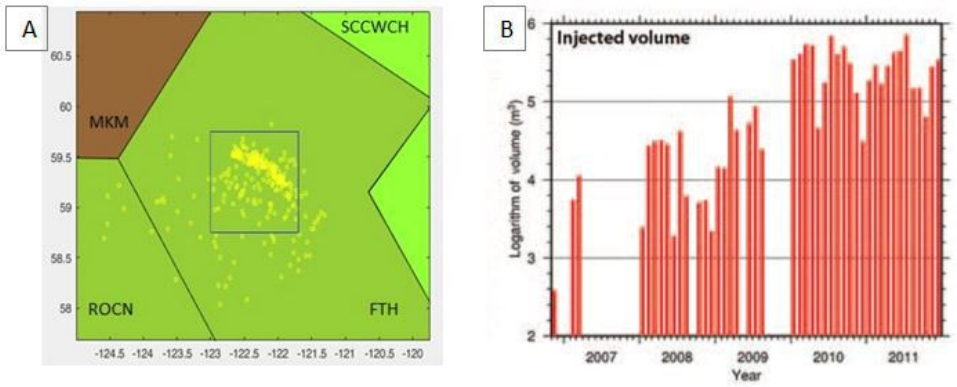

Figure 4.2: (A) Location of the seismic events (yellow dots) using the catalog from Farahbod et al. (2015b). The blue square indicates the seismic source area used in this study. The polygons in the picture represent some of the natural sources defined by the 2015 seismic-hazard model of Canada (Halchuk et al., 2014). FTH: Foothills, ROCN: Rocky Mountain fold/thrust belt north, MKM: Mackenzie Mountains, SCCWHC: Stable cratonic core western Canada (H model). (B) Monthly injection rates in the Horn River Basin. Notice the substantial increase in the injection volumes after Dec. 2009. From Farahbod et al. (2015a).

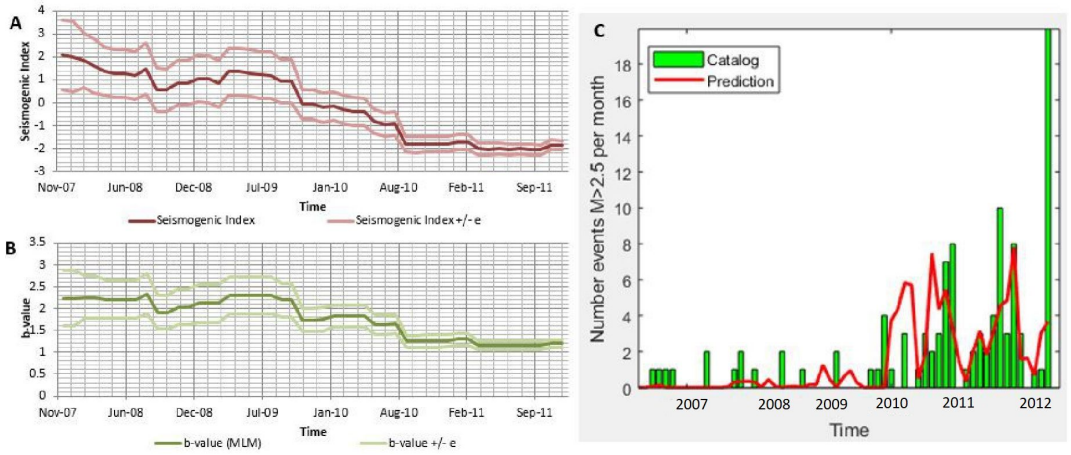

Figure 4.3: Evolution of the (A) Seismogenic Index, and (B) b-value at the Horn River Basin, using the earthquake catalog from Farahbod et al. (2015b) and injection rates from Farahbod et al. (2015a). (C) Monthly number of earthquakes with magnitude $M \geq 2.5$ given by the observed seismicity (green bars) and the Seismogenic Index predictions (red line).

(Dec. 2011).

To calculate the GR parameters, 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 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 MLM (Aki (1965), Wiemer and Wyss (1997)), that accounts for the magnitude binning. The cumulative a_c-values are calculated from the b-values and the number of earthquakes N larger than $M_c = 2.4$ in the catalog, using equation 4.1. The uncertainties in the GR parameters are also included. This is done by adding or subtracting the errors estimated for the b-values, and then using the N number of earthquakes from the catalog and eq. 4.1, to calculate the a-values. This results in 3 curves: middle, upper, and lower curves. For the estimations of the error in the b-values, we use the method described by Shi and Bolt (1982).

To calculate the seismogenic index at the endpoint of a time window, we calculate the total volume injected $Q_c(t)$ up to that time, and by knowing the corresponding cumulative a_c-value, we use equation 4.4 to obtain the Seismogenic Index value. We repeat this process for each endpoint. Figure 4.3 shows (A) the temporal evolution of the Seismogenic Index, and (B) the b-value at the Horn River Basin, using the time window methodology with moving endpoint.

The first test consists in evaluating the accuracy of this method to predict seismic rates for monthly earthquakes rates with magnitude $M \geq 2.5$. Using equation 4.3, we calculate the number of earthquakes with $M \geq 2.5$ per month, given the estimated Seismogenic Index value and the monthly injection rates. Figure 4.3 (C) shows the monthly number of earthquakes $M \geq 2.5$, using a Seismogenic Index of -1.82 and a b-value of 1.25. These values correspond to December 2010, when both parameters reach a more stable behavior (see figures 4.3 (A) and (B)). By comparing the predictions with the observed number of earthquakes, we found that the predictions follow the changing seismicity patterns. Furthermore, the total number of observed earthquakes $M \geq 2.5$ between Dec. 2006 and Dec 2011 was 121 events, in comparison with the 92.24 predicted events by the seismogenic Index. The most remarkable difference is

related to the underestimation of events in Dec. 2011. The Seismogenic Index predicted 4 events, but there were 32 events in that month. With the exception of this anomaly, the model predicts a reasonable number of earthquakes per month. If the events from December 2011 are not included in the analysis, the seismogenic index model predicts 88 events $M \geq 2.5$, which is very similar to the 89 observed events from December 2006 up to November 2011. The correlation coefficient between the predicted and observed number of earthquakes $M \geq 2.5$ is 0.51; however, the correlation coefficient increases to 0.60 if the events from December 2011 are not included.

The second test consists in evaluating how accurate are the Seismogenic Index predictions for specific magnitude ranges. We compare the annual magnitude-frequency distributions given by the GR parameters from the observed seismicity and the Seismogenic Index, as well as the probable number of earthquakes in a magnitude range derived from these GR parameters. We assume that the b-value and Seismogenic Index of one year are given by the b-value and Seismogenic Index from the last month of the previous year. The rationale behind this is to select the most up-to-date b-value available. On the other hand, the cumulative $a'_c(t)$- value will be given by adding the log of the future total volume to inject $\log(Q_c(t))$, and the Seismogenic Index Σ, see equation 4.6. For instance, we assume that the b-value and Seismogenic Index for 2011 are $b = 1.25$ and $SI = 1.82$, which are the parameters for Dec. 2010 (See figure 4.3). Also, knowing that the log of the total volume to inject in 2011 is $\log(Q_c(t)) = 6.54$, we have a forecasted cumulative $a'_c(t) = 4.71$ for 2011. Since the time duration of the forecast is one year, cumulative and annual a-values are equivalent.

We compare the GR parameters given by the predictions and the GR parameters from observed seismicity. To calculate the GR parameters per year from the observed seismicity, we first split the earthquake catalog into individual years. Then, by using a magnitude of completeness $M_c = 2.4$, we apply the MLM (Aki (1965), Wiemer and Wyss (1997)) and equation 4.1 to estimate the a-and b-values for each year. Again, for the estimations of the b-values error, we use the method described by Shi and Bolt (1982). Figure 4.4 (left column) shows a comparison between the GR parameters predicted by the Seismogenic

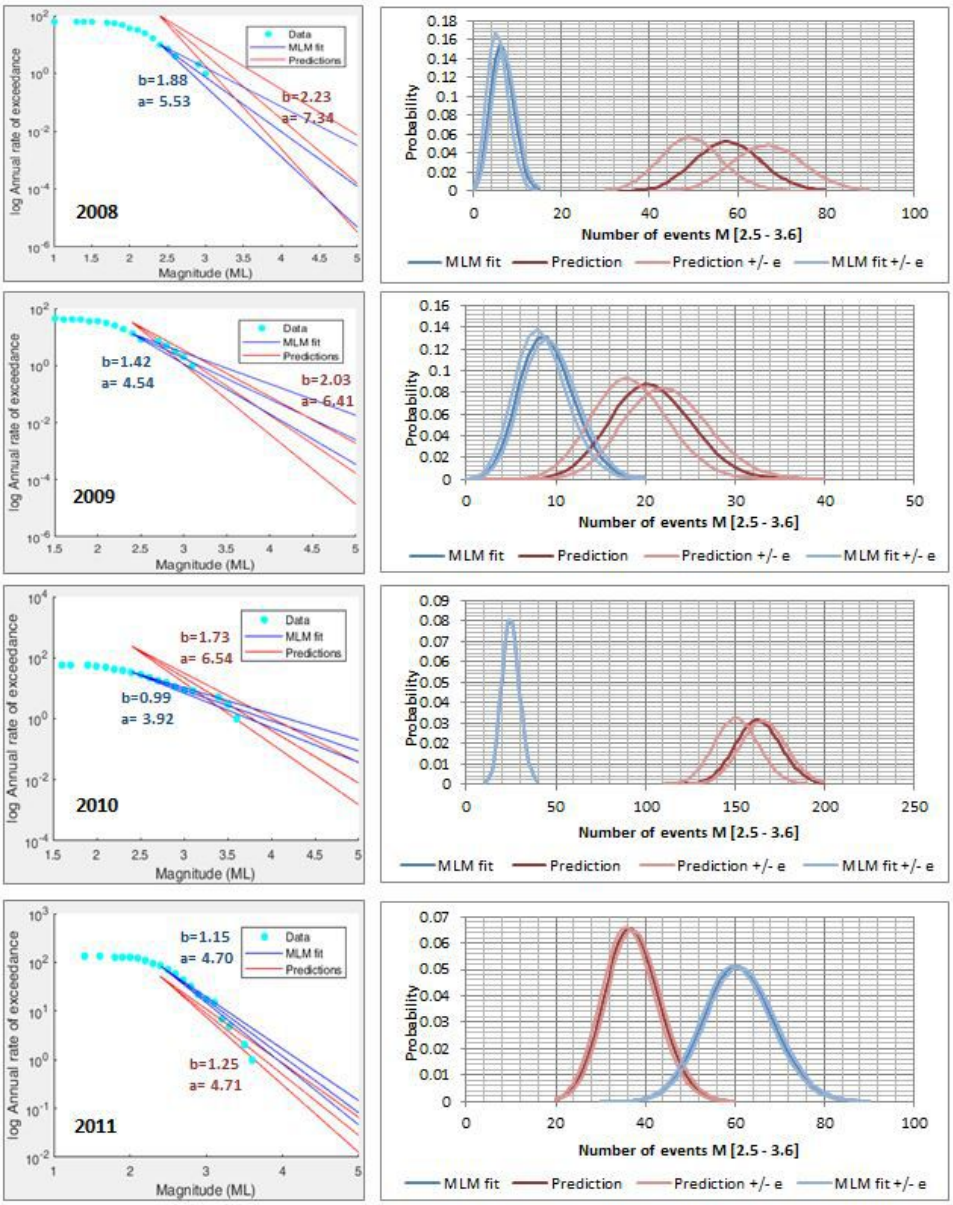

Figure 4.4: The column on the left shows a comparison between the GR parameters predicted by the Seismogenic Index (red line) and the GR parameters of the recorded seismicity (blue line, MLM fitting) for the years 2008, 2009, 2010, and 2011. The upper and lower curves in the annual rate of exceedance $\lambda_{exc}(m \geq m_j; t)$ vs. magnitude plot, shows the uncertainty of the GR parameter estimation. The column on the right shows the probable number of earthquakes, $P[N = n; t_a, t_b]$ of n, in the range $m = [2.5, 3.6]$, for the years 2008, 2009, 2010 and 2011, given by the GR parameters of the Seismogenic Index (red line, prediction) and the observed earthquake catalog (blue line, MLM fitting). The light red and blue curves are the probability distributions given by the upper and lower GR curves. MLM=Maximum Likelihood method.

Index and the GR parameters given by the observed seismicity for the years 2008, 2009, 2010 and 2011. We observe that there is a mismatch between the GR parameters from the predictions and the actual seismicity. This mismatch is larger in 2008 and 2010, as a consequence of uncertain GR parameters due to the small amount of initially observed events (2008), and a sudden change in the b-value (2010). For the year 2011 the a-values are very similar, $a = 4.70$ for the observed seismicity and $a = 4.71$ for the predictions. However, a change in the b-value from $b = 1.25$ to $b = 1.15$ causes the mismatch of observed versus predicted magnitude-frequency distributions in the year 2011.

Finally, we compare the probability $P[N = n; t_a, t_b]$ of n occurrences in a time interval $\Delta t = t_b - t_a$ using the observed and predicted GR parameters, eq 4.25. Figure 4.4 (right column) shows the probable number of earthquakes in the range $m = [2.5, 3.6]$, for the years 2008, 2009, 2010 and 2011, given by the GR parameters of the Seismogenic Index and the observed seismicity. We also include the distributions resulting from the upper and lower GR curves to count for the uncertainty. The largest mismatch between the observed and predicted number of events is observed for the years 2008 and 2010, where we had the largest difference between GR parameters. E.g. for the year 2010, the most likely number of predicted events in the range $m = [2.5, 3.6]$ is 165, which contrasts with the 26 events estimated by the observed GR parameters and the recorded 29 events in the catalog. On the other hand, for the year 2011, the mismatch is reduced, as a consequence of similar GR parameters. E.g. for the year 2011, the most likely number of predicted events in the range $m = [2.5, 3.6]$ is 38, which contrasts with the 60 events estimated by the observed GR parameters and the recorded 70 events in the catalog.

4.4.3 Forecasting seismic hazard using the Hydromechanical Nucleation model

The Hydromechanical Nucleation model requires multiple parameters that are unknown for the Horn River Basin. As a first approach, we use similar parameters to those described by Norbeck and Rubinstein (2018) in Oklahoma, with modifications to get a better fit between predictions and observed number

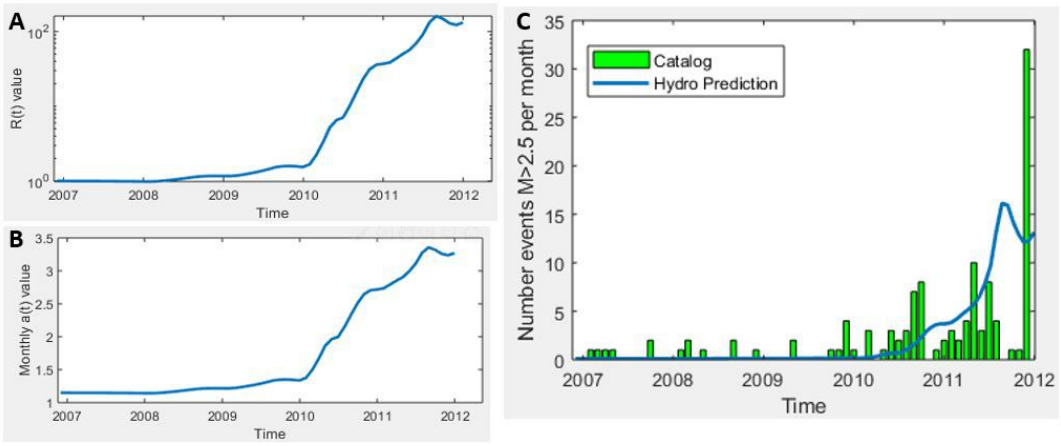

Figure 4.5: Evolution of the (A) $R(t)$ parameter, and (B) a-value (monthly) at the Horn River Basin, given by the Hydromechanical Nucleation approach. (C) Comparison between the monthly number of earthquakes larger than $m \geq 2.5$, given by the observed seismicity and the Hydromechanical Nucleation model.

of earthquakes per month. For the Hydromechanical Nucleation model in the Horn River Basin we assume the following parameters: Background stressing rate $\dot{s_0} = 0.5 \times 10^{-3} MPa.yr^{-1}$, direct effect parameter $\bar{a} = 0.0065$, effective normal stress at seismogenic depth $\bar{\sigma} = 20MPa$, average porosity for the Muskwa-Otter Park and Evie Formations $\phi = 0.05$ (BC Oil and Gas Commission, 2012, 2014), total reservoir compressibility $\beta = 3.2 \times 10^{-10} Pa^{-1}$, average thickness for the Muskwa-Otter Park, Evie Formations $h = 175m$ (BC Oil and Gas Commission, 2012, 2014), approximate area of the reservoir $A = 6000km^2$. We should notice that perform a sensitivity analysis to determine which reservoir property has a higher impact on the estimation of the number of earthquakes. We found that the total reservoir volume, defined by the area and the thickness of the reservoir, has a substantial impact on the expected earthquake rates.

For the background GR parameters, we assume a b-value of 0.86 ± 0.23, as described by the 2015 national seismic-hazard model of Canada (Halchuk et al., 2014) for the Foothills area, which contains the Horn River Basin. Based on the number of earthquakes identified between July 2002-July 2003 in the study made by Farahbod et al. (2015b), we assume $N = 1.58$ earthquakes larger than $m = 2.4$ per year. Using equation 4.1, with $b = 0.86$, we get an a-value for the background seismicity of $a = 2.26$. By re-scaling the a-values described by the 2015 national seismic-hazard model of Canada (Halchuk et al., 2014) for the

Foothills area, we get a considerable bias in the predictions. We think that the a-value given by the seismic-hazard model of Canada may be underestimated in this area due to the lack of seismometers prior to 2002.

Once the parameters for the Hydromechanical Nucleation model have been defined for the Horn River Basin, we proceed to calculate the pressurization rate $\dot{p}$, eq. 4.10, using the catalog of injected volumes per month, figure 4.2. Then, assuming that the Coulomb stressing rate is approximate to the pressurization rate, $\dot{s} \simeq \dot{p}$, we solve equation 4.10 to obtain $R(t)$, which is the ratio between the seismicity rate resulting from the injections $r(t)$, and the background seismicity rate r_0. Finally, by using 4.21, we add the $\log(R(t))$ to the background a-value in order to obtain the $a'(t)$-value resulting from the injection rates. Figure 4.5 shows (A) the evolution of the $R(t)$ parameter, and (B) the monthly a-value at the Horn River Basin, predicted by the Hydromechanical Nucleation model. To obtain the annual $a'(t)$-values, we simply add the equivalent N_0 of these $a'(t)$-values and return back by applying log. The uncertainties in the GR parameters are also included by adding and subtracting the errors estimated for the b-values.

For the first test, we calculate the number of earthquakes with $M \geq 2.5$ per month, see figure 4.3 (C). This is done by knowing the monthly $a'(t)$-values given by the Hydromechanical Nucleation approach, eq. 4.21. Contrary to the Seismogenic Index model, the Hydromechanical Nucleation approach tends to overestimate the number of events. The total number of observed earthquakes $M \geq 2.5$ between Dec. 2006 and Dec 2011 was 121 events, in comparison with the 189 predicted events by the Hydromechanical Nucleation approach. On the other hand, the predictions follow the changing seismicity patterns, although the changes in the seismicity rates are relatively smooth. The peak in the model replicates the high seismicity in late 2011. The correlation coefficient between predicted and observed number of earthquakes $M \geq 2.5$ is 0.55. It increases to 0.56 if the events from December 2011 are not included.

For the second test, we compare the annual magnitude-frequency distributions given by the GR parameters from the observed seismicity and the Hydromechanical Nucleation approach. We should clarify that the b-value for the prediction is assumed constant, and the only parameter that changes is

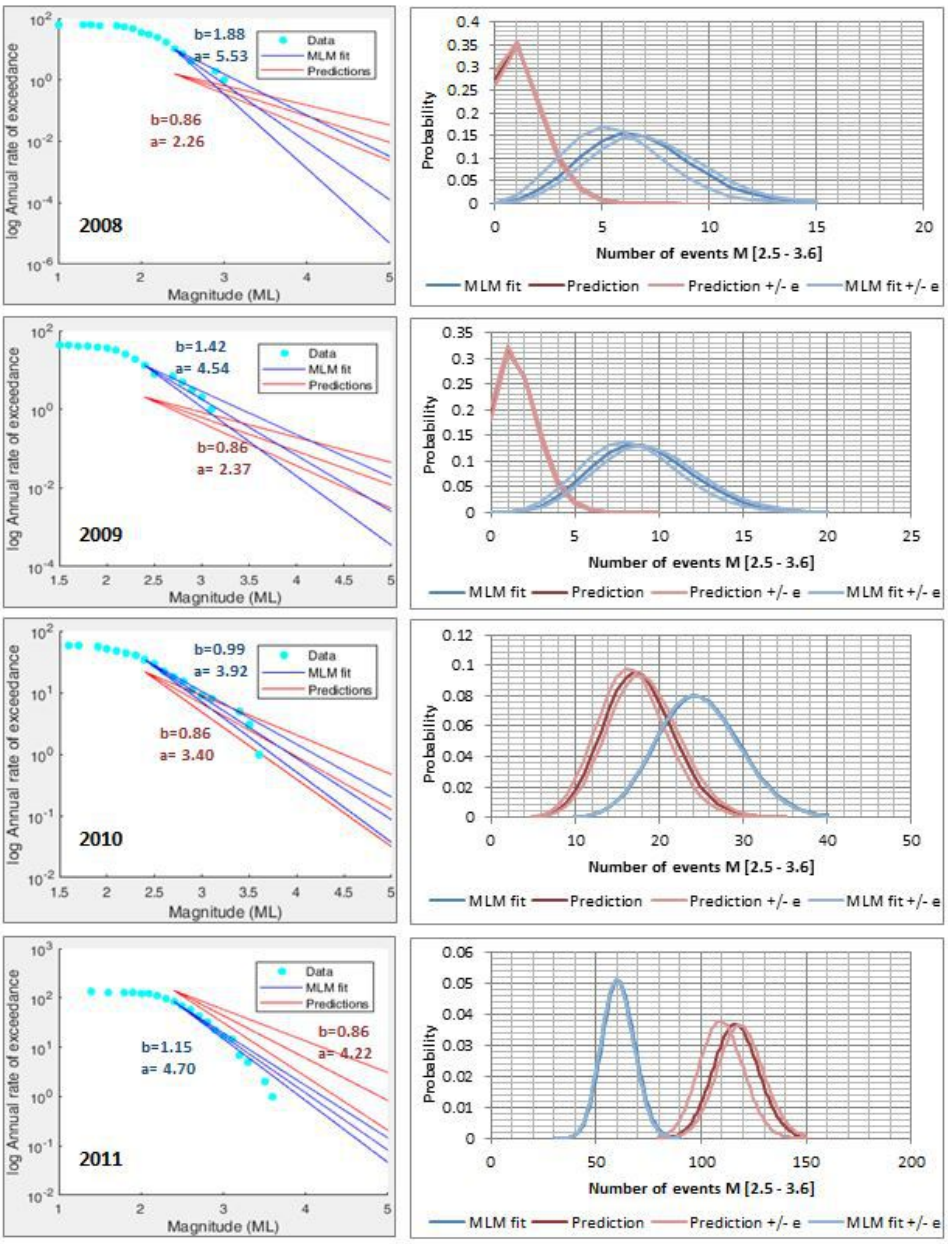

Figure 4.6: The column on the left shows a comparison between the GR parameters predicted by the Hydromechanical approach (red line) and the GR parameters of the recorded seismicity (blue line, MLM fitting), for the years 2008, 2009, 2010 and 2011. The upper and lower curves in the annual rate of exceedance $\lambda_{exc}(m \geq m_j; t)$ vs. magnitude plot, shows the uncertainty of the GR parameter estimation. The column on the right shows the probable number of earthquakes, $P[N = n; t_a, t_b]$ of n, in the range $m = [2.5, 3.6]$, for the years 2008, 2009, 2010 and 2011, given by the GR parameters of the Hydromechanical approach (red line) and the observed earthquake catalog (blue line, MLM fitting). The light red and blue curves are the probability distribution given by the upper and lower GR curves. MLM=Maximum Likelihood Method.

the $a'(t)$-value obtained from the $R(t)$ parameter. Figure 4.6 (Left column) shows a comparison between the GR parameters given by the Hydromechanical Nucleation model and the GR parameters of the observed seismicity for the years 2008, 2009, 2010 and 2011. We equally observe that there is a mismatch between the GR parameters from the prediction and the actual seismicity. For instance, for the year 2008, we assume $b = 0.86$, and predict $a = 2.26$, whereas the GR parameters for the seismicity in that year were $b = 1.88$ and $a = 5.53$. The assumption of $b = 0.86$ is based on the 2015 national seismic-hazard model of Canada (Halchuk et al., 2014) for the Foothills area.

Finally, as the most stringent test, we compare the probability $P[N = n; t_a, t_b]$ of n occurrences in a time interval $\Delta t = t_b - t_a$ using the observed and predicted GR parameters, eq 4.25. Figure 4.6 (Right column) shows the probable number of earthquakes in the range $m = [2.5, 3.6]$, for the years 2008, 2009, 2010 and 2011, given by the GR parameters of the Hydromechanical Nucleation model and the observed seismicity. Notice that our predictions rely on unchanged b-values, and this might not be appropriate in reality. Even small changes can lead to biased predictions, in some cases overestimating (2011) or underestimating the number of earthquakes (years 2008, 2009, 2010). For instance, for the year 2011, we assume $b = 0.86$ and predict $a = 4.22$ resulting in 118 earthquakes in the range $m = [2.5, 3.6]$. However, the GR parameters in the actual seismicity were $b = 1.15$ and $b = 4.70$, resulting in 60 earthquakes in the range $m = [2.5, 3.6]$.

To summarize, the Hydromechanical Nucleation model predicted 189 events $M \geq 2.5$ between December 2006 and December 2011 in the first test. In comparison, the Seismogenic Index predicted 92.24 events $M \geq 2.5$ in the same period of time. Overall, the correlation coefficient was slightly higher for the Hydromechanical Nucleation approach rather than the Seismogenic Index, 0.56 versus 0.51. However, if December 2011 is not included in the calculations, the Seismogenic Index shows a higher correlation, 0.60. For the second test, the GR parameters forecasted by the Hydromechanical Nucleation approach in 2011 were $b = 0.86$ and $a = 4.22$, resulting in 118 earthquakes in the range $m = [2.5, 3.6]$. On the other hand, the GR parameters forecasted by the Seismogenic Index in 2011 were $b = 1.2$ and $a = 4.70$, resulting in 38 earthquakes in the

range $m = [2.5, 3.6]$. Opposite trends are found for other years. For the second test, both models under- or overestimated the number of earthquakes for specific magnitude ranges.

4.4.4 Ground motion predictions

In the previous section, we showed that the physics-based models could lead to biased forecasts of magnitude-frequency distribution for this specific case study. However, we proceed to develop a full PSHA with these forecasted GR parameters to show the impact of incorrect predictions in the PSHA final products, particularly the seismic hazard curves and maps. Once the GR parameters per year have been estimated for the predictive models and the observed seismicity, we proceed to generate synthetic earthquake catalogs using the Monte Carlo Simulation method. We compute independent simulations for each model (Seismogenic Index, Hydromechanical Nucleation, and observed seismicity), for the years 2008, 2009, 2010, 2011. Each simulation comprises $N_r = 10,000$ realizations with a time duration of 1 year. We consider a magnitude bin size of $M_{bin} = 0.1$, and 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. To account for the uncertainties in the GR parameters, we simulate each weighted set of a-and b-values ($w = 0.68$ for the middle curve and $w = 0.16$ for the lower and upper curves) and then combine all simulations to obtain a unique catalog. For the source area, we define a square that covers all the events from the seismic catalog, see figure 4.2. We assume a spatially uniform random distribution of synthetic earthquakes.

To generate synthetic ground motion catalogs, we calculate source-to-site distances and use the GMPEs. For the Horn River Basin case, we use the GMPEs proposed by Atkinson et al. (2015) for induced events (A15 GMPEs). The A15 GMPEs are a modification of the GMPEs proposed by Atkinson (2015), which were designed for small-to-moderate events at short hypocentral distances. From the synthetic ground motion catalogs, we calculate the seismic hazard curves (equation 4.29) at the center of the source area, which is the location with higher expected ground motions.

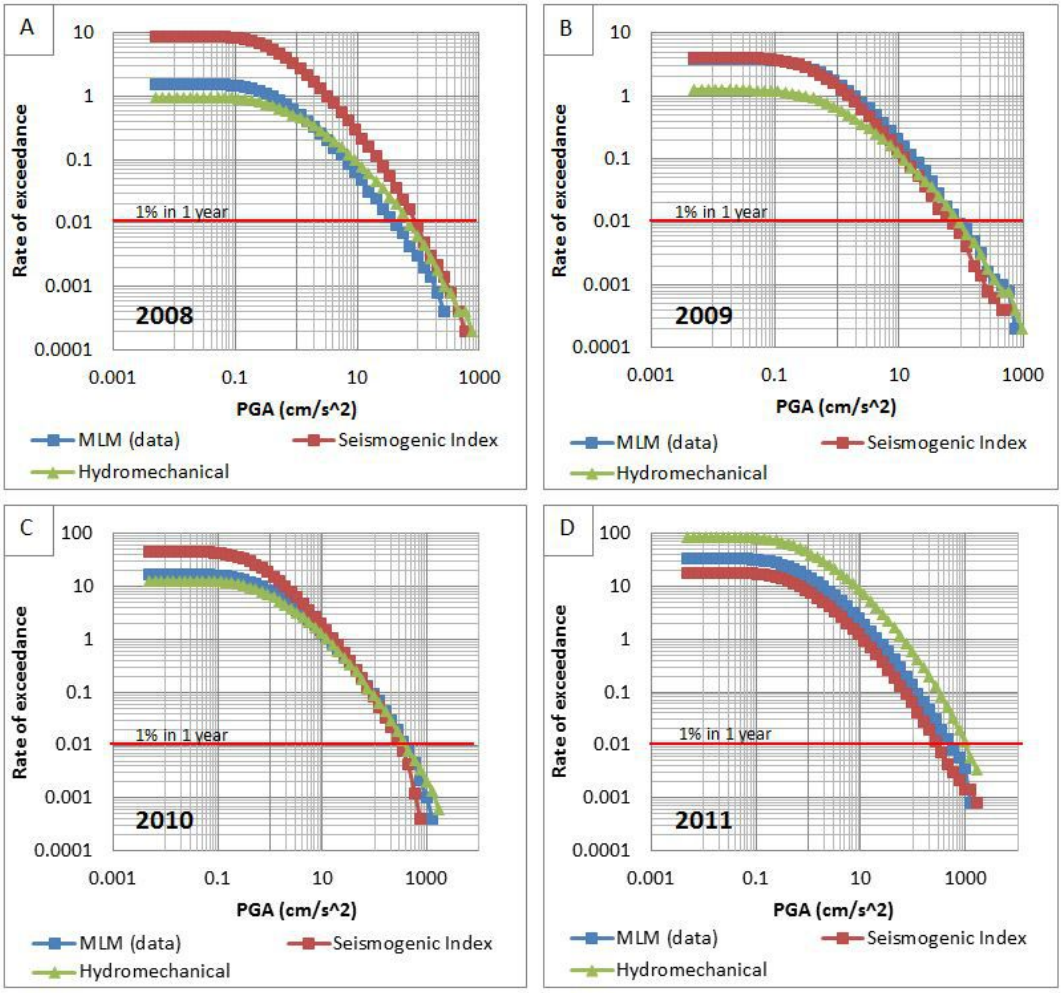

Figure 4.7: Seismic hazard curves (rate of exceedance $\lambda(gm \geq gm_j; t)$ as a function of ground motions) for the years (A) 2008, (B) 2009, (C) 2010, and (D) 2011 given by the Seismogenic Index forecast (red line), the Hydromechanical approach forecast (green line) and the observed seismicity (blue line, MLM fitting), for one location at the center of the seismic source area. The likelihood occurrence 1% in 1 year is indicated in the plot. MLM= Maximum Likelihood method.

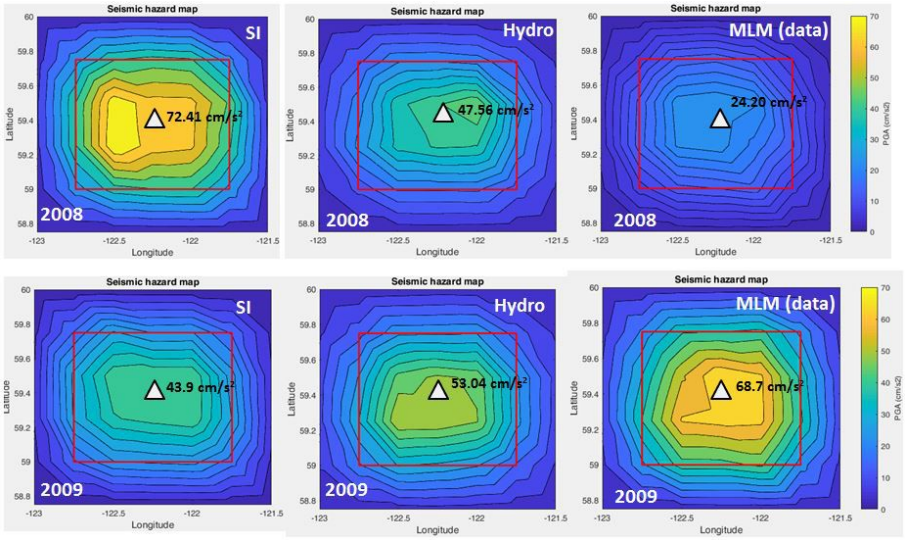

Figure 4.8: Seismic hazard maps for the years 2008 (first row from top to bottom) and 2009 (second row) given by the Seimogenic Index (SI), the Hydromechanical Nucleation model (Hydro) and the observed seismicity (MLM fitting). These maps show the 1% probability to reach or exceed a given ground motion in 1 year. The peak ground acceleration values (Value at the center of the figure) represents the maximum predicted acceleration. The red square represents the seismic source area.

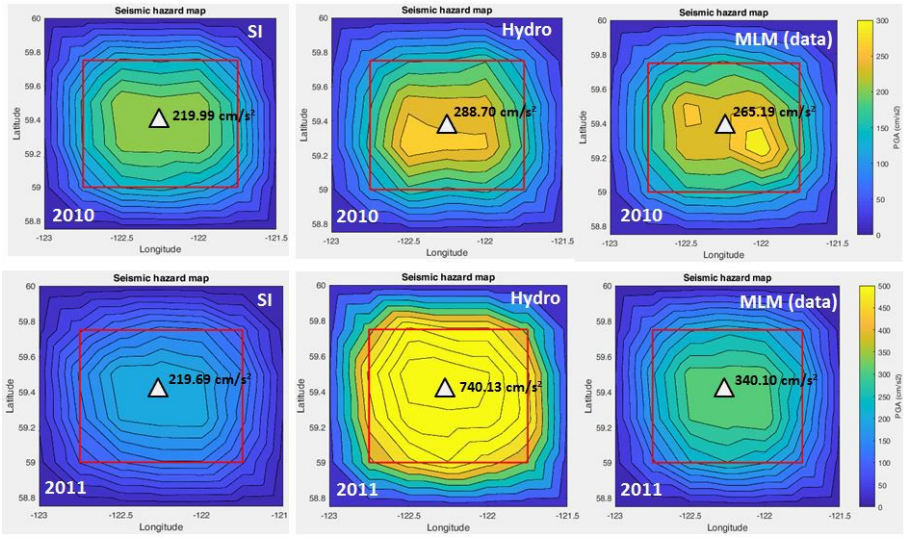

Figure 4.9: Seismic hazard maps for the years 2010 (first row from top to bottom) and 2011 (second row) given by the Seimogenic Index, the Hydromechanical Nucleation model (Hydro), and the observed seismicity (MLM fitting). These maps show the 1% probability to reach or exceed a given ground motion in 1 year. The red square represents the seismic source area.

Figure 4.7 shows the Seismic hazard curves (Rate of exceedance as a function of ground motions) for the years (A) 2008, (B) 2009, (C) 2010, and (D) 2011 given by the Seismogenic Index and the Hydromechanical Nucleation predictions, as well the actual observed seismicity (MLM fitting). The likelihood of 1% in 1 year (equation 4.30) is indicated in the plot. The likelihood of 1% in 1 year was used by Petersen et al. (2016), Petersen et al. (2017), and Petersen et al. (2018) in their short-term seismic hazard forecast, including induced seismicity in central and eastern USA. Like different GR parameters lead to different seismicity rates, we also expect differences in the ground motion predictions given by the predictive models and the observed seismicity. For instance, for the year 2008, both models overestimate the hazard. For 2008 there is a 1% probability in 1 year to reach a peak ground acceleration (PGA)) of 72.41 cm/s^2, 47.56 cm/s^2, and 24.20 cm/s^2 according to the Seismogenic Index, Hydromechanical Nucleation model and observed seismicity models, respectively. On the other hand, for the year 2011, the Hydromechanical Nucleation model overestimates the hazard, and the Seismogenic Index underestimates it. For the year 2011, there is a 1% probability in 1 year to reach a peak ground acceleration (PGA)) of 219.69 cm/s^2, 740.13 cm/s^2, and 340.10 cm/s^2 according to the Seismogenic Index, Hydromechanical Nucleation model and observed seismicity models, respectively.

To build the seismic hazard maps, we calculate the seismic hazard curves at different locations and then select the ground motions from a common rate of exceedance. Figures 4.8 and 4.9 show the Seismic hazard maps for the years 2008, 2009, 2010 and 2011 at the Horn River Basin given by the Seimogenic Index, the Hydromechanical approach and the observed seismicity (MLM fitting) models. These maps show the 1% probability to reach or exceed a given ground motion in 1 year. As mentioned before, having biased GR predictions lead to biased synthetic earthquake catalogs and ground motion predictions. For instance, for the year 2010, the Hydromechanical Nucleation model overestimates the hazard, and the Seismogenic Index underestimates it. In 2010, there is a 1% probability in 1 year to reach a peak ground acceleration (PGA) of 219.99 cm/s^2, 288.70 cm/s^2, and 265.19 cm/s^2 at the central location of the source area, according to the Seismogenic Index, the Hydromechanical Nu-

cleation model, and the observed seismicity, respectively. On the other hand, for the year 2009, both models underestimate the hazard. In 2009, there is a 1% probability in 1 year to reach a peak ground acceleration (PGA) of 43.9 cm/s^2, 53.04 cm/s^2, and 68.7 cm/s^2 at the central location of the source area, according to the Seismogenic Index, the Hydromechanical Nucleation model, and the observed seismicity, respectively.

4.5 Discussion

For the first test, both methods predict earthquake rates similar to the observed induced seismicity in the Horn River basin (number of earthquakes exceeding a certain magnitude per month). They are able to describe the changes in the overall seismicity patterns, in line with the anthropogenic activities. In the case of the Seismogenic Index, this model predicted 92.24 earthquakes $M \geq 2.5$ between Dec. 2006 and Dec 2011. On the other hand, the Hydromechanical Nucleation approach predicted 189 earthquakes $M \geq 2.5$ in the same period of time. In contrast, 121 earthquakes with $M \geq 2.5$ were recorded in the Horn River basin. For the first test, the correlation coefficient between observed and forecasted event rates was slightly higher for the Hydromechanical Nucleation (0.56) approach versus the Seismogenic Index (0.51).

For the second test, both models make inaccurate predictions for specific magnitude ranges (annual magnitude frequency distributions), resulting in an under- or overestimation of the hazard. For both models, predictions can be incorrect (under- or overestimation), and there is no consistency between both predictions, despite usage of the same input data. For instance, in the case of the Seismogenic Index, the most likely number of events in the range $m = [2.5, 3.6]$ for the year 2011 was 38, which underestimates the 60 events given by the observed GR parameters. On the other hand, the most likely number of events in the range $m = [2.5, 3.6]$ for the year 2009 was 21, which overestimates the 8 earthquakes given by the observed parameters. The same situation is observed in the predictions given by the Hydromechanical Nucleation approach. The most likely number of events in the range $m = [2.5, 3.6]$ for the year 2011 was 118, overestimating the 60 events given by the observed GR

parameters. On the other hand, the most likely number of events in the range $m = [2.5, 3.6]$ predicted for the year 2010 was 2, and the estimated number of earthquakes given by the observed parameters was 8. Both methods predict monthly rates with reasonable accuracy (test 1), but they struggle when forecasting magnitude-frequency distributions. The latter are the foundations for the largest-magnitude seismicity rates, having a significant impact on the PSHA estimations. Clearly, this study is a single evaluation of both predictive methods, and more case histories are needed to verify their robustness and accuracy.

One of the reasons why the predicted magnitude-frequency distributions are less precise is the assumption of constant b-values. A potential approach to include temporal variations in b-values due to industrial activities is to explore observed correlations between b-values and changes in stress. For instance, some studies (Scholz, 2015; Amitrano, 2003; Goebel et al., 2013) have found that the b-values are anticorrelated to the differential stress $\sigma_1 - \sigma_3$, that is, the difference between the maximum σ_1 and minimum σ_1 principal stress. Such systematic correlations have also been found in numerical studies (Amitrano, 2003; Van der Baan and Chorney, 2019). Furthermore, Van der Baan and Chorney (2019) also observed a decrease in b-value when approaching peak stress during a simulated triaxial deformation test. This pattern is also seen in laboratory experiments (Meredith et al., 1990). A potential avenue to predict changes in b-values due to industrial activities involving fluid injection or depletion is thus to combine the Seimogenic Index, and Hydromechanical Nucleation approaches with geomechanical models to compute changes in differential and absolute stress due to volumetric changes and the anticipated change in b-values. Some potential geomechanical models are described in Segall (1989) and Segall and Lu (2015). Furthermore, the geomechanical modeling performed by Segall (1989) found changes in the b-values between injection and postinjection stages.

A decrease in b-values is indicative of progressive internal damage, the prevalence of shear events, and the activation of faults that could cause large events. On the other hand, large b-values indicate tensile failure, opening, and closing of fractures characterized by a relatively large number of small-magnitude

events (Van der Baan and Chorney, 2019). For long-term hazard analysis, minor, short-lived changes in the b-value might not be that relevant because these values are averaged out over long time frames (e.g., 50 years). However, for short-term seismic hazard forecast, these changes in the b-value become more relevant.

4.6 Conclusion

We apply two methodologies, the Seismogenic Index and the Hydromechanical Nucleation model, in order to forecast the magnitude-frequency parameters related to fluid-injection seismicity at the Horn River Basin. For the first test, we compare the predicted number of earthquakes exceeding a certain magnitude per month with the observed number of earthquakes. Both methods predict earthquake rates similar to the observed induced seismicity in the Horn River basin. They are able to describe the changes in the overall seismicity patterns within a reasonable level. For the second test, both models make inaccurate predictions for specific magnitude ranges (annual magnitude frequency distributions), resulting in an under- or overestimation of the hazard but often with contradicting forecasts, despite using shared observations. The predictions tend to either underestimate or overestimate the seismic hazard, mainly due to the assumption of stationary b-values. Even if the predictions of a-values are reasonable, changes in the b-value can considerably change the annual prediction of magnitude-frequency distributions, which is one of the most critical factors to precisely determine the seismic hazard analysis. As expected, having biased magnitude-frequency distributions leads to biased synthetic earthquake catalogs and ground motion predictions. Previous studies have shown that the b-values are anticorrelated to differential stress. Furthermore, b-values show a decrease when approaching peak stress, reaching the lowest b-value when a mainshock occurs. Therefore, future physics-based models should be able to model changes in the b-values, based on operational and Geomechanical properties.

Chapter 5

Seismicity declustering and implications for seismic hazard analysis [1]

5.1 Introduction

Seismicity declustering is a widely used method in seismology, in particular for seismic hazard assessment and in earthquake prediction model (Van Stiphout et al., 2012; Petersen et al., 2016, 2017, 2018). It consists of classifying the events from an earthquake catalog into foreshock, mainshock and aftershock sequences (Van Stiphout et al., 2012). Some of the most common declustering methods (Gardner and Knopoff, 1974; Reasenberg, 1985) are based on the identification of mainshocks considering their magnitude and spatio-temporal distribution.

Seismicity declustering was initially used to shown that the mainshock sequences follow a temporal stationary Poisson distribution (Gardner and Knopoff, 1974; Marzocchi and Taroni, 2014). In other words, mainshock events are randomly distributed in time. Declustering thus produces a catalog of mainshocks that follow a Poisson distribution, which in turn is the underlying distribution

[1] Paper submitted to *Journal of Geophysical Research*.

assumed by most Probabilistic Seismic Hazard (PSHA) approaches (Cornell, 1968; Baker, 2008). Currently, the main rationale behind seismicity declustering is to differentiate between two groups of seismic sequences with clearly different characteristics. On the one hand, the background or mainshock events are a consequence of tectonic loading, and they are considered temporally independent. On the other hand, aftershocks are consequence of a parent event, either a mainshock or another triggered event, and they are temporally dependent on the precursor event (Van Stiphout et al., 2012).

Time-dependency might not be the only difference between mainshock and aftershock sequences. Some studies suggest different b-values between mainshock (declustered) and complete (non-declustered) earthquake catalogs (Utsu, 1966; Suyehiro and Sekiya, 1972). In contrast, other authors suggest that the observed difference is not statistically significant (Knopoff et al., 1982), it is simply a statistical artifact (Lombardi, 2003), or it might depend on the performed declustering method (Teng and Baker, 2019). Furthermore, some hazard studies do not apply seismicity declustering as it might reduce the hazard by excluding a large number of small-to-medium magnitude-size earthquakes (Atkinson et al., 2015; Langenbruch and Zoback, 2016; Ghofrani et al., 2019).

If mainshocks and aftershocks have different b-values, then this is likely to impact the hazard assessment, because generally a single b-value is assumed within a source region. Hazard predictions often most heavily depend on the magnitude-frequency distribution of the largest earthquakes, since these account for the largest ground motion. Once a b-value is determined, it is used to extrapolate occurrence rates to earthquake magnitudes that may not be well represented within the observed catalog. As a consequence, if aftershocks and mainshocks have different b-values, then using an averaged b-value may lead to biased hazard assessments. It is important to mention that different declustering methods might lead to different b-values for the mainshock sequence, as shown in the study implemented by Teng and Baker (2019) in Oklahoma. However, the purpose of this study is not to investigate which declustering method is most appropriate, but rather to show when declustering is recommended for the correct estimation of recurrence statistics for large events. Our final objective is to investigate if separating mainshocks and aftershocks leads to

more accurate hazard assessments. It would ultimately lead to more accurate estimates of the b-value for the largest magnitude events rendering a triage into Poissonian earthquake sequences of secondary importance.

We use synthetic earthquake catalogs based on the Epidemic-Type Aftershock Sequence (ETAS) methodology to determine the hazard implications of temporally declustering an earthquake catalog. From these synthetic catalogs, we obtain declustered (mainshock) and non-declustered (complete) earthquake catalogs, and study the impact of excluding aftershocks in the hazard analysis. We also allow a difference in b-values between the background and aftershock sequences in order to replicate more accurately the observed seismicity in some induced seismicity examples. We set the basis of when it is appropriate to temporally declustering an earthquake catalog based on the hazard estimations and we show that the Poisson statistics are still useful for the seismic hazard analysis, even if aftershock sequences are not removed. We apply the analysis strategy to a recent case of induced seismicity: Oklahoma, USA. We compare the hazard estimations given by declustered and non-declustered catalogs using both observed and simulated earthquake sequences. Finally, we end with general recommendations on the pros and cons of declustering for hazard assessment.

5.2 Theory

5.2.1 Simulation of earthquake catalogs following the Epidemic-Type Aftershock Sequence (ETAS) methodology

The Epidemic-Type Aftershock Sequence (ETAS) model (Ogata, 1988, 1998; Utsu et al., 1995; Zhuang et al., 2012), sometimes referred to as Epistemic-Type Aftershock Sequence, is a self-exciting point process in which every event can produce offspring of events. The model includes the background activity and the aftershock events produced by a parent event (either a background or an aftershock event). The rate of occurrence $\lambda(t)$ at time t is given by (Ogata,

1988, 1998; Utsu et al., 1995):

$$\lambda(t) = \mu + \sum_{t_i<t} g(t - t_i), \tag{5.1}$$

where μ is the background constant seismicity rate, which is assumed to follow a temporal Poisson distribution. On the other hand, $g(t - t_i)$ is the rate of activity at time t triggered by an event m_i at time t_i, and is given by:

$$g(t - t_i) = K \frac{e^{\alpha(m_i - M_{min})}}{(t - t_i + c)^p}, \tag{5.2}$$

where M_{min} is the minimum magnitude considered. Only events larger or equal than M_{min} are included in the data set. The parameters K, α, c and p are constants common to all aftershock sequences. $Ke^{\alpha(m_i - M_{min})}$ is referred to as the expected number of events triggered from an event of magnitude m_i, where K and α are constants (Zhuang et al., 2002). The parameters c and p are constants that describe the rate of decay given by the modified Omori's law (Omori, 1894; Utsu, 1961). The second term in eq. 5.1 represents the addition of all the rates of earthquakes from the aftershock activity, at a given time t.

There are two main methodologies to simulate an ETAS process: the thinning method and the branching process method (Zhuang and Touati, 2015). For the purpose of this study, we use the branching process method to simulate the ETAS model. This method consists of two main steps: (1) generation of the synthetic earthquake catalogs for the background seismicity, assuming a Poisson process (μ, eq. 5.1); (2) generation of the aftershock catalog, where each background event results in a cascade of directly-triggered aftershocks, aftershocks of those aftershocks, etc.

Simulation of the stationary background process

For the background seismicity rates, we will use expressions that are functions of the Gutenberg-Richter (GR) parameters, a- and b-values (Gutenberg and Richter, 1944). To obtain the background rate of earthquakes $\mu(M_{min} \leq m < M_{max})$ in the range $m = [M_{min}, M_{max})$, we multiply the probability of occurrence $P(M_{min} \leq m < M_{max})$, by the total expected number of earthquakes

$N(M_{min} \leq m \leq M_{max})$, resulting in (Reyes Canales and Van der Baan, 2019):

$$\mu(M_{min} \leq m < M_{max}) = P(M_{min} \leq m < M_{max})N(M_{min} \leq m \leq M_{max}), \tag{5.3}$$

where M_{max} is the maximum expected magnitude. The total expected number of earthquakes $N(M_{min} \leq m \leq M_{max})$ is given by:

$$N(M_{min} \leq m \leq M_{max}) = 10^{a_b - b_b M_{min}} - 10^{a_b - b_b M_{max}}. \tag{5.4}$$

In this equation, a_b and b_b are the parameters defined by the GR distribution (Gutenberg and Richter, 1944) for the background seismicity. Finally, the probability of occurrence $P(M_{min} \leq m < M_{max})$ is given by:

$$P(M_{min} \leq m < M_{max}) = F_M(M_{min}) - F_M(M_{max}), \tag{5.5}$$

where $F_M(m)$ denotes the cumulative distribution function for magnitude m, and is defined by:

$$F_M(m) = \frac{1 - 10^{-b_b(m - M_{min})}}{1 - 10^{-b_b(M_{max} - M_{min})}}. \tag{5.6}$$

Once the rate of earthquakes for the background seismicity μ has been defined, we proceed to simulate the synthetic earthquake catalogs by using the Monte-Carlo simulation methodology described by Reyes Canales and Van der Baan (2019).

Algorithm 1: Simulation of event times for the background seismicity, assuming a stationary Poisson process with rate $\mu(M_{min} \leq m < M_{max})$ (eq. 5.3) between times t_{start} and t_{end} (Zhuang and Touati, 2015; Reyes Canales and Van der Baan, 2019):

1. Set $t = t_{start}$, $P_p = 0$.
2. Generate r.
3. $t = t + \frac{-ln(r)}{\mu(M_{min} \leq m < M_{max})}$. If $t \geq t_{end}$, then stop.
4. Set $P_p = P_p + 1$ and set $t_P = t$.

5. Go to step 2.

The variable r is a random number obtained from a uniform distribution between [0,1], t_P is the vector that contains the desired arrival times and P_p the number of arrival times. Once we have generated the P_p arrival times in the time period $[t_{start}, t_{end}]$, we sample the GR distributions. 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):

$$m = \frac{-\ln(r)}{b_b \ln(10)} + M_{min}. \tag{5.7}$$

During the sampling, any $m > M_{max}$ is excluded (Truncated GR distribution), and the sampling is repeated until we get P values with $m \leq M_{max}$. Finally, a synthetic earthquake catalog for the background events $B = [(t_p, m_p) : p = 1, 2, ..., P_p]$ has been generated.

Simulation of the triggered events and branching process algorithm

There are multiple expressions to describe the rate of earthquakes triggered from a parent event, including Ogata's intensity rate (Ogata, 1988, 1998), Zhuang's conditional intensity form (Zhuang et al., 2002) and Harte's normalization rate (Harte, 2013). In this study , we will use the Harte's normalization rate to describe the decaying seismicity rate that follows an earthquake (Harte, 2013):

$$\nu(t) = \frac{K e^{\alpha(m_i - M_{min})}}{\left(\frac{t}{c} + 1\right)^p}, \tag{5.8}$$

where K, α, c and p are the ETAS parameters described previously. In order to simulate the triggered events associated to an event of magnitude m_i, we use a Monte Carlo simulation methodology assuming a non-stationary Poisson process (Zhuang and Touati, 2015; Reyes Canales and Van der Baan, 2019).

Algorithm 2: Simulation of event times for triggered events, assuming a non-stationary Poisson process with rate $\nu(t)$ between times t_0 and t_{max} (Zhuang and Touati, 2015; Reyes Canales and Van der Baan, 2019):

1. Consider ν^* such that $\nu^* \geq max(\nu(t))$.

2. Set $t = t_0$, $P_a = 0$.

3. Generate r.

4. $t = t + \frac{-ln(r)}{\nu^*}$. If $t \geq t_{max}$, then stop.

5. Generate r.

6. If $r \leq \nu(t)/\nu^*$, then set $P_a = P_a + 1$ and set $t_P = t$.

7. Go back to step 3.

In this case, r is a random number obtained from a uniform distribution between [0,1], t_0 is assumed to be the time for the parent event m_i, and t_{max} the maximum time for the aftershock generation. Again, for the generation of magnitudes m, we apply Monte Carlo sampling to the inverse cumulative distribution function (CDF) of the GR distribution (eq. 5.7), considering that the sampling is done with the b_a-value of the aftershock sequence.

In order to simulate a full ETAS branching process, we generate first the background catalog, as shown in Algorithm 1. Then, we generate the aftershocks of the background earthquakes, followed by the aftershocks of the aftershock, etc... in a loop that finishes when no more events larger than M_{min} are generated.

Algorithm 3: Simulating ETAS as a branching process (Modified from Zhuang and Touati (2015)):

1. Generate the background catalog $B(p) = [(t_p, m_p) : p = 1, 2, ..., P_p]$ as a stationary Poisson process, for a time period $t = [t_{start}, t_{end}]$ and background rate $\mu(M_{min} \leq m < M_{max})$. See Algorithm 1.

2. Set $p = 1$.

3. Define magnitudes $M = m_p$ and time $T = t_p$. If $p > P_p$, stop.

4. Generate the first offspring $G^{(0)} = [(t_i, m_i) : i = 1, 2, ..., L]$ for the background event p, with magnitude $M = m_p$ and time $T = t_p$. The event

time t_i is generated from a non-stationary Poisson simulation, with intensity rate $\nu(t)$, see Algorithm 2. The magnitude m_i is generated by applying Monte Carlo sampling to the inverse cumulative distribution function (CDF) of the GR distribution (eq. 5.7), using b_a.

5. Set $l = 0$.

6. For each event i, namely (t_i, m_i), in the catalog $G^{(l)}$, simulate its ${N_i}^{(l)}$ offspring, that is, ${O_i}^{(l)} = [({t_k}^{(i)}, {m_k}^{(i)}) : i = 1, 2, ..., {N_i}^{(l)}]$. Again, the event time ${t_k}^{(i)}$ is generated from a non-stationary Poisson simulation, with intensity rate $\nu(t)$, see Algorithm 2. The magnitude ${m_k}^{(i)}$ is generated by applying Monte Carlo sampling to the inverse cumulative distribution function (CDF) of the GR distribution (eq. 5.7), using b_a.

7. Set $G^{(l+1)} = \cup_{i \in G^{(l)}} {O_i}^{(l)}$. $\cup$ refers to the union symbol.

8. Remove all events with occurrence times are larger than t_{end} from $G^{(l+1)}$.

9. If $G^{(l+1)}$ is not empty, let $l = l+1$ and go to step 6. Else return $\cup_{j=0^l} G^{(l+1)}$.

10. Make $A^{(p)} = \cup_{j=0^l} G^{(l+1)}$, then make $p = p + 1$ and go back to step 3.

Finally, a complete catalog of events has been generated, with the variable $B(p)$ containing the background earthquakes, and $A^{(p)}$ the aftershock events. The complete catalog C is the combination of both the Background catalog B and the Aftershock catalog $A^{(p)}$. The process can be repeated N_r times in order to generate multiple realizations, which are going to describe the variability of the process. Notice that the simulated ETAS model has no history of events before time t_{start}, which might cause an underestimation in the number of earthquakes at the beginning of the simulation. To prevent this problem, a longer catalog than required can be simulated and then the early part of the simulation is neglected (Zhuang and Touati, 2015).

Notice that the branching process allows the use of different b-values for the background and aftershock sequences. Also, the setting described in Algorithm 3 is particularly useful in order to distinguish the different temporal clusters initiated by background events. In this study we make a distinction between background events and mainshocks. A background event is generated

by the stationary Poisson process, explained in Algorithm 1. A mainshock is the largest event in a temporal cluster, and it could be either the background event or one of the triggered aftershocks. Therefore, there will be a background catalog and a mainshock catalog. More details about these two catalogs are explained in the Implementation section. Notice as well that, by definition, the ETAS model allows triggered events larger than the parent event, but lower than M_{max}. We do not recommend limiting the magnitude of the triggered events, and force them to be lower than the parent event; this could have significant consequences on the global statistics. If all offspring events are limited to lower magnitudes, there will be a population increase of lower magnitude events, resulting in an artificial alteration of the magnitude-frequency distribution.

For simplicity, we assume that the background seismicity is stationary. However, it is possible to apply a non-stationary Poisson distribution, with time-dependent GR parameters for the background seismicity as explained by Reyes Canales and Van der Baan (2019).

5.2.2 Statistical quantities from synthetic earthquake catalogs

From the synthetic earthquake catalogs, it is possible to obtain relevant statistical quantities for the hazard analysis. The directly observed annual rate of exceedance $\lambda_{exc}(m \geq m_j)$ for a magnitude level, is obtained by counting the number of events with a magnitude larger than m_j and dividing by the time duration of the synthetic earthquake catalog:

$$\lambda_{exc}(m \geq m_j) = \frac{N_{m>m_j}}{\Delta t}, \tag{5.9}$$

where $N_{m>m_j}$ is the number of earthquakes larger than m_j from synthetic earthquake catalogs, and Δt the duration of the catalog, in years. By calculating the annual rate of exceedance for each realization, we will have generated a population with a mean annual rate of exceedance $m_{\lambda,exc}(m \geq m_j)$ for a magnitude

level, given by:

$$m_{\lambda,exc}(m \geq m_j) = \frac{\sum_{i=1}^{N_r} \lambda_{exc}^{(i)}(m \geq m_j)}{N_r}, \tag{5.10}$$

where $\lambda_{exc}^{(i)}(m \geq m_j)$ is the annual rate of exceedance for the i^{th} realization, and N_r the number of realizations. The standard deviation $\sigma_{\lambda,exc}(m \geq m_j)$ for a magnitude level is given by:

$$\sigma_{\lambda,exc}(m \geq m_j) = \sqrt{\frac{1}{1-N_r} \sum_{i=1}^{N_r} [\lambda_{exc}^{(i)}(m \geq m_j) - m_{\lambda,exc}(m \geq m_j)]^2}. \tag{5.11}$$

The standard deviation shows the variability of the annual rate of exceedance $\lambda_{exc}(m \geq m_j)$ for a magnitude level, given a group of synthetic earthquake catalogs. Likewise, it is possible to calculate the GR parameters for each realization, and study the population of $\hat{a}$-and $\hat{b}$-values given by the multiple realizations. The mean $m_{(\hat{a})}$ and $m_{(\hat{b})}$ for the $\hat{a}$ and $\hat{b}$-values, respectively, in a group of synthetic earthquake catalogs, are given by:

$$m_{(\hat{a})} = \frac{\sum_{i=1}^{N_r} \hat{a}^{(i)}}{N_r}, \tag{5.12}$$

$$m_{(\hat{b})} = \frac{\sum_{i=1}^{N_r} \hat{b}^{(i)}}{N_r}, \tag{5.13}$$

where $\hat{a}^{(i)}$ and $\hat{b}^{(i)}$ are the estimated $\hat{a}$ and $\hat{b}$-values for the i^{th} realization. In a similar way, the standard deviation $\sigma_{(\hat{a})}$ and $\sigma_{(\hat{b})}$ for the $\hat{a}$ and $\hat{b}$-values, respectively, are given by:

$$\sigma_{(\hat{a})} = \sqrt{\frac{1}{1-N_r} \sum_{i=1}^{N_r} [\hat{a}^{(i)} - m_{(\hat{a})}]^2}, \tag{5.14}$$

$$\sigma_{(\hat{b})} = \sqrt{\frac{1}{1-N_r} \sum_{i=1}^{N_r} [\hat{b}^{(i)} - m_{(\hat{b})}]^2}. \tag{5.15}$$

From average GR parameters, we can define the annual rate of exceedance $\lambda_{exc,GR}(m \geq m_j)$ for a magnitude level. This is given by the GR distribution

in terms of rate of exceedance (Gutenberg and Richter, 1944; Baker, 2008):

$$\log(\lambda_{exc,GR}(m \geq m_j)) = m_{(\hat{a})} - m_{(\hat{b})} m_j. \tag{5.16}$$

Rearranging, it produces:

$$\lambda_{exc,GR}(m \geq m_j) = 10^{m_{(\hat{a})} - m_{(\hat{b})} m_j}, \tag{5.17}$$

where $m_{(\hat{a})}$ and $m_{(\hat{b})}$ refer to the mean $\hat{a}$ and $\hat{b}$-values obtained from the synthetic catalog, 5.12 and 5.13.

Another useful statistical quantity is the probability $P[N = n; t_a, t_b]$ of n occurrences happening in a time interval $\Delta t = t_b - t_a$, which can be estimated by assuming a stationary Poisson process (Cornell, 1968; Baker, 2008):

$$P[N = n; t_a, t_b] = \frac{\lambda^n (t_b - t_a)^n e^{-\lambda(t_b - t_a)}}{n!}, \tag{5.18}$$

where λ is the rate of occurrence of events per time unit. For instance, λ can be equivalent to the annual rate of earthquakes $\lambda_{GR}(m_j \leq m < m_k)$ in the magnitude range $m = [m_j; m_k)$, from average GR parameters. This is defined by:

$$\lambda_{GR}(m_j \leq m < m_k) = P(m_j \leq m < m_k) N(m_j \leq m \leq m_k). \tag{5.19}$$

The total expected number of earthquakes $N(m_j \leq m \leq m_k)$ in the magnitude range $m = [m_j; m_k)$ is given by:

$$N(m_k \geq m \geq m_j) = 10^{m_{(\hat{a})} - m_{(\hat{b})} m_j} - 10^{m_{(\hat{a})} - m_{(\hat{b})} m_k}. \tag{5.20}$$

On the other hand, the probability of occurrence $P(m_j \leq m < m_k)$ is given by:

$$P(m_j \leq m < m_k) = F_M(m_j) - F_M(m_k), \tag{5.21}$$

where $F_M(m)$ denotes the cumulative distribution function, eq. 5.6. Finally, by using eq. 5.19 in eq. 5.18, we can estimate the probable number of earthquakes

in the magnitude range $m = [m_j; m_k)$ from the mean GR parameters given by the multiple realizations.

5.3 Implementation

In order to simulate the ETAS process described in the previous section, it is necessary to define the GR parameters for the background seismicity (a_b-and b_b-values), the ETAS parameters for the aftershock simulation (K, c, p, α), and the b_a-value for the aftershock sequence. The GR parameters can be calculated from observed mainshock and aftershock catalogs using Maximum Likelihood estimations (Aki, 1965; Wiemer and Wyss, 1997). The ETAS parameters can also be calculated from complete earthquake catalogs using Maximum Likelihood estimations (Ogata, 1998; Daley and Vere-Jones, 2003).

The first step of the ETAS process following the branching method consists in simulating the background sequence, see algorithm 1. Then, the aftershock catalog is generated by simulating all associated aftershocks triggered by a background event, see algorithms 2 and 3. We assume that the background events comprise almost all mainshocks. However, we can also generate a mainshock catalog by selecting the largest magnitude in a cluster of events triggered by a background event. Most of the time, the background event is the largest event in the cluster, but depending on the productivity of the sequence, the largest event (mainshock) can be one of the triggered aftershocks. Therefore, mainshock and background catalogs are different. We quantify this difference for some reasonable sets of ETAS parameters.

The ETAS simulation process can be repeated N_r times, generating multiple realizations that describe the variability of the process. Figure 5.1 shows a sketch of the branching process for the simulation of synthetic earthquake catalogs. From these catalogs, we can calculate statistical quantities that are relevant for the study of the hazard analysis: (1) the mean annual rate of exceedance $m_{\lambda,exc}(m \geq m_j)$ for a magnitude level (eq. 5.10); (2) the annual rate of exceedance $\lambda_{exc,GR}(m \geq m_j)$ (eq. 5.17) from mean $\hat{a}$-and $\hat{b}$-values (eq. 5.12 and 5.13); and (3) the probability $P[N = n; t_a, t_b]$ of n occurrences happening in a time interval $\Delta t = t_b - t_a$ (eq. 5.18), also from mean $\hat{a}$-and $\hat{b}$-values.

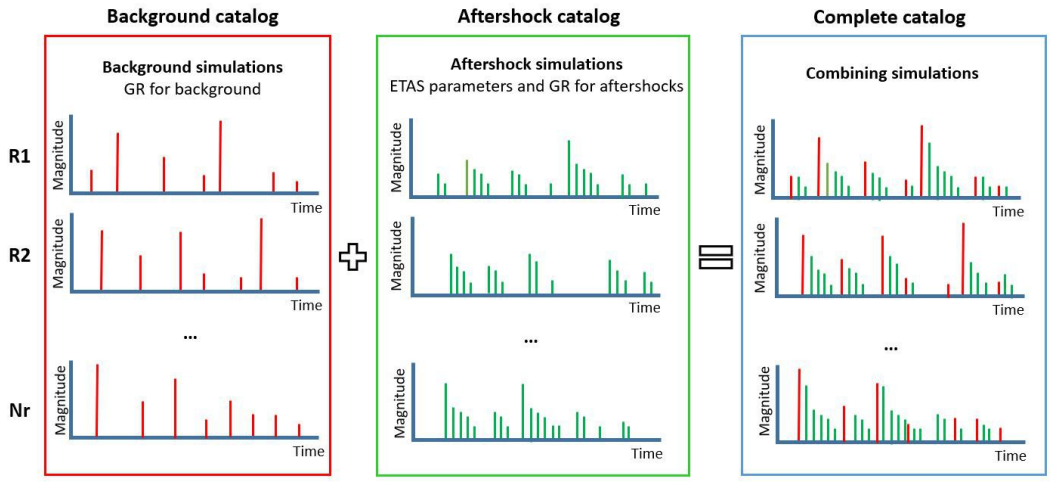

Figure 5.1: Sketch of the branching process used to simulate synthetic earthquakes, following the ETAS methodology. The background catalog (Left) is generated assuming a Poisson process. For each background earthquake, a sequence of aftershocks (Center)) are generated, following the Epidemic-type process. The complete catalog contains all events (Right). The x-axis (time) is not in real scale.

The directly observed annual rate of exceedance $\lambda_{exc}(m \geq m_j)$ for a magnitude level (Eq.5.9), is estimated by first counting the number of events with magnitude larger than m_j in one realization, and dividing by the time duration of the catalog. Repeating for all simulations and averaging, it results in the mean annual rate of exceedance $m_{\lambda,exc}(m \geq m_j)$ for a magnitude level, eq.5.10. It is directly observed from the simulations to account for the fact that the aftershocks may lead to nonlinear magnitude-frequency distributions. To count for the variability, we add and subtract the standard deviation to the mean annual rate of exceedance, $m_{\lambda,exc}(m \geq m_j) \pm \sigma_{\lambda,exc}$, eq. 5.11. This results in a upper and lower curve. Similarly, we can apply this process to the complete and mainshock catalogs, resulting in the mean annual rate of exceedance, $m_{\lambda,exc,c}(m \geq m_j)$ and $m_{\lambda,exc,m}(m \geq m_j)$ for the complete and mainschock catalogs, respectively. These represent the annual exceedances given by the parameters of the underlying generating processes.

In addition, we compute an annual rate of exceedance using estimated GR parameters, thereby assuming a linear magnitude-frequency distribution as usually done in PSHA. The annual rate of exceedance $\lambda_{exc,GR}(m \geq m_j)$

from mean $\hat{a}$-and $\hat{b}$-values, is obtained by estimating first the $\hat{a}$- and $\hat{b}$-values for individual realizations using Maximum Likelihood estimations (Aki, 1965; Wiemer and Wyss, 1997). By averaging, it results in the mean GR parameters $m_{(\hat{a})}$, $m_{(\hat{b})}$, eq. 5.12 and 5.13. These mean GR parameters $m_{(\hat{a})}$, $m_{(\hat{a})}$ are introduced into eq.5.16 and eq. 5.17 resulting in the rate of exceedance $\lambda_{exc,GR}(m \geq m_j)$ given estimated GR parameters. To count for variability, we can define upper and lower values for mean GR parameters, which result from adding and substracting the standard deviation, $\sigma_{(\hat{a})}$ and $\sigma_{(\hat{b})}$ (Eq. 5.14 and 5.15) to the mean $\hat{a}$- and $\hat{b}$-values, resulting in $m_{(\hat{a})} \pm \sigma_{\hat{a}}$, and $m_{(\hat{b})} \pm \sigma_{\hat{b}}$. From these upper and lower values we proceed to calculate the upper and lower annual rate of exceedance curves $\lambda_{exc,GR+\sigma}(m \geq m_j)$ and $\lambda_{exc,GR-\sigma}(m \geq m_j)$. Similarly, we can define the annual rate of exceedance $\lambda_{exc,GR,c}(m \geq m_j)$ and $\lambda_{exc,GR,m}(m \geq m_j)$, eq.5.16, for complete and mainshock catalogs, respectively. In this case we require mean GR parameters for complete catalogs, $m_{(\hat{a}_c)}$ and $m_{(\hat{b}_c)}$, and mainshock catalogs, $m_{(\hat{a}_m)}$ and $m_{(\hat{b}_m)}$ (Eq. 5.12 and 5.13). On the other hand, we can define $\sigma_{(\hat{a}_c)}$ and $\sigma_{(\hat{b}_c)}$, as well as $\sigma_{(\hat{a}_m)}$ and $\sigma_{(\hat{b}_c)}$, the standard deviations of estimated GR parameters for mainshock and complete catalogs, respectively.

To study the hazard estimation of large events given by the estimated GR parameters from both complete and mainshock catalogs, we calculate the probability $P[N = n; t_a, t_b]$ of n occurrences in the magnitude range $m = [m_j; m_k)$, happening in a time interval $\Delta t = t_b - t_a$, eq. 5.18. This is done by introducing the annual rate of earthquakes $\lambda_{GR,c}(m_j \leq m < m_k)$ and $\lambda_{GR,m}(m_j \leq m < m_k)$, for complete and mainshock catalogs, respectively, in eq. 5.18. To estimate the annual rate of earthquakes $\lambda_{GR}(m_j \leq m < m_k)$ in the magnitude range $m = [m_j; m_k)$, eq. 5.19, we use the mean GR parameters for complete ($m_{(\hat{a}_c)}$, $m_{(\hat{b}_c)}$) and mainshock ($m_{(\hat{a}_m)}$, $m_{(\hat{b}_m)}$) catalogs, obtained from simulations. We compare the predictions given by the mean GR from the complete and mainshock catalogs with the actual number of events in the synthetic catalogs. We can also obtain the probability of n number of events in a given magnitude range, in a period of time, directly from synthetic earthquake catalogs. This is obtained by counting the number of realizations with n number of events in a magnitude range, and then dividing by the number of realizations.

The rationale for comparing averaged predictions using estimated GR parameters with those computed directly from the generated catalog is to test if the assumption of a single linear magnitude-frequency distribution is appropriate given the different types of catalogs, that is, mainshocks only or mainshocks plus aftershocks. We use averaged predictions to account for sampling variances, that is, each realization may slightly deviate from the underlying statistical process, in particular for shorter sample populations.

5.4 Synthetic Examples

In order to study the impact of including or excluding non-stationary aftershocks in the seismic hazard analysis, we design two examples based on simulation of synthetic earthquake catalogs. For the first example, we identify the cases when declustering is appropriate, considering long and short-term seismicity durations. For the second example, we demonstrate that the recurrence of earthquakes can be estimated properly, irrespective if we have a Poissonian or a non-Poissonian process, as long as we have the appropriate GR parameters. Therefore, Poisson statistics are still valid, particularly for long-term seismicity.

For the **first example**, we consider two sets of simulations generated following the ETAS branching process described previously. The first set consists of 1,000 realizations of 50-year duration, representing a long-term seismicity case. The second set consists of 1,000 realizations of 5-year duration, representing a short-term seismicity case. For each set, we consider two cases where the b_b-value of the background seismicity is (1) equal or (2) smaller than the b_a-value of the aftershocks. We do not consider the case that the aftershocks have smaller b-values than the background seismicity since this leads to a large number of aftershocks with larger magnitudes than the background events. This breaks the assumption that the background sequence contains most of the mainshock events and that the mainshocks have the largest magnitudes in declustering approaches.

The ETAS parameters for these synthetic examples are defined as follows: $K = 1.5$, $c = 0.1$, $p = 1.3$, $\alpha = 1.5$. These values are loosely based on typical

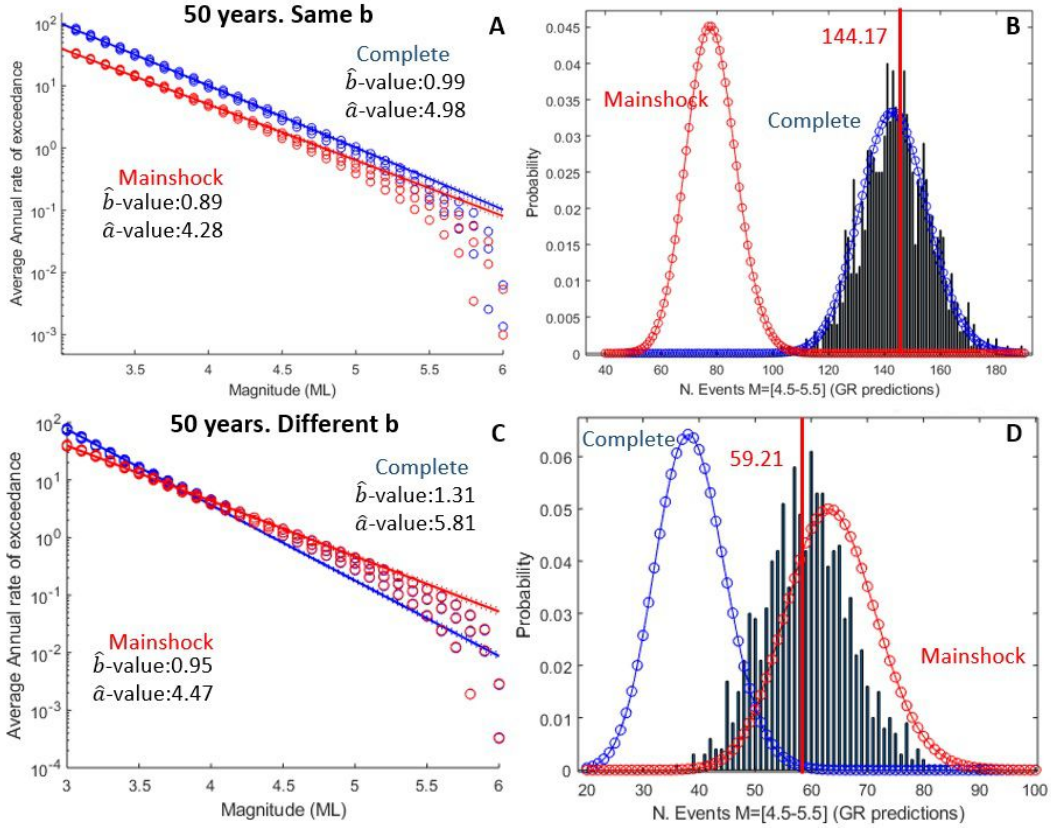

Figure 5.2: Results for first set of simulations (50 years), example 1. First row: $b_b = b_a = 1$, second row: $b_b = 1$ and $b_a = 2$. Figures (A) and (C) show the mean annual rate of exceedance ($m_{\lambda,exc}(m \geq m_j)$, dotted curves) directly extracted from synthetic earthquake catalogs and the annual rate of exceedance from estimated mean GR parameters ($\lambda_{exc,GR}(m \geq m_j)$, continuous curves), for mainshock (red) and complete (blue) catalogs. The upper and lower curves are the mean annual rates $\pm$ standard deviation. Figures (B) and (D) show the probability of n occurrences, $P[N = n; t_a, t_b]$ of n, in the magnitude range $m = [4.5; 5.5)$, in 50 years, for mainshocks (red) and complete (blue) catalogs. The blue bars show the probability of n occurrences obtained directly from the complete catalogs. The vertical red lines show the average number of events in the magnitude range $m = [4.5; 5.5)$ from the synthetic complete catalogs.

ranges found for ETAS parameters, which vary from 0.003 to 0.3 days for c, 0.9 to 1.4 for p, and 0.2 to 3.0 for α (Utsu et al., 1995). A value of $K = 1.5$ is selected in order to design an experiment with medium to high productivity of aftershocks. For the case with identical b-values, we set $b_b = b_a = 1$. For the case with different b-values, we set $b_b = 1$ and $b_a = 2$. Finally, for the background seismicity, we set $N_0 = 40,000$, which is equivalent to $a_b = 4.60$. The maximum time for the aftershock generation t_{max}, see algorithm 2, is set as $t_{max} = 200$ days for all parent events. Finally, the minimum and maximum magnitudes to simulate are: $M_{min} = 3.0$ and $M_{max} = 6.0$. Once these parameters are defined, we proceed to simulate the background and aftershock sequences. Then, we derive the complete and mainshock sequences for each case.

The central dotted curves in figures 5.2(a), 5.2(c), 5.3(a) and 5.3(c) represent the directly observed mean annual rate of exceedance $m_{\lambda,exc,c}(m \geq m_j)$ and $m_{\lambda,exc,m}(m \geq m_j)$ for the complete and mainshock catalogs, respectively, estimated by counting events with magnitudes $m \geq m_j$ in the simulated samples (Eq. 5.10). The upper and lower dotted curves are the mean annual rate of exceedance plus or minus the standard deviation, $m_{\lambda,exc}(m \geq m_j) \pm \sigma_{\lambda,exc}$. The central continuous curves in figures 5.2(a), 5.2(c), 5.3(a) and 5.3(c) represent the annual rate of exceedance from mean GR parameters, $\lambda_{exc,GR,c}(m \geq m_j)$ and $\lambda_{exc,GR,m}(m \geq m_j)$, for complete and mainshock catalogs, respectively, estimated from simulated samples (Eq. 5.17). The upper and lower dashed curves represent the upper and lower annual rate of exceedance curves $\lambda_{exc,GR+\sigma}(m \geq m_j)$ and $\lambda_{exc,GR-\sigma}(m \geq m_j)$, obtained from mean GR parameters $\pm$ standard deviations, $\sigma_{(\hat{a})}$ and $\sigma_{(\hat{b})}$.

Figures 5.2 (A) and (B) show the results for the 50-year seismicity simulations when $b_b = b_a = 1$, and figures 5.2 (C) and (D) show the results for the 50-year seismicity simulations when $b_b = 1$ and $b_a = 2$. The mean $m_{(\hat{a})}$ and $m_{(\hat{b})}$, estimated by averaging $\hat{a}$-and $\hat{b}$-values over all realizations, are summarized in table 5.1, together with standard deviations $\sigma_{(\hat{a})}$ and $\sigma_{(\hat{b})}$.

Figures 5.2 (B) and (D) show the probability of n occurrences in the magnitude range $m = [4.5; 5.5)$ in 50 years, for mainshock (red) and complete (blue) catalogs. This is done by introducing the annual rate of earthquakes

$\lambda_{GR,c}(m_j \leq m < m_k)$ and $\lambda_{GR,m}(m_j \leq m < m_k)$, for complete and mainshock catalogs, respectively, in eq. 5.18. Again, to estimate the annual rate of earthquakes $\lambda_{GR}(m_j \leq m < m_k)$ in a magnitude range $m = [m_j; m_k)$, eq. 5.19, we use the mean GR parameters for complete ($m_{(\hat{a}_c)}$, $m_{(\hat{b}_c)}$) and mainshock ($m_{(\hat{a}_m)}$, $m_{(\hat{b}_m)}$) catalogs. The blue bars in both figures represent the probability of n number of events in the magnitude range $m = [4.5; 5.5)$, in a 50-year period, obtained directly from complete synthetic catalogs. This is obtained by counting the number of realizations with n number of events in the magnitude range $m = [4.5; 5.5)$, and then dividing by the number of realizations. When $b_b = b_a = 1$, the most likely number of events is $n = 78$, given the mean GR parameters of the mainshock catalog, and $n = 142$, given the mean GR parameters of the complete catalog. From the simulations, the average number of events in the magnitude range $m = [4.5; 5.5)$ is 144.17 (1441660 events within the magnitude range $m = [4.5; 5.5)$ in 1000 realizations, see red vertical line). When $b_b = 1$ and $b_a = 2$, the most likely number of event is $n = 62$, given the mean GR parameters of the mainshock catalog, and $n = 38$, given the mean GR parameters of the complete catalog. From the simulations, the average number of events in the magnitude range $m = [4.5; 5.5)$ is 59.21 (592090 events within the magnitude range $m = [4.5; 5.5)$ in 1000 realizations). Notice, that the GR parameters obtained from the mainshock simulations tend to slightly overpredict the number of earthquakes within the magnitude range $m = [4.5; 5.5)$. This results from the use of non-truncated GR curves, which contrasts with the established truncation at $M_{max} = 6.0$ for the synthetic catalogs.

Similarly, figures 5.3 (A) and (B) show the results for the 5-year seismicity simulations when $b_b = b_a = 1$, and figures 5.3 (C) and (D) show the results for the 5-year seismicity simulations when $b_b = 1$ and $b_a = 2$. When $b_b = b_a = 1$, the most likely number of events is $n = 7$, given the mean GR parameters of the mainshock catalog, and $n = 12$, given the mean GR parameters of the complete catalog. From the simulations, the average number of events in the magnitude range $m = [4.5; 5.5)$ is 13.41 (134070 events within the magnitude range $m = [4.5; 5.5)$ in 1000 realizations). When $b_b = 1$ and $b_a = 2$, the most likely number of events is $n = 6$, given the mean GR parameters of the mainshock catalog, and $n = 3$, given the mean GR parameters of the

Table 5.1: Summary of the mean GR parameters ($m_{(\hat{a})}$ and $m_{(\hat{b})}$) and standard deviations ($\sigma_{(\hat{b})}$ and $\sigma_{(\hat{b})}$) for the two sets of simulations: 1000 synthetic earthquake catalogs in each set, with 5 and 50 years duration for sets 1 and 2, respectively. We show the results from both mainshock and complete simulated catalogs.

ETAS model	5 yr sim $b_b = b_a$	5 yr sim $b_b >> b_a$	50 yr sim $b_b = b_a$	50 yr sim $b_b >> b_a$
$m_{(\hat{a}_c)} \pm \sigma_{(\hat{a}_c)}$ (Complete)	4.94 ± 0.134	5.72 ± 0.179	4.98 ± 0.040	5.81 ± 0.054
$m_{(\hat{b}_c)} \pm \sigma_{(\hat{b}_c)}$ (Complete)	0.99 ± 0.045	1.28 ± 0.058	0.99 ± 0.013	1.31 ± 0.018
$m_{(\hat{a}_m)} \pm \sigma_{(\hat{a}_m)}$ (Mainshock)	4.30 ± 0.177	4.50 ± 0.20	4.28 ± 0.055	4.47 ± 0.054
$m_{(\hat{b}_m)} \pm \sigma_{(\hat{b}_m)}$ (Mainshock)	0.90 ± 0.058	0.96 ± 0.060	0.89 ± 0.017	0.95 ± 0.019

complete catalog. From the simulations, the average number of events in the magnitude range $m = [4.5; 5.5)$ is 6.68 (6677 events within the magnitude range $m = [4.5; 5.5)$ in 1000 realizations).

In the cases where $b_b = b_a = 1$, the annual rate of exceedance curve for mainshock catalogs do not intersect the annual rate of exceedance curve for complete catalogs. It does consistently show lower rates for all considered magnitudes; therefore, declustering is not recommended. On the other hand, when $b_b = 1$ and $b_a = 2$, the curves intersect at magnitude $m = 3.7$, and the annual rate of exceedance curve for mainshock catalogs consistently shows higher values for large magnitude events; therefore, declustering is recommended. These results are confirmed by comparing the probability of large events given by GR parameters from complete and mainshock catalogs with the average number of events in the synthetic catalogs. When $b_b = b_a = 1$, the mean GR parameters from the complete catalogs estimate better the number of events in the magnitude range $m = [4.5; 5.5]$ from the synthetic catalogs. When $b_b = 1$ and $b_a = 2$, the mean GR parameters from mainshock catalogs estimate better the number of events in the magnitude range $m = [4.5; 5.5)$ from the synthetic catalogs. Based on these simulations, when the b-value of the complete catalog is considerably larger than the b-value of the mainshock catalog, declustering is an useful tool to get a better fit of the GR parameters for large events, which

ultimately contribute more to the seismic hazard.

The need for declustering in case of different b-values can also be seen in the magnitude-frequency distributions which display a kink in the complete catalogs if the aftershocks and mainshocks have different b-values (Figures 5.2 (C) and 5.3 (C)) but is essentially linear if both b-values are identical (Figures 5.2 (A) and 5.3 (A)). The deviation from a linear slope at the highest magnitudes occurs because a truncated GR distribution is used.

We also compare the background and mainshock catalogs to determine how frequent, given this particular setting, an aftershock is larger than the background event. For the short-term seismicity and when $b_b = b_a = 1$, we found that in 17.7% of the cases there is an aftershock, from the cluster of offsprings, with a larger magnitude than the parent background event. However, when $b_b = 1$ and $b_a = 2$, we found that this likelihood reduces to 10.2%. For the long-term seismicity, when $b_b = b_a = 1$, there is a 19.5% of probabilities to have an aftershock larger than the background event. When $b_b = 1$ and $b_a = 2$, this probability decreases to 11.6%. We notice that these likelihoods strongly depends on both the aftershock productivity given by the ETAS parameters and the b_a-value of the aftershock sequence. Also, considering that the background and mainshock catalogs are relatively similar (from 80 % up to 95 % of the mainshock events in the catalog are background events), we can assume that the GR of the mainshock catalog is similar to the GR of the background catalog.

One remarkable observation from these results is that the Poisson simulations and statistics are still useful for the seismic hazard analysis of long term seismicity. In cases when declustering is recommended, the mainshock catalog describes better the occurrence rate of large events. Because the mainshocks are generated with a Poissonian process, all Poisson-based analytical statistics such as the annual rate of exceedance, eq.5.9, are applicable. In cases where declustering is not recommended, the complete catalog does not follow a Poisson distribution because of the presence of non-stationary aftershock sequences as modeled by the ETAS process. Nonetheless, analytic expressions based on Poissonian statistics, such as the probability of n occurrences within a certain magnitude range, eq.5.18, still produce representative predictions.

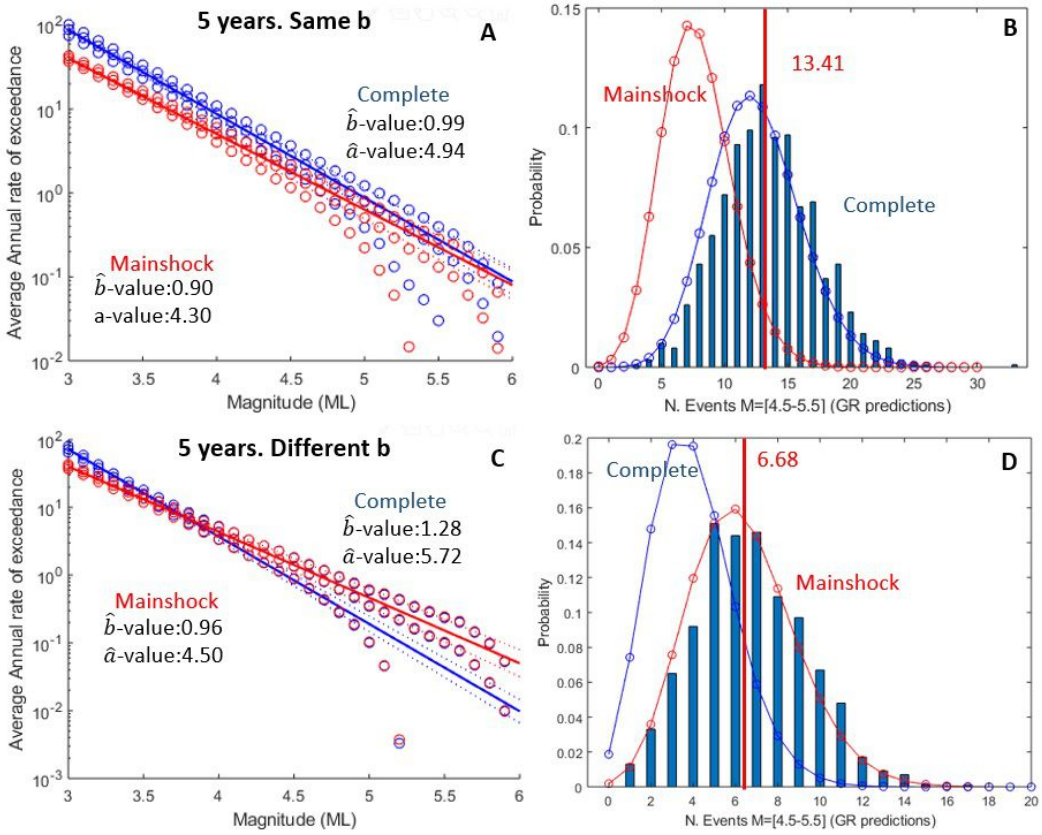

Figure 5.3: Results for second set of simulations (5 years), example 1. First row: $b_b = b_a = 1$, second row: $b_b = 1$ and $b_a = 2$. Figures (A) and (C) show the mean annual rate of exceedance ($m_{\lambda,exc}(m \geq m_j)$, dotted curves) directly extracted from synthetic earthquake catalogs and the annual rate of exceedance from estimated mean GR parameters ($\lambda_{exc,GR}(m \geq m_j)$, continuous curves), for mainshock (red) and complete (blue) catalogs. The upper and lower curves are the mean annual rates $\pm$ standard deviation. Figures (B) and (D) show the probability of n occurrences, $P[N = n; t_a, t_b]$ of n, in the magnitude range $m = [4.5; 5.5)$, in 5 years, for mainshocks (red) and complete (blue) catalogs. The blue bars show the probability of n occurrences obtained directly from the complete catalogs. The vertical red lines show the average number of events in the magnitude range $m = [4.5; 5.5)$ from the synthetic complete catalogs.

To exemplify this point, we design a **second example** where we compare the hazard statistics given by ETAS and stationary Poisson simulations. First, we simulate 1000 realizations of 1 and 50 years duration, following the ETAS process (ETAS catalog). The ETAS parameters are identical to previous simulations, with same GR parameters for the background process. Then, we simulate 1000 realizations of 1 and 50 years duration, but assuming a stationary Poisson process (Mainshock Poissonian, MP catalog). The GR parameters of the MP catalog are $b = 0.99$ and $a = 4.98$, which are identical to the average GR parameters obtained from the 50-year synthetic ETAS catalogs.

The central dotted curves in figures 5.4(a) and 5.4(c) represent the mean annual rate of exceedance $m_{\lambda,exc}(m \geq m_j)$ obtained from the complete ETAS (blue) and MP (red) catalogs. The upper and lower dotted curves are the mean annual rate of exceedance $\pm$ standard deviation. The central continuous curves represent the annual rate of exceedance from the mean GR parameters, $\lambda_{exc,GR}(m \geq m_j)$ for the complete ETAS (blue) and MP (red) catalogs. The upper and lower dashed curves represent the upper and lower annual rate of exceedance curves $\lambda_{exc,GR+\sigma}(m \geq m_j)$ and $\lambda_{exc,GR-\sigma}(m \geq m_j)$. For the 1-year case, the mean GR parameters ($m_{(\hat{a})}$ and $m_{(\hat{b})}$) and standard deviations ($\sigma_{(\hat{a})}$ and $\sigma_{(\hat{b})}$) are $\hat{a} = 5.08 \pm 0.35$ and $\hat{b} = 1.05 \pm 0.12$ for the MP catalog, and $\hat{a} = 5.010 \pm 0.36$ and $\hat{b} = 1.03 \pm 0.14$ for the ETAS catalog. Likewise, for the 50-year case, the mean GR parameters ($m_{(\hat{a})}$ and $m_{(\hat{b})}$) and standard deviations ($\sigma_{(\hat{a})}$ and $\sigma_{(\hat{b})}$), are $\hat{a} = 4.97 \pm 0.039$ and $\hat{b} = 0.99 \pm 0.013$ for the MP catalog, and $\hat{a} = 4.98 \pm 0.040$ and $\hat{b} = 0.99 \pm 0.013$ for the ETAS catalog.

Figure 5.4 (B) and (D) show the probability of n occurrences in the magnitude range $m = [4.5; 5.5)$, in 1 and 50 years, for MP (red) and ETAS (blue) catalogs. The blue bars in both figures represent the probability distribution of n events obtained directly from complete synthetic catalogs. For the 1-year simulation case, the most likely number of events is $n = 2$, using the mean GR parameters obtained from both MP and ETAS catalogs. From the simulations, the average number of events in the magnitude range $m = [4.5; 5.5)$ is 2.19 (2187 events within the magnitude range $m = [4.5; 5.5)$ in 1000 realizations, see red vertical line). For the 50-year simulation case, the most likely number of events is $n = 146$ and 142, using the mean GR parameters obtained from

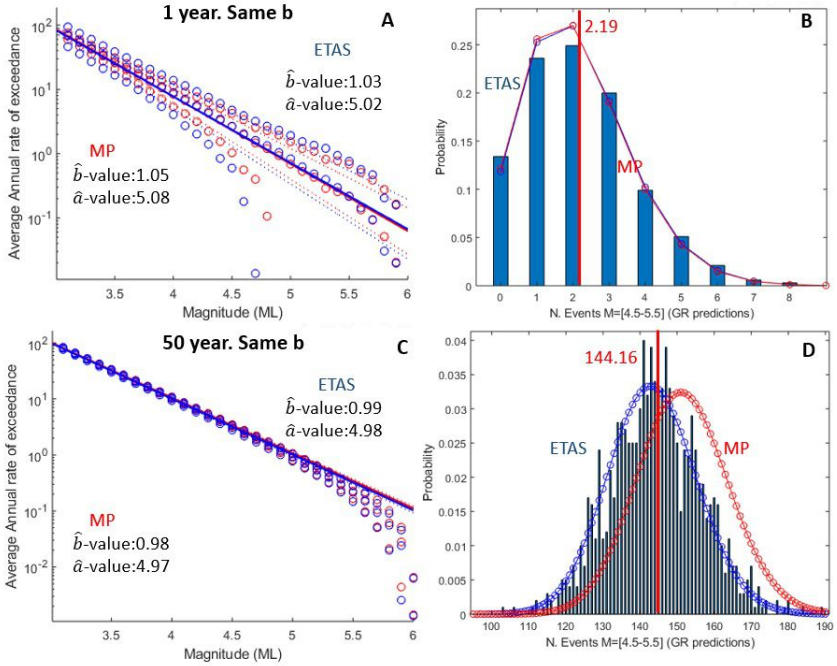

Figure 5.4: Comparison between results from MP and ETAS simulations. First row: 1-year simulation, second row: 50-year simulation. Figures (A) and (C) show the mean annual rate of exceedance ($m_{\lambda,exc}(m \geq m_j)$, dotted curves) from synthetic earthquake catalogs, and the annual rate of exceedance from estimated mean GR parameters ($\lambda_{exc,GR}(m \geq m_j)$, continuous curves), for MP simulations (red) and ETAS simulations (blue). Figures (B) and (D) show the probability of n occurrences, $P[N = n; t_a, t_b]$ of n, in the magnitude range $m = [4.5; 5.5)$ in 1 and 50 years, for MP simulations (red) and ETAS simulations (blue). The blue bars show the probability of n occurrences obtained directly from the complete catalogs. The vertical red lines represents the average number of events in the magnitude range $m = [4.5; 5.5)$ in the synthetic complete catalogs.

both MP and ETAS catalogs, respectively. From the simulations, the average number of events in the magnitude range $m = [4.5; 5.5)$ is 144.17 (1441660 events within the magnitude range $m = [4.5; 5.5)$ in 1000 realizations, see red vertical line).

From these results, it is clear that both models show similar hazard results. The only remarkable difference is the larger variability in the annual rate of exceedance from the ETAS simulations during the 1-year simulations. Therefore, for long-term seimicity, assuming a stationary Poisson process is still valid, which is particularly useful considering the longer time required to run ETAS simulations and the complex parameter estimation techniques that the ETAS model needs. Note however that we assumed stationary parameters for both the Poisson and ETAS processes. This is not valid for predicting the hazard associated with human-induced seismicity which is a non-stationary process and thus requires time-dependent GR parameters as described by (Reyes Canales and Van der Baan, 2019). ETAS is still a very useful tool for short-term seismicity hazard analyses, especially to estimate the hazard of the aftershocks triggered by a major tectonic event.

5.5 Application to a recent case of induced seismicity: Oklahoma, USA

The recent increase, peak and decline of the seismic activity in Oklahoma has been one of the most studied cases of induced seismicity worldwide (Ellsworth, 2013; Van der Baan and Calixto, 2017; Langenbruch and Zoback, 2016). The seismic activity in Oklahoma has been associated to large volumes of salt-water injection in the Arbuckle formation (Van der Baan and Calixto, 2017). The injection rates were stable prior to January 2012, then there was a considerable increased that peaked in January 2015. After that, the injection rates in the Arbuckle formation have declined until recent times. The number of induced earthquakes also followed a similar pattern of increase, peak and decline, with a 9 months lag with respect to the injection volumes (Langenbruch and Zoback, 2016; Van der Baan and Calixto, 2017).

In this study, we use the declustered and non-declustered earthquake catalogs for the 2017 Central and Eastern U.S. short-term seismic hazard model (Petersen et al., 2017). From this catalog, we only take the seismicity in Oklahoma between January 2009 and December 2016. The number of events in the non-declustered (complete) catalog is 21781, in the magnitude range $m = [1.28; 5.59]$. For the declustered catalog there are 814 events, in the magnitude range $m = [2.55; 5.59]$. The chosen algorithm to perform the declustering was the Gardner and Knopoff (1974) method, and the magnitudes in these catalogs are local magnitudes m_l. The largest induced-related earthquakes that have occurred in Oklahoma are the Prague $m_w = 5.6$ ($m_l = 5.59$) in November 2011, the Fairview $m_w = 5.1$ ($m_l = 5$) in February 2016, and the Pawnee $m_w = 5.8$ ($m_l = 5.55$) in September 2016 (Van der Baan and Calixto, 2017; Langenbruch and Zoback, 2016). Figure 5.5 shows the location of earthquakes in Oklahoma with magnitudes $m \geq 2.7$ from declustered and complete catalogs, as well as the cumulative number of earthquakes, from January 2009 until December 2016.

We apply the ETAS branching process simulation to the Oklahoma data, in order to assess the impact of including or excluding aftershock sequences in the hazard analysis. As a first approach, we assume stationary synthetic earthquake catalogs, as well as stationary GR and ETAS parameters, even though the induced seismicity in Oklahoma shows a clear non-stationarity pattern. However, these estimated parameters will represent the average non-stationary rates between 2009 and 2016. The rationale for doing this is that estimations of non-stationary ETAS parameters is beyond the scope of this article.

5.5.1 Estimation of the GR and ETAS parameters

In order to estimate the GR and ETAS parameters, we use the non-declustered and declustered earthquake catalogs for the 2017 Central and Eastern U.S. short-term seismic hazard model (Petersen et al., 2017). For our purposes, we will use the seismicity covering Oklahoma, from January 2009 until December 2016. Petersen et al. (2016), Petersen et al. (2017), and Petersen et al. (2018) apply the Gardner and Knopoff (1974) declustering method. It is important

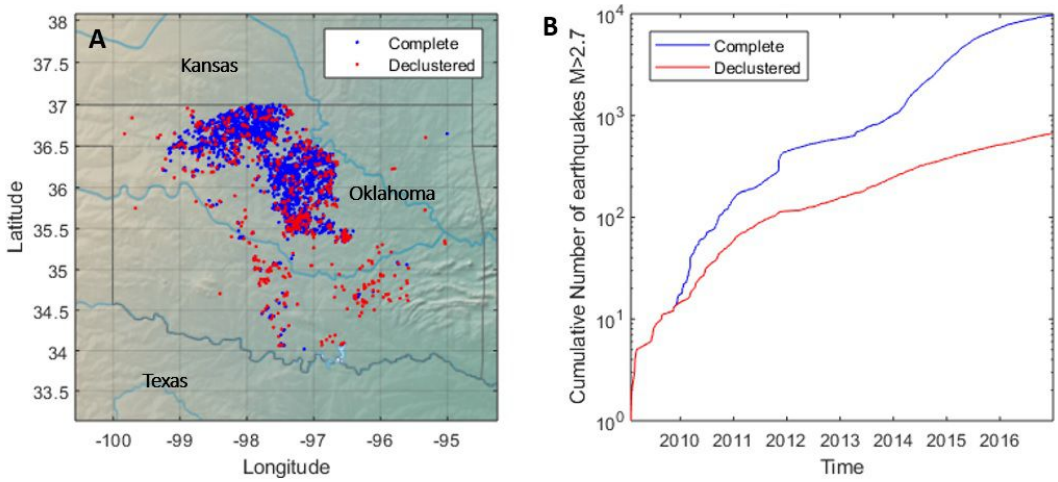

Figure 5.5: (A) Location of earthquakes with magnitude $m \geq 2.7$ from declustered (red) and complete (blue) catalogs. (B) Cumulative number of earthquakes in Oklahoma with magnitudes $m \geq 2.7$, using declustered (red) and complete (blue) catalogs, from Jan. 2009 until Dec. 2016

to mention that different declustering methods might lead to different GR parameters, as shown by Teng and Baker (2019). However, the Gardner and Knopoff (1974) declustering method is the standard declustering methodology performed by the United State Geological Survey (USGS), and the resulting b-value ($b = 1.0$) from the mainshock catalog fits reasonably well the observed distribution of large events ($M \geq 4.5$).

We first estimated the GR parameters from the non-declustered (complete) catalog, using Maximum Likelihood estimations (MLM) (Aki, 1965; Wiemer and Wyss, 1997). For the error estimations of the b-value we use the method described by (Shi and Bolt, 1982). We acknowledge that using a different magnitude of completeness (M_c) may lead to different estimations of b-values (Akinci et al., 2018). However, we are confident with the magnitude of completeness value used in this analysis, $M_c = 2.7$, which has been supported by multiple studies in Oklahoma (Petersen et al., 2016; Teng and Baker, 2019). The a_c- and b_c-values from the complete catalog equal $a_c = 7.392$ and $b_c = 1.596$, respectively, with a b-value error of 0.015, see figure 5.6 (A). We also estimate the GR parameters from the declustered (mainshock) catalog, using MLM. The a_m- and b_m-values from the declustered catalog equal $a_m = 4.62$ and $b_m = 1.00$, respectively, with a b-value error of 0.035, see figure 5.6 (A).

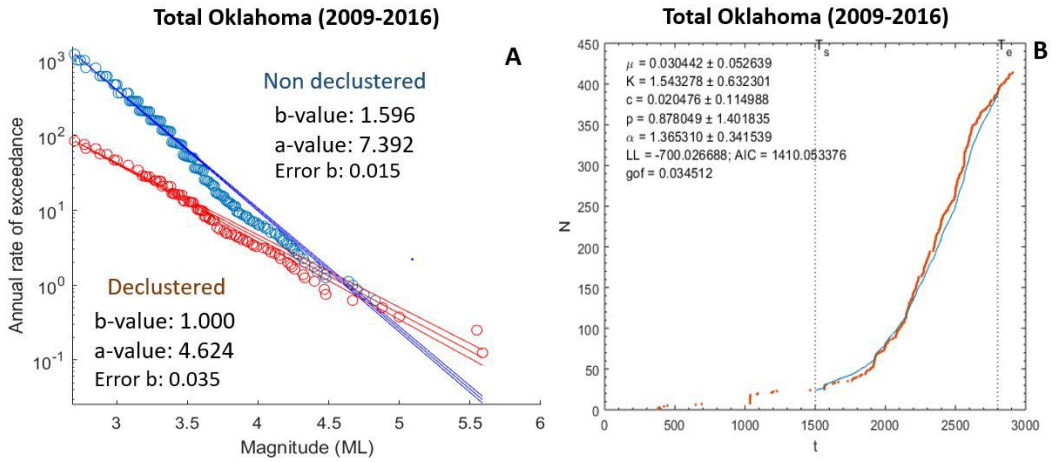

Figure 5.6: (A) Estimation of the GR parameters in Oklahoma, using the MLM, for declustered and non-declustered catalogs in Oklahoma (2009-2016). The upper and lower curves represent the b-value $\pm$ error. Notice the very different slopes for the declustered and non-declustered catalogs. The dots represents the annual rate of exceedance per magnitude level for declustered (red) and Non-declustered (blue) catalogs. (B) Estimation of the ETAS parameters in Oklahoma, using MLM. The orange curve represents the cumulative number of events $m \geq 3.5$ and the blue line shows the fitted model from estimated ETAS parameters.

For the first simulation approach, we assume that the b-values of the background and aftershock sequences are equal to the b-value of the complete catalog, $b_b = b_a = b_c = 1.596$. To keep the same number of mainshock events with $m \geq 3.5$, we set $a_b = 6.60$ for the mainshock sequence. For the second simulation approach, we assume that the a-value and b-values of the background sequence are identical to the a-value and b-value of the observed declustered catalog (mainshock), that is $a_b = a_m = 4.60$ and $b_b = b_m = 1.0$. On the other hand, for the b-value of the aftershock sequence, we estimate $b_a = 1.66$, which is obtained by applying MLM to the residual aftershock sequence resulting from the declustering methodology. The minimum and maximum magnitudes to simulate are: $M_{min} = 2.7$ and $M_{max} = 6.0$. We choose a relatively low M_{max} in order to compare the simulations with the actual number of observed large earthquakes (The largest event in the catalog is $M = 5.6$). However, we acknowledge that a higher M_{max} might be appropriate.

We use MLM for the estimation of ETAS parameters ((Ogata, 1998; Daley and Vere-Jones, 2003), R. Shcherbakov, pers. comm., 2019). To get stable results, we use a $M_{min,ETAS} = 3.5$ for the ETAS parametrization. The temporal interval used for the curve fitting is defined as: $t_{start} = 1500$ and $t_{end} = 2800$ days, which correspond to February 2013 and January 2016, respectively. The estimated ETAS parameters for the induced seismicity in Oklahoma are: $c = 0.0204$, $p = 0.8780$, $\alpha = 1.36$ and $\mu = 0.0304$, see figure 5.6 (B). For a better adjustment with observed seismicity, we set the productivity of the aftershock sequence to $K = 1.70$. As before we use the GR parameters to obtain the annual rate of the mainshocks, instead of using the background seismicity from the ETAS process. Both rates agree within a 15% margin of difference. The maximum time for the aftershock generation is set at $t_{max} = 200$ days for all parent events. In this example, we do not consider necessary going beyond $tmax > 200$ days because we would be increasing the computational cost to obtain similar global statistical results.

5.5.2 Hazard Analysis from ETAS simulations and observed seismicity

In order study the impact of including or excluding aftershocks in the seismic hazard analysis in Oklahoma, we consider two sets of simulations. For the first set of simulations, we assume the same b-value for the mainshock and aftershock sequences $b_m = b_a = 1.596$. For the second set of simulations, we assume a different b-value between the mainshock and aftershock sequences $b_m = 1.0$ and $b_a = 1.66$. Each set consists of 10 realizations with a duration of 9 years each. However, the first year is neglected (8 years to analyze) due to the underestimation of events caused by the lack of history before t_{start}. As mentioned before, the simulations assume stationary GR and ETAS parameters.

The central dotted curves in figures 5.7(A) and 5.7(B) represent the mean annual rate of exceedance $m_{\lambda,exc,c}(m \geq m_j)$ and $m_{\lambda,exc,m}(m \geq m_j)$ for the complete and mainshock catalogs, respectively, estimated by counting events with magnitudes $m \geq m_j$ in the simulated samples. The upper and lower dotted curves are the mean annual rate of exceedance plus or minus the standard deviation, $m_{\lambda,exc}(m \geq m_j) \pm \sigma_{\lambda,exc}$. The central continuous curves in figures 5.7(A) and 5.7(B) represent the annual rate of exceedance from the mean GR parameters, $\lambda_{exc,GR,c}(m \geq m_j)$ and $\lambda_{exc,GR,m}(m \geq m_j)$, for complete and mainshock catalogs, respectively, estimated from the simulations. The upper and lower dashed curves represent the upper and lower annual rate of exceedance curves $\lambda_{exc,GR+\sigma}(m \geq m_j)$ and $\lambda_{exc,GR-\sigma}(m \geq m_j)$, obtained from the mean GR parameters $\pm$ their standard deviations, $\sigma_{(\hat{a})}$ and $\sigma_{(\hat{b})}$.

Figure 5.7 (A) shows the results for the first set of simulations. The mean a- and b-values for the complete and mainshock catalogs are: $m_{(a_c)} = 6.95$ and $m_{(b_c)} = 1.57$ for the complete catalog, and $m_{(a_m)} = 5.49$ and $m_{(b_m)} = 1.34$, for the mainshock catalog. Figure 5.7 (A) show the results for the second set of simulations. The mean a- and b-values for the complete and mainshock catalogs are: $m_{(a_c)} = 7.32$ and $m_{(b_c)} = 1.62$ for the complete catalog, and $m_{(a_m)} = 4.55$ and $m_{(b_m)} = 0.98$, for the mainshock catalog. Finally, for a better comparison between the simulations and observed seismicity, we plot again the annual rate of exceedance from observed declustered and non-declustered

catalogs in Oklahoma, see figure 5.7 (C). The reason why we assume $b_m = b_a$ for the first set of simulation, is to confirm if it is still reasonable to simulate mainshock and complete catalogs using the same b-value through all the simulation. However, the only way we could replicate the clear intersection in the magnitude-frequency plot between GR from mainshock and complete catalogs is by considering different b-values between mainshock and aftershock sequences.

The second row of figure 5.7 shows the probability of n occurrences in the magnitude range $m = [5.0; 6.0)$, given by the sets of simulations and observed seismicity. When $b_m = b_a = 1.596$, the most likely number of events are $n = 0$ and $n = 1$, given the mean GR parameters of the simulated mainshock and complete catalogs, respectively. When $b_m = 1.0$ and $b_a = 1.66$ the most likely number of events are $n = 3$ and $n = 1$, given the mean GR parameters of the simulated mainshock and complete catalogs, respectively. Finally, using the GR parameters of the observed seismicity, the most likely number of events are $n = 3$ and $n = 2$, given the GR parameters of the mainshock catalog and complete catalog, respectively. 3 earthquakes with local magnitude $m \geq 5$ in Oklahoma occurred between 2009 and 2016 (Petersen et al., 2017).

Figure 5.7 thus show that the declustering is advisable for a more representative hazard assessment in Oklahoma because the b-value of the aftershocks and mainshocks is different, leading to a non-linear magnitude-frequency distribution. The GR parameters estimated for the complete (non-declustered) catalog yield a lower likelihood of an $m \geq 5$ event occurring than if the GR parameters are estimated from the declustered catalog (Fig. 5.7 F).

5.6 Discussion

One important discussion topic is the declustering methodology and how they might result in different b-values for the mainshock sequence. The purpose of this study is not to show which declustering method is most appropriate, but rather to show when declustering might be a useful tool for the correct estimation of recurrence characteristics for large events. We acknowledge that different declustering methods might lead to different b-values for the main-

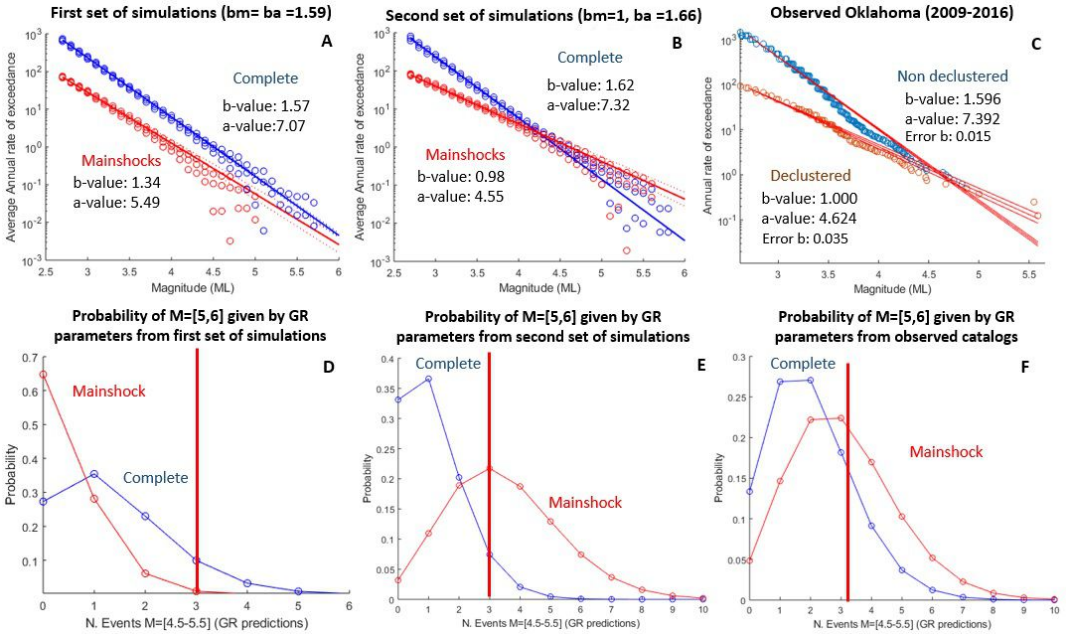

Figure 5.7: **First row**: (A) and (B) show the mean annual rate of exceedance ($m_{\lambda,exc}(m \geq m_j)$, dotted curves) from the synthetic catalogs and annual rate of exceedance from the mean GR parameters ($\lambda_{exc,GR}(m \geq m_j)$, continuous curves), using the estimated parameters for the induced seismicity in Oklahoma. We consider the same b-value for the mainshock and aftershock events (A), or different b-values for the mainshock and aftershock sequences (B). The upper and lower dashed curves are the mean annual rates $\pm$ their standard deviation. The estimation of the GR parameters in Oklahoma for declustered and non-declustered catalogs is plotted again for comparative purposes(C). **Second row**: Probability of n occurrences, $P[N = n; t_a, t_b]$ of n, in the magnitude range $m = [5.0; 6.0)$, in a 8-year period, given by the complete and mainshock catalogs from (D) the mean GR parameters from the first set of simulations ($b_m = b_a = 1.596$), (E) the mean GR parameters from the second set of simulations ($b_m = 1.0$ and $b_a = 1.66$) and (F) the GR parameters from the observed seismicity. The number of earthquakes in Oklahoma with local magnitude larger than $m \geq 5$ is $n = 3$ (red vertical lines).

shock sequence, as shown in the study implemented by Teng and Baker (2019) in Oklahoma. One reason for obtaining different b-values is that some declustering methods tend to reduce the number of earthquakes more than others. In one extreme, we have the Gardner and Knopoff (1974) methodology, which is one of the methods that leads to the largest reduction in earthquake numbers. On the other hand, we have the Reasenberg (1985) methodology, which tends to reduce less the number of earthquakes and, in the case of Oklahoma, keeps a similar b-value between declustered and non-declustered catalogs. One of the advantages of the synthetic simulations shown in this study is that we do not have to perform any declustering methodology in order to know with complete certainty the background/mainshock sequences. Therefore, from a theoretical point of view and without relying on any declustering method to determine the mainshock sequence, we identify the cases when declustering is more appropriate in order to estimate the recurrence statistics of the large events (mainshocks).

Based on the synthetic examples when the b_m-value of the mainshocks is considerably smaller than the b_a-value of the aftershock sequence, and the results from observed induced seismicity in Oklahoma, seismicity declustering is highly recommended for the hazard analysis. It yields better estimates of the b-value of the largest events which may not be well represented in the observed catalog. However, declustering is inappropriate in catalogs with a small number of earthquakes or when the b_b-value of the background is similar to b_a-value of the aftershock sequence, as shown in the synthetic examples. Similar findings about discouraging the use of declustering methodologies are described by Boyd (2012). They found that the inclusion of aftershocks causes ground motions that are exceeded at probability levels of engineering interest to increase by about 10 % but could be as high as 20 % if variations in aftershock productivity are accounted. However, they perform this analysis using earthquake catalogs from California, where the b_b-value of the background is similar to b_a-value of the aftershocks (Boyd, 2012; Teng and Baker, 2019). Note that we are only evaluating temporal declustering. For instance, Marzocchi and Taroni (2014) have shown the relevance of spatial declustering in order to avoid underestimations in the hazard.

One question that must be evaluated is how different should be the b-value of the mainshocks and aftershock sequence in order to decluster or not. It would depend on the catalog and, ultimately, the analyst. From the Oklahoma example, the difference of b-values between complete and mainshock catalogs is 0.59, and it is enough to advocate for declustering. Without declustering, we would have underestimated the number of events larger than $M = 4.5$. It ultimately depends on the analyst to decide which large magnitudes cannot be underestimated. For instance, let us say that there is a kink in the magnitude-frequency distribution, and the GR curves from complete and mainshock catalogs intersect in a magnitude lower than a defined magnitude threshold; that would be a clear indication that the b-values are different enough to consider declustering. Nonetheless, the true bottleneck is not that main- and aftershocks sequences may have different b-values, but that a single b-value is used to extrapolate hazard predictions for large-magnitude events that may be poorly sampled in the recorded catalogs due to limited observation periods compared with the recurrence time of large earthquakes.

A tell-tale sign that the b-values of mainshocks and aftershocks may be different is the appearance of kinks in the magnitude-frequency distributions (Figures 5.2 (C), 5.3 (C), 5.6 (A), and 5.7 (B)). These are absent if the mainshocks and aftershocks are characterized by identical b-values (Figures 5.2 (A), 5.3 (A) and 5.7 (A)). On the other hand, the suggestion of increasing b-values for aftershock sequences has been observed in recent studies done in the Amatrice–Norcia (Italy) and Kumamoto (Japan) Earthquakes (Gulia and Wiemer, 2019). However, further investigation is necessary in order to understand why, in some cases, the b-values between background and aftershock sequences may be different, independently on the performed declustering method. This may also lead to an enhanced understanding of the underlying earthquake physics.

5.7 Conclusions

If mainshocks and aftershocks are characterized by different b-values, declustering leads to improved hazard assessments, since it allows for better estimation of the magnitude-frequency distribution of the largest events. Conversely, if

mainshocks and aftershocks have similar b-values then declustering is inadmissible since it eliminates larger events that contribute to the long-term hazard. Assuming Poissonian distributions in hazard predictions does not lead to inaccurate long-term hazard predictions, even if time-varying aftershock sequences are present in the catalog. Short-term hazard due to aftershocks is, however, better estimated by evaluating appropriate non-stationary models such as the ETAS process.

Chapter 6

Towards building seismicity scenarios in areas susceptible to induced earthquakes in Alberta

6.1 Introduction

The first induced seismicity cases in Alberta date back to the 1970s, related to conventional oil and gas production (Van der Baan and Calixto, 2017). Gas extraction at the Strachan field, Rocky Mountain House (Wetmiller, 1986; Baranova et al., 1999), has been associated with the peaks of seismicity observed in the late 1970s and early 1980s in that region. Since the mid-1990s, seismic activity has been observed at the Cordel field, Brazeau River area, linked to waste water disposal activities (Schultz et al., 2014). Even though the history of induced seismicity in Alberta is over four decades long, induced earthquakes have brought more attention since the recent increase in the number of seismic events linked to the development of unconventional resources.

Since 2010 there has been an increase in the number of earthquakes in the province of Alberta (Atkinson et al., 2016; Van der Baan and Calixto, 2017). Most of these events are associated with the hydraulic fracturing activities in the Duvernay Formation near Fox Creek (Schultz et al., 2015). These activities have been linked to events as large as magnitude $M = 4.8$ (Alberta Geological

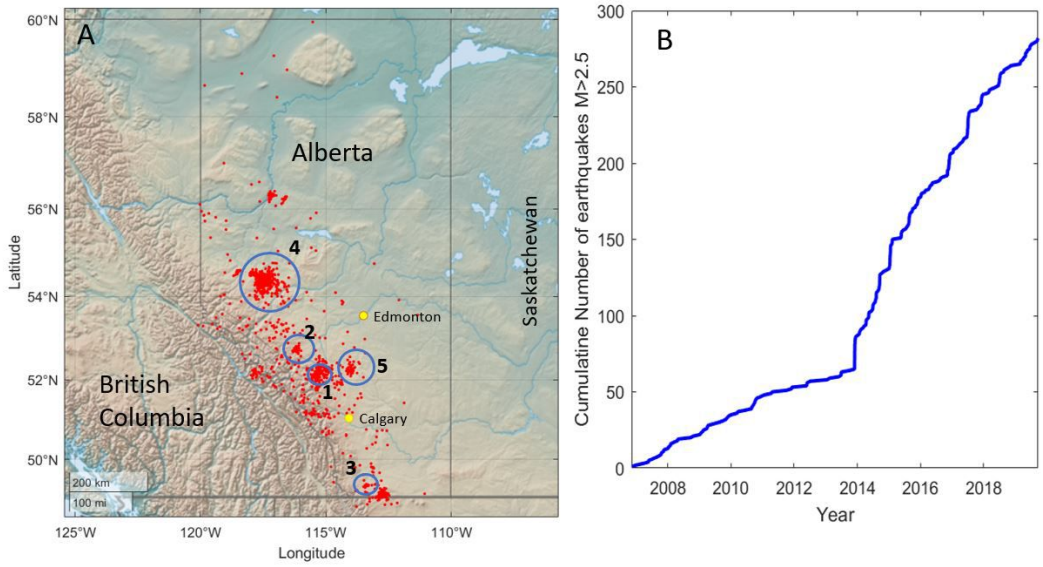

Figure 6.1: (A) Location of earthquakes in Alberta, using the catalog from Alberta Geological Survey (2019) from Oct. 2006 until Oct. 2019. The different induced seismicity cases are shown in circles: (1) Rocky Mountain House, (2) Brazeau River, (3) Cardston seismicity cluster, (4) Duvernay Formation (Fox Creek area), (5) Duvernay Formation (Red Deer area). (B) Cumulative number of earthquakes in Alberta with magnitudes $m \geq 2.5$. Notice the considerable change in the cumulative number of earthquakes, particularly after 2013.

Survey, 2019). In March 2019, a seismic event $M = 4.2$ was recorded near the city of Red Deer (Alberta Geological Survey, 2019), and it has been linked to hydraulic fracturing operations in the area. Figure 6.1 shows a map of the recorded events in the province of Alberta since Oct. 2006 (A) and the equivalent cumulative number of earthquakes with magnitudes $m \geq 2.5$(B). This map includes both natural and induced events. Since Oct. 2006, more than 275 events with magnitudes $m \geq 2.5$ have been recorded by the Alberta Geological Survey. Notice that most of the events are located near the foothills and the Rocky Mountains, as well as areas prone to induced seismicity like the Duvernay Formation near Fox Creek.

Unsurprisingly, these induced events have to lead to increasing concerns over seismic hazard (Atkinson et al., 2015). Previous hazard assessments (Atkinson et al., 2015) have shown that induced events generate hazard potentially higher than the natural seismic hazard, especially in areas with small-to-moderate natural background seismicity like the province of Alberta. Furthermore, the

hazard related to future induced seismicity relies on unknown forthcoming seismicity patterns. As a result, it has been necessary to develop seismicity scenarios that can give us insights into future seismicity patterns in areas known to be prone to induced earthquakes. Such a scenario evaluation is the first stage before developing a full seismic hazard analysis for induced seismicity.

In this chapter, we first present an overview of the reported induced seismicity cases in Alberta. Then, we propose a methodology to build seismicity scenarios in areas susceptible to induced earthquakes. This methodology consists of the following steps: (1) Use of retrospective Probabilistic Seismic Hazard Analysis (PSHA) maps to identify areas affected by induced seismicity, and quantitatively evaluate the temporal changes in the seismic hazard; (2) Identify operational parameters that can act as proxies to foreseen seismicity rates in areas prone to induced earthquakes; (3) Forecast appropriate proxies based on projections of anthropogenic activities; (4) Building future seismicity scenarios using forecasts of operational proxies. This type of scenario building, based on projected operations in susceptible areas to induced seismicity, can give us insights into future seismicity patterns. Government and industry can evaluate the areas that will likely experience an increase or decrease of induced events and, therefore, develop strategies based on future seismicity scenarios.

6.2 Induced seismicity cases in Alberta and current regulatory framework

1. **Rocky Mountain House:** the seismic events associated with the secondary recovery of gas at the Strachan Field, near Rocky Mountain House, is the first case of induced seismicity recorded in Alberta (Wetmiller, 1986; Van der Baan and Calixto, 2017). Large amounts of gas were removed at the Leduc Formation, reducing the vertical stresses and facilitating fault activation (Baranova et al., 1999). The seismicity started in the mid-1970s, peaking in 1985 with 40 earthquakes larger than $M > 3.0$ in that year (Stern et al., 2013). The number of earthquakes recorded in the area has decreased since then, with four earthquakes larger than

$M > 3.0$ in 2011. The largest induced event had a magnitude of $M = 4.0$ (Stern et al., 2013).

2. **Brazeau River Area:** the seismic events at the Cordel Field, Brazeau River area, have been linked to waste water disposal activities in the Rundle Group (Schultz et al., 2014). The induced seismicity in the area started in the mid-1990s and is still ongoing (Schultz et al., 2014). However, the annual number of events has decreased since 2006, in line with a reduction of the volumes injected. The largest induced event had a magnitude of $M = 4.0$ (Schultz et al., 2014).

3. **Cardston Swarm:** hydraulic fracturing activity in the Exshaw Formation near Cardston, has been linked to a swarm of events recorded between December 2011 and March 2012 (Schultz et al., 2015). More than 60 events were recorded, and the largest induced event had a magnitude $M = 3.0$ (Schultz et al., 2015). The area had no previously known seismic activity, and the timing of the operations temporally correlates with the timing of the swarm. After the hydraulic fracturing activity was over, no more seismic events were recorded in the area. Further studies (Galloway et al., 2018) have related these induced events with the activation of basement faults, where the earthquakes occurred.

4. **Duvernay Formation:** Induced seismic activity related to hydraulic fracturing activity in the Duvernay Formation started in December 2013, west of Fox Creek (Schultz et al., 2018). Since then, this area has become one of the most seismic active regions of the province, with over 220 events larger than $M > 2.5$ (Schultz et al., 2017). However, most of the events are of small magnitude, and only four earthquakes correspond to events with magnitude larger than magnitude $M = 4.0$ (Alberta Geological Survey, 2019). The largest event occurred in January 2016, with a recorded local magnitude of $M = 4.8$.

   Schultz et al. (2017, 2018) conclude that the events were triggered by pore pressure increases in response to hydraulic fracturing activity. They also related these induced events with the activation of Precambrian basement faults in the For Creek area. More recently, induced events related to

hydraulic fracturing activities at the Duvernay Formation have also been recorded near Red Deer (Alberta Geological Survey, 2019). In March 2019, an event $M = 4.2$ was recorded in this area (Alberta Geological Survey, 2019), getting attention from the media and causing concerns in the local population (Snowdon, 2019).

These seismic events, particularly those related to the hydraulic fracturing operations in the Duvernay Formation, have caused increasing concerns over induced seismic hazard (Atkinson et al., 2015). In response, the Alberta Energy Regulator (AER) introduced a traffic light protocol as part of the monitoring and reporting requirements to operate in specific areas known to be susceptible to induced seismicity (Alberta Energy Regulator, 2015). The traffic light protocol specifies that (1) the operators do not need to change their strategies if the magnitude of induced events are below a defined magnitude (Green light); (2) they have to apply mitigation strategies if the magnitude of the induced events is between a defined magnitude range (Yellow light); (3) or they have to cease operations if there is an induced event exceeding a threshold magnitude (Red light) (Alberta Energy Regulator, 2015; Kao et al., 2018). Three Subsurface Orders have been mandated to limit the impact of potential induced events from hydraulic fracturing up to date. These Sub-Surface Orders set different traffic light protocols based on their geology and surface structures (Alberta Energy Regulator, 2019b).

Subsurface Order no. 2 stipulates that the Duvernay operators in the Fox Creek area must report all adjacent yellow light events ($M \geq 2.0$) and implement mitigation strategies. If a red light event occurs during the operation ($M \geq 4.0$), the operator must cease the operations (Alberta Energy Regulator, 2015). Subsurface Order no. 6 defines two areas near the Brazeau dam: (1) One area within 5 km of the Brazeau dam that prohibits hydraulic fracturing targeting the Duvernay Formation or below; (2) a second area within 3 km of the Brazeau Dam that prohibits hydraulic fracturing in formations above the Duvernay Formation. Subsurface Order no. 6 also stipulates a traffic light protocol, where operators must report all yellow light events ($M \geq 1.0$) and implement mitigation strategies. The operator must cease the operations if a red light event occurs ($M \geq 2.5$)(Alberta Energy Regulator, 2019b). Subsurface

Order no. 7 prohibits hydraulic fracturing within 5 km of the Dickson Dam if operations are targeting the Duvernay Formation or below. It also stipulates a traffic light protocol, where yellow light events are set at $M \geq 1.0$, and the threshold for red light events is defined at $M \geq 3.0$ (Alberta Energy Regulator, 2019c). Hydraulic fracturing is considerably limited in these areas defined by Subsurface order no. 6 and 7 in order to reduced any induced seismic hazard near a critical facility.

6.3 Building seismicity scenarios in areas susceptible to induced earthquakes in Alberta

I propose a methodology to build seismicity scenarios in areas susceptible to induced earthquakes. First, we need to identify the areas that have experienced considerable seismicity changes due to the emergence of induced events. To do this, I recommend the use of retrospective PSHA maps in areas with induced seismic activity, since they quantitatively describes the seismic hazard changes over time. Second, we need to identify the operational parameters that can act as proxies to foreseen seismicity changes in areas prone to induced earthquakes. Third, projections of future anthropogenic operations are required, in particular, the operational parameters correlated to the induced seismicity. Finally, by using the identified proxies, we can build future seismicity scenarios in areas prone to induced seismicity.

6.3.1 Previous Probabilistic Seismic Hazard Analysis in Alberta

PSHA for natural seismicity in Alberta: The Geological Survey of Canada regularly updates the seismic hazard maps for all provinces in Canada, including the province of Alberta. For the 2015 Seismic Hazard map of Canada, The Geological Survey of Canada defines four seismic areas in Alberta based on historical catalogs (Natural Resources Canada, 2015). The two most active seismic source areas, the Rocky Mountain Fold/Thrust North (ROCN) and

the Rocky Mountain Fold/Thrust South (ROCS), comprise the mountainous southwestern section of the province. A third seismic source area, the Foothills (FTH), links the relatively active mountainous sources with the stable craton. Finally, the Stable Cratonic Core Western Canada (SCCWHC) comprises most of the northeastern section of the province and is characterized by low seismicity. According to the 2015 Seismic Hazard map of Canada, there is a probability of 2% in 50 years to reach a PGA of approximately 0.05 g for SCCWH (low hazard), 0.10 g for FTH and ROCN (moderate to low hazard), and 0.15 g for ROCS (moderate hazard). In this case, g refers to the gravitational acceleration (980 cm/s^2). Figure 6.2 (A) shows a simplified seismic hazard map for the province of Alberta, in terms of low to high seismic hazard. Figure 6.2 (B) shows the areas that encompass the natural seismic sources described in the 2015 Seismic Hazard map of Canada (Natural Resources Canada, 2015).

These seismic hazard maps are built by using historical catalogs that explicitly exclude induced seismicity. This is useful since it is important to characterize the background seismicity previous to any anthropogenic activity. These seismic hazard maps work as baselines for the hazard analysis of induced seismicity cases. However, these maps do not count for the hazard related to induced seismicity, which in many areas has become the primary source of hazard (e.g., Fox Creek, Alberta (Atkinson et al., 2015)). Therefore, induced seismic sources have to be included in order to characterize the seismic hazard in the province accurately.

PSHA including induced seismicity in Alberta: Ghofrani et al. (2019) elaborate retrospective Probabilistic Seismic Hazard Analysis (PSHA) maps for Alberta, including the induced seismic activity from 2011 to 2018. The Ghofrani et al. (2019) analysis is closely related to previous studies in central and eastern United States (Petersen et al., 2016, 2017, 2018), where they develop a one-year hazard forecast based on catalogs with recorded induced events. Even though these studies are limited to short-term hazard predictions, due to the assumption of temporal stationarity, they can provide insights into the temporal evolution of the seismic hazard.

For the retrospective seismic hazard analysis, Ghofrani et al. (2019) define a single seismic source area namely the Western Canadian Sedimenatary Basin

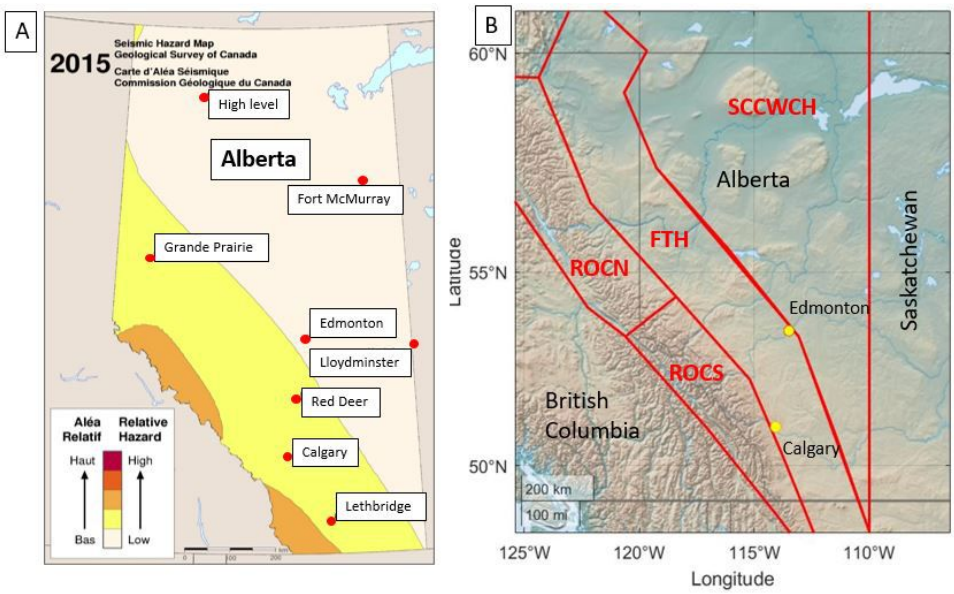

Figure 6.2: (A) Simplified seismic hazard map for the province of Alberta (only natural seismicity). A smoothed-seismicity model has been applied for this example. Modified from Natural Resources Canada (2015). (B) Natural seismic sources areas (red polygons) defined by Natural Resources Canada (2015) for the province of Alberta. FTH=Foothills; ROCN= Rocky Mountain Fold/Thrust North; ROCS=Rocky Mountain Fold/Thrust South; SCCWHC= Stable Cratonic Core Western Canada, H model.

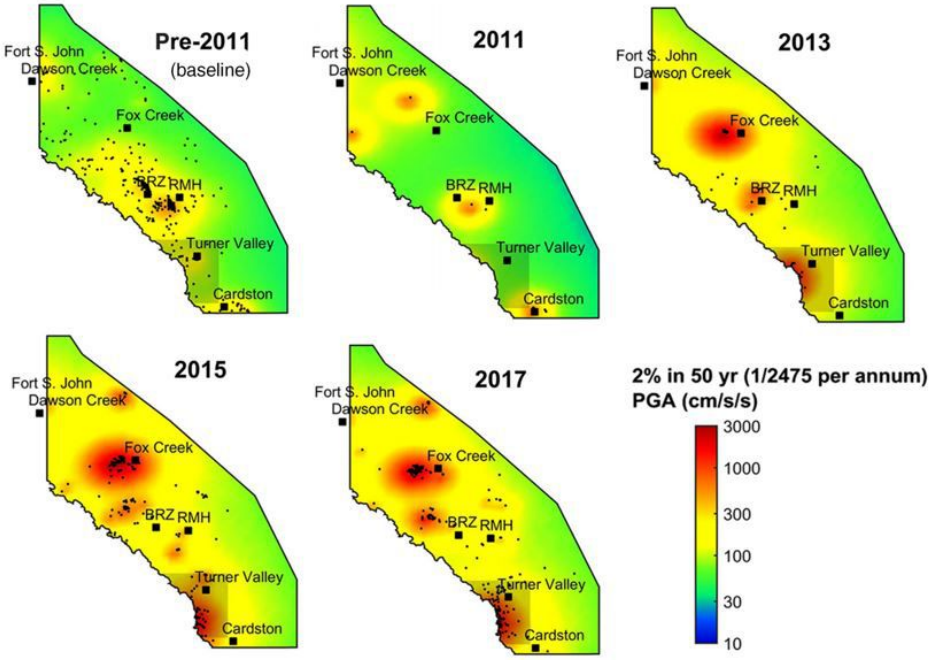

Figure 6.3: 2% in 50 years seismic hazard maps pre-2011 (baseline seismicity), and for the years 2011, 2013, 2015, 2017 in southern and central Alberta. These maps include both natural and induced seismicity. Notice how the hazard increases in 2013, 2015 and 2017, particularly near Fox Creek, as a result of increasing induced events in the area. The shaded region near Turner Valley corresponds to areas where the catalog is contaminated by blasts. BRZ= Brazeau cluster; RMH= Rocky Mountain House. Modified from Ghofrani et al. (2019).

(WCSB). The Gutenberg-Richter (GR) parameters for this area, a-and b-values, are originally defined by the earthquake catalogs before 2011 (pre-2011, baseline) and the natural seismicity models described by the 2015 Seismic Hazard map of Canada (Natural Resources Canada, 2015). Then, the annual a-values change depending on the increase or decrease in the number of earthquakes $M \geq 3$ per year. The catalogs used to estimate the changes in the a-value per year, include natural and induced seismicity, as well as blasts. They estimated a range of b-values, $b = [0.75, 0.89, 1.05]$, and the b-value is assumed constant over time. Ghofrani et al. (2019) also proposed a range of $M_{max} = [6.0, 7.5]$, which is also similar to the maximum magnitude expected for natural seismicity in the Western Canadian Sedimentary Basin (Halchuk et al., 2014).

Once the GR parameters are defined per year, Ghofrani et al. (2019) implemented a Monte Carlo simulation approach for PSHA to generate the seismic hazard maps in Alberta (Musson, 2000; Assatourians and Atkinson, 2013). The GR parameters are assumed constant over a 50-year time-frame, in order to get the 2 % probability in 50 years maps (figure 6.3). For the Ground Motion Prediction Equations (GMPEs), they use a four-brach GMPE suite designed by Atkinson (2015), for small-to-moderate earthquakes at a short hypocentral distance. Ghofrani et al. (2019) use a smoothed-seismicity model for the distribution of synthetic events. This model considers the seismic activity density a continuous function in space, such that earthquakes tend to be located near clusters of previous earthquakes (Assatourians and Atkinson, 2019). Therefore, even though it is a single source area with a set of GR parameters, the seismicity is going to be concentrated in areas where there has been previous seismicity, in accordance with the historical catalogs. Figure 6.3 shows the retrospective seismic hazard maps for southern and central Alberta. Notice how the seismic hazard increases in areas like Fox Creek or the Brazeau cluster, in line with increasing induced seismicity.

These retrospective PSHA maps are appropriate to quantitatively evaluate the temporal hazard changes in areas affected by induced earthquakes. These hazard changes are expressed in probabilities to reach or exceed an acceleration level in a period of time. Therefore, we can identify what areas have experienced the most significant hazard change and prioritize them for the seismicity

scenario building. However, future seismic hazard evaluations in Alberta should consider the following aspects:

1. Blasts should be excluded from the seismic catalogs. Ghofrani et al. (2019) recognize that these events might contaminate the seismic hazard maps, particularly in the area west of Turner Valley (figure 6.3). Blasts generate areas with anomalous high hazard, which do not correspond with reality. Also, the inclusion of blasts might alter global statistics by increasing the population of small magnitude earthquakes. This is particularly important since a single set of GR parameters is defined for all the WCSB seismic source areas.

2. Instead of the standard likelihood level of 2 % in 50 years used by Ghofrani et al. (2019), we recommend the 1% likelihood level in 1 year proposed by Petersen et al. (2016, 2017, 2018) in central and eastern United States. A short time frame of 1 year is more appropriate for short-term seismicity, particularly since the hazard has been calculated on an annual basis. Stationary a-and b-values over 50 years are not realistic given the variations in the industrial activity.

3. Defining multiple seismic source areas for induced seismicity cases, instead of a single seismic source area as defined by Ghofrani et al. (2019). The use of multiple seismic source areas with unique GR parameters has been used in the seismic hazard analysis in central and eastern United States (Petersen et al., 2016, 2017, 2018). The new temporal source areas should have their own set of GR parameters, instead of using regional averages and then compressing earthquakes into tight zones. Considering multiple areas is important, particularly if the induced seismic source areas have different GR parameters, a-and b-values. If this is the case, using a single set of a-and b-values obtained from regional averages, could result in misleading local hazard estimations. A more traditional approach should lead to a better match between observed and predicted local seismicity.

4. A validation step is required to make sure that the predicted local seismicity in the Ghofrani et al. (2019) model honors the local observations.

This validation can be done by comparing the rate of earthquakes between predicted and observed earthquakes, particularly in locations with induced seismicity. A more stringent validation should compare magnitude-frequency distributions between local predictions and observed earthquakes.

6.3.2 Identifying appropriate proxies to describe seismicity changes in areas prone to induced seismicity

Several studies (Van der Baan and Calixto, 2017; Langenbruch and Zoback, 2016; Schultz et al., 2018) have shown correlations between operational parameters and the number of induced events in areas prone to induced seismicity. In some cases, both the oil production and the total volume of salt water disposed are strongly correlated to the number of induced earthquakes. Van der Baan and Calixto (2017) found that increased seismicity in Oklahoma has an 85 % correlation with oil production. However, this correlation was not found in other induced seismicity cases, including Alberta. On the other hand, (Schultz et al., 2018) found a strong correlation between the total volume injected by hydraulic-fracturing wells and the observed number of earthquakes in the Duvernay Formation near Fox Creek, Alberta. Figure 6.4 shows the logarithm of the number of earthquakes $M \geq 1.3$ versus the logarithm of the cumulative injection volumes using all Duvernay hydraulic fracturing pads (A), and using only seismogenic hydraulic fracturing pads (B). As expected, the correlation is higher when only suspected seismogenic wells are considered.

If there are clear correlations between operational parameters and the number of induced earthquakes, they can be used as proxies to describe the overall future seismicity in areas prone to induced earthquakes. These correlations can be estimated for the different induced seismicity cases in Alberta. For instance, Schultz et al. (2018) plot the logarithm of the number of earthquakes as a function of the logarithm of the cumulative injected volume for the Fox Creek induced seismicity area. They found a linear trend in the logarithmic domain, and a high correlation between the cumulative injected volume and the number

of earthquakes. This correlation can be used as a proxy to describe seismicity changes; however, further options should be consider (e.g. production rates) and select the operational proxies that best correlate with the observed seismicity. This correlation can be used as a proxy to describe seismicity changes; however, further options should be considered and, ultimately, select the proxies that best correlate with the observed seismicity.

These correlations can be estimated in areas prone to induced seismicity in Alberta: (1) Rocky Mountain House; (2) Brazeau River; (3) Cardston area; (4) Duvernay Formation at Fox Creek; (5) Duvernay Formation at Red Deer. The number of induced events at the Rocky Mountain House and the Brazeau river has declined since they peaked in 1985 and 2006, respectively. Furthermore, the PSHA maps generated by Ghofrani et al. (2019) show a considerable reduction in the seismic hazard, in line with the reduction of the anthropogenic activity. For the case of the Cardston cluster, the seismicity finished once the hydraulic fracturing activities concluded, and the seismic hazard only peaked in this area for the year 2011 (see figure 6.3). Special attention is required in the Fox Creek and Red Deer areas. These areas are under special regulations (Subsurface orders no. 2 and 7), where operators are mandated to monitor adjacent seismic activities during the operations, and follow the traffic light protocol (Alberta Energy Regulator, 2019b,c). Currently, there are ongoing hydraulic fracturing activities in these areas prone to induced seismicity, and, as shown in the PSHA maps generated by Ghofrani et al. (2019), the Fox Creek area has shown the higher since hazard in the province since 2012.

6.3.3 Estimating appropriate proxies based on future anthropogenic activities

Correlations between operational parameters and induced seismicity can be used to forecast future seismicity scenarios. For instance, the curves resulting from correlations plots (see figure 6.4), can be projected into higher cumulative injection volumes and, given a future volume to inject, obtain the equivalent number of earthquakes. However, these correlations depend on future operational parameters, and the intensity of future operations depends on multiple

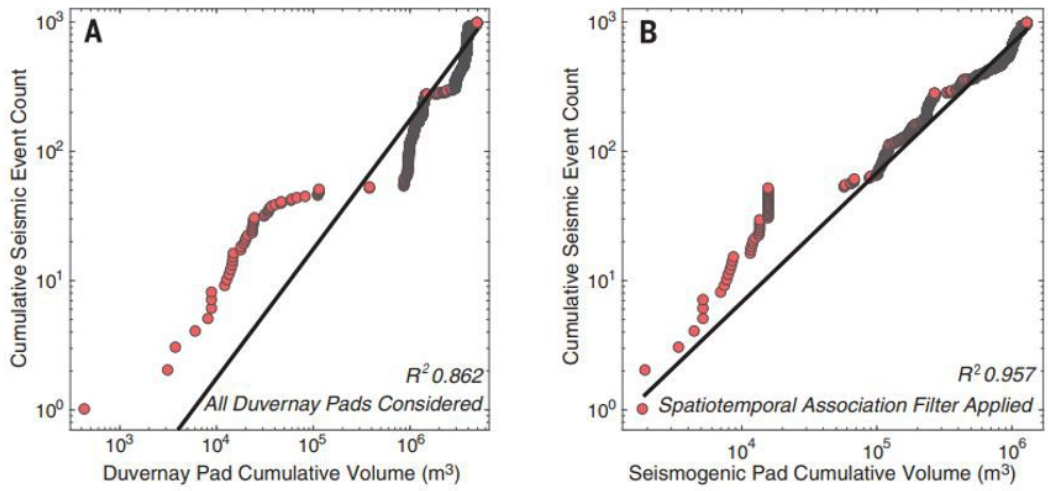

Figure 6.4: Number of earthquakes $M \geq 1.3$ versus cumulative injection volumes, in (A) all Duvernay hydraulic fracturing Pads considered, and (B) only seismogenic hydraulic fracturing pads considered. From Schultz et al. (2018).

factors, including changes in the technology of extraction of hydrocarbons, future economic trends, and governmental decisions. Therefore, it is important to establish correlations between operational parameters and factors that ultimately lead the anthropogenic activity, like economic and technical trends. For instance, Van der Baan and Calixto (2017) shown how the number of induced earthquakes in Oklahoma fell in early 2016 as a consequence of the reduction in injected volumes, which it turns to be originated by the 2015 oil price drop and regulations from governmental authorities in Oklahoma.

Forecasting future anthropogenic activities that depend on economic or technical trends is a challenge itself; however, multiple public and private institutions elaborate forecasts for the energy sector. For instance, the Alberta Energy Regulator (AER) issues annual reports that provide information about the state of reserves, supply, and demand outlook for the energy resources in Alberta. Figure 6.5 shows the Alberta average daily production of conventional crude oil by density. On the other hand, figure 6.6 shows the Alberta average daily production and number of conventional crude oil wells placed on total production. Notice that both trends show a rapid increase in production and the number of horizontal wells between 2011 and 2014, which finished in 2015 as a consequence of the oil price drop. Many of these horizontal wells

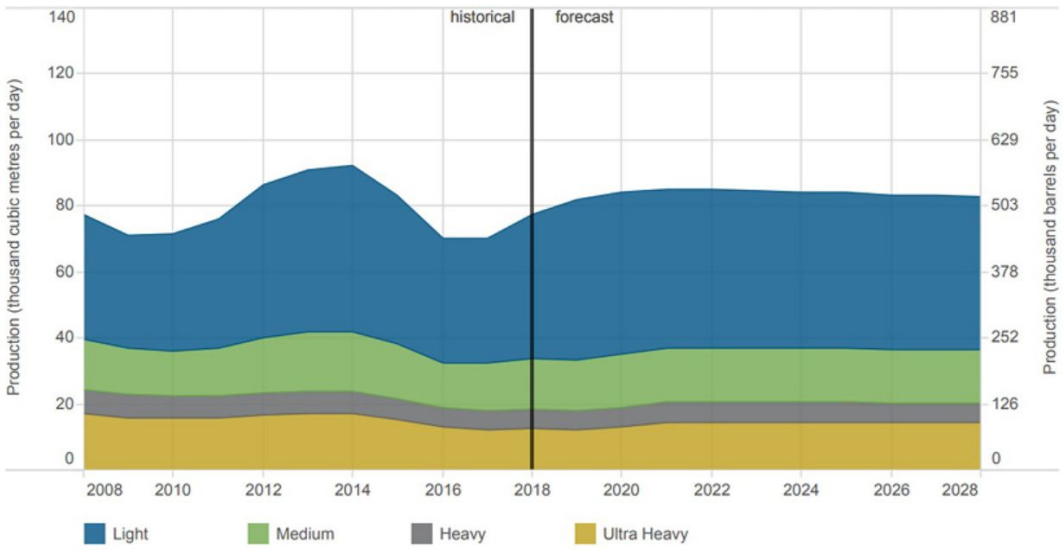

Figure 6.5: Historical and forecast average daily production of conventional crude oil by density in Alberta. In this chart, conventional crude oil production also includes the oil produced by hydraulic fracturing activities. From (Alberta Energy Regulator, 2019a)

are related to the increase of hydraulic fracturing treatments. However, these statistics show all horizontal and vertical wells in the province, and just a small number of wells are related to induced activity (Atkinson et al., 2016).

6.3.4 Seismicity scenario evaluation, physics-based models, and seismic hazard forecasts

If quantitative correlations between human activities (e.g., number of horizontal wells drilled) and operational parameters (e.g., injection volumes) are established in areas susceptible to induced seismicity, the forecasts of human activities can be used to build scenarios from the operational parameters. Ultimately, these scenarios of operational parameters will lead to future seismicity scenarios, using correlations between operational parameters and induced seismicity. For instance, we would expect a positive correlation between the number of wells and the cumulative injected volumes. In a seismogenic area, if the number of wells increases, the injection volumes will likely increase, and the expected number of induced earthquakes are going to increase as well. In

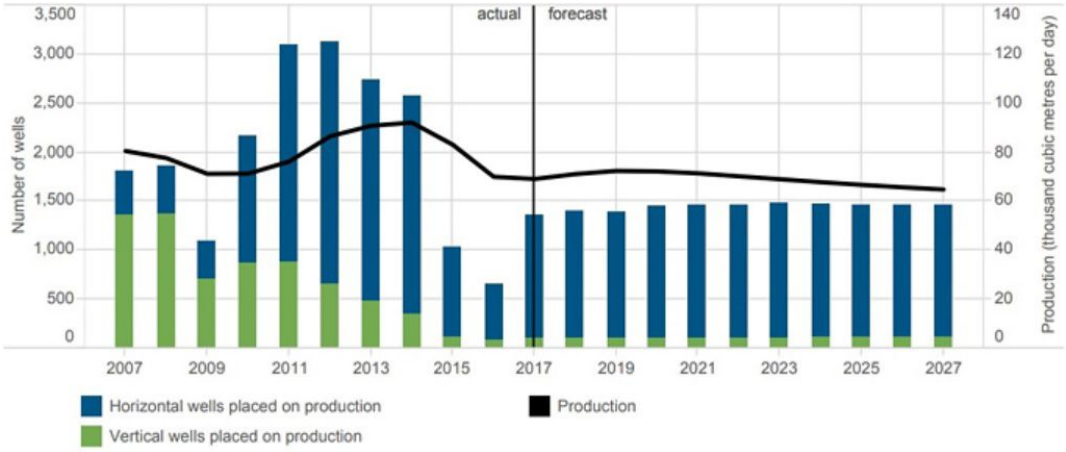

Figure 6.6: Historical and forecast average daily production and number of conventional crude oil wells placed on total production in Alberta. In this chart, conventional crude oil production also includes the oil produced by hydraulic fracturing activities. From Alberta Energy Regulator (2019a)

contrast, there are cases like the Horn River Basin, where the hydraulic fracturing activities stopped, and the induced events ceased shortly. In this case, the conclusion of the hydraulic fracturing activities was related to the 2012 drop in the gas prices (Van der Baan and Calixto, 2017). Finally, to capture better the uncertainty in the human activity forecast, at least three scenarios should be considered: low, medium (base), or high activity. Then, each human activity scenario will correspond to a low, medium, or high seismic activity case.

On the other hand, the physics-based model, like the seismogenic Index and the Hydromechanical Nucleation approach (Shapiro et al., 2010; Norbeck and Rubinstein, 2018) can be equally used to predict seismicity related to industrial activity. Both the seismogenic index model and the Hydromechanical Nucleation model allow time-dependent cumulative a-values. This cumulative a-value increases in line with the total volume injected. Therefore, and by using the appropriate correlations, the forecasts of human activities can be used to build scenarios of future injection volumes. Ultimately, the future injection volumes will dictate future seismicity rates. Both models have been used to forecast the number of earthquakes larger than $M \geq 3$ per month in Oklahoma (see figure 6.7). From our results in the Horn River Basin (see chapter 4), we found that the physics-based model can provide insights into

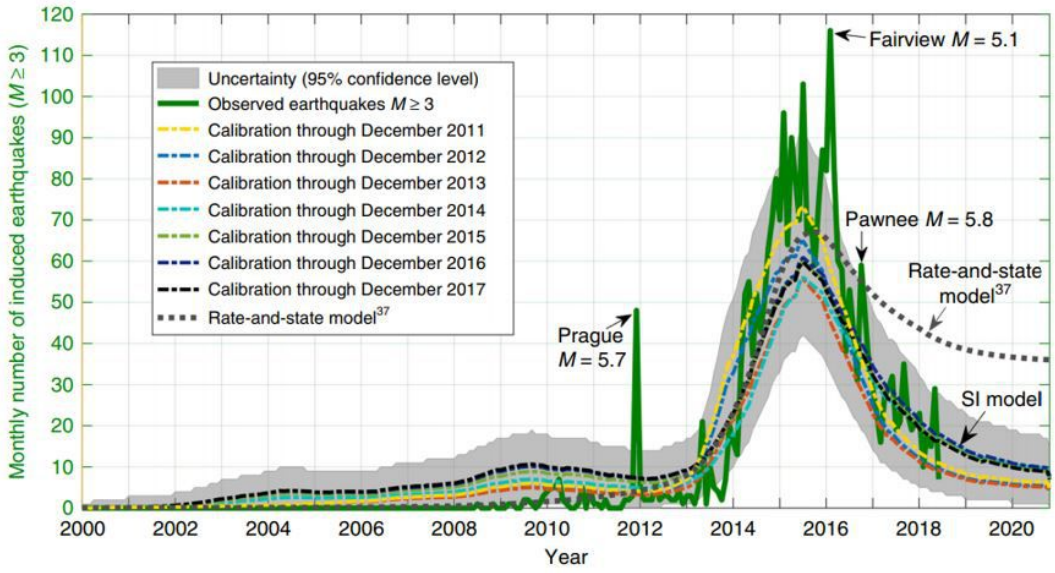

Figure 6.7: Forecasted seismicity rates (Earthquakes $M \geq 3$ per month)in Oklahoma, given by the Seismogenic Index model performed by Langenbruch et al. (2018). The figure shows multiple forecasted seismicity rates resulting from the Seismogenic Index model calibrated using different temporal endpoints (Dec. 2011 - Dec. 2017). From Langenbruch et al. (2018).

the temporal evolution of the overall seismicity patterns, and it can be used to build different seismicity scenarios as well. The decision of either using physics-based models or correlations between operational parameters and the number of induced events will depend on the performance of each method. A performance evaluation of both methods should include: (1) a comparison between predicted and observed number of earthquakes, and (2) a comparison of correlation coefficients between predicted and observed seismicity.

These seismicity scenarios, either built using correlations with operational parameters or physics-based models, can provide us insights into future seismicity patterns in areas known to be prone to induced seismicity. Government and industry can evaluate the areas that will likely experience an increase or decrease in induced events and, therefore, develop strategies based on future seismicity scenarios. For instance, if the seismicity scenarios indicate a clear increase in the number of earthquakes in areas prone to induced seismicity, it would be prudent to improve seismic monitoring by installing more stations. Also, operators could consider mitigation strategies beforehand in order to re-

duce the possible increase in the number of earthquakes and associated hazard. On the other hand, if the scenarios indicate a clear decrease in the number of earthquakes, installing more seismological stations might not be necessary, or a reduction could be recommended.

Building seismicity scenarios is one of the first steps to forecast seismic hazard for induced seismicity. From the predicted number of earthquakes, given by the seismicity scenarios, and by assuming a b-value, it is possible to make predictions of the likelihood to reach large magnitude events. For instance, after predicting the number of earthquakes in Oklahoma using a modified version of the Seismogenic Index, (Langenbruch and Zoback, 2016) predicted the probability to reach large magnitude events for the upcoming years. By knowing the changes in the a-values, they describe the changes in the number of earthquakes, and by assuming a given b-value, they use the GR magnitude-frequency distribution to estimate the rate of earthquakes of large magnitude events. Then, by knowing these rates and assuming a Poisson distribution, they predicted the probability of occurrence of large events (See chapter 3). However, assuming a constant b-value might not be appropriate; if the b-values change over time, the forecast of large events could be biased (See chapter 4). Therefore, any predictions of large magnitude events based on these current methods, should be taken cautiously.

Ultimately, once the GR parameters are more certain, particularly the b-values, a fully PSHA for induced seismicity can be developed. The reader is referred to the discussion section in Chapter 4, where we describe some possible strategies to predict short-term changes in the b-value by including the evaluation of geomechanical properties. As explained before, PSHA is used by governments and industry in applications for life and property safety, such as developing building code requirements, deciding the security criteria for critical facilities like dams, hydroelectric plants, nuclear plants, and determining earthquake insurance rates (Baker, 2008; Mulargia et al., 2017). Evaluating the hazard related to induced events is critical for a proper PSHA in areas where the anthropogenic activity is a major source of seismicity.

6.4 Discussion

Due to the changing nature of the induced seismicity and the uncertainties in the forecasts of human activity, I recommend the assessment of short-term seismicity scenarios in areas susceptible to induced earthquakes in Alberta. These seismicity scenarios can be updated annually by including revised: (1) retrospective PSHA maps, to quantify changes in the seismic hazard; (2) correlations between operational parameters and induced seismicity; (3) earthquake catalogs to calibrate the physics-based models; (4) Expected intensity of the human activity, and forecasts in the amount of operations. Considering long-term seismicity scenario building is too uncertain at this point. It should be clarified that these seismicity scenarios are built for global trends in relatively large areas affected by induced seismicity, and they are not intended to predict the seismicity scenarios for a particular operation. Also, successful mitigation strategies implemented by operators can alter the validation of the past correlations between operational parameters and induced seismicity. This is one of the reasons to keep updating correlations that reflects possible improvements in the mitigation of induced seismicity.

I also suggest annual updates of the PSHA maps. As mentioned before, these maps can provide useful insights into the temporal evolution of the seismic hazard related to anthropogenic activity. It can also provide short-term forecasts, as proposed by Petersen et al. (2016, 2017, 2018) for 2016, 2017, and 2018 seismic hazard maps in central and eastern United States. However, one of the problems of this model is that the forecasts are based on the seismicity from previous years. Since the peak of seismicity in 2015-2016, the hazard analysis for later years tends to be overestimated (Petersen et al., 2018), due to the continuous decline in the number of induced earthquakes. It should be noticed that the findings in chapter 5 about the necessity to perform seismicity declustering in an earthquake catalog can be included in this analysis. However, I should mention that the application of seismicity declustering might not be advisable due to the relatively low number of induced earthquakes in the province of Alberta, in comparison with other cases like Oklahoma.

Other approaches should be evaluated in order to identify areas susceptible to induced seismicity, as well as include additional proxies to describe seismicity

changes. For instance, Pawley et al. (2018) implemented a machine-learning algorithm to identify tectonic, geomechanical, and hydrological proxies suspected to control induced seismicity. They found that the proximity to the basement, in situ stress, proximity to fossil reef margins, lithium concentration, and rate of natural seismicity, are the geological factors with the highest correlation to well-activity associated with induced earthquakes. Pawley et al. (2018) also elaborated maps showing the areas with higher seismic activation potential in the Duvernay Formation, Alberta, based on the subsurface proxies mentioned before. Therefore, these maps can be used to delineate the areas that may be prone to induced earthquakes. In addition, if there are quantitative relations between these subsurface proxies and the number of induced events, they can be used to describe seismicity changes in areas prone to induced earthquakes.

Building seismicity scenarios is one of the first steps to fully forecast seismic hazard for induced seismicity. PSHA for future induced seismicity requires precise predictions of GR parameters. This is one of the reasons to establish correlations between operational parameters and seismicity patterns, as well as introduce physics-based models in the seismic hazard analysis. Both methods can be used to generate seismicity rates that depend on human activities. However, these models are limited to short-term forecasts (from a couple of months up to a year), due to the non-stationarity behavior of seismic parameters like the b-value. This is perhaps one of the biggest challenges in PSHA for induced seismicity that has to be properly addressed. Some possible strategies to predict short-term changes in the b-value include identification and evaluation of geomechanical proxies (See discussion chapter 4). An additional limitation of these seismicity scenarios is that the new induced seismicity areas, with no historical seismicity, would be rated with low hazard. This is one of the reasons to update the seismicity scenarios annually, and include the new areas with induced seismicity.

6.5 Conclusions

In this chapter, I first present an overview of the reported induced seismicity cases in Alberta. Then, I describe a methodology to build seismicity scenarios

in areas susceptible to induced earthquakes. This methodology is based on (1) Identifying areas that have experienced considerable seismicity changes, using retrospective PSHA maps; (2) Identify the operational parameters that can act as proxies to foreseen seismicity changes in areas prone to induced earthquakes; (3) Project future anthropogenic operations, in particular, the operational parameters correlated to induced activity; (4) Build future seismicity scenarios in areas prone to induced seismicity, using the identified proxies and projected future anthropogenic operations. These short-term (e.g., annual) induced seismicity scenarios can give us insights into future seismicity patterns based on projected operations in areas known to be prone to induced seismicity.

Government and industry can evaluate the areas that will likely experience an increase or decrease in induced events and, therefore, develop future strategies based on seismicity scenarios. For instance, if the seismicity scenarios indicate a clear increase in the number of earthquakes in areas prone to induced seismicity, it would be necessary to improve the seismic monitoring or design mitigation strategies to reduce the number of earthquakes and associated hazard. Building seismicity scenarios is also one of the first steps to fully develop seismic hazard forecasts for induced seismicity. Current methods can be used to generate seismicity rates that depend on human activities, but they are limited due to the non-stationarity behavior of seismic parameters like the b-value. Further studies are required in order to obtain precise predictions of GR parameters, which is currently one of the biggest challenges in PSHA for induced seismicity.

Chapter 7

Conclusions and suggested direction for future research

Due to the increase in human-induced seismicity, specially in geologically stable basins in North America, it has been necessary to quantify the seismic hazard related to anthropogenic activity. In this thesis, we developed a PSHA methodology capable of quantifying the seismic hazard related to induced seismicity. We identify the challenges in the implementation of PSHA for induced events, including: (1) the non-stationary behavior of the induced seismicity, (2) the estimation of future GR parameters for induced seismicity, (3) the debate around applying seismicity declustering.

In chapter 3, we developed a methodology to compute synthetic earthquake catalogs for non-stationary seismicity using Monte Carlo simulations. We also developed the associated occurrence earthquake statistics for non-stationary seismicity like induced seismicity. Non-stationary Gutenberg-Richter (GR) parameters are included to assess the hazard for this type of seismicity. In both synthetic and real case examples, tests showed agreements between analytical predictions and numerical results.

In chapter 4, we applied two methodologies to predict the GR parameters related to injection-induced seismicity. We make a complete seismic hazard analysis using GR parameters from physics-based models (Seismogenic Index and Hydromechanical Nucleation approach) and the GR parameters given by

the observed seismicity. Our preliminary results show that the predictions of both methods follow the observed induced seismicity patterns, particularly the increase and decrease in the monthly number of earthquakes, in line with the anthropogenic activity. However, by comparing the annual GR parameters, the predictions tend to either underestimate or overestimate the seismic hazard, mainly due to the assumption of stationary b-values.

In chapter 5, we identified the cases where seismicity declustering is advisable. If mainshocks and aftershocks have considerably different b-values, declustering leads to improved hazard assessments, since it allows for better estimations of magnitude-frequency distribution of the largest events. On the other hand, if mainshocks and aftershocks have similar b-values, declustering is not recommended since it eliminates large magnitude events that significantly contribute to the seismic hazard. We also show that assuming Poissonian distributions in hazard predictions does not lead to inaccurate long-term hazard predictions, even if time-varying aftershock sequences are present in the catalog.

7.1 Direction for future research

There are still multiple challenges that have to be properly addressed, including the prediction of the maximum magnitude for induced events, or the development of more accurate physics-based models for the estimation of earthquake occurrence. Below, we listed some of the suggested directions for future research:

1. It is necessary to improve the forecasting of GR parameters for injection-induced earthquakes, particulary the b-value. The Seismogenic Index and the Hydromechanical Nucleation models are useful tools for the prediction of overall seismicity rates. However, the predicted GR parameters tend to show biased results. One of the reasons is that both models rely on constant b-values, which may not be appropriate for short-term seismicity. I think that future physics-based models should allow temporal variation of the b-values. Also, a better understanding of the stress conditions could lead to a better hazard prediction. Some studies (Scholz, 2015;

Amitrano, 2003; Goebel et al., 2013; Van der Baan and Chorney, 2019) have found that the b-values are anticorrelated to differential stress $\sigma_1 - \sigma_3$. Therefore, the monitoring and modelling of geomechanical properties, like differential stress or stress-drops, could work as a proxy to predict changes in the b-value.

2. Another critical aspect that should be addressed is the maximum magnitude M_{max} related to injection-induced seismicity. Different approaches have been proposed to predict the maximum magnitude M_{max}, including: (1) geometrical approaches based on inferred dimensions of the stimulated reservoir volume (Shapiro et al., 2011); (2) a linear relationship between maximum seismic moment and total injected volume (McGarr, 2014); and (3) a probabilistic approach based in seismicity rates (Van der Elst et al., 2016). Future seismic hazard studies should evaluate different expected maximum magnitudes given by these approaches. However, it is important to consider the geological features that limit the maximum magnitude of the earthquakes, like the surface area of the seismogenic faults.

3. I recommend the assessment of short-term seismicity scenarios in areas susceptible to induced earthquakes in Alberta, as described in chapter 6. This type of scenario building, based on projected operations in areas prone to induced seismicity, can give us insights into future seismicity patterns. Building seismicity scenarios is one of the first steps to fully forecast seismic hazard for induced seismicity. I also recommend the assessment of retrospective annual seismic hazard analysis in areas prone to induced earthquakes, like the province of Alberta. These annual seismic hazard assessments can provide useful insights into the temporal evolution of the seismic hazard related to anthropogenic activity.

4. I recommend the assessment of annual seismic hazard maps in areas prone to induced earthquakes, like the province of Alberta. These annual seismic hazard assessments can provide useful insights into the temporal evolution of the seismic hazard related to anthropogenic activity as well as quantitatively evaluate the seismic hazard changes. Physics-based mod-

els can be included in the hazard analysis; however, they should be taken cautiously due to changes in the seismic parameters. A good strategy is to monitor the induced seismicity continuously and detect sudden changes in the GR parameters; thereafter, update the seismic hazard. These annual maps can be useful tools for regulators, operators, and the general public to evaluate the anthropogenic operations causing induced seismicity.

Bibliography

Aki, K., 1965, Maximum likelihood estimate of b in formula logN=a-bM and its confidence limits: Bulletin of the Earthquake Research Institute, **43**, 237–239.

Akinci, A., M. P. Moschetti, and M. Taroni, 2018, Ensemble smoothed seismicity models for the new Italian probabilistic seismic hazard map: Seismological Research Letters, **89**, 1277–1287.

Alberta Energy Regulator, 2015, Subsurface order no. 2: URL: https://www.aer.ca/documents/orders/subsurface-orders/SO2.pdf.

———, 2019a, ST98: Alberta Energy Outlook: https://www.aer.ca/providing-information/data-and-reports/statistical-reports/st98.

———, 2019b, Subsurface order no. 6: URL: https://www.aer.ca/documents/orders/subsurface-orders/SO6.pdf.

———, 2019c, Subsurface order no. 7: URL: https://www.aer.ca/documents/orders/subsurface-orders/SO7.pdf.

Alberta Geological Survey, 2019, Interactive maps: Alberta Earthquakes (2006-2019): https://ags.aer.ca/data-maps-models/interactive-maps.htm.

Amitrano, D., 2003, Brittle-ductile transition and associated seismicity: Experimental and numerical studies and relationship with the b value: Journal of Geophysical Research: Solid Earth, **108**, 1–15.

Anagnos, T., and A. S. Kiremidjian, 1988, A review of earthquake occurrence models for seismic hazard analysis: Probabilistic Engineering Mechanics, **3**, 3–11.

Assatourians, K., and G. Atkinson, 2013, EqHaz: An open-source probabilistic seismic-hazard code based on the Monte Carlo simulation approach: Seismological Research Letters, **84**, 516–524.

———, 2019, Implementation of a smoothed-seismicity algorithm in Monte Carlo PSHA software EQHAZ and implications for localization of hazard in the Western Canada sedimentary basin: Seismological Research Letters, **90**, 1407–1419.

Atkinson, G., 2015, Ground-motion prediction equation for small-to-moderate events at short hypocentral distances, with application to induced-seismicity hazards: Bulletin of the Seismological Society of America, **105**, 981–992.

Atkinson, G., H. Ghofrani, and K. Assatourians, 2015, Impact of induced seismicity on the evaluation of seismic hazard: Some preliminary considerations: Seismological Research Letters, **86**, 1009–1021.

Atkinson, G. M., D. W. Eaton, H. Ghofrani, D. Walker, B. Cheadle, R. Schultz, R. Shcherbakov, K. Tiampo, J. Gu, R. M. Harrington, Y. Liu, M. Van der Baan, and H. Kao, 2016, Hydraulic Fracturing and Seismicity in the Western Canada Sedimentary Basin: Seismological Research Letters, **87**, 632–647.

Baker, J., 2008, An Introduction to Probabilistic Seismic Hazard Analysis (PSHA): Technical report, Stanford Education. URL: https://web.stanford.edu/~bakerjw/Publications/Baker_(2008)_Intro_to_PSHA_v1_3.pdf.

———, 2013, Introduction To Probabilistic Seismic Hazard Analysis: Technical report, Stanford Education.

Baranova, V., A. Mustaqeem, and S. Bell, 1999, A model for induced seismicity caused by hydrocarbon production in the Western Canada Sedimentary Basin: Canadian Journal of Earth Sciences, **36**, 47–64.

BC Oil and Gas Commission, 2012, Investigation of Observed Seismicity in the Horn River Basin: Technical report.

———, 2014, Horn River Basin Unconventional Shale Gas Play Atlas: Technical report.

Bourne, S., S. Oates, J. Bommer, B. Dost, J. Van Elk, and D. Doornhof, 2015, A Monte Carlo Method for Probabilistic Hazard Assessment of Induced Seismicity due to Conventional Natural Gas Production: Bulletin of the Earthquake Research Institute, **105**, 1721–1738.

Bourne, S., S. Oates, and J. Van Elk, 2018, in the Groningen gas field and its implications for controlling seismic: Geophysical Journal International, **213**,

1693–1700.

Bourne, S. J., S. J. Oates, J. Van Elk, and D. Doornhof, 2014, A seismological model for earthquakes induced by fluid extraction from a subsurface reservoir: Journal of Geophysical Research: Solid Earth, **119**, 8991–9015.

Boyd, O., 2012, Including foreshocks and aftershocks in time-independent probabilistic seismic-hazard analyses: Bulletin of the Seismological Society of America, **102**, 909–917.

Brodsky, E., and L. Lajoie, 2013, Anthropogenic seismicity rates and operational parameters at the Salton Sea Geothermal Field: Science, **341**, 543–546.

Convertito, V., N. Maercklin, N. Sharma, and A. Zollo, 2012, From induced seismicity to direct time-dependent seismic hazard: Bulletin of the Seismological Society of America, **102**, 2563–2573.

Cornell, C., 1968, Engineering seismic risk analysis: Bulletin of the Seismological Society of America, **58**, 1583–1606.

Daley, D. J., and D. Vere-Jones, 2003, An introduction to the theory of Point processes: Volume 1: Elementary theory and methods, second edi ed.: Springer.

Dieterich, J. H., 1994, A constitutive law for rate of earthquake production and its application to earthquake clustering: Journal of Geophysical Research: Solid Earth, **99**, 2601–2618.

Ellsworth, W., 2013, Injection-Induced Earthquakes: Science, **341**, 1–8.

Farahbod, A., H. Kao, J. F. Cassidy, and D. Walker, 2015a, How did hydraulic-fracturing operations in the Horn River Basin change seismicity patterns in northeastern British Columbia , Canada ?: The Leading Edge, **34**, 658–663.

Farahbod, A., H. Kao, D. Walker, and J. F. Cassidy, 2015b, Investigation of regional seismicity before and after hydraulic fracturing in the Horn River Basin , northeast British Columbia: Canadian Journal of Earth Sciences, 112–122.

Galloway, E., T. Hauck, H. Corlett, D. Pană, and R. Schultz, 2018, Faults and associated karst collapse suggest conduits for fluid flow that influence hydraulic fracturinginduced seismicity: Proceedings of the National Academy of Sciences of the United States of America, **115**, 10003–10012.

Gardner, J. K., and L. Knopoff, 1974, Is the sequence of earthquakes in southern California, with aftershocks removed, Poissonian?: Bulletin of the Seismological Society of America, **64**, 1363–1367.

Ghofrani, H., G. Atkinson, R. Schultz, and K. Assatourians, 2019, Short-Term Hindcasts of Seismic Hazard in the Western Canada Sedimentary Basin Caused by Induced and Natural Earthquakes: Seismological Research Letters, **90**, 1420–1435.

Ghofrani, H., and G. M. Atkinson, 2016, A preliminary statistical model for hydraulic fracture-induced seismicity in the Western Canada Sedimentary Basin: Geophysical Research Letters, **43**, 164–172.

Goebel, T. H., D. Schorlemmer, T. W. Becker, G. Dresen, and C. G. Sammis, 2013, Acoustic emissions document stress changes over many seismic cycles in stick-slip experiments: Geophysical Research Letters, **40**, 2049–2054.

Gulia, L., and S. Wiemer, 2019, Real-time discrimination of earthquake foreshocks and aftershocks: Nature, **574**, 193–199.

Gutenberg, B., and C. Richter, 1944, Frequency of earthquakes in California: Bulletin of the Seismological Society of America, **34**, 591–610.

Halchuk, S., T. I. Allen, J. Adams, G. C. Rogers, S. Halchuk, T. I. Allen, J. Adams, and G. C. Rogers, 2014, Fifth Generation Seismic Hazard Model Input Files as Proposed to Produce Values for the 2015 National Building Code of Canada: Technical report, Geological Survey of Canada.

Harte, D., 2013, Bias in fitting the ETAS model: a case study based on the New Zeland Seimicity: Geophysical Journal International, **192**, 390–412.

Hincks, T., W. Aspinall, R. Cooke, and T. Gernon, 2018, Oklahoma's induced seismicity strongly linked to wastewater injection depth: Science, **1255**, 1251–1255.

Hornbach, M., H. Deshon, W. Ellsworth, B. Stump, C. Hayward, C. Frohlich, H. Oldham, J. Olson, M. Magnani, C. Brokaw, and J. Luetgert, 2015, Causal factors for seismicity near Azle , Texas: Nature Communications, **6**, 1–11.

Horne, R., 1995, Modern well test analysis: A computer aided approach, 2nd editio ed.: Palo Alto, California. Petroway, Inc.

Kao, H., R. Visser, B. Smith, and S. Venables, 2018, Performance assessment of the induced seismicity traffic light protocol for northeastern British Columbia

and western Alberta: Leading Edge, **37**, 117–126.

Keranen, K. M., M. Weingarten, G. Abers, B. A. Bekins, and S. Ge, 2014, Sharp increase in central Oklahoma seismicity since 2008 induced by massive wastewater injection: Science, **345**, 448–451.

Knopoff, L., Y. Kagan, and R. Knopoff, 1982, b values for foreshocks and aftershocks in real and simulated earthquake sequences: Bulletin of the Seismological Society of America, **72**, 1663–1676.

Kramer, S., 1996, Geotechnical Earthquake Engineering: Prentice Hall.

Langenbruch, C., C. Dinske, and S. A. Shapiro, 2011, Inter event times of fluid induced earthquakes suggest their Poisson nature: Geophysical Research Letters, **38**, 1–6.

Langenbruch, C., M. Weingarten, and M. D. Zoback, 2018, Physics-based forecasting of man-made earthquake hazards in Oklahoma and Kansas: Nature Communications, **9**, 1–10.

Langenbruch, C., and M. D. Zoback, 2016, How will induced seismicity in Oklahoma respond to decreased saltwater injection rates?: Science Advances, **3**, 1–10.

Lombardi, M., 2003, The Maximum Likelihood estimator of b-value for mainshocks: Bulletin of the Seismological Society of America, **93**, 2082–2088.

Marzocchi, W., and M. Taroni, 2014, Some thoughts on declustering in probabilistic seismic-hazard analysis: Bulletin of the Seismological Society of America, **104**, 1838–1845.

McGarr, A., 2014, Maximum magnitude earthquake induced by fluid injection: Journal of Geophysical Research, **119**, 1008–1019.

McGarr, A., D. Simpson, and L. Seeber, 2002, Case histories of induced and triggered seismicity: International Geophysics, **81**, 647–661.

Meredith, P., I. Main, and C. Jones, 1990, Temporal variations in seismicity during quasi-static and dynamic rock failure: Tectonophysics, **175**, 249–268.

Mulargia, F., P. B. Stark, and R. J. Geller, 2017, Why is Probabilistic Seismic Hazard Analysis (PSHA) still used?: Physics of the Earth and Planetary Interiors, **264**, 63–75.

Musson, R., 2000, The use of Monte Carlo simulations for seismic hazard assessment in the UK: Annali di Geofisica, **43**, 1–9.

Natural Resources Canada, 2015, 2015 National Building Code of Canada seismic hazard maps: URL: http://earthquakescanada.nrcan.gc.ca/hazard-alea/zoning-zonage/NBCC2015maps-en.php.

Norbeck, J. H., and J. L. Rubinstein, 2018, Hydromechanical Earthquake Nucleation Model Forecasts Onset, Peak, and Falling Rates of Induced Seismicity in Oklahoma and Kansas: Geophysical Research Letters, **45**, 2963–2975.

Ogata, Y., 1988, Statistical models for earthquake occurrences and residual analysis for point processes: Journal of the American Statistical Association, **83**, 9–27.

———, 1998, Space-time point-process models for earthquake occurrences: Annals of the Institute of Statistical Mathematics, **50**, 379–402.

Ogata, Y., and J. Zhuang, 2006, Space – time ETAS models and an improved extension: Tectonophysics, Elseviver, **413**, 13–23.

Omori, F., 1894, On Aftershocks of earthquakes: Journal of the College of Science, Imperial University of Tokio, **7**, 111–120.

Pawley, S., R. Schultz, T. Playter, H. Corlett, T. Shipman, S. Lyster, and T. Hauck, 2018, The Geological Susceptibility of Induced Earthquakes in the Duvernay Play: Geophysical Research Letters, **45**, 1786–1793.

Petersen, M., C. S. Mueller, M. P. Moschetti, S. M. Hoover, A. L. Llenos, W. L. Ellsworth, A. J. Michael, J. L. Rubinstein, A. F. McGarr, and K. S. Rukstales, 2016, 2016 One-Year Seismic Hazard Forecast for the Central and Eastern United States from Induced and Natural Earthquakes: USGS Open File Report, **2016-1035**, 1–50.

Petersen, M., C. S. Mueller, M. P. Moschetti, S. M. Hoover, K. S. Rukstales, D. E. McNamara, R. A. Williams, A. M. Shumway, P. M. Powers, P. S. Earle, A. L. Llenos, A. J. Michael, J. L. Rubinstein, J. H. Norbeck, and E. S. Cochran, 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.

———, 2018, 2018 one-year seismic hazard forecast for the central and eastern United States from induced and natural earthquakes: Seismological Research Letters, **89**, 1049–1061.

Pollard, D., and R. Fletcher, 2005, Fundamentals of Structural Geology: Cam-

bridge University Press.

Reasenberg, P., 1985, Second-order moment of central California seismicity, 1969-82: Journal of Geophysical Research, **90**, 5479–5495.

Reyes Canales, M., and M. Van der Baan, 2019, Including Non-stationary Magnitude-Frequency distributions in Probabilistic Seismic Hazard Analysis: Pure and Applied Geophysics, **176**, 2299–2319.

Roche, V., M. Grob, T. Eyre, and M. Van der Baan, 2015, Statistical characteristics of microseismic events and in-situ state of stress in the Horn River Basin: Presented at the 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.

———, 2015, On the stress dependence of the earthquake b value: Geophysical Research Letters, **42**, 1399–1402.

Schultz, R., G. Atkinson, D. W. Eaton, Y. J. Gu, and H. Kao, 2018, Hydraulic fracturing volume is associated with induced earthquake productivity in the duvernay play: Science, **359**, 304–308.

Schultz, R., S. Mei, D. Pană, V. Stern, Y. Gu, A. Kim, and D. Eato, 2015, The cardston earthquake swarm and hydraulic fracturing of the exshaw formation (Alberta bakken play): Bulletin of the Seismological Society of America, **105**, 2871–2884.

Schultz, R., V. Stern, and Y. Gu, 2014, Journal of Geophysical Research : Solid Earth relation to a nearby disposal well: Journal of Geophysical Research: Solid Earth, **119**, 3410–3423.

Schultz, R., R. Wang, Y. Gu, K. Haug, and G. Atkinson, 2017, A seismological overview of the induced earthquakes in the Duvernay play near Fox Creek, Alberta: Journal of Geophysical Research: Solid Earth, **122**, 492–505.

Segall, P., 1989, Earthquakes triggered by fluid extraction: Geology, **17**, 942–946.

Segall, P., and S. Lu, 2015, Injection-induced seismicity : Poroelastic and earthquake nucleation effects: Journal of Geophysical Research: Solid Earth, **120**, 5082–5103.

Shapiro, S., C. Dinske, C. Langenbruch, and F. Wenzel, 2010, Seismogenic index and magnitude probability of earthquakes induced during reservoir

fluid stimulations: Leading Edge, **29**, 304–309.

Shapiro, S., O. Krüger, C. Dinske, and C. Langenbruch, 2011, Magnitudes of induced earthquakes and geometric scales of fluid-stimulated rock volumes: Geophysics, **76**, 55–64.

Shi, Y., and B. Bolt, 1982, The standart 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: Technical report, Columbia Education.

Snowdon, W., 2019, 'A very short, sharp shock': Earthquake rattles central Alberta (CBC news): URL: https://www.cbc.ca/news/canada/edmonton/earthquake-red-deer-alberta-1.5041580.

Stern, V., R. Schultz, L. Shen, Y. Gu, and D. Eaton, 2013, Alberta Earthquake Catalogue, Version 1.0: September 2006 through December 2010: AER report.

Suyehiro, S., and H. Sekiya, 1972, Foreshocks and Earthquake Prediction: Tectonophysics, Elseviver, **14**, 219–225.

Teng, B. G., and J. W. Baker, 2019, Seismicity declustering and hazard analysis of the Oklahoma–Kansas Region: Bulletin of the Seismological Society of America, **109**, 2356–2366.

Utsu, T., 1961, Statistical study on the occurrence of aftershocks: Geophys. Mag., **30**, 521–605.

———, 1966, A statistical significance Test of the difference in b-value between two earthquake groups: Journal of Physics of Earth, **14**, 37–40.

Utsu, T., Y. Ogata, and R. S. Matsu'ura, 1995, The Centenary a Decay of the Omori Formula Activity for Law of Aftershock: J. Phys. Earth, **43**, 1–33.

Van der Baan, M., and F. Calixto, 2017, Human-induced seismicity and large-scale hydrocarbon roduction in the USA and Canada: Geochemistry Geophysics Geosystems, **18**, 2467–2485.

Van der Baan, M., and D. Chorney, 2019, Insights from micromechanical modeling of intact rock failure: Event characteristics, stress drops and force networks: Journal of Geophysical Research: Solid Earth, **124**, 12,955–12,980.

Van der Elst, N., M. Page, D. Weiser, T. Goebel, and S. Hosseini, 2016, Jour-

nal of Geophysical Research : Solid Earth with permeability measurement: Journal of Geophysical Research: Solid Earth, **121**, 4575–4590.

Van Stiphout, T., J. Zhuang, and D. Marsan, 2012, Models and Techniques for Analyzing Seismicity: Community Online Resource for Statistical Seismicity Analysis, **Theme V**, 1–25. doi:10.5078/corssa– 52382934.

Wetmiller, R. J., 1986, Earthquakes near Rocky Mountain House, Alberta, and their relationship to gas production facilities (Canada).: Canadian Journal of Earth Sciences, **23**, 172–181.

Wiemer, S., and M. Wyss, 1997, Mapping the frequency-magnitude distribution in asperities: An improved technique to calculate recurrence times?: Journal of Geophysical Research: Solid Earth, **102**, 15115–15128.

Wyss, M., 1979, Estimating maximum expectable magnitude of earthquakes from fault dimensions: Geology, 336–340.

Zhuang, J., D. Harte, M. J. Werner, S. Hainzl, and S. Zhou, 2012, Basic models of seismicity: Temporal models Basic models of seismicity: temporal models, Community Online Resource for Statistical Seismicity.

Zhuang, J., Y. Ogata, and D. Vere-Jones, 2002, Stochastic declustering of space-time earthquake occurrences: Journal of the American Statistical Association, **97**, 369–380.

Zhuang, J., and S. Touati, 2015, Theme V - Models and Techniques for Analyzing Seismicity Stochastic simulation of earthquake catalogs: Community Online Resource for Statistical Seismicity Analysis, 1–34.