Posted on 30 Nov 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.1002/essoar.10503731/v2 — This a preprint and has not been peer reviewed. Data may be preliminary.

Statistical Modeling and Characterization of Induced Seismicity within the Western Canada Sedimentary Basin

Sidhanth Kothari[1,1], Robert Shcherbakov[2,2], and Gail M. Atkinson[1,1]

[1]University of Western Ontario
[2]Western University

November 30, 2022

Abstract

In western Canada, there has been an increase in seismic activity linked to anthropogenic energy-related operations including conventional hydrocarbon production, wastewater fluid injection, and, more recently, hydraulic fracturing (HF). Statistical modeling and characterization of the space, time, and magnitude distributions of the seismicity clusters is vital for a better understanding of induced earthquake processes and development of forecasting models. In this work, a statistical analysis of the seismicity in the Western Canada Sedimentary Basin was performed across past and present time periods by utilizing a compiled earthquake catalogue for Alberta and eastern British Columbia. Specifically, the inter-event space-time distance distributions of earthquakes were studied using the nearest-neighbour distance (NND) method. Additionally, the frequency-magnitude statistics and aftershock parameters of several clusters were analyzed using the Gutenberg-Richter relation and the epidemic type aftershock sequence model. The results suggest that recent regional changes in the NND distributions, namely, a disproportionate increase in loosely and tightly clustered seismic activity over time, are unnatural and likely due to the rise in HF operations for the development of unconventional resources. It is concluded that both these loosely and tightly clustered earthquake subpopulations differ measurably from what may be the region's tectonic seismic activity. Additionally, HF treatments have a greater probability of triggering swarm-like sequences that sharply spike the seismicity rate and are characterized by larger Gutenberg-Richter b-values. In contrast, conventional production and wastewater disposal operations largely trigger loosely clustered activity with more typical magnitude-occurrence distributions.

Statistical Modeling and Characterization of Induced Seismicity within the Western Canada Sedimentary Basin

Sidhanth Kothari[1], Robert Shcherbakov[1], and Gail Atkinson[1]

[1]Western University, London, Ontario, Canada

Corresponding author: Sidhanth Kothari (skothar3@uwo.ca)

Key Points:

- Parts of western Canada have seen an increase in clustered seismic activity coinciding with the growing usage of hydraulic fracturing
- The regional seismicity can be subdivided into three main classes of inter-event distance using the nearest-neighbour distance method
- Analysis of earthquake clusters occurring near different types of anthropogenic operations reveal key differences in scaling and structure

Abstract

In western Canada, there has been an increase in seismic activity linked to anthropogenic energy-related operations including conventional hydrocarbon production, wastewater fluid injection and more recently hydraulic fracturing (HF). Statistical modeling and characterization of the space, time and magnitude distributions of the seismicity clusters is vital for a better understanding of induced earthquake processes and development of forecasting models. In this work, a statistical analysis of the seismicity in the Western Canada Sedimentary Basin was performed across past and present time periods by utilizing a compiled earthquake catalogue for Alberta and eastern British Columbia. Specifically, the inter-event space-time distance distributions of earthquakes were studied using the nearest-neighbour distance (NND) method. Additionally, the frequency-magnitude statistics and aftershock parameters of several clusters were analyzed using the Gutenberg-Richter relation and the epidemic type aftershock sequence model. The obtained results suggest that recent regional changes in the NND distributions, namely a disproportionate increase in loosely and tightly clustered seismic activity over time, are unnatural and likely due to the rise in HF operations for the development of unconventional resources. It is concluded that both these loosely and tightly clustered earthquake subpopulations differ measurably from what may be the region's tectonic seismic activity. Additionally, HF treatments have a greater probability of triggering swarm-like sequences that sharply spike the seismicity rate and are characterized by steeper frequency-magnitude distributions. Conventional production and wastewater disposal operations largely trigger loosely clustered activity with more typical magnitude-occurrence rates.

Plain Language Summary

In western Canada, there has been an increase in earthquake activity linked to industrial activities including fossil fuel extraction, wastewater disposal and more recently hydraulic fracturing. Statistical modeling of earthquake phenomena is important for the understanding of the specific mechanisms involved in triggering earthquakes. In this work, a statistical analysis of the recorded earthquakes in the Western Canada Sedimentary Basin is performed. The results of this study suggest that there are discrete statistical differences between the natural, tectonic earthquakes and those triggered artificially by human activity. Additionally, hydraulic fracturing operations appear capable of triggering swarm-like sequences that temporarily increase the

earthquake rate and tend to occur at small-to-moderate magnitudes. Conventional fuel production and wastewater disposal operations largely trigger earthquakes that are loosely clustered together in space and time across a broader magnitude range.

1 Introduction

1.1 Induced Seismicity

Human activities, such as hydrocarbon production, reservoir impoundment, mining and geothermal energy extraction, can alter subsurface stress regimes through a variety of mechanisms, including the withdrawal or injection of fluid, reservoir compaction, excess surface loading, and ground subsidence (Grigoli et al., 2017; Doglioni, 2018; Keranen & Weingarten, 2018). In some cases, these stress perturbations result in earthquakes, particularly in areas characterized by higher states of stress and/or preexisting, well-oriented faults. This phenomenon is referred to as induced seismicity and is contributing to an increase in seismic hazard in certain parts of the world (Segall et al., 1995; Lei et al., 2008; Atkinson, 2017; McClure et al., 2017; Eaton et al., 2018; Brudzinski & Kozłowska, 2019).

In the central United States, for example, numerous earthquakes have been attributed to wastewater disposal wells operating in close proximity (Ellsworth, 2013; Llenos & Michael, 2013; van der Elst et al., 2013; Keranen et al., 2014; Hornbach et al., 2016; Schoenball & Ellsworth, 2017). These wells inject large quantities of excess flow-back fluid from associated oil and gas production operations deep into underground reservoirs. This creates a pore pressure front that travels outward and may interact with fault structures, particularly in the crystalline basement where most of the associated seismic events occur (Horton, 2012; Ellsworth, 2013; Keranen et al., 2014; McClure et al., 2017; Shah & Keller, 2017). In the year 2016 alone, Oklahoma observed three earthquakes of moment magnitude (**M**) >5, including an **M**5.8 event in Pawnee that was the largest event recorded in the state’s history. All three events occurred within 10 km of wastewater disposal wells and had moment releases that scaled with the net volume of near-field injection (McGarr & Barbour, 2017).

In parts of western Canada, there has been a notable rise in the seismic rate coinciding with the implementation of unconventional extraction technology developed for the production of oil and gas, known as hydraulic fracturing (HF) or “fracking” (B.C. Oil and Gas Commission, 2012, 2014; Schultz, Stern, Novakovic, et al., 2015; Atkinson et al., 2016; Bao & Eaton, 2016; Deng et al., 2016; Ghofrani & Atkinson, 2016; Schultz et al., 2016, 2017, 2018; Wang et al., 2016; Eaton et al., 2018; Zhang et al., 2019). During the HF process, fractures are created or enhanced within a target formation holding desired hydrocarbons, typically tight (low permeability) sedimentary layers, by the pumping of chemical slurry into segments of the rock over several stages. It is increasingly common for wells to be drilled at a deviated or horizontal angle as they approach a reservoir, in order to engage a larger portion of the source rock than would have been reached vertically. These technological advancements have prompted a dramatic growth in the number of possible fracture stages per wellbore as well as increased the average total volume of high-pressure fluid injected (King, 2010) and its areal extent. Between 2010 and 2018, approximately 20,000 HF wells had been drilled horizontally within the Western Canada Sedimentary Basin (WCSB) (Atkinson et al., 2020). The associated rise in induced seismicity in western Canada appears highly clustered near some of these HF operations and cannot be fully accounted for by the deployment of denser seismic monitoring networks and more sensitive instruments (Schultz, Stern, Gu, et al., 2015; Atkinson et al., 2016; Cui & Atkinson, 2016). Furthermore, recent studies have demonstrated that the hazard related to induced seismicity, including HF, is potentially much greater than that of natural seismicity, particularly in areas characterized by low to moderate tectonic activity (Atkinson, 2017; Lee et al., 2019; Atkinson et al., 2020; Langenbruch et al., 2020).

1.2 Earthquake Clustering

Cases of induced seismicity commonly appear as earthquake clusters and can manifest as both mainshock-aftershock burst sequences and as seismic swarms. For example, the large wastewater disposal-induced events in Oklahoma triggered typical aftershock behavior, temporarily increasing the seismic rate due to the transfer of stress and brittle failure of the crust (Keranen et al., 2014; McGarr & Barbour, 2017). On the other hand, the HF-induced clustering near Youngstown, Ohio, and injection-related events in central Arkansas near Guy and Greenbrier displayed swarm-like characteristics, where the events were of similar magnitude and could not be attributed to any dominant mainshock (Horton, 2012; Llenos & Michael, 2013; Skoumal et al., 2015). Studies have shown that both natural and anthropogenic changes to the subsurface fluid content can enhance or induce earthquake sequences via subsidence and/or changes in Coulomb fault stress and pore pressure conditions, especially near critically oriented structures (Segall, 1985; Langenbruch & Shapiro, 2010; Brodsky & Lajoie, 2013; Kumazawa & Ogata, 2013; Keranen et al., 2014; Goebel et al., 2015; Schoenball et al., 2015; Bao & Eaton, 2016; Kettlety et al., 2019, 2020). The nature of clustering observed within a region may be attributed to its rheological structure, in the framework of viscoelastic deformation (Ben-Zion & Lyakhovsky, 2006). In this context, a medium with low levels of heat flow and/or less fluid content correlate with higher viscosity and the conditions of brittle rheology, resulting in "burst-like" cracking of the crust and subsequent aftershock clustering. The converse is attributed to lower viscosity lithospheres of more brittle-ductile rheology (higher levels of heat and/or more fluid content), where failure is more likely to result in swarms of inter-linked events related to factors such as local fluid balance, destabilizing aseismic slip, and inter-event triggering (Zaliapin & Ben-Zion, 2016; Scuderi et al., 2017; Martínez-Garzón et al., 2018).

Identification of earthquake clustering involves a separation of the independent background rate from dependent event sequences (Gardner & Knopoff, 1974; Reasenberg, 1985; Baiesi & Paczuski, 2004; Console et al., 2010; Ader & Avouac, 2013; Zaliapin & Ben-Zion, 2013a; Schaefer et al., 2017). Due to the innumerable factors involved in the tectonic process, background seismicity may be approximated as random and modeled as a time-stationary, space-inhomogeneous marked Poisson process. Within this framework, rates of seismicity are assumed to vary in space but not in time and data points (seismic events) are marked by their magnitudes. Clustering, on the other hand, cannot be represented by a Poisson model, as the earthquake rate does not remain constant and depends in part on prior events. The separation of background and clustered earthquake phenomena is an important and non-trivial task required not only in cluster analysis but also in seismic hazard assessment, where catalogues are typically de-clustered in order to delineate source zones and assess recurrence parameters. The practical distinction of clustered and background seismicity should not restrict the consideration for potential interplay between them, particularly when external factors, such as anthropogenic activity, are involved. Induced seismicity has been observed to increase both the background rate and clustering productivity within affected regions (Lombardi et al., 2010; Llenos & Michael, 2013; Schoenball et al., 2015; Maghsoudi et al., 2016, 2018; Zaliapin & Ben-Zion, 2016; Vasylkivska & Huerta, 2017; Martínez-Garzón et al., 2018). It is plausible that a rise in the former subsequently affects a rise in the latter.

For example, Llenos & Michael (2013) characterized both natural and fluid injection-induced swarms in Arkansas, namely the natural 1980s Enola sequence and the wastewater disposal-related 2010-2011 Greenbrier sequences, by applying the epidemic type aftershock sequence (ETAS) model (Ogata, 1988, 1989; Zhuang et al., 2004). The ETAS model estimates the time dependent seismic rate using the summation of a constant background term with a parameterized Omori-type aftershock kernel. Llenos & Michael (2013) found that the Enola and Greenbrier swarms could not be modeled using the same set of parameter values, with the induced cluster resulting in both a higher background rate and elevated aftershock productivity relative to the natural swarm. The authors proposed that variation in the absolute values of ETAS parameters may be a way to distinguish between natural and human-induced seismicity within the same region, particularly changes in the background rate parameter and magnitude-weighting factor. A comparable investigation was performed over the geothermal sites of Salton Sea and Brawley in southern California, and obtained similar results (Llenos & Michael, 2016).

Zaliapin & Ben-Zion (2013a, 2013b) analyzed multiple southern California earthquake catalogues, which contain a large amount of both natural (tectonic and magmatic) and man-made (geothermal energy production-related) seismicity, using the nearest-neighbour distance (NND) method (described in detail in section 3). Briefly, the NND approach links events to their closest ancestor, i.e. their “nearest neighbour”, based on a space, time, and magnitude-dependent metric. Events are separated into clusters (those that are strongly linked to their nearest neighbour) and background seismicity (those that are only weakly linked), whereby variation in relative mixing proportions between the two populations may be evaluated. Clusters may then be classified further into mainshock-aftershock “burst-like” sequences or inter-event triggered “swarm-like” sequences, using their distinguishable structural characteristics. Zaliapin & Ben-Zion (2013a, 2013b), along with Hicks (2011), found that a natural separation occurs between clustered and background events for many regional catalogues, as well as for worldwide seismicity. Furthermore, the authors found that events within the clustered mode largely exhibit characteristics of either burst or swarm-like sequences. Their studies agreed well with the viscoelastic damage model, where a higher degree of inter-event triggering and swarm-like clustering was found within more ductile regions, such as geothermal settings or areas prone to magmatic or dike intrusion (e.g. Sagiya et al., 2002; Morita et al., 2006; Farrell et al., 2009), whereas more brittle rheology tended toward a higher proportion of burst-like sequences.

1.3 Study Area and Motivation

This study focuses on Alberta and eastern British Columbia, where seismicity has been low historically, with the majority of tectonic events occurring along the foreland belt of the Rocky Mountain range (Rogers & Horner, 1991). However, isolated incidents of induced spatiotemporally clustered seismicity have been documented. Baranova et al. (1999) and Eaton & Mahani (2015) linked the earthquake clustering near Rocky Mountain House, Alberta, to a depletion of pore pressure and an accumulation of compaction-related stress beneath the nearby Strachan gas extraction field between the 1970s and 1990s. In the 1980s and 1990s, fluid injection for secondary oil recovery and wastewater disposal had most likely triggered earthquake clustering northwest of Fort St. John, B.C. (Horner et al., 1994). More recently, HF and wastewater injection has taken place within the Montney and Horn River Basin shale formations; these operations are suspected to be contributing to the growth of seismic activity in that area based on spatiotemporal links to well activity as well as identification of nearby subsurface fault channels (B.C. Oil and Gas Commission, 2012, 2014). Further, several studies have provided evidence relating the extended HF-related fluid injection within the Duvernay formation to the recent seismicity west of Fox Creek, Alberta, via pore pressure increase and poroelastic stress perturbation. These events began in December 2013 as a series of clusters near Crooked Lake and continue to transpire intermittently to the present day (Schultz, Stern, Novakovic, et al., 2015; Bao & Eaton, 2016; Wang et al., 2016; Schultz et al., 2017; Zhang et al., 2019).

While many other studied regions tend to have a single dominant mechanism driving most of the induced earthquake activity, such as large-scale wastewater disposal in the central United States (Horton, 2012; Ellsworth, 2013; Llenos & Michael, 2013; van der Elst et al., 2013) or the geothermal energy operations in southern California (Brodsky & Lajoie, 2013; Zaliapin & Ben-Zion, 2013a), the WCSB is characterized by a multiplicity of local triggering mechanisms. The low natural occurrence rate has allowed for a relatively straightforward identification of the recent surge in unnatural seismicity (e.g. Atkinson et al., 2016), in comparison to the situation in other regions (e.g. Schoenball et al., 2015). Moreover, the increased seismicity may be particularly consequential in low-hazard regions where facilities were not designed for high seismic levels (Atkinson et al., 2020). These factors offer significant motivation to analyze the regional changes in earthquake space, time, and magnitude distributions statistically. It is also worthwhile to compare the different clusters and their potential triggers, particularly past cases of conventional production and disposal-related earthquakes versus the recent seismicity triggered by hydraulic fracturing. In this paper, we aim to characterize the regional and clustered earthquakes using data from the Composite Alberta Seismicity Catalogue (CASC, described in section 2) via the application of the nearest-neighbour distance method. The objectives are twofold. The first is to demonstrate fundamental differences in the regional seismicity distributions over time and the second is to illuminate specific features of induced seismicity clustering that could be expected or recognized when performing certain types of operations, particularly hydraulic fracturing.

In the following sections, background information is provided on the regional earthquake catalogue and on four particular areas of earthquake clustering suspected to have been triggered by nearby human operations (section 2). The nearest-neighbour distance methodology is described (section 3). Analysis of the regional catalogue over time is presented (section 4); scaling relationships and statistical properties of the four clustering sites are discussed (section 5). A discussion and comparison with other, similar regional studies (section 6) is followed by a brief summary (section 7).

2 Data and Study Regions

2.1 Earthquake Catalogue

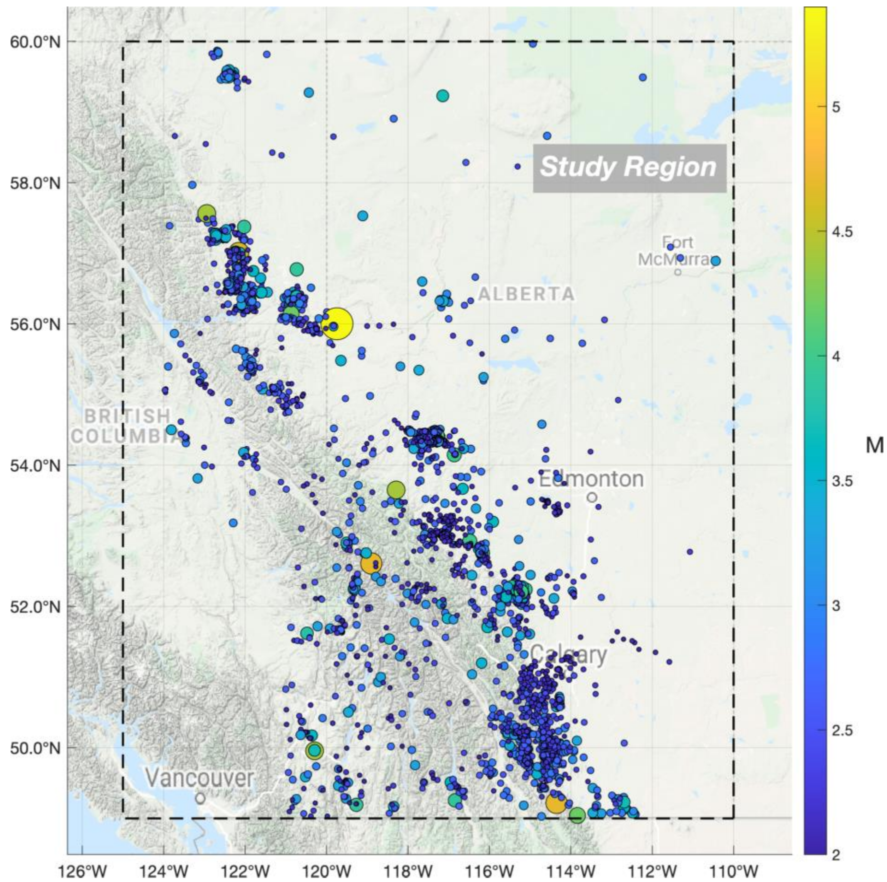


***Figure 1**: Map of **M**2+ seismic events documented by the Composite Alberta Seismicity Catalogue (CASC) from 1975-2018. Marker size and colour indicate earthquake moment magnitude.*

The Composite Alberta Seismicity Catalogue (CASC), available online at www.inducedseismicity.ca, contains seismic event records from the early 1900s through to the present. The spatial boundaries of the region encompass a rectangular area spanning $[49° - 60°N]$ and $[110° - 125°W]$, including the entirety of Alberta and a portion of eastern British Columbia (Figure 1). The CASC is compiled from several contributing agencies operating across Alberta and eastern British Columbia, including the Geological Survey of Canada and Earthquakes Canada (www.seismo.nrcan.gc.ca), the Alberta Geological Survey and the Regional Alberta Observatory for Earthquake Studies (ags.aer.ca), the Canadian Rockies and Alberta Network (ds.iris.edu), and the TransAlta/Nanometrics Network (www.nanometrics.ca). The catalogue contains information concerning the date, time, estimated geographic location, magnitude, magnitude scale, and moment magnitude conversion for each detected event occurrence (for a detailed discussion on the compilation of the CASC, see Cui et al. (2015)). The database is estimated to be complete to the moment magnitude $M \approx 3$ level from 1985 onward (Adams & Halchuk, 2003; Cui & Atkinson, 2016). Seismic network coverage is generally spatiotemporally inhomogeneous and so local completeness levels over a given time period may be substantially lower than the regional completeness; this matter is explored in the cluster analyses in section 5. Accurate depth estimations remain a difficult task for most networks and, as a consequence, depths listed in this catalogue have large errors or are only specified as default values. While hypocentral locations would be beneficial for statistical analyses in three dimensions, they are not critical. The methods used in this study require relatively few input requirements; only the magnitudes, epicentral locations and times of occurrence are used from the database.

Some potential artifacts related to catalogue inconsistencies are important to note. First, many of the seismic recordings within the CASC are nontectonic and instead a product of quarry and mining blasts. These events are generally flagged by network personnel based on several criteria, including event time (blasts occur during daylight hours), proximity to active mines and quarries, and specific waveform characteristics (typically compressional first motions and high frequency spectra) (Cui et al., 2015; Schultz, Stern, Gu, et al., 2015; Cui & Atkinson, 2016). In this study, all flagged events were removed from the catalogue beforehand. However, recent blast events (after 2014), southwest of Calgary, had not yet been flagged by the network at the time of access (last accessed June 2020, www.inducedseismicity.ca) and were hence included in the analyses. A second potential artifact is that the CASC is spatially limited to the Alberta region and only a portion of northeastern B.C. This explains the lack of recorded events surrounding the Vancouver area in Figure 1. Readers interested in documented seismicity west of Alberta are referred to the National Earthquakes Database (database link).

2.2 Induced Earthquake Clusters

The Rocky Mountain House cluster (RMHC)

One of the first significant instances of induced seismicity within the WCSB occurred approximately 25 km southwest of the town of Rocky Mountain House, Alberta. The area had been historically quiescent before the onset of conventional gas production within the Duvernay formation in the early 1970s (Rogers & Horner, 1991). It became active predominantly from 1975-1992, lagging production rates by several years and returning to apparent background activity by the year 2000. Wetmiller (1986) found that the majority of events occurred roughly within a 15 km radius and were concentrated close to the Strachan and Ricinus gas fields. Baranova et al. (1999) proposed that the earthquakes were triggered due to long-term compaction-related changes in the stress field caused by the extraction of fluid. We refer to this collection of events between 1975-2000 as the RMHC (Figure S1).

The Montney clusters (MC1 & MC2)

Conventional oil and gas production has occurred within the Montney formation, which stretches across northeastern British Columbia and northwestern Alberta, since the 1950s. Associated wastewater disposal wells have been active from the 1960s and are suspected to have triggered two distinct earthquake clusters (Horner et al., 1994; B.C. Oil and Gas Commission, 2014). The first began in 1984, north of the town Fort St. John, B.C., and the second began in 2003, west of Halfway Ranch, B.C. We collectively refer to these two clusters, occurring predominantly between 1984-2009, as the MC1.

Since the mid-to-late 2000s, with the development of horizontal drilling and HF technology, the Montney trend has attracted significant interest for its siltstone and shale gas reserves. By 2018, thousands of natural gas wells were active in the area, operating along the formation's northwestern margin as well as to the southeast near Dawson Creek, B.C. In addition, more than 15 wastewater disposal wells have been drilled since 2005, bringing the formation's total to over 100 (B.C. Oil and Gas Commission, 2014). Since then, substantial low-to-moderate seismicity has been recorded in the area, due to the augmentation of the local seismic network and possibly the increase in subsurface human activity. The distribution of events occurring in the formation has changed over time, as the dominant triggering mechanism shifted from wastewater injection to HF (B.C. Oil and Gas Commission, 2014; Atkinson et al., 2016). We refer to the recently recorded seismicity in this area, from 2010 to 2018, as the MC2.

The Fox Creek Cluster (FCC)

Conventional production in central Alberta, primarily within the Duvernay, Swan Hills and Leduc formations, has been occurring since the 1960s and resulted in minor associated seismicity apart from the clustering near Rocky Mountain House. However, in December 2013, earthquakes began occurring approximately 30-40 km west of the town of Fox Creek, Alberta, where HF wells had recently been drilled in order to access the Duvernay's reservoirs of tight shale. Several hundreds of these wells have been drilled since 2012, near the recent clustering, and a large proportion were drilled horizontally or at a deviated angle. Seismic activity began as a few distinct sequences near Crooked Lake, Alberta, and continues to form further clusters up to the present day (Bao & Eaton, 2016; Clerc et al., 2016; Deng et al., 2016; Schultz et al., 2016, 2017, 2018; Wang et al., 2016; Eaton et al., 2018; Zhang et al., 2019). We refer to this group of events, from 2013 to January 2020, as the FCC.

We focus our study on these four clusters, as their seismicity is sufficiently rich for analysis. New clusters continue to emerge, such as those in the area near Red Deer, Alberta (e.g. Schultz et al., 2020).

3 The Nearest-Neighbour Distance Method

The nearest-neighbour distance (NND) method is a statistical approach to earthquake cluster identification and classification, first formulated by Baiesi & Paczuski (2004) and expanded significantly by Zaliapin et al. (2008) and Zaliapin & Ben-Zion (2013a, 2013b, 2016). Its purpose is to link together and characterize *event families* or *sequences* using a rescaled inter-event distance metric termed the nearest-neighbour distance η, which is defined below as space, time and magnitude dependent. This method is applied in sections 4 and 5 in order to describe the regional and local inter-event distance distributions within the WCSB as well as to statistically categorize the types of seismic clustering observed.

3.1 The Rescaled Inter-Event Distance Metric η

The inter-event distance values η_{ij} are defined based on the spatiotemporal distance between each event pair within the catalogue as well as on the magnitude of the event that occurred *first* (the potential *parent* event *i*). Specifically, each event *j* is assigned values η_{ij} based on its relationship with all other events *i* as

$$\eta_{ij} = \begin{cases} t_{ij}(r_{ij})^{d_f}10^{-bm_i}, t_{ij} > 0 \\ \infty, \quad t_{ij} \leq 0 \end{cases}, \qquad [1]$$

where $t_{ij} = t_j - t_i$ is the time in days between event *j* and event *i*. Note that event *j* must succeed event *i* in order for the quantity t_{ij} to be positive, otherwise $\eta_{ij} = \infty$.

The inter-event spatial distance r_{ij} is computed between epicenters using the Haversine formula for great-circle distance (or arc length) in kilometers

$$r_{ij} = 2r_e \arcsin \sqrt{\sin^2 \frac{(\varphi_i - \varphi_j)}{2} + \cos\varphi_i \cos\varphi_j \sin^2 \frac{(\lambda_i - \lambda_j)}{2}}. \qquad [2]$$

In Equation [2], r_e is the Earth's radius estimated as 6378.14 km, and (φ_i, λ_i) and (φ_j, λ_j) are the latitudinal and longitudinal coordinates of events *i* and *j*, respectively.

d_f is the fractal spatial dimension of earthquake epicenter distribution. In 2 dimensions, for both local and worldwide epicentral distributions, d_f has been found to vary approximately between 1.2 and 1.6 (Sadovsky et al., 1984; Kagan, 1991; Kosobokov & Mazhkenov, 2013). m_i is the magnitude of the i^{th} event and *b* is the Gutenberg-Richter *b*-value, which approximates the exponential distribution of magnitude scaling according to

$$N(\geq M) = 10^{a-b(M-M_0)}; M \geq M_0, \qquad [3]$$

where $N(\geq M)$ is the number of detected events greater than or equal to magnitude *M* and *a* reflects the rate of seismicity over the time period considered.

The scalar distance η_{ij} may be expressed in terms of its rescaled temporal and spatial components by defining

$$T_{ij} = t_{ij} 10^{-\frac{bm_i}{2}};$$
$$R_{ij} = r_{ij}{}^{d_f} 10^{-\frac{bm_i}{2}}. \qquad [4]$$

By this formulation, $\eta_{ij} = T_{ij} R_{ij}$. Once nearest-neighbour values η have been determined for each event j, where $\eta_j = \min_{i\, i<j} \eta_{ij}$, the η distribution, the joint distribution of (T, R) as well as their individual histograms can be plotted to observe possible modality in the temporal and/or spatial distance between events. Hicks (2011) showed that a bimodal distribution in η, interpreted as a distinction between background and clustered events, is an intrinsic property of both worldwide and regional seismicity. The clustered mode is identified as the subpopulation of events occurring at small η values (small inter-event distances); these events are considered to be strongly linked to their nearest neighbours. The background mode is identified as the set of events occurring at larger η values (larger inter-event distances); these events are considered to be only weakly linked to their nearest neighbours.

3.2 Formal Analysis of Modality in the η Distribution

Although the subpopulations of clustered and background events may be apparent upon visual inspection, it is useful to define them rigorously considering a Gaussian mixture model (GMM), as detailed in Hicks (2011). A GMM is defined as a composition of normal density functions, each with a mean, covariance and mixing proportion (or weight). The parameters for these component functions are estimated using the 2-step expectation-maximization algorithm (Hastie et al., 2009). The approach uses an initial guess for the set of parameters to then:

a) Calculate Bayesian probabilities for each data point as a possible member of each mode. This is the expectation step.

b) Estimate the model parameters for each mode through their maximum likelihood function, using the probabilities determined in the expectation step as weights. This is the maximization step.

This process is iterated until it converges to the optimal estimation of the means, standard deviations and weights. The number of modes in the distribution is determined using information criteria. The model that minimizes the information criteria is considered the best fit. Two such criteria were used in this study, the Akaike and Bayesian information criteria (Akaike, 1974; Schwarz, 1978). The threshold value $\log_{10} \eta_{thresh}$, which separates the background mode from the clustered mode, is chosen as the intersection point between the resulting component densities.

3.3 Event-Family Classification

By removing all weak links from the NND distribution, the clustered mode may be further discretized into hierarchical families based on the strong links between parents and offspring events. The largest event in each family is classified as the mainshock; if there is more than one largest-magnitude event then the first is considered the mainshock. Events in the sequence that occur before the mainshock are called foreshocks and occur after are called aftershocks. An aftershock's *generation,* or *order,* is determined by its hierarchical distance from the mainshock. For example, an aftershock whose parent event is the mainshock is considered 1st generation, while an aftershock whose parent event is a 1st generation aftershock is considered 2nd generation, and so on.

These families can be categorized as aftershock sequences, seismic swarms, a combination of the two, or neither, based on statistical parameters introduced by Zaliapin & Ben-Zion (2013b). These parameters quantify the structural characteristics of each event family and can determine their structure type. The terminology of a rooted tree-graph is employed, which considers the first event in a sequence as the *root* of the tree, the downward-directed edges connecting events as *branches* and the end-nodes (earthquakes that have no further offspring) as *leaves*. The size *N* is computed as the number of events in a sequence. The *leaf depth* d is calculated by counting the number of branches connecting each leaf back to its root. The *average leaf depth* $\langle d \rangle$ of a particular tree provides an indication of its shape, with larger $\langle d \rangle$ potentially indicating higher levels of event chaining and a deeper structure, and smaller $\langle d \rangle$ implying low orders of event offspring and a shallower structure. Therefore, swarm sequences are expected to have larger $\langle d \rangle$ values than bursts, even given a similar number of leaves. However, since $\langle d \rangle$ scales with the sequence size *N*, a normalized leaf depth δ, where $\delta = \frac{\langle d \rangle}{\sqrt{N}}$, is also calculated. The *inverted branching number* B_I of a tree is computed as the number of parent events divided by the total number of branches. A maximum B_I value of unity indicates a perfect path shape and only a single leaf within the structure. Smaller values of B_I imply more offspring from fewer parents, i.e. a more burst-like formation. The *magnitude differential* Δm is defined as the difference between the designated mainshock and second-largest event in a sequence. Generally, aftershock sequences tend to have larger Δm values than swarms, given a similar size *N*. In addition, we compute the *spatial area* A (in km^2) and *time period* t_D (in days) covered by each family. However, these parameters also scale with *N* and are included for observational purposes only.

4 Regional Analysis of NND Distributions within the WCSB

In this section, we applied the nearest-neighbour distance (NND) method to the regional Composite Alberta Seismicity Catalogue (CASC; detailed in section 2). The entire time period, between 1975-2018, was analyzed, as were separated time windows, between 1975-2009 and 2010-2018. The separation between time frames was chosen to correspond with the rapid rise in hydraulic fracturing implementation that occurred within the region after 2009 (e.g. Atkinson et al., 2016, 2020)

Zaliapin & Ben-Zion (2013a, 2013b) showed that NND distributions are stable for cut-off magnitudes below the regional completeness magnitude ($M_c \approx 3.0$ in this case); we therefore applied a magnitude cut-off of $M_0 = 2.0$ to enrich sampling. Figures S5-8 show that the results found here hold for cut-off magnitudes up to the regional completeness level. The same constant values $d_f = 1.5$ and $b = 1.0$ were used for each subset of the regional data to illustrate the fundamental differences observed over time without prior parameterization.

4.1 Entire Time Period (1975-2018)

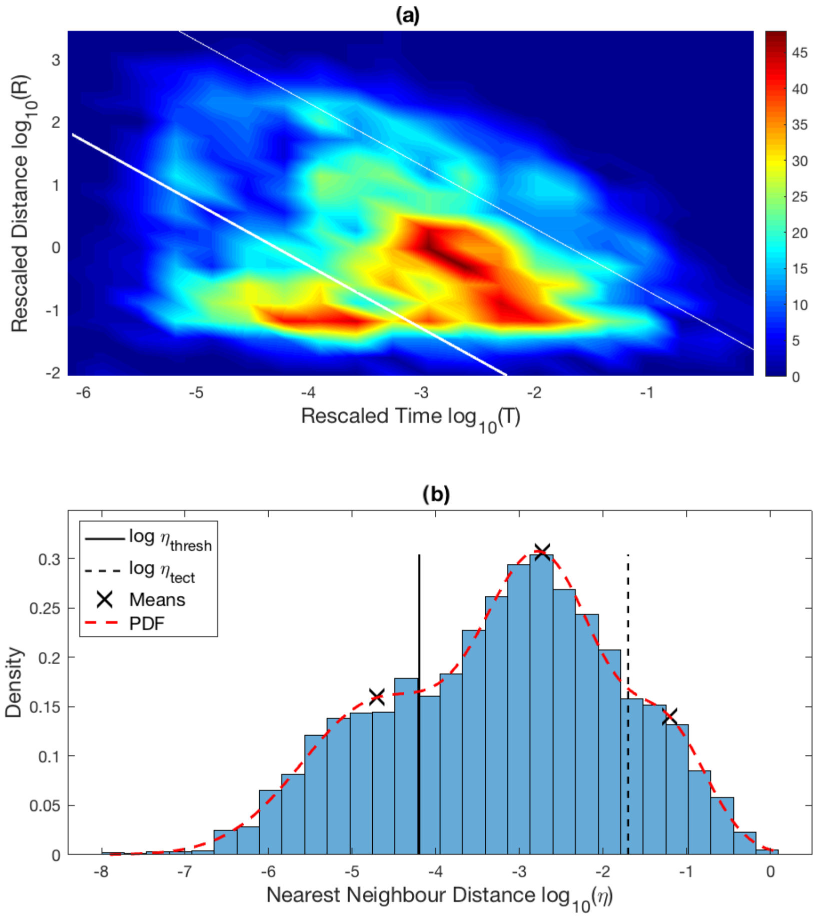


Figure 2: *NND distribution of the regional WCSB catalogue (1975-2018) using* $M_0 = 2.0$. *a) Joint distribution of the temporal and spatial components (T,R). Bold white line indicates the threshold* $log_{10}\,\eta_{thresh}$ *between tightly clustered and loosely clustered components. Thin white line indicates the threshold* $log_{10}\,\eta_{tect}$ *between loosely clustered and deep-background components. Colour bar indicates frequency of inter-event distance occurrence. b) Normalized density of* η *values. Solid black line is* $log_{10}\,\eta_{thresh}$ *and dashed black line is* $log_{10}\,\eta_{tect}$. *Dashed red line is the normalized probability density function of the Gaussian mixture and black crosses are the component means.*

Figure 2 shows the 2-dimensional (T,R) distribution and the normalized density of nearest-neighbour distances η for the entire study period. There are two prevailing modes within the joint distribution (Figure 2a), which differ in size and shape. The clustered mode, to the bottom-left of the plot beneath the bold white line, is oriented somewhat horizontally, while the background mode, in between the two white lines, covers a broader range along the T and R axes and is faintly oriented along the downward diagonal. This observation is consistent with most findings in other studies, which showed the existence and distinct shaping of clustered and background earthquake subpopulations in several tectonic and induced areas (Hicks, 2011; Zaliapin & Ben-Zion, 2013a, 2016; Schoenball et al., 2015). Notably, in these cases the background mode was reminiscent of a time-stationary, space-inhomogeneous marked Poisson process, which forms a unimodal distribution concentrated along a downward diagonal at large *T* and *R*, while the clustered mode typically occupies an ellipse at smaller *T* and *R* values than would be expected from a Poisson process (Zaliapin & Ben-Zion, 2013a). The chief disparity between different regions, including within the WCSB, lies in the relative intensities, or mixing proportions, of each mode; these proportions tend to reflect the nature of seismicity occurring (discussed further below). More difficult to see in Figure 2a is a third subpopulation, to the upper-right above the thin white line, which is much less concentrated and sprawls along a downward diagonal over large values of *T* and *R*. The one-dimensional η distribution (Figure 2b) reflects this tri-modality clearly, with the third mode at large η perhaps more apparent.

The results of the Gaussian mixture model analyses indicate that the optimal mixture estimation, i.e. a minimization of the information criteria, is a 3-component, instead of the typically observed 2-component (clustered and background), distribution. This may be observed in Figure 2b, where the dashed red line and black crosses represent the resulting probability density function and 3 component means, respectively.

The intersection point between the two dominant modes occurs at $\log_{10} \eta_{thresh} = -4.2$, shown by the bolded white line in Figure 2a, and solid black line in Figure 2b. Out of the 3531 total events analyzed, 29% were found in the clustered domain, which is henceforth referred to as the *tightly clustered* mode. 56% were located in the middle mode, which is henceforth referred to as the *loosely clustered* mode, and 15% were located in the third mode at large η, henceforth referred to as the *deep background*. The intersection between the loosely clustered and deep background modes occurs at $\log_{10} \eta_{tect} = -1.7$, shown by the thin white line in Figure 2a, and dashed black line in Figure 2b. No definitive trends were apparent in the individual *T* or *R* distributions (not shown), with the exception of an increasing proportion of small *R* distances that tended to stack at the limits of network location resolution, giving the tightly clustered mode its azimuthal shape. This agrees with another observation of Zaliapin & Ben-Zion (2013a), who state that the modality in η cannot be fully explained by marginal trends present in either *T* or *R* but is in fact dependent on the association between the two, as seen in the 2-dimensional joint distribution.

4.1.1 Comparative Analysis of Mixture Components

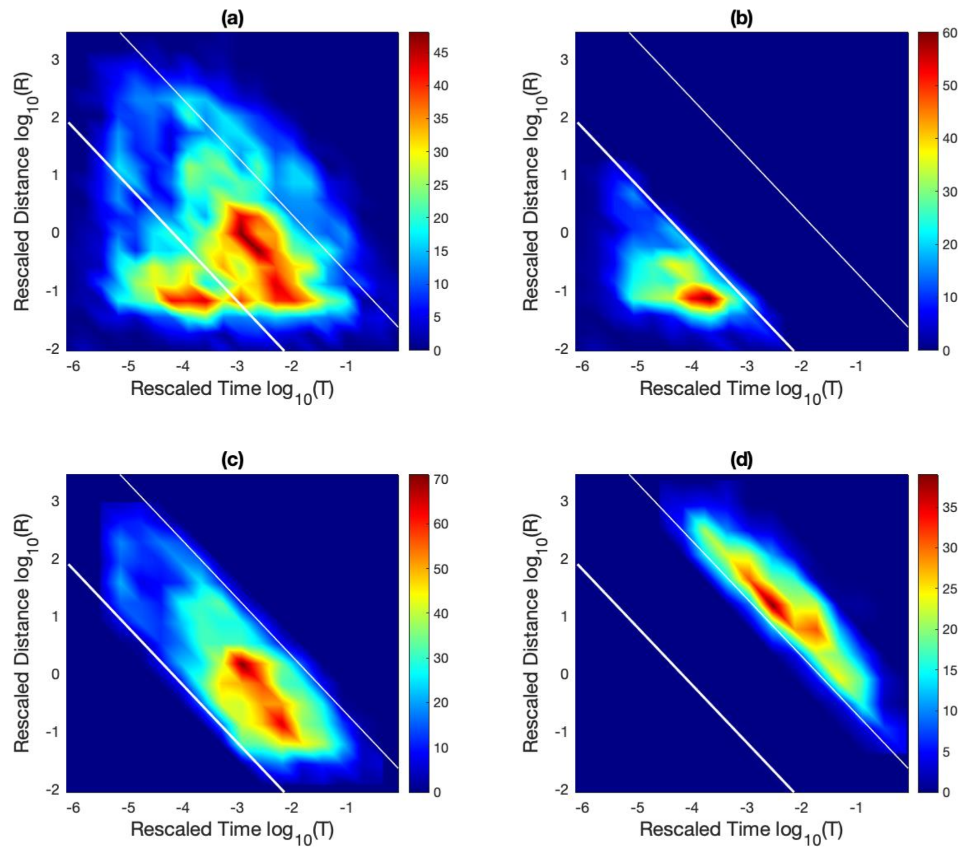


***Figure 3**: Joint (T,R) distributions of individual modes for the regional WCSB catalogue (1975-2018) using $M_0 = 2.0$. Colour bar indicates frequency of inter-event distance occurrence. Bold white line indicates the threshold $log_{10}\,\eta_{thresh}$ between tightly clustered and loosely clustered components. Thin white line indicates the threshold $log_{10}\,\eta_{tect}$ between loosely clustered and deep background components. a) Entire distribution. b) Tightly clustered mode. c) Loosely clustered mode. d) Deep background.*

Figure 3 plots the modal decomposition of the joint distribution, where each subpopulation is plotted separately, removing the dependence on mixing proportion. As mentioned above, if a distribution does not intensify along the bisecting diagonal (constant η) and instead forms an elliptical cloud, then a trend may exist in the data as a deviation from Poisson behavior. As expected, the tightly clustered subpopulation deviates substantially from the diagonal and forms an ellipse within the sub-region $\{-5 < \log_{10} T < -3.5 \mid -1.5 < \log_{10} R < -0.5\}$. The loosely clustered mode is faintly concentrated along the diagonal yet also forms a cloud within the sub-region $\{-3.5 < \log_{10} T < -1 \mid -1.5 < \log_{10} R < 1\}$. By contrast, the deep background is distributed very closely along the diagonal and stretches almost its complete length, most clearly resembling a Poisson process.

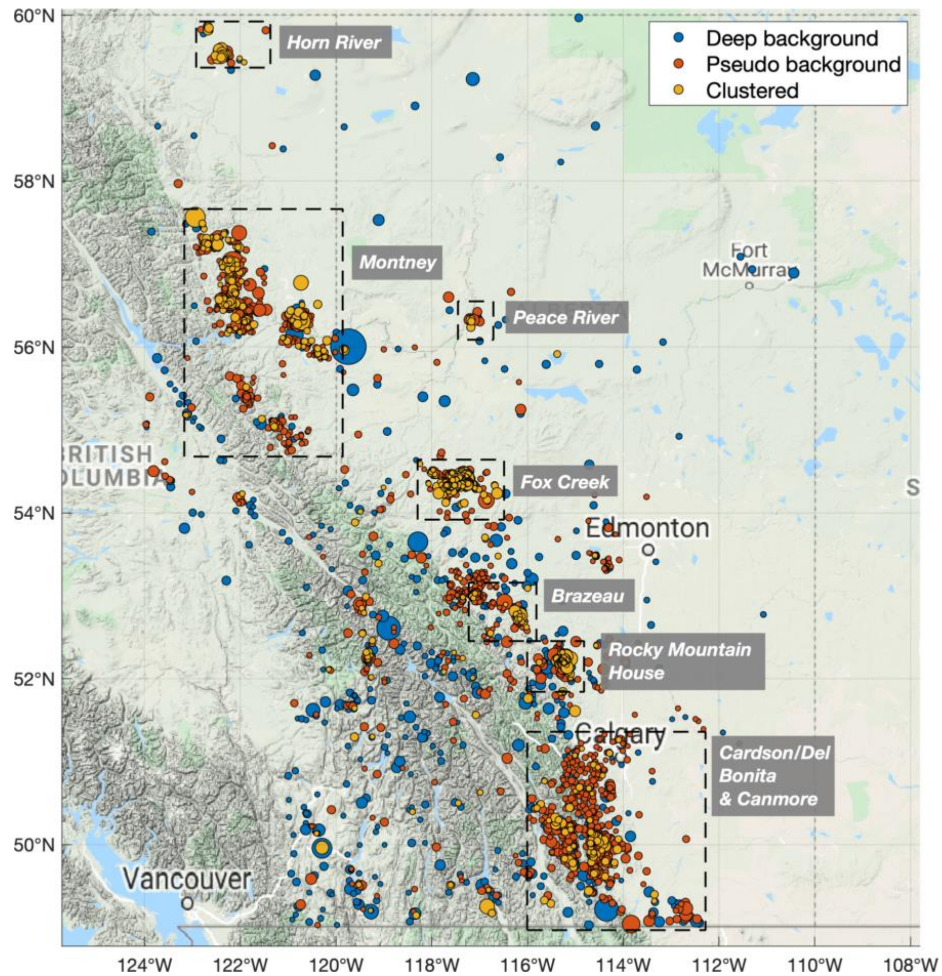


Figure 4*: Spatial map of earthquakes for the regional WCSB catalogue (1975-2018) using* $M_0 = 2.0$*, represented in terms of their nearest-neighbour distance categorization. Blue markers are the deep background, orange markers are the loosely clustered events and yellow markers are the tightly clustered events. Marker size indicates magnitude. Dashed boxes surround areas of suspected induced clustering (see (Atkinson et al., 2016)).*

A spatial map of the catalogued earthquakes is displayed in Figure 4. Blue markers represent deep background events, while orange and yellow markers represent the loosely clustered and tightly clustered events, respectively. Dashed boxes surround individual areas of suspected induced activity. According to the NND method, both the blue and orange markers are classified as the "background" portion of events. However, the two subpopulations are not representative of the same type of seismicity. As a physical representation of the results in Figure 3, the blue markers are substantially more evenly distributed and seemingly reflective of the natural tectonic background (the majority occur along the foreland belt of the Rocky Mountain range), whereas an obvious spatial dependence on the distribution of orange and yellow markers is indicative of clustering. The orange and yellow markers overwhelmingly dominate the dashed boxes.

We also observe that the background events which occurred farther away from their own potential parent were less likely to initiate a clustering sequence, with deep background events substantially less likely to trigger future earthquakes than loosely clustered events. Out of 320 non-single event families, 281 (88%) began from the loosely clustered mode and 39 (12%) began from the deep background. Out of the 35 identified *significant* sequences, which we arbitrarily define as containing five or more events, 33 (94%) initiated from loosely clustered ancestors while only 2 (6%) initiated from the deep background. This correlation holds for all family sizes; as the number of events within a family *increases*, the likelihood of the family to originate from the loosely clustered mode also *increases*. On the other hand, it is true that there is a larger population of loosely clustered events, which could partially explain the disparity. However, even relative to the mixing proportion of each mode, a non-single event family is twice as likely, and a significant event family is nearly four times as likely to have originated from the loosely clustered mode rather than in the deep background. This may be observed in Figure 4; overlap occurs substantially between loosely clustered (orange) and tightly clustered (yellow) markers, but rarely occurs between deep background (blue) and tightly clustered (yellow) markers. Overall, the differences between the loosely clustered and deep background subpopulations appear noteworthy and demonstrable.

4.2 Separated Time Periods: 1975-2009 and 2010-2018

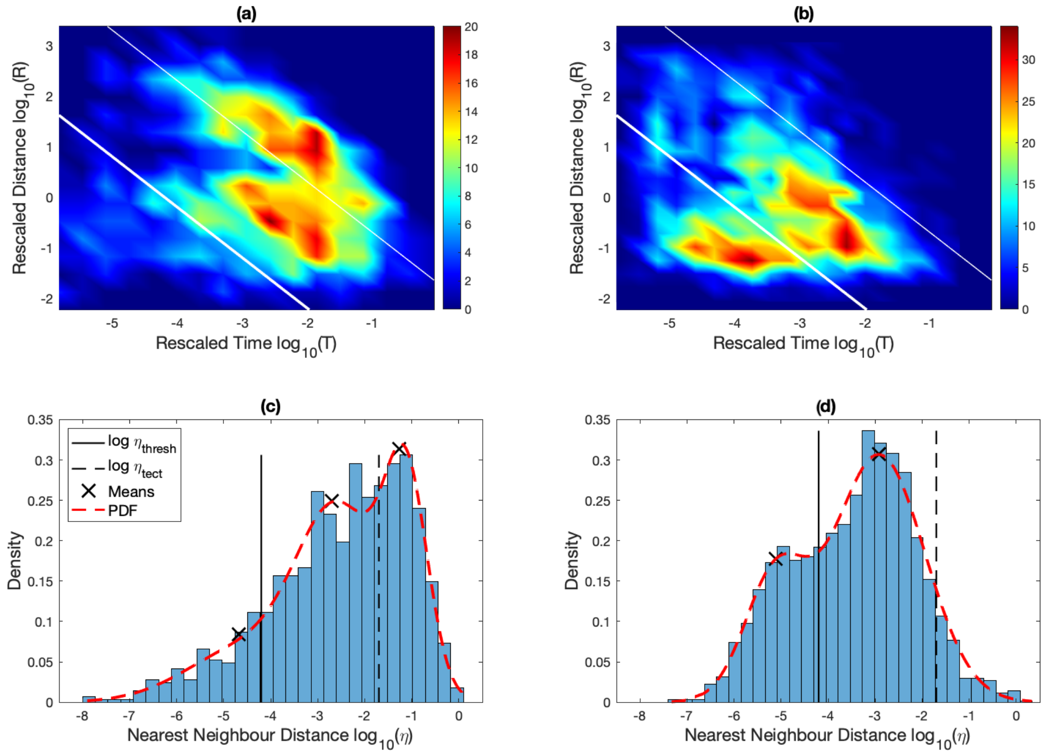


***Figure 5**: Comparison of NND distributions of the regional WCSB catalogue across time using $M_0 = 2.0$. (a,b) Joint distributions of the temporal and spatial components (T,R). Bold white line indicates the threshold $log_{10}\,\eta_{thresh}$ between tightly clustered and loosely clustered components. Thin white line indicates the threshold $log_{10}\,\eta_{tect}$ between loosely clustered and deep background components. Colour bar indicates frequency of occurrence. (c,d) Normalized densities of η values. Solid black line is $log_{10}\,\eta_{thresh}$ and dashed black line is $log_{10}\,\eta_{tect}$. Dashed red line is the normalized probability density function of the Gaussian mixture and black crosses are the component means. (a, c) 1975-2009, both background modes are dominant. (b, d) 2010-2018, deep background shrinks while tightly clustered mode appears. The loosely clustered subpopulation is common to both time frames.*

Figure 5 presents the *(T,R)* and η distributions of the WCSB for separated time intervals. From 1975-2009, the regional catalogue is characterized by natural activity and isolated cases of induced clustering due to conventional oil and gas production and associated wastewater disposal, primarily within the Duvernay and Montney formations (Wetmiller, 1986; Rogers & Horner, 1991; Horner et al., 1994; Baranova et al., 1999; Schultz et al., 2014). The resulting space-time inter-event distance distribution shows that both background modes are dominant, containing 83% of all events analyzed, indicating mainly single events and loose clustering, but little tight clustering (Figure 5a and Figure 5c). In contrast, the regional catalogue between 2010-2018 is characterized by sparse natural activity, due to the shorter time-frame, and large amounts of induced activity attributed to the sharp rise in the implementation of horizontally drilled HF operations within the Duvernay, Montney, Cardium and Bakken formations, among others (B.C. Oil and Gas Commission, 2012, 2014; Schultz, Mei, et al., 2015; Schultz, Stern, Novakovic, et al., 2015; Atkinson et al., 2016; Bao & Eaton, 2016). The joint distribution changes strikingly; the deep background subpopulation nearly disappears and a tightly clustered mode arises, indicating both substantial loosely clustered and tightly clustered seismicity (Figure 5b and Figure 5d). An important factor to consider is the variation in seismic monitoring capability over time; many more stations were operational during the recent time period, which undoubtedly had an effect on the distributions. We show below that the changes in mixing proportion across time are vastly disproportionate and cannot be entirely attributed to improvement in network detection.

The Gaussian mixture model analysis over the first time period detects the same three modes identified over the entire time frame. Over the second period, it picks out only the tightly clustered and loosely clustered components. This is understandable, as the deep background shrinks and ultimately contains only 7.5% of the mixing proportion from 2010-2018. Overall, there is a 27% drop in the relative number of deep background events and a 10% increase in the proportion of loosely clustered events between the two time periods. There is a 17.5% increase in the relative population of tightly clustered events. The changes in rate are disproportionate; while the yearly detected **M**2+ deep background rate roughly doubles, the loosely and tightly clustered rates increase tenfold and seventeen-fold, respectively. The disproportionality holds above the estimated regional completeness level $M_c = 3.0$; for magnitudes above this threshold, the deep background rate is almost doubled across the two time periods, while the loosely and tightly clustered rates both increase nearly six-fold (Figure S8). We perform a two-sample Kolmogorov-Smirnov test on the distributions of η values between the separated time periods, in order to test the null hypothesis that the differences observed are within sampling errors and the two samples are actually from the same continuous distribution. The test rejects the hypothesis at the 99% confidence level.

These results suggest that the changes in the inter-event distance distribution across Alberta and eastern B.C. over time are statistically significant and not naturally occurring. The decreased mixing proportion of deep background events indicate that the majority of the recent seismicity is not tectonic, while the increase in tightly clustered seismic activity correlates temporally with the rise in horizontally drilled HF treatments within the region. Figure 4 shows that clustering is highly spatially correlated with human activity as well, with smaller inter-event distances (i.e. orange and yellow markers) transpiring chiefly near areas flagged as suspicious. The statistical properties of several of these suspicious areas are explored in detail in the next section.

5 Analysis and Comparison of Isolated Seismic Clusters

Here we examine four separate cases of seismic clustering within the Western Canada Sedimentary Basin (WCSB): the Rocky Mountain House cluster (RMHC) – induced by conventional natural gas extraction, the Montney clusters (MC1 and MC2) – triggered by wastewater disposal and hydraulic fracturing, and the Fox Creek Cluster (FCC) – induced by hydraulic fracturing (these clusters are described in more detail in section 2; see Figures S1-4 for related event maps).

5.1 Frequency-Magnitude Statistics

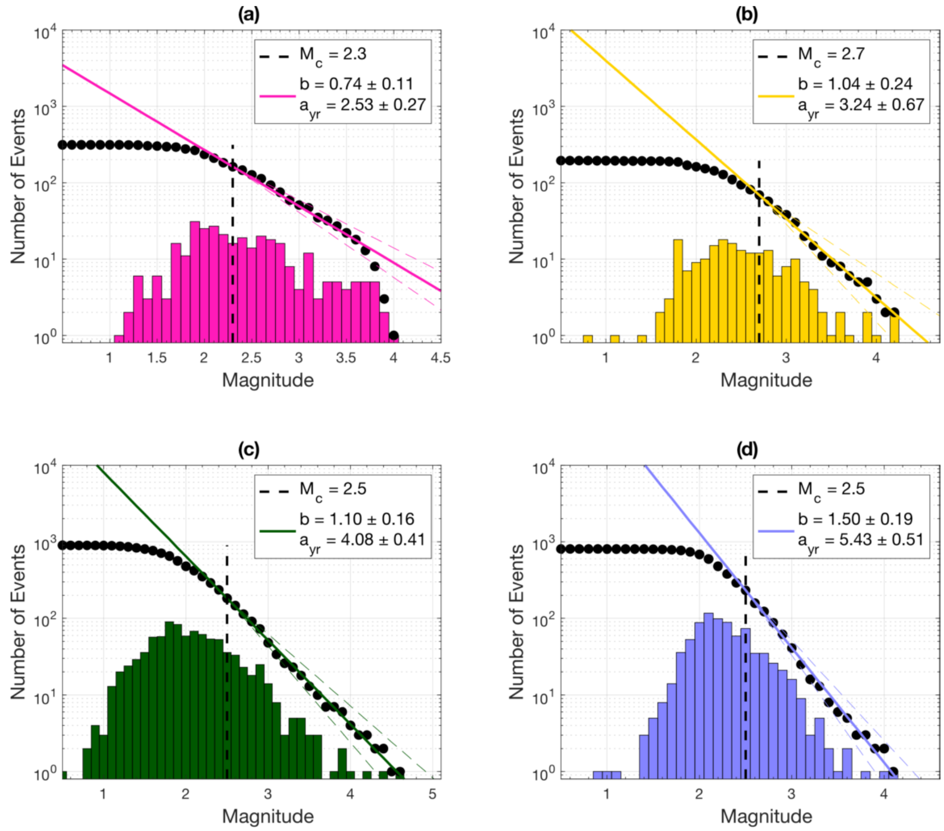


Figure 6: *Frequency-magnitude distributions and estimated Gutenberg-Richter parameters for each investigated cluster. Coloured histograms represent the non-cumulative magnitude distributions and black dots represent the cumulative magnitude distributions. Dashed black lines indicate the determined magnitude of completeness. Coloured lines represent the modeled GR relations. a) Rocky Mountain House cluster. b) Montney cluster 1. c) Montney cluster 2. d) Fox Creek cluster.*

We begin by assessing the frequency-magnitude distributions (FMDs) of each cluster and computing their Gutenberg-Richter scaling parameters (Figure 6). The FMD plots the cumulative and non-cumulative frequencies of earthquake magnitude occurrence in log-linear space. We use the maximum likelihood method to determine the *b*-values along with their 95% uncertainties (Aki, 1965; Utsu, 1966; Shi & Bolt, 1982). The local completeness level in each case was determined using a suite of catalogue-based methods, namely the method of maximum curvature, the goodness-of-fit test, and the method of *b*-value stability (Wiemer & Wyss, 2000; Cao & Gao, 2002).

The RMHC, which was induced by conventional gas extraction, is characterized by a very broad distribution of event magnitudes resulting in a very low *b*-value (Figure 6a, $b =$

0.74 ± 0.11). This contrasts with the Fox Creek cluster (FCC), induced by hydraulic fracturing, whose FMD is very steep and contains many more small magnitude events and fewer large magnitude events. As a consequence, it is described by a very high b-value (Figure 6d, $b = 1.50 \pm 0.19$). Both the wastewater disposal and hydraulic fracturing-induced Montney clusters (MC1 and MC2) have more typical magnitude distributions (Figure 6b, $b = 1.04 \pm 0.24$ and Figure 6c, $b = 1.10 \pm 0.16$). The a-values are normalized to reflect the yearly-detected seismicity rate. It is unsurprising that the more recent, densely populated clusters (MC2 and FCC) have the highest yearly a-values, whereas the RMHC and MC1 both span longer periods of time and are smaller overall, resulting in lower yearly a-values.

5.2 Nearest-Neighbour Distance Distributions

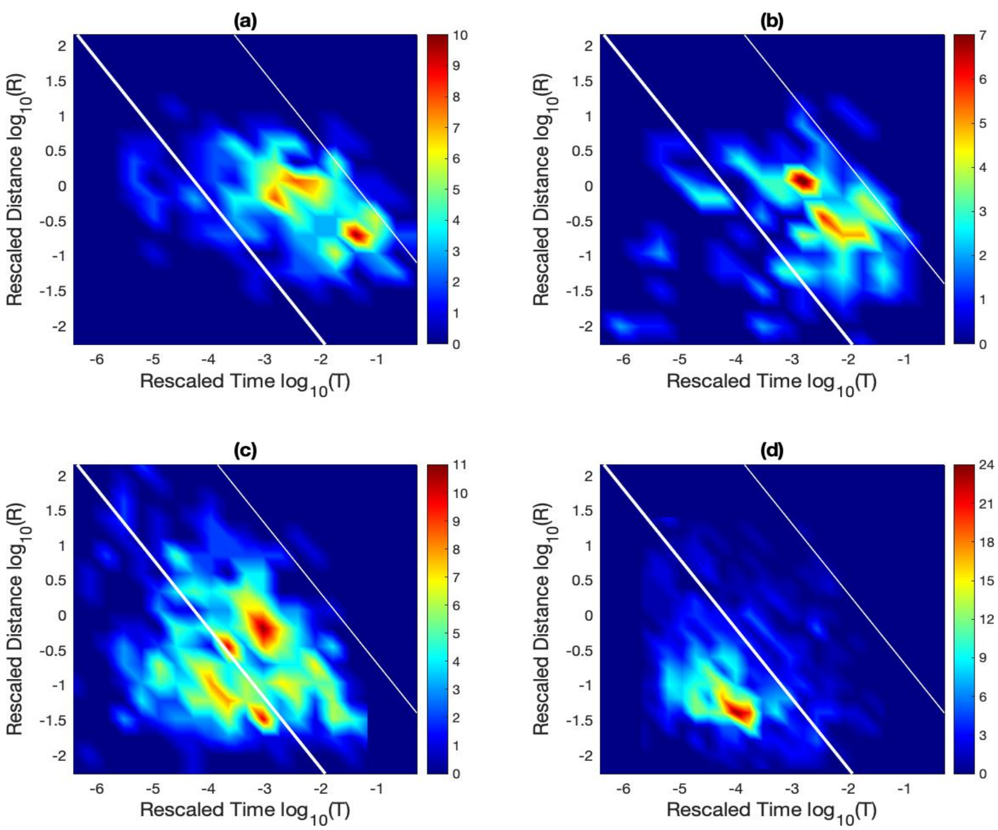


***Figure 7**: Joint NND distributions of the temporal and spatial inter-event distances (T,R) for each investigated cluster. Bold white line indicates the threshold $log_{10}\,\eta_{thresh}$ between tightly and loosely clustered components. Thin white line indicates the threshold $log_{10}\,\eta_{tect}$ between the loosely clustered and deep background. Colour bar reflects inter-event distance occurrence frequency. a) Rocky Mountain House cluster. b) Montney cluster 1. c) Montney cluster 2. d) Fox Creek cluster.*

Next, we use the *b*-values determined above to parameterize and plot the joint space-time nearest-neighbour distance (NND) distributions for each cluster, which reflect the mixing proportions of loosely and tightly clustered earthquakes (Figure 7). As an initial observation, the modal locations of all four clusters provide confirmation of some of the regional interpretations formed in section 4. In particular, we suggested that the regional clustering observed within the WCSB occurred within a distinct subset of the multidimensional inter-event NND space, namely within the $\log_{10}\eta \leq \log_{10}\eta_{tect}$ realm (i.e. within the loosely and tightly clustered domains). The remaining earthquakes were evenly distributed and occurred largely within the $\log_{10}\eta > \log_{10}\eta_{tect}$ realm (i.e. within the deep background domain). From Figure 7, it is indeed the case that all of the investigated clusters occupy the $\log_{10}\eta \leq \log_{10}\eta_{tect}$ realm.

However, *within* the $\log_{10}\eta \leq \log_{10}\eta_{tect}$ realm, several distinctions between the clusters regarding their mixing proportions can be made. The RMHC (Figure 7a) and MC1 (Figure 7b) distributions are visually similar, containing dominant populations of loosely clustered events. We can estimate the proportions of loosely clustered background events μ_{GMM}, based on a bimodal Gaussian mixture model (disregarding the deep background), which is computed as a fraction of the total seismicity within each cluster. This yields $\mu_{GMM} = 0.86$ for the RMHC and $\mu_{GMM} = 0.74$ for the MC1. The majority of earthquakes in these clusters occur somewhat closely together in space and time, but mainly as separate instances that rarely trigger future earthquakes. This loosely clustered activity of single events may reflect the type of seismicity expected to occur within the WCSB due to steady, long-term alterations to the subsurface stress field, such as long-term gas extraction and wastewater disposal (Wetmiller, 1986; Horner et al., 1994; Baranova et al., 1999; B.C. Oil and Gas Commission, 2014). Conversely, both the MC2 (Figure 7c) and FCC (Figure 7d) distributions contain distinct modes within the tightly clustered ($\log_{10}\eta < \log_{10}\eta_{thresh}$) domain. The existence of these modes indicates that a significant proportion of earthquakes are occurring very closely together and are possibly occurring in direct response to a triggering mechanism (previous earthquakes and/or external forces). The proportion of loosely clustered events in the MC2, estimated by the bimodal Gaussian mixture model, reduces to $\mu_{GMM} = 0.41$. The FCC is further distinguishable from the other clusters by its apparent lack of a loosely clustered subpopulation ($\mu_{GMM} = 0.09$). This is quite significant, as it implies that either: (i) the small population of background events occurring in this area is capable of triggering massive amounts of tightly clustered activity; or (ii) other, external triggering factors are contributing to the unnatural levels of event sequencing. Based on the spatiotemporal correlations between the increased rates of seismicity and the rise in horizontally drilled HF operations made here and in other studies, it is logical to connect HF as one of the probable causative mechanisms (discussed further below).

5.3 Event-Family Classification

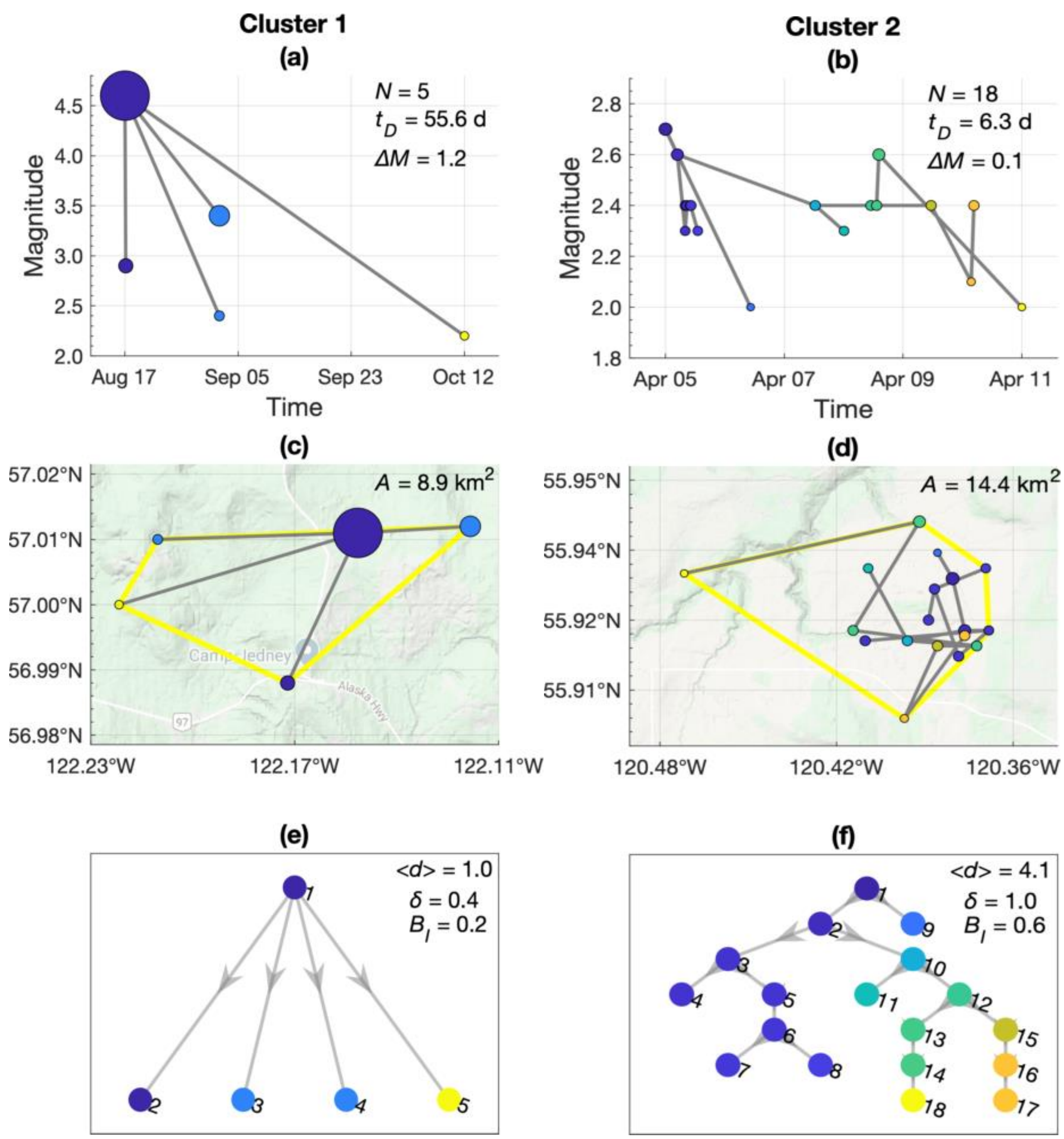


***Figure 8**: Characterizations of "burst" and "swarm" sequences found in the MC2. Circles represent earthquakes and are coloured chronologically from darkest to lightest; grey lines represent the strong links connecting them. (a,b) Event family magnitudes vs. times of occurrence.(c,d) Spatial maps of events; yellow border outlines the hull area occupied by the sequence. (e,f) Topological trees, which branch downward in time.*

In this subsection, we discretize each of the RMHC, MC1, MC2, and FCC's tightly clustered NND subpopulations into hierarchical families based on the strong links between parents and offspring events. We then categorize these families as aftershock sequences, seismic swarms, a combination of the two, or neither, based on statistical parameters outlined by Zaliapin & Ben-Zion (2013b) (see section 3.3 for a detailed description of these parameters). We consider only those families consisting of three or more events; event-pairs are discarded, as are the single events that chiefly populate the background mode.

The RMHC and MC1 are each composed of six significant sequences across 26-year periods; these sequences are small, with the largest containing five and eight events, respectively. The MC2 is composed of ten families over a 9-year period, the largest containing 17 events, while FCC is comprised of 16 sequences over a 6-year period, the largest containing 80 events. Figure 8 shows examples of two families identified in the MC2, along with their associated parameters, displayed as time-magnitude sequences, spatial maps and time-oriented, topological trees. Cluster 1 is a typical aftershock sequence while Cluster 2 is a swarm. Intuitively, Cluster 1 is characterized by a *burst* or *spray-like* shape, involving a dominant mainshock causing several lower magnitude, 1st generation offspring in multiple directions. Conversely, Cluster 2 displays a *linear* or *path-like* shape where the events are chained together gradually, the mainshock is less distinct and is succeeded by multiple generations of aftershocks.

***Table 1**: Mean and median parameter values for all significant event families within each investigated cluster.*

Cluster	**Number of Families**		**N**	$\langle d \rangle$	δ	B_I	A **(km^2)**	t_D **(days)**	ΔM
RMHC	6	Mean	3.67	1.22	0.64	0.50	8.40	1.33	0.57
		Median	3.5	1	0.58	0.5	5.93	1.00	0.25
MC1	6	Mean	4.17	1.83	0.92	0.68	23.45	23.60	0.35
		Median	3.5	2	0.855	0.58	4.42	5.52	0.25
MC2	10	Mean	7.60	2.76	1.08	0.67	26.33	24.83	0.27
		Median	5.5	3	1.125	0.71	7.19	5.23	0.1
FCC	16	Mean	27.13	7.60	1.54	0.66	20.97	32.09	0.22
		Median	20.5	6.19	1.27	0.65	17.16	15.05	0.1

Table 1 presents the mean and median parameter values for all of the detected sequences (see Figures S9-12 for the structural representations of all sequences, analogous to Figure 8). The RMHC is almost entirely composed of small bursts; these sequences have a relatively large mainshock and are followed by few (four or less) aftershocks. The mean and median values of the magnitude differential Δm are large while the normalized leaf depth δ and inverted branching number B_I are small, highlighting the spray-like nature of the sequences. The MC1 is comprised of a mixture of small bursts and swarms. Its mean and median Δm are slightly smaller and its topological parameters are larger than the RMHC's, indicating more swarm-like behavior and possibly some level of inter-event triggering. The MC2 is more swarm-like than the MC1; it consists of a large swarm, a large burst and smaller sequences. The mean and median topological parameters are larger than both previous clusters while the magnitude differential is smaller. This signifies that, apart from the large burst, the mainshocks are generally less distinct, the sequences are graphically deeper and there are more parent events and fewer leaves. The mean Δm is skewed somewhat by the large burst, which explains the disparity between it and the median value. The FCC is overwhelmingly swarm-like; 13 of the 16 sequences are chain-like in time (mean and median δ are large) and contain similarly sized events (mean and median Δm are small). The B_I values are similar to the MC1 and MC2. The remaining 3 families in the FCC consist of a large burst and two swarm-bursts (where there is a large mainshock but also many generations of foreshocks and aftershocks chained together; see Figure S12, clusters 5 and 6). It is important to note that while the average leaf depth $\langle d \rangle$ scales with the sequence size *N*, both δ and B_I do not.

The spatial extents A and timeframes t_D are slightly more difficult to contextualize as these parameters scale with *N*. The sequences within the RMHC are quite consistent; they all cover small spatial areas and decay rapidly (within days). The MC1 suffered from occasional improper location recording, where several events were placed in the same spot, resulting in inaccurate area calculations. Its time periods are variable; the two largest families are bursts and persist for over a month while the remaining smaller sequences degenerate within days. The MC2 is consistent apart from a large burst, which covers a large area and persists for several months. The remaining sequences are much more constrained spatiotemporally, including a large swarm sequence, which explains the disparity between mean and median A and t_D values. Finally, the FCC's swarm sequences are similar, with comparable A and t_D values. The largest families cover spatial areas between 30-50 km^2 and span approximately a month's time, while the smaller ones cover 10 km^2 or less and decay within two weeks. The lone identified burst sequence in the FCC occupies a comparable area but spans over half a year's time (Figure S12, cluster 4).

5.4 The Epidemic Type Aftershock Sequence Model

Finally, we use the point-process epidemic-type aftershock sequence (ETAS) model to compare the rates of seismicity, aftershock parameters and model-fitting quality between the four clusters. The ETAS model, developed by Ogata (1988, 1989), is a stochastic branching model based on an extension of the Omori-Utsu law of aftershocks (Omori, 1894; Utsu, 1961). It aims to represent the seismicity rate as the summation of a constant background term and a parameterized aftershock kernel. The rate is modeled as a function of previous activity; at time t, it is conditioned by all events M_i that satisfy $t_s \leq t_i < t$, where t_s is the start of a target window. According to the model, each seismic event has the potential to trigger its own aftershock sequence. Given the completeness magnitude M_c, the earthquake rate λ at time t is

$$\lambda(t|H_t) = \mu + \sum_{t_i<t} \frac{K e^{\alpha(M_i - M_c)}}{\left(\frac{t - t_i}{c} + 1\right)^p}, \qquad [8]$$

where H_t is the conditional earthquake occurrence history prior to time t within the target window. μ represents the independent background rate, K is the aftershock productivity, p controls the observed power law-based rate of aftershock decay, and c is the time offset between the mainshock and start of decay. The parameter α governs the degree of aftershock cascading for a given magnitude. Larger values of α imply a greater sensitivity to magnitude in the generation of aftershocks, which has been observed for great earthquakes (Ogata, 1992; Omi et al., 2014). Conversely, smaller α values reduce the significance of event magnitude on aftershock triggering. This characteristic has been linked to swarm sequences where mainshocks are less distinct (Mogi, 1963; Utsu, 1970; Ogata, 1988). Estimation of the set of parameters $\varphi = \{\mu, K, \alpha, c, p\}$ may be obtained by maximizing the log-likelihood function for $\lambda(t)$ (see Ogata, 1989).

The quality of the ETAS model fit is generally evaluated based on a transformation of occurrence times, where a new set of times τ is defined such that τ_i is the cumulative conditional intensity function at time t_i. A plot of the cumulative ETAS model rate versus the transformed time thus results in an increasing function with constant unit slope. If the model fits the data well, then a plot of the observed cumulative event count versus the transformed time will match the cumulative model with rate close to unity. We determine the model's quality-of-fit (*QOF*) to each cluster by quantifying the deviations of the observed cumulative plot from the ideal function in transformed time; the *QOF* value is computed as the normalized area present between the two plots. The maximum area value is 0.5, corresponding to the worst possible fit to the sequence; a value closer to 0 implies a better fit.

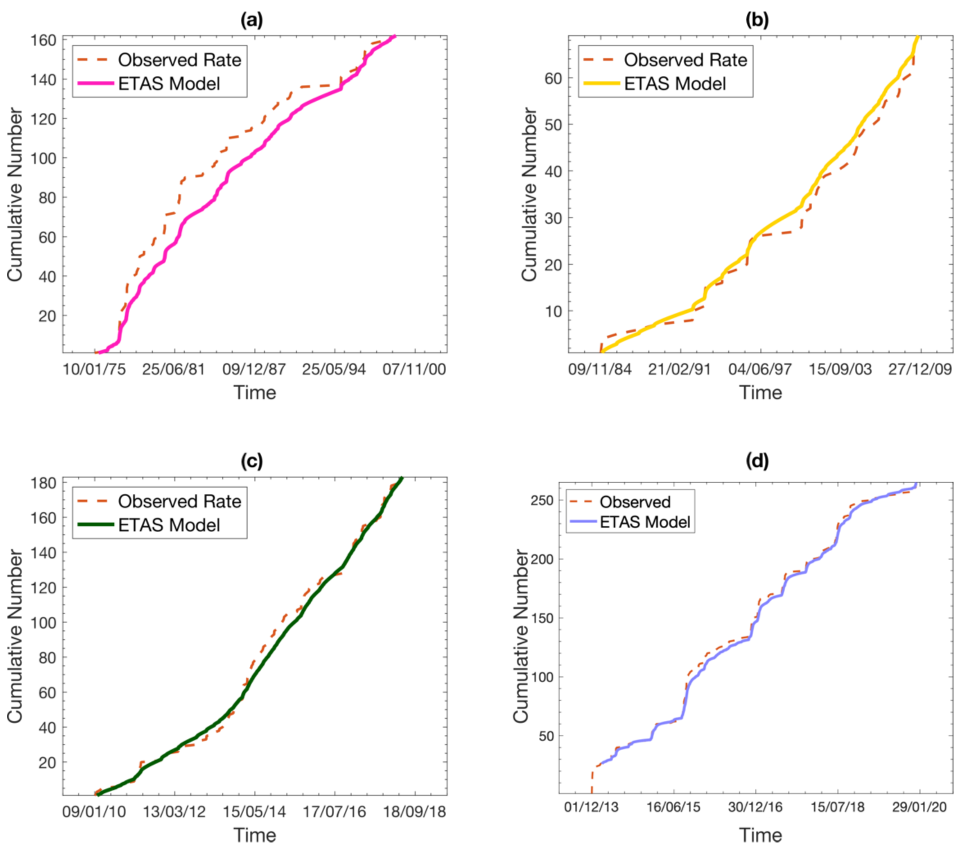


Figure 9*: ETAS Models in original time for each investigated cluster. Dashed orange lines are the observed cumulative rates of seismicity. Solid coloured lines are the optimized models. Dates are given in dd/mm/yy format. a) Rocky Mountain House cluster. b) Montney cluster 1. c) Montney cluster 2. d) Fox Creek cluster.*

Figure 9 presents the ETAS model and observed seismicity rates for each cluster, using events above the magnitude of completeness thresholds determined in subsection 5.1. The corresponding background and aftershock parameters are provided in Table 2. The model successfully converged for the RMHC, MC1, and MC2 but failed to converge to realistic parameter values for the FCC. We therefore constrained the background rate parameter μ to nearly zero and reran the simulation, achieving convergence and an improved *QOF*. Justification for this approach is provided below. The *QOF* varies for each cluster. The RMHC (Figure 9a) is by far the worst fit; the model underpredicts the rate across the entire time frame. The MC1 (Figure 9b) fit is fair; this time, the model slightly overpredicts the rate. The MC2 (Figure 9c) is fit well, while the FCC (**Figure 9**d, fixed μ) is fit very well. The type of seismicity within each cluster, as illustrated by the NND model in the previous subsections, may provide some explanation for the differing *QOF*s. The FCC is composed of large, distinct clustering sequences that are easily distinguishable across time. The ETAS model appears to be correctly modeling these individual spikes in the rate. Conversely, the RMHC is composed of a loosely clustered type of seismicity that produces few clearly defined sequences and mainly manifests as elevated background activity. Consequently, the ETAS model may have encountered difficulty in replicating the gradually increasing rate with an absence of aftershock sequences. The steady RMHC rate may be more suitably modeled by an increased background parameter μ without the aftershock kernel. The improved *QOF* of the MC2 over the MC1 could be further evidence that the ETAS model requires defined sequences with which to optimize its fit, although the MC1 dataset is likely too small for a realistic comparison.

Table 2: *ETAS model parameters for each investigated cluster.*

Cluster	M_c	N	μ	K	c	p	α	QOF
RMHC	2.3	162	0.0047	0.26	0.24	1.06	0.76	0.072
MC1	2.7	69	0.0033	3.09	0.0019	0.81	0.84	0.042
MC2	2.5	183	0.016	1.0	0.022	0.84	0.39	0.020
FCC*	2.5	242	0.000002	0.42	0.62	1.25	0.47	0.017

* μ is held constant.

It is problematic to compare the parameter values between all the clusters, as the model may not be accurately representing the rate in the cases of the RMHC and MC1. For the MC2 and FCC, the ETAS model seems able to accurately capture the swarm activity in addition to the aftershock sequences for which it was intended. 21 out of the 26 event sequences identified within the MC2 and FCC do not have distinct mainshocks but are instead composed of similarly sized events. This appears to be reflected in the optimized ETAS α parameter, which governs the dependence on magnitude in the generation of aftershocks. α is low for both clusters, indicating that the model may be correctly identifying the persistent nature of the swarm sequences, which continue to produce generations of offspring despite the lack of an obvious mainshock. The p parameter, which controls the sequence decay rate, is low for the MC2 and high for the FCC. This might again be attributable to the more gradual rate increase within the MC2 (Figure 9c) compared to the sharp spikes due to the large, tightly constrained sequences observed within the FCC (Figure 9d).

Finally, within the FCC, a better fit and a realistic convergence are achieved by constraining the background rate μ. Generally, the ETAS μ parameter is optimized to represent independent seismicity as a non-zero constant, which allows for the generation of the conditional aftershock rate. Consequently, μ may relate to the nearest-neighbour loosely clustered subpopulation, as they both represent the subset of background earthquakes occasionally preceding clustered activity. While the three other investigated clusters contained high amounts of loosely clustered activity, likely resulting from long-term stress perturbations caused by fluid extraction and injection, the FCC is comprised of tightly clustered sequences and relatively few background events. This suggests that external factors may be contributing directly to the triggering of seismicity. The absence of the loosely clustered population was used as motivation for the fixed $\mu \approx 0$ constraint. When μ was similarly fixed on a trial basis for the other clusters, either the QOF was worse or the model did not converge. This does makes sense, as the seismicity in those areas is largely made up of an elevated background with few significant sequences, and so assuming an absence of background seismicity would be inaccurate. The comparatively high-quality fit of the model to the FCC also provides some confirmation that the ETAS model seems to perform better on a dataset containing separable, tightly connected sequences. For seismic clusters resembling an elevated background rate, on the other hand, it may not be able to predict what events are or are not part of an earthquake sequence.

6 Discussion

Many regional studies of anthropogenic seismicity have described an elevation in background seismicity during and immediately after potential earthquake-inducing processes such as geothermal energy production, wastewater disposal, hydraulic fracturing etc. (Lombardi et al., 2010; Llenos & Michael, 2013; Schoenball et al., 2015; Maghsoudi et al., 2016; Zaliapin & Ben-Zion, 2016; Schoenball & Ellsworth, 2017; Martínez-Garzón et al., 2018). In sections 4 and 5, we found evidence that an induced background subpopulation of earthquakes in the WCSB was elevated near areas of human activity, which could be separated from the tectonic rate in the nearest-neighbour distance (NND) distributions. This elevated background was more specifically a collection of earthquakes that was loosely clustered in certain areas over limited time frames; its inter-event space-time distances were measurably lower than the regional Poisson-like background but larger than typical clustering activity. This is likely to have been the case in other regions as well. For example, Zaliapin & Ben-Zion (2016) used the NND method to examine seismicity within selected regions across the state of California; these regions were either (i) dominated by induced activity related to geothermal operations, (ii) a mixture of tectonic and induced earthquakes, or (iii) characterized entirely by tectonic events. In every instance, the anthropogenic seismicity contained a background mode that was situated much closer to the clustered mode than the natural activity. The separation threshold value between subpopulations was smallest for type (i) areas and largest for areas of type (iii); this difference spanned several orders of magnitude. Similar observations were made by Schoenball et al. (2015), who assessed the seismicity surrounding and within the Coso geothermal field in southern California. The induced seismicity again contained a background element positioned closer to the clustered mode compared to the surrounding earthquakes, which were attributed to tectonic and magmatic sources. Interestingly, the background mode held a dominant mixing proportion during the coproduction intervals and also extended further downward along the diagonal towards larger T and smaller R, suggesting that a large proportion of the induced activity was acting as an elevated Poisson process of independent events. Vasylkivska & Huerta (2017) studied the rapidly increasing earthquake rate in Oklahoma associated with large-scale wastewater injection. During the pre-injection time interval, only a single component was present in the two-dimensional NND distribution, located far in the upper right section of the space. This mode largely dissipated during the co-injection intervals and was replaced by clustered and background components located closer to the center. As discussed in these studies, induced seismicity tends to increase earthquake rates locally and across limited time spans. This is reflected quite clearly in the joint NND distributions, where populations of the elevated induced background manifest more centrally and closer to the clustered population. This effect is starkly evident in regions governed by a low natural seismicity rate like the WCSB and Oklahoma, where the tectonic and induced background components can be separated by inspection, but is also discernable in naturally active areas such as California.

We also observed a tightly clustered earthquake subpopulation in the NND distributions, which occurred at very small inter-event distance values and appeared highly correlated with the loosely clustered background. We found that tightly clustered event sequences were nearly four times as likely to be strongly linked to a loosely clustered event compared to a natural background event. Both groups of earthquakes made up the entirety of discernable induced clustering in the region and both grew disparately in relative proportion over recent times, coinciding with the growing use of hydraulic fracturing technology. However, we discovered that different triggering mechanisms produced differences in the mixing proportions of these earthquake groups, with hydraulic fracturing operations resulting in much larger proportions of tightly clustered earthquakes than conventional hydrocarbon production and wastewater disposal. Based on the results in section 5, several further insights may be gathered regarding the induced seismic clustering within the WCSB. First, wherever fluid injection was suspected as the seismogenic mechanism, either through wastewater disposal or HF (i.e. the MC1, MC2 and FCC), the clustering appeared more swarm-like and reminiscent of ductile failure. The lone case of fluid extraction-related seismicity, the RMHC, resulted in an elevated background rate and occasional burst-like sequencing, more suggestive of brittle failure. Second, HF operations seemed to trigger greater swarm-like behavior than wastewater injection. Results of the NND and ETAS model applications indicated that the MC2 and FCC contained higher levels of swarm seismicity than did the MC1. This may have been a consequence of the differences in injection volume and rate between cases, as well as the horizontal orientation of many new HF wells, which allows fluid and stress perturbations to be forced through a much larger volume of rock in the short term (King, 2010; Smith & Montgomery, 2015). There is also a possibility that HF-induced swarms can occur predominantly at lower magnitude levels, as evidenced by the elevated Gutenberg-Richter *b*-value within the FCC. Third, HF is capable of, albeit less frequently, triggering larger aftershock sequences, where the migrating fluid or stress perturbation may be traveling into the crystalline basement and interacting with critically stressed faults, similar to the triggering mechanism attributed to wastewater disposal. Fourth, these aftershock sequences appear to result in connected earthquakes that span longer time frames than do the swarms of comparable sequence size. The swarms within the MC2 and FCC are almost all tightly constrained in space and time relative to their size, which may correlate with the spatial and temporal extent of their associated stimulating HF operations. This also suggests that the two types of clusters are caused by different mechanisms; the bursts by fluid intermingling with critically oriented faults resulting in a large event, which *then* triggers multiple offspring events in a conventional aftershock manner, versus the swarms where no distinct mainshock is present, yet multiple offspring continue to transpire as the pumped fluid repeatedly disturbs nearby faults (Schultz, Stern, Novakovic, et al., 2015; Bao & Eaton, 2016; Eaton, 2018). A possible example of swarms being directly related to HF activity is the notable resemblance between the identified swarms in the FCC and the largest sequence in the MC2 near Dawson Creek, B.C. (Figure S12, clusters 1, 2, and 3, and Figure S11, cluster 1). They each occurred within kilometers of active HF operations and are structurally similar; the Dawson Creek swarm is a smaller-scale version of the FCC sequences (comparable events chained together with a rapid decay time, relative to its size). It is possible that the likenesses in the fundamental structuring of these clusters may be reflective of their shared triggering mechanism. The disparities in their size and scope may be due to factors such as different pumping rates/pressures/times or total volume of fluid injected, in addition to local geologic factors.

A critical issue with the models employed in this study are their dependence on earthquake sampling statistics, which vary in representative accuracy and precision based on sample size and variance. The associated uncertainties for the Gutenberg-Richter and ETAS parameter estimates can be quantified, although the ETAS parameter errors computed in this study were unstable and hence removed; this needs to be addressed in future work. The NND calculations are less well constrained. Zaliapin & Ben-Zion (2015) examined the potential artifacts of catalogue inconsistencies on the results of their cluster analyses within southern California. They found that location errors and short-term incompleteness can lead to an overestimation of background seismicity and a corresponding underestimation of clustered earthquakes. This phenomenon appears independent of any particular cluster identification technique, as long as it is based upon parent-offspring relationships. With respect to the long-term regional completeness, Zaliapin & Ben-Zion (2013a) showed that NND distributions remain stable (up to a certain magnitude cut-off, after which the distribution becomes unimodal and contains only single events). In this study, we nevertheless introduce some bias by applying a cut-off below the regional estimated completeness level to enrich sampling. The networks contributing to the Composite Alberta Seismicity Catalogue have undergone numerous changes and improvements across time, particularly near areas of suspected induced activity including along the Duvernay and Montney formations (for details, we refer to Adams & Halchuk (2003) and Cui & Atkinson (2016)). This implies that the differences observed in earthquake distribution over time and between clusters may be partly the result of a changing M_c. However, we point to the disparate changes in rate between the earthquake subpopulations found in section 4, which hold for cut-off magnitudes up to the regional estimated completeness (Figure S8), as well as the suspicious spatial relationship between the clustered events and human operations as evidence that the recent changes in earthquake distribution are significant, and not entirely due to the enhancement of the regional network.

7 Conclusions

A statistical analysis of catalogued seismicity within the Western Canada Sedimentary Basin (WCSB) was performed. We conducted a regional study in section 4, analyzing space-time-magnitude inter-event nearest-neighbour distance (NND) distributions across Alberta and eastern B.C. over time, beginning from the first observed instance of induced activity in 1975 up to the nearly present hydraulic fracturing (HF) related activity in 2018. Analysis over the entire time frame revealed the existence of a tri-modal inter-event distance distribution, where events generally appeared to occur either: (i) very closely together in space and time (within the tightly clustered mode); (ii) moderately close together (within the loosely clustered mode); or (iii) as a stationary, space-inhomogeneous Poisson process (the deep background). Analysis over separated time intervals demonstrated that a disproportionate increase in both the loosely and tightly clustered earthquake components occurred between 1975-2009 and 2010-2018, where the first interval predates the broad-scale implementation of HF technology. The first two modes of inter-event distance distribution are believed to reflect HF operation, wastewater disposal and conventional hydrocarbon production-related earthquakes. The third mode is inferred to represent natural background seismicity, based on its uniform spatial distribution and large inter-event distances (resembling a point process). The spatial distribution of the natural background events contrasts with the localized loosely and tightly clustered subpopulations occurring at smaller inter-event distance values. We posit that the majority of induced activity occurs within these clustered modes, and that their increasing prevalence within the region is attributable to the growing usage of HF technology.

In section 5, we investigated four cases of induced seismic clustering within the WCSB. Frequency-magnitude statistics were assessed using the Gutenberg-Richter (GR) relation. Inter-event space-time-magnitude distributions and individual clustering properties were evaluated with the nearest-neighbour distance (NND) method and the epidemic type aftershock sequence (ETAS) model. It was determined that the seismicity triggered by conventional gas extraction near Rocky Mountain House (the RMHC) and the wastewater disposal-related earthquakes near Fort St. John (the MC1) primarily manifested as discrete events, loosely clustered in space and time, that occupied the middle mode of the inter-event NND distributions. These events were characterized by low-to-moderate GR *b*-values and poorer fits to the ETAS model. The few tightly clustered event sequences that did occur were small and decayed rapidly. The more recent clustering along the Montney formation (the MC2) also occupied the loosely clustered domain but contained an additional mode within the tightly clustered domain. These earthquakes were characterized by a slight increase in *b*-value from the previous period and were better fit by the ETAS model, with a reduced magnitude-sensitivity parameter. We identified more event sequences over this period, which were larger and more swarm-like. Finally, near Fox Creek (the FCC), substantial HF-related activity occurred almost entirely within the tightly clustered domain, in stark contrast with the previous clusters. These earthquakes formed a very steep frequency-magnitude distribution with a high *b*-value and fit the ETAS model exceptionally well, particularly when the background rate parameter was constrained to nearly zero, with a low magnitude sensitivity comparable to the MC2. The many detected event sequences were large, distinctly separable and overwhelmingly swarm-like.

Acknowledgments and Data

This work is supported by the Natural Sciences and Engineering Research Council of Canada and is a part of the Canadian Induced Seismicity Collaboration. The seismic database was

compiled from the Composite Alberta Seismicity Catalogue, available online at www.inducedseismicity.ca (last accessed June 2020).

References

Adams, J., & Halchuk, S. (2003). Fourth Generation Seismic Hazard Maps of Canada: Values for Over 650 Canadian Localities Intended for the 2005 National Building Code of Canada. https://doi.org/10.4095/214223

Ader, T. J., & Avouac, J.-P. (2013). Detecting periodicities and declustering in earthquake catalogs using the Schuster spectrum, application to Himalayan seismicity. Earth and Planetary Science Letters, 377–378(Journal Article), 97–105. https://doi.org/10.1016/j.epsl.2013.06.032

Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

Aki, K. (1965). Maximum Likelihood Estimate of b in the Formula logN=a-bM and its Confidence Limits. Bull. Earthq. Res. Inst. Tokyo Univ, 43, 237–239.

Atkinson, G. M. (2017). Strategies to prevent damage to critical infrastructure due to induced seismicity. FACETS, 2(1), 374–394. https://doi.org/10.1139/facets-2017-0013

Atkinson, G. M., Eaton, D. W., Ghofrani, H., Walker, D., Cheadle, B., Schultz, R., et al. (2016). Hydraulic Fracturing and Seismicity in the Western Canada Sedimentary Basin. Seismological Research Letters, 87(3), 631–647. https://doi.org/10.1785/0220150263

Atkinson, G. M., Eaton, D. W., & Igonin, N. (2020). Developments in understanding seismicity triggered by hydraulic fracturing. Nature Reviews Earth & Environment, 1(5), 264–277. https://doi.org/10.1038/s43017-020-0049-7

Baiesi, M., & Paczuski, M. (2004). Scale-free networks of earthquakes and aftershocks. Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics, 69(6). https://doi.org/10.1103/PhysRevE.69.066106

Bao, X., & Eaton, D. W. (2016). Fault activation by hydraulic fracturing in western Canada. Science (New York, N.Y.), 354(6318), 1406–1409. https://doi.org/10.1126/science.aag2583

Baranova, V., Mustaqeem, A., & Bell, S. (1999). A model for induced seismicity caused by hydrocarbon production in the Western Canada Sedimentary Basin. Canadian Journal of Earth Sciences, 36(1), 47–64. https://doi.org/10.1139/cjes-36-1-47

B.C. Oil and Gas Commission. (2012). Investigation of observed seismicity in the Horn River Basin. Victoria, B.C.: BC Oil and Gas Commission.

B.C. Oil and Gas Commission. (2014). Investigation of observed seismicity in the Montney Trend. Victoria, B.C.: BC Oil and Gas Commission.

Ben-Zion, Y., & Lyakhovsky, V. (2006). Analysis of aftershocks in a lithospheric model with seismogenic zone governed by damage rheology. Geophysical Journal International, 165(1), 197–210. https://doi.org/10.1111/j.1365-246X.2006.02878.x

Brodsky, E. E., & Lajoie, L. J. (2013). Anthropogenic Seismicity Rates and Operational Parameters at the Salton Sea Geothermal Field. Science, 341(6145), 543–546. https://doi.org/10.1126/science.1239213

Brudzinski, M. R., & Kozłowska, M. (2019). Seismicity induced by hydraulic fracturing and wastewater disposal in the Appalachian Basin, USA: a review. Acta Geophysica, 67(1), 351–364. https://doi.org/10.1007/s11600-019-00249-7

Cao, A., & Gao, S. S. (2002). Temporal variation of seismic b-values beneath northeastern Japan island arc. Geophysical Research Letters, 29(9). https://doi.org/10.1029/2001GL013775

Clerc, F., Harrington, R. M., Liu, Y., & Gu, Y. J. (2016). Stress drop estimates and hypocenter relocations of induced seismicity near Crooked Lake, Alberta. Geophysical Research Letters, 43(13), 6942–6951. https://doi.org/10.1002/2016GL069800

Console, R., Jackson, D. D., & Kagan, Y. Y. (2010). Using the ETAS Model for Catalog Declustering and Seismic Background Assessment. Pure and Applied Geophysics, 167(6), 819–830. https://doi.org/10.1007/s00024-010-0065-5

Cui, L., & Atkinson, G. M. (2016). Spatiotemporal Variations in the Completeness Magnitude of the Composite Alberta Seismicity Catalog (CASC). Seismological Research Letters, 87(4), 853–863. https://doi.org/10.1785/0220150268

Cui, L., Fereidoni, A., & Atkinson, G. (2015). Compilation of Composite Alberta Seismicity Catalog (CASC) for Earthquake Hazard from Induced Seismicity in Alberta.

Deng, K., Liu, Y., & Harrington, R. M. (2016). Poroelastic stress triggering of the December 2013 Crooked Lake, Alberta, induced seismicity sequence. Geophysical Research Letters, 43(16), 8482–8491. https://doi.org/10.1002/2016GL070421

Doglioni, C. (2018). A classification of induced seismicity. Geoscience Frontiers, 9(6), 1903–1909. https://doi.org/10.1016/j.gsf.2017.11.015

Eaton, D. W. (2018). Passive seismic monitoring of induced seismicity: fundamental principles and application to energy technologies. Cambridge University Press.

Eaton, D. W., & Mahani, A. B. (2015). Focal Mechanisms of Some Inferred Induced Earthquakes in Alberta, Canada. Seismological Research Letters, 86(4), 1078–1085. https://doi.org/10.1785/0220150066

Eaton, D. W., Igonin, N., Poulin, A., Weir, R., Zhang, H., Pellegrino, S., & Rodriguez, G. (2018). Induced Seismicity Characterization during Hydraulic-Fracture Monitoring with a Shallow-Wellbore Geophone Array and Broadband Sensors. Seismological Research Letters, 89(5), 1641–1651. https://doi.org/10.1785/0220180055

Farrell, J., Husen, S., & Smith, R. B. (2009). Earthquake swarm and b-value characterization of the Yellowstone volcano-tectonic system. Journal of Volcanology and Geothermal Research, 188(1), 260–276. https://doi.org/10.1016/j.jvolgeores.2009.08.008

Gardner, J. K., & Knopoff, L. (1974). Is the sequence of earthquakes in Southern California, with aftershocks removed, Poissonian? Bulletin of the Seismological Society of America, 64(5), 1363–1367.

Ghofrani, H., & Atkinson, G. M. (2016). A preliminary statistical model for hydraulic fracture-induced seismicity in the Western Canada Sedimentary Basin. Geophysical Research Letters, 43(19), 172. https://doi.org/10.1002/2016GL070042

Goebel, T. H. W., Hauksson, E., Aminzadeh, F., & Ampuero, J.-P. (2015). An objective method for the assessment of fluid injection-induced seismicity and application to tectonically active regions in central California. Journal of Geophysical Research: Solid Earth, 120(10), 7013–7032. https://doi.org/10.1002/2015JB011895

Grigoli, F., Cesca, S., Priolo, E., Rinaldi, A. P., Clinton, J. F., Stabile, T. A., et al. (2017). Current challenges in monitoring, discrimination, and management of induced seismicity related to underground industrial activities: A European perspective. Reviews of Geophysics, 55(2), 310–340. https://doi.org/10.1002/2016RG000542

Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). The EM Algorithm. In The elements of statistical learning: data mining, inference, and prediction (2nd ed., pp. 236–243). New York: Springer.

Hicks, A. L. (2011). Clustering in multidimensional spaces with applications to statistical analysis of earthquake clustering. ProQuest Dissertations Publishing.

Horner, R. B., Barclay, J. E., & MacRae, J. M. (1994). Earthquakes and Hydrocarbon Production in the Fort St. John Area of Northeastern British Columbia. Canadian Journal of Exploration Geophysics, 30(1), 38–50.

Horton, S. (2012). Disposal of Hydrofracking Waste Fluid by Injection into Subsurface Aquifers Triggers Earthquake Swarm in Central Arkansas with Potential for Damaging Earthquake. Seismological Research Letters, 83(2), 250–260. https://doi.org/10.1785/gssrl.83.2.250

Kagan, Y. Y. (1991). Fractal dimension of brittle fracture. Journal of Nonlinear Science, 1(1), 1–16. https://doi.org/10.1007/BF01209146

Keranen, K. M., & Weingarten, M. (2018). Induced Seismicity. Annual Review of Earth and Planetary Sciences, 46(1), 149–174. https://doi.org/10.1146/annurev-earth-082517-010054

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

Kettlety, T., Verdon, J. P., Werner, M. J., Kendall, J. M., & Budge, J. (2019). Investigating the role of elastostatic stress transfer during hydraulic fracturing-induced fault activation. Geophysical Journal International, (Journal Article). https://doi.org/10.1093/gji/ggz080

Kettlety, T., Verdon, J. P., Werner, M. J., & Kendall, J. M. (2020). Stress Transfer From Opening Hydraulic Fractures Controls the Distribution of Induced Seismicity. Journal of

Geophysical Research: Solid Earth, 125(1), e2019JB018794. https://doi.org/10.1029/2019JB018794

King, G. E. (2010). Thirty Years of Gas Shale Fracturing: What Have We Learned? In SPE-133456-MS (p. 50). SPE: Society of Petroleum Engineers. https://doi.org/10.2118/133456-MS

Kosobokov, V., & Mazhkenov, S. A. (2013). On Similarity in the Spatial Distribution of Seismicity. In Computational Seismology and Geodynamics (Vol. 22). https://doi.org/10.1029/CS001p0006

Kumazawa, T., & Ogata, Y. (2013). Quantitative description of induced seismic activity before and after the 2011 Tohoku-Oki earthquake by nonstationary ETAS models. Journal of Geophysical Research: Solid Earth, 118(12), 6165–6182. https://doi.org/10.1002/2013JB010259

Langenbruch, C., & Shapiro, S. A. (2010). Decay rate of fluid-induced seismicity after termination of reservoir stimulations. Geophysics, 75(6), MA53–MA62. https://doi.org/10.1190/1.3506005

Langenbruch, C., Ellsworth, W. L., Woo, J.-U., & Wald, D. J. (2020). Value at Induced Risk: Injection-Induced Seismic Risk From Low-Probability, High-Impact Events. Geophysical Research Letters, 47(2), e2019GL085878. https://doi.org/10.1029/2019GL085878

Lee, K.-K., Ellsworth, W. L., Giardini, D., Townend, J., Ge, S., Shimamoto, T., et al. (2019). Managing injection-induced seismic risks. Science, 364(6442), 730. https://doi.org/10.1126/science.aax1878

Lei, X., Yu, G., Ma, S., Wen, X., & Wang, Q. (2008). Earthquakes induced by water injection at ~3 km depth within the Rongchang gas field, Chongqing, China. Journal of Geophysical Research - Solid Earth, 113(B10), B10310. https://doi.org/10.1029/2008JB005604

Llenos, A. L., & Michael, A. J. (2013). Modeling Earthquake Rate Changes in Oklahoma and Arkansas: Possible Signatures of Induced Seismicity. Bulletin of the Seismological Society of America, 103(5), 2850–2861. https://doi.org/10.1785/0120130017

Llenos, A. L., & Michael, A. J. (2016). Characterizing potentially induced earthquake rate changes in the Brawley Seismic Zone, southern California. Bulletin of the Seismological Society of America, 106(5), 2045–2062. https://doi.org/10.1785/0120150053

Lombardi, A. M., Cocco, M., & Marzocchi, W. (2010). On the Increase of Background Seismicity Rate during the 1997-1998 Umbria-Marche, Central Italy, Sequence: Apparent Variation or Fluid-Driven Triggering? Bulletin of the Seismological Society of America, 100(3), 1138–1152. https://doi.org/10.1785/0120090077

Maghsoudi, S., Eaton, D. W., & Davidsen, J. (2016). Nontrivial clustering of microseismicity induced by hydraulic fracturing. Geophysical Research Letters, 43(20), 10-10,679. https://doi.org/10.1002/2016GL070983

Maghsoudi, S., Baró, J., Kent, A., Eaton, D., & Davidsen, J. (2018). Interevent Triggering in Microseismicity Induced by Hydraulic Fracturing. Bulletin of the Seismological Society of America, 108(3A), 1133–1146. https://doi.org/10.1785/0120170368

Martínez-Garzón, P., Zaliapin, I., Ben-Zion, Y., Kwiatek, G., & Bohnhoff, M. (2018). Comparative Study of Earthquake Clustering in Relation to Hydraulic Activities at Geothermal Fields in California. Journal of Geophysical Research: Solid Earth, 123(5), 4041–4062. https://doi.org/10.1029/2017JB014972

McClure, M., Gibson, R., Chiu, K.-K., & Ranganath, R. (2017). Identifying potentially induced seismicity and assessing statistical significance in Oklahoma and California: Induced Seismicity Statistics. Journal of Geophysical Research: Solid Earth, (Journal Article). https://doi.org/10.1002/2016JB013711

McGarr, A., & Barbour, A. J. (2017). Wastewater Disposal and the Earthquake Sequences During 2016 Near Fairview, Pawnee, and Cushing, Oklahoma: Induced Earthquake Sequences During 2016. Geophysical Research Letters, 44(18), 9330–9336. https://doi.org/10.1002/2017GL075258

Mogi, K. (1963). Some Discussions on Aftershocks, Foreshocks and Earthquake Swarms : the Fracture of a Semi-infinite Body Caused by an Inner Stress Origin and Its Relation to the Earthquake Phenomena (Third Paper). Bull. Earthq. Res. Inst. Tokyo Univ, 41, 615–658.

Morita, Y., Nakao, S., & Hayashi, Y. (2006). A quantitative approach to the dike intrusion process inferred from a joint analysis of geodetic and seismological data for the 1998 earthquake swarm off the east coast of Izu Peninsula, central Japan. Journal of Geophysical Research: Solid Earth, 111(B6). https://doi.org/10.1029/2005JB003860

Ogata, Y. (1988). Statistical Models for Earthquake Occurrences and Residual Analysis for Point Processes. Journal of the American Statistical Association, 83(401), 9–27. https://doi.org/10.1080/01621459.1988.10478560

Ogata, Y. (1989). Statistical model for standard seismicity and detection of anomalies by residual analysis. Tectonophysics, 169(1), 159–174. https://doi.org/10.1016/0040-1951(89)90191-1

Ogata, Y. (1992). Detection of precursory relative quiescence before great earthquakes through a statistical model. Journal of Geophysical Research, 97(B13), 19845. https://doi.org/10.1029/92JB00708

Omi, T., Ogata, Y., Hirata, Y., & Aihara, K. (2014). Estimating the ETAS model from an early aftershock sequence. Geophysical Research Letters, 41(3), 850–857. https://doi.org/10.1002/2013GL058958

Omori, F. (1894). On the aftershocks of earthquakes. Journal of the College of Science, Imperial University of Tokyo, (7), 111–200.

Reasenberg, P. (1985). Second-order moment of central California seismicity, 1969–1982. Journal of Geophysical Research: Solid Earth, 90(B7), 5479–5495. https://doi.org/10.1029/JB090iB07p05479

Rogers, G. C., & Horner, R. B. (1991). An Overview of Western Canadian Seismicity. In D. B. Slemmons, E. R. Engdahl, M. D. Zoback, & D. D. Blackwell (Eds.), Neotectonics of North America (Vol. 1, pp. 69–76). Geological Society of America.

Sadovsky, M. A., Golubeva, T. V., Pisarenko, V. F., & Shninman, M. G. (1984). Characteristic dimension of rocks and hierarchical properties of seismicity. Izy. Acad. Sci. USSR. Phys. Solid Earth, 20, 87–95.

Sagiya, T., Stein, R. S., & Toda, S. (2002). Evidence from the ad 2000 Izu islands earthquake swarm that stressing rate governs seismicity. Nature, 419(6902), 58–61. https://doi.org/10.1038/nature00997

Schaefer, A. M., Daniell, J. E., & Wenzel, F. (2017). The smart cluster method: Adaptive earthquake cluster identification and analysis in strong seismic regions. Journal of Seismology, 21(4), 965–985. https://doi.org/10.1007/s10950-017-9646-4

Schoenball, M., & Ellsworth, W. L. (2017). A Systematic Assessment of the Spatiotemporal Evolution of Fault Activation Through Induced Seismicity in Oklahoma and Southern Kansas. Journal of Geophysical Research: Solid Earth, 122(12), 10,189-10,206. https://doi.org/10.1002/2017JB014850

Schoenball, M., Davatzes, N. C., & Glen, J. M. G. (2015). Differentiating induced and natural seismicity using space-time-magnitude statistics applied to the Coso Geothermal field. Geophysical Research Letters, 42(15), 6221–6228. https://doi.org/10.1002/2015GL064772

Schultz, R., Stern, V., & Gu, Y. J. (2014). An investigation of seismicity clustered near the Cordel Field, west central Alberta, and its relation to a nearby disposal well. Journal of Geophysical Research: Solid Earth, 119(4), 3410–3423. https://doi.org/10.1002/2013JB010836

Schultz, R., Stern, V., Gu, Y. J., & Eaton, D. (2015). Detection Threshold and Location Resolution of the Alberta Geological Survey Earthquake Catalogue. Seismological Research Letters, 86(2A), 385–397. https://doi.org/10.1785/0220140203

Schultz, R., Stern, V., Novakovic, M., Atkinson, G., & Gu, Y. J. (2015). Hydraulic fracturing and the Crooked Lake Sequences: Insights gleaned from regional seismic networks. Geophysical Research Letters, 42(8), 2750–2758. https://doi.org/10.1002/2015GL063455

Schultz, R., Mei, S., Pană, D., Stern, V., Gu, Y. J., Kim, A., & Eaton, D. (2015). The Cardston Earthquake Swarm and Hydraulic Fracturing of the Exshaw Formation (Alberta Bakken Play). Bulletin of the Seismological Society of America, 105(6), 2871–2884. https://doi.org/10.1785/0120150131

Schultz, R., Corlett, H., Haug, K., Kocon, K., MacCormack, K., Stern, V., & Shipman, T. (2016). Linking fossil reefs with earthquakes: Geologic insight to where induced seismicity occurs in Alberta. Geophysical Research Letters, 43(6), 2534–2542. https://doi.org/10.1002/2015GL067514

Schultz, R., Wang, R., Gu, Y. J., Haug, K., & Atkinson, G. (2017). A seismological overview of the induced earthquakes in the Duvernay play near Fox Creek, Alberta. Journal of Geophysical Research: Solid Earth, 122(1), 492–505. https://doi.org/10.1002/2016JB013570

Schultz, R., Atkinson, G., Eaton, D. W., Gu, Y. J., & Kao, H. (2018). Hydraulic fracturing volume is associated with induced earthquake productivity in the Duvernay play. Science, 359(6373), 304–308. https://doi.org/10.1126/science.aao0159

Schultz, R., Skoumal, R. J., Brudzinski, M. R., Eaton, D., Baptie, B., & Ellsworth, W. (2020). Hydraulic Fracturing Induced Seismicity. Reviews of Geophysics, e2019RG000695. https://doi.org/10.1029/2019RG000695

Schwarz, G. (1978). Estimating the Dimension of a Model. Ann. Statist., 6(2), 461–464. https://doi.org/10.1214/aos/1176344136

Scuderi, M. M., Collettini, C., & Marone, C. (2017). Frictional stability and earthquake triggering during fluid pressure stimulation of an experimental fault. Earth and Planetary Science Letters, 477(Journal Article), 84–96. https://doi.org/10.1016/j.epsl.2017.08.009

Segall, P. (1985). Stress and subsidence resulting from subsurface fluid withdrawal in the epicentral region of the 1983 Coalinga Earthquake. Journal of Geophysical Research, 90(B8), 6801. https://doi.org/10.1029/JB090iB08p06801

Segall, P., Grasso, J., & Mossop, A. (1995). Poroelastic stressing and induced seismicity near the Lacq gas field, southwestern France. International Journal of Rock Mechanics and Mining Sciences and Geomechanics Abstracts, 32(4), A152–A152. https://doi.org/10.1016/0148-9062(95)96941-4

Shi, Y., & Bolt, B. (1982). The standard error of the magnitude-frequency b value. Bulletin of the Seismological Society of America, 72(5), 1677.

Skoumal, R. J., Brudzinski, M. R., & Currie, B. S. (2015). Earthquakes Induced by Hydraulic Fracturing in Poland Township, Ohio. Bulletin of the Seismological Society of America, 105(1), 189–197. https://doi.org/10.1785/0120140168

Smith, M. B., & Montgomery, C. T. (2015). Hydraulic fracturing. Boca Raton, Florida: CRC Press.

Utsu, T. (1961). A statistical study of the occurrence of aftershocks. Geophysical Magazine, (30), 521–605.

Utsu, T. (1966). A Statistical Significance Test of the Difference in b-value between Two Earthquake Groups. Journal of Physics of the Earth, 14(2), 37–40. https://doi.org/10.4294/jpe1952.14.37

Utsu, T. (1970). Aftershocks and earthquake statistics (II) - Further investigation of aftershocks and other earthquake sequences based on a new classification of earthquake sequences. J. Fac. Sci. Hokkaido Univ., 3, 197–266.

Vasylkivska, V. S., & Huerta, N. J. (2017). Spatiotemporal distribution of Oklahoma earthquakes: Exploring relationships using a nearest-neighbor approach. Journal of Geophysical Research: Solid Earth, 122(7), 5395–5416. https://doi.org/10.1002/2016JB013918

Wang, R., Gu, Y. J., Schultz, R., Kim, A., & Atkinson, G. (2016). Source analysis of a potential hydraulic-fracturing-induced earthquake near Fox Creek, Alberta. Geophysical Research Letters, 43(2), 564. https://doi.org/10.1002/2015GL066917

Wetmiller, R. J. (1986). Earthquakes near Rocky Mountain House, Alberta, and their relationship to gas production facilities. Canadian Journal of Earth Sciences, 23(2), 172–181. https://doi.org/10.1139/e86-020

Wiemer, S., & Wyss, M. (2000). Minimum Magnitude of Completeness in Earthquake Catalogs: Examples from Alaska, the Western United States, and Japan. Bulletin of the Seismological Society of America, 90(4), 859–869. https://doi.org/10.1785/0119990114

Zaliapin, I., & Ben-Zion, Y. (2013a). Earthquake clusters in southern California I: Identification and stability. Journal of Geophysical Research: Solid Earth, 118(6), 2847–2864. https://doi.org/10.1002/jgrb.50179

Zaliapin, I., & Ben-Zion, Y. (2013b). Earthquake clusters in southern California II: Classification and relation to physical properties of the crust. Journal of Geophysical Research: Solid Earth, 118(6), 2865–2877. https://doi.org/10.1002/jgrb.50178

Zaliapin, I., & Ben-Zion, Y. (2015). Artefacts of earthquake location errors and short-term incompleteness on seismicity clusters in southern California. Geophysical Journal International, 202(3), 1949–1968. https://doi.org/10.1093/gji/ggv259

Zaliapin, I., & Ben-Zion, Y. (2016). Discriminating Characteristics of Tectonic and Human-Induced Seismicity. Bulletin of the Seismological Society of America, 106(3), 846–859. https://doi.org/10.1785/0120150211

Zaliapin, I., Gabrielov, A., Keilis-Borok, V., & Wong, H. (2008). Clustering analysis of seismicity and aftershock identification. Physical Review Letters, 101(1), 018501. https://doi.org/10.1103/PhysRevLett.101.018501

Zhang, H., Eaton, D. W., Rodriguez, G., & Jia, S. Q. (2019). Source-Mechanism Analysis and Stress Inversion for Hydraulic-Fracturing-Induced Event Sequences near Fox Creek, Alberta. Bulletin of the Seismological Society of America, 109(2), 636–651. https://doi.org/10.1785/0120180275

Zhuang, J., Ogata, Y., & Vere-Jones, D. (2004). Analyzing earthquake clustering features by using stochastic reconstruction. Journal of Geophysical Research - Solid Earth, 109(B5), B05301-n/a. https://doi.org/10.1029/2003JB002879

Journal of Geophysical Research: Solid Earth

Supporting Information for

Statistical Modeling and Characterization of Induced Seismicity within the Western Canada Sedimentary Basin

Sidhanth Kothari[1], Robert Shcherbakov[1], and Gail Atkinson[1]

[1]Western University, London, Ontario, Canada

Contents of this file

Figures S1 to S17

Introduction

This supporting information provides additional context to both the regional (section 4) and the cluster (section 5) analyses in the main paper. In particular, Figures S1-4 show the event maps for each cluster discussed in section 2 and analyzed in section 5. Figure 5 shows the estimated completeness level of the Composite Alberta Seismicity Catalogue between 2010-2018. Figures S6-9 present the additional tests detailed in section 4, namely the nearest-neighbour distance (NND) analyses for randomized catalogues, sensitivity tests for variations in the Gutenberg-Richter *b*-value and fractal dimension (d_f), and the likelihood-ratio test to assess the validity of a three-component Gaussian mixture model. Figures S10-13 give the regional NND analyses results at higher cutoff magnitudes, which confirm the existence of three earthquake subpopulations and a transformation in earthquake distribution between 1975-2009 and 2010-2018 (analogous to Figures 2 and 5). Finally, Figures S14-17 give the structural representations and statistical parameters of the event families identified in section 5.3 (analogous to Figure 8).

1975 - 2000

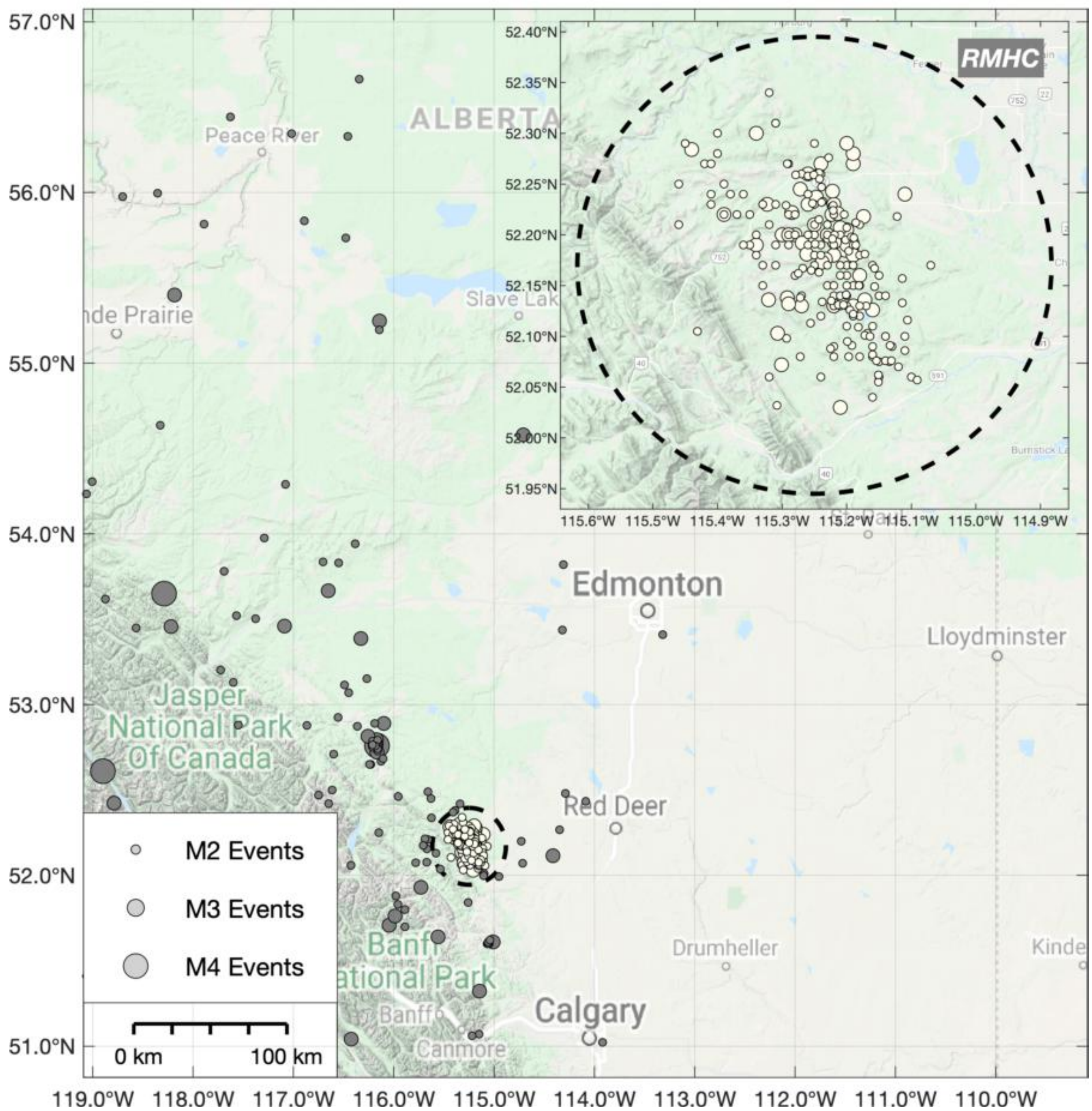


Figure S1. Map of the Rocky Mountain House cluster (RMHC) study area between 1975-2000. Dashed circle represents a 20 km radius from the coordinates [115.24°W, 52.17°N]. Markers are seismic events. White markers are the data points used for analysis.

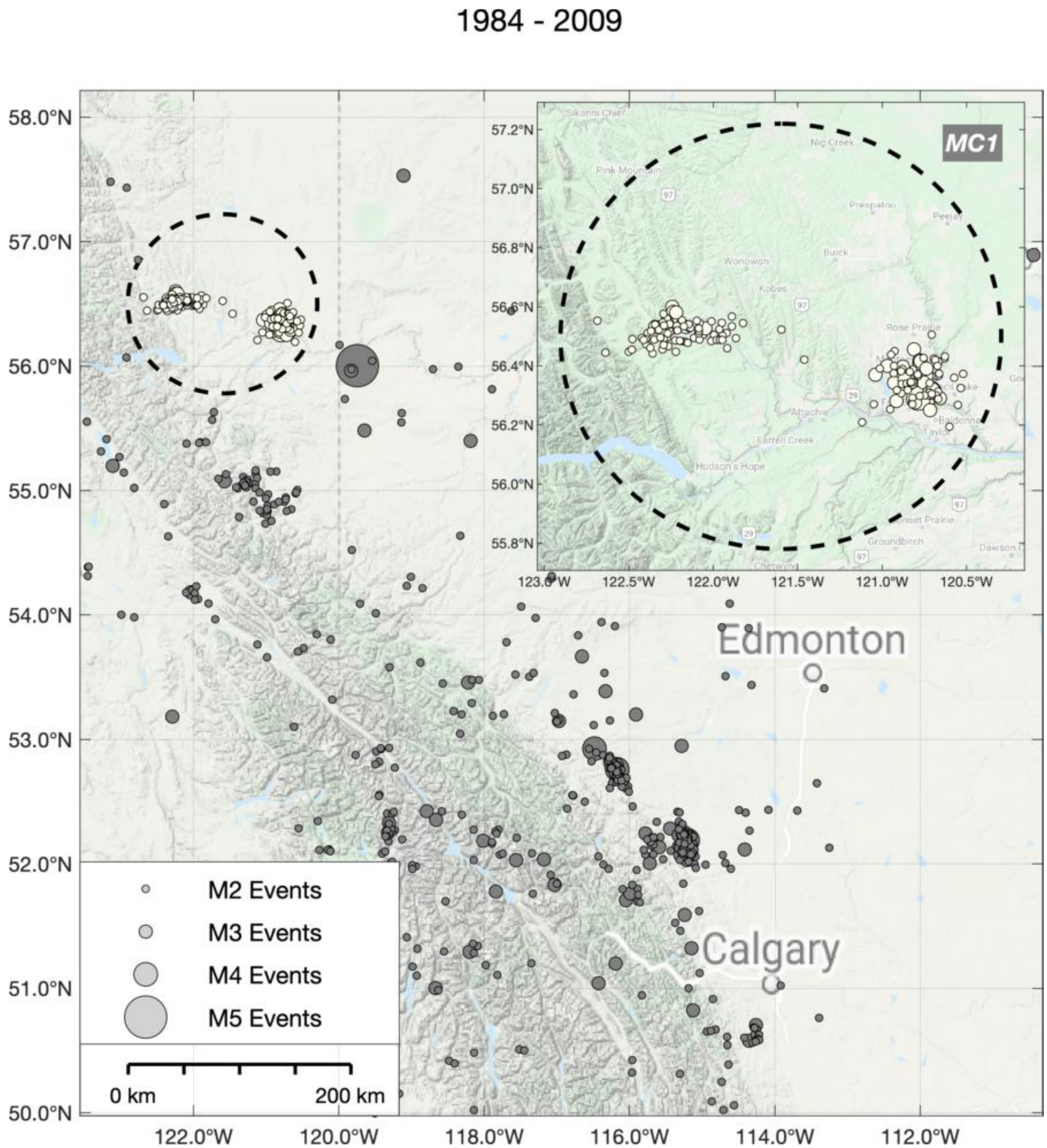


Figure S2. Map of the Montney cluster 1 (MC1) study area between 1984-2009. Dashed circle represents a 75 km radius from the coordinates [121.6°W, 56.5°N]. Markers are seismic events. White markers are the data points used for analysis.

2010 - 2018

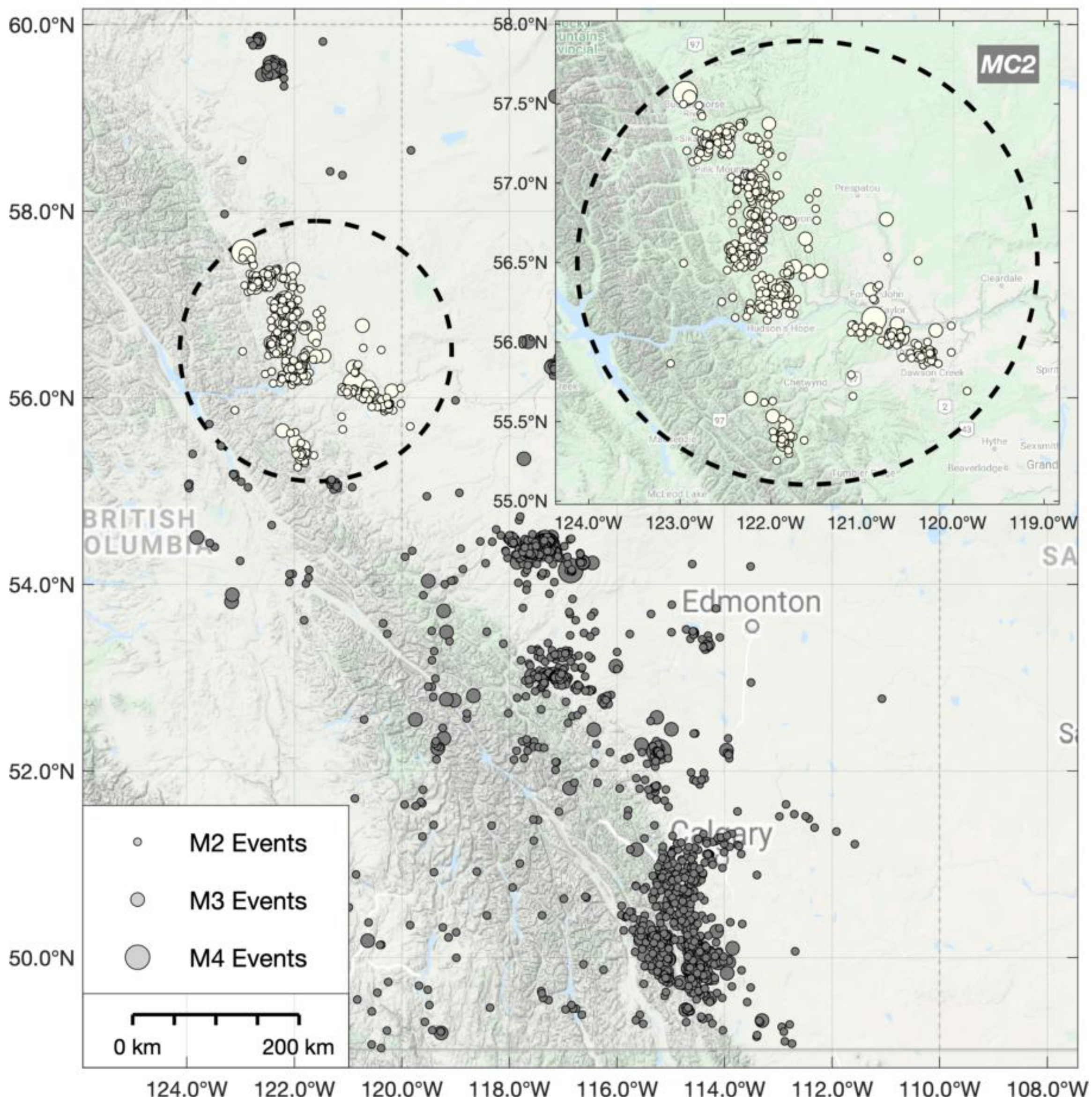


Figure S3. Map of the Montney cluster 2 (MC2) study area between 2010 and 2018. Dashed circle represents a 150 km radius from the coordinates [121.6°W, 56.5°N]. Markers are seismic events. White markers are the data points used for analysis.

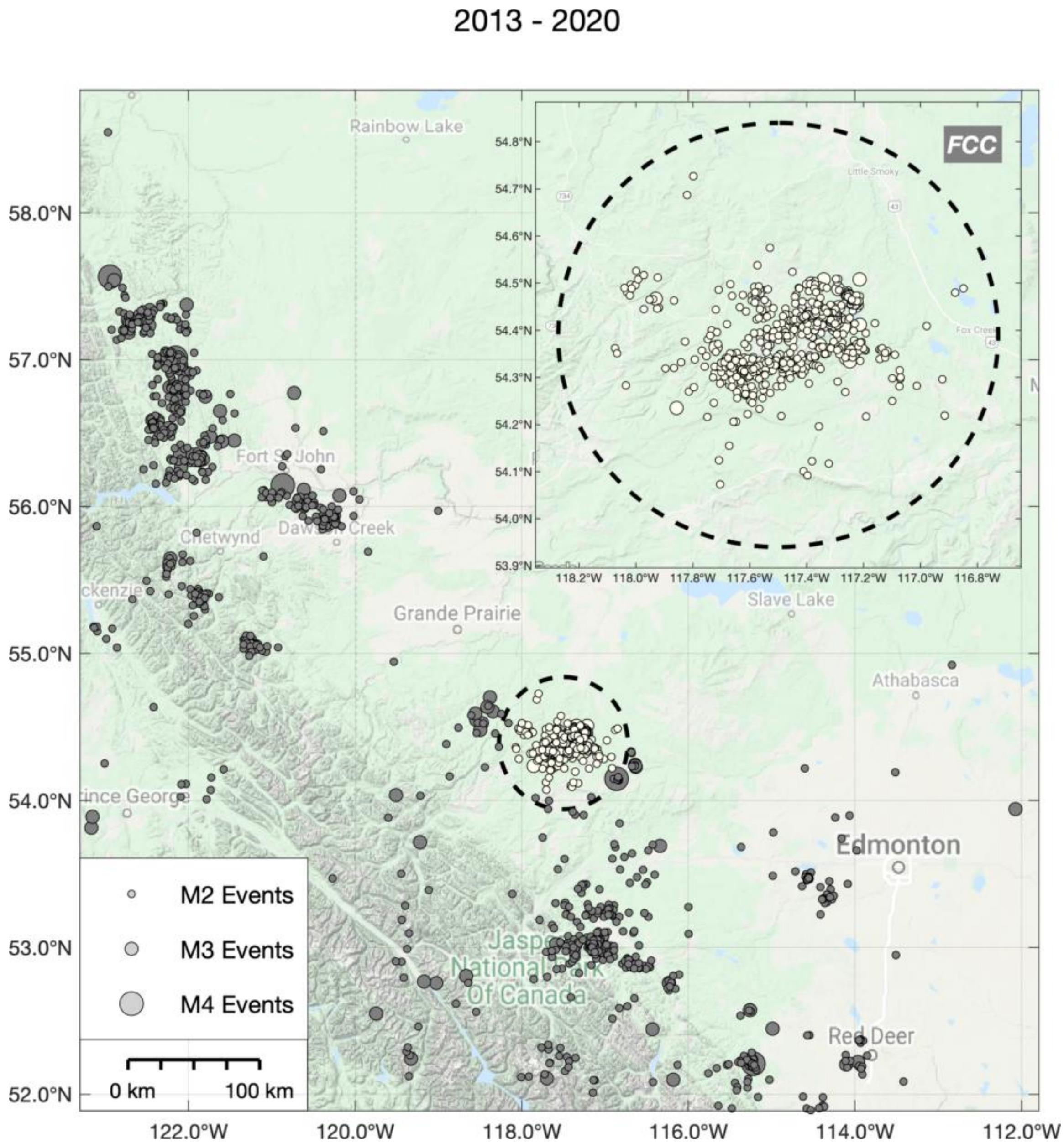


Figure S4. Map of the Fox Creek cluster (FCC) study area between 2013 and January 2020. Dashed circle represents a 45 km radius from the coordinates [117.4°W, 54.4°N]. Markers are seismic events. White markers are the data points used for analysis.

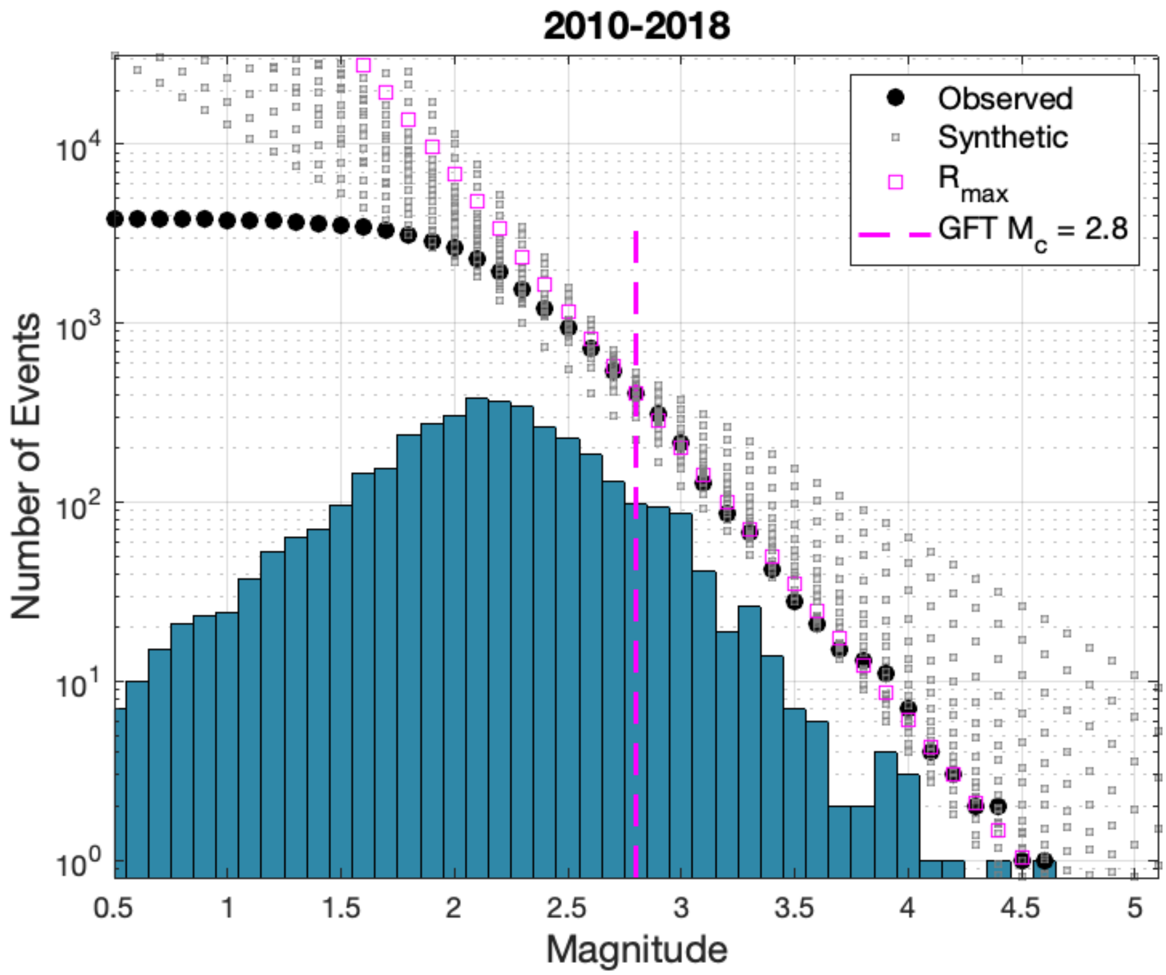


Figure S5. Estimated completeness of the Composite Alberta Seismicity Catalogue, between 2010-2018, using the goodness-of-fit test (GFT) of Wiemer & Wyss (2000). Squares represent synthetic distributions drawn from the (exponential) Gutenberg-Richter relation, for a range of *b*-values. Pink squares represent the synthetic distribution that maximizes the goodness-of-fit (*R*), i.e. that minimizes the residual between it and the cumulative frequency-magnitude distribution. Grey squares represent the rejected synthetic distributions. Vertical dashed pink line indicates the corresponding completeness threshold.

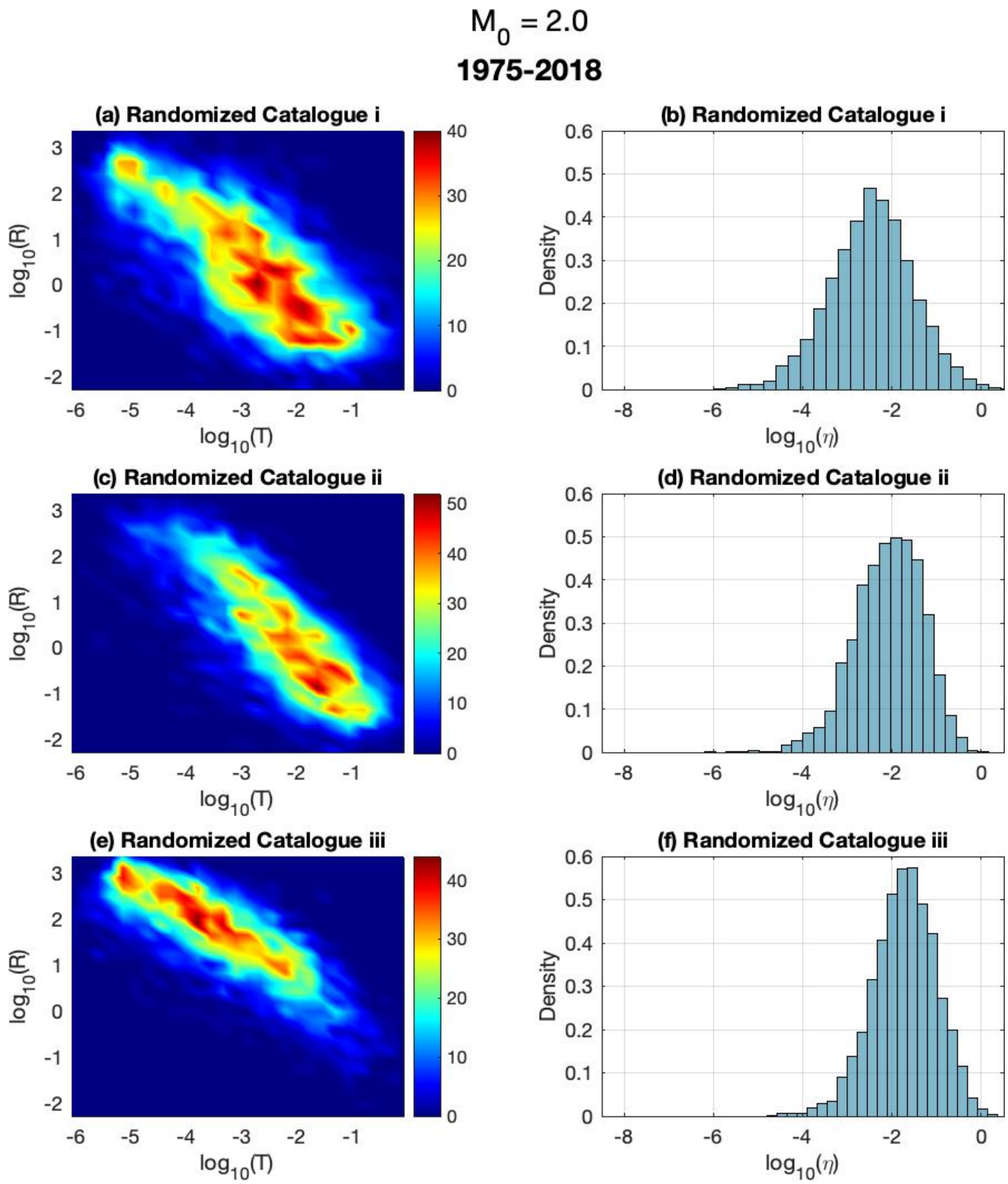


Figure S6. Nearest-neighbour distance distributions of three randomized versions of the Composite Alberta Seismicity Catalogue (CASC). (a, b) Event times and event locations are shuffled. (c, d) Event times are uniformly distributed within the temporal limits of the CASC and event locations are shuffled. (e, f) Original event times are kept and event locations are uniformly distributed within the latitudinal and longitudinal limits of the CASC.

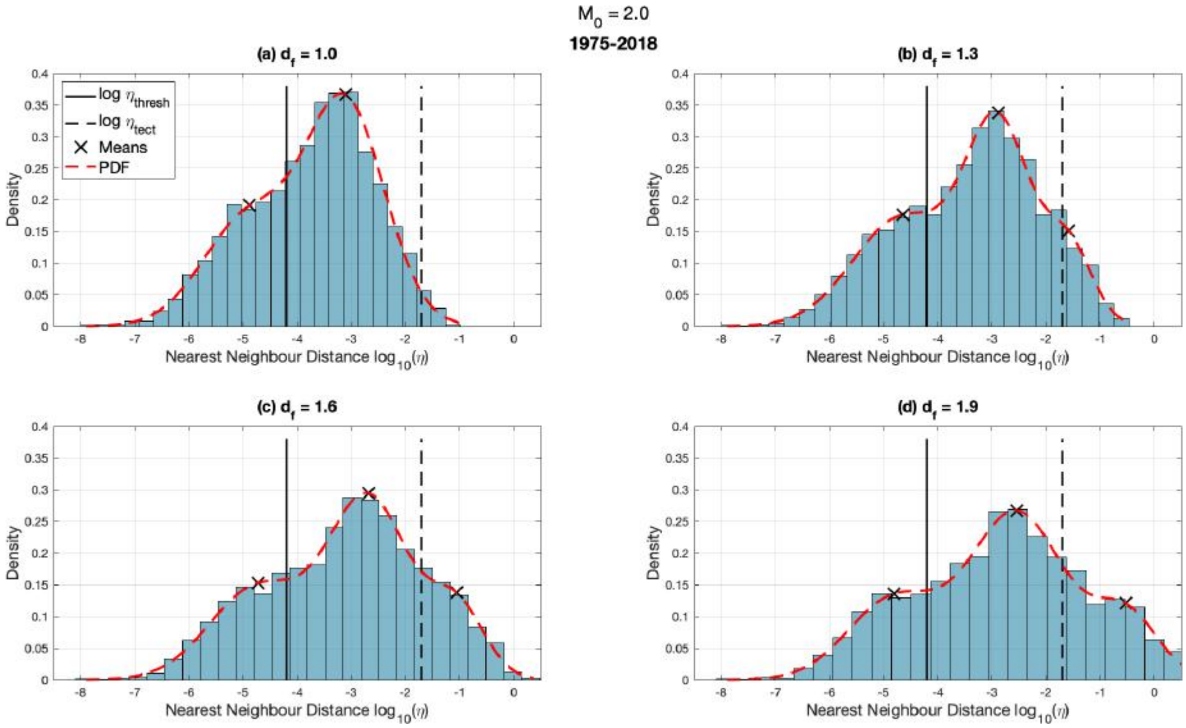


Figure S7. One-dimensional η distributions, analogous to Figure 2b, using a range of fractal dimension (d_f) values. As d_f increases, the distribution spreads and trimodality becomes more apparent.

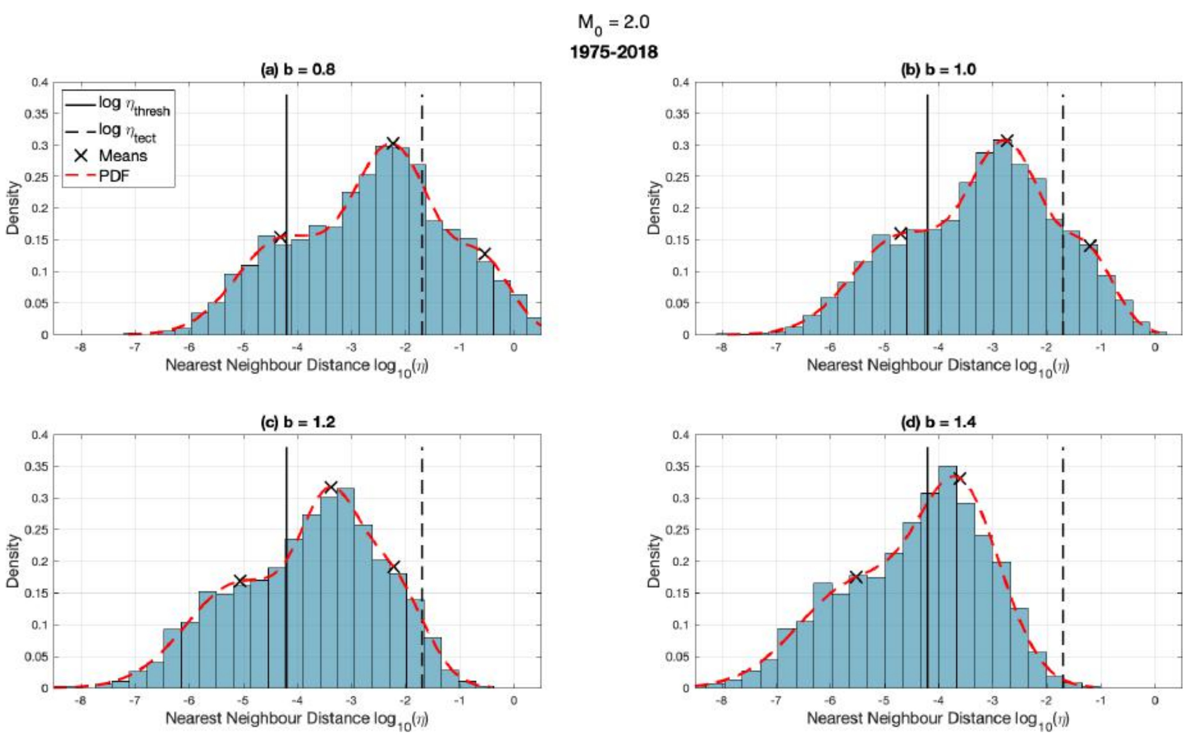


Figure S8. One-dimensional η distributions, analogous to Figure 2b, using a range of Gutenberg-Richter *b*-values. As *b* increases, the distribution gets squeezed and shifts leftward, obscuring the distinctive background mode at large η.

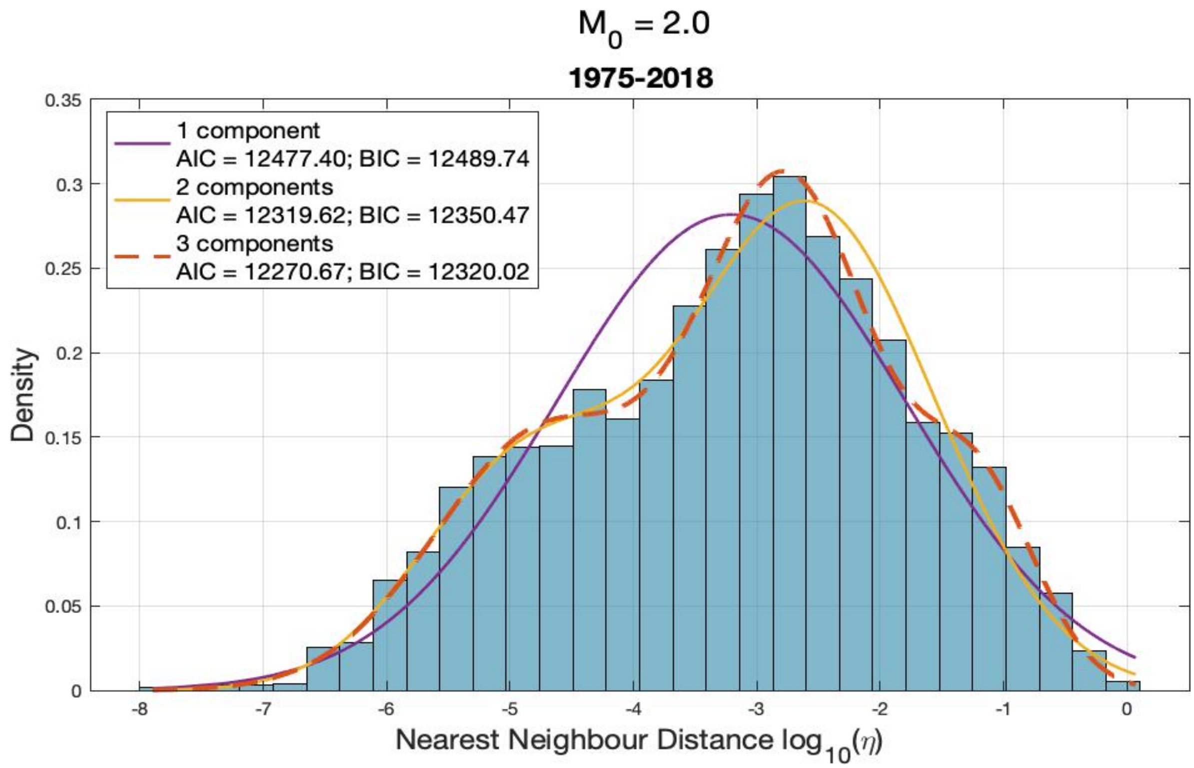


Figure S9. Gaussian mixture models, with 1, 2, and 3 components, for the regional distribution of inter-event distances η. AIC and BIC are their respective Akaike and Bayesian information criteria. The lowest AIC and BIC values correspond to the best fitting model; here they are minimized for a three-component mixture.

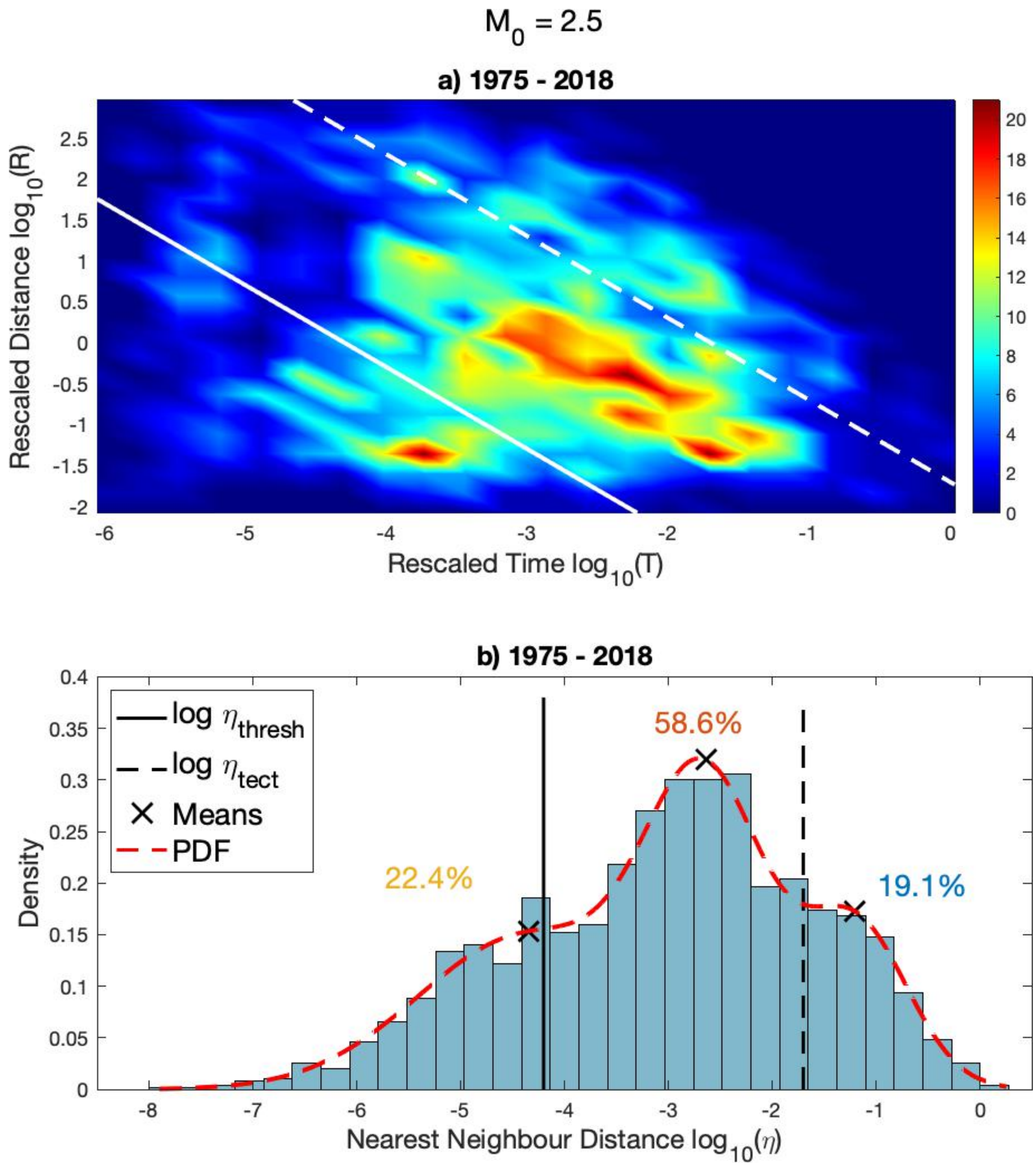


Figure S10. Nearest-neighbor distance distribution of the regional CASC dataset from 1975-2018, using a cutoff magnitude of M_0 = 2.5 (analogous to Figure 2). Percentages in subplot (b) reflect the modal mixing proportions. Trimodality remains distinguishable at M_0 = 2.5.

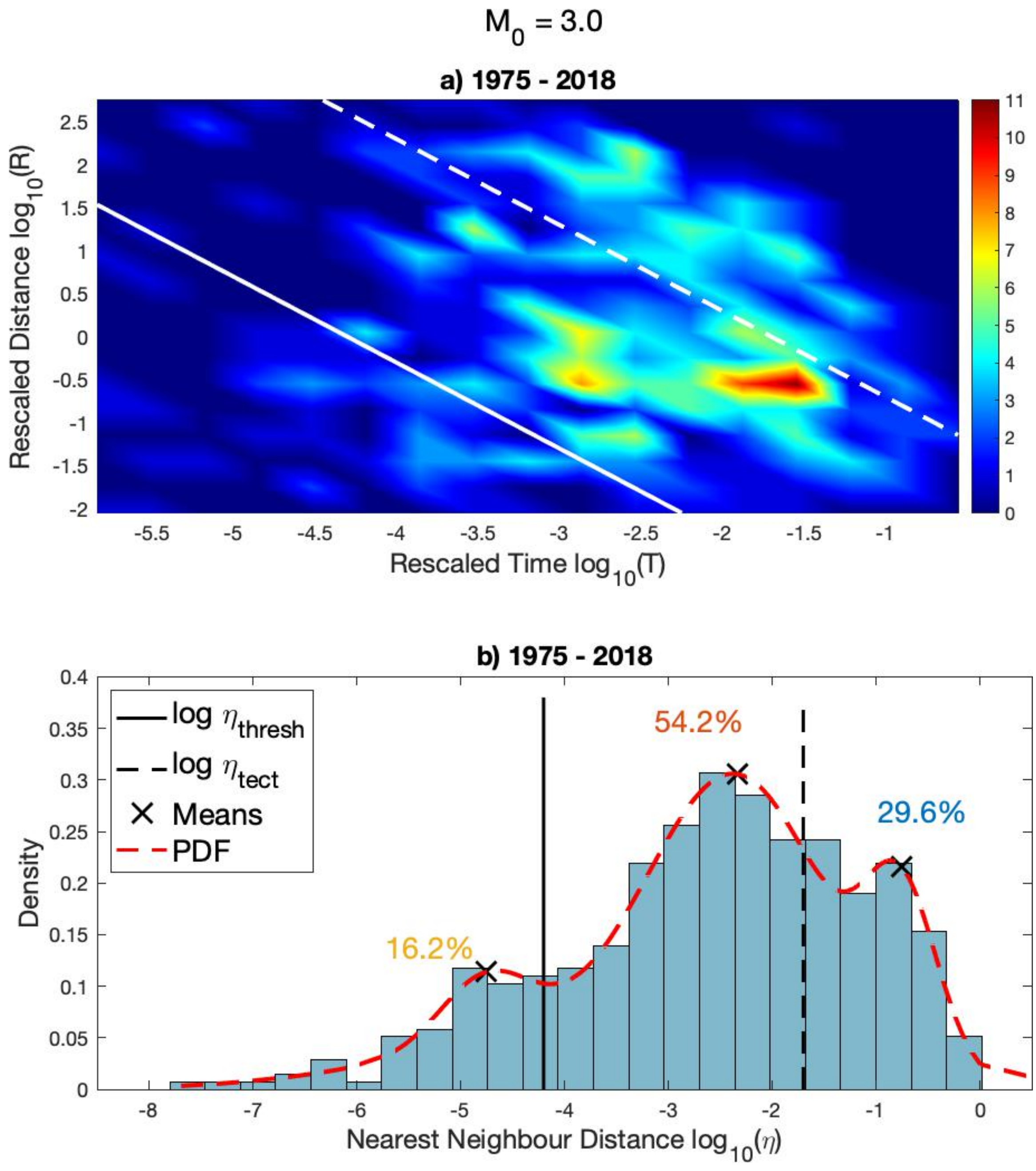


Figure S11. Nearest-neighbor distance distribution of the regional WCSB catalogue from 1975-2018, using a cutoff magnitude of M_0 = 3.0 (analogous to Figure 2). Percentages in subplot (b) reflect the modal mixing proportions. Trimodality remains distinguishable at M_0 = 3.0.

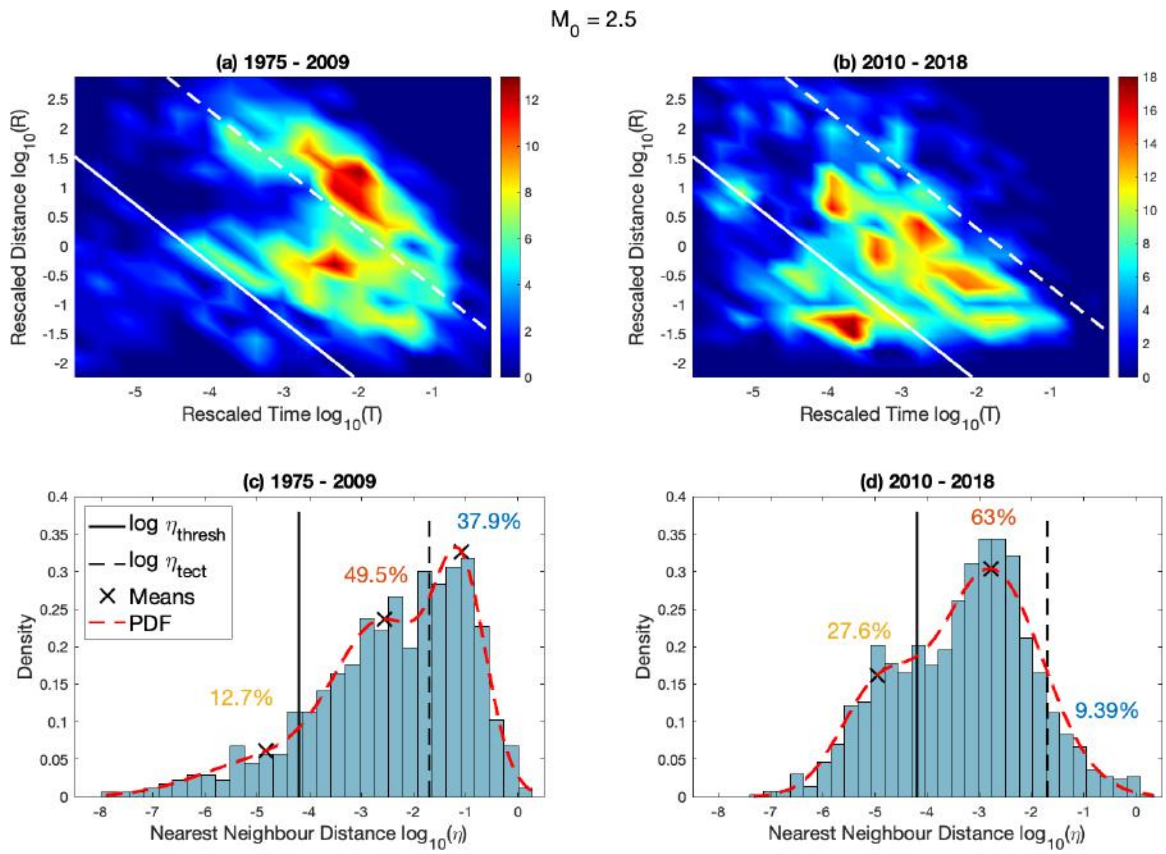


Figure S12. Comparison of nearest-neighbor distance distributions of the regional WCSB catalogue across time using a cutoff magnitude of M_0 = 2.5 (analogous to Figure 5). (a, c) 1975-2009. (b, d) 2010-2018. Percentages in subplots (c, d) reflect the modal mixing proportions.

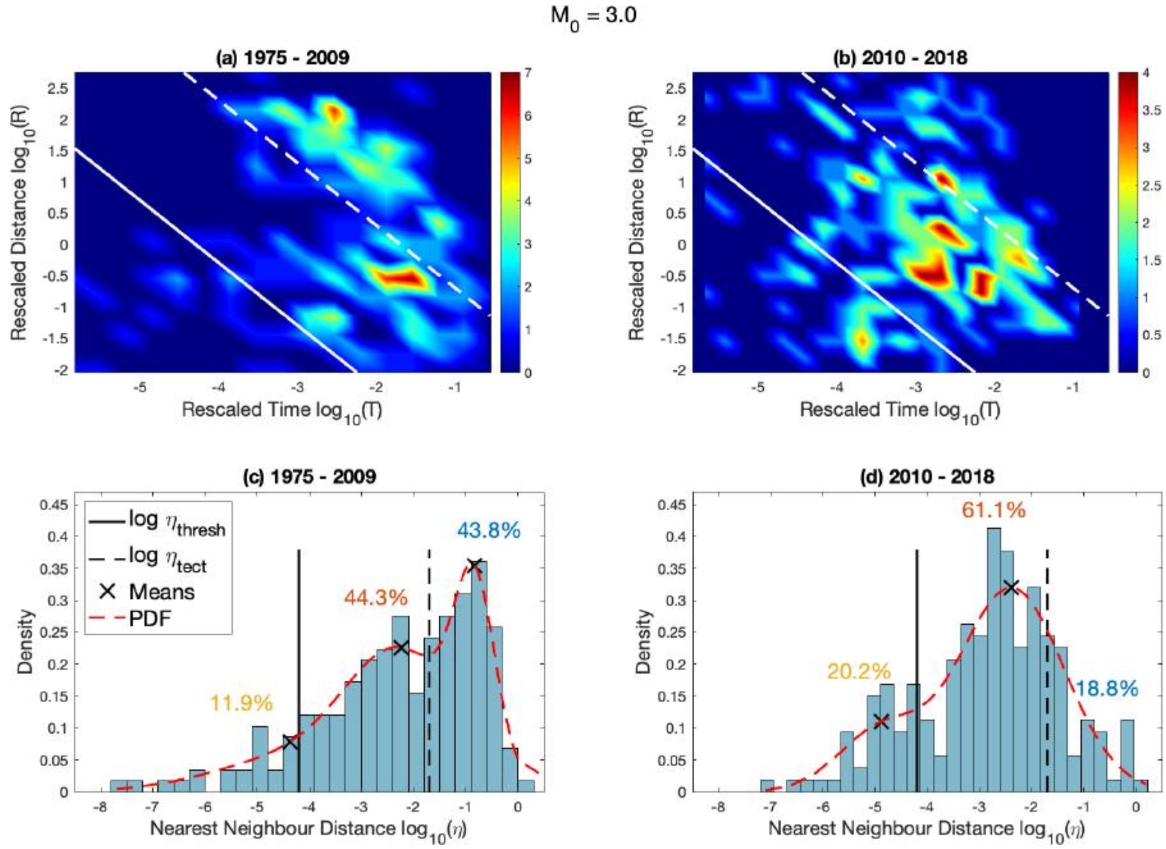


Figure S13. Comparison of nearest-neighbor distance distributions of the regional WCSB catalogue across time using a cutoff magnitude of M_0 = 3.0 (analogous to Figure 5) (a, c) 1975-2009. (b, d) 2010-2018. Percentages in subplots (c, d) reflect the modal mixing proportions.

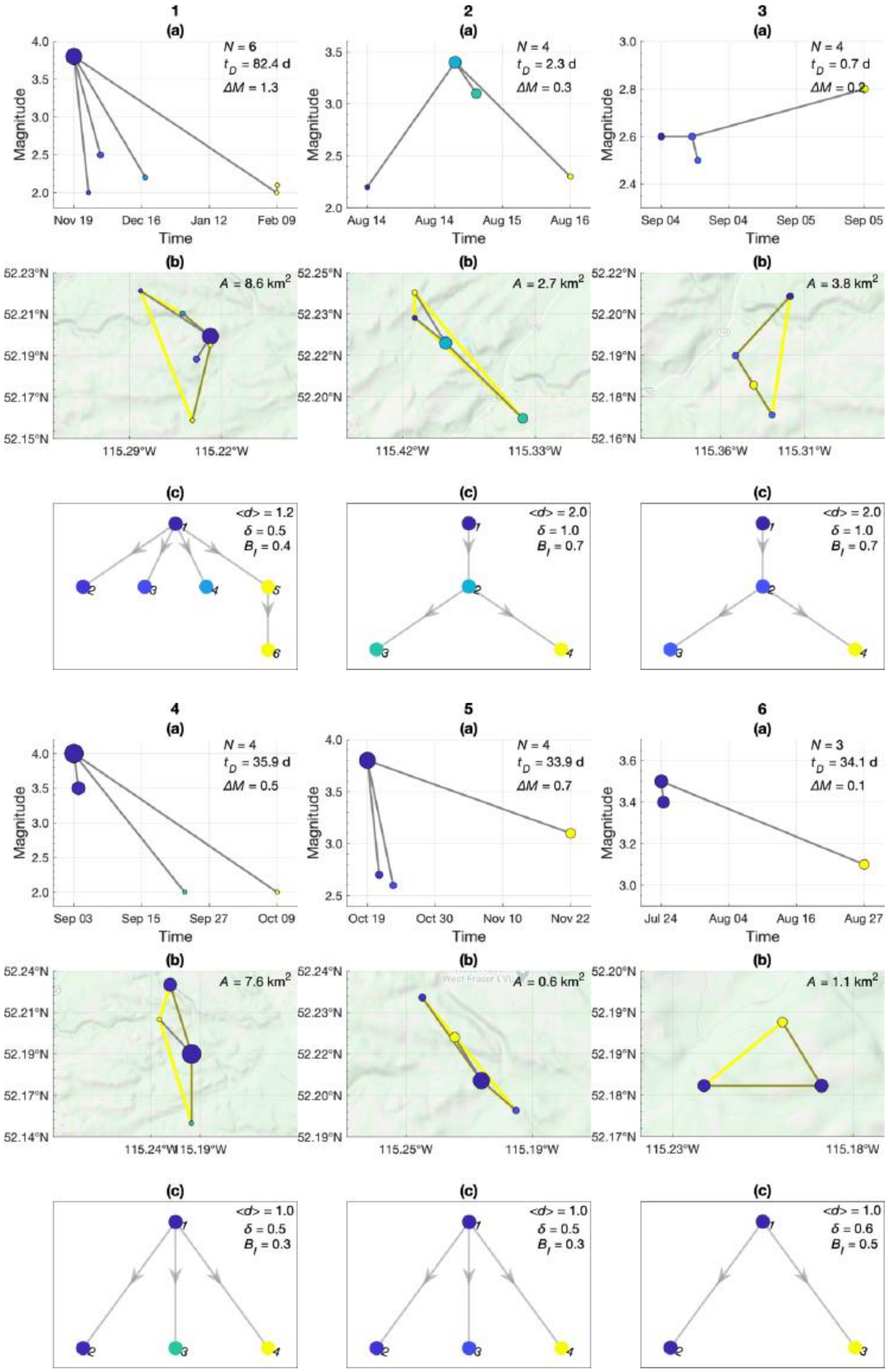


Figure S14. Event family structures in the Rocky Mountain House cluster (RMHC). a) Moment magnitude vs. time in days. b) Spatial map; yellow border outlines the hull area occupied by the sequence. c) Directed tree graph in dimensionless space. Data points are coloured chronologically from darkest to lightest.

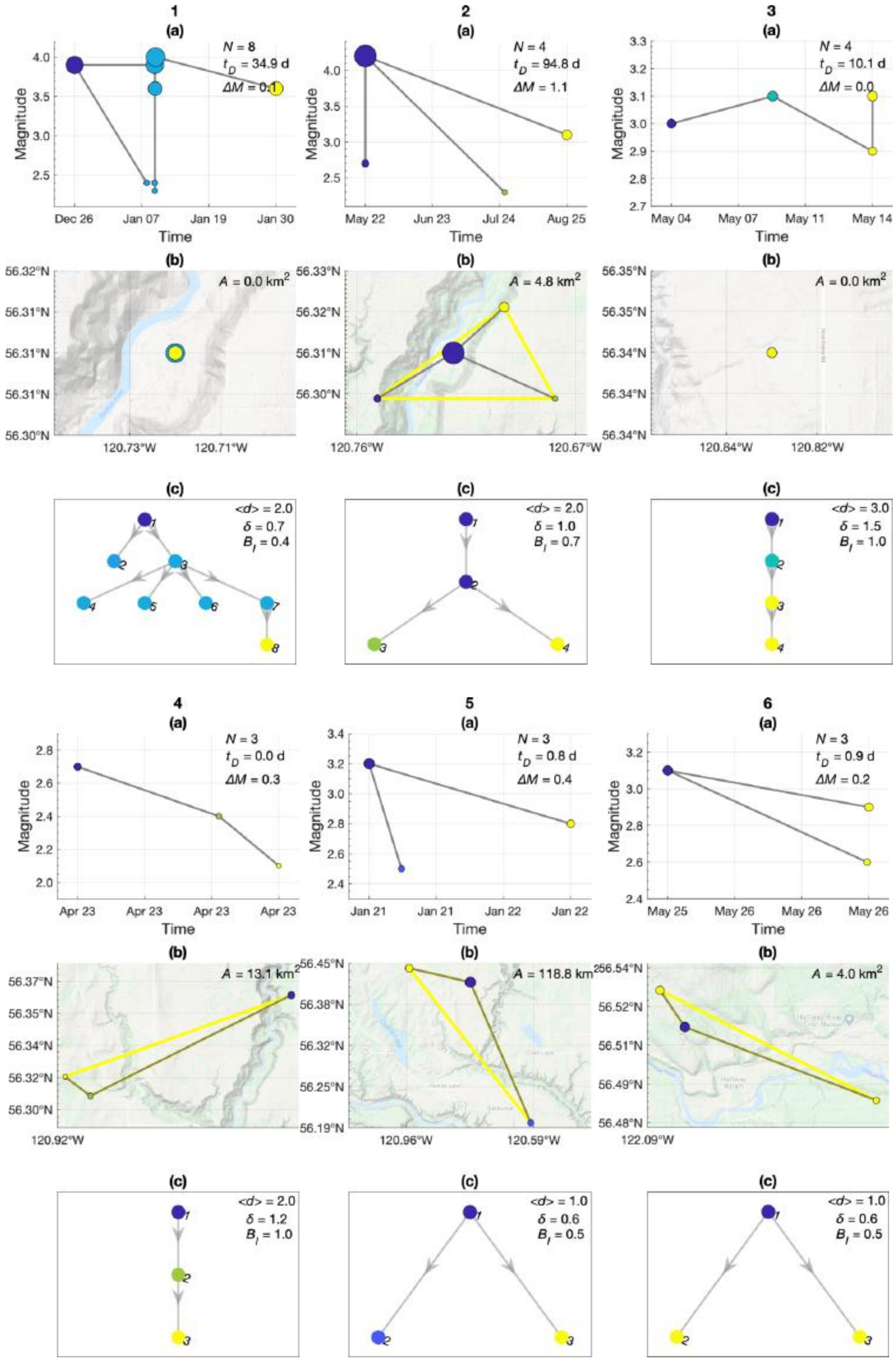
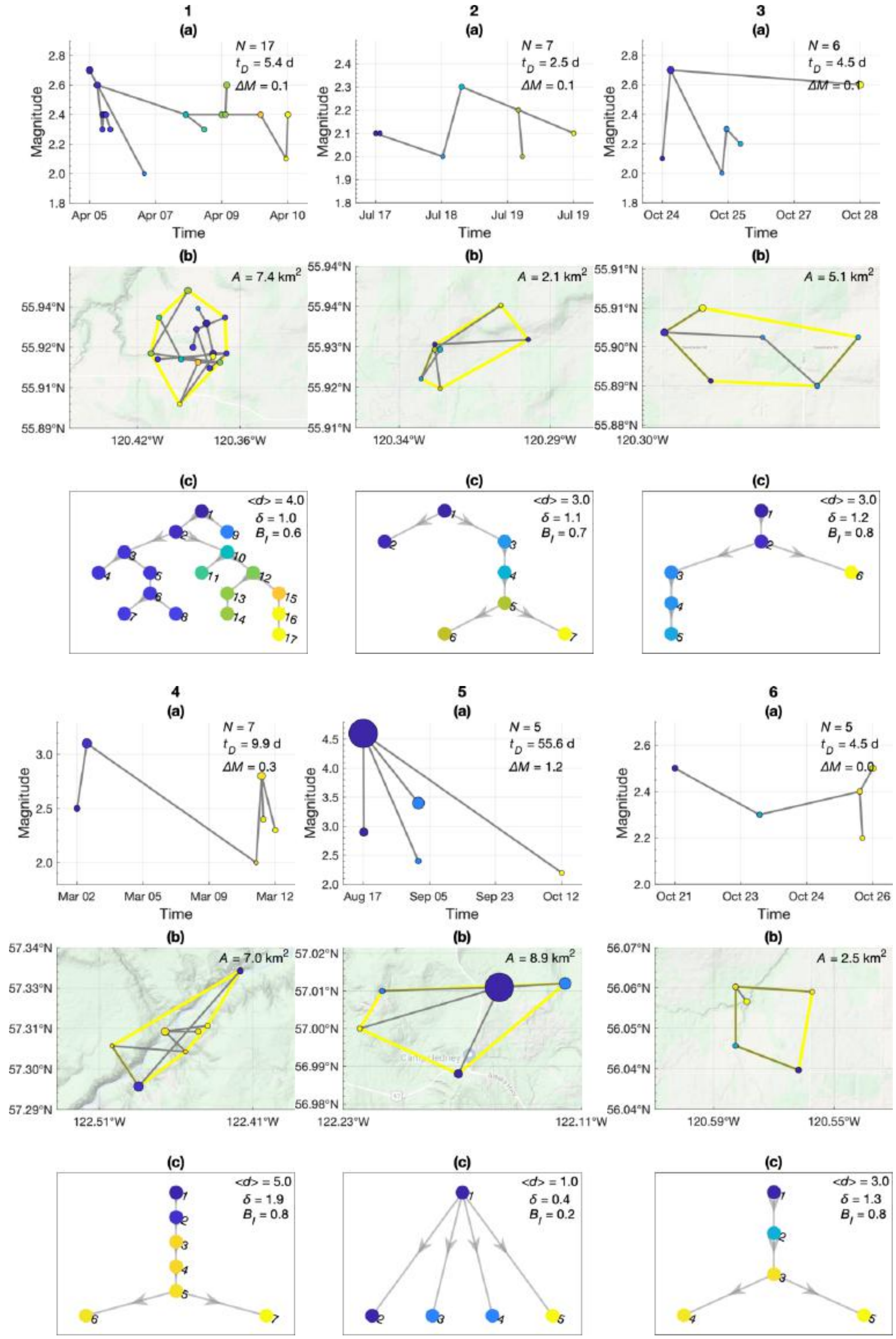


Figure S15. Event family structures in the Montney cluster 1 (MC1). a) Moment magnitude vs. time in days. b) Spatial map; yellow border outlines the hull area occupied by the sequence. c) Directed tree graph in dimensionless space. Data points are coloured chronologically from darkest to lightest.


1
(a)
N = 17
t_D = 5.4 d
ΔM = 0.1
Magnitude
Time
Apr 05
Apr 07
Apr 09
Apr 10
(b)
A = 7.4 km²
55.94°N
55.92°N
55.91°N
55.89°N
120.42°W
120.36°W
(c)
<d> = 4.0
δ = 1.0
B_l = 0.6
2
(a)
N = 7
t_D = 2.5 d
ΔM = 0.1
Magnitude
Time
Jul 17
Jul 18
Jul 19
Jul 19
(b)
A = 2.1 km²
55.94°N
55.93°N
55.92°N
55.91°N
120.34°W
120.29°W
(c)
<d> = 3.0
δ = 1.1
B_l = 0.7
3
(a)
N = 6
t_D = 4.5 d
ΔM = 0.1
Magnitude
Time
Oct 24
Oct 25
Oct 27
Oct 28
(b)
A = 5.1 km²
55.91°N
55.90°N
55.89°N
55.88°N
120.30°W
(c)
<d> = 3.0
δ = 1.2
B_l = 0.8
4
(a)
N = 7
t_D = 9.9 d
ΔM = 0.3
Magnitude
Time
Mar 02
Mar 05
Mar 09
Mar 12
(b)
A = 7.0 km²
57.34°N
57.33°N
57.31°N
57.30°N
57.29°N
122.51°W
122.41°W
(c)
<d> = 5.0
δ = 1.9
B_l = 0.8
5
(a)
N = 5
t_D = 55.6 d
ΔM = 1.2
Magnitude
Time
Aug 17
Sep 05
Sep 23
Oct 12
(b)
A = 8.9 km²
57.02°N
57.01°N
57.00°N
56.99°N
56.98°N
122.23°W
122.11°W
(c)
<d> = 1.0
δ = 0.4
B_l = 0.2
6
(a)
N = 5
t_D = 4.5 d
ΔM = 0.0
Magnitude
Time
Oct 21
Oct 23
Oct 24
Oct 26
(b)
A = 2.5 km²
56.07°N
56.06°N
56.05°N
56.04°N
56.04°N
120.59°W
120.55°W
(c)
<d> = 3.0
δ = 1.3
B_l = 0.8

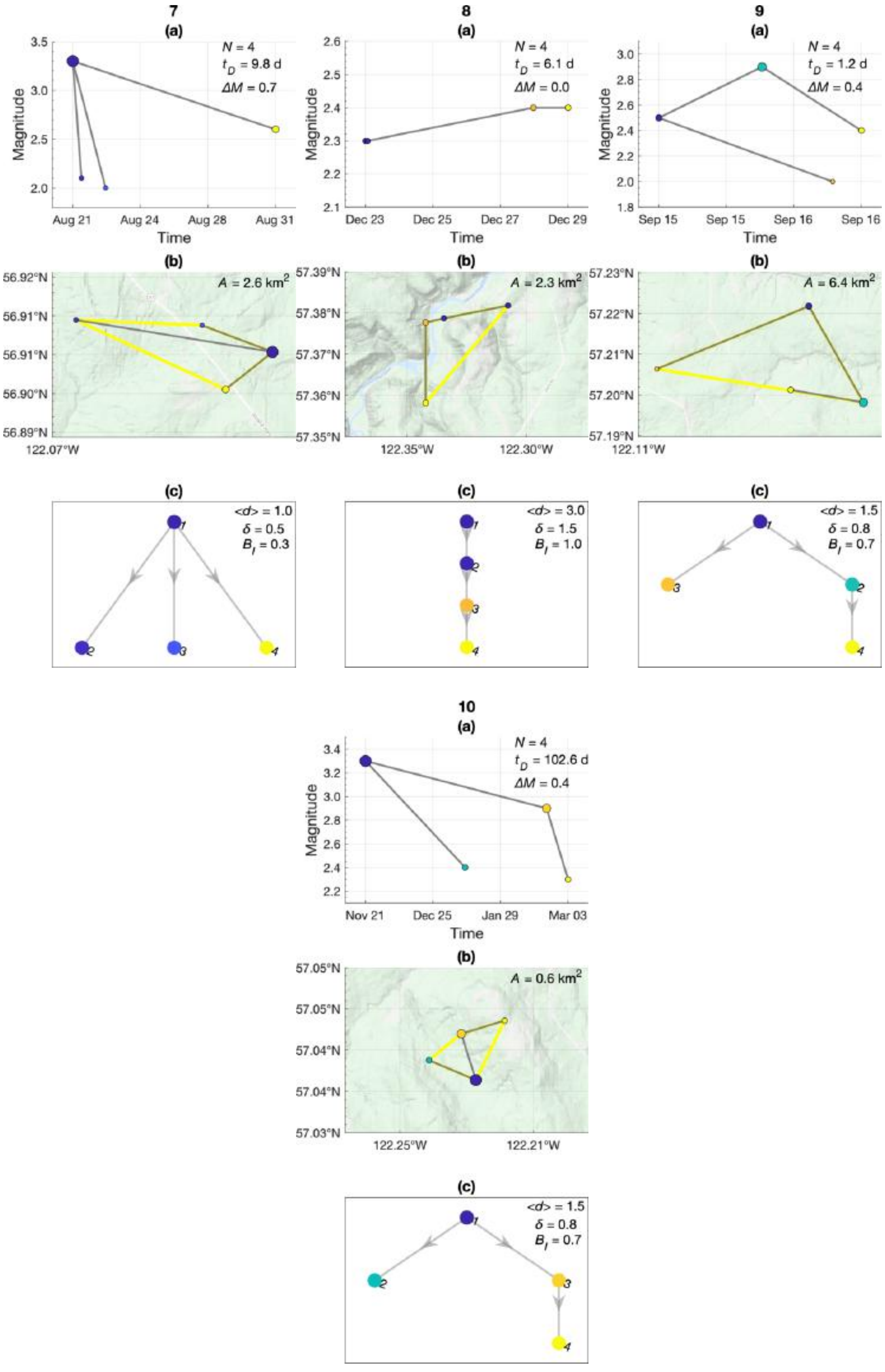
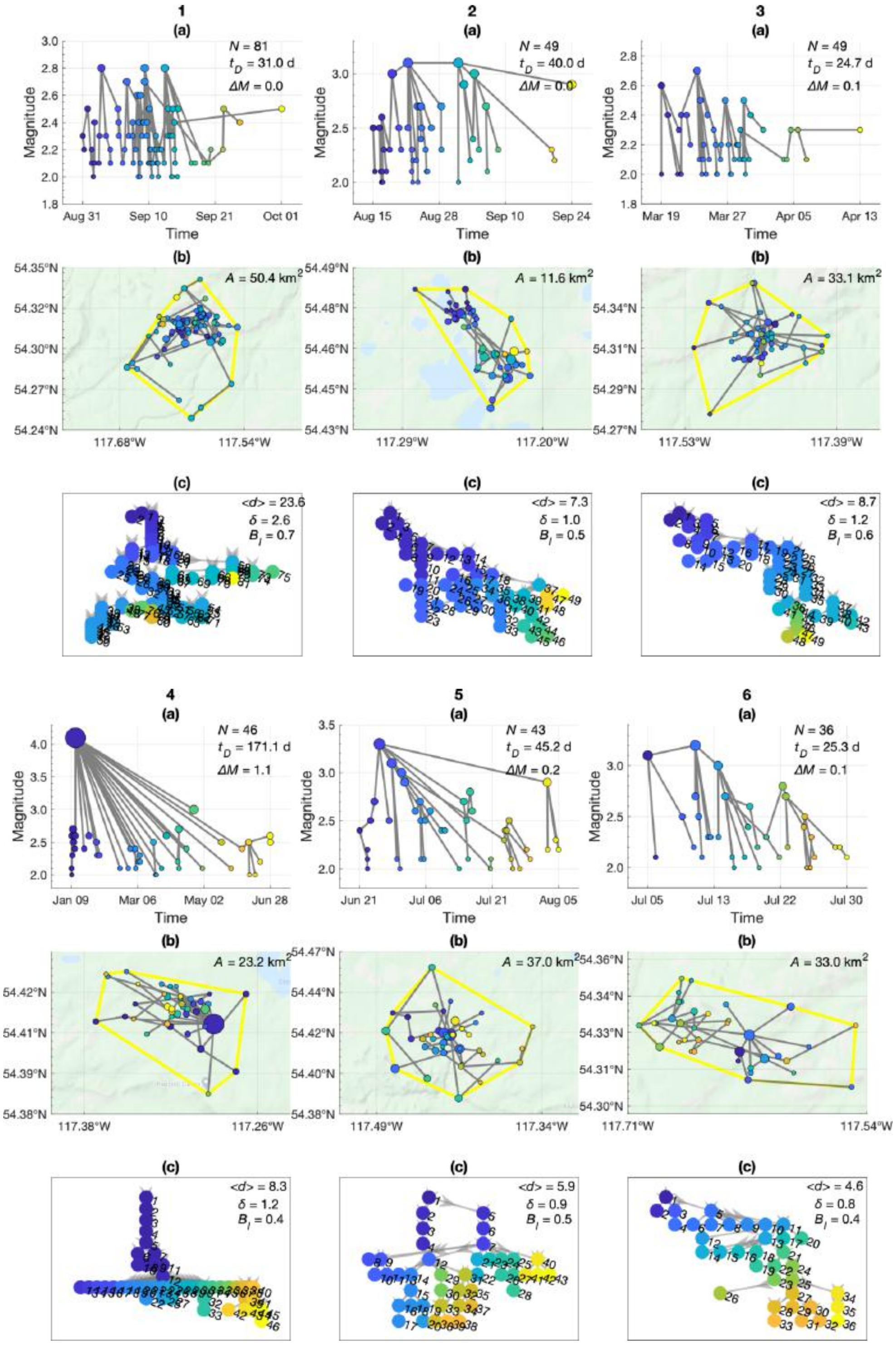
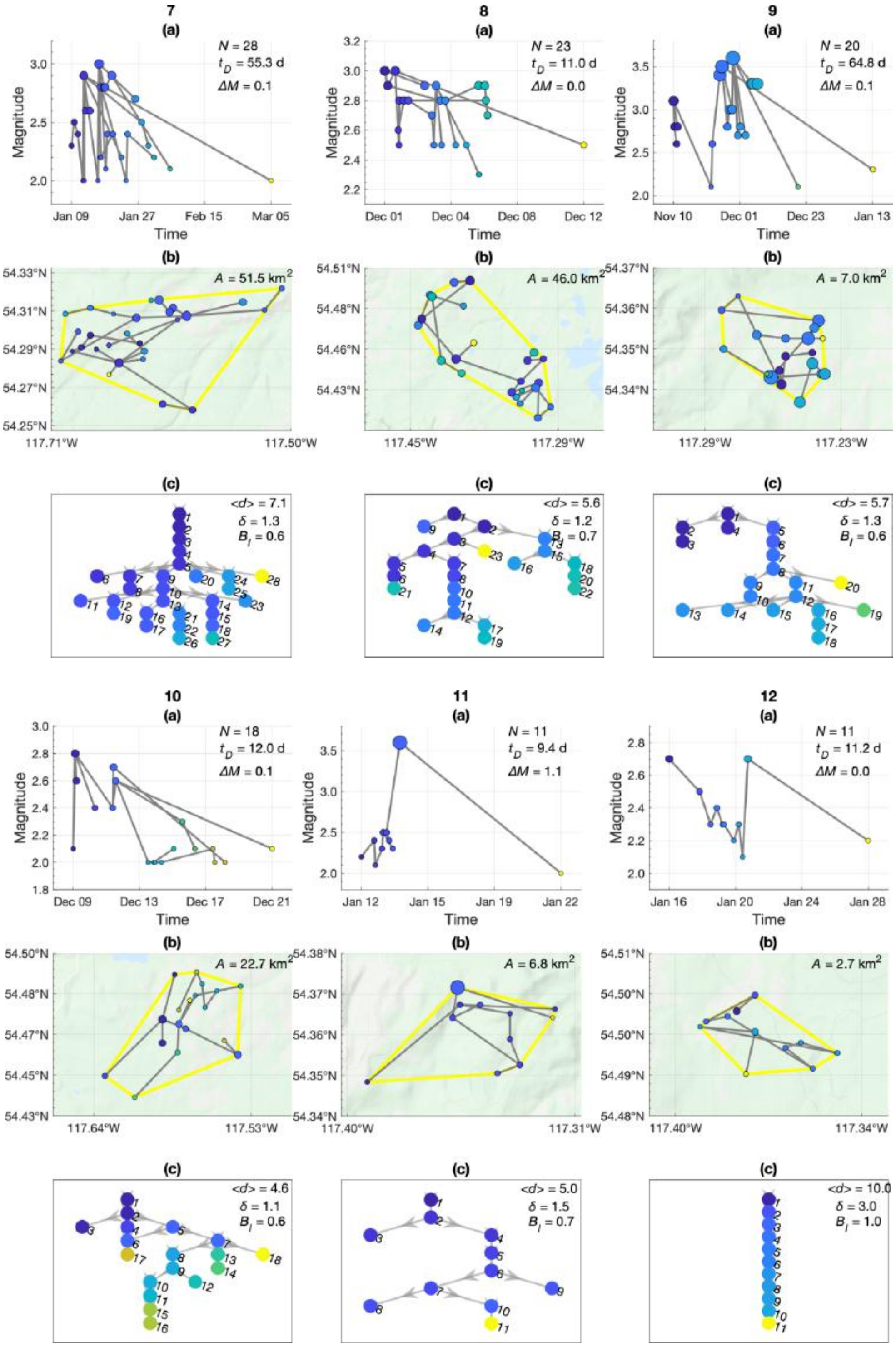


Figure S16. Event family structures in the Montney cluster 2 (MC2). a) Moment magnitude vs. time in days. b) Spatial map; yellow border outlines the hull area occupied by the sequence. c) Directed tree graph in dimensionless space. Data points are coloured chronologically from darkest to lightest.


1
(a)
N = 81
t_D = 31.0 d
ΔM = 0.0
Magnitude
Time
Aug 31
Sep 10
Sep 21
Oct 01
(b)
A = 50.4 km²
54.35°N
54.32°N
54.30°N
54.27°N
54.24°N
117.68°W
117.54°W
(c)
<d> = 23.6
δ = 2.6
B_l = 0.7
2
(a)
N = 49
t_D = 40.0 d
ΔM = 0.0
Magnitude
Time
Aug 15
Aug 28
Sep 10
Sep 24
(b)
A = 11.6 km²
54.49°N
54.48°N
54.46°N
54.45°N
54.43°N
117.29°W
117.20°W
(c)
<d> = 7.3
δ = 1.0
B_l = 0.5
3
(a)
N = 49
t_D = 24.7 d
ΔM = 0.1
Magnitude
Time
Mar 19
Mar 27
Apr 05
Apr 13
(b)
A = 33.1 km²
54.34°N
54.31°N
54.29°N
54.27°N
117.53°W
117.39°W
(c)
<d> = 8.7
δ = 1.2
B_l = 0.6
4
(a)
N = 46
t_D = 171.1 d
ΔM = 1.1
Magnitude
Time
Jan 09
Mar 06
May 02
Jun 28
(b)
A = 23.2 km²
54.42°N
54.41°N
54.39°N
54.38°N
117.38°W
117.26°W
(c)
<d> = 8.3
δ = 1.2
B_l = 0.4
5
(a)
N = 43
t_D = 45.2 d
ΔM = 0.2
Magnitude
Time
Jun 21
Jul 06
Jul 21
Aug 05
(b)
A = 37.0 km²
54.47°N
54.44°N
54.42°N
54.40°N
54.38°N
117.49°W
117.34°W
(c)
<d> = 5.9
δ = 0.9
B_l = 0.5
6
(a)
N = 36
t_D = 25.3 d
ΔM = 0.1
Magnitude
Time
Jul 05
Jul 13
Jul 22
Jul 30
(b)
A = 33.0 km²
54.36°N
54.34°N
54.33°N
54.31°N
54.30°N
117.71°W
117.54°W
(c)
<d> = 4.6
δ = 0.8
B_l = 0.4


7
(a)
N = 28
t_D = 55.3 d
ΔM = 0.1
Magnitude
Time
Jan 09
Jan 27
Feb 15
Mar 05
(b)
A = 51.5 km²
117.71°W
117.50°W
(c)
<d> = 7.1
δ = 1.3
B_l = 0.6
8
(a)
N = 23
t_D = 11.0 d
ΔM = 0.0
Magnitude
Time
Dec 01
Dec 04
Dec 08
Dec 12
(b)
A = 46.0 km²
117.45°W
117.29°W
(c)
<d> = 5.6
δ = 1.2
B_l = 0.7
9
(a)
N = 20
t_D = 64.8 d
ΔM = 0.1
Magnitude
Time
Nov 10
Dec 01
Dec 23
Jan 13
(b)
A = 7.0 km²
117.29°W
117.23°W
(c)
<d> = 5.7
δ = 1.3
B_l = 0.6
10
(a)
N = 18
t_D = 12.0 d
ΔM = 0.1
Magnitude
Time
Dec 09
Dec 13
Dec 17
Dec 21
(b)
A = 22.7 km²
117.64°W
117.53°W
(c)
<d> = 4.6
δ = 1.1
B_l = 0.6
11
(a)
N = 11
t_D = 9.4 d
ΔM = 1.1
Magnitude
Time
Jan 12
Jan 15
Jan 19
Jan 22
(b)
A = 6.8 km²
117.40°W
117.31°W
(c)
<d> = 5.0
δ = 1.5
B_l = 0.7
12
(a)
N = 11
t_D = 11.2 d
ΔM = 0.0
Magnitude
Time
Jan 16
Jan 20
Jan 24
Jan 28
(b)
A = 2.7 km²
117.40°W
117.34°W
(c)
<d> = 10.0
δ = 3.0
B_l = 1.0

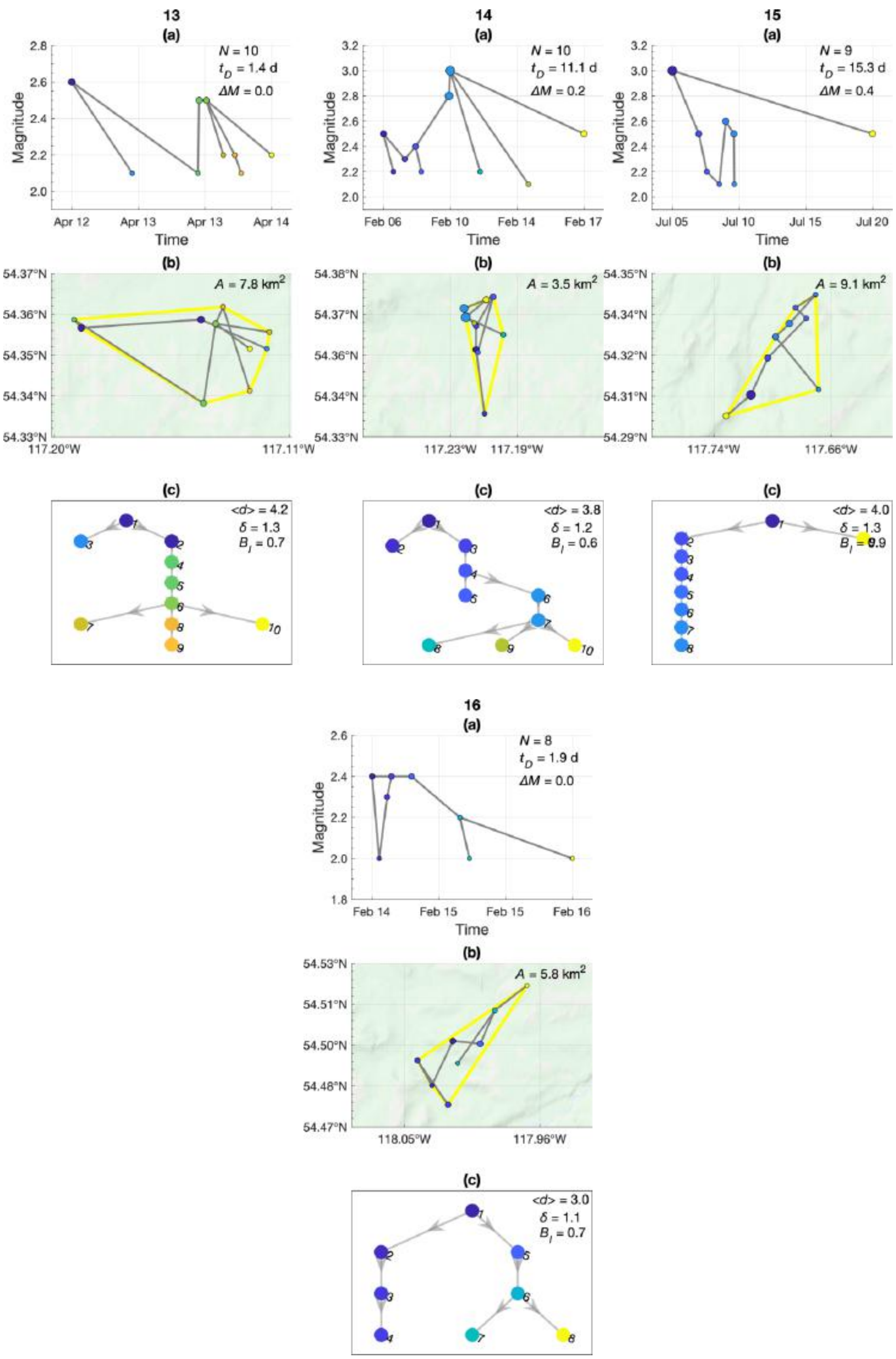


Figure S17. Event family structures in the Fox Creek cluster (FCC). a) Moment magnitude vs. time in days. b) Spatial map; yellow border outlines the hull area occupied by the sequence. c) Directed tree graph in dimensionless space. Data points are coloured chronologically from darkest to lightest.