Natural Resources Canada | Ressources naturelles Canada

GEOLOGICAL SURVEY OF CANADA
OPEN FILE 8831

A comprehensive earthquake catalogue for northeastern British Columbia: the northern Montney trend from 2017 to 2020 and the Kiskatinaw seismic monitoring and mitigation area from 2019 to 2020

R. Visser, H. Kao, R.M.H. Dokht, A.B. Mahani,
and S. Venables

2021

Canada

GEOLOGICAL SURVEY OF CANADA
OPEN FILE 8831

A comprehensive earthquake catalogue for northeastern British Columbia: the northern Montney trend from 2017 to 2020 and the Kiskatinaw seismic monitoring and mitigation area from 2019 to 2020

R. Visser[1,2], H. Kao[1,3], R.M.H. Dokht[1], A.B. Mahani[2], and S. Venables[4]

[1]Geological Survey of Canada, 9860 West Saanich Road, Sidney, British Columbia
[2]Geoscience BC, 750 West Pender Street, Vancouver, British Columbia
[3]School of Earth and Ocean Sciences, University of Victoria, 9882 Ring Road, Victoria, British Columbia
[4]BC Oil and Gas Commission, 2950 Jutland Road, Victoria, British Columbia

2021

Permanent link: https://doi.org/10.4095/329078

This publication is available for free download through GEOSCAN (https://geoscan.nrcan.gc.ca/).

Recommended citation
Visser, R., Kao, H., Dokht, R.M.H., Mahani, A.B., and Venables, S., 2021. A comprehensive earthquake catalogue for northeastern British Columbia: the northern Montney trend from 2017 to 2020 and the Kiskatinaw seismic monitoring and mitigation area from 2019 to 2020; Geological Survey of Canada, Open File 8831, 1 .zip file. https://doi.org/10.4095/329078

Publications in this series have not been edited; they are released as submitted by the author.

Table of Contents

Abstract

To increase our understanding of induced seismicity, we develop and implement methods to enhance seismic monitoring capabilities in northeastern British Columbia (NE BC). We deploy two different machine learning models to identify earthquake phases using waveform data from regional seismic stations and utilize an earthquake database management system to streamline the construction and maintenance of an up-to-date earthquake catalogue. The completion of this study allows for a comprehensive catalogue in NE BC from 2014 to 2020 by building upon our previous 2014—2016 and 2017—2018 catalogues. The bounds of the area where earthquakes were located were between 55.5˚N—60.0˚N and 119.8˚W—123.5˚W. The earthquakes in the catalogue were initially detected by machine learning models, then reviewed by an analyst to confirm correct identification, and finally located using the Non-Linear Location (NonLinLoc) algorithm. Two distinct sub-areas within the bounds consider different periods to supplement what was not covered in previously published reports – the Northern Montney Trend (NMT) is covered from 2017 to 2020 while the Kiskatinaw Seismic Monitoring and Mitigation Area (KSMMA) is covered from 2019 to 2020. The two sub-areas are distinguished by the BC Oil & Gas Commission (BCOGC) due to differences in their geographic location and geology. The catalogue was produced by picking arrival phases on continuous seismic waveforms from 51 stations operated by various organizations in the region. A total of 17,908 events passed our quality control criteria and are included in the final catalogue. Comparably, the routine Canadian National Seismograph Network (CNSN) catalogue reports 207 seismic events – all events in the CNSN catalogue are present in our catalogue. Our catalogue benefits from the use of enhanced station coverage and improved methodology. The total number of events in our catalogue in 2017, 2018, 2019, and 2020 were 62, 47, 9579 and 8220, respectively. The first two years correspond to seismicity in the NMT where poor station coverage makes it difficult to detect small magnitude events. The magnitude of completeness within the KSMMA (M_L = ~0.7) is significantly smaller than that obtained for the NMT (M_L = ~1.4). The new catalogue is released with separate files for origins, arrivals, and magnitudes which can be joined using the unique ID assigned to each event.

1. Introduction

Since the early 2000s, when unconventional oil and gas development began in NE BC, there has been a significant number of seismic events that occurred in the Montney Trend. Over 13,000 unique seismic events were detected from 2000 until 2019 (Figure 1), with the largest recorded magnitude of 4.6 being recorded twice, first in August of 2015 then in November of 2018. In recent years, the BCOGC, Natural Resources Canada (NRCan), Geoscience BC, and several universities have conducted monitoring and research operations within the study area to address the increased public concern and manage the risk imposed by induced seismicity. These efforts have led to the densification of seismic networks in the area, development of local velocity models and a local magnitude correction formula to improve earthquake location and magnitude accuracy, respectively, along with many other scientific contributions (Mahani and Kao, 2019; Mahani et al., 2021). From January through May 2020, the University of Calgary deployed 13 stations under the network code EO in the KSMMA. The EO stations supplement the existing networks (PQ, 1E, and XL), and continue to improve seismic monitoring in NE BC (Figure 2).

The BCOGC has designated two seismic monitoring areas in NE BC. These areas were suggested to demark the bounds where oil and gas related induced seismicity from primarily hydraulic fracturing has historically occurred and establish effective seismic monitoring in these locations. It is worth mentioning that magnitude 4+ events have occurred in both the KSMMA and the NMT (Mahani et al., 2017; Kao et al., 2018; Peña Castro et al., 2020; Roth et al., 2020). The two areas are distinguished in our report for a couple of reasons: (1) the previously published catalogues are not as complete for the NMT as they are for the KSMMA, and (2) the earthquakes that occur in each area need to be handled differently due to differences in the station coverage and geology, including different local velocity models. While the KSMMA has a dense network of stations allowing for many events to be detected and the earthquake locations to be of high accuracy with the average length of the major axis of the hypo-ellipse error being just ~700m, the error in the NMT, by comparison, is much higher at around 5000m.

2. Method – Developing an Advanced Earthquake Location System

2.1. The Need for Speed and Accuracy

The number of seismic stations available for monitoring in NE BC has greatly increased over the last 5 years, from fewer than 10 stations within 50 km of the Montney Trend to over 50 at the beginning of 2021. While the increase in the number of stations has had positive effects on the quality of seismic monitoring, it presents a challenge to maintaining the speed and efficiency of seismic analysis without increasing the personnel involved – this is because: (1) with more stations, a larger dataset needs to be processed to identify events, (2) more seismic events are detected since the minimum number of phases required to locate an event is met more often, and (3) more phases need to be uniquely identified for each event. Considering these factors, if the quality of monitoring is maintained, the amount of time required to analyze events can grow exponentially with increasing numbers of stations. The routine activities associated with seismic analysts represent good targets for automation. The time and resource savings are easy to quantify for routine processes and up-front costs associated with the development of the automation solutions can easily be justified. From experience, the main areas where time has been spent in seismic analysis, from most time consuming to least, are picking or adjusting phase arrival times, scanning waveforms, manually interacting with views of waveforms in earthquake analysis software, locating earthquakes, and calculating magnitude. We focused our efforts on addressing the most time-consuming tasks while keeping in mind the priority of ensuring that the quality of the catalogue is not unduly affected. The speed and accuracy of monitoring are both crucial to regulators who need access to high-quality and up-to-date information in order to effectively perform industry oversight and protect the public. Likewise, speed and accuracy are important for researchers to make observations and conduct science in a timely manner, the faster that publications can go out, the faster that others can build on the previous research. This is especially significant for issues that may involve public safety such as induced seismicity concerning seismic hazard and mitigation.

2.2. Phase Detection

The traditional way of conducting seismic monitoring has been effective, but is labour intensive, both with training and conducting routine seismic monitoring. For phase detection, an analyst will look at a computer screen and scroll through time windows of waveforms from seismic stations. These waveforms are usually bandpass-filtered to make the arrivals of seismic phases most visible. Due to the characteristic nature of the phases that appear on seismic traces, there have been a number of statistics developed to attempt to automate the phase detection process. A few of these in no particular order are the signal-to-noise ratio (SNR), Akaike-Information-Criterion (AIC), kurtosis, and scintillation index (Kao et al., 2007; Akram & Eaton, 2016). These generally take advantage of the fact that the background noise at the recording sites remains below a certain level during and immediately after the phase onset. These statistics usually require a threshold value which, when exceeded, will determine the phase onset time. If the threshold is set too low, it can be easily triggered by noise unrelated to seismicity, and, on the other hand, if set too high, it may not be triggered by an event at all. These methods by default cannot discriminate between *P* or *S* phases, so some higher-level programming logic must be employed to enable this capacity. With the advent of machine learning, there are several advantages over conventional techniques, with the caveat that they can be slower to process. The key advantages that lead us to utilize machine learning in our system instead of traditional earthquake detection techniques include: (1) lower rates of false phase detections from machine learning models relative to SNR and kurtosis statistics (we offer this anecdotally from our experience), (2) reliable discrimination between *P* and *S* phases without requiring any advanced logic to be programmed, and (3) fewer manual adjustments of phase arrival times are required with machine learning relative to statistics. We developed a Long Short-Term Memory (LSTM) phase picker, using earthquake statistics as input features with the idea that, when combined in a machine learning model, they can help discriminate earthquake signal from noise better than any one statistic on its own (Figure 3.).

2.2.1. Statistics for the LSTM

The SNR is defined as the short-term average divided by the long-term average of the waveform signal. The long-term average is representative of the noise, while the short-term average is the signal component. This statistic works reliably on its own and is used commonly to identify earthquake *P* phases. There are difficulties using the SNR on stations where there is active site noise, which is common at stations around oil and gas operations in NE BC. There is also the challenge of setting the optimal threshold value to identify phases, since the threshold value may be different from station to station and region to region. We use a signal time window of 1 second and a noise time window length of 10 seconds.

Kurtosis is a statistic that describes the shape of the probability distribution and is a measure of the tailedness of the distribution. The kurtosis of a normal distribution is 3; distributions with a kurtosis greater than 3 indicate more outliers than the normal distribution, such is the case when arrival phases first appear on the waveform and generate a non-Gaussian wavefield (Baillard et al., 2014). For our purpose, we used the adjusted version which is the kurtosis minus 3 to establish 0 as the baseline. Any negative kurtosis values were zeroed since they were not useful to identify phase onsets. We use two different lengths of kurtosis time windows since they provide complimentary perspectives on the distribution that can reveal different onset. A shorter time window provides a higher temporal resolution for identifying impulsive phase arrivals; on the other hand, longer time windows are less sensitive to varying noise levels and can help reveal emergent signals. In addition to the kurtosis, we calculate the kurtosis rate, *kr*, as:

$$kr_n = \frac{k_n - k_{n-5}}{5}, \qquad (1)$$

where k_n is the kurtosis value of the n-th sample. The kurtosis rate has the added benefit of making the signal spikes, caused by phase onsets, even more clear. The kurtosis rates are calculated for short- and long-time windows of 1 and 5 s, respectively.

The scintillation index (SI) is a parameter originally proposed to measure the intensity of energy bursts of radio waves in the ionosphere. The SI is the square root of the normalized variance of the intensity of the signal and has successfully been used to identify earthquakes and tremor signals (Kao et al., 2007). Similar to the SNR and kurtosis, the SI generally serves as a diagnostic function measuring changes in waveform pattern, which makes it one of the key features in our model for characterization and reliable detection of local events. We use a time window of 8 seconds to calculate the SI.

2.2.2. Model Architecture and Training Data

To create the model, we tried both Gated Recurrent Unit (GRU) and Long-Short-Term-Memory (LSTM) architectures since they have a similar control flow and can be easily converted from one to another. In the end, we obtain the lowest loss on training and validation datasets with an LSTM with the following setup: bidirectional LSTM with 20 units and a sigmoid activation layer, dropout layer (with a dropout rate of $\alpha = 0.2$), second bidirectional LSTM with 5 units and a sigmoid activation layer, dropout layer ($\alpha = 0.2$), and finally, a dense layer with a SoftMax activation function. The loss function used is categorical cross-entropy with the Adam optimization function, a stochastic gradient descent method with adaptive learning rate optimization. The dataset is split between a training and validation dataset, with 80% for training and 20% for validation. The final model is the one with the lowest loss on the validation dataset after 10 training epochs. The model with the lowest validation loss is the one trained at 4 epochs.

The model is trained to differentiate if noise, a *P* phase arrival, or an *S* phase arrival is present at any time sample except for the first or final 10 samples of a sequence. We do not train the edges of the sequences because we posit that picks in these margins will be of lower quality due to not having as much information as is present in the middle of the sequences. To increase the processing speed, we calculate the statistics on the raw waveform then downsample the statistics to 20 Hz. We train the model using sequences with 100 samples or a total of 5 seconds of data. We

experiment with sequence lengths greater than 100, but encounter the exploding/vanishing gradient problem that is common with LSTMs. Prior to calculating the statistics, we filter the waveform with a 1-15 Hz bandpass filter to remove unwanted frequency information. Since the samples with phase picks are relatively rare in comparison to samples without, we smooth the labels around the *P* and *S* picks and weight samples with *P* and *S* picks higher to further account for the class imbalance. The model outputs are P and S phase detection probabilities at every time step with values ranging from 0 to 1.

2.2.3. Model Results

The trained model succeeds at detecting the majority of *P* and *S* phases. We compare the phase picks of SNR, kurtosis rate, and model predictions with 987 manually detected phases for the period December 1, 2020, to December 14, 2020, on stations from the 1E network. The model is found to be able to successfully identify approximately ~87% of the phase picks in the catalogue within 0.5 seconds of the manually reviewed arrival times. We encounter a false-to-real detection ratio of ~77. When we compare this to the SNR measurements with a long-term window of 10 seconds, a short-term window of 1 second and a threshold of 2 for the same period, it finds ~40% of the picks with a false-to-real detection ratio of over 1000, while the kurtosis (centered at 0) with a time window of 1 second and a threshold of 7 finds nearly 55% of arrivals with a false-to-real detection ratio of over 1000. In addition, the deep learning-based model achieves significantly smaller time errors. For matching arrivals within 0.5 seconds from the reviewed picks, the LSTM model results in an average time error of 0.07 s while the SNR and kurtosis picks show average errors of 0.33 and 0.21 s, respectively. Lower time error reduces the time required to manually adjust phases, and allows more restrictive phase association criteria to be implemented to reduce false events. Reducing the number of false picks relative to real picks speeds up processing and reduces the number of false events.

2.2.4. Phase Picker Settings

To create the catalogue, we use two phase pickers: the LSTM-based phase picker and EQTransformer (Mousavi et al, 2020). We use both to ensure that we obtain as many high-quality phases as possible and get multiple perspectives on the same waveforms. Both of those phase pickers identify *P* and *S* phases. The LSTM picker, which is used only for earthquakes in the KSMMA, has probability thresholds of 0.8 for *P* phases and 0.9 for *S* phases. On the other hand, the EQTransformer model is used for both the KSMMA and the NMT since it is trained on a significantly larger dataset with stations at greater distances. The probability threshold for EQTransformer is set at 0.2 for *P* phases and 0.3 for *S* phases. The low rate of false phase detections with EQTransformer allows the low probability threshold values to be employed to detect more events.

2.3. Phase Association

Once phases are identified, the next step for traditional seismic analysis is going through phases, removing false or mis-timed phase picks, and choosing which phases should be associated with each other. This is often performed in an earthquake location software such as Antelope or SeisComP. We automate this task through the use of a travel time grid search which helps us identify phases that could be associated with each other at potential source locations throughout the study area. Grid nodes are established across the study area at 1-km horizontal intervals and at different depths ranging from 0 to 10 km at 2-km intervals. For each arrival, we first determine a theoretical origin time for every grid node. Then for other arrivals, we calculate their corresponding origin times at each node with a time allowance of +/- 2 s. If the other arrivals come within the time allowance and have at least 7 phases matched in the KSMMA or 5 phases matched in the NMT at any grid node and origin time, the event is deemed to be real and passed onwards to location processing. If multiple grid nodes exceed the threshold, the one with the most associated phases is chosen. Theoretical travel times are calculated using local velocity models for the KSMMA and the NMT (Mahani et al., 2020). We use different phase counts as thresholds for the different areas because of the difference in the number of stations in each area. Through the association step, we get rid of the majority of false-positive phases

and false events while allowing for minor mis-timed phase picks which can later be corrected. Some legitimate earthquakes with a low number of phases and/or with significant picking issues might be eliminated using this method.

2.4. Event Location

Like phase associations, the traditional earthquake location techniques are commonly performed using an earthquake location software such as Antelope or SeisComP. In this study, a combination of automatic and manual processes is used to locate events. First, initial locations are determined automatically by the open-source software NonLinLoc (Lomax et al., 2000). If solutions converge and are not located on the border of the study area (which indicated that the true locations are likely to be outside the bounds of the study area), they are kept. The remaining events are then prepared to be manually reviewed. The Antelope DBLOC2 program is used to manually inspect waveforms – to add missing phases, adjust phase timings, and confirm that earthquake solutions fall within the study area. We manually ensure that *P* phases are picked on the vertical components while *S* phases are picked on a horizontal component. In this study, we limit the theoretical travel times to epicentral distances of up to 150 km to avoid instances with multiple prominent, indirect phase arrivals. After the manual review, the arrivals from events located with Antelope are given to NonLinLoc to obtain probabilistic earthquake locations.

The locations for automatic and post-review events are performed by exporting the phase arrival information into NonLinLoc's input format. The locations are determined using the Oct-Tree algorithm, an importance sampling-based technique, where the area is divided into coarse grid nodes and misfit is calculated at each node (Lomax & Curtis, 2001). The node with the lowest misfit is chosen as the centre for a new cube with smaller bounds than the origin grid node spacing. New cubes are continuously created until a set number of iterations are reached or the grid spacing is smaller than a predefined threshold. If an earthquake solution falls on a node on the edge of the grid, it is removed since the event is likely to have occurred outside of the study area.

2.5. The Database Management System

The Earthquake Database Management System (EQDBMS) is created using MySQL which allows multiple catalogues to be combined without needing to deal with duplicate events and is accessible using SQL queries. Once earthquakes are located, the event details are output into the EQDBMS. The system associates events from the different catalogues by comparing origin times that are within +/- 5 seconds of each other, and the associated events are eventually prioritized via a user-defined hierarchy. Our catalogue priorities, from the least to the most important, are automatic NMT, automatic KSMMA, manually reviewed NMT, and manually reviewed KSMMA. For the events located in the overlapping area between the two sub-regions, the KSMMA is prioritized over the NMT due to an enhanced station coverage in the Kiskatinaw monitoring area. Once events are associated and the hierarchy of importance is defined, the event with the highest priority is labelled as the preferred event and is output to the final catalogue. These events are then able to be queried easily using SQL.

2.6. Magnitudes

Magnitudes are determined using the ObsPy Python package. Single-station magnitude estimates are determined from arrivals that are extracted from the EQDBMS for new events that do not yet have an event magnitude, using the magnitude distance correction formula suggested for NE BC (Mahani and Kao, 2020). An event magnitude is determined as the median value of the station magnitudes which are calculated from the vertical component seismograms high-pass filtered at 1 Hz. Once calculated, the event magnitudes are updated in the EQDBMS.

3. Results

3.1 Spatiotemporal Distribution of Seismicity

In the KSMMA during 2019 and 2020, there were a total of 17,597 seismic events detected with an average of approximately 730 events per month (Figure 4). We observe the highest and lowest seismicity rates in March 2020 and July 2019, respectively, with

the corresponding number of events being 2,600 and 100 (Figure 5). The differences between seismicity during different months are assumed to be attributed to the spatiotemporal distribution of the injection operations associated with the development of unconventional hydrocarbon resources (Atkinson et al., 2016; Kao et al., 2018). In the KSMMA, seismicity clustered tightly, presumably around injection activity, and the epicentral locations appears to follow a NW-SE trend. There were more earthquakes during 2019 than 2020, with 9467 observed during that period compared to 8130 during 2020. This could be due to a slowdown in oil and gas operations during 2020 relative to 2019 or because operators may be targeting locations known to be aseismic to mitigate their risk.

In the NMT, from 2017 through 2020, there was an average of 7 events per month with a total of 314 events (Figure 6). The month with the highest amount of seismicity was February 2019 with 24 events detected. There are five months with fewer than 5 events detected. The reason there were much fewer earthquakes in the NMT than the KSMMA is likely partially due to fewer stations in the area, making it more difficult to detect small earthquakes there; it is also possible there were fewer oil & gas operations in the area compared with the KSMMA. Seismicity in the NMT appears to be sporadic with no obvious dominant trend regarding the orientation of regional tectonic structures. On an annual basis, seismicity was the highest during 2019 with 113 events observed, while 2017, 2018, and 2020 had 62, 47, and 92 events observed, respectively.

3.2 Quality Control Criteria

In order to only include high-quality earthquake locations in the catalogue, we use the major axis error of the location solution's hypo-ellipse. Since the NMT and KSMMA fall in two distinct areas with differences in station coverage, we have separate quality control criteria for each area. In the NMT, the station coverage is sparse relative to the KSMMA; therefore, we allow greater location uncertainty and set a threshold value of 20 km for the length of the major axis of the location's hypo-ellipse. On the other hand, for the KSMMA area, denser station coverage enables higher location accuracy and a lower threshold of 10 km.

3.3 Magnitude of Completeness

The magnitudes of completeness within the KSMMA and NMT are quite different due to different seismic station densities in the two areas. The NMT has fewer than 10 stations within 20 km of its bounds, while the KSMMA has almost 40. As a result, the magnitude of completeness in the KSMMA is ~0.7 M_L while the magnitude of completeness in the NMT is ~1.4 (Figure 7.).

4. Conclusion

To compile the most complete catalogue possible with the resources available, we have undertaken several initiatives to streamline and improve the earthquake location process. These include the deployment of machine-learning models to pick seismic arrival phases and the development of an earthquake database management system. In this approach, though an analyst is still required to verify earthquake arrival phases, all other steps have been automated. The resulting catalogue extends the existing seismic record in NE BC with an additional 17,908 events. The vast majority of those events occurred within the KSMMA from 2019-2020 with 17,597, while the NMT had just 314 events detected from 2017-2020. The drastic difference in the number of earthquakes is partly due to the limited number of seismic stations available locally in the NMT, the different subsurface structural settings, and the much lower number of HF operations in the NMT. The magnitude of completeness is ~1.4 in the NMT and ~0.7 in the KSMMA.

5. References

Akram, J., & Eaton, D. W. (2016). A review and appraisal of arrival-time picking methods for downhole microseismic data. *Geophysics, 81*(2), KS71-KS91.

Atkinson, G. M., Eaton, D. W., Ghofrani, H., Walker, D., Cheadle, B., Schultz, R., ... & Kao, H. (2016). Hydraulic fracturing and seismicity in the Western Canada Sedimentary Basin. *Seismological research letters*, *87*(3), 631-647.

Baillard, C., Crawford, W. C., Ballu, V., Hibert, C., & Mangeney, A. (2014). An automatic kurtosis-based P-and S-phase picker designed for local seismic networks. *Bulletin of the Seismological Society of America*, *104*(1), 394-409.

Kao, H., Thompson, P. J., Rogers, G., Dragert, H., & Spence, G. (2007). Automatic detection and characterization of seismic tremors in northern Cascadia. *Geophysical Research Letters*, *34*(16), 1–6. https://doi.org/10.1029/2007GL030822

Kao, H., Visser, R., Smith, B., & Venables, S. (2018). Performance assessment of the induced seismicity traffic light protocol for northeastern British Columbia and western Alberta. *The Leading Edge*, *37*(2), 117-126.

Lomax, A., & Curtis, A. (2001). Fast, probabilistic earthquake location in 3D models using oct-tree importance sampling. In *Geophys. Res. Abstr* (Vol. 3, p. 955).

Lomax, A., Virieux, J., Volant, P., & Berge-Thierry, C. (2000). Probabilistic earthquake location in 3D and layered models. In *Advances in seismic event location* (pp. 101-134). Springer, Dordrecht.

Mahani, A. B., Schultz, R., Kao, H., Walker, D., Johnson, J., & Salas, C. (2017). Fluid injection and seismic activity in the northern Montney play, British Columbia, Canada, with special reference to the 17 August 2015 M w 4.6 induced earthquake. *Bulletin of the Seismological Society of America*, *107*(2), 542-552.

Mahani, A. B., & Kao, H. (2019). Accurate determination of local magnitude for earthquakes in the Western Canada Sedimentary basin. *Seismological Research Letters*, *90*(1), 203–211. https://doi.org/10.1785/0220180264

Mahani, A. B., Malytskyy, D., Visser, R., Hayes, M., Gaucher, M., & Kao, H. (2021). Well-log-based velocity and density models for the montney unconventional resource play in Northeast British Columbia, Canada, applicable to induced seismicity monitoring and research. *Seismological Research Letters*, *92*(2), 886–894. https://doi.org/10.1785/0220200213

Mousavi, S. M., Ellsworth, W. L., Zhu, W., Chuang, L. Y., & Beroza, G. C. (2020). Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nature Communications*, *11*(1), 1–12. https://doi.org/10.1038/s41467-020-17591-w

Peña Castro, A. F., Roth, M. P., Verdecchia, A., Onwuemeka, J., Liu, Y., Harrington, R. M., ... & Kao, H. (2020). Stress chatter on a fracture network reactivated by hydraulic fracturing.

Roth, M. P., Verdecchia, A., Harrington, R. M., & Liu, Y. (2020). High-resolution imaging of hydraulic-fracturing-induced earthquake clusters in the Dawson-Septimus area, Northeast British Columbia, Canada. *Seismological Research Letters*, *91*(5), 2744-2756.

Visser, R., Kao, H., Smith, B., Goerzen, C., Kontou, B., Dokht, R. M. H., … Mahani, A. B. (2020). A comprehensive earthquake catalogue for the Fort St. John – Dawson Creek region, British Columbia, 2017 – 2018. *Geological Survey of Canada, Open File*, *8718*, 1–28.

Visser, R., Smith, B., Kao, H., Babaie Mahani, A., Hutchinson, J., & McKay, J. E. (2017). A comprehensive earthquake catalogue for northeastern British Columbia and western Alberta, 2014-2016. *Geological Survey of Canada, Open File*, *8335*(February 2018), 1–28.

Appendix 1: Earthquake origins

Column	Type	Description
id	int	Unique id given to each origin. Used to join arrivals and station magnitudes.
Latitude	float(4)	Latitude in degrees North.
Longitude	float(4)	Longitude in degrees East.
Depth	float(1)	Depth of event in km.
Mag	float(1)	Median of station magnitudes.
MagType	string	All rows are Mlv - local magnitude determined on the vertical waveform component.
Datetime	datetime	The date and time of the event origin.
LocationError	float(2)	Standard Error - RMS of the travel time residuals of the arrivals used for location.
MajaxError	float(2)	Maximum horizontal uncertainty of 68% confidence ellipse in km
MinaxError	float(2)	Minimum horizontal uncertainty of 68% confidence ellipse in km.
Azimuth	float(1)	Azimuth of major axis of 68% confidence ellipse in degrees.
VelocityModel	string	Local velocity model used to locate event. Either NMT or KSMMA.
Ndef	int	Number of phases used to locate event.
Nmag	int	Number of station magnitudes used to determine event magnitude.

Appendix 2: Earthquake arrivals

Column	Type	Description
id	int	Unique id given to each origin. Used to join origins, arrivals, and magnitudes.
Datetime	datetime	The date and time of the phase arrival.
Sta	string	The station code used to identify the station.
Chan	string	The channel code used to identify on which channel the arrival was picked.
Phase	string	The phase code used to identify the phase – either *P* or *S*.

Appendix 3: Earthquake magnitudes

Column	Type	Description
id	int	Unique id given to each origin. Used to join origins, arrivals, and magnitudes.
Datetime	datetime	The date and time that the station amplitude value was taken.
Net	string	The network code to identify the network.
Sta	string	The station code used to identify the station.
Chan	string	The channel code used to identify on which channel the arrival was picked.
Mag	float(2)	The local station magnitude value calculated on the vertical component.

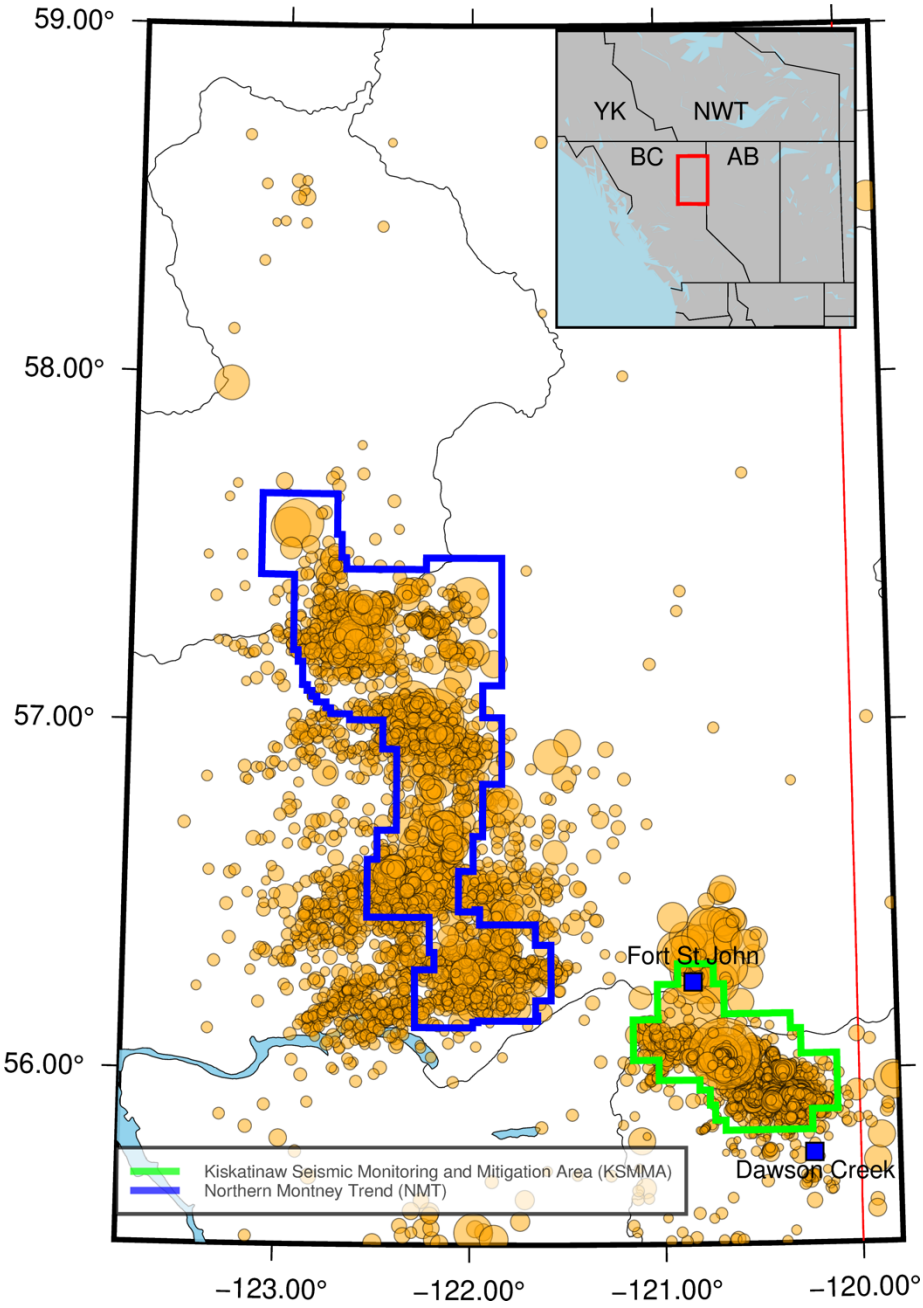

Figure 1. Previously published earthquakes from 2000-2018 (Earthquakes Canada, Visser et al., 2017, Visser et al., 2020.) The blue and green lines mark the seismic monitoring areas in NE BC established by the BC Oil & Gas Commission.

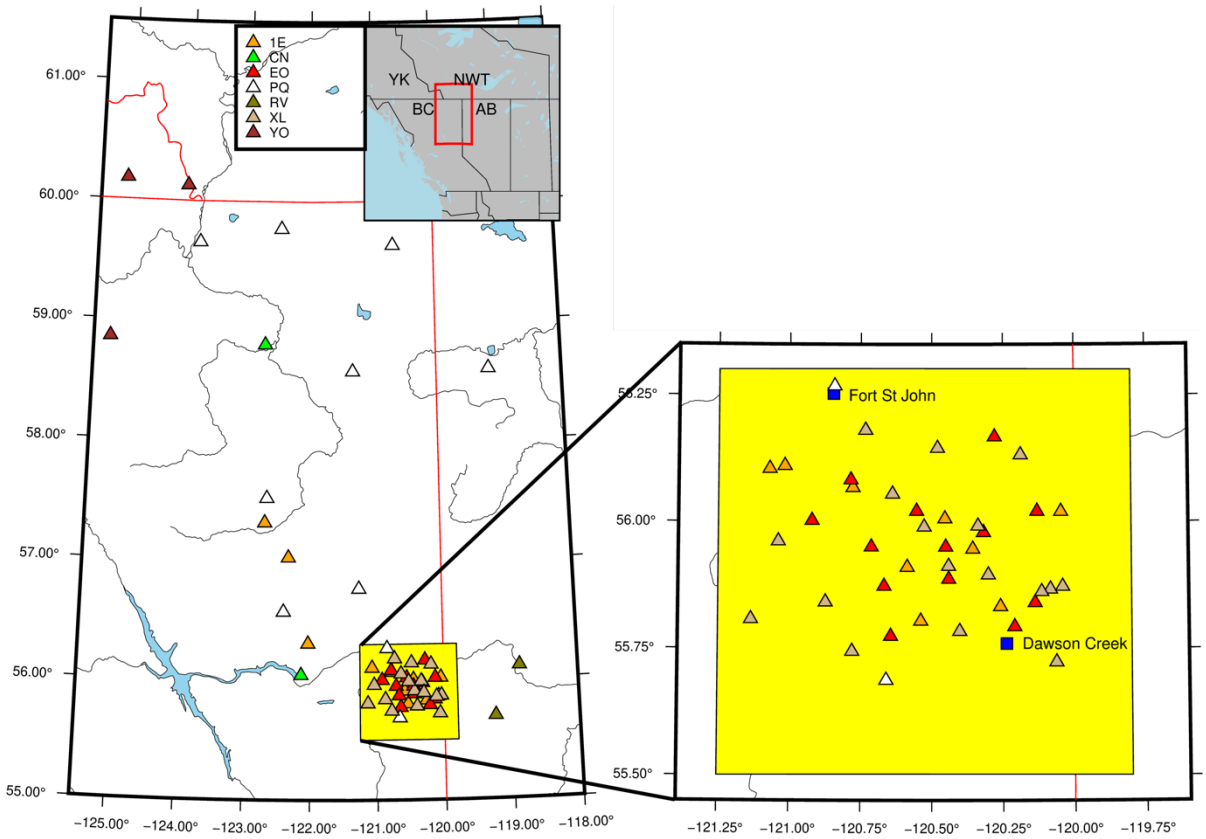

Figure 2. Distribution of seismic stations in NE BC from 2019-2020. The colour of the station indicates the corresponding network. The networks were deployed and are maintained by different groups or combinations of groups.

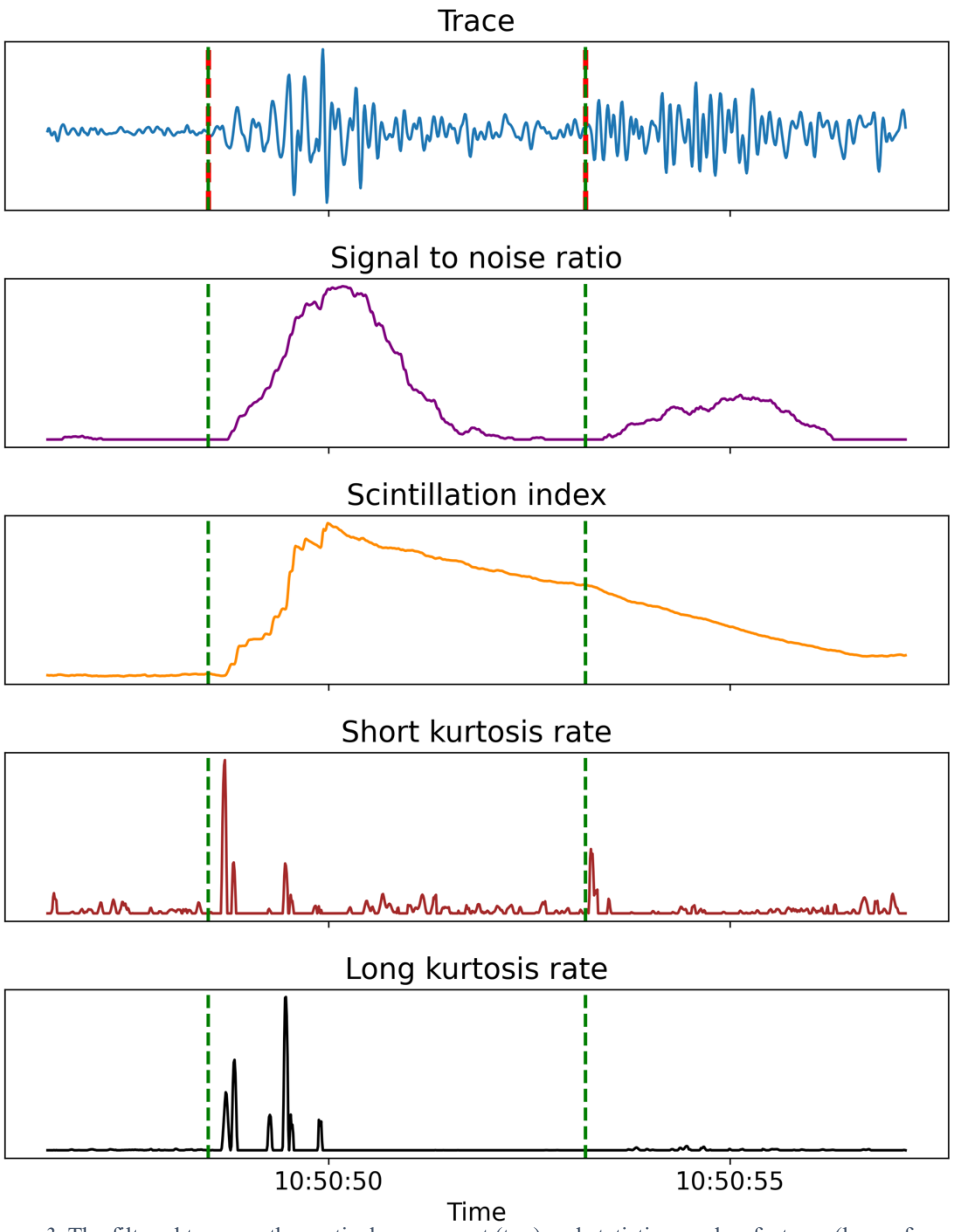

Figure 3. The filtered trace on the vertical component (top) and statistics used as features (lower four plots) in the machine learning model. The manually reviewed arrival times are shown as red dashed lines. The LSTM phase picks are shown as green dashed lines.

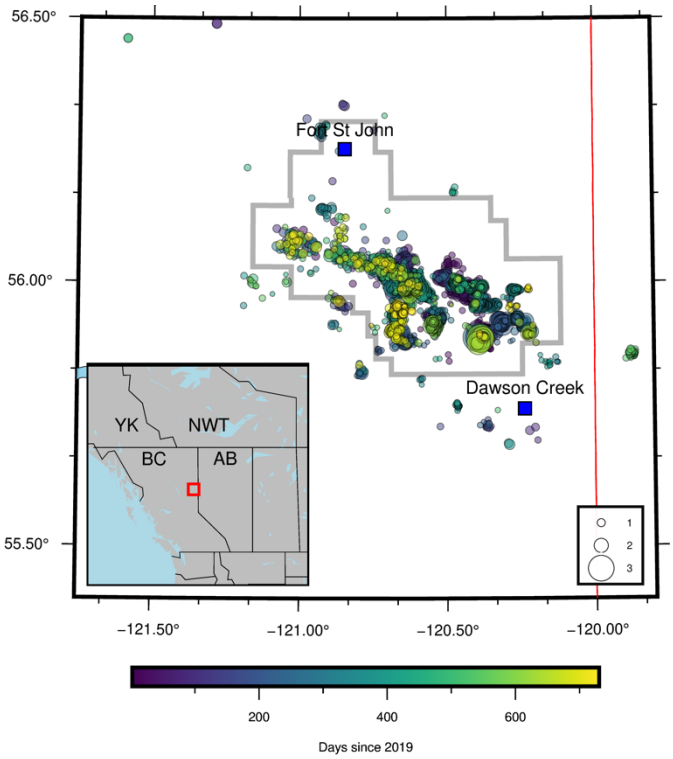

*Figure 4. E*arthquakes detected in the KSMMA from 2019-2020. The colour shows when the earthquake occurred in the number of days since January 1st, 2019. The size of the circles illustrates the local magnitude of the seismic events.

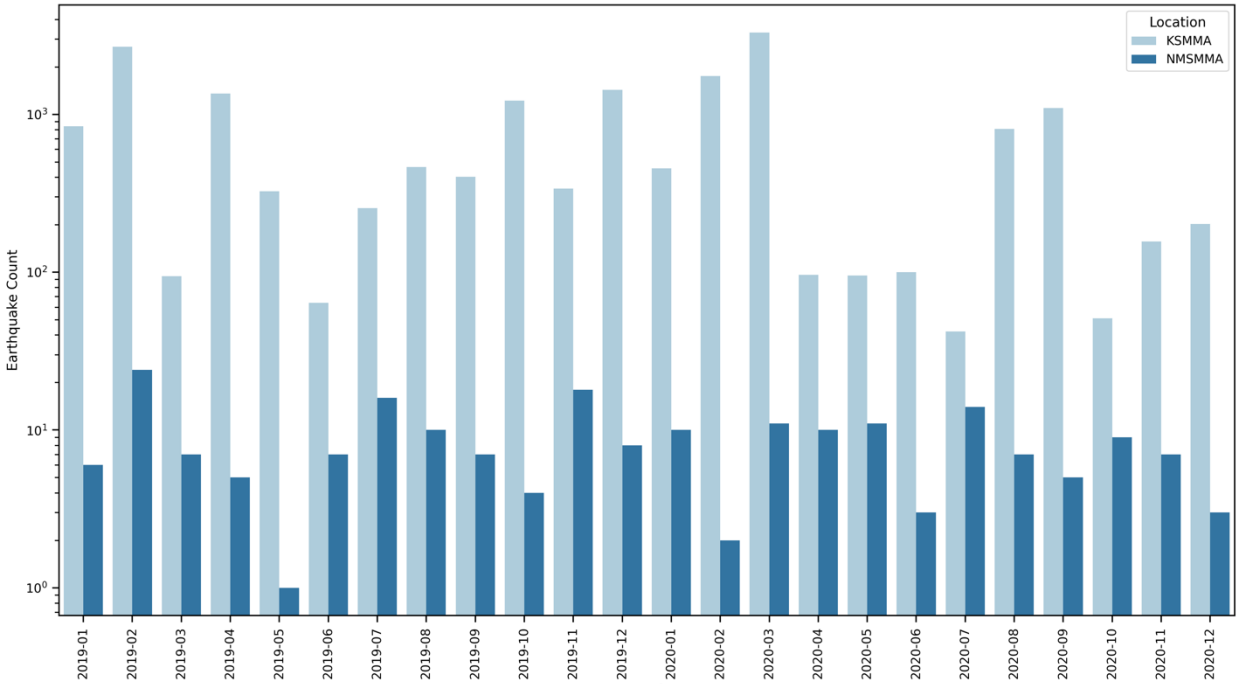

Figure 5. Number of earthquakes per month in the KSMMA (light blue) and NMSMMA (dark blue) seismic monitoring areas within NE BC from 2019-2020.

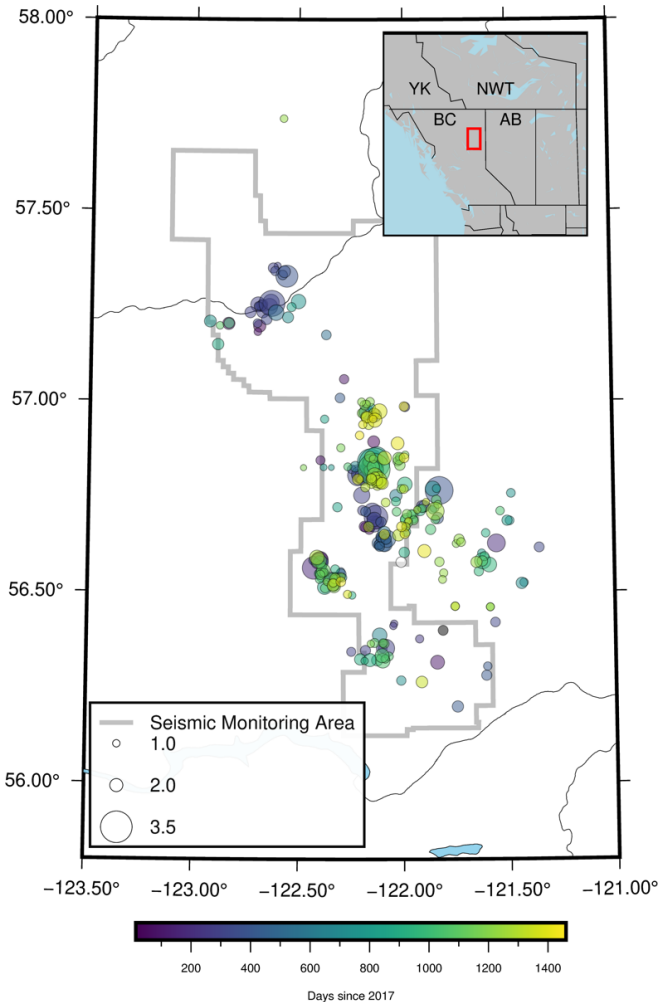

Figure 6. Earthquakes that occurred within the NMSMMA from 2017-2020. Colour indicates when the earthquakes occurred in days since January 1st, 2017. The size of the circle illustrates the local magnitude of the seismic events.

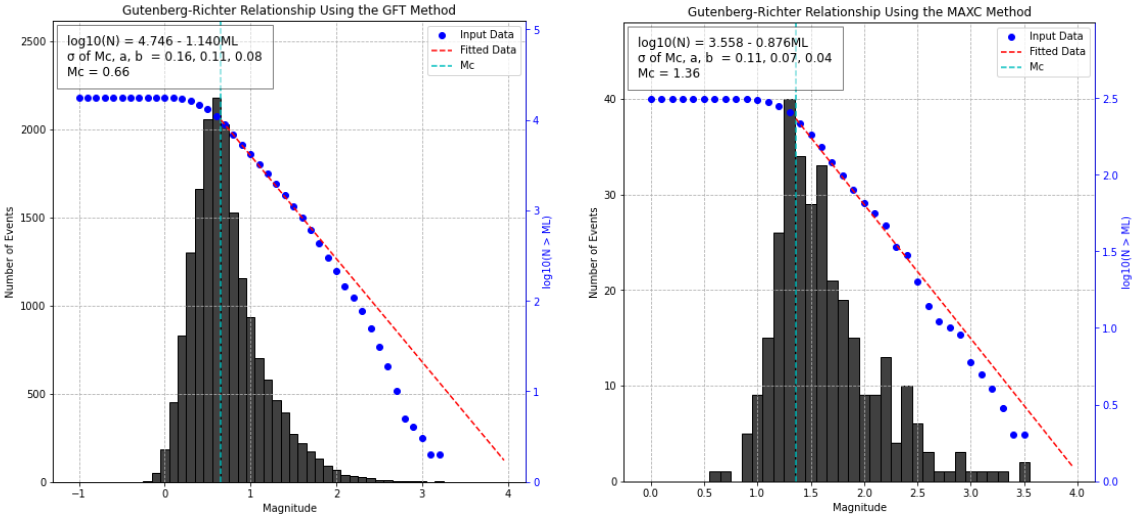

Figure 7. Earthquake frequency-magnitude distributions for the KSMMA (left) and the NMSMMA (right). The red dashed lines represent the estimated Gutenberg-Richter (GR) relationships.