



RESEARCH LETTER

10.1029/2022GL099995

Key Points:

- We use supervised machine learning to investigate the relationship between hydraulic fracturing operation parameters and induced seismicity
- Geological properties and a limited number of operational parameters predominantly influence the probability of an induced earthquake
- The approach has the potential to guide detailed investigations of injection parameters critical for inducing earthquakes

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

K. B. Kemna,
kilian.kemna@rub.de

Citation:

Kemna, K. B., Roth, M. P., Wache, R. M., Harrington, R. M., & Liu, Y. (2022). Small magnitude events highlight the correlation between hydraulic fracturing injection parameters, geological factors, and earthquake occurrence. *Geophysical Research Letters*, 49, e2022GL099995. <https://doi.org/10.1029/2022GL099995>

Received 9 JUN 2022

Accepted 19 OCT 2022

Author Contributions:

Conceptualization: Kilian B. Kemna
Data curation: Kilian B. Kemna, Marco P. Roth, Ricarda M. Wache
Formal analysis: Kilian B. Kemna, Marco P. Roth, Ricarda M. Wache
Investigation: Kilian B. Kemna
Methodology: Kilian B. Kemna, Marco P. Roth, Ricarda M. Wache
Project Administration: Kilian B. Kemna
Software: Kilian B. Kemna, Marco P. Roth, Ricarda M. Wache
Supervision: Rebecca M. Harrington, Yajing Liu

© 2022. The Authors.

This is an open access article under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Small Magnitude Events Highlight the Correlation Between Hydraulic Fracturing Injection Parameters, Geological Factors, and Earthquake Occurrence

Kilian B. Kemna¹ , Marco P. Roth¹ , Ricarda M. Wache¹, Rebecca M. Harrington¹ , and Yajing Liu² 

¹Institute of Geology, Mineralogy and Geophysics, Ruhr University Bochum, Bochum, Germany, ²Department of Earth and Planetary Sciences, McGill University, Montréal, Québec, Canada

Abstract Hydraulic fracturing (HF) operations are widely associated with induced seismicity in the Western Canadian Sedimentary Basin. This study correlates injection parameters of 12,903 HF stages in the Kiskatinaw area in northeast British Columbia with an enhanced catalog containing 40,046 earthquakes using a supervised machine learning approach. It identifies relevant combinations of geological and operational parameters related to individual HF stages in efforts to decipher fault activation mechanisms. Our results suggest that stages targeting specific geological units (here, the Lower Montney formation) are more likely to induce an earthquake. Additional parameters positively correlated with earthquake likelihood include target formation thickness, injection volume, and completion date. Furthermore, the COVID-19 lockdown may have reduced the potential cumulative effect of HF operations. Our results demonstrate the value of machine learning approaches for implementation as guidance tools that help facilitate safe development of unconventional energy technologies.

Plain Language Summary Hydraulic fracturing (HF), a technique used in unconventional energy production, increases rock permeability to enhance fluid movement. Its use has led to an unprecedented increase of associated earthquakes in the Western Canadian Sedimentary Basin in the last decade, among other regions. Numerous studies have investigated the relationship between induced earthquakes and HF operations, but the connection between specific geological and operational parameters and earthquake occurrence is only partly understood. Here, we use a supervised machine learning approach with publicly available injection data from the British Columbia Oil and Gas Commission to identify influential HF parameters for increasing the likelihood of a specific operation inducing an earthquake. We find that geological parameters, such as the target formation and its thickness, are most influential. A small number of operational parameters are also important, such as the injected fluid volume and the operation date. Our findings demonstrate an approach with the potential to develop tools to help enable the continued development of alternative energy technology. They also emphasize the need for public access to operational data to estimate and reduce the hazard and associated risk of induced seismicity.

1. Introduction

Permeability enhancement in unconventional oil and gas production in tight-shale formations through hydraulic fracturing (HF) has been clearly associated with induced earthquakes in many areas, such as the Sichuan Basin, China (e.g., Lei et al., 2019), central and eastern United States (e.g., Skoumal et al., 2020), and the Western Canadian Sedimentary Basin (WCSB) (e.g., Schultz et al., 2015). The WCSB in particular experienced a drastic increase in HF-associated induced seismicity, including earthquakes with magnitudes as large as M_w 4.6 (e.g., Babaie Mahani et al., 2017), and nine M3+ events in the Kiskatinaw area since 2010 (Figure 1a).

Most wells in Kiskatinaw target the Montney Formation (Figure 1a1), a fine-grained, westward thickening siltstone wedge that reaches ~300 m (Davies, 1997; Edwards et al., 1994). Seismicity in Kiskatinaw occurs in swarm-like sequences that are spatio-temporally correlated with HF-well treatments (e.g., Schultz et al., 2020). HF wells commonly consist of multiple trajectories, each with multiple HF stages spaced at intervals of ~50 m (BCOGC, 2022). The heightened seismic response to industrial activity in the area prompted seismic station densification efforts that enabled the development of detailed earthquake catalogs (e.g., Roth et al., 2020). It also led to the BC Oil and Gas Commission (BCOGC) designation of the *Kiskatinaw Seismic Monitoring and*

Visualization: Kilian B. Kemna
Writing – original draft: Kilian B. Kemna
Writing – review & editing: Marco P. Roth, Ricarda M. Wache, Rebecca M. Harrington, Yajing Liu

Mitigation Area (referred to as Kiskatinaw, BCOGC, 2018), which requires mandatory publication of detailed injection parameters for individual HF stages. A comparison of earthquake density to active well locations using the dense and nearly full-azimuthal seismic station coverage in this area (Figure 1a) shows that only a small fraction of wells or stages (~15%) induce earthquakes above the detection threshold. Therefore, identifying the combination of relevant geological and HF parameters and quantifying their correlation with earthquake occurrence is key to mitigating hazard associated with induced earthquakes (Ghofrani & Atkinson, 2020).

Here, we use a supervised machine learning algorithm to correlate injection parameters from 12,903 HF stages with detectable seismicity in the Kiskatinaw area. We first enhance the existing earthquake catalog using an AI-based phase-arrival picker combined with template matching to lower the magnitude of completeness (M_C) and increase the number of detected earthquakes. We then add a binary target variable to the HF-stage data set based on whether a particular stage may be spatio-temporally associated with an earthquake from the enhanced catalog. Next, we use tree-based models (random forests and gradient boosted trees) combined with a tree-explainer to interpret the models. Our interpretation focuses on the importance of single injection parameters, parameter interdependence, and their distribution in different geological subformations. We show that the target formation thickness and distance to the Precambrian basement exert primary influence on seismicity occurrence. The majority of individual stage injection parameters exert a comparatively weaker influence, with the exception of fluid volume and treatment date. Our results highlight the benefits of machine learning approaches in determining the relative importance of the available set of HF-stage parameters for generating induced earthquakes. They also highlight the potential to decipher mechanisms driving induced fault activation where more detailed HF-stage parameters are available. This study also underlines the significance of open-data access to enable induced seismicity research aimed at mitigating seismic hazard and associated risks.

In the following, we first describe the catalog enhancement, followed by the machine-learning workflow. We then discuss the results for each stage parameter on a collective basis and the results for individual stages, while comparing our results to findings from previous studies.

2. Method

The machine learning approach determines the relative importance of both geological and operational parameters on a per-stage basis and their potential association with earthquakes in an enhanced catalog. Catalog enhancement uses a multi-station matched filter (MMF, Figure S1 in Supporting Information S1) on an initial earthquake catalog (Roth et al., 2020) combined with the AI-based phase arrival picker EQTransformer (EQT, Figure S2 in Supporting Information S1, Mousavi et al., 2020) from 12 July 2017 to 31 December 2020. We locate the detections and calculate magnitudes as done for the initial catalog and relocate earthquakes using hypoDD (Waldhauser, 2000). We then obtain detailed HF-stage injection information from the BCOGC database (BCOGC, 2022) for 865 HF wells within the spatial area shown in Figure 1 that were actively stimulated during the catalog period. We extract the start-/end-times and injected fluid volumes for individual stages, and associate geological parameters to each stage, such as the target formation and thickness. We follow BCOGC (2012) by separating the Montney Formation into Lower and Upper units. Davies et al. (2018) further distinguishes the Montney Formation into Lower, Lower Middle, Upper Middle, and Upper units, where the Lower and Lower Middle Montney correspond to the Lower Montney and the Upper Middle and Upper to the Upper Montney of BCOGC (2012). We then spatio-temporally associate earthquakes $\geq M_C$ in the enhanced catalog to individual stages (horizontal event-stage distance ≤ 5 km, 5-day window from the HF treatment, Figure 1b, Figure S4 in Supporting Information S1). The 5-day time-window limit is intended to account for cases of delayed triggering and could potentially lead to misassociation. However, as later discussed in the Results, 82% of events occur within 3 hr of a specific stage. The Supporting Information also provides further details of catalog enhancement, stage information retrieval, and earthquake-stage association.

Following earthquake-stage association, we apply the machine learning approach with workflow shown in Figure 1c. The approach sets the 13-parameter stage parameter data set as independent variables and the condition of whether or not the stage has an associated earthquake as the dependent variable. We use a stratified split to generate a training (80%) and test (20%) set, where the stratified split ensures training and test sets contain equal numbers of stages with and without earthquakes. We leverage a repeated k-Fold cross validator with five splits and 10 repetitions on the training set to obtain a validation set and estimate confidence intervals. We then follow Howard and Guggen (2020) using a random forest classifier to remove insignificant columns in the

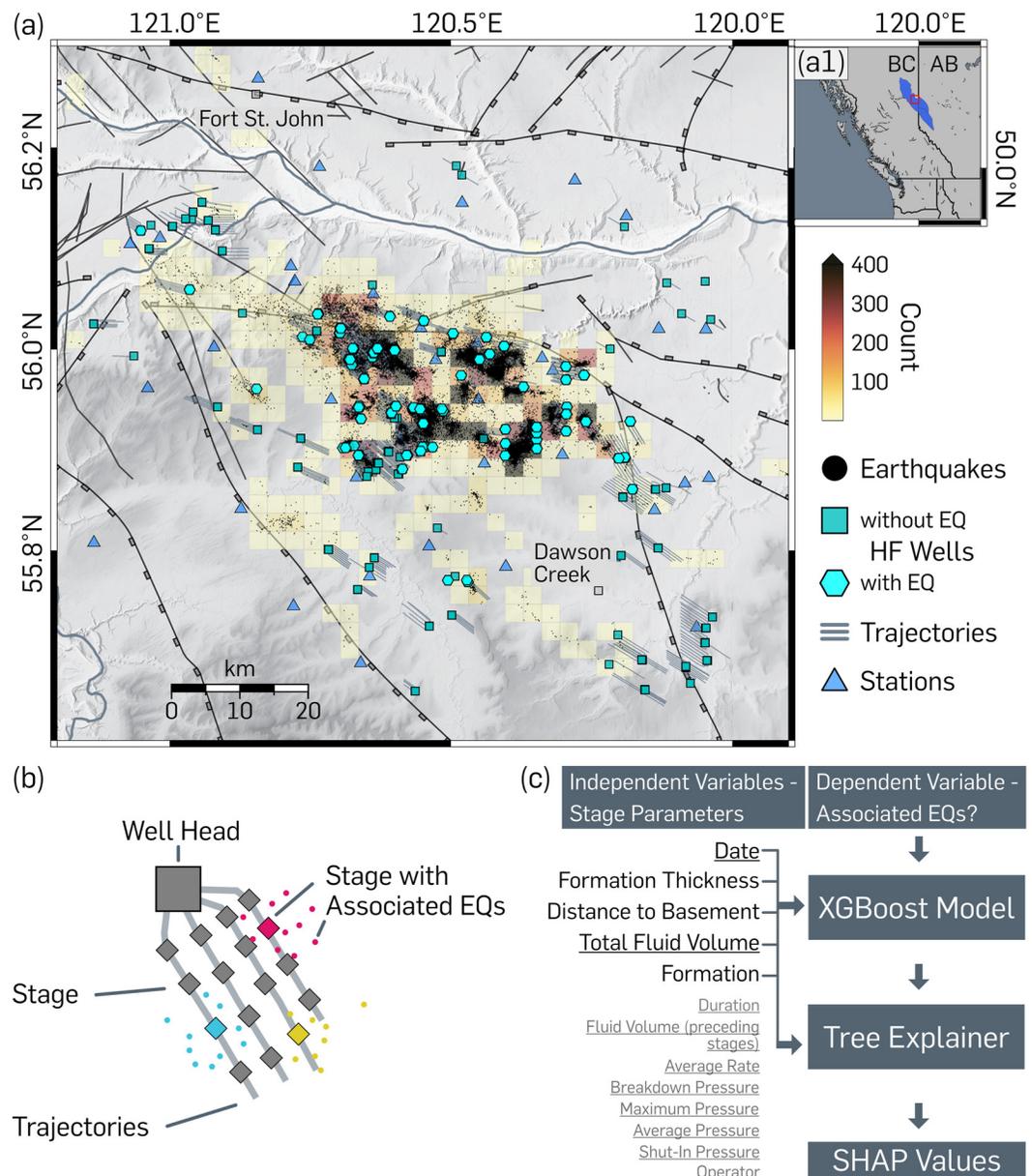


Figure 1. (a) Kiskatinaw study area, northeastern British Columbia (BC). Red box in the overview map (a1) shows the study area within the Montney Formation (BC and Alberta) in blue, as reported by BC Oil and Gas Commission. Black points denote earthquakes in the enhanced catalog, with colorbar indicating the number of earthquakes in each 2×2 km grid. Hydraulic fracturing (HF) wells with/without associated earthquakes (hexagons/squares) with respective well trajectories (gray lines) demonstrate a variable seismic response that is resolved by the dense seismic station coverage (blue triangles). Mapped faults (barbed lines) follow Norgard (1997); Berger et al. (2009); Cui et al. (2017) and Davies et al. (2018). (b) Schematic HF well (gray square - well head, gray lines - well trajectories, diamonds - HF stages) with associated seismicity as points color-coded according to their respective associated stages. (c) Supervised machine learning approach. Dark rectangles at the top denote the model's independent (input) and the dependent (output) variables. Left column shows stage parameters used as independent variables, with parameters of negligible importance shown in lighter gray (see main text and Supporting Information for details). Underlined parameters denote operational (in contrast to geological) parameters. The XGBoost (Chen & Guestrin, 2016) model is trained as described in the main text and the tree-explainer (Lundberg et al., 2020) calculates the SHapley Additive exPlanations values.

independent variable to reduce the likelihood of overfitting the data with excessive numbers of independent variables (Breiman, 2001, Figure 1c, Figure S5 in Supporting Information S1). We use the gradient boosted tree classifier XGBoost (Chen & Guestrin, 2016) to increase accuracy and support missing values present in the

well reports (Figure S6 in Supporting Information S1). A test of other gradient boosted tree classifiers or neural networks demonstrated no significant increase in accuracy. We use the optimization framework Optuna (Akiba et al., 2019) to optimize the hyperparameters of XGBoost. We measure the performance of the model using the Area under the Receiver Operating Characteristic Curve (ROC-AUC) score, which is a good measure of separability (Fawcett, 2006). Text S4 in Supporting Information S1 provides extensive details regarding the machine learning workflow and ROC-AUC score.

We use SHapley Additive exPlanations (SHAP), Lundberg et al., 2020) to explain the model and its predictions. For each stage, we calculate a SHAP value as a measure of the contribution of each single parameter (input) to the output of the model, where the sum of the total resulting SHAP values from all parameters indicates the model prediction. A positive or negative SHAP sum corresponds to a stage being likely or unlikely to induce an earthquake, respectively. We also compare SHAP values of various HF parameters over all stages to interpret their relative importance as well as whether they are positive or negative (i.e., increase or decrease the likelihood of inducing an earthquake).

3. Results

Earthquake catalog enhancement adds an additional 21,404 events using MMF and 10,260 events using EQTransformer to the initial catalog of 8,382 events (Roth et al., 2020). The following discussion considers additional detected and relocated events with a maximum horizontal error of 10 km (median horizontal error: 0.4 km). Catalog enhancement reduces the magnitude completeness M_c from ~ 1.6 to ~ 0.2 (Figure S3 in Supporting Information S1). We infer the observed magnitude-frequency distribution likely results from a summation of multiple earthquake clusters related to individual HF wells with variable M_c and b-values. However, a more detailed investigation of clusters would be required to draw any conclusions. We make no further interpretations of M_c and the b-value here, as it is not the focus of this work.

A total of 39,156/40,046 ($\sim 98\%$) enhanced catalog events associate with an individual stage, and 4,502/12,903 stages ($\sim 35\%$) have at least one associated earthquake. Seismicity occurs predominantly in swarm-like clusters, as noted in Section 1 (Figure S3b in Supporting Information S1), where 82% of associated events occur within 3 hr of the start a specific HF stage. Delayed earthquakes are comparatively few (16%), and 83% of all delayed events occur within 3 hr of the end of a stage (Figure S4 in Supporting Information S1). Considering events exclusively within a specific treatment period reduces the number of stages with associated seismicity by $\sim 8\%$ to 4,163 but does not change the results of the machine learning models. The reduction of the initial 13 parameters to five in the independent variable (Figure 1c) using a random forest classifier (Section 2, Figure S5, Text S4 in Supporting Information S1) does not decrease the accuracy of the model. The fitted XGBoost model ROC-AUC score of 0.89 for both the validation and test sets suggests the model robustly generalizes the relation between stage parameters and earthquake occurrence with negligible overfitting (Figure S6 in Supporting Information S1).

The SHAP-value distribution in Figure 2a for individual stage parameters shows all predictions calculated using all stages and the trained XGBoost model. Figure 2a1 shows the mean absolute SHAP values for all stages separated by their target formation. Both the individual and mean values demonstrate the importance of formation thickness, hydraulic-fracturing date, fluid volume, distance to the basement, and target formation. The following discussion describes the dependence of SHAP values with respect to the four most important stage parameters (formation thickness, date, fluid volume, and distance to basement).

Figure 2b shows the dependence of both individual stage and mean SHAP values on the formation thickness, where undifferentiated sediments lie stratigraphically above the Upper Montney Formation with no available thickness information. The correlation ranges from negative to small positive values for thicknesses up to 150 m located primarily in the Upper Montney. Formation thickness correlates positively in stages in the Lower Montney and Upper Montney, and reaches a maximum in the Lower Montney at a thickness of ~ 200 m. The correlation decreases rapidly at thicknesses ≥ 225 m for the Lower Montney. Figure 2a1 emphasizes the importance of formation thickness in the Lower Montney.

The HF-treatment date also correlates positively with the SHAP values for all formations up to the COVID-19 lockdown (Figure 2c). The last completed stage in the database before the lockdown was on 2020-03-29 and the first stage completed after the lockdown occurred on 2020-08-06. All formations correlated negatively with SHAP values following the suspended operational period, particularly the Lower Montney. The SHAP values

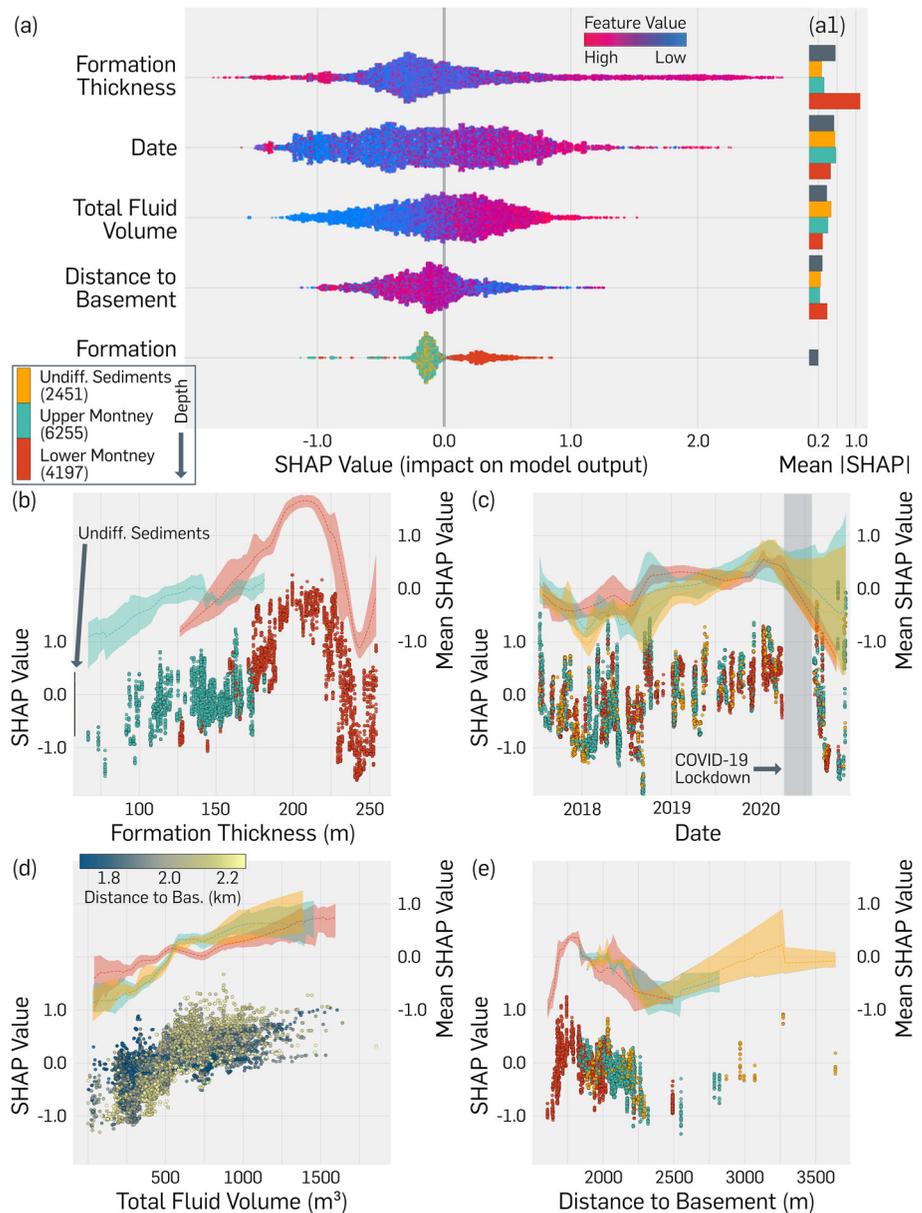


Figure 2. Overview of SHapley Additive exPlanations (SHAP) values from all hydraulic fracturing (HF) wells and the XGBoost model. (a) SHAP summary plot for parameters where each point is related to the prediction from the XGBoost model for an individual stage. Point color reflects respective feature value in the colorbar for all parameters except formation (formation color shown by the colors in the lower left). (a1) Mean absolute SHAP value for each parameter using all stages (dark gray), and stages in each formation. Table S1 in Supporting Information S1 provides all SHAP values. (b) Formation thickness dependence with points colored by the same formation colors as in (a). Note: stages in the undifferentiated sediments have no calculated thickness. Colored bars above the points from a locally weighted scatterplot smoothing (LOWESS) for stages in the same formation (color as in (a)). LOWESS lines correspond to the y-axis on the right. The 2σ confidence intervals are calculated from 100 random samples. (c) Dependence plot for completion date with colors and symbols as in (b). The gray rectangle outlines the COVID-19 lockdown period. (d) Dependence plot for the total injected fluid volume with colors and symbols as in (b), but with distance to the Precambrian basement shown by the respective colorbar. (e) Dependence plot for stage distance to the Precambrian basement with colors and symbols as in (b).

exhibit a more variable range following the lockdown, where fewer HF stages were completed comparatively to the period before. We further discuss the influence of the injection in Section 4.

The total injected fluid volume per stage shows a distinct positive correlation with the SHAP values, with positive SHAP values for volumes $\geq 500 \text{ m}^3$ (Figure 2d). SHAP values of the Upper Montney and undifferentiated

sediments show stronger dependence on the injected fluid volume compared to the Lower Montney. They are also negative for volumes $\sim \leq 500 \text{ m}^3$ and positive for larger volumes, showing a slightly positive correlation with fluid volume. SHAP values for the Lower Montney are roughly zero with a slight positive correlation at higher volumes, and are positive for volumes $\sim \geq 1,000 \text{ m}^3$. Considering both the distance to the basement together with injection volume also highlights variable behavior (colorbar, Figure 2d). For example, the increase of SHAP values at 500 m^3 is more pronounced for stages located $\geq 2,000 \text{ m}$ from the basement.

The stage distance to the basement negatively correlates with the SHAP values for distances $\geq 1,700 \text{ m}$, and SHAP values are primarily negative for distances $\sim \geq 2,000 \text{ m}$ (Figure 2e). The SHAP values are highest at distances of $\sim 1,700 \text{ m}$ in the Lower Montney, and decrease sharply with decreasing distance.

Figure 3a summarizes the correlation of the SHAP-value sums with the maximum observed magnitude (MOM) for all stages, as well as the mean formation SHAP value as a function of magnitude. Panel (b) shows the composition of the SHAP sum for seven representative stages in (a), including four stages (1–4) with an associated earthquake, and three (5–7) without. Two (1–2) of the four stages with an associated earthquake (also the largest magnitude earthquakes in our study area) are deemed likely to occur according to the model (“true positives”), while two (3–4) are deemed unlikely to occur (“false negatives”). The three stages without an associated earthquake include one stage (5) where the model predicts no earthquake (“true negative”) and two stages (6,7) where the model predicts an earthquake (“false positives”). SHAP sums show no clear correlation with magnitude, but the sum is positive for stages with a magnitude ≥ 0.5 for all formations. In addition, the locally-weighted scatterplot smoothing curves show the highest predictability for the Lower Montney (Figure 3a). In contrast, the number of false positives is also higher in the Lower Montney (see boxplots in Figure 3a).

4. Discussion

This study employs a machine-learning model to quantify the dependence of seismic propensity on detailed HF-stage parameters in the Kiskatinaw area using a significantly enhanced earthquake catalog from multi-year continuous monitoring. We find that geological parameters (formation thickness and vertical distance to basement) and operational parameters (HF treatment date and total injection volume) are key parameters for generating seismogenic HF stages. The remaining operational parameters tested appear to be of negligible importance. Furthermore, we find that stages targeting the Lower Montney formation are more likely to induce earthquakes than stages targeting shallower formations (Upper Montney, undifferentiated sediments). We emphasize that feature importance does not entail a causal relationship with predictive value. Instead, the model and subsequent interpretations are intended to provide a starting point for further analyses that put operational parameters in geological context.

Our findings largely agree with similar studies focused on HF wells targeting the Montney Play. Wozniakowska and Eaton (2020) estimated the Seismogenic Activation Potential based on 6,466 HF wells (rather than specific stages) using logistical regression on published earthquake catalogs between 2006 and 2020 in Kiskatinaw and the North Peace Ground Motion Monitoring Area (NPGMMA), north-west from Kiskatinaw. They established injection depth and distance to the Cordilleran thrust belt as key parameters in their model. The importance of injection depth is also observed in Oklahoma (e.g., Hincks et al., 2018). The local focus on Kiskatinaw in this study implies that distances to the thrust belt are all similar for all HF stages considered, and are therefore not evaluated here. While injection depth is also essential here, we mapped the injection depth to specific digital geological models, as the geological units exhibit a significant NE-SW oriented dipping trend across Kiskatinaw (Figure S7 in Supporting Information S1). Amini and Eberhardt (2021) employed a similar approach using tree-based machine learning models (as XGBoost, among others) with detailed information for 95,807 HF stages in the NPGMMA and Kiskatinaw with a combination of published catalogs containing 9,949 events, similar to Wozniakowska and Eaton (2020). They found that geological parameters are more important than operational parameters, similar to our results. In particular, both studies found a negative correlation with the distance to the basement, as demonstrated in Figure 2e, indicating that injection depths closer to the crystalline basement are more likely to induce earthquakes. Wang et al. (2022) follow a similar approach to this one (XGBoost model with SHAP values) while focusing on the Northern Montney Play (NMP), which includes the NPGMMA. They found that the distance to the Cordilleran thrust belt (similar to Wozniakowska and Eaton (2020)), and the total injected fluid volume (similar to this study and to Amini et al. (2021)) are most important. They also found that treating pressures and the vertical distance to the Debolt formation are key parameters. However, our results suggest that

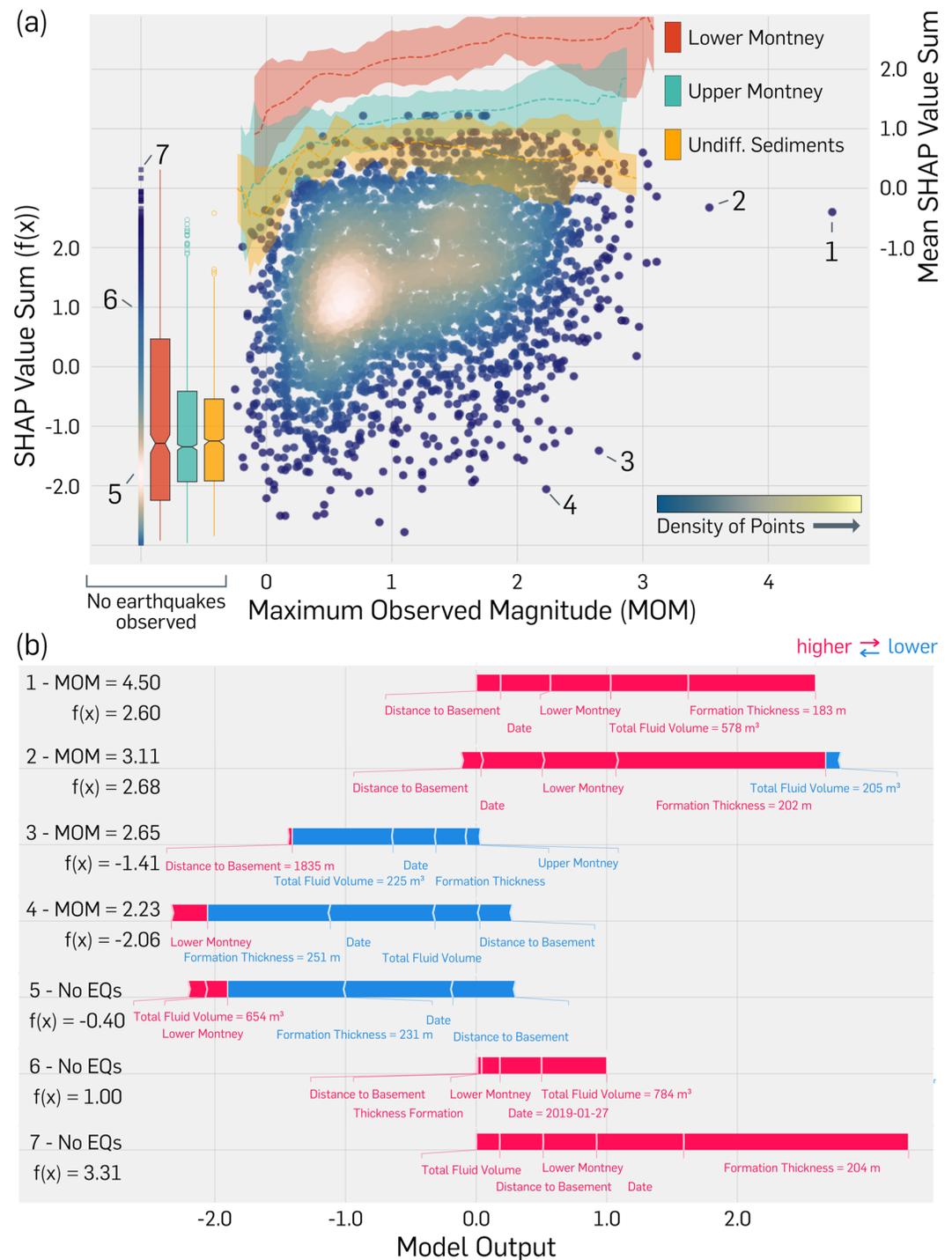


Figure 3. SHAPley Additive exPlanations (SHAP) value sum for all HF stages and the XGBoost model. (a) SHAP value sum versus maximum observed magnitude (MOM) from the associated earthquakes or if no earthquake is detected (i.e., stages with no associated earthquakes are contained in the rectangle and forming a line over “no earthquakes observed”). The colorbar shows the point density for each subset of points. The lines colored by their respective formation result from a locally weighted scatterplot smoothing (LOWESS) over the MOM based on stages in each formation. The LOWESS lines correspond to the y-axis on the right with 2σ confidence intervals from 100 random samples. The colored boxplots on the left are calculated from stages with no observed earthquakes in each formation with the same colors used for the LOWESS lines. The notch of each boxplot represents the 2σ confidence interval from 10,000 random bootstrap samples. (b) The composition of the SHAP sum value for seven representative stages (marked by numbers 1–7 in (a)). Red and blue colors correspond to a positive and negative value for the specific parameter. Table S2 in Supporting Information S1 provides all parameter values.

treating pressure bears no significant importance relative to other parameters, including the vertical distance to the Precambrian basement. The differences found here might highlight a variable seismic response to operations in the NMP relative to Kiskatinaw in the south. In contrast to the studies by Wozniakowska and Eaton (2020) and Amini and Eberhardt (2021), we do not use the maximum horizontal compression orientation, SHmax, due to a lack of resolution in the small study region. (Published values in Kiskatinaw are based limited observations of ~2 wells (Bell & Grasby, 2012; Heidbach et al., 2018; Shen et al., 2019)). A recent induced earthquake focal mechanism study in Kiskatinaw (Roth et al., 2022) found predominantly left-lateral strike-slip focal mechanisms with nodal planes at shallow angles to the regional SHmax, suggesting low spatial variability in SHmax in our study area. We note that SHmax is a parameter worth investigating in cases where observations exist at sufficient spatial resolution (e.g., from borehole measurements) near injection wells and/or seismicity. The abundance of strike-slip faulting mechanisms within the local study area also makes it difficult to use structural elements with potentially vertical throw (e.g., Wozniakowska et al., 2021) to identify correlations.

In contrast to previous models noted above, our model includes additional parameters. It suggests thickness of the target formation (for both Lower and Upper Montney) bears significant importance for the likelihood of inducing earthquakes (Figure 2b), particularly in the Lower Montney (Figure 2a1)). The overall correlation is positive, meaning a thicker target formation is more likely to host an induced earthquake; the correlation becomes negative when formation thickness increases to values ≥ 200 m. The positive correlation is also consistent with the theoretical model of Galis et al. (2017), which assumes that most induced earthquakes are self-arrested ruptures caused by local pore-pressure perturbations propagating on pre-stressed faults. One notable aspect is that the model takes the area of intersection between the reservoir and a pre-existing fault into account. The positive observed correlation with the formation thickness is consistent with an increase in the probability of having a potential pre-existing fault intersect a thicker reservoir or target formation while also increasing the size of an intersection. The more likely intersection with a fault should subsequently increase the likelihood of inducing an earthquake. We discuss below why the reverse trend in thicknesses ≥ 200 m does not fit the above interpretation, including what other factors could influence the observed negative correlation.

The HF treatment date also exhibits an important role. For example, stages completed in late 2019 to early 2020 are more likely to induce an earthquake compared with stages in 2018 (Figure 2c). The likelihood reverses after the seismically quiescent COVID-19 lockdown period (e.g., Salvage & Eaton, 2021). The positive correlation of increasing likelihood with time could point to a potential cumulative effect of injection, leading to increasing pore pressure, poroelastic, and/or Coulomb static stress changes (Brown & Ge, 2018). The lockdown likely relaxed stress perturbations to some degree, decreasing the probability of inducing earthquakes afterward. Oklahoma and Kansas also exhibit similar time-dependent effects where forecasted seismicity rates decrease with injection rates (Langenbruch et al., 2018). However, the time-dependent effects in the central US are related to wastewater-disposal-induced earthquakes attributed to pore-pressure perturbations over km length scales (e.g., Peterie et al., 2018). In contrast, perturbations for HF-induced earthquakes are likely more localized due to lower permeability, and may drive stress changes by both poroelastic and pore-pressure diffusion (e.g., Deng et al., 2016). For example, temporal poroelastic stress changes generate cumulative static Coulomb stress changes, which increase the likelihood of an induced earthquake by subsequent HF stages. We note that the interpretation related to any cumulative effect is speculative, albeit consistent with the seismicity rates following the lockdown. Confirming the interpretation would require testing other factors, such as the dependence on changes in stimulation strategies or other geological parameters. The lack of additional information in this data set prohibits further testing here. However, it does show the merit of testing the validity of cumulative effects on data sets using a longer time window where HF operations are sustained over longer periods (Figure S3a in Supporting Information S1). The time-dependent observation highlights the potential benefit of the machine learning approach to identify essential injection parameters that should be studied in greater detail to identify fault activation processes.

The positive correlation with total injected volume and negative correlation with the basement distance has also been observed using similar approaches (Amini & Eberhardt, 2021; Wang et al., 2022; Wozniakowska & Eaton, 2020). Studies also correlate the volume to maximum magnitude (e.g., Galis et al., 2017; McGarr, 2014) and seismicity rate (e.g., Schultz et al., 2020). Figure 2d further highlights the detailed and variable response of the target formation, where correlation is positive for all formations. Where fluid volumes $\sim \leq 500$ m³ appear to be less important for the shallower formations (Lower Montney, undifferentiated sediments), they increase

drastically for volumes $\sim \geq 500 \text{ m}^3$. The relationship could imply a minimum fluid volume is needed to reach deeper layers, presumably with more pre-existing faults. Once a minimum fluid volume is exceeded, it becomes important for estimating the likelihood of inducing an earthquake. Another explanation could be related to M_c . For example, Schultz et al. (2018) show that a minimum fluid volume is necessary to increase probability of causing an event above M_c . The observed increased likelihood associated with larger fluid volumes in shallower layers combined with larger basement distances (colorbar in Figures 2d and 2e) suggests the necessary combination of conditions to generate earthquake inducing stress perturbations in or near the basement. Similar dependences on the distance to the basement have also been observed for HF-induced earthquakes in the United States (Skoumal et al., 2018). Again, the combined importance of fluid volume and distance to the basement needs to be studied in more detail in future work, and goes beyond the scope of this paper.

One potential shortcoming of machine learning models is that the selection and availability of parameters may fail to represent the truth correctly or may introduce non-negligible bias. For example, the inverse correlation of the SHAP values with formation thicknesses $\geq 200 \text{ m}^3$ (Figure 2b) contradicts the theoretical model of Galis et al. (2017), which could suggest that the spatio-temporal distribution of stages biases the data. A possible explanation is that the observed trend of decreasing SHAP value with increasing formation thickness coincidentally reflects a higher number of HF stages targeting thicker formations completed after the COVID-19 lockdown. The temporal coincidence could result in a decreased likelihood of an induced earthquake after the lockdown period that overshadows the likelihood associated with thicker formations (Figure 2c). However, we do not observe a clear spatio-temporal distribution supporting the above hypothesis, as stage completions occur at arbitrary times in any spatial region (Figures S7, S8 in Supporting Information S1). In contrast, stages in the eastern part of the study area show significantly lower SHAP sum values compared to the center (Figure S9 in Supporting Information S1). The lower eastern SHAP values also correspond to higher formation thickness of the Lower Montney and lower distances to the basement (Figure S7 in Supporting Information S1). The spatial distribution of stages could explain the observed inverse correlation with formation thickness, and it is possible that other stage parameters that were not included in the model change in the same area. We infer from the distribution of mapped faults that the parameter changes might be related to changing geological factors, such as proximity to pre-existing faults (Figure S9 in Supporting Information S1). Another potential bias could arise from the possibility that HF operations target thicker formations less frequently, resulting in less pronounced cumulative effects that can not be ruled out with the limited time window of our data set. Furthermore, our data set consists exclusively of stages that are *a priori* associated with seismicity. The prohibitive number of PDF files that would need to be processed manually to include non-seismogenic stages make the latter potential bias inevitable (see Section 2).

The unavailability of other potentially important parameters is another potential source of bias. As discussed above, the local SHmax orientation is likely influential, but lacks sufficient spatial accuracy in the study area. Another influential (unavailable) indicator could be lithium concentration in formation waters, which could be used as a proxy for vertical, fault-related fluid flow (Pawley et al., 2018). Formation overpressure has also been found to correlate with induced seismicity, but is also unavailable here (e.g., Eaton & Schultz, 2018). Many potentially influential parameters could explore similar false positive and false negative cases shown as in Figure 3 to identify important stage parameters to be included in future models. In addition, false predictions offer the potential to identify methodological limitations in cases where more detailed information about the interaction between injection and seismicity is available. An additional potential bias might be unbalanced variable distributions, for example, the abundance of events, or lack thereof, within a formation unit. Since hypocentral resolution of our enhanced catalog does not permit association to formation units, we are unable to quantify the effect with this data set, but it demonstrates an avenue of future use.

We note that we do not calculate seismic risk based on our machine learning models or attempt to predict the MOM, as done by previous studies. The fitting error using a regression model corresponds to $\sim \geq 0.5$ and includes significant outliers, as the larger magnitude range is not quantitatively well constrained (where 8 stages events have $M_{Lmax} \geq 3$ and 544 stages with $M_{Lmax} \geq 2$). Therefore, the time frame considered will not provide a regression model for a robust risk assessment. The MOM in Kiskatinaw may be more strongly influenced by the local distribution of pre-existing faults or fluid conduits (Peña Castro et al., 2020), which such models are unable to take into account. Such limitations would also apply to using the total number of earthquakes or total moment release as the prediction target. In addition, the observed varying M_c of different seismicity clusters would further bias the input data.

5. Conclusion

We investigate the relationship between an enhanced catalog of 40,046 earthquakes and detailed injection parameters for 12,903 HF stages using a supervised machine learning approach in the Kiskatinaw area, British Columbia, Canada. We use the gradient boosted decision tree classifier XGBoost (ROCAUC score 0.89) combined with the tree-explainer SHAP to assess the importance of specific injection parameters on the likelihood of an HF stage to induce a detectable earthquake. We find that:

1. Geological parameters show higher importance than operational parameters.
2. The thickness of the target formation, the completion date, the total injected fluid volume, the vertical distance to the Precambrian basement, and the target formation are the most influential parameters in order of relative importance.
3. Stages in the Lower Montney Formation have a higher likelihood of inducing an earthquake compared to the Upper Montney or shallower undifferentiated sediments.
4. Thicker target formations are generally more likely to host a seismogenic stage.
5. The increasing probability of HF stages over time may suggest the significance of cumulative effects of pore pressure, poroelastic, and/or Coulomb static stress changes in areas with HF. Cumulative effects seem to partially abate in Kiskatinaw during the COVID-19 lockdown in 2020.
6. Larger injected fluid volumes are more likely to induce an earthquake. Stages in shallower layers (Lower Montney, undifferentiated sediments) show a significant increase in importance for volumes $\sim \geq 500 \text{ m}^3$. The latter observation could point to a minimum fluid volume needed to reach deeper layers with more pre-existing faults.
7. Stages closer to the Precambrian basement are generally more likely to host a seismogenic stage.
8. The accuracy of our model is highest in the Lower Montney compared to the Upper Montney and undifferentiated sediments, whereas the false positive rate is also higher in the Lower Montney.

This study highlights a fundamental application of machine learning models toward induced seismicity: identifying essential parameters to be studied in greater detail in efforts to reduce the overall risk of industrial operations. Machine learning models also have a general potential to provide new information about the mechanisms driving induced seismicity. Although these models do not consider the geological context, they exhibit a potential to guide detailed studies focusing on the most relevant parameters. Furthermore, false-positive and false-negative predictions could be used in detailed studies of specific stages to provide further insights on parameters that might drive earthquake generation.

We emphasize that the open-access availability of HF stage information through BCOGC enabled this study. Open data availability is crucial to enabling scientific study to decipher the correlation between induced seismicity and HF (or other alternative energy production) operations, as the risk associated with induced seismicity is of great public concern. Publicly available data, as well as dense regional seismic networks to study potential relationships using small earthquakes will help, in ideal cases, to identify the processes leading to (and avoiding) damaging earthquakes. All the above steps will only help increase public acceptance and tolerance of the manageable hazard and risks associated with alternative energy production.

Data Availability Statement

Waveform data used in this study are archived at IRIS under network codes XL, 1E, PQ, and EO (e.g., <https://ds.iris.edu/gmap/XL>). Well data are provided by British Columbia Oil and Gas Commission (BCOGC; <https://files.bco.gc.ca/thinclient/>, last accessed May 2022). Topographic information in Figure 1 comes from the Canadian Digital Elevation Model (CDEM) <https://open.canada.ca/data/en/dataset/7f245e4d-76c2-4caa-951a-45d1d2051333> (last accessed May 2022). The open-source Python package Obspy 1.2.2 was used for data processing (Krischer et al., 2015) available at: <https://github.com/obspy/obspy> (last accessed May 2022). The multi-station-matched filter detection uses the Python package EQCorrscan 0.4.1 (Chamberlain et al., 2017), available at: <https://github.com/eqcorrscan/EQcorrscan> (last accessed March 2022). We used EQTransformer 0.1.59 (Mousavi et al., 2020, <https://github.com/smousavi05/EQTransformer>, last accessed May 2022) to generate phase arrivals and REAL 1.0 for phase association (Zhang et al., 2019, <https://github.com/Dal-mzhang/REAL>, last accessed May 2022). During the machine learning workflow we used several functions from scikit-learn 1.0.2 (Pedregosa et al., 2011, <https://scikit-learn.org/stable/>, last accessed May 2022) and fastai 2.5.3

(Howard & Gugger, 2020, <https://docs.fast.ai>, last accessed May 2022). We further used XGBoost 1.5.1 (Chen & Guestrin, 2016, <https://xgboost.readthedocs.io/en/stable/>, last accessed May 2022) and SHAP 0.40.0 (Lundberg et al., 2020, <https://github.com/slundberg/shap>, last accessed May 2022). Several figures were created using GMT 6.3.0 (Wessel et al., 2019), available at: <https://docs.generic-mapping-tools.org/latest/> (last accessed May 2022). The other figures were created using Matplotlib 3.5.1 (Hunter, 2007) available at: <https://matplotlib.org/> (last accessed May 2022). Many colormaps from Scientific Color Maps 7 were used during plotting, available at: <https://doi.org/10.5281/zenodo.5501399> (last accessed May 2022).

Acknowledgments

We would like to thank Honn Kao, Bei Wang, and Hongyu Yu for discussion and constructive comments during the preparation of this manuscript. We thank Editor G. Prieto, Ryan Schultz, and three anonymous reviewers for their comments. This project is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) project number 428868223, and Natural Sciences and Engineering Research Council of Canada (NSERC) Strategic Partnership Grant for Projects 494141-2016.

References

- Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining* (pp. 2623–2631). Association for Computing Machinery. <https://doi.org/10.1145/3292500.3330701>
- Amini, A., & Eberhardt, E. (2021). *Machine-learning analysis of factors influencing induced seismicity susceptibility in the Montney play area*. Geoscience BC, Summary of Activities 2020 (Report 2021-02), 45–56. Retrieved from https://www.geosciencebc.com/i/pdf/SummaryofActivities2020/EW/Project%202019-014_EWSOA2020.pdf
- Amini, A., Mehrabifard, A., & Eberhardt, E. (2021). *Development of an induced seismicity susceptibility framework and map for NEBC using an integrated machine learning and mechanistic validation approach*. Geoscience BC, Geoscience BC Project 2019-014 (Report 2021-11). Retrieved from https://www.geosciencebc.com/i/project_data/GBCReport2021-11/Geoscience%20BC%20Report%202021-11.pdf
- Babaie Mahani, A., 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. <https://doi.org/10.1785/0120160175>
- BCOGC. (2012). Montney formation play atlas NEBC (Tech. Rep.). BC oil and gas commission. Retrieved from <https://www.bco.gc.ca/files/reports/Technical-Reports/montney-format-play-atlas-nebc.pdf>
- BCOGC. (2018). Kiskatinaw seismic monitoring and mitigation area special project order (Tech. Rep.). BC oil and gas commission. Retrieved from <https://www.bco.gc.ca/files/reports/Technical-Reports/Kiskatinaw-Seismic-Monitoring-and-Mitigation-Area-KSMMA-Special-Project-Order-18-90-001-Amendment-1.pdf>
- BCOGC. (2022). BC oil and gas commission - eLibrary. Retrieved from <https://www.bco.gc.ca/energy-professionals/online-systems/elibrary/>
- Bell, J. S., & Grasby, S. E. (2012). The stress regime of the Western Canadian Sedimentary Basin. *Geofluids*, 12(2), 150–165. <https://doi.org/10.1111/j.1468-8123.2011.00349.x>
- Berger, Z., Mushayandevu, M., & Boast, M. (2009). *The contribution of HRAM data to exploration in Canada: An overview* (Vol. 36, pp. 40–45). Canadian Society of Petroleum Geologists, Reservoir.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Brown, M. R. M., & Ge, S. (2018). Small earthquakes matter in injection-induced seismicity. *Geophysical Research Letters*, 45(11), 5445–5453. <https://doi.org/10.1029/2018GL077472>
- Chamberlain, C. J., Hopp, C. J., Boese, C. M., Warren-Smith, E., Chambers, D., Chu, S. X., et al. (2017). EQcorrscan: Repeating and near-repeating earthquake detection and analysis in Python. *Seismological Research Letters*, 89(1), 173–181. <https://doi.org/10.1785/0220170151>
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 785–794). Association for Computing Machinery. <https://doi.org/10.1145/2939672.2939785>
- Cui, Y., Miller, D., Schiarizza, P., & Diakow, L. J. (2017). *British Columbia digital geology*. British Columbia Ministry of Energy, Mines and Petroleum Resources, British Columbia Geological Survey Open File 2017-8, 9p. Data version 2019-12-19. Retrieved from <https://www2.gov.bc.ca/gov/content/industry/mineral-exploration-mining/british-columbia-geological-survey/geology/bcdigitalgeology>
- Davies, G. (1997). The Triassic of the Western Canada Sedimentary Basin: Tectonic and stratigraphic framework, paleogeography, paleoclimate and biota. *Bulletin of Canadian Petroleum Geology*, 45(4), 434–460. <https://doi.org/10.35767/gscpgbull.45.4.434>
- Davies, G., Watson, N., Moslow, T., & Maceachern, J. (2018). Regional subdivisions, sequences, correlations and facies relationships of the Lower Triassic Montney Formation, West-Central Alberta to northeastern British Columbia, Canada — With emphasis on role of paleostructure. *Bulletin of Canadian Petroleum Geology*, 66, 23–92.
- 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>
- Eaton, D. W., & Schultz, R. (2018). Increased likelihood of induced seismicity in highly overpressured shale formations. *Geophysical Journal International*, 214(1), 751–757. <https://doi.org/10.1093/gji/ggy167>
- Edwards, D. E., Barclay, J., Gibson, D., Kville, G., & Halton, E. (1994). Triassic strata of the Western Canada Sedimentary Basin. In G. D. Mossop, & I. Shetsen (Eds.), *Geological atlas of the western Canada Sedimentary Basin* (pp. 259–275). Canadian Society of Petroleum Geologists and Alberta Research Council. Retrieved from https://static.ags.aer.ca/files/document/Atlas/chapter_16.pdf
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861–874. <https://doi.org/10.1016/j.patrec.2005.10.010>
- Galis, M., Ampuero, J. P., Mai, P. M., & Cappa, F. (2017). Induced seismicity provides insight into why earthquake ruptures stop. *Science Advances*, 3(12), eaap7528. <https://doi.org/10.1126/sciadv.aap7528>
- Ghofrani, H., & Atkinson, G. M. (2020). Activation rate of seismicity for hydraulic fracture wells in the Western Canada Sedimentary Basin. *Bulletin of the Seismological Society of America*, 110(5), 2252–2271. <https://doi.org/10.1785/0120200002>
- Heidbach, O., Rajabi, M., Cui, X., Fuchs, K., Müller, B., Reinecker, J., et al. (2018). The World Stress Map database release 2016: Crustal stress pattern across scales. *Tectonophysics*, 744, 484–498. <https://doi.org/10.1016/j.tecto.2018.07.007>
- Hincks, T., Aspinall, W., Cooke, R., & Gernon, T. (2018). Oklahoma's induced seismicity strongly linked to wastewater injection depth. *Science*, 359(6381), 1251–1255. <https://doi.org/10.1126/science.aap7911>
- Howard, J., & Gugger, S. (2020). Fastai: A layered API for deep learning. *Information*, 11(2), 108. <https://doi.org/10.3390/info11020108>
- Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3), 90–95. <https://doi.org/10.1109/MCSE.2007.55>
- Krischer, L., Megies, T., Barsch, R., Beyreuther, M., Lecocq, T., Caudron, C., & Wassermann, J. (2015). ObsPy: A bridge for seismology into the scientific Python ecosystem. *Computational Science & Discovery*, 8(1), 014003. <https://doi.org/10.1088/1749-4699/8/1/014003>

- Langenbruch, C., Weingarten, M., & Zoback, M. D. (2018). Physics-based forecasting of man-made earthquake hazards in Oklahoma and Kansas. *Nature Communications*, 9(1), 3946. <https://doi.org/10.1038/s41467-018-06167-4>
- Lei, X., Wang, Z., & Su, J. (2019). The December 2018 ML 5.7 and January 2019 ML 5.3 earthquakes in South Sichuan Basin induced by shale gas hydraulic fracturing. *Seismological Research Letters*, 90(3), 1099–1110. <https://doi.org/10.1785/0220190029>
- Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., et al. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, 2(1), 56–67. <https://doi.org/10.1038/s42256-019-0138-9>
- McGarr, A. (2014). Maximum magnitude earthquakes induced by fluid injection: Limits on fluid injection earthquakes. *Journal of Geophysical Research: Solid Earth*, 119(2), 1008–1019. <https://doi.org/10.1002/2013JB010597>
- 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), 3952. <https://doi.org/10.1038/s41467-020-17591-w>
- Norgard, G. T. (1997). Structural inversion of the Middle Triassic Halfway Formation, Monias field, northeast British Columbia. *Bulletin of Canadian Petroleum Geology*, 45(4), 614–623. <https://doi.org/10.35767/GSCPGBULL.45.4.614>
- Pawley, S., Schultz, R., Playter, T., Corlett, H., Shipman, T., Lyster, S., & Hauck, T. (2018). The geological susceptibility of induced earthquakes in the Duvernay play. *Geophysical Research Letters*, 45(4), 1786–1793. <https://doi.org/10.1002/2017GL076100>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12(85), 2825–2830. Retrieved from <http://jmlr.org/papers/v12/pedregosa11a.html>
- Peña Castro, A. F., Roth, M. P., Verdecchia, A., Onwuewemeka, J., Liu, Y., Harrington, R. M., et al. (2020). Stress chatter via fluid flow and fault slip in a hydraulic fracturing-induced earthquake sequence in the Montney formation, British Columbia. *Geophysical Research Letters*, 47(14), e2020GL087254. <https://doi.org/10.1029/2020GL087254>
- Peterie, S. L., Miller, R. D., Intfen, J. W., & Gonzales, J. B. (2018). Earthquakes in Kansas induced by extremely far-field pressure diffusion. *Geophysical Research Letters*, 45(3), 1395–1401. <https://doi.org/10.1002/2017GL076334>
- Roth, M. P., Kemna, K. B., Harrington, R. M., & Liu, Y. (2022). Source properties of hydraulic-fracturing-induced earthquakes in the Kiskatinaw area, British Columbia, Canada. *Journal of Geophysical Research: Solid Earth*, 127(3), e2021JB022750. <https://doi.org/10.1029/2021JB022750>
- 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. <https://doi.org/10.1785/0220200086>
- Salvage, R. O., & Eaton, D. W. (2021). Unprecedented quiescence in resource development area allows detection of long-lived latent seismicity. *Solid Earth*, 12(3), 765–783. <https://doi.org/10.5194/se-12-765-2021>
- 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.aaa0159>
- Schultz, R., Skoumal, R. J., Brudzinski, M. R., Eaton, D., Baptie, B., & Ellsworth, W. (2020). Hydraulic fracturing-induced seismicity. *Reviews of Geophysics*, 58(3), e2019RG000695. <https://doi.org/10.1029/2019RG000695>
- 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>
- Shen, L. W., Schmitt, D. R., & Haug, K. (2019). Quantitative constraints to the complete state of stress from the combined borehole and focal mechanism inversions: Fox Creek, Alberta. *Tectonophysics*, 764, 110–123. <https://doi.org/10.1016/j.tecto.2019.04.023>
- Skoumal, R. J., Barbour, A. J., Brudzinski, M. R., Langenkamp, T., & Kaven, J. O. (2020). Induced seismicity in the Delaware Basin, Texas. *Journal of Geophysical Research: Solid Earth*, 125(1), e2019JB018558. <https://doi.org/10.1029/2019JB018558>
- Skoumal, R. J., Brudzinski, M. R., & Currie, B. S. (2018). Proximity of Precambrian basement affects the likelihood of induced seismicity in the Appalachian, Illinois, and Williston Basins, central and eastern United States. *Geosphere*, 14(3), 1365–1379. <https://doi.org/10.1130/GES01542.1>
- Waldhauser, F. (2000). A double-difference earthquake location algorithm: Method and application to the northern Hayward fault, California. *Bulletin of the Seismological Society of America*, 90(6), 1353–1368. <https://doi.org/10.1785/0120000006>
- Wang, B., Kao, H., Dokht, R. M. H., Visser, R., & Yu, H. (2022). Delineating the controlling factors of hydraulic fracturing-induced seismicity in the northern Montney play, northeastern British Columbia, Canada, with machine learning. *Seismological Research Letters*, 93(5), 2439–2450. <https://doi.org/10.1785/0220220075>
- Wessel, P., Luis, J. F., Uieda, L., Scharroo, R., Wobbe, F., Smith, W. H. F., & Tian, D. (2019). The generic mapping tools version 6. *Geochemistry, Geophysics, Geosystems*, 20(11), 5556–5564. <https://doi.org/10.1029/2019GC008515>
- Wozniakowska, P., Eaton, D., Deblonde, C., Mort, A., & Ardakani, O. (2021). Identification of regional structural corridors in the Montney play using trend surface analysis combined with geophysical imaging. *https://doi.org/10.4095/328850*
- Wozniakowska, P., & Eaton, D. W. (2020). Machine learning-based analysis of geological susceptibility to induced seismicity in the Montney formation, Canada. *Geophysical Research Letters*, 47(22), e2020GL089651. <https://doi.org/10.1029/2020GL089651>
- Zhang, M., Ellsworth, W. L., & Beroza, G. C. (2019). Rapid earthquake association and location. *Seismological Research Letters*, 90(6), 2276–2284. <https://doi.org/10.1785/0220190052>

References From the Supporting Information

- Aki, K. (1965). Maximum likelihood estimate of b in the formula $\log N = a - bM$ and its confidence limits. *Bulletin of the Earthquake Research Institute, Tokyo University*, 43, 237–239.
- Gibbons, S. J., & Ringdal, F. (2006). The detection of low magnitude seismic events using array-based waveform correlation. *Geophysical Journal International*, 165(1), 149–166. <https://doi.org/10.1111/j.1365-246X.2006.02865.x>
- Kokoska, S., & Zwillinger, D. (2000). *CRC standard probability and statistics tables and formulae*. CRC Press. <https://doi.org/10.1201/b16923>
- Nembrini, S., König, I. R., & Wright, M. N. (2018). The revival of the Gini importance? *Bioinformatics*, 34(21), 3711–3718. <https://doi.org/10.1093/bioinformatics/bty373>
- Schultz, R., Ellsworth, W. L., & Beroza, G. C. (2022). Statistical bounds on how induced seismicity stops. *Scientific Reports*, 12(1), 1184. <https://doi.org/10.1038/s41598-022-05216-9>
- Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding machine learning: From theory to algorithms* (1st ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9781107298019>
- Strobl, C., Boulesteix, A.-L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, 8(1), 25. <https://doi.org/10.1186/1471-2105-8-25>

- 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>
- Wright, G., McMechan, M., & Potter, D. (1994). Structure and architecture of the western Canada Sedimentary Basin. In G. D. Mossop & I. Shetsen (Eds.), *Geological atlas of the estern Canada Sedimentary Basin* (pp. 25–40). Canadian Society of Petroleum Geologists and Alberta Research Council. Retrieved from https://cspg.org/common/Uploaded%20files/pdfs/documents/publications/atlas/geological/atlas_03_structure_and_architecture_of_the_WCSB.pdf